Lecture Notes in Electrical Engineering 336

Zhidong Deng
Hongbo Li
Editors

Proceedings of the 2015 Chinese Intelligent Automation Conference

Intelligent Information Processing

Springer

Lecture Notes in Electrical Engineering

Volume 336

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Zhidong Deng · Hongbo Li
Editors

Proceedings of the 2015 Chinese Intelligent Automation Conference

Intelligent Information Processing

Springer

Editors
Zhidong Deng
Tsinghua University
Beijing
China

Hongbo Li
Department of Computer Science
and Technology
Tsinghua University
Beijing
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-3-662-46468-7 ISBN 978-3-662-46469-4 (eBook)
DOI 10.1007/978-3-662-46469-4

Library of Congress Control Number: 2015932968

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Information Theory and Its Relation to Machine Learning

Bao-Gang Hu

Abstract In this position paper, I first describe a new perspective on machine learning (**ML**) by four basic problems (or levels), namely "What to learn?", "How to learn?", "What to evaluate?", and "What to adjust?". The paper stresses more on the first level of "What to learn?", or "Learning Target Selection". Toward this primary problem within the four levels, I briefly review the existing studies about the connection between information theoretical learning (**ITL** [1]) and machine learning. A theorem is given on the relation between the empirically-defined similarity measure and information measures. Finally, a conjecture is proposed for pursuing a unified mathematical interpretation to learning target selection.

Keywords Machine learning · Learning target selection · Entropy · Information theory · Similarity · Conjecture

> *From the Tao comes one, from one comes two, from two comes three, and from three comes all things.* [2].
> —by Lao Tzu (ca. 600–500 BCE).
> *Nature is the realization of the simplest conceivable mathematical ideas.* [3].
> —by Albert Einstein (1879–1955).

1.1 Introduction

Machine learning learning is the study and construction of systems that can learn from data. The systems are called *learning machines*. When Big Data emerges increasingly, more learning machines are developed and applied in different domains. However, the ultimate goal of machine learning study is *insight*, not machine itself. By the term insight I mean *learning mechanisms* in descriptions of

B.-G. Hu (✉)
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: hubg@nlpr.ia.ac.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_1

mathematical principles. In a loose sense, learning mechanisms can be regarded as the natural entity. As the "*Tao* (道)" reflects the most fundamental of the universe by Lao Tzu (老子), Einstein suggests that we should pursue the simplest mathematical interpretations to the nature. Although learning mechanisms are related to the subjects of psychology, cognitive, and brain science, this paper stresses on the exploration of mathematical principles for interpretation of learning mechanisms. Up to now, we human beings are still far away from deep understanding ourself of learning mechanisms in terms of mathematical principles. It is the author's belief that "mathematical-principle-based machine" might be more important and critical than "brain-inspired machine" in the study of machine learning.

The purpose of this position paper is to put forward a new perspective and a novel conjecture within the study of machine learning. In what follows I will present four basic problems (or levels) in machine learning. The study on information theoretical learning is briefly reviewed. A theorem between the empirically defined similarity measures and information measures are given. Based on the existing investigations, a conjecture is proposed in this paper.

1.2 Four Basic Problems (or Levels) in Machine Learning

For information processing by a machine, in the 1980s, Marr [4] proposed a novel methodology by three distinct yet complementary levels, namely "Computational theory", "Representation and algorithm", and "Hardware implementation", respectively. Although the three levels are "coupled" loosely, the distinction is of great necessity to isolate and solve problems properly and efficiently. In 2007, Poggio [5] described another set of three levels on learning, namely "Learning theory and algorithms", "Engineering applications", and "Neuroscience: models and experiments", respectively. Apart from showing a new perspective, one of the important contributions of this methodology is adding a *closed loop* between the levels. These studies are enlightening because they show that complex objects or systems should be addressed by decompositions with different, yet basic, problems. The methodology is considered to be *reductionism* philosophically.

In this paper, I propose a novel perspective on machine learning by four levels shown in Fig. 1.1. The levels correspond to four basic problems. The definition of each level is given below.

Definition 1.1 "What to learn" is a study on identifying learning target(s) to the given problem(s), which will generally involve two distinct sets of representations (Fig. 1.2) defined below.

Definition 1.2 "Linguistic representation" reflects a *high-level* description in a *natural language* about the expected learning information. This study is more related to linguistics, psychology, and cognitive science.

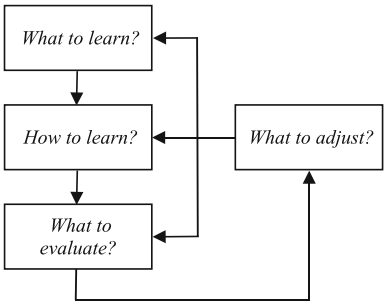


Fig. 1.1 Four basic problems (or levels) in machine learning

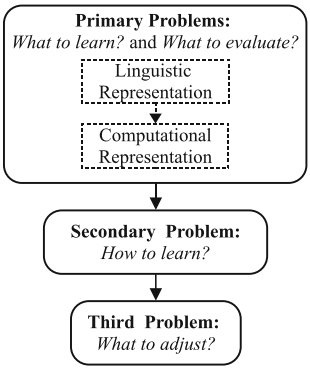


Fig. 1.2 Design flow according to the basic problems in machine learning

Definition 1.3 "Computational representation" is to define the expected learning information based on *mathematical notations*. It is a relatively *low-level* representation which generally includes objective functions, constraints, and optimization formations.

Definition 1.4 "How to learn?" is a study on learning process design and implementations. Probability, statistics, utility, optimization, and computational theories will be the central subjects. The main concerns are generalization performance, robustness, model complexity, computational complexity/cost, etc. The study may include physically realized system(s).

Definition 1.5 "What to evaluate?" is a study on "evaluation measure selection" where evaluation measure is a mathematical function. This function can be the same or different with the objective function defined in the first level.

Definition 1.6 "What to adjust?" is a study on dynamic behaviors of a machine from adjusting its component(s). This level will enable a machine with a functionality of "evolution of intelligence".

The first level is also called "learning target selection". The four levels above are neither mutually exclusive, nor collectively exhaustive to every problems in machine learning. We call them *basic* so that the extra problems can be merged

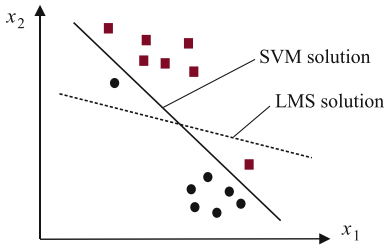


Fig. 1.3 Learning target selection within linearly separated dataset. (after [8] in Figs. 5–17). *Black Circle* Class 1, *Ruby Square* Class 2

within one of levels. Figures 1.1 and 1.2 illustrate the relations between each level in different contexts, respectively. The problems within four levels are all interrelated, particularly for "What to learn?" and "What to evaluate?" (Fig. 1.2). "How to learn?" may influence to "What to learn?", such as convexity of the objective function or scalability to learning algorithms [6] from a computational cost consideration. Structurally, "What to adjust?" level is applied to provide the multiple closed loops for describing the interrelations (Fig. 1.1). Artificial intelligence will play a critical role via this level. In the "knowledge driven and data driven" model [7], the benefits of utilizing this level are shown from the given examples by *removable singularity hypothesis* to "Sinc" function and *prior updating* to Mackey-Glass dataset, respectively. Philosophically, "What to adjust?" level remedies the intrinsic problems in the methodology of *reductionism* and offers the functionality power for being *holism*. However, this level receives even less attention while learning process holds a *self-organization* property.

I expect that the four levels show a novel perspective about the basic problems in machine learning. Take an example shown in Fig. 1.3 (after Duda et al. [8], Figs. 5–17). Even for the linearly separable dataset, the learning function using least mean square (**LMS**) does not guarantee a "minimum-error" classification. This example demonstrates two points. First, the computational representation of LMS is not compatible with the linguistic representation of "minimum-error" classification. Second, whenever a learning target is wrong in the computational representation, one is unable to reach the goal from Levels 2 and 3. Another example in Fig. 1.4 shows why we need two sub-levels in learning target selection. For the given character (here is *Albert Einstein*), one does need a linguistic representation to describe "(un)likeness" [9] between the original image and caricature image. Only when a linguistic representation is well-defined, is a computational measure of similarity *possibly proper* in caricature learning. The meaning of *possibly proper* is due to the difficulty in the following definition.

Definition 1.5 "Semantic gap" is a difference between the two sets of representations. The gap can be linked by two ways, namely a *direct way* for describing a connection from linguistic representation to computational representation, and an *inverse way* for a connection opposite to the direct one.

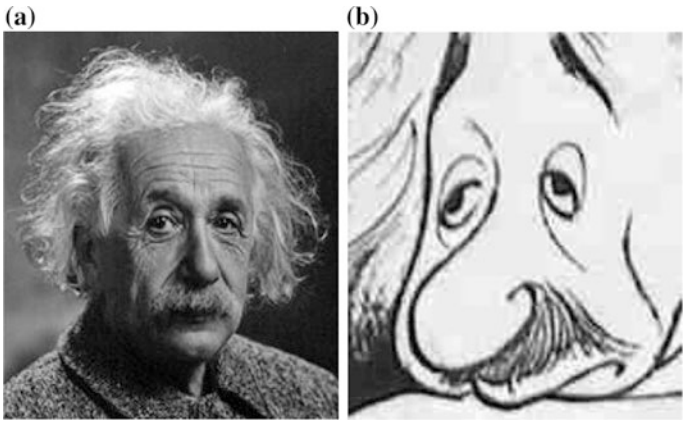


Fig. 1.4 Example of "What to learn?" and a need of defining a linguistic representation of similarity for the given character. **a** Original image (http://en.wikipedia.org/wiki/Albert_Einstein). **b** Caricature image drawn by A. Hirschfeld (http://www.georgejgoodstadt.com/goodstadt/hirschfeld.dca)

In this paper, I extend the definition of the gap in [10] by distinguishing two ways. The gap reflects one of the critical difficulties in machine learning. For the direct-way study, the difficulty source mostly comes from *ambiguity* and *subjectivity* of the linguistic representation (say, on *mental* entity), which will lead to an *ill-defined* problem. While sharing the same problem, an inverse-way study will introduce an extra challenge called *ill-posed* problem, in which there is no *unique* solution (say, from a 2D image to 3D objects).

Up to now, we have missed much studies on *learning target selection* if comparing with a study of *feature selection*. When "What to learn?" is the most primary problem in machine learning, we do need a systematic, or comparative, study on this subject. The investigations from [11, 12] into *discriminative* and *generative* models confirm the importance of learning target selection in the vein of computational representation. From the investigations, one can identify the advantages and disadvantages of each model for applications. A better machine gaining the benefits from both models is developed [13]. Furthermore, the subject of "What to learn?" will provide a strong driving force to machine learning study in seeking "the fundamental laws that govern all learning processes" [14].

Take a decision rule about "Less costs more"[1] for example. Generally, Chinese people classify object's values according to this rule. In *Big Data* processing, the useful information, which often belongs to a *minority class*, is extracted from massive datasets. While an English idiom describes it as "Finding a needle in a

[1] This rule is translated from Chinese saying, "Suddenly" in Pinyin "Wu Yi Xi Wei Gui". The translation is modified from the English phase "Less is more" which usually describes *simplicity* in design.

haystack", the Chinese saying refers to "Searching a needle in a sea (Needle in a haystack)". Users may consider that an error from a *minority class* will cost heavier than that from a *majority class* in their searching practices. This consideration will derive a decision rule like "Less costs more". The rule will be one of the important strategies in Big Data processing. Two questions can be given to the example. What is the *mathematical principle* (or *fundamental law*) for supporting the decision rule of "Less costs more"? Is it a Bayesian rule? Machine learning study does need to answer the questions.

1.3 Information Theoretical Learning

Shannon introduced "entropy" concept as the basis of information theory [15]:

$$H(Y) = -\sum_{y} p(y) \log_2 p(y), \tag{1.1}$$

where Y is a discrete random variable with *probability mass function* $p(y)$. Entropy is an expression of *disorder* to the information. From this basic concept, the other *information measures* (or *entropy functions*) can be formed (Table 1.1), where $p(t, y)$ is the *joint distribution* for the target random variable T and prediction random variable Y, and $p(t)$ and $p(y)$ are called *marginal distributions*. We call them *measures* because some of them do not satisfy the *metric* properties fully, like *KL divergence* (asymmetric). Other measures from information theory can be listed as learning criteria, but the measures in Table 1.1 are more common and sufficiently meaningful for the present discussion.

We can divide the learning machines, in view of "mathematical principles", within two groups. One group is designed based on the empirical formulas, like *error rate* or *bound*, *cost* (or *risk*), *utility*, or *classification margins*. The other is on information theory [1, 16]. Therefore, a systematic study seems necessary to answer the two basic questions below [17]:

Table 1.1 Some information formulas and their properties as learning measures

Name	Formula	(Dis)similarity	(A)symmetry
Joint information	$H(T, Y) = -\sum_{t}\sum_{y} p(t, y) \log_2 p(t, y)$	Inapplicable	Symmetry
Mutual information	$I(T, Y) = \sum_{t}\sum_{y} p(t, y) \log_2 \frac{p(t,y)}{p(t)p(y)}$	Similarity	Symmetry
Conditional entropy	$H(Y\|T) = -\sum_{t}\sum_{y} p(t, y) \log_2 p(y\|t)$	Dissimilarity	Asymmetry
Cross entropy	$H(T; Y) = -\sum_{z} p_t(z) \log_2 p_y(z)$	Dissimilarity	Asymmetry
KL divergence	$KL(T, Y) = \sum_{z} p_t(z) \log_2 \frac{p_t(z)}{p_y(z)}$	Dissimilarity	Asymmetry

Q1. When one of the principal tasks in machine learning is to process data, can we apply entropy or information measures as a generic learning target for dealing with uncertainty of data in machine learning?
Q2. What are the relations between information learning criteria and empirical learning criteria, and the advantages and limitations in using information learning criteria?

Regarding the first question, Watanabe [18, 19] proposed that "learning is an entropy-decreasing process" and pattern recognition is "a quest for minimum entropy". The principle behind entropy criteria is to transform disordered data into ordered one (or *pattern*). Watanabe seems to be the first "to cast the problems of learning in terms of minimizing properly defined entropy functions" [20], and throws brilliant light on the learning target selection in machine learning.

In 1988, Zellner theoretically proved that Bayesian theorem can be derived from the optimal information processing rule [21]. This study presents a novel, yet important, finding that Bayesian theory is rooted in information and optimization concepts. Another significant contribution is given by Principe and his collaborators [1, 22] for the proposal of Information Theoretical Learning (**ITL**) as a generic learning target in machine learning. We consider ITL will stimulate us to develop new learning machines as well as "theoretical interpretations" of learning mechanisms. Take again the example of the decision rule about "Less costs more". Hu [23] demonstrates theoretically that Bayesian principle is unable to support the rule. When a minority class approximates to a zero population, Bayesian classifiers will tend to misclassify the minority class completely. The numerical studies [23, 24] show that *mutual information* provides positive examples to the rule. The classifiers based on *mutual information* are able to protect a minority class and automatically balance the error types and reject types in terms of population ratios of classes. These studies reveal a possible mathematical interpretation of learning mechanism behind the rule.

1.4 (Dis)similarity Measures in Machine Learning

When mutual information describes similarity between two variables, the other information measures in Table 1.1 are applied in a sense of dissimilarity. For better understanding of them, their graphic relations are shown in Fig. 1.5. If we consider the variable T provides a ground truth statistically (that is, $p(t) = (p_1, \ldots, p_m)$ with the *population rate* $p_i (i = 1, \ldots, m)$ is known and fixed), its entropy $H(T)$ will be the baseline in learning. In other words, when the following relations hold:

$$\begin{aligned} I(T,Y) = H(T;Y) = H(Y;T) = H(Y) = H(T), \text{or} \\ KL(T,Y) = KL(Y,T) = H(T|Y) = H(Y|T) = 0, \end{aligned} \tag{1.2}$$

we call the measures *reach* the baseline of $H(T)$.

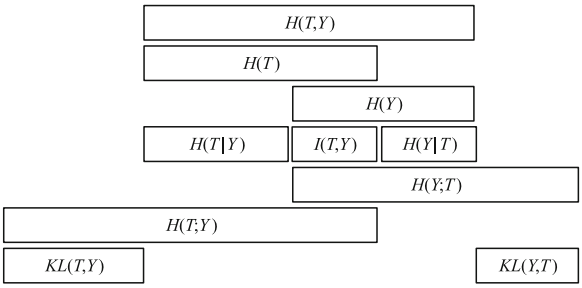


Fig. 1.5 Graphic relations among joint information, mutual information, marginal information, conditional entropy, cross entropy, and *KL* divergences (modified based on [25] by including cross entropy and *KL* divergences)

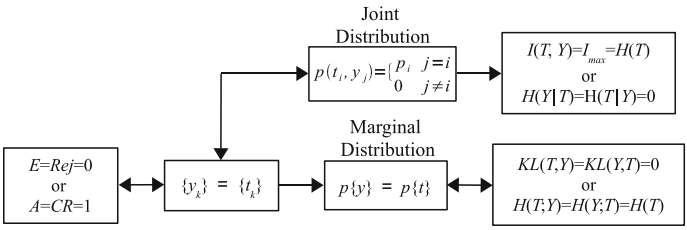


Fig. 1.6 Relations between exact classifications and mutual information, conditional entropy, cross entropy, and *KL* divergences

Based on the study in [26], further relations are illustrated in Fig. 1.6 between exact classifications and the information measures. We apply the notations of *E*, *Rej*, *A*, *CR* for the *error, reject, accuracy, and correct recognition rates*, respectively. Their relations are given by:

$$\begin{aligned} CR + E + Rej &= 1, \\ A &= \frac{CR}{CR + E}. \end{aligned} \tag{1.3}$$

The form of $\{y_k\} = \{t_k\}$ in Fig. 1.6 describes an equality between the label variables in every sample. For a finite dataset, the empirical forms should be used for representing the distributions and measures [26]. Note that the link using "↔" indicates a *two-way* connection for equivalent relations, and "→" for a *one-way* connection. Three important aspects can be observed from Fig. 1.6:

I. The *necessary* condition of exact classifications is that all the information measures reach the baseline of $H(T)$.
II. When an information measure reaches the baseline of $H(T)$, it does not *sufficiently* indicate an exact classification.
III. The different locations of one-way connections result in the interpretations *why* and *where* the sufficient condition exists.

Although Fig. 1.6 only shows the relations to the information measures listed in Table 1.1 for the classification problems, its observations may extend to other information measures as well as to the other problems, like clustering, feature selection/extraction, image registrations, etc. When we consider machine learning or pattern recognition to be a process of data in a *similarity* sense (any dissimilarity measure can be transformed into similarity one [26]), one important theorem exists to describe their relations.

Theorem 1.1 *Generally, there is no one-to-one correspondence between the empirically-defined similarity measures and information measures.*

The proof is neglected in this paper, but it can be given based on the study of bounds between entropy and error (cf. [27] and references therein). The significance of Theorem 1.1 implies that an optimization of information measure may not guarantee to achieve an optimization of the empirically defined similarity measure.

1.5 Final Remarks

Machine learning can be exploited with different perspectives depending on the study goals of researchers. For in-depth understanding of the learning mechanisms mathematically, we can take learning machines as *human's extended sensory perception*. This paper stresses on identifying the primary problem in machine learning from a novel perspective. I define it as "What to learn?" or "learning target selection". Furthermore, two sets of representations are specified, namely "linguistic representation" and "computational representation". While a wide variety of computational representations have been reported in learning targets, we can argue if there exists a unified, yet fundamental, principle behind them. Towards this purpose, this paper extends the Watanabe's proposal [18, 19] and the studies from Zellner [21] and Principe [1] to a "conjecture of learning target selection" in the following descriptions.

Conjecture 1.1 *In a machine learning study, all computational representations of learning target(s) can be interpreted, or described, by optimization of entropy function(s).*

I expect that the proposal of the conjecture above will provide a new driving force not only for seeking fundamental laws governing all learning processes [14] but also for developing improved learning machines [28] in various applications.

Acknowledgments This work is supported in part by NSFC (No. 61273196).

References

1. Principe JC (2010) Information theoretic learning: Renyi's entropy and kernel perspectives. Springer, New York
2. Lao T (ca. 500 BCE) Tao Te Ching "Tao Te Ching"
3. Norton JD (2000) Nature is the realization of the simplest conceivable mathematical ideas: Einstein and the canon of mathematical simplicity. Stud Hist Philos Mod Phys 31:135–170
4. Marr D (2010) Vision. A computational investigation into the human representation and processing of visual information. The MIT Press, Cambridge
5. Poggio T (2007) How the brain might work: The role of information and learning in understanding and replicating intelligence. In: Jacovitt G et al (eds) Information: science and technology for the new century. Lateran University Press, New York, pp 45–61
6. Bengio Y, LeCun Y (2007) Scaling learning algorithms towards AI. Large-Scale Kernel Mach 34:1–41
7. Hu BG, Qu HB, Wang Y, Yang SH (2009) A generalized constraint neural networks model: associating partially known relationships for nonlinear regressions. Inf Sci 179:1929–1943
8. Duda RO, Hart PE, Stork D (2001) Pattern classification, 2nd edn. Wiley, New York
9. Brennan SE (1985) Caricature generator: the dynamic exaggeration of faces by computer. Leonardo 40:392–400
10. Smeulders AW, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380
11. Rubinstein YD, Hastie T (1997) Discriminative vs informative learning. KDD 5:49–53
12. Ng A, Jordan MI (2002) On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. In: NIPS
13. Bishop CM, Lasserre J (2007) Generative or discriminative? Getting the best of both worlds. In: Bernardo JM et al (eds) Bayesian Statistics, vol 8. Oxford University Press, Oxford, pp 3–23
14. Mitchell TM (2006) The discipline of machine learning. Technical Report CMU-ML-06-108, Carnegie Mellon University
15. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
16. Yao YY (2003) Information-theoretic measures for knowledge discovery and data mining. In: Karmeshu (ed) Entropy measures, maximum entropy principle and emerging applications. Springer, Berlin, pp 115–136
17. Hu BG, Wang Y (2008) Evaluation criteria based on mutual information for classifications including rejected class. Acta Automatica Sinica 34:1396–1403
18. Watanabe S (1980) Pattern recognition as a quest for minimum entropy. Pattern Recognit 13:381–387
19. Watanabe S (1981) Pattern recognition as conceptual morphogenesis. IEEE Trans Pattern Anal Mach Intell 2:161–165
20. Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 21:660–674
21. Zellner A (1988) Optimal information processing and Bayes's theorem. Am Stat 42:278–284
22. Principe JC, Fisher JW III, Xu D (2000) Information theoretic learning. In: Haykin S (ed) Unsupervised adaptive filtering. Wiley, New York, pp 265–319
23. Hu BG (2014) What are the differences between Bayesian classifiers and mutual-information classifiers? IEEE Trans Neural Netw Learn Syst 25:249–264
24. Zhang X, Hu BG (2014) A new strategy of cost-free learning in the class imbalance problem. IEEE Trans Knowledge Data Eng 26:2872–2885
25. Mackay DJC (2003) Information theory, inference, and learning algorithms. Cambridge University Press, Cambridge
26. Hu BG, He R, Yuan XT (2012) Information-theoretic measures for objective evaluation of classifications. Acta Automatica Sinica 38:1170–1182

27. Hu BG, Xing HJ (2013) A new approach of deriving bounds between entropy and error from joint distribution: case study for binary classifications. arXiv:1205.6602v1[cs.IT]
28. He R, Hu BG, Yuan XT, Wang L (2014) Robust recognition via information theoretic learning. Springer, Heidelberg

Chapter 2
Robust Covariance Intersection Fusion Steady-State Kalman Filter with Uncertain Parameters

Wenjuan Qi, Xuemei Wang, Wenqiang Liu and Zili Deng

Abstract For the linear discrete time-invariant system with uncertain parameters and known noise variances, a robust covariance intersection (CI) fusion steady-state Kalman filter is presented by the new approach of compensating the parameter uncertainties by a fictitious noise. Based on the Lyapunov equation approach, it is proved that for the prescribed upper bound of the fictitious noise variances, there exists a sufficiently small region of uncertain parameters; such that its actual filtering error variances are guaranteed to have a less-conservative upper bound. This region is called the robust region. By the searching method, the robust region can be found. Its robust accuracy is higher than that of each local robust Kalman filter. A Monte-Carlo simulation example shows its effectiveness and the good performance.

Keywords Covariance intersection fusion · Robust Kalman filter · Uncertain parameters · Fictitious noise approach · Robust region

2.1 Introduction

Multisensor information fusion Kalman filtering has been applied to many fields, such as signal processing, data fusion, and target tracking. For Kalman filtering fusion, there are two basic fusion methods: The centralized and distributed fusion methods. For the distributed fusion method, the three-weighted state fusion approaches weighted by matrices, diagonal matrices, and scalars have been presented. In order to compute the weights, the cross-covariances among the local filtering errors are required. However, in many practical applications, the computation of the cross-covariance is very difficult [1]. In order to overcome this limitation, the covariance intersection fusion algorithm has been presented [2].

W. Qi · X. Wang · W. Liu · Z. Deng (✉)
Department of Automation, Heilongjiang University, 150080 Harbin, China
e-mail: dzl@hlju.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_2

In this paper, a robust CI fusion steady-state Kalman filter is presented for system with uncertain parameters and known noise variances. Two important approaches used to develop the robust Kalman filter are the Riccati equation approach [3] and the linear matrix inequality (LMI) approach [4]. More research references on this topic are using these two approaches; however, in this paper, a new approach is presented by compensating the uncertain parameters by a fictitious noise which converts the system with uncertain parameters into the system with noise variance uncertainties [5].

This paper extends the robust CI fusion Kalman filter with uncertain noise variances [5] to the robust CI fusion Kalman filter with uncertain parameters. Compared with the suboptimal Kalman filter without fictitious noise, the proposed robust Kalman filter can significantly improve the filtering performance, and its robust accuracy is higher than that of each local robust Kalman filter.

2.2 Local Robust Steady-State Kalman Filter

Consider the multisensor uncertain system with model parameters uncertainties

$$x(t+1) = (\Phi_e + \Delta\Phi)x(t) + \Gamma w(t) \tag{2.1}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, \ldots, L \tag{2.2}$$

where t is the discrete time, $x(t) \in R^n$ is the state to be estimated, $y_i(t) \in R^{m_i}$ is the measurement of the ith subsystem, $w(t) \in R^r$, $v_i(t) \in R^{m_i}$ are uncorrelated white noises with zero means and known variances Q and R_i, respectively. Φ_e, Γ, and H_i are known constant matrices with appropriate dimensions. L is the number of sensors. $\Phi = \Phi_e + \Delta\Phi$ is uncertain transition matrix, $\Delta\Phi$ is the uncertain parameter disturbance. Assume that Φ and Φ_e are stable matrices.

$\xi(t)$ is a uncertain fictitious white noise with zeros mean and upper-bound variance $\Delta_\xi > 0$, which is used to compensate the uncertain model parameter error term $\Delta\Phi x(t)$ in (2.1), so that the systems (2.1) and (2.2) with uncertain model parameters can be converted into the following worst-case conservative system with known model parameters and noise variances Q, R_i, and Δ_ξ.

$$x_e(t+1) = \Phi_e x_e(t) + w_e(t),\ w_e(t) = \Gamma w(t) + \xi(t) \tag{2.3}$$

$$y_{ei}(t) = H_i x_e(t) + v_i(t),\ i = 1, \ldots, L \tag{2.4}$$

Assume that each conservative subsystem is completely observable and completely controllable. The conservative local steady-state optimal Kalman filters are given as

$$\hat{x}_{ei}(t|t) = \Psi_i \hat{x}_{ei}(t-1|t-1) + K_i y_{ei}(t) \tag{2.5}$$

$$\Psi_i = [I_n - K_i H_i]\Phi_e,\ K_i = \Sigma_i H_i^{\mathrm{T}}\left(H_i \Sigma_i H_i^{\mathrm{T}} + R_i\right)^{-1},\ P_i = [I_n - K_i H_i]\Sigma_i \tag{2.6}$$

where I_n is an $n \times n$ identity matrix, Ψ_i is a stable matrix, and Σ_i satisfies the steady-state Riccati equation

$$\Sigma_i = \Phi_e\left[\Sigma_i - \Sigma_i H_i^{\mathrm{T}}(H_i \Sigma_i H_i^{\mathrm{T}} + R_i)^{-1} H_i \Sigma_i\right]\Phi_e^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} + \Delta_\xi \tag{2.7}$$

where the symbol T denotes the transpose. Define $\tilde{x}_{ei}(t|t) = x_e(t) - \hat{x}_{ei}(t|t)$, applying (2.3) and (2.5), we have

$$\begin{aligned}\tilde{x}_{ei}(t|t) = {} & \Psi_i \tilde{x}_{ei}(t-1|t-1) + [I_n - K_i H_i]\Gamma w(t-1) \\ & + [I_n - K_i H_i]\xi(t-1) - K_i v_i(t)\end{aligned} \tag{2.8}$$

Applying (2.8) yields that the conservative local filtering error variances P_i and cross-covariance P_{ij} satisfy the conservative Laypunov equation

$$\begin{aligned}P_{ij} = {} & \Psi_i P_{ij} \Psi_j^{\mathrm{T}} + [\mathrm{I}_n - \mathrm{K}_i \mathrm{H}_i]\Gamma Q \Gamma^{\mathrm{T}}\left[I_n - K_j H_j\right]^{\mathrm{T}} \\ & + [\mathrm{I}_n - \mathrm{K}_i \mathrm{H}_i]\Delta_\xi\left[I_n - K_j H_j\right]^{\mathrm{T}} + K_i R_{ij} K_j^{\mathrm{T}} \delta_{ij}, \quad i,j = 1,\ldots,L\end{aligned} \tag{2.9}$$

where δ_{ij} is the Kronecker δ function, $\delta_{ii} = 1$, $\delta_{ij} = 0$ ($\mathrm{i} \neq \mathrm{j}$).

Remark 2.1 Notice that in (2.5), the conservative measurements $y_{ei}(t)$ are unavailable, only the actual measurements $y_i(t)$ are known. Therefore, replacing the conservative measurements $y_{ei}(t)$ with the known actual measurements $y_i(t)$, we obtain the actual local Kalman filters as

$$\hat{x}_i(t|t) = \Psi_i \hat{x}_i(t-1|t-1) + K_i y_i(t) \tag{2.10}$$

From (2.10) and (2.11) we have

$$\begin{aligned}\tilde{x}_i(t|t) = {} & \Psi_i \tilde{x}_i(t-1|t-1) + [I_n - K_i H_i]\Delta\Phi x(t-1) \\ & + [I_n - K_i H_i]\Gamma w(t-1) - K_i v_i(t)\end{aligned} \tag{2.11}$$

So the actual local filtering error variances and cross-covariance are given as

$$\begin{aligned}\bar{P}_{ij} = {} & \Psi_i \bar{P}_{ij} \Psi_j^{\mathrm{T}} + [\mathrm{I}_n - \mathrm{K}_i \mathrm{H}_i]\Delta\Phi X \Delta\Phi^{\mathrm{T}}\left[I_n - K_j H_j\right]^{\mathrm{T}} \\ & + [\mathrm{I}_n - \mathrm{K}_i \mathrm{H}_i]\Gamma Q \Gamma^{\mathrm{T}}\left[I_n - K_j H_j\right]^{\mathrm{T}} + [\mathrm{I}_n - \mathrm{K}_i \mathrm{H}_i]\Delta\Phi C_j \Psi_j^{\mathrm{T}} \\ & + \Psi_i C_i^{\mathrm{T}} \Delta\Phi^{\mathrm{T}}\left[I_n - K_j H_j\right]^{\mathrm{T}} + K_i R_{ij} K_j^{\mathrm{T}} \delta_{ij}\end{aligned} \tag{2.12}$$

where $X = \mathrm{E}\left[x(t)x^{\mathrm{T}}(t)\right], C_i = \mathrm{E}\left[x(t)\tilde{x}_i^{\mathrm{T}}(t|t)\right]$. From (2.1), X satisfies the following Lyapunov equation

$$\mathrm{X} = \Phi \mathrm{X} \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} \tag{2.13}$$

Applying (2.1) and (2.11), we have the Lyapunov equation

$$C_i = \Phi C_i \Psi_i^{\mathrm{T}} + \Phi X \Delta\Phi^{\mathrm{T}}[I_n - K_i H_i]^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}}[I_n - K_i H_i]^{\mathrm{T}} \tag{2.14}$$

Lemma 2.1 [6] Consider the Lyapunov equation with U to be a symmetric matrix

$$P = FPF^{\mathrm{T}} + U \tag{2.15}$$

If the matrix F is stable (all its eigenvalues are inside the unit circle) and U is positive (semi)definite, then the solution P is unique, symmetric, and positive (semi-)definite.

Theorem 2.1 *For uncertain systems* (2.1) *and* (2.2) *with uncertain parameters, the actual local steady-state Kalman filter* (2.10) *is robust in the sense that there exists a region* $\Re_{\Delta\Phi}^{(i)}$*, such that for all admissible uncertain model parameter* $\Delta\Phi \in \Re_{\Delta\Phi}^{(i)}$*, the corresponding actual filtering variances* $\bar{P}_i$ *have the upper-bound* P_i*, i.e.,*

$$\bar{P}_i < P_i \tag{2.16}$$

and $\Re_{\Delta\Phi}^{(i)}$ *is called the robust region of the local robust Kalman filter* (2.10).

Proof Define $\Delta P_i = P_i - \bar{P}_i$, subtracting (2.12) from (2.9) yields

$$\Delta P_i = \Psi_i \Delta P_i \Psi_i^{\mathrm{T}} + U_i(\Delta\Phi) \tag{2.17}$$

$$\begin{aligned} U_i(\Delta\Phi) = {} & [\mathrm{I}_n - \mathrm{K}_i\mathrm{H}_i]\Delta_\xi[I_n - K_i H_i]^{\mathrm{T}} - [\mathrm{I}_n - \mathrm{K}_i\mathrm{H}_i]\Delta\Phi X \Delta\Phi^{\mathrm{T}}[I_n - K_i H_i]^{\mathrm{T}} \\ & - [\mathrm{I}_n - \mathrm{K}_i\mathrm{H}_i]\Delta\Phi C_i \Psi_i^{\mathrm{T}} - \Psi_i C_i^{\mathrm{T}} \Delta\Phi^{\mathrm{T}}[I_n - K_i H_i]^{\mathrm{T}} \end{aligned} \tag{2.18}$$

From (2.6) yields $I_n - K_i H_i = P_i \Sigma_i^{-1}$, so we have $\det[I_n - K_i H_i] = \det P_i \det \Sigma_i^{-1} \neq 0, [I_n - K_i H_i]$ is invertible. Since $\Delta_\xi > 0$, then $U_{0i} = [\mathrm{I}_n - \mathrm{K}_i\mathrm{H}_i]\Delta_\xi[I_n - K_i H_i]^{\mathrm{T}} > 0$. According to the property of the continuous function, as $\Delta\Phi \to 0$, we have $U_i(\Delta\Phi) \to U_{0i} > 0$. Hence there exists a sufficiently small region $\Re_{\Delta\Phi}^{(i)}$, such that for all admissible $\Delta\Phi \in \Re_{\Delta\Phi}^{(i)}$, we have $U_i(\Delta\Phi) > 0$. Applying Lemma 2.1 yields $\Delta P_i > 0$, i.e., (2.16) holds, and $\Re_{\Delta\Phi}^{(i)}$ is called the robust region of uncertain parameters for the local robust Kalman filter (2.10). The proof is completed.

2.3 Robust CI Fusion Steady-State Kalman Filter

For multisensor uncertain time-invariant systems (2.1) and (2.2), the robust steady-state CI-fused Kalman filter is presented as

$$\hat{x}_{CI}(t|t) = P_{CI}\sum_{i=1}^{L}\omega_i P_i^{-1}\hat{x}_i(t|t) \tag{2.19}$$

$$P_{CI} = \left[\sum_{i=1}^{L}\omega_i P_i^{-1}\right]^{-1},\ \sum_{i=1}^{L}\omega_i = 1,\ \omega_i \geq 0 \tag{2.20}$$

The optimal weighting coefficientsω_i are obtained by minimizing the performance index

$$J = \min_{\omega_i}\mathrm{tr}P_{CI} = \min_{\substack{\omega_i \in [0,1]\\ \omega_1+\cdots+\omega_L=1}}\mathrm{tr}\left\{\left[\sum_{i=1}^{L}\omega_i P_i^{-1}\right]^{-1}\right\} \tag{2.21}$$

The actual error variances are given as

$$\bar{P}_{CI} = P_{CI}\left[\sum_{i=1}^{L}\sum_{j=1}^{L}\omega_i P_i^{-1}\bar{P}_{ij}P_j^{-1}\omega_j\right]P_{CI} \tag{2.22}$$

It is proved that [7] the local robustness (2.16) yields the robustness of the CI fuser for all $\Delta\Phi \in \Re_{\Delta\Phi}^{CI} = \bigcap_{i=1}^{L}\Re_{\Delta\Phi}^{(i)}$

$$\bar{P}_{CI} \leq P_{CI} \tag{2.23}$$

where the symbol $\cap$ denotes the intersection of sets.

Theorem 2.2 [8] *The local and CI fusion robust Kalman filters have the following robust accuracy relations*

$$\mathrm{tr}\bar{P}_i < \mathrm{tr}P_i, i = 1,\ldots,L \tag{2.24}$$

$$\mathrm{tr}\bar{P}_{CI} \leq \mathrm{tr}P_{CI} \leq \mathrm{tr}P_i, i = 1,\ldots,L \tag{2.25}$$

Remark 2.2 Taking the trace operation for (2.16), we have $\mathrm{tr}\bar{P}_i < \mathrm{tr}P_i, i = 1,\ldots,L$. The trace $\mathrm{tr}P_i$ is called robust accuracy or global accuracy of a robust Kalman filter, the trace $\mathrm{tr}\bar{P}_i$ is called its actual accuracy. Theorem 2.1 shows that the robust accuracy of the CI fusion Kalman filter is higher than that of each local robust Kalman filter. The actual accuracy of the local or CI fuser is higher that its robust accuracy.

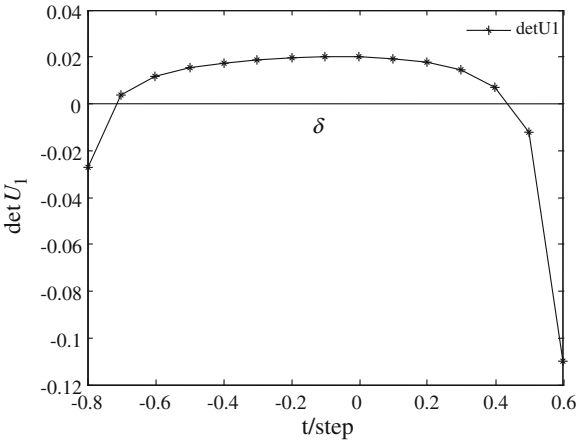


Fig. 2.1 The robust region of the local robust Kalman filter $\hat{x}_1(t|t)$

2.4 Simulation Example

Consider a 2-sensor time-invariant system with uncertain model parameters

$$x(t+1) = (\Phi_e + \Delta\Phi)x(t) + \Gamma w(t) \tag{2.26}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, 2 \tag{2.27}$$

In the simulation, we take $\Phi_e = \begin{bmatrix} 0.43 & 0.32 \\ 0.56 & 0 \end{bmatrix}, \Delta\Phi = \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix}, \Gamma = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, H_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $H_2 = I_2, Q = 1, R_1 = 1, R_2 = \mathrm{diag}(6, 0.36)$, δ is the uncertain parameter. The simulation results are shown in the following. From Figs. 2.1 and 2.2, the necessary and sufficient condition of $U_i(\delta) > 0$ is that $\det U_i(\delta) > 0$, so we can

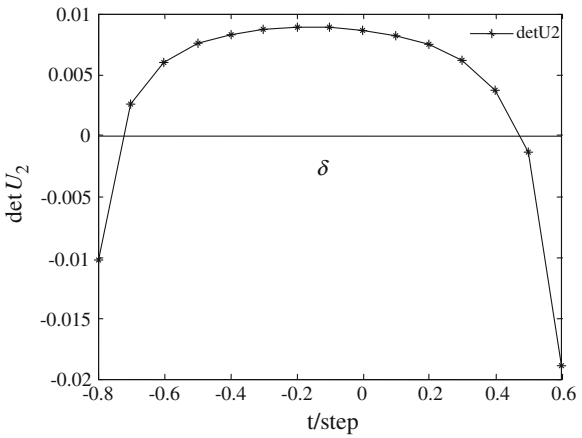


Fig. 2.2 The robust region of the local robust Kalman filter $\hat{x}_2(t|t)$

Table 2.1 The robust and actual accuracy comparison of $\mathrm{tr}P_i$ and $\mathrm{tr}\bar{P}_i$, $i = 1, 2, CI$

$\mathrm{tr}P_1(\mathrm{tr}\bar{P}_1)$	$\mathrm{tr}P_2(\mathrm{tr}\bar{P}_2)$	$\mathrm{tr}P_{CI}(\mathrm{tr}\bar{P}_{CI})$
1.1477 (0.7412)	1.4713 (1.1673)	1.1059 (0.6299)

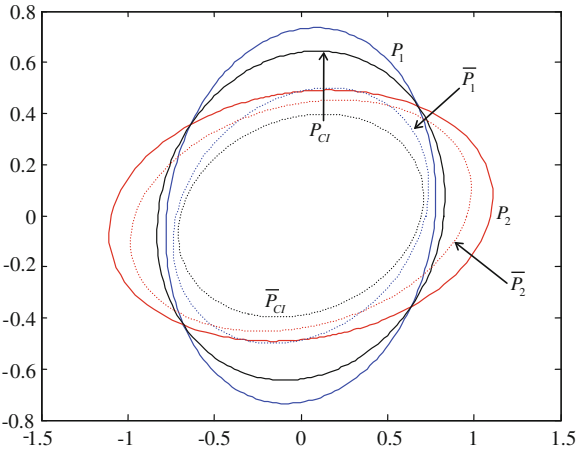


Fig. 2.3 The covariance ellipses of the local and CI fusion robust Kalman filters

obtain that the robust region of $\hat{x}_1(t|t)$ is $\Re_\delta^{(1)} : -0.72 < \delta < 0.43$, the robust region of $\hat{x}_2(t|t)$ is $\Re_\delta^{(2)} : -0.73 < \delta < 0.48$, so the robust region of CI fuser is $\Re_\delta^{CI} = \Re_\delta^{(1)} \cap \Re_\delta^{(2)} : -0.72 < \delta < 0.43$.

When $\delta = 0.2$ in the robust region, the traces comparisons of the conservative and actual filtering error variances are given in Table 2.1, which verify the accuracy relations (2.24) and (2.25).

In order to give a geometric interpretation of the matrix accuracy relations, the covariance ellipse of variance $P$is defined as the locus of points $\left\{x : x^{\mathrm{T}}P^{-1}x = c\right\}$, where P is $n \times n$ the variance matrix and $x \in R^n$ and $c$is a constant. Generally, we select $c = 1$ without loss of generality. It has been proved in [8] that $P_1 \leq P_2$ is equivalent to that the covariance ellipse of P_1 is enclosed in that of P_2.

The matrix accuracy relations are given based on the covariance ellipses as shown in Fig. 2.3. From Fig. 2.3, we see that the ellipse of $\bar{P}_i$ is enclosed in that of P_i, the ellipse of $\bar{P}_{CI}$ is enclosed in that of P_{CI}. These verify that the accuracy relations (2.16) and (2.23) hold.

In order to verify the above theoretical results for the accuracy relation, taking the Monte-Carlo simulation with 1,000 runs, the mean-square error (MSE) curves of the local and CI-fused Kalman filters are shown in Fig. 2.4; we see that the values of the $\mathrm{MSE}_i(t)$, $i = 1, 2, CI$ are close to the corresponding $\mathrm{tr}\bar{P}_i$ and the accuracy relations (2.24–2.25) hold.

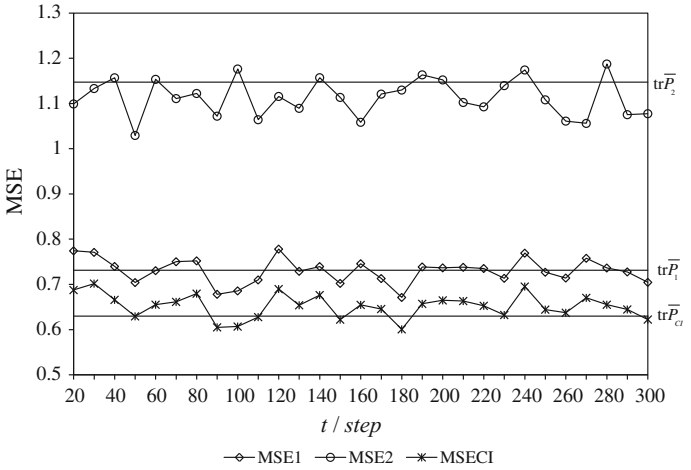


Fig. 2.4 The MSE curves of the local and CI fusion robust Kalman filters

2.5 Conclusion

For multisensor systems with uncertain parameters and known noise variances, the local and CI-fused steady-state Kalman filters are presented by the new approach of compensating the parameters uncertainties by a fictitious noise. It is proved that the local and CI-fused Kalman filters are robust for all admissible uncertain parameters in the robust region, this is, the actual filtering error variances have a less-conservative upper bound, and the robust accuracy of the CI fuser is higher than those of the local robust Kalman filters. When the fictitious noise variance is prescribed, by the searching method, the robust region can be found.

Acknowledgments This work is supported by the National Natural Science Foundation of China under Grant NSFC-60874063 and NSFC-60374026.

References

1. Sun XJ, Gao Y, Deng ZL, Li C, Wang JW (2010) Multi-model information fusion Kalman filtering and white noise deconvolution. Inf Fusion 11:163–173
2. Julier SJ, Uhlmann JK (2007) Using covariance intersection for SLAM. Robot Auton Syst 55 (1):3–20
3. Qu XM, Zhou J (2013) The optimal robust finite-horizon Kalman filtering for multiple sensors with different stochastic failure rates. Appl Math Lett 26(1):80–86
4. Yang F, Li Y (2012) Robust set-membership filtering for systems with missing measurement: a linear matrix inequality approach. IET Signal Process 6(4):341–347
5. Qi WJ, Zhang P, Feng WQ, Deng ZL (2013) Covariance intersection fusion robust steady-state Kalman filter for multisensor systems with unknown noise variances. Lect Notes Electr Eng 254:835–860

6. Kailath T, Sayed AH, Hassibi B (2000) Linear Estimation. Prentice Hall, New York
7. Deng ZL, Zhang P, Qi WJ, Gao Y, Liu JF (2013) The accuracy comparison of multisensor covariance intersection fuser and three weighting fusers. Information Fusion 14:177–185
8. Qi WJ, Zhang P, Deng ZL (2014) Robust weighted fusion Kalman filters for multisensor time-varying systems with uncertain noise variances. Signal Process 99:185–200

Chapter 3
Robust Centralized Fusion Steady-State Kalman Predictor with Uncertain Parameters

Xuemei Wang, Wenqiang Liu and Zili Deng

Abstract For multisensor time-invariant systems with uncertain parameter and known noise variances, the centralized fusion robust steady-state Kalman predictor based on the minimax robust estimation principle is presented by a new approach of compensating the parameter uncertainties by fictitious noise. Using the Lyapunov equation, it is proved that the variances of its actual prediction error variances have a conservative upper bound when the uncertainty of parameters is restricted in a sufficiently small region, which is called the robust region of the parameter uncertainties. It is also proved that the robust accuracy of the centralized fuser is higher than that of each local robust Kalman predictor. A simulation example shows how to search the robust region and shows its good performances.

Keywords Robust · Kalman predictor · Uncertain parameters · Centralized fusion · Lyapunov equation approach

3.1 Introduction

Multisensor information fusion has been applied to many fields, including military affairs, navigation, guidance, remote sensing, signal processing, target tracking. There exist two basic fusion methods: one is the centralized fusion approach [1],

X. Wang · W. Liu · Z. Deng (✉)
Department of Automation, Heilongjiang University, XueFu Road 74, Heilongjiang University, Electronic and Engineering College 130, 150080 Harbin, China
e-mail: dzl@hlju.edu.cn; dengzili889@163.com; dengzili890@163.com

X. Wang
Department of Computer Science and Technology, ChengDong College of Northeast Agricultural University, 150025 Harbin, China

W. Liu
Department of Computer and Information Engineering, Heilongjiang University of Science and Technology, 150022 Harbin, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_3

which can give a globally optimal state estimate by directly combing the local measurement equations to obtain an augmented measurement equation. The other is the distributed fusion approach [2–5], which can combine or weight the local Kalman estimators to obtain a global optimal or suboptimal state estimator.

The standard Kalman filtering is only suitable for the systems with exactly known model. For uncertain systems with the uncertainties of model parameters and/or noise variances, the performance of the Kalman filter will degrade or the filter may be divergent [6]. However, since the system model is usually an approximation to a physical situation in many applications, the research on robust Kalman filters for uncertain systems received great attention. An important class of robust Kalman filtering problems is to find a Kalman filter such that its actual filtering error variances yielded by all admissible uncertainties are guaranteed to have a minimal upper bound [7]. Such a Kalman filter is called robust Kalman filter, and such property is called robustness. There are two main approaches to solve this problem, i.e., the Riccati equation approach [8] and the linear matrix inequality (LMI) approach [9]. The limitation of the above robust Kalman filters is that only model parameters are assumed to be uncertain, while the noise variances are assumed to be exactly known.

Centralized fusion steady-state robust Kalman filter [10] for multisensor systems with uncertainty of noise variances, the local and centralized fusion robust steady-state Kalman filter are presented.

In this paper, we consider the problem of designing the local and centralized fusion robust steady-state Kalman predictors for systems with uncertain parameters and known noise variances by a fictitious noise-based compensation technique. The uncertainty of parameters is compensated by introducing a fictitious noise with upper bound variance. Further, we can obtain the robust region by the searching method. Finally, it is proved that the robust accuracy of the centralized fuser is higher than that of the local robust Kalman predictor.

3.2 Local and Centralized Fusion Robust Kalman Predictors

Consider the multisensor system with uncertain parameters

$$x(t+1) = (\Phi_e + \Delta\Phi)x(t) + \Gamma\omega(t) \tag{3.1}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, \ldots, L \tag{3.2}$$

$$\Phi = \Phi_e + \Delta\Phi \tag{3.3}$$

where t is the discrete time, $x(t) \in R^n$ is the state to be estimated, $y_i(t) \in R^{m_i}$ is the measurement of the ith subsystem, $\omega(t) \in R^r$ is the input noise, $v_i(t) \in R^{m_i}$ is the measurement noise, and they are mutually uncorrelated white noises with zero

means and known variances Q and R_i. Φ is the true transition matrix, Φ_e is a known estimate of Φ and $\Delta\Phi$ is the uncertain parameter disturbance matrix. Γ, H_i, Q and R_i are known constant matrices with appropriate dimensions.

$$\Delta\Phi \in \Re_{\Delta\Phi} \tag{3.4}$$

$\Re_{\Delta\Phi}$ can be found or prescribed.

From (3.1)

$$x(t+1) = \Phi_e x(t) + \Delta\Phi x(t) + \Gamma\omega(t) \tag{3.5}$$

Introducing a fictitious white noise $\xi(t)$ with zero mean and known upper bound variance Δ_ξ of variances, which compensates the model error term $\Delta\Phi x(t)$, then we have the worst-case conservative multisensor system

$$x(t+1) = \Phi_e x(t) + \xi(t) + \Gamma\omega(t) \tag{3.6}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, \ldots, L \tag{3.7}$$

where $\omega(t)$ and $v_i(t)$ have the known true variances Q and R_i.

The conservative centralized fused system is given as

$$x(t+1) = \Phi_e x(t) + \xi(t) + \Gamma\omega(t) \tag{3.8}$$

$$y_c(t) = H_c x(t) + v_c(t) \tag{3.9}$$

$$y_c(t) = \left[y_1^{\mathrm{T}}(t), \ldots, y_L^{\mathrm{T}}(t)\right]^{\mathrm{T}},\ H_c = \left[H_1^{\mathrm{T}}, \ldots, H_L^{\mathrm{T}}\right]^{\mathrm{T}},\ v_c(t) = \left[v_1^{\mathrm{T}}(t), \ldots, v_L^{\mathrm{T}}(t)\right]^{\mathrm{T}} \tag{3.10}$$

where the symbol T denotes the transpose. $v_c(t)$ is the conservative fused noise, and has the variance $R_c = diag(R_1, \ldots, R_L)$.

The conservative centralized fusion Kalman predictor is given as

$$\hat{x}(t+1|t) = \Psi_c \hat{x}(t|t-1) + \mathrm{K}_c y_c(t) \tag{3.11}$$

where $y_c(t)$ is conservative measurement, and

$$\Psi_c = \Phi_e - K_c H_c,\ K_c = \Phi_e \Sigma_c H_c^{\mathrm{T}} \left(H_c \Sigma_c H_c^{\mathrm{T}} + R_c\right)^{-1} \tag{3.12}$$

where Ψ_c is stable, the conservative prediction error variance Σ_c satisfies the Riccati equation

$$\Sigma_c = \Phi_e \left[\Sigma_c - \Sigma_c H_c^{\mathrm{T}} \left(H_c \Sigma_c H_c^{\mathrm{T}} + R_c\right)^{-1} H_c \Sigma_c\right] \Phi_e^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} + \Delta_\xi \tag{3.13}$$

From (3.6), (3.9), and (3.11), we easily obtain the conservative prediction error system

$$\tilde{x}(t+1|t) = \Psi_c \tilde{x}(t|t-1) + \Gamma\omega(t) + \xi(t) - K_c v_c(t) \tag{3.14}$$

where $\tilde{x}(t+1|t) = x(t+1) - \hat{x}(t+1|t)$, $x(t+1)$ is the conservative state in (3.6), $\hat{x}(t+1|t)$ is the conservative Kalman predictor in (3.11). $v_c(t)$ is the conservative fused noise with variance R_c.

This yields the conservative variance Σ_c satisfies the Lyapunov equation

$$\Sigma_c = \Psi_c \Sigma_c \Psi_c^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} + \Delta_\xi + K_c R_c K_c^{\mathrm{T}} \tag{3.15}$$

Now we find the actual prediction error

$$\tilde{x}(t+1|t) = x(t+1) - \hat{x}(t+1|t) \tag{3.16}$$

where $x(t+1)$ is the true state given by (3.1), and $\hat{x}(t+1|t)$ is the actual Kalman predictor (3.11) with $y_c(t)$ is the actual measurement, i.e., $y_c(t) = \left[y_1^{\mathrm{T}}(t), \ldots, y_L^{\mathrm{T}}(t)\right]^{\mathrm{T}}$, where $y_i(t)$ is the actual measurement, which is available, and which is yielded from (3.1) and (3.2). Hence from (3.1), (3.9) and (3.11) we obtain

$$\tilde{x}(t+1|t) = \Psi_c \tilde{x}(t|t-1) + \Delta\Phi x(t) + \Gamma\omega(t) - K_c v_c(t) \tag{3.17}$$

where $v_c(t)$ is the actual fused noise with variance $\bar{R}_c = diag(\bar{R}_1, \ldots, \bar{R}_L)$.

Thus we obtain the actual predictor error variance $\bar{\Sigma}_c$ satisfies the Lyapunov equation

$$\bar{\Sigma}_c = \Psi_c \bar{\Sigma}_c \Psi_c^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} + K_c R_c K_c^{\mathrm{T}} + \Delta\Phi X \Delta\Phi^{\mathrm{T}} + \Delta\Phi \underline{C} \Psi_c^{\mathrm{T}} + \Psi_c \underline{C}^{\mathrm{T}} \Delta\Phi^{\mathrm{T}} \tag{3.18}$$

where we defind the steady-state cross-covariance

$$\underline{C} = E\left[x(t)\tilde{x}^{\mathrm{T}}(t|t-1)\right] = E\left[x(t+1)\tilde{x}(t+1|t)^{\mathrm{T}}\right] \tag{3.19}$$

From (3.1), (3.17) and (3.19) we obtain the Lyapunov equation

$$\underline{C} = \Phi \underline{C} \Psi_C^{\mathrm{T}} + \Phi X \Delta\Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} \tag{3.20}$$

with the definition $X = E[x(t)x^{\mathrm{T}}(t)]$. From (3.1) we have

$$X = \Phi X \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} \tag{3.21}$$

$$\Phi = (\Phi_e + \Delta\Phi) \tag{3.22}$$

Theorem 3.1 *For multisensor system* (3.1, 3.2 *and* 3.3) *with uncertain parameters and known noise variances, the actual centralized fusion steady-state Kalman predictor* (3.11) *with the actual fused measurement* $y_c(t)$*, is robust in the sense that for the prescribed upper bound* $\Delta_\xi > 0$ *of fictitious noise variances, there exists a sufficiently small region* $\Re_{\Delta\Phi}$*, such that for all admissible uncertain disturbance* $\Delta\Phi \in \Re_{\Delta\Phi}$*, we have*

$$\bar{\Sigma}_c < \Sigma_c \tag{3.23}$$

which is called the robustness of robust Kalman predictor.

Proof Letting $\Delta\Sigma_c = \Sigma_c - \bar{\Sigma}_c$, from (3.15) and (3.18) we have the Lyapunov equation

$$\Delta\Sigma_c = \Psi_c \Delta\Sigma_c \Psi_c^{\mathrm{T}} + \Delta_\xi - \Delta\Phi X \Delta\Phi^{\mathrm{T}} - \Delta\Phi \underline{C} \Psi_c^{\mathrm{T}} - \Psi_c \underline{C}^{\mathrm{T}} \Delta\Phi^{\mathrm{T}} \tag{3.24}$$

Defining

$$U = \Delta_\xi - \Delta\Phi X \Delta\Phi^{\mathrm{T}} - \Delta\Phi \underline{C} \Psi_c^{\mathrm{T}} - \Psi_c \underline{C}^{\mathrm{T}} \Delta\Phi^{\mathrm{T}} \tag{3.25}$$

Since $\Delta\Phi \to 0$, $U \to \Delta_\xi > 0$, hence there exists a sufficiently small region $\Re_{\Delta\Phi}$, such that for all $\Delta\Phi \in \Re_{\Delta\Phi}$,we have

$$U > 0 \tag{3.26}$$

Form (3.24–3.26) we obtain $\Delta\Sigma_c > 0$, i.e.,

$$\bar{\Sigma}_c < \Sigma_c \tag{3.27}$$

The proof is completed. □

Remark 3.1 Similar to the derivation of the centralized fusion robust Kalman predictor, for the system (3.1–3.3), we can also obtain the corresponding local robust Kalman predictors $\hat{x}_i(t+1|t), i = 1, \ldots, L$ with the actual variance $\bar{\Sigma}_i$ and the conservative upped bounds Σ_i, and similar to the derivation of Theorem 3.1, we have the robustness $\bar{\Sigma}_i < \Sigma_i$.

Theorem 3.2 *The local and centralized robust Kalman predictors have the robust accuracy relation*

$$\Sigma_c < \Sigma_i, i = 1, \ldots, L \tag{3.28}$$

$$\mathrm{tr}\bar{\Sigma}_c < \mathrm{tr}\Sigma_c < \mathrm{tr}\Sigma_i, i = 1, \ldots, L \tag{3.29}$$

where the symbol tr *denote the trace of matrix.*

Proof For the worst-case conservative system (3.6) and (3.9), applying [10] yields (3.28) to hold. Taking the trace operations to (3.23) and (3.28) yield (3.29).The proof is completed. □

Remark 2 The trace $\mathrm{tr}\Sigma_c$ is called as robust accuracy, and the trace $\mathrm{tr}\bar{\Sigma}_c$ is called as actual accuracy. This shows that the actual accuracy of a robust Kalman predictor is higher than its robust accuracy. The robust accuracy of the centralized fuser is higher than that of each local robust Kalman predictor.

3.3 Simulation Example

Consider two-dimensional 2-sensors time-invariant system (3.1–3.3) with uncertain parameter, where $x(t) = [x_1(t), x_2(t)]^T$ is the state. In the simulation, we take $\Phi_e = \begin{bmatrix} 0.3 & -0.5 \\ 1 & 0.5 \end{bmatrix}, \Delta\Phi = \begin{bmatrix} \delta & 0 \\ 0 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} -6 \\ 1 \end{bmatrix}, H_1 = [1 \quad 1], H_2 = [1 \quad 1], Q = 1.5,$ $R_1 = 20, R_2 = 3.5$.

The conservative prediction error variances of the local robust and centralized robust fused steady-state Kalman predictors are given in Table 3.1.

When the fictitious noise variance $\Delta_\xi = \alpha I_2$ is prescribed, the robust region of uncertainty in the state matrix can be obtained by the searching method. When $\alpha = 3.6$, from Table 3.2, we can obtain that the robust region of centralized robust fused Kalman prediction is $-0.3 < \delta < 0.1$, which ensures $\det \Delta\Sigma_c > 0$, which yields $\Delta\Sigma_c > 0$. Similarly, the robust regions of the local robust Kalman prediction are

Table 3.1 The conservative prediction error variances of the local and centralized robust steady-state Kalman predictors

Σ_1	Σ_2	Σ_c
$\begin{bmatrix} 68.6106 & -0.6727 \\ -0.6727 & 21.9944 \end{bmatrix}$	$\begin{bmatrix} 64.1738 & -4.2876 \\ -4.2876 & 10.6171 \end{bmatrix}$	$\begin{bmatrix} 63.9180 & -4.5165 \\ -4.5165 & 10.0066 \end{bmatrix}$

Table 3.2 The determinants of $\Delta\Sigma_\theta$, $\theta = 1, 2, c$ with respect to δ

δ	$\det \Delta\Sigma_1$	$\det \Delta\Sigma_2$	$\det \Delta\Sigma_c$
−0.5	−54.2566	−28.7639	−28.5661
−0.4	−24.2111	−8.1500	−7.9089
−0.3	−1.3032	7.5152	7.7730
−0.2	14.5833	18.3569	18.6248
−0.1	22.5712	23.8877	24.1846
0.0	20.4353	22.7732	23.1547
0.1	3.8500	12.2712	12.8509
0.2	−35.1000	−13.0677	−12.0807

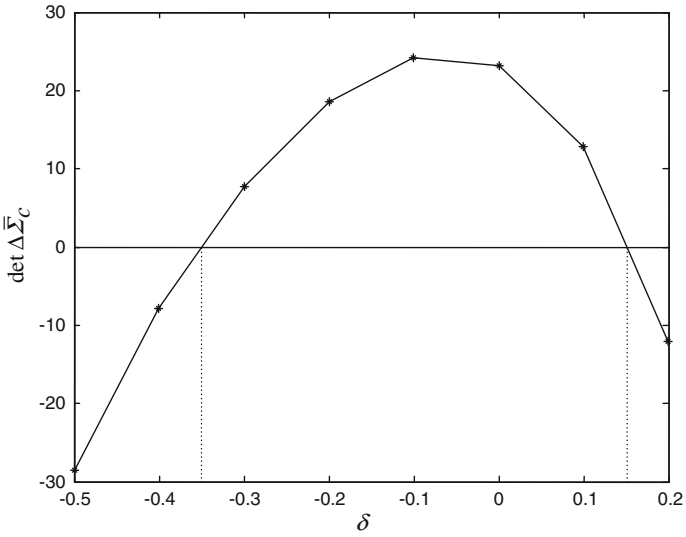


Fig. 3.1 Robust region of the centralized robust Kalman predictor

$-0.2 < \delta < 0.1$ and $-0.3 < \delta < 0.1$ respectively, which ensures $\det \Delta\Sigma_i > 0$, which yields $\Delta\Sigma_i > 0$.

From Fig. 3.1, we have a parabola going downwards. It can obtain a more precise robust region of centralized fusion robust Kalman predictor $-0.3534 < \delta < 0.1607$ by dichotomy, so that $\det \Delta\bar{\Sigma}_c > 0$ and $\bar{\Sigma}_c < \Sigma_c$ in this robust region.

The 1,000 Monte Carlo runs are performed. The MSE curves of the local and centralized robust steady-state Kalman predictors are shown in Fig. 3.2.

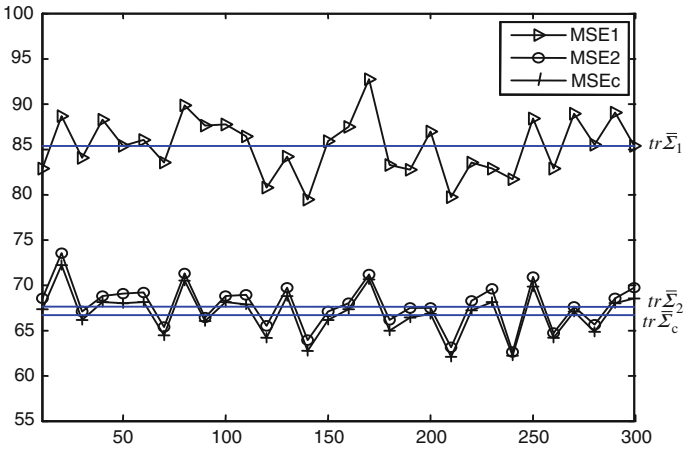


Fig. 3.2 MSE curves of the local robust and centralized robust fused Kalman predictors

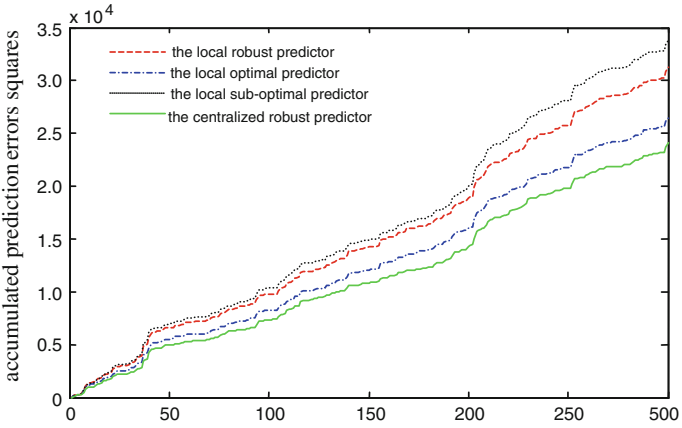


Fig. 3.3 The curves of accumulated prediction errors squares for component $x_1(t)$

From Fig. 3.2, we see that the values of $\mathrm{MSE}_\theta(t)$ are close to the corresponding $\mathrm{tr}\,\Sigma_\theta$, $\theta = 1, 2, c$ and the accuracy relation (3.29) holds. The curves of accumulated prediction errors squares for component $x_1(t)$ are shown in Fig. 3.3. From Fig. 3.3, we can see that the actual accuracy of centralized robust fused steady-state Kalman predictor is superior to others.

3.4 Conclusion

For multisensor systems with uncertainty parameters, a new robust Kalman prediction approach of compensating parametric uncertain by fictitious noise was presented. The problem is converted into the robust Kalman prediction problem for the system with uncertain noise variances. The local and centralized robust steady-state Kalman prediction algorithms are presented. Based on the Lyapunov equation, it is proved that the robustness of local and centralized robust fusion Kalman predictor, i.e., the actual predictor error variance have a conservative upper bound for all the admissible uncertainties. This approach is different from the Riccati equation approach and the linear matrix inequality (LMI) approach. The simulation results show that actual accuracy of centralized robust fusion Kalman predictor is higher than those of the local optimal Kalman predictor and the local suboptimal predictor. The simulation shows at how to search the robust region and shows its good performances.

Acknowledgment This work is supported by the Natural Science Foundation of China under grant NSFC-60874063 and NSFC-60374026.

References

1. Anderson BDO, Moore JB (1979) Optimal Filtering. Prentice Hall, Englewood Cliffs
2. Hashemipour HR, Rog A, Laub AJ (1988) Decentralized structures for parallel Kalman filtering. IEEE Trans Autom Control 33(1):88–94
3. Carlson NA (1990) Federated square root filter for decentralized parallel processes. IEEE Trans Aerosp Electron Syst 26(3):517–525
4. Saber RO (2007) Distributed Kalman filtering for sensor network, in: Proceeding of the 46th IEEE Conference on Decision and Control : 5492-5498
5. Mutambara GO (1998) Decentralized estimation and control for multisensor systems. CRC Press, Boca Raton
6. Lewis FL, Xie LH, Popa D (2008) Optimal and Robust Estimation, 2nd edn. CRC Press, New York
7. Xiong K, Wei CL, Liu LD (2012) Robust Kalman filtering for discrete-time nonlinear systems with parameter uncertainties. Aerosp Sci Technol 18:15–24
8. Qu XM, Zhou J (2013) The optimal robust finite-horizon Kalman filtering for multiple sensors with different stochastic failure rates. Appl Math Lett 26(1):80–86
9. Yang F, Li Y (2012) Robust set-membership filtering for systems with missing measurement:a linear matrix inequality approach. IET Signal Process 6(4):341–347
10. Zhang P, Qi W, Deng Z (2013) Centralized fusion steady-state robust Kalman filter For uncertain multisensor systems. In: Proceedings of 2013 Chinese intelligent automation conference, lecture notes in electrical engineering, vol 255, pp 219–226

Chapter 4
Robust Weighted Measurement Fusion Kalman Filter with Uncertain Parameters and Noise Variances

Chunshan Yang and Zili Deng

Abstract For the multisensor time-invariant system with both the uncertainties noise variances and parameters, by introducing a fictitious white noise to compensate the uncertain parameters, based on the minimax robust estimation principle and the Lyapunov equation method, a robust weighted measurement fusion Kalman filter is presented. It is proved that for prescribed upper bound variance of fictitious noise, there exists a sufficiently small robust region of uncertain parameter perturbances, such that its actual filtering error variances are guaranteed to have a conservative upper bound. A simulation example shows how to search the robust region, and shows its good performances.

Keywords Uncertain parameters · Uncertain noise variances · Fictitious white noise · Weighted measurement fusion · Minimax robust Kalman filter

4.1 Introduction

Multisensor information fusion Kalman filtering has been applied to many fields, [1, 2]. One of the key assumptions in Kalman filtering is that the model parameters and noise variances are exactly known. But in many applications, this condition cannot always hold, thus the performance of the Kalman filter may degraded or an

C. Yang · Z. Deng (✉)
Department of Automation, Heilongjiang University, XueFu Road 74, Heilongjiang University, Electronic and Engineering College 130, Harbin 150080, China
e-mail: dengzili891@163.com

C. Yang
e-mail: dzl@hlju.edu.cn

C. Yang
Department of Computer Science and Technology, ChengDong College of Northeast Agricultural University, Harbin 150025, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_4

inexact model may cause the filter divergence [3]. This has stirred up many studies on robust Kalman filter design.

So far, robust Kalman filters for systems with uncertain parameters have been designed, two important approaches are the Riccati equation approach [3] and linear matrix inequality (LMI) approach [4]. The robust Kalman filters for systems with uncertain noise variances have been designed [5, 6], a Lyapunov equation approach is presented to prove the robustness of the proposed robust Kalman filters. Up to now, the robust Kalman filters for uncertain systems both in noise variances and model parameters are seldom considered.

In this paper, we consider these two uncertainties for multi-sensor invariant system. By introducing a fictitious white noise to compensate the uncertain model parameter, the uncertain system can be converted into the worse-case conservative system with known parameters and uncertain noise variance. Using the minimax robust estimation principle, weighted least squares method, a robust weighted measurement fusion Kalman filter is presented based on the worst-case conservative system with the conservative upper bounds of noise variances. Furthermore, the robustness of the proposed robust Kalman filters is proved by Lyapunov equation approach.

4.2 Weighted Measurement and Local Robust Steady-State Kalman Filter

Consider the true discrete system with uncertain noise variances and uncertain model parameters.

$$x(t+1) = (\Phi_e + \Delta\Phi)x(t) + \Gamma w(t) \tag{4.1}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, \ldots, L \tag{4.2}$$

where state $x(t) \in R^n$, measurement of the ith subsystem $y_i(t) \in R^{m_i}$. $w(t) \in R^r$ and $v_i(t)$ are uncorrelated white noises with zero means and uncertain actual variances $\bar{Q}$ and $\bar{R}_i$, respectively. Assume that Q and R_i are conservative upper bounds of $\bar{Q}$ and $\bar{R}_i$, i.e.,

$$\bar{Q} \le Q,\ \bar{R}_i \le R_i,\ i = 1, \ldots, L \tag{4.3}$$

$\Phi_e,\ \Gamma,\ H_i$ are known constant matrices. $\Phi = \Phi_e + \Delta\Phi$ is the true transition matrix. $\Delta\Phi$ is the uncertain perturbances of model parameter matrix and satisfies that

$$\Delta\Phi \in \Re_{\Delta\Phi} \tag{4.4}$$

And each subsystem is completely observative and completely controllable.

A fictitious white noise $\xi(t)$ with zero mean and upper bound variance $\Delta_\xi > 0$ is used to compensate the uncertain model parameter error term $\Delta\Phi x(t)$ in (4.1), then the system (4.1) and (4.2) is transformed into the following worse-case conservative system with known parameters and uncertain noise variance

$$x(t+1) = \Phi_e x(t) + \Gamma w(t) + \xi(t) \tag{4.5}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1, \ldots, L \tag{4.6}$$

Assume that each measurement matrix H_i has a common $m \times n$ right factor H [7], i.e., $H_i = M_i H,\ i = 1, \ldots, L$ and define $M^{(0)} = \left[M_1^{\mathrm{T}}, \ldots, M_L^{\mathrm{T}}\right]^{\mathrm{T}}$, where the symbol T denotes the transpose. Assume that $M^{(0)}$ is of full-column rank. The centralized fusion measurement equation is given as

$$y_c(t) = H_c x(t) + v_c(t) \tag{4.7}$$

$$H_c = \left[H_1^{\mathrm{T}}, \ldots, H_L^{\mathrm{T}}\right]^{\mathrm{T}},\ v_c(t) = \left[v_1^{\mathrm{T}}(t), \ldots, v_L^{\mathrm{T}}(t)\right]^{\mathrm{T}} \tag{4.8}$$

and $v_c(t)$ has the conservative and actual variance matrix

$$R_c = diag(R_1, \ldots, R_L),\ \bar{R}_c = diag(\bar{R}_1, \ldots, \bar{R}_L) \tag{4.9}$$

Applying the WLS method, Eq. (4.7) can be converted into

$$y_M(t) = Hx(t) + v_M(t) \tag{4.10}$$

where $y_M(t)$ is the conservative weighted fusion measurement, $v_M(t)$ is the fused measurement white noise, such that

$$y_M(t) = \sum_{i=1}^{L} \left[M_i^{\mathrm{T}} R_i^{-1} M_i\right]^{-1} \sum_{i=1}^{L} M_i^{\mathrm{T}} R_i^{-1} y_i(t) \tag{4.11}$$

$$v_M(t) = \sum_{i=1}^{L} \left[M_i^{\mathrm{T}} R_i^{-1} M_i\right]^{-1} \sum_{i=1}^{L} M_i^{\mathrm{T}} R_i^{-1} v_i(t) \tag{4.12}$$

$v_M(t)$ has the conservative and actual variances matrix [5]

$$R_M = \left[M^{(0)T} R_c^{-1} M^{(0)}\right]^{-1} \tag{4.13}$$

$$\bar{R}_M = \sum_{i=1}^{L} \left[M_i^{\mathrm{T}} R_i^{-1} M_i\right]^{-1} \sum_{i=1}^{L} M_i^{\mathrm{T}} R_i^{-1} \bar{R}_i R_i^{-1} M_i \sum_{i=1}^{L} \left[M_i^{\mathrm{T}} R_i^{-1} M_i\right]^{-1} \tag{4.14}$$

For the conservative system (4.5) and (4.11), we have the conservative weighted measurement fusion steady-state Kalman filter

$$\hat{x}_M(t|t) = \Psi_M \hat{x}_M(t-1|t-1) + K_M y_M(t) \tag{4.15}$$

$$\Psi_M = [I_n - K_M H]\Phi_e,\ K_M = \Sigma_M H^{\mathrm{T}}\left[H\Sigma_M H^{\mathrm{T}} + R_M\right]^{-1} \tag{4.16}$$

Σ_M satisfies the steady-state Riccati equation

$$\Sigma_M = \Phi_e\left[\Sigma_M - \Sigma_M H^{\mathrm{T}}\left(H\Sigma_M H^{\mathrm{T}} + R_M\right)^{-1}\right]\Phi_e^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} + \Delta_\xi \tag{4.17}$$

From (4.5), (4.11), and (4.15), we have

$$\tilde{x}_M(t|t) = \Psi_M \tilde{x}_M(t-1|t-1) + [I_n - K_M H][\Gamma w(t-1) + \xi(t-1)] - K_M v_M(t) \tag{4.18}$$

So we have the conservative filtering error variance

$$P_M = \Psi_M P_M \Psi_M^{\mathrm{T}} + [I_n - K_M H]\left(\Gamma Q \Gamma^{\mathrm{T}} + \Delta_\xi\right)[I_n - K_M H]^{\mathrm{T}} + K_M R_M K_M^T \tag{4.19}$$

Now we find the actual filter error variance

$$\tilde{x}_M(t|t) = x(t) - \hat{x}_M(t|t) \tag{4.20}$$

where $x(t)$ is the true state given in (4.1), $\hat{x}_M(t|t)$ is the actual Kalman filter (4.15) with $y_M(t)$ is the actual fused measurement (4.11) with $y_i(t)$ is the actual $y_i(t)$ are define by (4.1) and (4.2).

Notice that the actual system (4.1)–(4.2) and conservative system (4.5)–(4.6) have the same weighted measurement fusion equation as (4.10)–(4.11).

For the actual system, $x(t)$ is defined by (4.1) and $y_M(t)$ is defined by the actual measurement based on (4.1) and (4.2). Hence we have the actual error

$$\tilde{x}_M(t|t) = \Phi_e x(t-1) + \Delta\Phi x(t-1) + \Gamma w(t-1) - \hat{x}_M(t|t) \tag{4.21}$$

From (4.1), (4.2) and (4.15) we have

$$\begin{aligned}\tilde{x}_M(t|t) = {} & \Psi_M \tilde{x}_M(t-1|t-1) + [I_n - K_M H]\Gamma w(t-1) \\ & + [\mathrm{I}_n - K_M H]\Delta\Phi x(t-1) - K_M v_M(t)\end{aligned} \tag{4.22}$$

So we have the actual filtering error variance

$$\begin{aligned}\bar{P}_M = {} & \Psi_M \bar{P}_M \Psi_M^T + [I_n - K_M H]\left(\Gamma \bar{Q} \Gamma^{\mathrm{T}} + \Delta\Phi X \Delta\Phi^{\mathrm{T}}\right)[I_n - K_M H]^{\mathrm{T}} \\ & + [I_n - K_M H]\Delta\Phi C \Psi_M^{\mathrm{T}} + \Psi_f C^{\mathrm{T}} \Delta\Phi^{\mathrm{T}}\left[I_n - K_f H\right]^{\mathrm{T}} + K_M \bar{R}_M K_M^T\end{aligned} \tag{4.23}$$

where $X = E[x(t)x^{\mathrm{T}}(t)]$, $C = E[x(t)\tilde{x}^{\mathrm{T}}(t|t)]$, X and C satisfy the following equation, respectively

$$X = \Phi X \Phi^{\mathrm{T}} + \Gamma \bar{Q} \Gamma^{\mathrm{T}} \tag{4.24}$$

$$C = \Phi C \Psi_M^{\mathrm{T}} + \Phi X \Delta\Phi^{\mathrm{T}}[I_n - K_M H]^{\mathrm{T}} + \Gamma \bar{Q} \Gamma^{\mathrm{T}}[I_n - K_M H]^{\mathrm{T}} \tag{4.25}$$

Lemma 4.1 [8] *Consider the Lyapunov equation with U being a symmetric matrix*

$$P = FPF^{\mathrm{T}} + U \tag{4.26}$$

If the matrix F is stable and U is positive (semi-)definite, then the solution P is unique, symmetric, and positive (semi-)definite.

Theorem 4.1 *For multisensor uncertain system* (4.1) *and* (4.2) *with uncertain model parameters and uncertain noise variances, the actual steady-state Kalman filter are robust in the sense that for all admissible model parameters* $\Delta\Phi$ *satisfying* (4.4) with δ_φ *being a sufficiently small position number, we have*

$$\bar{P}_M < P_M \tag{4.27}$$

Proof Define $\Delta P_M = P_M - \bar{P}_M$, subtracting (4.23) from (4.19) yields the Lyapunov equation $\Delta P_M = \Psi_M \Delta P \Psi_M^{\mathrm{T}} + V_M$ with the definition

$$V_M = [I_n - K_M H]\Gamma(Q - \bar{Q})\Gamma^{\mathrm{T}}[I_n - K_M H]^{\mathrm{T}} + K_M(R - \bar{R}_M)K_M^T + \bar{V}_M \tag{4.28}$$

$$\begin{aligned} \bar{V}_M = {} & [I_n - K_M H]\Delta_\xi[I_n - K_M H]^{\mathrm{T}} - [I_n - K_M H]\Delta\Phi X \Delta\Phi^{\mathrm{T}}[I_n - K_M H]^{\mathrm{T}} \\ & - [I_n - K_M H]\Delta\Phi C \Psi_M^{\mathrm{T}} - \Psi_M C^{\mathrm{T}} \Delta\Phi^{\mathrm{T}}[\mathrm{I}_n - K_M H]^{\mathrm{T}} \end{aligned} \tag{4.29}$$

From $[I_n - K_M H] = P_M \Sigma_M^{-1}$, we have that $\det[I_n - K_M H] = \det\left(P_M \Sigma_M^{-1}\right) \neq 0$, so that $[I_n - K_M H]$ is invertible, and

$$V_0 = [I_n - K_M H]\Delta_\xi[I_n - K_M H]^{\mathrm{T}} > 0 \tag{4.30}$$

□

From (4.29) and (4.30), when $\Delta\Phi \to 0$, then $\bar{V}_M \to V_0$. Hence there exists a sufficiently small robust region $\Re_{\Delta\Phi}$ of uncertain $\Delta\Phi$, such that for all $\Delta\Phi \in \Re_{\Delta\Phi}$, it follows that $\bar{V}_M > 0$. Applying Lemma 4.1 to Lyapunov equation $\Delta P_M = \Psi_M \Delta P \Psi_M^{\mathrm{T}} + V_M$ yields $\Delta P_M \geq 0$, i.e., $\bar{P}_M < P_M$. The proof is completed.

Similarly, for the local subsystem (4.1) and (4.2) we can obtain the robust local steady-state Kalman filter $\hat{x}_i(t|t)$ with the conservative and actual variances P_i and $\bar{P}_i$, and with the robustness $\bar{P}_i < P_i$, $P_M < P_i$, $i = 1, \ldots, L$.

Corollary 4.1 *We have the accuracy relations*

$$\bar{P}_i < P_i,\ \bar{P}_M < P_M \le P_i,\ i = 1, \ldots, L \tag{4.31}$$

$$\bar{P}_c < P_c,\ tr\bar{P}_M < trP_M < trP_i,\ i = 1, \ldots, L \tag{4.32}$$

Proof For the conservative system, we have $P_M = P_c$, where P_c is conservative variance of the centralized fusion Kalman filter for system (4.5) and (4.7). From $P_c \le P_i$, we have $P_M \le P_i,\ i = 1, \ldots, L$. The proof is completed. □

4.3 Simulation Example

Consider the 2-sensor invariant tracking system (4.1) and (4.2) with $\Phi_e = \begin{bmatrix} 0.8 & 0.3 \\ 0.5 & 0 \end{bmatrix}$, $\Delta\Phi = \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix}$, $\Gamma = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $H_1 = [1 \quad 0]$, $H_2 = [0 \quad 1]$, $Q = 1.5$, $R_1 = 2.5$, $\bar{Q} = 1.0$, $\bar{R}_1 = 2.0$, $R_2 = 4.5$, $\bar{R}_2 = 3.8$. The simulation results are given in the following. δ is uncertain perturbances of parameter.

The common right factor we select is $H = I_2$. Taking the conservative upper bound of the compensating fictitious noise variance as $\Delta_\xi = 0.5I_2$. The values of determinant $\bar{V}_M$ changed with the uncertainty δ are shown in Fig. 4.1. From Fig. 4.1, the robust region of uncertainty δ is $\Re_{\Delta\Phi} = \{\delta | \det \bar{V}_M > 0\} = (-0.806, 0.171)$, which ensures $\Delta P_M > 0$, i.e., $\bar{P}_M < P_M$.

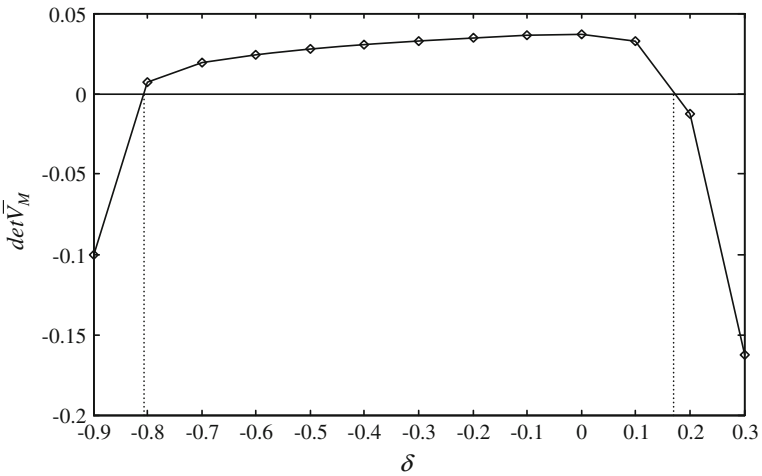


Fig. 4.1 The robust region of the fused robust Kalman filter

When $\bar{Q}$ varies from 0 to Q, the changes of robust region of the uncertainty with $\bar{Q}$ are given in Fig. 4.2. From Fig. 4.2, we can obtain that when $\bar{Q}$ varies from 0 to Q, the robust region of the fused Kalman filter narrows.

A three-dimensional figure of the robust region of the Kalman filter is given in Fig. 4.3. From Fig. 4.3, we can see that the robust region of the fused robust Kalman filter how changes over $\bar{Q}$ and δ.

Taking $\Delta_{\xi} = 0.5I_2$, $\delta = 0.1$ in the robust region, the comparisons of filtering performance among the weighted measurement fusion optimal, robust, and suboptimal Kalman filters are given in Fig. 4.4. From Fig. 4.4, we can see that the performance of suboptimal Kalman filter is clearly worse than that of the other two filters, because it does not consider the uncertainty of model parameter and noise variances, so suboptimal Kalman filter leads to serious performance loss.

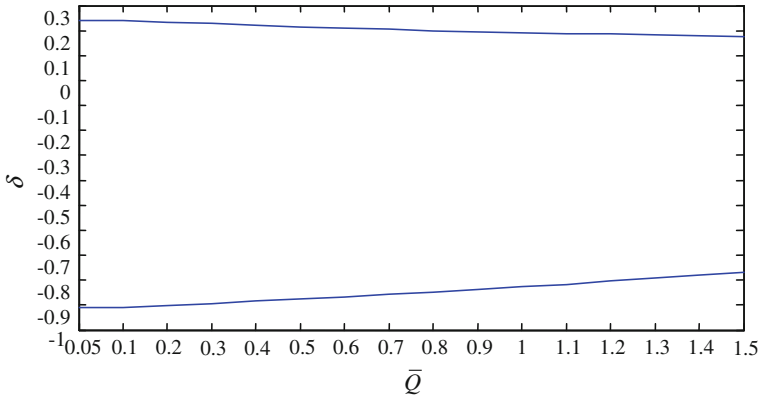


Fig. 4.2 The robust region of the fused robust Kalman filter changes over $\bar{Q}$

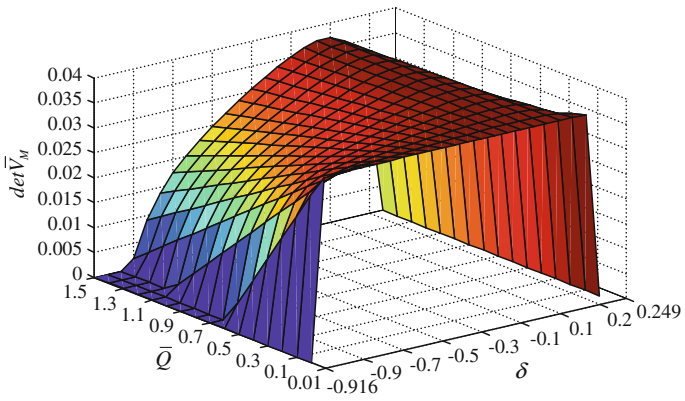


Fig. 4.3 The robust region of the fused robust Kalman filter changes over $\bar{Q}$ and δ

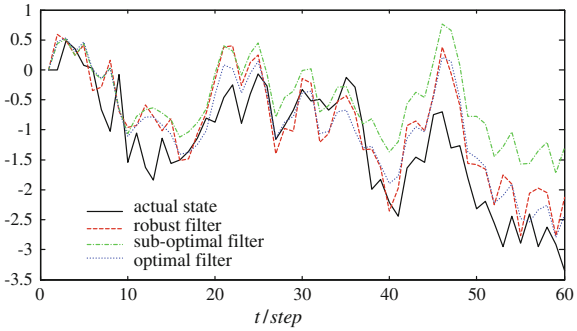


Fig. 4.4 The comparison among the optimal, robust, and suboptimal Kalman filters

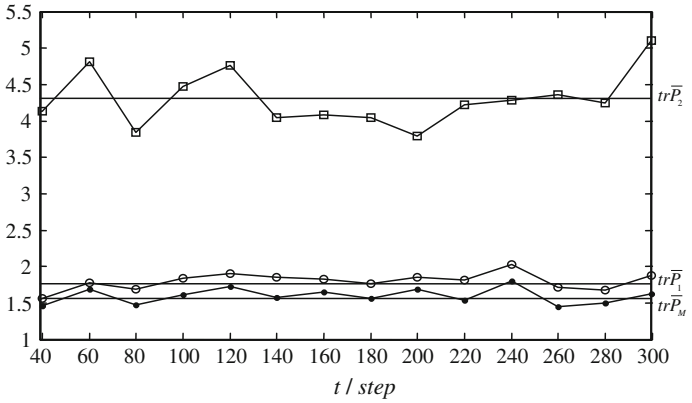


Fig. 4.5 The MSE curves of local and fused robust Kalman filters

In order to verify the above theoretical accuracy relations, Fig. 4.5 gives the mean square error (MSE) curves with $\rho = 200$ Monte Carlo simulation runs. According to the ergodicity [9], we have

$$\mathrm{MSE}_\theta(t) \to tr\bar{P}_\theta, \text{ as } t \to \infty, \ \rho \to \infty, \ (\theta = 1, 2, M) \tag{4.33}$$

From Fig. 4.5, we can see that when $t \to \infty$, the values of $MSE(t)$ are close to the corresponding theoretical values $tr\bar{P}_\theta$, which verifies the robust accuracy relation (4.31).

4.4 Conclusion

For the multisensor system with uncertain parameters and noise variances, using a fictitious noise approach to compensate parameter uncertainties, a robust weighted measurement fusion Kalman filter has been presented based on the worst-case conservative system with the conservative upper bounds of noise variances. Based on the Lyapunov equation approach, its robustness is proved, and their robust accuracy is higher than that of each local robust Kalman filter. A search approach for finding the robust region is given.

Acknowledgments This work is supported by the Natural Science Foundation of China under grant NSFC-60874063 and NSFC-60374026.

References

1. Shalom YB, Li XR, Kirubarajan T (2001) Estimation with applications to tracking and navigation. Wiley, New York
2. Sun S, Deng Z (2004) Multi-sensor optimal information fusion Kalman filter. Automatica 40 (6):1017–1023
3. Sriyananda H (1972) A simple method for the control of divergence in Kalman filter algorithms. Int J Control 16(6):1101–1106
4. Ebinara Y, Hagivara T (2005) A dilated LMI approach to robust performace analysis of linear time-invariant uncertain systems. Automatica 41(11):1933–1941
5. Qi W, Zhang P, Deng Z (2014) Robust weighted fusion Kalman filters for multisensor time-varing systems with uncertain noise variances. Sig Process 99:185–200
6. Qi W, Zhang P, Nie G-H, Deng Z (2014) Robust weighted fusion Kalman predictors with uncertain noise variances. Digit Signal Proc 30:37–54
7. Ran CJ, Hu YS, Gu L, Deng ZL (2008) Correlated measurement fusion steady-state Kalman filtering algorithms and their optimality. Acta Automatica Sinica 34:233–239
8. Kailath T, Sayed AH, Hassibi B (2000) Linear estimation. Prentice Hall, New York
9. Ljung L (1999) System identification. Theory for user, 2nd edn. Practice Hall PTR, Tsinghua University Press, Beijing

Chapter 5
Face Recognition Based on Maximum Sparse Coefficients of Object Region

Zineng Xu, Hongjun Li, Xiangyu Jin and Ching Y. Suen

Abstract Face recognition is an active topic in recognition systems, while face occlusion is one of the most challenging problems for recognition. Recently, robust sparse coding achieved the state-of-the-art performance, especially when dealing with occluded images. However, robust sparse coding is known that only guarantees the coefficient is global sparse when solving sparse coefficients. In this paper, we enable the elements in the object region to approximate global maximum by fitting the distribution of elements in the object region with successful recognition. The efficacy of the proposed approach is verified on publicly available databases. Furthermore, our method can achieve much better performance when the training samples are limited.

Keywords Face recognition · Maximum sparse coefficient · Occlusion

5.1 Introduction

Recently, face recognition (FR) approaches based on linear representations have led to state-of-the-art performance [1]. The most representative approach is sparse representation-based classification (SRC) [1]. However, SRC is not robust to contiguous occlusion such as sunglasses and scarf. Therefore, many related works [2–4] have been developed. In particular, the robust sparse coding (RSC) [2] method has achieved very good performance in FR with various occlusions. In order to construct a more robust model for sparse coding of face images, RSC finds

Z. Xu · H. Li (✉) · X. Jin
School of Electronic Information Engineering, Nantong University,
Nantong 226019, China
e-mail: lihongjun103@126.com

H. Li · C.Y. Suen
Centre for Pattern Recognition and Machine Intelligence,
Concordia University, Montreal, QC H3G 1M8, Canada

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_5

a maximum likelihood estimation solution of the coding coefficients and obtains high recognition accuracy; however, it is also very time-consuming, like SRC. Recently, Yang et al. [5] proposed a fast and robust FR method, which is faster than SRC and RSC.

Motivated by the recent success of RSC, we proposed a novel approach to improve the recognition rates by approximating the ideal sparse coefficients.

5.2 Related Works

Wright et al. [1] proposed a sparse representation-based classification scheme for FR. Let $D = [D_1, D_2, \ldots, D_s]$ be the set of training samples from all the s classes, where D_i is the subset of the training samples from class i. Denote y a testing sample. In SRC, y is sparsely coded over D via l_1-minimization:

$$\underset{\alpha}{\operatorname{argmin}}\{||y - D\alpha||_2^2 + \lambda||\alpha||_1\} \tag{5.1}$$

where λ is a scalar constant, α is the coding vector of y over D.

Yang et al. [2] proposed RSC to solve this problem by LASSO:

$$\min_{\alpha} ||y - D\alpha||_2^2 \quad \text{s.t.} \quad ||\alpha||_1 \leq \sigma \tag{5.2}$$

where $\sigma > 0$ is a constant. They used a weighted regression function to measure the representation residual. The final equation can be described as:

$$\min_{\alpha}\{||W^{1/2}(y - D\alpha)||_2^2 + \lambda||\alpha||_1\} \tag{5.3}$$

where W is a diagonal matrix, each element is the weight assigned to each pixel of the test image y.

5.3 Proposed Method

5.3.1 Motivation

In sparse coding, the ideal sparse coefficients α can be described as follows:

$$\alpha = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,k_i}, 0, \ldots, 0]^T \tag{5.4}$$

where k_i is the number of class i. We can know that RSC uses l_1-regularized least squares [6] to solve the sparse coefficients. Although this method can achieve good sparse representation, it only ensures the coefficients global sparse. In practice, we

hope the nonzero elements of sparse coefficients α can be concentrated in the object region as far as possible, and the remaining values are as small as possible. RSC is supervised, which means we know the labels of testing images. Thus we can easily identify the location of object region, and it is possible to control the distribution of sparse coefficients. We analyze how RSC method solves sparse coefficients. In RSC, when the testing image y belongs to class i, the solution of sparse coefficients can be performed as follows:

$$\alpha^{y\in i} = [\overbrace{0_{1,1}, \alpha_{1,2}\ldots, \alpha_{1,k_1}}^{\text{region 1}}, \ldots, \overbrace{\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,k_i}}^{\text{object region } i}, \ldots \underbrace{\alpha_{m,1}, 0_{m,2}, \ldots, \alpha_{m,k_m}}_{\text{region } m}, \ldots, \underbrace{0_{s,1}, \alpha_{s,2}, \ldots, \alpha_{s,k_s}}_{\text{region } s}]^T \tag{5.5}$$

where $\alpha_{1,k_1}, \ldots, \alpha_{i,k_i}, \ldots, \alpha_{m,k_m}, \ldots, \alpha_{s,k_s}$ are nonzero elements of coefficients vector $\alpha^{y\in i}$, s is the number of classes. Denote $\mathrm{SOR}_x = \alpha_{x,1} + \alpha_{x,2} + \cdots + \alpha_{x,k}$ as the sum of elements of coefficient vector in the region x, x = 1,2, …, s.

5.3.2 Maximum Sparse Coefficients of Object Region

We analyze the distribution of coefficients α in the object region with successful recognition. Figure 5.1 shows some distributions of elements in object regions. We can see that the value of nonzero elements in each object region is exponential growth. Considering that $\alpha < 1$, we choose logistic function to fit the distribution of elements in the object region. The basic formula of logistic function is

$$L(r) = \frac{1}{1 + p^{-r}} \tag{5.6}$$

where r is independent variable, p is a constant. Here, we modify the model in Eq. (5.6) to solve our problem.

The new formula can be described as follows:

$$L(\alpha_{\mathrm{or}}) = \begin{cases} \frac{c}{1+p^{(a-b\alpha_{\mathrm{or}})}}, & \max(\alpha_{\mathrm{or}}) < z \\ \frac{c_1}{1+p_1^{(a_1-b_1\alpha_{\mathrm{or}})}}, & z \le \max(\alpha_{\mathrm{or}}) < z_1 \\ \frac{c_2}{1+p_2^{(a_2-b_2\alpha_{\mathrm{or}})}}, & z_1 \le \max(\alpha_{\mathrm{or}}) < 1 \end{cases} \tag{5.7}$$

where $a, a_1, a_2, b, b_1, b_2, c, c_1, c_2, p, p_1, p_2, z, z_1$ and z_2 are constants, α_{or} is coefficient in the object region. From the result in Fig. 5.2, we can identify these parameters as follows: $a = 2.02$, $b = 20.01$, $c = 0.40$, $p = 8.01$; $a_1 = 4.00$, $b_1 = 20.00$, $c_1 = 0.50$, $p_1 = 5.99$; $a_2 = 3.31$, $b_2 = 10.52$, $c_2 = 0.74$, $p_2 = 4.05$.

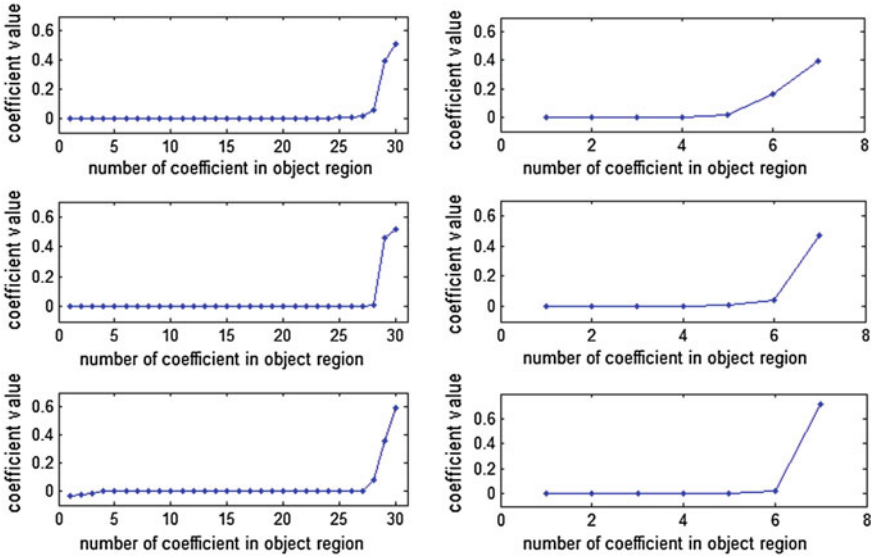


Fig. 5.1 Distribution of coefficients in object regions of different testing images

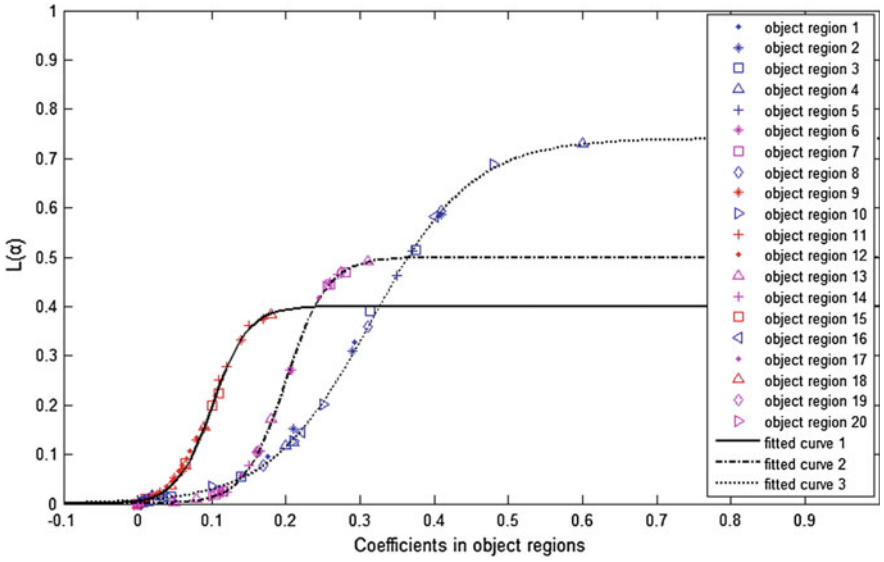


Fig. 5.2 Distribution of object regions with successful recognition and its fitted *curves*

For simplicity, we let $z = 0.2$, $z_1 = 0.35$ according to the range of independent variable. In order to globally maximize SOR_i, we need to find the maximum SOR_x. If $x \neq i$, we use the following formula to decrease its value:

$$M(\alpha_{\mathrm{mr}}) = \frac{\mathrm{SOR}_{\max} - \Delta\mathrm{SOR}_i}{\mathrm{SOR}_{\max}} \alpha_{\mathrm{mr}} \tag{5.8}$$

where α_{mr} is the coefficient in the maximum region, $\mathrm{SOR}_{\max}$ is the maximum region, $\Delta\mathrm{SOR}_i$ is the increment in the object region. Using Eqs. (5.7) and (5.8), we can ensure SOR_i to be $\mathrm{SOR}_{\max}$. Finally, the sparse coefficients can be described as follows:

$$\alpha = [\alpha_{\mathrm{r1}}, \alpha_{\mathrm{r2}}, \ldots, L(\alpha_{\mathrm{ri}}), \ldots, M(\alpha_{\mathrm{rs}}), \ldots, \alpha_{\mathrm{rk}-1}, \alpha_{\mathrm{rk}}]^T \tag{5.9}$$

where α_{rj} is the coefficient in region j, j = 1, 2, 3, …, k. $\alpha_{\mathrm{ri}} = \alpha_{\mathrm{or}}$, $\alpha_{\mathrm{rs}} = \alpha_{\mathrm{mr}}$.

5.3.3 Algorithm

We summarize the overall procedure of our algorithm as follows:

Step 1. Input: Normalized test sample y, dictionary D = [D_1, D_2,…,D_s]

Step 2. Solve α: $\min_{\alpha}\{||W^{1/2}(y - D\alpha)||_2^2 + \lambda||\alpha||_1\}$

a. $j = 1$, Initialize W_j and residual e_j
b. Initialize α
c. Identify $\mathrm{SOR}_{\max}$ and $\alpha_{\max}$
d. If $(\mathrm{SOR}_i < \mathrm{SOR}_{\max})$ and $(\alpha_{i-\max} < \alpha_{\max})$

$$\alpha = [\alpha_{r1}, \alpha_{r2}, \ldots, L(\alpha_{\mathrm{ri}}), \ldots, M(\alpha_{\mathrm{rs}}), \ldots, \alpha_{\mathrm{rk}-1}, \alpha_{\mathrm{rk}}]^T \tag{5.10}$$

Else go to next step. ($\alpha_{i-\max}$ is the maximal element in the object region)

e. Sparse coding by l_1-regularized least squares: α, then go back to c.
f. Compute residual: $e_j(y) = ||y - D\alpha||_2^2$
g. Update weights: W_j
h. $j = j + 1$, go back to b. until the maximal number of iterations is reached.

Step 3. Output: the identity of y as $\mathrm{Identity}(y) = \mathrm{argmin}\{e_j\}$

5.4 Experimental Results

In this section, we carry out experiments on benchmark face databases (AR [7] and Extended Yale B [8]) to demonstrate the performance of our algorithm. In all experiments, we compare our method with some popular methods such as CRC_RLS [9], FDDL [10], and RSC [2].

5.4.1 Recognition Without Occlusion

Extended Yale B database: The Extended Yale B database consists of 2414 frontal-face images of 38 individuals. We used the cropped and normalized 45 × 40 face images. Figure 5.3 shows some facial images from the Extended Yale B database. Table 5.1 shows the recognition rates versus feature dimension by CRC_RLS, FDDL, RSC and our method.

Fig. 5.3 Samples of different illuminations from the Extended Yale B database

Table 5.1 Face recognition rates on the Extended Yale B database

Algorithm	Feature dimension		
	84	120	240
CRC_RLS (%)	96.6	97.2	98.4
FDDL (%)	93.1	96.4	98.0
RSC (%)	**98.8**	99.3	99.3
Our method (%)	98.7	**99.6**	**99.6**

Fig. 5.4 Samples from the AR database

AR database: The AR database consists of over 4,000 frontal images from 126 individuals. Figure 5.4 shows some facial images for example. The comparison of our method with its competing methods is given in Table 5.2. Table 5.3 shows that our method achieves much higher recognition rates than other three methods in different numbers of training samples. Our method performs better than RSC when there are few training samples.

Table 5.2 Face recognition rates on the AR database

Algorithm	Feature dimension			
	30	54	120	300
CRC_RLS (%)	64.6	80.6	90.1	93.7
FDDL (%)	53.3	77.4	86.4	91.4
RSC (%)	70.3	86.7	94.6	96.1
Our method (%)	**71.0**	**91.1**	**98.0**	**98.9**

Table 5.3 Recognition rates of different numbers of training samples on the AR database

Num	Algorithm			
	CRC_RLS (%)	FDDL (%)	RSC (%)	Our method (%)
1	59.4	–	65.0	**66.9**
2	62.0	65.4	68.0	**73.1**
3	64.3	68.0	72.6	**77.4**
4	76.1	71.0	77.6	**84.3**
5	86.6	81.0	86.9	**91.9**
6	93.9	91.3	95.4	**98.3**
7	93.7	91.4	96.1	**98.9**

5.4.2 Recognition with Real Disguise

In this section, we evaluate the robustness of our method to real disguise. We select a subset of the AR database consisting of both neutral and corrupted images in this experiment. The following three scenarios are considered for performance evaluation: Sunglasses, Scarf, and Sunglasses + Scarf. Figure 5.5 shows one person's images with different disguises from training and testing images. Table 5.4 shows that our recognition rates are higher than other methods. The experimental results also demonstrate the good robustness of our method to face occlusion.

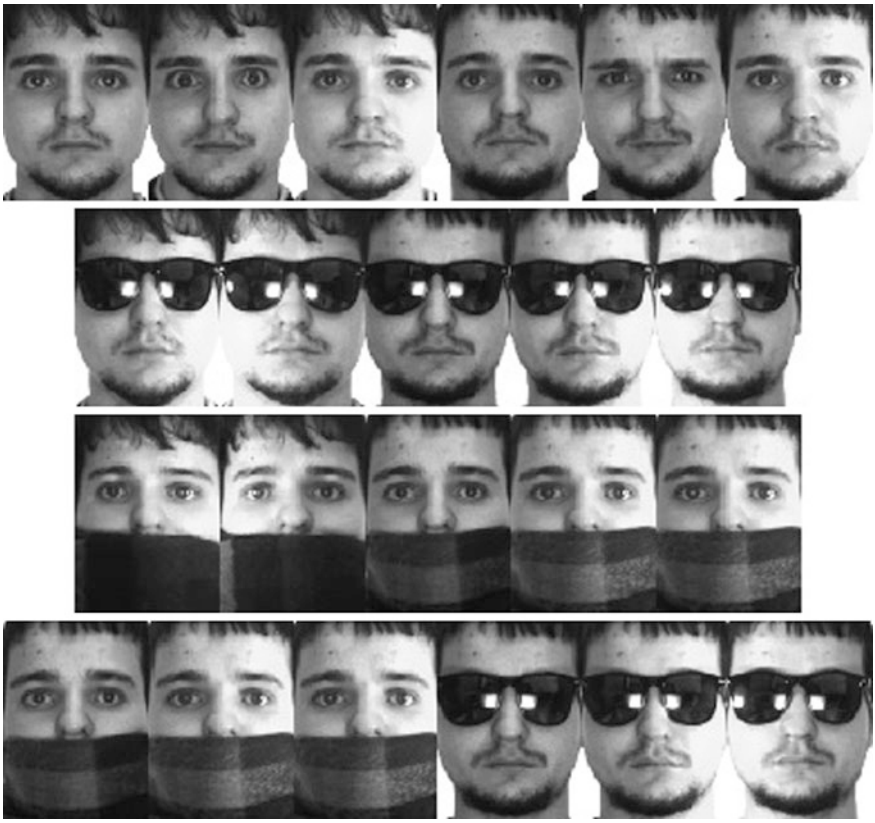


Fig. 5.5 The first row is training images, the rest are testing images

Table 5.4 Face recognition rates on the AR database with disguise occlusion

Algorithm	Sunglasses	Scarf	Sunglasses + scarf
CRC_RLS (%)	47.8	46.8	51.3
RSC (%)	91.6	88.8	92.2
Our method (%)	**91.8**	**90.0**	**92.7**

5.5 Conclusion

In this paper, motivated by the recent success of RSC, we present a novel method to recognize face image with illumination variations and occlusion. We add constraints and use logistic function to optimize the sparse coefficients. Our extensive experimental results on benchmark face databases show that the proposed method is effective and robust. Compared with RSC, our method achieves higher recognition rates, even when the training samples are limited.

Acknowledgments This work is supported by the National Natural Science Foundation of China (NO.61171077), University Science Research Project of Jiangsu Province (NO.12KJB510025), Nantong University Undergraduate Training Program for Innovation (NO.2013067), and the Natural Sciences and Engineering Research Council of Canada.

References

1. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
2. Yang M, Zhang L, Yang J et al (2013) Regularized robust coding for face recognition. IEEE Trans Image Process 22(5):1753–1766
3. Ou WH, You XG, Tao DC et al (2014) Robust face recognition via occlusion dictionary learning. Pattern Recogn 47(4):1559–1572
4. Li HJ, Nobile N, Suen CY (2014) Face recognition based on discriminative dictionary with multilevel feature fusion. Lect Notes Artif Intell 8774:252–263
5. Yang M, Feng ZZ, Shiu SCK et al (2014) Fast and robust face recognition via coding residual map learning based adaptive masking. Pattern Recogn 47(2):535–543
6. Goldstein T, Osher S (2009) The split Bregman method for L1-regularized problems. SIAM J Imaging Sci 2(2):323–343
7. Martinez A, Benavente R (1998) The AR face database. Technical report 24, CVC
8. Georghiades A, Belhumeur P, Kriegman D (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Machine Intell 23(6):643–660
9. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of ICCV, pp 471–478
10. Yang M, Zhang L, Feng X et al (2014) Sparse representation based fisher discrimination dictionary learning for image classification. Int J Comput Vision 109(3):209–232

Chapter 6
Hybrid Dependency Parser with Segmented Treebanks and Reparsing

Fuxiang Wu and Fugen Zhou

Abstract We propose a hybrid dependency parsing pipeline which combines transition-based parser and graph-based parser, and use segmented treebanks to train transition-based parsers as subparsers in front end, and then propose a constrained Eisner's algorithm to reparse their outputs. We build the pipeline to investigate the influence on parsing accuracy when training with different segmentations of training data and find a convenient method to obtain parsing reliability score while achieving state-of-the-art parsing accuracy. Our results show that the pipeline with segmented training dataset could improve accuracy through reparsing while providing parsing reliability score.

Keywords Hybrid dependency parsing · Constrained Eisner's algorithm · Parsing reliability score · Transition-based parser · Graph-based parser

6.1 Introduction

A good amount of research has been devoted to parsing technology, due to the importance of dependency parsing, and many natural language applications, such as information retrieval and Q&A system [10], employing it as a base component, and their performance may highly rely on the parsing result.

Recent methods of dependency parsing can be divided into two classes: data-driven methods and rule-based methods. Data-driven methods usually are statistical parser and use some machine learning algorithms to catch the statistical features of data in order to produce syntactic relations of words in sentences.

F. Wu (✉) · F. Zhou
Image Processing Center, Beihang University, Beihang, China
e-mail: fxwuedu@buaa.edu.cn

F. Zhou
e-mail: zhfugen@buaa.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_6

Most of the state-of-the-art parsers are statistical parser, parsing accuracy of which highly relies on the quality and quantity of treebank [3, 6, 11]. There are several treebanks in China for syntactic parsing, such as Penn Chinese Treebank (CTB) [8] and Chinese Dependency Treebank (CDT). The CTB is constituency annotation and was retrieved from Xinhua Newswire, Hong Kong news, Sinorama and ACE broadcast news, while the CDT is a dependency treebank which was retrieved from People's Daily newswire stories. Occurrence frequency of syntactic substructure (subtree) in one treebank may vary from another. Some would frequently occur in the treebank but others would not. The rare syntactic substructures of some sentences in the treebank are well formed for human, but would be abandoned when there are some common syntactic substructures which are conflict with the rare ones and prevent the parser from handling rare syntactic structures correctly. This would lead to label attachment recall rate degradation partly.

In order to examine the assumption, we segment treebank into k-parts in one round, and train a transition-based parser (subparser) for each part, and then a result set generated by the subparsers is compared with *k*-best generated by parser trained with full treebank. We learn that label attachment recall rate of k-segments would be higher than corresponding *k*-best's, and this confirms the assumption. Depending on this phenomenon, we further employ a parser to post-reparse the subparsers output. Since transition-based parser and graph-based parser have different training and inference algorithms [5, 7] and have different behaviors, we construct the post-reparser with constrained Eisner's algorithm [2, 4] to find maximum spanning trees (MST). The experiment shows that the pipeline could improve the parsing accuracy while computing the parsing reliability score.

6.2 Parsing Pipeline

The pipeline in this paper addresses the general structural prediction problem, which map an input sentence $x \in \mathbf{X}$ to an output dependency structure $y \in \mathbf{Y}$, which is composed of edge e,

$$e = \begin{cases} \langle i, j, \leftarrow, l \rangle \\ \langle i, j, \rightarrow, l \rangle \end{cases} \tag{6.1}$$

where i and j are relation endpoints, l is dependent label in label set $\mathbf{L}$. We employ transition-based parser with beam search [9] as subparser and use MST parser with conditional random field as post-reparser. In CRF model, the output y probability would be,

$$p(y|x) = \exp(f(y, x) \cdot \boldsymbol{\lambda}) / Z(x) \tag{6.2}$$

where $f(y,x)$ maps y and x to a feature vector, $\boldsymbol{\lambda}$ is a corresponding weight vector, and $Z(x)$ is the normalization factor. For a sentence x, the parsing result y is calculated by finding the highest probability one among the all possible results,

$$O(x) = \arg\max_{y \in \mathbf{PSET}(x)} p(y|x) \tag{6.3}$$

where $\mathbf{PSET}(x)$ denotes the set of the possible result for the sentence x.

The pipeline is composed of a training procedure and a parsing procedure. The training procedure mainly creates a set of subparsers which are transition-based parser. And in parsing procedure, we first use this set of subparsers to achieve a result set, and then reparse the set to compute the best output.

6.2.1 Training

The training procedure is as follows:

- Segmenting treebank into equally sized sub-treebanks $\boldsymbol{\Omega} = \{b_i\}_{i=1,\ldots,N}$
- Training subparser t_i with sub-treebank $b_i \in \Omega$ to build a set of subparsers $\boldsymbol{\Gamma}_N = \{t_i\}_{i=1,\cdots,N}$.
- Training the MST parser T with the whole treebank to calculate the weight vector $\boldsymbol{\lambda}_T$ for features.

Through this procedure, we get a trained model $\{N, \boldsymbol{\Gamma}_N, \boldsymbol{\lambda}_T, f_T\}$, where N is the number of subparsers trained by segmented treebanks; f_T is a feature extract function of MST parser, built by feature temples.

6.2.2 Post-reparsing

In parsing step, a set of result $\mathbf{R} = \{r_i\}_{i=1,\ldots,N}$, which is different from N-best result, have been generated by subparser in $\boldsymbol{\Gamma}_N$ for an input sentence x. We use them to constrain the searching space for the sentence (far small than the full searching space), and then employ constrained Eisner's algorithm to extract the best result. The constraint scores are obtained as follows:

$$s_c(e, \mathbf{R}) = \begin{cases} f_T(e) \cdot \boldsymbol{\lambda}_T & \text{if} \quad e \in r_i, i = 1, \ldots, N \\ 0 & \text{else} \end{cases} \tag{6.4}$$

where $f_T(e)$ maps edge e to a feature vector. And mixture score is as follows:

$$s_{\text{mixc}}(e, \mathbf{R}, \alpha) = \alpha \cdot f_T(e) \cdot \boldsymbol{\lambda}_T + (1 - \alpha) \cdot s_c(e, \mathbf{R}) \tag{6.5}$$

Table 6.1 Pseudo-code for constrained Eisner's algorithm

Allocate $E[N][N][2][2][N_L]$ array for edge's scores
Initialize each score in the array to zero
for i in $[1,\cdots,N]$
  for j in $[1,\cdots,N]$
    $t = i + j$
    if $t > N$ then break
    $E[i][j][\leftarrow][1][l] = \max_{i\le k<j}(\max_{u\in L} E[i][k][\rightarrow][0][u] +$
      $\max_{u\in L} E[k][j][\leftarrow][0][u] + s_{mixc}(\langle i,j,\leftarrow,l\rangle, \mathbf{R}, \alpha)$
    $E[i][j][\rightarrow][1][l] = \max_{i\le k<j}(\max_{u\in L} E[i][k][\rightarrow][0][u] +$
      $\max_{u\in L} E[k][j][\leftarrow][0][u] + s_{mixc}(\langle i,j,\rightarrow,l\rangle, \mathbf{R}, \alpha)$
    $E[i][j][\leftarrow][0][l] = \max_{i\le k<j}(\max_{u\in L} E[i][k][\leftarrow][0][u] + \max_{u\in L} E[k][j][\leftarrow][1][u]$
    $E[i][j][\rightarrow][0][l] = \max_{i\le k<j}(\max_{u\in L} E[i][k][\rightarrow][1][u] + \max_{u\in L} E[k][j][\rightarrow][0][u]$
Where N_L is the size of label set L.

where α is a mixture factor which controls the strength of constraint from the result set, given a sentence $S = w_0 w_1 \cdots w_N$ and the corresponding $\mathbf{R}$. The post-reparsing procedure is as follows (Table 6.1).

6.2.3 Reliability Score of Dependency Relation

With the pipeline, we can get a set of subparsers $\mathbf{\Gamma}_N$, in which each subparser is trained by different parts of training corpora. Because each part of corpora can be seen as unseen data from other part, we can assume that the parsing result of each subparser for a sentence is supported by the corresponding part of training corpora. Thus, reliability score can be calculated like a weighted voting scheme [1] as follows:

$$c(e_i) = \sum_{\varepsilon\in\mathbf{E}_i, \varepsilon = e_i} \exp(\xi \cdot f_T(\varepsilon)\cdot \boldsymbol{\lambda}_T) \Big/ \sum_{\varepsilon\in R_i} \exp(\xi \cdot f_T(\varepsilon)\cdot \boldsymbol{\lambda}_T) \tag{6.6}$$

where $\mathbf{E}_i$ is a set of dependency relationships $\langle i,\cdot,\cdot,\cdot\rangle$, which is ith relationship of dependency structure in set $\mathbf{R}$, ξ is an adjusting factor, and when $\xi = 0$, $c(e_i)$ is normal voting score for edge e_i.

Table 6.2 The training, development, and test data for CTB6

	File index	Sentences
Training	1–1129; 2019–2923	24,092
Dev	2924–3012; 3108–3145	1191
Test	1130–1151; 2000–2018; 3013–3107	2846

Table 6.3 The test results of the baseline parsers

	ZPar	MSTParser1	MSTParser2	crfParser
LAS	0.824361	0.775656	0.763978	0.782913
UAS	0.83939	0.81595	0.82032	0.8024

6.3 Baseline and Experiments

This section presents the pipeline experiments of segmentation and post-reparsing. Before this, we only evaluate the pipeline with Chinese Penn Treebank corpora as heavy computation cost for the CRF training without loss of generality. We split sentences in the Penn Treebank 6.0 into training, development, and test set as Table 6.2, and then employ the head-finding rule to translate them into dependency structures.

Baseline parsers are ZPar[1] dependency parser, MSTParser[2], and crfParser,[3] which are open source projects and have achieved state-of-the-art accuracy. They are trained with the training data in Table 6.2, and use their default feature temple, respectively. The test results are as follows (Table 6.3).

where MSTParser1 and crfParser are first order graph-based parsers, MSTParser is second order parser, and ZPar is transition-based dependency parser.

In order to explore the phenomenon brought by corpora segmentation, we segment the training data into parts with different number, namely, $\mathbf{\Theta} = [2, 3, 4, 6, 12]$, and then build the set of subparser $\mathbf{\Gamma}_K$ for each $k \in \Theta$.

6.3.1 Attachment Recall Rate

Labeled/unlabeled attachment recall rate (LAR/UAR) is the ratio of correct labeled/unlabeled attachment among the dependency structure of result set,

[1] http://sourceforge.net/projects/zpar/.

[2] http://sourceforge.net/projects/mstparser/.

[3] http://sourceforge.net/projects/crfparser/.

$$\begin{cases} \mathrm{LAR}(\Pi) = \sum\limits_{(R_K, r_c) \in \Pi} \delta_L(\mathbf{R}_K, r_c) \Big/ \sum_{(\mathbf{R}_K, r_c) \in \Pi} |r_c| \\ \mathrm{UAR}(\Pi) = \sum\limits_{(R_K, r_c) \in \Pi} \delta_U(\mathbf{R}_K, r_c) \Big/ \sum_{(\mathbf{R}_K, r_c) \in \Pi} |r_c| \end{cases}$$

where $\mathbf{R}_K$ and r_c are a set of parsing result and gold dependency structure for a sentence, $\mathbf{\Pi}$ generated from test data is a set of $(\mathbf{R}_K, r_c)$, $|r_c|$ is the number of relationship in dependency r_c, function $\delta_L(\mathbf{R}_k, r_c)$ counts the correct dependency relationships with label in r_c which coexist in result set $\mathbf{R}_k$, and function δ_U counts similarly without label.

We calculate LAR and UAR for the k-segment's result set R_K and baseline parser's k-best result. The relationship between number of parts and attachment recall rate is as follows.

With data segmentation, we can achieve higher LAR and UAR then k-best parsing result, this means that the dataset $\mathbf{R}_K$ would cover more correct dependency relationship than the k-best dataset. It would be beneficial to postprocessing in the pipeline, such as reparsing and reranking, with small searching space.

6.3.2 Post-reparsing

In post-reparsing state, we analyze the result set generated by the set of subparser $\mathbf{\Gamma}_N$ or the N-best result made by baseline parser to get final parsing result, and their accuracy is as follows.

For each $k \in \mathbf{\Theta}$, we search the best $\alpha \in [0, 1]$ for calculating the LAS/UAS of each k-segment's or k-best result. The highest LAS of 2-segmentation is 83.2112, and is 0.7751 % higher than the ZPar in baseline parsers, and 2-best's is 83.1275 %. From Fig. 6.2, we could see that the LAS of k-segment and k-best is lower than the ZPar's when $k > 2$, meanwhile, k-segment's LAS is lower than k-best's. The reason may be the postparser, which is first-order minimum spanning tree parser with local features, is not powerful enough to utilize the higher label attachment recall rate. That is why 2-segment's LAS is higher than 2-best's. From Fig. 6.1, we could find that k-segment's LAR ascends faster than k-best's, and the k-best's LAS descend slower than k-segment's since $k > 3$ in Fig. 6.2, this is also shown that the post-parser needs a finer design. Besides, we employ reliability score $c(e_i)$ to rerank the result set $\mathbf{R}_K$, the result is as follows:

From Fig. 6.3, we can find that using reliability score $c(\boldsymbol{e}_i)$ to directly select dependency relationship is feasible. Their LAS/UAS are higher than each element in k-segment's result set, but their output may be not a tree, and need further process. The LAS/UAS of reranking result is lower than the baseline ZPar, this may also due to weakness of the postparser as well.

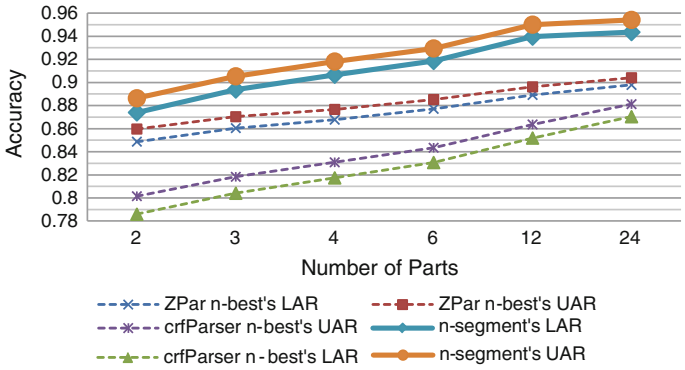


Fig. 6.1 The labeled/unlabeled attachment recall rate (LAR/UAR)—number of parts curve

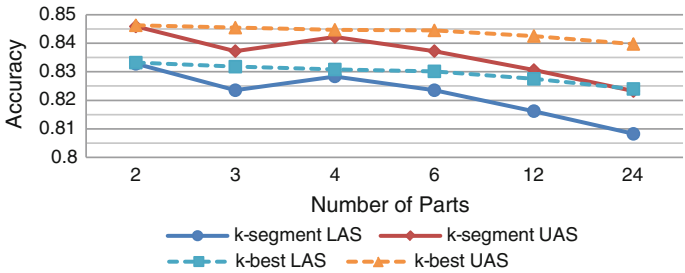


Fig. 6.2 The LAS/UAS of reparsing for *k*-segment's and *k*-best result set

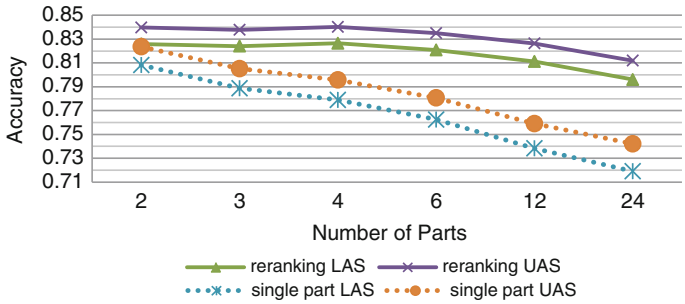


Fig. 6.3 The LAS/UAS of reranking for *k*-segment's result set. The dot line is the maximum LAS/UAS of element in *k*-segment's set

6.4 Conclusions and Future Work

We build a hybrid parsing pipeline, which employs transition-based dependency parser as subparser in front end, and then use graph-based dependency parser in next stage. Finally, we investigate the influence on the pipeline with different *k*-segment dataset. From the experiment, we found that using segmentation of training data would largely improve the labeled/unlabeled attachment recall rate with some final LAS/UAS drop, and the result set generated by the subparsers with high attachment recall rate could be used to calculate reliability score, such as simple voting scheme used in this paper. Besides, the hybrid pipeline could improve the final LAS/UAS when using 2-segment. But it cannot further improve the accuracy due to weak postparser. In future, we would try to use more sophisticated parser as postparser to explore the searching space constructed by the subparsers effectively.

References

1. Collins M (2002) Ranking algorithms for named-entity extraction: boosting and the voted perceptron. In: Proceedings of the 40th annual meeting on association for computational linguistics (ACL' 02), pp 489–496
2. Eisner J (1996) Three new probabilistic models for dependency parsing: an exploration. In: Proceedings of the 16th international conference on computational linguistics (COLING-96), pp 340–345
3. Li ZH, Liu T, Che WX (2012) Exploiting multiple treebanks for parsing with quasi-synchronous grammars. In: Proceedings of the 50th annual meeting of the association for computational linguistics (ACL' 12), pp 675–684
4. McDonald R, Pereira F (2006) Online learning of approximate dependency parsing algorithms. In: Proceedings of the 11th international conference of the European chapter of the association for computational linguistics (EACL 2006), pp 81–88
5. Nivre J, McDonald R (2008) Integrating graph-based and transition-based dependency parsers. In: Proceedings of the 46th annual meeting of the association for computational linguistics, pp 950–958
6. Niu ZY, Wang HF, Wu H (2009) Exploiting heterogeneous treebanks for parsing. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP, pp 46–54
7. Plank B, Noord GV (2010) Grammar-driven versus data-driven: which parsing system is more affected by domain shifts? In: Proceedings of the 2010 workshop on NLP and linguistics: finding the common ground (NLPLING' 10), pp 25–33
8. Xue NW, Xia F, Chiou FD, Palmer M (2005) The Penn Chinese treebank: phrase structure annotation of a large corpus. Nat Lang Eng 11(2):207–238
9. Zhang Y, Clark S (2011) Syntactic processing using the generalized perceptron and beam search. Comput Linguist 37(1):105–151
10. Zhou GY, Cai L, Zhao J, Liu K (2011) Phrase-based translation model for question retrieval in community question answer archives. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies (HLT' 11), pp 653–662
11. Zhou GY, Zhao J (2013) Joint inference for heterogeneous dependency parsing. In: The 51st annual meeting of the association for computational linguistics, pp 104–109

Chapter 7
Accelerated Rendering and Fast Reconstruction of EEG Data in Real-Time BCI

Ning Wang, Peng Lu, Lipeng Zhang, Shijie Li and Hanghang Hu

Abstract In real-time BCI (Brain Computer Interface), the *ITRs* (Information Transmission Rates) is one of the most common criteria for evaluating the performance of the whole system, and one of the key factors is the duration of one single trial. This paper aims at improving the *ITRs* by decreasing the duration. Accelerated rendering is used for drawing raw EEG (Electroencephalogram) data in presentation thread, and thread scheduling based on adaptive one-sided fuzzy inference and the mechanism of mutual exclusion and synchronization with semaphore is adopted to recombine intervening data blocks in reconstruction thread.

Keywords Real-time BCI · The mechanism of mutual exclusion · EEG data · Thread scheduling · Fuzzy inference

7.1 Introduction

The target of BCI (Brain Computer Interface) is to transform awareness into real-time commands of controlling external devices and communication tools [1, 2]. Compared with offline BCI, continuous and real-time judgement of brain status is needed in online BCI. The key strategy of a high-performance BCI system is the immediate processing of raw EEG (Electroencephalogram) data [3]. By far, the usual policy is adopting general software platform, for instance, LabVIEW

N. Wang · P. Lu (✉) · L. Zhang · S. Li · H. Hu
The 27th Research Institute of China Electronics Technology Group Corporation, Zhengzhou 450047, China
e-mail: lupeng@zzu.edu.cn

N. Wang · P. Lu · L. Zhang · S. Li · H. Hu
School of Electric Engineering, Zhengzhou University, Zhengzhou 450001, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_7

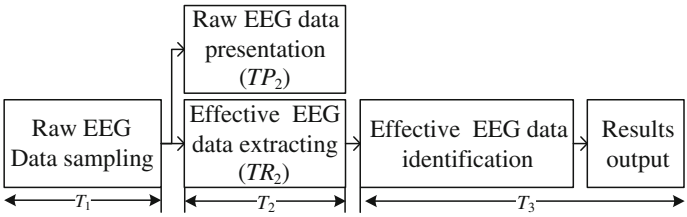


Fig. 7.1 General procedures of online BCI

components are used to handle data streams in document [4], and BCI2000 is adopted in document [5]. While general software packages fixed function interfaces or mix too many other function modules, their flexibility and efficiencies are limited for online BCI system when facing complicated problems. This paper focuses on solving the instantaneity of online BCI through deeper threads mechanism.

The main performance index of an online BCI system is the ITRs (Information Transmission Rates) calculated with the following equation (unit: bits/min),

$$\mathrm{ITRs} = B * 60/T, \tag{7.1}$$

where

$$B = \mathrm{lb}N + P\mathrm{lb}P + (1-P)\mathrm{lb}[(1-P)/(N-1)], \tag{7.2}$$

indicates the amount of information transmission in a single trial, N indicates the classification number of awareness recognition, P indicates the accuracy rate of recognition, and T is the period (unit: s) of a single trial. When N is certain, ITRs have positive relationship with P and inverse relationship with T. Thus decreasing T is one of the keys to improve ITRs.

In general, online BCI system [6] with synchronous stimulation, one single trial can be decomposed into three stages $T1$, $T2$, and $T3$, as shown in Fig. 7.1. The system samples raw EEG data in $T1$, presents and reconstructs them in $T2$, and analyzes them in $T3$. Before effective EEG data comes in the next trial, the whole procedure has to be finished in time. Document [7] lists each parameter's influences on ITRs, which still assumes that each trial runs serially.

In our experiment, we found that stage $T2$, $T3$ always consume more time than stage $T1$, and there are even losses of effective EEG data in some trials. The primary reason is that a single thread cannot finish the handling of analysis in the time between the front and rear two sections of effective EEG data. In synchronous BCI, before the next section of effective EEG data comes, the tasks in $T2$ and $T3$ should be completed in time. This paper focuses on the time factor for ITRs and aims at decreasing $T2$ through concurrent thread mechanisms.

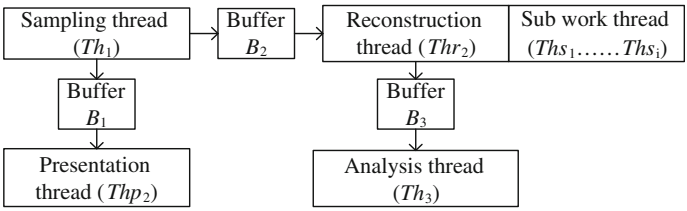


Fig. 7.2 Threading module structure

7.2 Method

Through mechanism of thread concurrency, we decompose a single trial into more subtasks further, allowing the front subtask enter next trial while the rear subtask is executing, and make concurrent subtasks execute among adjacent trials.

Decompose the whole online BCI system into 3 main concurrent modules, with the second module including 2 main threads: the presentation thread and the reconstruction thread. Duration of the presentation thread and the reconstruction thread is our research target. Accelerated rendering is used for drawing raw EEG data in presentation thread, and thread scheduling based on adaptive one-sided fuzzy inference [8] and the mechanism of mutual exclusion and synchronization [9] with semaphore is adopted to recombine intervening data blocks in reconstruction thread.

The thread module structure of the online BCI system is designed as Fig. 7.2. There are four main threads, namely the sampling thread Th1, the presentation thread Thp2, the reconstruction thread Thr2, and the analysis and recognition thread Th3. Besides that, there are more subthreads dominated by Thr2 for fitting better the complicated change of EEG data size.

7.2.1 Accelerated Rendering of Raw EEG Data

In general, EEG acquisition system such as Neuroscan, the raw EEG data is presented in a method of superposition by column from left to right with the ordinary GDI (Graphics Device Interface) drawing [10]. We found that the method of superposition by column only draws in a local area each time; thus it has a relative small workload.

This kind of presentation has two main shortages:

1. Each drawing process contains an operation of clearing the local area so as to cover the data curves of last drawing. Moreover, as the number of EEG sampling channels increases, and the sampling rate speeds up, this phenomenon becomes more serious. Because the GDI drawing is based on CPU [11], more operations means more burdens for CPU.

2. The procedure of superposition by column is continuous in time but not in spatial presentation; from Fig. 7.1, we can see that later EEG data may appear on the left of earlier EEG data. If the EEG data curves can be presented from left to right all the time, then the presentation will look more intuitive.

Based on the two factors proposed, the accelerated rendering technology is an available choice. As most computers support GPU-accelerated rendering currently [12], delivering the presentation to GPU and taking full advantage of the graphics card can contribute to fluent presentation of large-scale EEG data and relieve the burden of CPU to some extent. What is more, this can decrease duration of stage $T2$, and reserve more CPU resources for the analysis process of stage $T3$.

Due to the fixed drawing area in a practical EEG system, a circular buffer of fixed size is feasible, and in the mode of hardware acceleration.

Assume that the sampling rate is S, the EEG channel number and column number of each data block transmitted from the network is N and C, the pixel width and height of the drawing area is W and H, and the duration for the presentation of the whole drawing area is Tp2 (unit: s).

The linguistic descriptions for the procedure of accelerated rendering are as follows:

1. Buffer construction. Construct a circular buffer $B1$ with a length of S*Tp2*N as the frame buffer unit, and initialize the data pointer pHead and pTail to point the head of $B1$; construct a vector buffer with a length of S*Tp2 for the accelerated rendering of graphics card.
2. Matrix transposition. Due to the fact that EEG data of all channels exist continuously in each data block from the network, a matrix transposition is needed so that the EEG data of the same channel can be attached directly, as shown in Fig. 7.2.
3. Data block connection. Now the data block can be attached directly, so put it into $B1$ directly and move the tail pointer at the same time. When the length of EEG data in $B1$ is S*Tp2, transpose it again to a matrix with row N and column S*Tp2.
4. Coordinate mapping. According to the position of the drawing area, map the EEG data values to the screen coordinates and send them to the vector buffer. Then take each row of data in the vector buffer as one unit and draw them to the hardware off-screen buffer; after drawing, send them to hardware frame buffer to finish the accelerated rendering of one row.
5. Pointer handling. When the tail pointer pTail arrives to the tail of $B1$, it automatically moves from the head next time. When pTail exceeds the head pointer pHead, pHead automatically moves to the next data block after pTail.

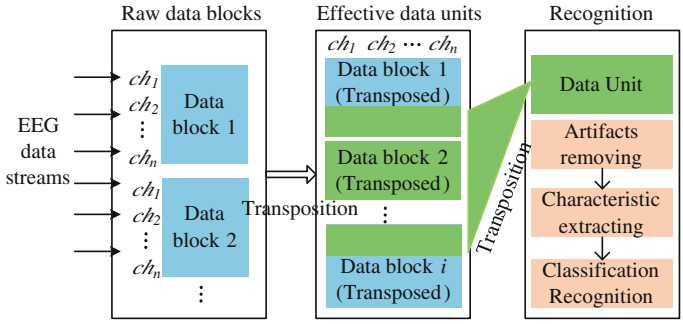


Fig. 7.3 Reconstruction of effective EEG data

7.2.2 Thread Scheduling Against Variable EEG Data

Reconstruction of effective EEG data is a procedure of recombining the data in working status before recognition. The duration of this procedure is marked as Tr2 in Fig. 7.1, and the target of this section is to decrease Tr2.

Assume the column number of effective EEG data unit for one-time recognition is L, due to the uncertain appearing time of event mark value in raw EEG data block, an effective EEG data unit may not start exactly at the beginning of an EEG data block, so the number of raw EEG data block needed to recombine an effective EEG data unit should be $L/C + 1$, in which C is the column number of a raw EEG data block mentioned in Sect. 7.2.1. The flowchart for reconstruction of effective EEG data is shown in Fig. 7.3, in which chi indicates the ith sampling channel.

In Fig. 7.3, the green fill area represents an effective EEG data unit, and the processing steps of reconstruction are as follows:

1. Buffer construction. Construct a second buffer named B_2 for storing an effective data unit, and initialize the data pointer pHead and pTail to point the head of B_2;
2. Matrix transposition. Read the event mark value in the raw EEG data block, and transpose the data block from a matrix with N rows and C columns to another with C rows and N columns.
3. Data blocks recombining. The data blocks transposed can be put in B_2 and stored sequentially. Move the head pointer pHead to the position pos_1 where the event value begins to appear, and the tail pointer pTail to the position $pos_2 = pos_1 + k*C$ when the following kth data block comes, so the length between pHead and pTail is always an integer multiple of that for a data block. When the length between pHead and pTail is L, transpose the matrix of the whole data unit again and use the effective data unit with row N and column L for analysis afterward.
4. Pointer reset. Move pHead and pTail to the head of B_2 and begin the next reconstruction of another effective data unit.

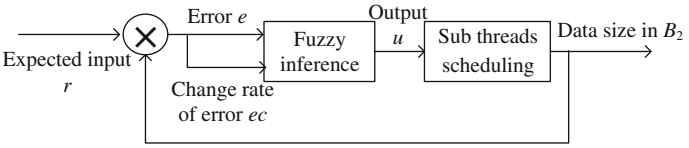


Fig. 7.4 Thread scheduling based on one-sided fuzzy inference

Different sampling rate and the data reading speed will lead to increase and decrease variations of data size. Our experiments show that, due to the accelerated rendering process, the EEG data in buffer $B1$ can always be consumed in time, whereas in buffer $B2$, the EEG data size may generate all kinds of increase and decrease variations.

Predicting the changes of data size effectively, and handling the EEG data flexibly, is the key to prevent the space complexity from increasing, reduce system delay, and improve the efficiency of the system. The online BCI system has a typical environment of time-varying multitasking, and compared with traditional methods, fuzzy inference has a prominent advantage in predicting data variations [13]. Therefore, a method of fuzzy inference is designed to solve the problems of predicting data streams variations in nonlinear environment, as shown in Fig. 7.4. The reconstruction process is fulfilled by the reconstruction thread Thr2 and more subworking threads Ths. The input $r = 0$ indicates the desired data size in buffer $B2$, the error e indicates the current actual data size, the error rate ec = de/dt indicates the change rate of data streams, and u indicates number of desired subworking threads. Data size and the number of subworking threads cannot be negative, as a result, the structure above is a procedure of one-sided fuzzy inference.

Set M as the initial total number of subworking threads, K as the number of the subworking threads inferred by fuzzy inference. Take e, ec, and u as the input and output of fuzzy inference, use positive domain to fuzzily e and u, and use two-sided domain to fuzzily ec, and adopt the triangle membership function and the Mamdani minimax reasoning method to design one-sided fuzzy inference rules, as shown in Table 7.1.

Adopt gravity method for the defuzzification process, and assume the ith fuzzy rule is:

$$R_i: \text{ If } x \text{ is } A_i \text{ and } y \text{ is } \mathrm{B}_i, \text{ then } z \text{ is } \mathrm{C}_i,$$

Table 7.1 One-sided fuzzy inference rules

e	ec						
	NB	NM	NS	Z	PS	PM	PB
Z	Z	Z	Z	Z	PS	PS	PM
PS	Z	Z	Z	PS	PS	PM	PM
PM	Z	Z	PS	PS	PM	PM	PB
PB	Z	PS	PS	PM	PM	PB	PB

where Ai, Bi, and Ci are, respectively, the fuzzy subset of input variables x, y, and output variable z, thus fuzzy set of z can be obtained through (7.3);

$$R_{\mathrm{i}} = (A_i \text{ and } B_i) \rightarrow C_i;\ R = \bigcup_{i=1}^{n} R_i; C^T = (A^T \times B^T) \circ R = \bigcup_{i=1}^{n} C_i^T, \tag{7.3}$$

where A^T, B^T are the fuzzy sets of inputs. The implication operation "$\rightarrow$" fuzzy "and" operator, the synthetic operation "$\circ$" adopts maximum–minimum method, and the minimum, minimum indicate fuzzy "and," "or" operator. Accurate output z_0 can be converted from fuzzy quantity through (7.4).

$$z_0 = \frac{\sum_{i=1}^{n} z_i \int \mu_C(z_i)}{\sum_{i=1}^{n} \int \mu_C(z_i)}, \tag{7.4}$$

where $\int \mu_C(z_i)$ is the area of conclusion membership function for the ith rule, $z_i (i = 1, 2, \ldots, \mathrm{n})$ indicates the center of each conclusion membership function.

For Mamdani fuzzy inference process with double inputs and one output, set the error e and its change rate ec as input, and u as output, considering the influences of quantization factor k_e, k_{ec}, and scaling factor k_u on the system is not monotonous, influences of different stages are distinct and restrict for each other. Dynamic correction factors are adopted to adjust quantization and scaling factors k_e, k_{ec}, and k_u, the adjustment rules are as follows:

$$\begin{aligned} k_e(n+1) &= k_e(n) + \delta_1 \\ k_{ec}(n+1) &= k_{ec}(n) + \delta_2 \\ k_u(n+1) &= k_u(n) + \delta_3 \end{aligned}, \tag{7.5}$$

where $\delta_1, \delta_2, \delta_3$ are, respectively, the dynamic correction factors of the quantization factor k_e, k_{ec}, and scaling factor k_u.

In order to make the adjustment rules play a part in the whole domain, use another variable ε as a threshold value to make the control effect more accurate for small errors, so as to ensure the control accuracy. As $e > 0$, Thr2's online automatic adjustment strategies of parameters are set as follows:

$$\begin{aligned} &\mathrm{Rule}_1 : \text{If } e > 0 \text{ and ec} > 0 \text{ then} \\ &\delta_1 = \Delta k_e, \delta_2 = 0, \delta_3 = -\Delta k_u; \\ &\mathrm{Rule}_2 : \text{If } e > 0 \text{ and ec} < 0 \text{ and } e > \varepsilon \text{ then} \\ &\delta_1 = \Delta k_e, \delta_2 = 0, \delta_3 = \Delta k_u; \\ &\mathrm{Rule}_3 : \text{If } e > 0 \text{ and ec} < 0 \text{ and } e < \varepsilon \text{ then} \\ &\delta_1 = -\Delta k_e, \delta_2 = \Delta k_{ec}, \delta_3 = -\Delta k_u; \end{aligned}$$

In rules proposed above, $\Delta k_e, \Delta k_{\mathrm{ec}}, \Delta k_u$ are, respectively, the minimum increment of k_e, k_{ec}, and k_u, whose values can be set according to actual situations.

7.2.3 Restricted Access to Shared Resources

Concurrent multitasks can make full use of system resources, and improve the performance of online BCI data analysis system [14]. However, influenced by scheduling properties of the operating system [15], intermediate results are not ordered. In order to get true awareness instructions, a reliable mechanism of mutual exclusion and synchronization to ensure the correct recombination of raw EEG data is necessary.

For the analysis process in real-time BCI, the procedure of EEG data processing is that Th1 receives data and put it in $B2$, Thr2 control the scheduling of subthreads Ths, Ths read, recombine EEG data to the form of data units and put them in buffer $B3$. Th3 reads data units from $B3$ and identify them. Direct production–consumption relationship exists between Th1 (P) and Ths ($C1$, $C2$, …, Cn), Ths ($P1$, $P2$, …, Pn) and Th3 (C).

Since Th1, Ths, and Th3 are related to the access of shared resources in $B2$, $B3$, a reliable mechanism of mutual exclusion and synchronization is the key to manage shared resources. Therefore, to solve the mutex and synchronization problem, the system kernel object such as the mutex lock, the event, and semaphore object are adopted.

Mutual exclusion and synchronization access rules on $B2$ are as follows:

1. $B_2 = 0$, all Ths get into synchronous waiting state;
2. $B_2 \neq 0$, one Ths consumes data block in a method of mutual exclusion.

Set up a semaphore object between Thr2 and Ths, and initialize the resource count with 0 and the maximum resource count with M; Set a mutex lock among multiple Ths with its initial state signaled. For thread Thr2, the result K of fuzzy inference determines the implement times of V.

Mutual exclusion and synchronization access rules on $B3$ are as follows:

1 Multiple Ths put data units into B_3 in the correct sequence exclusively;
2 $B_3 = 0$, Th_3 get into the synchronous waiting state.

Set a mutex lock (hMutex2) for multiple Ths with initial state signaled; Set an event object (hEvent2) between Ths and Th3 with initial state nonsignaled; Set two global variables (TR, TW) protected by the mutex as the token numbers to control the correct recombination of intermediate results. Each Ths carries the token number information through a TLS (Thread Local Storage) [16] variable with the same name.

7.3 Results and Discussion

Subject selections: Select graduate students with healthy physical and normal vision correction, and the ratio of male to female is 1:1. Subjects and the EEG acquisition system are in a shielded room. Indoor lighting is darker to reduce the EOG (Electro-Oculogram) artifacts and distraction caused by surrounding environment.

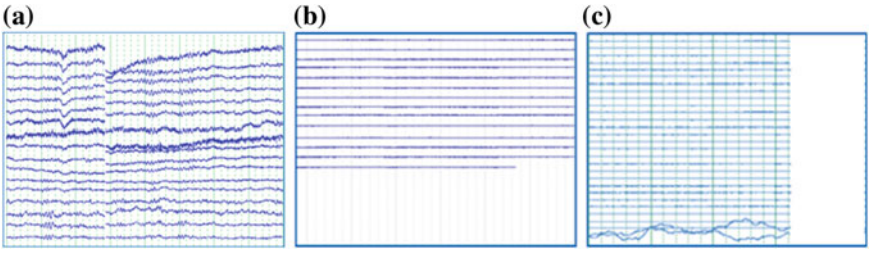


Fig. 7.5 Delay of screen refresh. **a** Neuroscan's normal drawing status. **b** Neuroscan's row drawing delay in repainting. **c** Our client's column drawing delay in repainting with ordinary drawing

System structure configurations: CPU Intel (R) Celeron (R) 2.5 GHz, 2.5 GHz; RAM: 4.00 GB; Operating system platform: Windows 7, 32 bit; Signal stimulation equipment: E-Prime2.0; Data acquisition system: Neuroscan 4.5; Offline analysis environment: MATLAB R2010a.

Experimental paradigms: Stimulation time of each trial: 3 s; Stimulus onset asynchrony: 2 s; Presentation order: random; Use 10–20 system electrode cap with 64 channels, and choose original raw EEG data of C3, C4, FC3, and FC4 channels for multimodal analysis.

We chose six students (male:female = 3:3), and let each one have two online synchronous stimulus experiments. In each one's two experiments, one was with ordinary drawing and thread scheduling, while the other was with accelerated rendering and adaptive thread scheduling. Moreover, the latter experiment's effective EEG data units were immediately saved as the experiment was in progress, so that we could verify the accuracy of online BCI experiment results through offline analysis. Each experiment contains 80 trials.

First, we compared the two kinds of drawing effects and durations of completing a whole window screen repainting.

Figure 7.5 displays the effect of ordinary drawing method. In normal running state, this kind of presentation through superposition by column seems not problematic; however, when encountering the restoring of a window from the minimized state, the window screen began to show an evident sign of stuck drawing process due to tremendous EEG data.

In contrast, Fig. 7.6 shows the accelerated rendering for all kinds of status. Dynamic drawing processes reflect the rapidity and stability of accelerated rendering.

Moreover, the duration of drawing time for a whole window screen is shown in Fig. 7.7 for the six subjects, from which, we can see that the accelerated rendering consumes significantly less time than the other.

Then we set a monitoring window to display dynamic changes of data size in $B2$ and corresponding number of active subthreads, and compared thread scheduling under different conditions, including the ordinary methods just according to the EEG data size in $B2$, and the adaptive methods based on the adaptive one-sided

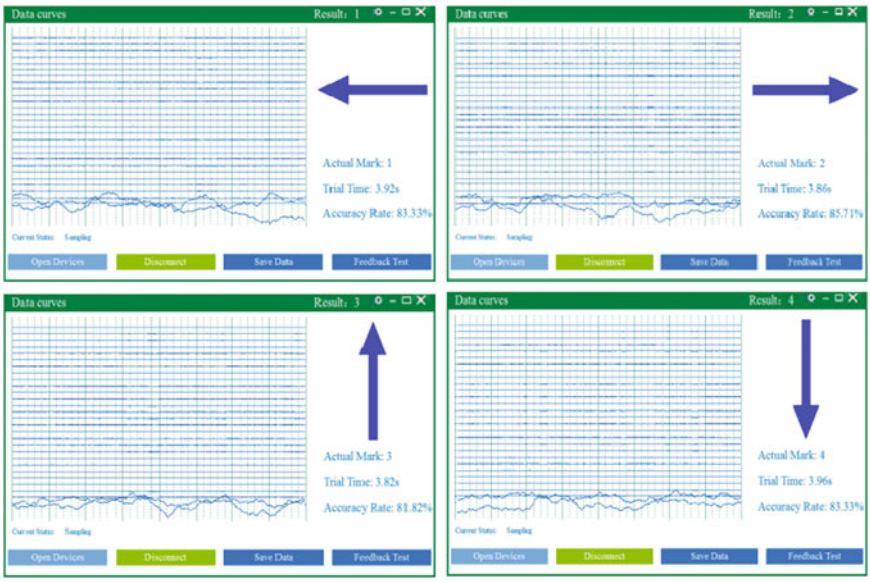


Fig. 7.6 Presentation of accelerated rendering

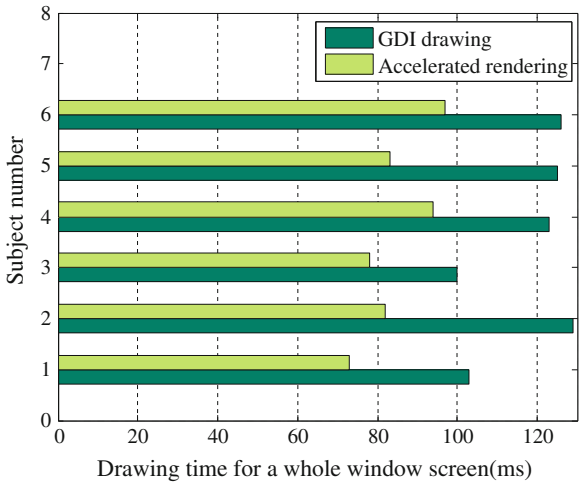


Fig. 7.7 Comparison of two kinds of drawing time for a whole window screen repainting

fuzzy inference according to both the EEG data size and its variation trend. Figure 7.8 shows dynamic variation of data size in $B2$ and number of active subthreads following time under condition of ordinary thread scheduling.

Next, we tested the adaptive fuzzy inference process in the main working thread Thr2, and provided a set of suitable parameters according to the actual situation. The basic parameters include basic domain and fuzzy subset domain of the error e,

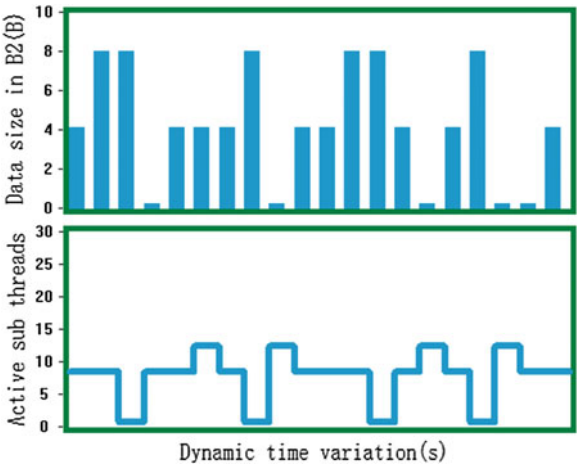


Fig. 7.8 Thread scheduling through ordinary methods

Table 7.2 Setting of domain, quantization factor, and scale factor

Variable	Basic domain	Fuzzy subset domain	Quantization/scale factor
e	[0, 5]	[0, 6]	$k_e = 1.2$
ec	[−20, 20]	[−6, 6]	$k_{ec} = 0.3$
u	[0, 30]	[0, 6]	$k_u = 0.2$

the error rate of change ec, and the output u, and the corresponding quantization, scale factor k_e, k_{ec}, and k_u, as shown in Table 7.2.

The minimum increments of k_e, k_{ec}, and k_u were set as $\Delta k_e = 0.5k_e$, $\Delta k_{de} = 0.25k_{de}$, $\Delta k_u = 0.125k_u$, the switch threshold of error change was set as $\varepsilon = 3$, and the parameters of the fuzzy inference process were adjusted according to Rule1, Rule2, and Rule3 proposed above in Sect. 7.2.2. Afterward, we re-ran the system and observed dynamic variation of data size in $B2$ and corresponding number of active subthreads following time under adaptive fuzzy inference, as shown in Fig. 7.9.

Compare Figs. 7.9 with 7.8, data size in $B2$ changes obviously and K keeps a trend of oscillation for ordinary thread scheduling, whereas on condition of adaptive fuzzy inference, data size in $B2$ becomes relatively stable, and with the increase of K from the initial stage, only some but small overshoot appears, moreover, K is always approaching a stable state following time variation.

At last, we compared the average delay time of the two kinds of online handling methods in each trial. The raw EEG data was saved for offline analysis at the same time. Final statistical results are shown in Fig. 7.10 and Table 7.3, in which the offline recognition time amounts to the serial time of one trial, i.e., $T1 + T2 + T3$, considering some overlaps existing between two adjoining trials on condition of online real-time concurrence, the duration of one trial in online BCI is an average of all trials from the start time to the end time.

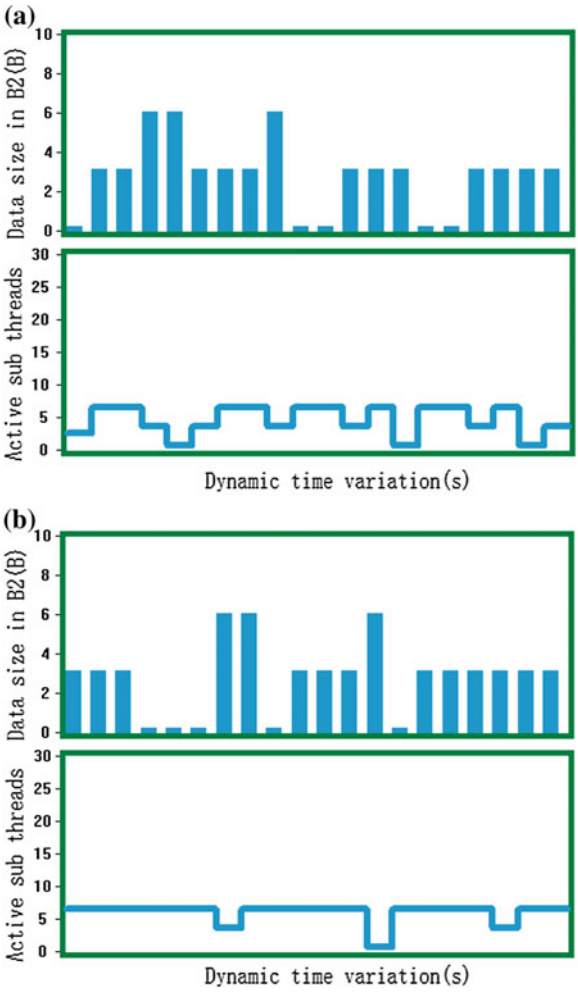


Fig. 7.9 Thread scheduling through adaptive fuzzy inference. **a** System running in the initial stage. **b** System running in the stable stage

Based on the same recognition algorithms, the procedure of offline analysis amounts to implementation of serial trials, thus has the longest time, as can be seen from Fig. 7.10. Concurrence of online BCI data analysis system makes the delay time overlapped for each trial, thus reduces the average delay time of each trial. What is more, from Fig. 7.8 and Table 7.3 we can see that ordinary thread scheduling is less adaptive to high-speed EEG data streams, thus has a relatively low recognition rate due to some losses of effective EEG data, whereas adaptive thread scheduling decreases the average delay time to about 2.1 s with a comparative accuracy rate, which is significant enough to the improvement of ITRs.

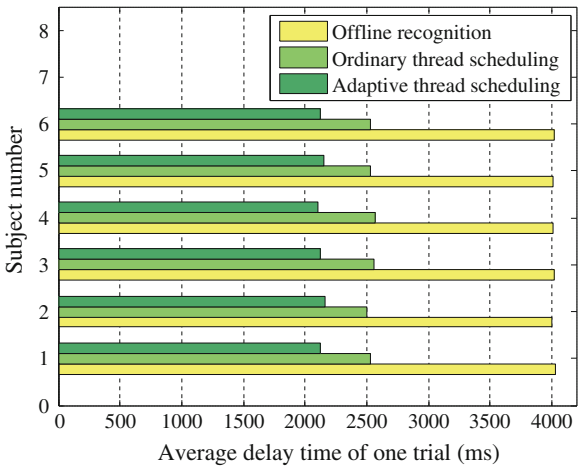


Fig. 7.10 Delay time of one trial under different conditions

Table 7.3 Result comparison for 80 trials

Subject number	Ordinary online recognition		Adaptive online recognition		Offline recognition	
	Correct results	Accuracy rate (%)	Correct results	Accuracy rate (%)	Correct results	Accuracy rate (%)
1	64	80.00	64	80.00	65	81.25
2	61	76.25	65	81.25	66	82.50
3	59	73.75	64	80.00	66	82.50
4	60	75.00	64	80.00	65	81.25
5	59	73.75	66	82.50	67	83.75
6	60	75.00	64	80.00	65	81.25
Average		75.62		80.63		82.08

Figure 7.11 is an overall running status of the whole system showing the event mark value and the result on condition of adaptive thread scheduling, in which the bars indicate the actual event mark values and the curves indicate the recognition results. The whole system runs stably and keeps relatively short delay time from beginning to end.

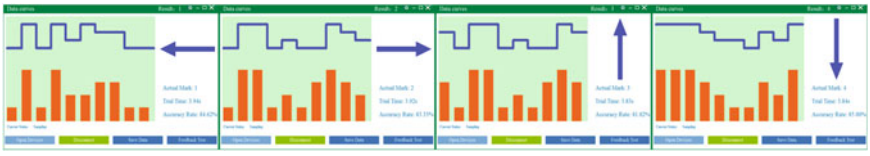


Fig. 7.11 Running status of our system showing the event mark value and the result

7.4 Conclusion

Accelerated rendering and fast reconstruction proposed by this paper is an effective way to improve the rapidity and stability of online real-time BCI system. Through the mechanism of thread concurrency, each module can make full use of the system resources; through accelerated rendering, the presentation of raw EEG data can be more fluent; and through fast reconstruction, effective EEG data unit can be analyzed timely. What is more, thread scheduling based on adaptive one-sided fuzzy inference in the process of reconstruction can enhance the robustness of the whole system for responding to the complicated data streams on condition of different sampling rate. As the statistical results show, integrative design scheme decreases the average delay time of one single trial, and improve the ITRs.

Acknowledgments Fund Project: The National Natural Science Fund (NO.60841004, 60971000, 61172152).

References

1. Allison BZ, Brunner C, Altstätter C et al (2012) A hybrid ERD/SSVEP BCI for continuous simultaneous two dimensional cursor control. J Neurosci Methods 209(2):209–307
2. Geng T, Gan JQ, Hu H (2010) A self-paced online BCI for mobile robot control. Int J Adv Mechatron Syst 2(1–2):28–35
3. Liu T, Yang P, Peng X, Huang Y, Yao D (2009) Real-time brain-computer interface system based on motor imagery. J Electron Sci Technol Chin 7(1)
4. Yu X (2012) Real time brain-computer interface based on alpha rhythms in electroencephalography. J Chongqing Univ Technol (Nat Sci) 26(7):89–93, 2012
5. Henderson A (2010) A design for a middleware communications layer between an industrial robotic arm and the BCI2000 software package. In: Proceedings of Florida conference on recent advances in robotics
6. Chin Z, Ang K, Wang C, Guan C, Zhang H (2009) Multi-class filter bank common spatial pattern for four-class motor imagery BCI. In: Engineering in medicine and biology society, annual international conference of the IEEE, vol 1, pp 571–574
7. Yuan P, Gao X, Allison B et al (2013) A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces. J Neural Eng 10(2) (Article ID 026014)
8. Leite D, Ballini R, Costa P, Gomide F (2012) Evolving fuzzy granular modeling from nonstationary fuzzy data streams. Evolving Syst 3(2):65–79
9. Katz G, Peled D (2008) Genetic programming and model checking: synthesizing new mutual exclusion algorithms, automated technology for verification and analysis. Springer, Berlin, pp 33–47
10. Han L, Kong Q, Yang F, Li W (2012) Visualization of multi-beam bathymetric data based on GDI. Sci Surveying Mapp 4:051
11. Li Q, Hai T (2010) The study on GDI/GDI + rendering function defects and how to avoid them. In: 2010 2nd international conference on Information engineering and computer science. IEEE, pp 1–5
12. Kutter O, Shams R, Navab N (2009) Visualization and GPU-accelerated simulation of medical ultrasound from CT images. Comput Methods Programs Biomed 94(3):250–266

13. Long Z, Liang X, Yang L (2010) Some approximation properties of adaptive fuzzy systems with variable universe of discourse. Inf Sci 180(16):2991–3005
14. Gebhart M, Johnson DR, Tarjan D et al (2012) A hierarchical thread scheduler and register file for energy-efficient throughput processors. ACM Trans Comput Syst 30(2):8
15. Zhuravlev S, Blagodurov S, Fedorova A (2010) Addressing shared resource contention in multicore processors via scheduling. ACM SIGARCH Comput Archit News 38(1):129–142
16. Carribault P, Pérache M, Jourdren H (2011) Thread-local storage extension to support thread-based MPI/OpenMP applications, OpenMP in the Petascale Era. Springer, Berlin pp 80–93

Chapter 8
An HDR Image Encoding Method Compatible with LDR Image Format

Binling Luo, Shuting Cai, Daolin Hu, Shaojia Wen, Ming Yin and Simin Yu

Abstract Because of its excellent performance, the High Dynamic Range (HDR) display becomes more and more popular in many fields. The problem is that most of the display devices available nowadays are still for Low Dynamic Range (LDR) content only. As a result, finding a way to achieve the compatibility between LDR content and HDR content is necessary. In this paper, we provide an improved version of the JPEG-HDR model which allows typical drivers to read and display a tone-mapped image while the HDR information is carried in "sub-band marker". On the other hand, we can recover the HDR image with the help of inverse tone-mapping operator (iTMO) and "sub-band marker". It is the backwards-compatible enhancement to HDR. When an application is not designed for HDR imaging, it displays only the tone-mapped image, allowing the user to have access to part of the content. When an application is designed for HDR imaging, we can recover the original HDR image by decoding process.

Keywords High dynamic range image · JPEG compatible · Inverse tone-mapping operator

8.1 Introduction

In the last two decades, high dynamic range (HDR) imaging has revolutionized the field of computer graphics and other related industry. However, the traditional displays nowadays are not suit for HDR content; in this condition, the transforming between low dynamic range (LDR) and HDR image is necessary. Tone-mapping operator (TMO) adapts the dynamic range of HDR content to suit the lower dynamic range available on a given display; inverse tone-mapping operator (iTMO)

B. Luo · S. Cai (✉) · D. Hu · S. Wen · M. Yin · S. Yu
School of Automation, Guangdong University of Technology, Guangzhou 510006, Guangdong Province, People's Republic of China
e-mail: shutingcai@gdut.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_8

performs an inverse process of tone mapping. In the past few years, many TMOs and iTMOs have been proposed [1].

However, they are one-way transformation, and cannot realize the compatible between LDR and HDR image, which limit their usage in certain displayers. Ward and Simmons [2] proposed a backward compatible JPEG-HDR model, which is an extension of the JPEG compression scheme to HDR images. The main idea is to tone-map HDR images and to encode them using JPEG. Additional information such as Y-residuals and TMO parameters to recover the compressed range are stored in a spatial function. Another attempt on designing HDR encoding scheme that is backward compatible with JPEG was done by Chen et al. [3]. The difference is that Chen adds an iTMO to calculate the residues. This idea was extended by Korshunov et al. [4, 5], they provided an algorithm that allow one to have both lossless and lossy compression, their method accommodate different TMOs and such a decoding scheme that allows displaying the same HDR image differently, depending on its content and given context, besides.

In this paper, we provide an extended version for the scheme mention above. We agree the idea of having JPEG backward compatible HDR compression and format, but save R, G, B channels residual to improve the image quality instead of only saving Y channel residual which employed in the above-mention methods. Another main difference of our method is that we propose a new iTMO algorithm aiming to minimize the residual; a well-designed iTMO will lead to very efficient compression and improve the image quality.

The following sections will describe our method in details. Section 8.2 introduces our method to extend JPEG-HDR model in details. We will present the experiment results in Sect. 8.3, there will be a comparison between our algorithm and JPEG-HDR model. Finally, we will give a brief conclusion for our algorithm.

8.2 Our Extended Version for JPEG-HDR Model

In this paper, we provide an extended version of JPEG-HDR model which allows standard device to read and display a tone-mapped image, while the HDR information is carried in a "sub-band marker."

The first step of our method is to adapt the dynamic range of HDR image to suit the traditional display, this process is called tone mapping. Then we read in the tone-mapped image and run iTMO to produce a new HDR image. Subsequently, we calculate the R, G, B differences between the original and the new HDR image, they are residuals. The residuals can significantly improve the quality of the result when we recover the HDR image. Residuals are compressed in JPEG2000 formats and embed into tone-mapped file which was in JPEG standard. Besides, the related TMO parameter is save in this JPEG file to enable the subsequent decoding process. These are the encoding process of our method. At this point, our JPEG file is marked with the HDR information and can be presented on traditional displayers.

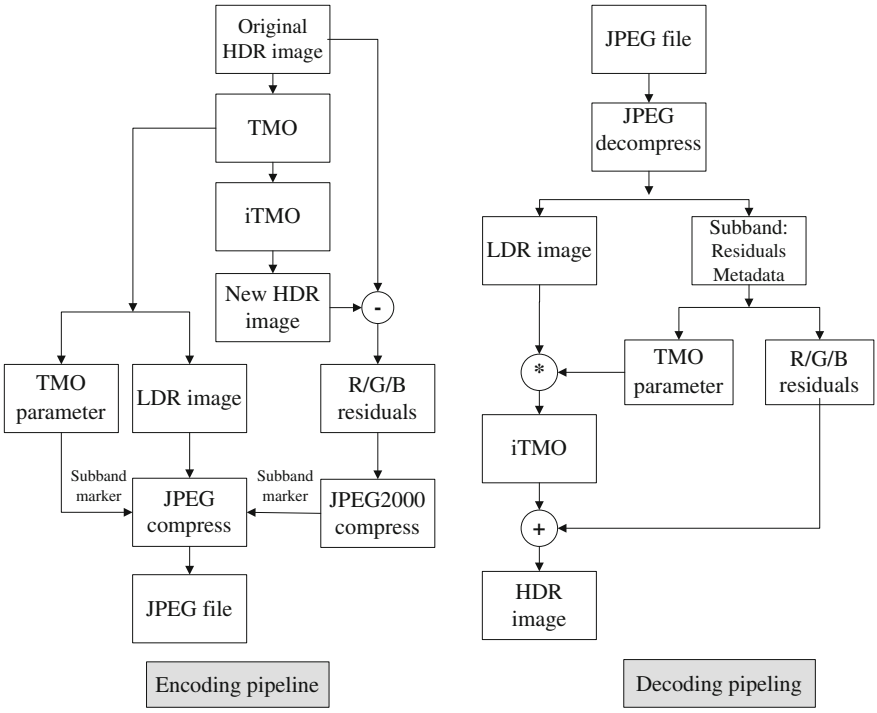


Fig. 8.1 Our extended version for JPEG-HDR model

When we want to recover HDR image, we first decompress JPEG file to read LDR image, as well as the additional information: Residual data and TMO parameter. We run iTMO process with the help of TMO parameter, and then we get a new HDR image. We add the residuals to the new HDR image to improve the quality of the final image. We show our framework as Fig. 8.1.

8.2.1 Encoding

Encoding process aims at producing tone-mapped image which allows typical drivers to read and display, while the HDR information is carried in "sub-band marker." The image is stored in JPEG formats. It takes much less space than traditional HDR content, so it is well-suited to sending by or posting on the web.

8.2.1.1 Tone-Mapping Operator

The encoding starts with tone-mapping HDR image. In our study, we select bilateral filter TMO which is proposed by Durand and Dorsey [6] since it can

preserve local contrast without losing fine details; and according to the study [2] conduct by Ward et at., bilateral filter TMO performs better than histogram adjustment TMO [7], global photographic TMO [8], and gradient operator [9] for compression purpose.

To begin with, LDR image is decomposed into luminance and chromaticity. At this point, use a bilateral filter to separate LDR image into high-frequency and low-frequency parts, which are called the detail layer and the base layer, respectively. We apply exponential TMO [1] to adapt the dynamic range of the base layer since it deals with medium dynamic range content reasonably well and also its simplicity. Exponential operator is defined as:

$$L_d(x) = 1 - \exp\left(-\frac{k \cdot L_w(x)}{\text{Lwa}}\right) \tag{8.1}$$

$$\text{Lwa} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(L_w(x_i) + \varepsilon)\right), \tag{8.2}$$

where, x_i is the ith pixel of the picture, N is the number of pixels in the image, L_d is the luminance of LDR image, L_w is the luminance of HDR image, $k \in (0, 1]$ is a user parameter, and Lwa is the geometric average of HDR image. ε is a small nonnegative value. The last step is to combine tone-mapped base layer, detail layer, and chromaticity to form a LDR image.

8.2.1.2 Inverse Tone-Mapping Operator

Inspired by bilateral filter TMO, we provide an iTMO algorithm base on frequency separation with the goal of preserving edges and local contrast. This is achieved by computing in the frequency domain instead of the spatial domain. We show the pipeline in Fig. 8.2.

To start with, LDR image is decomposed into luminance and chromaticity. The luminance channel passes a bilateral filter, obtaining a low-frequency part, which is called base layer. The detail layer is the high-frequency part, which is calculated by dividing the luminance channel by the base layer. For the base layer, we inverse Eq. (8.1) directly to expand the dynamic range, now, we have formula (8.3) for the enhancement:

$$L_w(x) = \text{Lwa} \cdot \frac{Ln(1 - L_d(x))}{-k}, \tag{8.3}$$

where Ln is the natural logarithm symbol, $k \in (0,1]$ is a user parameter, and Lwa is the geometric average of HDR image. We want to solve $L_w(x)$ from LDR image, there is an unknown parameter: Law. However, we can calculate Lwa value by

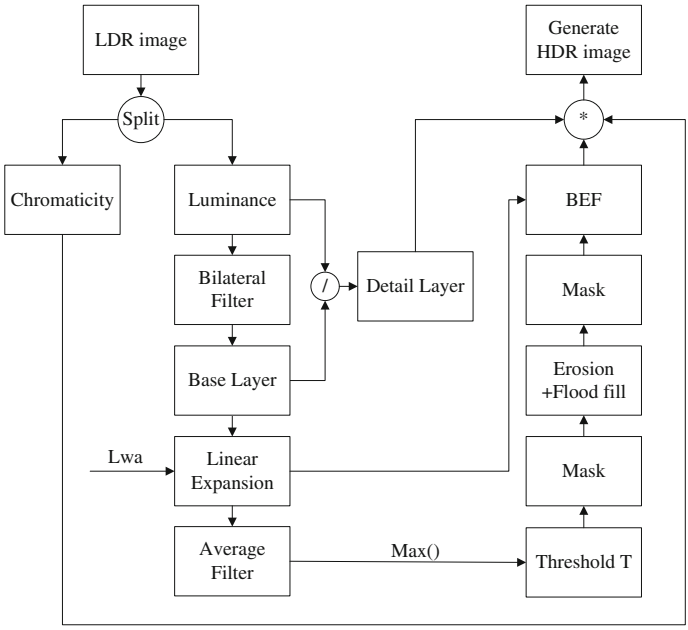


Fig. 8.2 The pipeline of our iTMO

formula (8.2) in TMO step. Lwa was set to global variable and saved in the program; by these way, we can read Lwa value when we run iTMO in encoding process.

Since the truncated information in very bright regions of LDR image usually lost, as a solution, we apply a brightness enhancement function (BEF) in these areas. Rempel et al. [10]. have proposed an automatic approach for computing BEF. They applied a threshold for LDR image to generate a binary mask M1 (the value of the pixel set to one means it is a very bright pixel). A flood-fill algorithm is applied to M1 until the edge is reached (the edge is defined by an edge stopping function). Our work is base on that of Rempel's but more robust. The threshold for our binary mask M1 is computed based on the characteristics of the image instead of a fixed value for all images. Usually, an indoor picture and an outdoor picture need different thresholds. To calculate the threshold T, an average filter with size m = max (width, height)/250 + 1 is applied to the luminance channel and then we calculate the maximum luminance T. T is the threshold we defined. Actually, this method has already been used to divide specular regions and diffuse regions in image processing [11]. Mask M1 defined the very bright regions; in these regions, we enhance the luminance value by [0, a].

Finally, we combine the expanding base layer, the detail layer, and the chromaticity by multiplying each one to form a new HDR image.

8.2.1.3 Calculate and Compress Residuals

There is deviation between the original and new HDR image, we call it residuals. The original HDR image's R, G, B value subtract that of the new HDR image's can calculate the residuals. According to the study [4] conducted by Korshunov et at., JPEG2000 can achieve a better compression efficiency compared with the compression schemes capable of encoding HDR images; so we compress the R, G, B residuals in JPEG2000 standard and take it as the header marker of our JPEG file.

8.2.1.4 JPEG Compression

In our study, we use JPEG standard to compress the tone-mapped image. Residuals and TMO parameter: Lwa are embed into the header of the JPEG file. These markers enable the decoding process (Fig. 8.3).

8.2.2 Decoding Process

The decoding process is to recover the HDR content from JPEG file, by which we can present the image on HDR displayer. To start with, we decompress JPEG file to get LDR data and header marker: Residuals and Lwa. Residuals are the difference between the original and new HDR data after quantization. Lwa is a TMO parameters, they are needed to expand the signal back. The iTMO program which was introduced in Sect. 8.2.1.2 is run for a second time, and then we got a new HDR image. We add the residuals to the new HDR image to complement the deviation. Residuals can significantly improve the quality of the final image because the spatial quantization and bit reduction can introduce quantization errors, and this can be noticed in the form of noise, enhancement of blocking and ringing, banding artifacts, etc. By this way, we can recover the HDR image reasonably well.

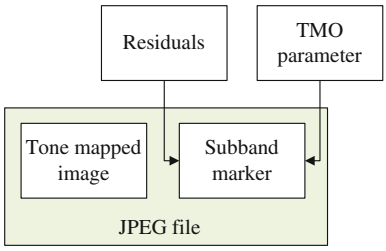


Fig. 8.3 Our JPEG file

8.3 Experiments and Evaluations

We conduct the experiment with Microsoft Visual Studio 2010 and MATLAB R2012b. We validated our algorithm in PSNR value compare against JPEG-HDR. Table 8.1 is our experiment result. In which, rate is the ratio of desired image size to the raw image size and nbits is to set the number of bits per pixel in Jasper.

From Table 8.1, we can see our algorithm produces overwhelming results in PSNR value for all images. Comparing to JPEG-HDR model, our algorithm increases PSNR value by 10–37 dB. Besides, it consumes a small amount of memory, reduce the data size of the original image by 40–94 %.

Figure 8.4 presents the images used in our study, Since HDR images cannot be showed here naturally due to the limitations of the print medium, we show the normalized HDR image by colorizing the luminance levels; we arrange the blue/cyan/green/yellow/red colors in ascending order to represent the luminance level 0.00/0.25/0.50/0.75/1.00, respectively, in which red color presents the maximum luminance level and blue color presents the minimum luminance level.

Table 8.1 The comparison between our method and JPEG-HDR model

Image	Compression	Data size (bytes)	PSNR R (dB)	PSNRG (dB)	PSNR B (dB)	Average (dB)
Room	Original HDR image	7,438,076				
	JPEG-HDR (quality = 100)	2,262,404	46.896	37.420	41.402	41.906
	Our method Rate = 1, nbits = 16 bits	2,452,413	72.458	72.217	70.416	71.697
	Our method Rate = 1, nbits = 12 bits	1,148,431	62.691	61.499	50.545	58.245
	Our method rate = 0.8, nbits = 12 bits	379,713	51.742	54.058	52.240	52.680
Bridge-studio	Original HDR image	19,376,169				
	JPEG-HDR (quality = 100)	5,784,513	30.526	45.557	35.717	37.267
	Our method rate = 1, nbits = 16 bits	11,407,576	74.514	74.087	75.641	74.748
	Our method rate = 1, nbits = 12 bits	5,504,541	71.444	71.913	68.338	70.565
	Our method rate = 0.8, nbits = 12 bits	1,594,575	65.319	66.170	64.941	65.477

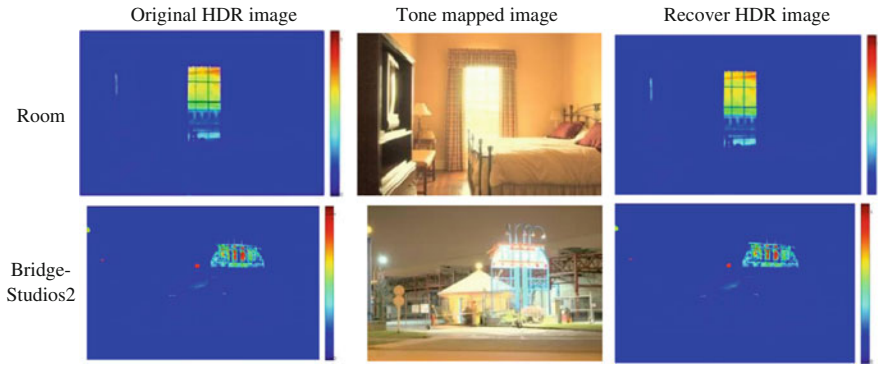


Fig. 8.4 Related images

8.4 Conclusion

In this paper, we propose an effective compatible model which is an extended vision for JPEG-HDR. The novelties of our method are that we provide an effective iTMO and calculate R, G, B residuals to improve the image quality. The method is backward compatible because the additional information is encoded using extra application markers of JPEG format. When an application is not designed for HDR imaging, it displays only the tone-mapped image. When an application is designed for HDR imaging, we can recover it by decoding process. Experiment result proves that our algorithm has better performance than JPEG-HDR.

Acknowledgments This work is supported by NSF China under grants No. 61201392, 61025017, 61172023; by the Specialized Research Foundation of Doctoral Subjects of Education Ministry under Grant 20114420110003

References

1. Banterle F, Artusi A, Debattista K, Chalmers A (2011) Advanced high dynamic range imaging. A K Peters Ltd, Massachusetts
2. Ward G, Simmons M (2004) Sub-band encoding of high dynamic range imagery. In: APGV'04: Proceedings of the 1st symposium on applied perception in graphics and visualization, ACM Press, New York, USA, pp 83–90
3. Chen M et al (2006) JPEG compatible coding of high dynamic range imagery using tone mapping operators. In: Picture coding symposium, pp 22–28
4. Korshunov P, Ebrahimi T (2012) A JPEG backward-compatible HDR image compression. In: SPIE/hdr mmspg
5. Korshunov P, Ebrahimi T (2013) Context dependent JPEG backward compatible high dynamic range image compression. SPIE/Optical Eng 52(10)
6. Durand F, Dorsey J (2002) Fast bilateral filtering for the display of high-dynamic-range images. ACM Trans Graph 21(3):257–266

7. Larson GW, Rushmeier H, Piatko C (1997) A visibility matching tone reproduction operator for high dynamic range scenes. IEEE Trans Visual Comput Graphics 3(4):291–306
8. Reinhard E, Stark M, Shirley P, Ferwerda J (2002) Photographic tone reproduction for digital images. ACM Trans Graph 21(3):267–276
9. Fattal R, Lischinski D, Werman M (2002) Gradient domain high dynamic range compression. ACM Trans Graph 21(3):249–256
10. Rempel AG, Trentacoste M, Seetzen H, David Young H, Heidrich W, Whitehead L, Ward G (2007) LDR2HDR: on-the-fly reverse tone mapping of legacy video and photographs. ACM Trans Graph 26(3):39
11. Meylan L, Daly S, Süsstrunk S (2006) The reproduction of specular highlights on high dynamic range displays. In: IST/SID 14th color imaging conference, Scottsdale, AZ, USA, pp 333–338

Chapter 9
A Prediction Method for Wind Speed Based on the Correlation Analysis of Measured Data of Adjacent Wind Turbine

Yinsong Wang and Ziqing Su

Abstract Wind speed prediction is very important for the control of wind power generation. Former studies for wind speed prediction are limited to measuring their own historical data to achieve the future wind speed prediction. Compared with this situation, this paper presents a prediction method to get the specified point's wind speed predicting value from the adjacent point's historical wind speed data. Using Pearson correlation coefficient to show the relativity between the adjacent measuring points, the method of regression analysis based on correlation coefficient is introduced, and we chose the wind speed data of ten points from an actual wind field to simulate and verify this method. Simulation result shows that this method can guarantee the accuracy of wind forecasting, and it can improve the redundancy and reliability of wind measuring equipments effectively.

Keywords Wind speed forecast · Pearson correlation coefficient · Regression equation · Time-series model

9.1 Introduction

In recent years, wind power generation developed rapidly, and to predict wind speed accurately is very important for the allocation of power grid resources [1]. Meanwhile, wind speed forecast is also essential for the advance adjustment of the pitch angle and the maximum wind power tracking [2].

An erratum chapter can be found under DOI 10.1007/978-3-662-46469-4_64

Y. Wang (✉) · Z. Su
Tsinghua University, Beijing, China
e-mail: 920028274@qq.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_9

The literature [3] put forwards a kind of method to predict the short-term wind speed by using time-series model. The literature [4] selected the Pearson correlation coefficient as basis to measure the relevance between different wind speed series, and put forward a prediction method using the artificial neural network. But it did not compare the predicted results with the results of other methods, and cannot show the feasibility of that method very well.

The literature [5] pointed out that the rotating components of cup anemometers used in wind farms are easy to damage, and cannot overcome the dust and strong wind and other inclement weather, this will affect the accuracy and reliability of the measurements of wind speed. So if we can predict the specified wind speed using the wind speed data which is related to it, then we can improve the redundancy and reliability of wind speed measuring devices largely.

This paper presents a method of regression analysis based on Pearson correlation coefficient, and selects the wind speed data of ten neighboring measurement points of a wind farm in the Netherlands. At last, we use the model proposed in this paper to predict the wind speed of the specified point.

9.2 The Correlation of Wind Speed

9.2.1 Pearson Correlation Coefficient

Pearson correlation coefficient is a reflection of the degree of correlation between the two linear variables, if we choose v_b as the wind speed of the adjacent wind power generator, v_t as the wind speed of the specified wind power generator which is to be predicted. The Pearson correlation coefficient between them can be defined as [4]:

$$r_{v_b,v_t} = \frac{n\sum v_{b_i}v_{t_i} - \sum v_{b_i}\sum v_{t_i}}{\sqrt{n\sum v_{b_i}^2 - \left(\sum v_{b_i}\right)^2}\sqrt{n\sum v_{t_i}^2 - \left(\sum v_{t_i}\right)^2}} \tag{9.1}$$

n is the sample size and r is the coefficient of correlation between the two variables.

Pearson correlation coefficients can be used to measure the linear relationship between two variables. The value of $|r_{v_bv_t}|$ is between 0 and 1, when $0.7 \le |r_{v_bv_t}| < 1$, it can be considered there is a high linear correlation between the two wind series.

9.2.2 Correlation Between the Measured Wind Speed

This article chose ten adjacent turbines' wind speed data from a wind farm in the Netherlands, and randomly selected the No. 2 machine to be the specified generator; we selected the 50 measuring points' and the 120 points' historical data to calculate

Table 9.1 Pearson correlation coefficients between the nine other generators and the No. 2 generator

Turbine no.	50 sampling points	120 sampling points
1	**0.9126**	**0.8799**
2	**1.0000**	**1.0000**
3	0.8597	0.8009
4	0.9013	0.8532
5	0.8550	0.7989
6	**0.9043**	**0.8680**
7	0.8973	0.8520
8	0.8230	0.7903
9	0.8532	0.8021
10	0.7826	0.7373

Pearson correlation coefficients between the nine other generators and the No. 2 generator, the results are shown in Table 9.1.

Seen from the table, in the two set of data, the correlation coefficients of the No. 1 generator are both the highest, and the No. 6 generator's are the second. So we can consider the correlation among No. 1, No. 2, and No. 6 generators are closer. In the future prediction, we will choose the No. 1, No. 2, and the No. 6 generators to be the main objects of study.

9.3 Regression Models

9.3.1 Regression Equation Based on the Correlation

Correlation coefficient has a wide application in the regression analysis of the data sequence [6]. For the wind sequences studied in this paper, when r_{v_b,v_t} is high enough, we can fit out wind sequence V_T from V_B by using regression equation. We suppose the sample size to be n, then the regression equation between wind sequence V_T and V_B can be shown as:

$$\hat{v}_t = \hat{\sigma} + \hat{\lambda} v_b \tag{9.2}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} (v_{b_i} - \bar{v}_b)(v_{t_i} - \bar{v}_t)}{\sum_{i=1}^{n} (v_{b_i} - \bar{v}_b)^2} \tag{9.3}$$

$$\hat{\sigma} = \bar{v}_t - \hat{\lambda} \bar{v}_b \tag{9.4}$$

$$\bar{v}_b = \frac{\sum_{i=1}^{n} v_{b_i}}{n} \tag{9.5}$$

$$\bar{v}_t = \frac{\sum_{i=1}^{n} v_{t_i}}{n} \tag{9.6}$$

$\hat{v}_t$ is the approximate wind speed of the specified wind generator which is fitted out of the adjacent generator's actual wind speed v_b. $\hat{\sigma}$ and $\hat{\lambda}$ are the two regression coefficients, $\bar{v}_b$ and $\bar{v}_t$ are the average values of the two wind series.

9.3.2 Alternative Wind Sequence

We suppose that V_B is known and we cannot get V_T through direct measurement, so we cannot fit the wind series $\hat{V}_T$ (hereinafter referred as hereinafter referred to as "alternative wind sequence") directly. This is mainly because we need to use $\bar{v}_t$ when we are calculating $\hat{\sigma}$ and $\hat{\lambda}$.

But wind sequence is also a random sequence, the linear relationship between two adjacent measuring points' wind sequences only depends on the correlation coefficient, and it will not change as the time changes. And the linear relationship between two sequences only depends on $\hat{\sigma}$ and $\hat{\lambda}$, for the two wind sequences which have the higher correlation coefficients, the results of calculating $\hat{\sigma}$ and $\hat{\lambda}$ with wind speed data of different times should be almost the same.

So we can calculate $\hat{\sigma}$ and $\hat{\lambda}$ by feeding the two adjacent historical data into formula (9.3) and (9.4). Then we feed the data of adjacent point wind sequence V_B into formula (9.2) to get the alternative wind sequence $\hat{V}_T$. Thus, we can replace V_P with $\hat{V}_T$ to predict the specified point's wind speed with time-series model.

9.4 Wind Speed Prediction Based on Autoregressive and Moving Average Model

The Autoregressive and moving average model, which is also called the ARMA model, is a usual kind of time-series model used in prediction. The general form of ARMA (p, q) model can be expressed as [7]:

$$Y_t = \sum_{j=1}^{p} \omega_j Y_{t-j} + e_t - \sum_{j=1}^{q} \xi_j e_{t-j} \tag{9.7}$$

$\omega_1, \omega_2 \cdots \omega_p, e_t, \xi_1, \xi_2 \cdots \xi_q$ are the model parameters, e_t is a white noise signal.

Since the order (p, q) is generally low, so we choose the trial method to determine the order of the model. Select the (p, q) value one by one from low to

high, then estimate the parameters, and check whether the model is accepted. If not, readjust the value of (p, q) and repeat the process until the model is accepted.

After several attempts, we found that when p is taken as 3 and q is taken as 2 the model has the highest predictive accuracy, so we select the ARMA (3, 2) model.

After the order is determined, the next step is to estimate the parameters of which the most commonly used method is least squares estimation, the principle is described as follows. The wind speed model can be described as [8]:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_n x_{in} + e_i \tag{9.8}$$

$y_1, \cdots, y_n$ are the observed data, the arguments $x_{i1}, \cdots x_{in}$ are all known, $\beta_1, \cdots \beta_n$ are the parameters to be estimated, e_i is the error, the matrix form is

$$\begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} x_{12} \cdots x_{1n} \\ x_{21} x_{22} \cdots x_{2n} \\ \\ x_{n1} x_{n2} \cdots x_{nn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \\ e_n \end{bmatrix} \tag{9.9}$$

If you make the error sum squares [8]:

$$Q(\beta) = Q(\beta_1, \beta_2, \cdots, \beta_n) \sum (y_t - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_n x_{in})^2 = \sum e_i * e_i$$

minimum, it is called the least squares parameter estimates at this time.

9.5 Simulation Analysis

At last, we selected the measured data from a wind field in Netherlands (http://www.knmi.nl) and used MATLAB to simulate and test the prediction model. In order to prevent the difference in numbers of sampling points to bring errors, this paper selected 50 sampling points and 120 sampling points, respectively, and calculated the Pearson correlation coefficients between each wind turbine. We chose the No. 2 generator to be the selected generator, as shown in Table 9.1, the correlation between No. 1, No. 2, and No. 6 generator is stronger, so we chose the No. 1, No. 2, and No. 6 generator to be the main objects to study.

Then input the wind speed data, and selected the No. 2 generator as the reference and calculated the alternative wind sequence $\hat{V}_T$ between each generator and the No. 2 generator. $\hat{\sigma}$ and $\hat{\lambda}$ are calculated with the historical data over the same period last year of each wind generator and the No. 2 generator. When predicting, we chose the ARMA (3, 2) model to predict the No. 2 generator's wind speed one step further, and selected the relative mean error (MRE) as the evaluation index to check the prediction results. Tables 9.2 and 9.3, respectively, reflected the prediction errors of choosing 50 sampling points and 120 sampling points.

Table 9.2 Prediction errors of 50 sampling points

Turbine no.	MRE (%)	Correlation coefficient
1	**4.56**	**0.9126**
2	**3.18**	**1.0000**
3	5.40	0.8597
4	4.98	0.9013
5	5.67	0.8550
6	**4.79**	**0.9043**
7	5.43	0.8973
8	5.79	0.8230
9	5.98	0.8532
10	7.89	0.7826

Table 9.3 Prediction errors of 120 sampling points

Turbine No.	MRE (%)	Correlation coefficient
1	**5.93**	**0.8799**
2	**4.21**	**1.0000**
3	6.81	0.8009
4	6.39	0.8532
5	7.39	0.7989
6	**6.16**	**0.8680**
7	7.37	0.8520
8	7.61	0.7903
9	7.53	0.8021
10	9.86	0.7373

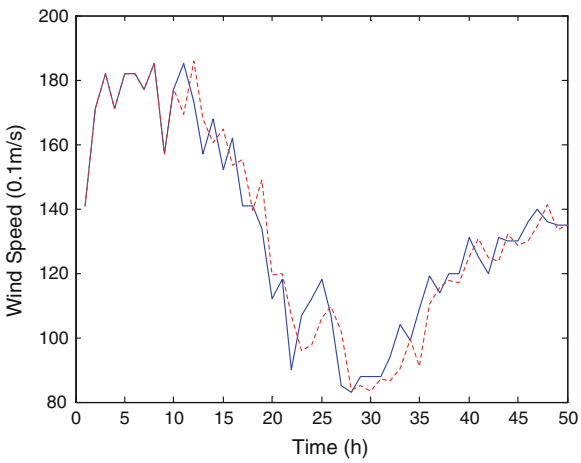


Fig. 9.1 Predicted *curves* of the traditional method (50 sampling points)

Through comparing the data in the table, we can see that the prediction errors of each wind generator reduce with the increase of the correlation coefficient. Meanwhile, the prediction from the No. 1 generator to the No. 2 generator among which has the highest correlation coefficient has the minimum error.

At the same time, we compared the prediction result of the traditional time-series model with the results of the prediction method put forward in this paper. Figures 9.1 and 9.2 show the prediction results of the traditional time-series model, the blue solid line represents the actual wind speed curve of the No. 2 generator and the red dotted line represents the predicted wind speed curve.

Figures 9.3 and 9.4 are the prediction curves of the No. 1 generator and Figs. 9.4 and 9.5 are the prediction curves of the No. 6 generator. Through comparing the prediction results of the method put forward in this paper and the results of the traditional time-series model, we can find that traditional method has the smaller

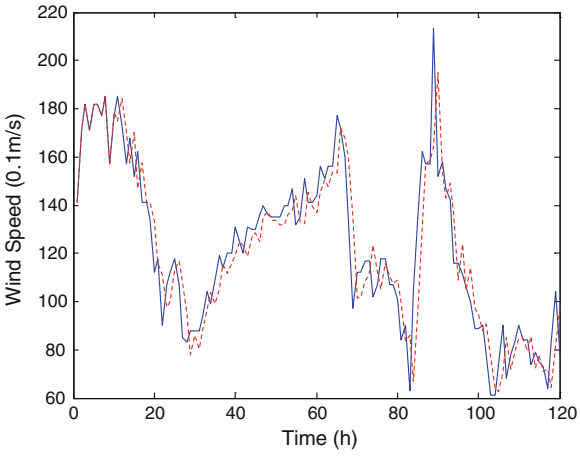


Fig. 9.2 Predicted *curves* of the traditional method (120 sampling points)

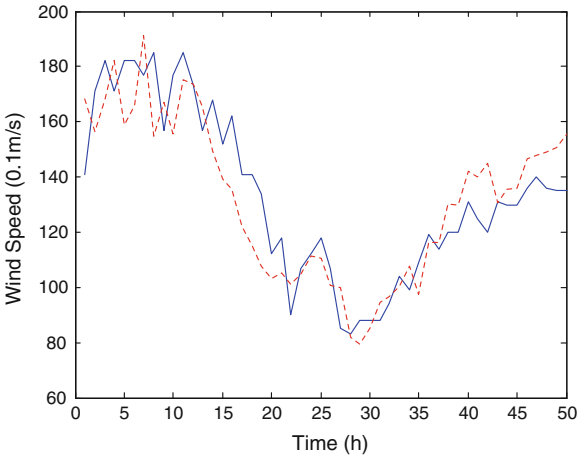


Fig. 9.3 Predicted *curves* from the No. 1 generator to the No. 2 generator (50 sampling points)

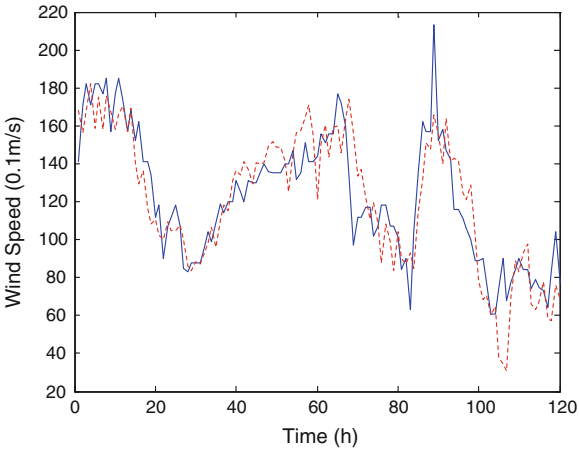


Fig. 9.4 Predicted *curves* from the No. 1 generator to the No. 2 generator (120 sampling points)

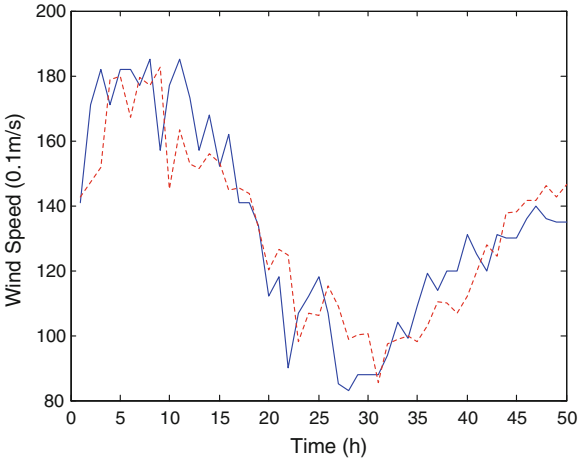


Fig. 9.5 Predicted *curves* from the No. 6 generator to the No. 2 generator (50 sampling points)

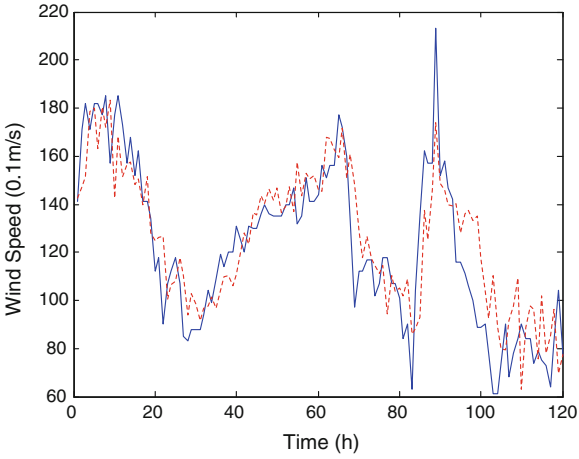


Fig. 9.6 Predicted *curves* from the No. 6 generator to the No. 2 generator (120 sampling points)

error, respectively, are 3.18–4.21 %. But the prediction error of our method is just about 2 % higher than the traditional time-series model, and this is completely acceptable (Fig. 9.6).

9.6 Conclusion

Simulation results show that the method to replace the V_T with the alternative wind sequence $\hat{V}_T$ is feasible. Then we compared the results of the method put forward in this paper with the results of traditional time-series model, and found there is not very big difference between the prediction errors of the two methods. So we can improve the redundancy and reliability of wind measuring equipments without extra investment.

References

1. Yang X, Sun H (2011) The study of time series model wind speed prediction. Power Eng 31 (3):203–208 (in Chinese)
2. Guo P (2010) Autoregressive moving average model to predict wind speed maximum wind energy strategy. North China Electr Power Univ 37(2) (in Chinese)
3. Shao P, Sun Y, Liang L (2008) Short-term wind farm wind speed prediction based on ARMA model. Grid Clean Energy 24(7):52–55 (in Chinese)
4. Kusiak A, Li W (2010) Estimation of wind speed: a data-driven approach. J Wind Eng Ind Aerodyn 98(10):559–567
5. Yan R, Jiang C (2011) Study of wind speed and direction sensor. Inner Mongolia Petrochemical Ind 21:019 (in Chinese)
6. Jiang S (1999) Correlation coefficients used in the regression analysis. Shanghai Univ Electr Power 15(1):34239 (in Chinese)
7. Bao N (2007) Wind power gas turbine complementary several key problems in power system research. Tsinghua University (in Chinese)
8. Zheng C (2009) Two ARMA model parameter estimation methods and residual model application. Yanshan University (in Chinese)

Chapter 10
A Novel Method Based on Data Visual Autoencoding for Time-Series Classification

Chen Qian, Yan Wang and Lei Guo

Abstract A variety of techniques based on numerical characteristics are currently presented for mining time-series data. However, we find that time-series data generally contain curves sharing some set of visual characteristics and features. These characteristics offer a deeper understanding of time-series data, and open up a potential new technique for time-series analysis. Particularly beneficial from recent advances in deep neural networks, representations and features can be automatically learnt by deep learning architectures such as autoencoders. Based on that, our work proposes a novel method, named time-series visualization (TSV), to efficiently detect visual characteristics from curves of time-series data and use these characteristics for intelligent analysis. Architecture and algorithm of TSV based on stacked autoencoders are introduced in this paper. Further, important factors affecting the performance of TSV are discussed based on empirical results. Through empirical evaluation, it is demonstrated that TSV has better efficiency and higher classification accuracy on analyzing the datasets with significant curve feature.

Keyword Time series · Autoencoder · Classification · Input dropout · TSV

10.1 Introduction

In the last decade, interest in mining time-series data is like an explosion which, in turn, resulted in lots of researches proposed to introduce new techniques to index, classify, cluster, and segment time series. However, most of these techniques have limited performance because the form of time-series data is inconstant but their focus is mainly on numerical characteristics of data.

Similarity measure is one of the most important ways toward mining time-series data. The most straightforward similarity measure for time series is the Euclidean

C. Qian · Y. Wang (✉) · L. Guo
School of Automation Science and Electrical Engineering,
Beihang University, Beijing 100191, People's Republic of China
e-mail: w-yan@buaa.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_10

Distance (ED) [1], which has two advantages: Linear complexity and parameter-free. However, ED is quite sensitive to noise and misalignments which means it is unable to handle the time-shifting series. Inspired by this motivation, Berndt and Clifford [2] introduced dynamic time warping (DTW) which can be used to measure the similarity between time series with local shifts. However, DTW is too slow to be of practical use, even though it provides good measuring accuracy [3]. Thus, many methods have been proposed based on DTW to improve the efficiency of DTW [4–6]. In addition, longest common subsequences (LCSS) was proposed based on the model introduced by André-Jönsson and Badal [7], which was another group of similarity measures. Other famous examples for this category include edit distance on real sequence (EDR) [8] and edit distance with real penalty (ERP) [9]. However, most of these similarity measures focus on numerical characteristics, which make them quite sensitive to changes of time-series data.

Interestingly, for human, it is intuitional to identify the similarity of time series through curves rather than the real data. Based that, we think if it is possible to get a good representation of time series from the curves and use it for time-series analysis. Fortunately, autoencoder provides a potential way to achieve that autoencoder is a learning circuit to encode the inputs into some representations that are as close as possible to outputs [10]. It was first proposed by Hinton and his group in the 1980s, but with the recent revival of interest in "deep networks," [11] autoencoder is coming back to the center stage. We believe that autoencoder and human visual system are quite similar in some aspects [12]; and in this paper, we try to construct a new method named time-series visualization (TSV) for time-series classification based on good representations learnt from curves of time-series data. First, representations can be learnt by stacked autoencoder (SAE) during the pretraining process. Since the learning performance of normal SAE architecture seems not good enough for image patches, dropout is introduced into input layer to reduce model complexity. Experimental results show that input dropout improves 86 % training accuracy and reduces 16 % running time. Second, a normal neural network classifier is trained by using encoding weights as initial connect weights between input layer and hidden layer during the training process. Finally, the trained classifier can be applied to classify similar time-series data.

The remainder of this paper is organized as follows: Section. 10.2 discusses the architecture of SAE with an input dropout that is used in the rest of the paper. Section 10.3 describes the architecture of TSV. Then, the experimental results and comparisons are presented on classification for time-series data. Finally, Sect. 10.4 presents our conclusions.

10.2 The Architecture of TSV for Time-Series Classification

The architecture of TSV for time-series classification is given in Fig. 10.1 intuitively.

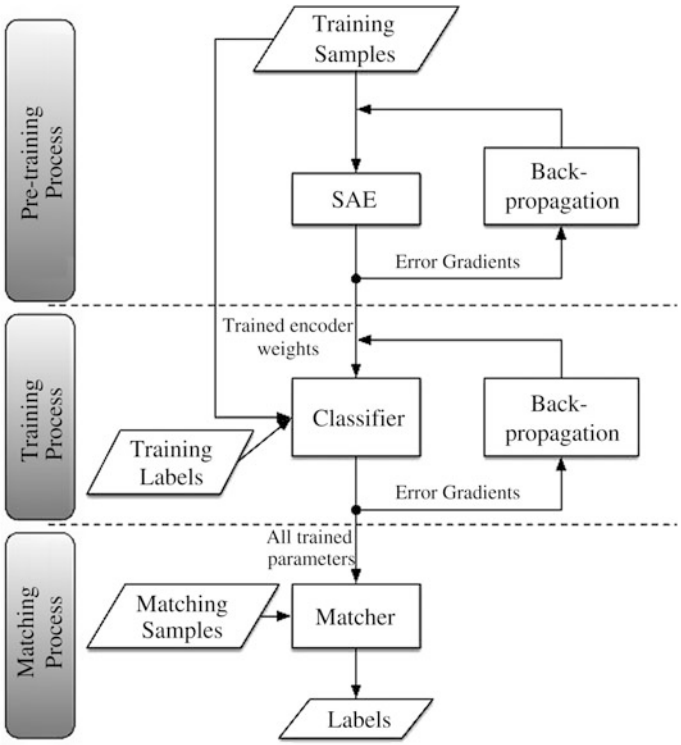


Fig. 10.1 The architecture of TSV for time-series classification

Three processes are executed successively in an overall similarity matching of TSV. It is worth noting that not all the parameters trained from the previous level will be transported to the next level as shown in Fig. 10.1. For example, only the trained weights between input layer and hidden layer of SAE will be used in the next training of the NN classifier.

Here we define an input $X = [x_1, x_2 \ldots x_n]^T$, and the autoencoder transforms the input X into an output $Y = [y_1, y_2 \ldots y_n]^T$ with learnt representation. In order to drive a general architecture of autoencoder network, a three-layer neural network architecture is applied. However, autoencoders are distinguished from more general neural networks by the fact that their outputs are desired to be the same to their inputs. The hidden layer detects features in input data. Then the corresponding decoder takes encodings, and attempts to reconstruct the original input.

Dropout was proposed by Hinton et al. [13] as an approach to improve the performance of fully connected neural network layers. When the dropout is applied in a fully connected layer, each element of the layer is kept with probability p, otherwise set to 0 with probability $(1 - p)$. Our dropout algorithm is modified by introducing dropout probability $(1 - p)$ into input layer. The architecture and introduction of SAE with input dropout is specified below.

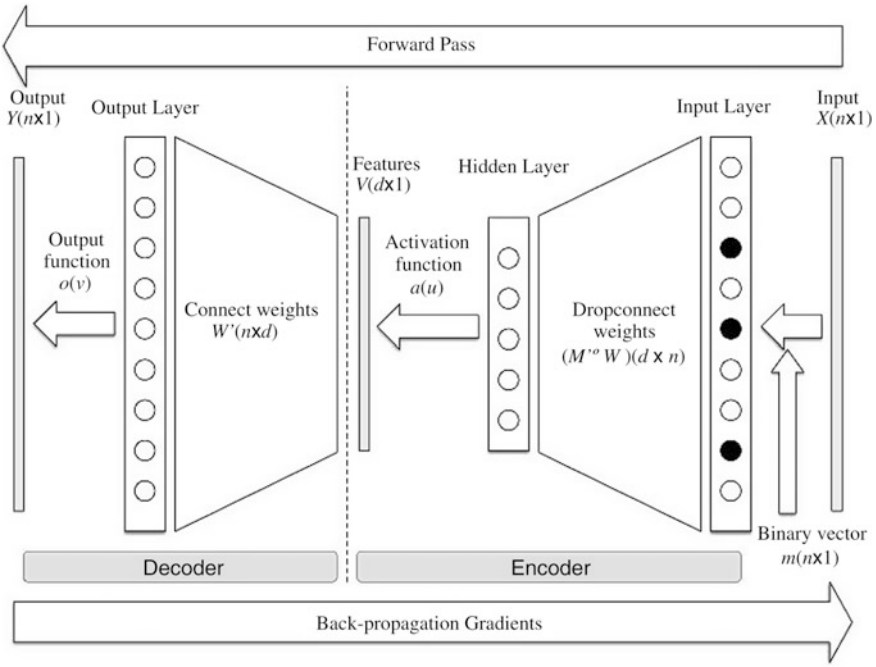


Fig. 10.2 The architecture of SAE with input dropouts

Encoder. The deterministic mapping f_θ that transforms an input into hidden features is called the encoder [14]. Each input vector X and the weight matrix W followed by a nonlinear activation function $a(u)$ such as tanh, sigmoid, or relu, which can be expressed as (Fig. 10.2):

$$V = f_\theta(X) = a(WX + b),$$

where V is a feature matrix extracted by the encoder. f_θ is an affine mapping and its parameter set is $\theta = \{W, b\}$, where W is a $d \times n$ weight matrix and b is an offset vector of hidden dimensionality d. Because we have introduced dropout probability $(1 - p)$ into input layer, f_θ can be rewritten as

$$V = a(W(mX) + b) = a((M' \circ W)X + b),$$

where m is a binary vector of size n with each element m_j drawn independently from Bernoulli(p), and $M^{'}$ is a $d \times n$ drop connect weight matrix with a same binary value in each row. Then we made an approximation

$$\sum_{M'} a((M' \circ W)X + b) \approx a(\sum_{M'} ((M' \circ W)X + b))$$

Table 10.1 Training SAE with input dropouts

Initialization: input $X \rightarrow$ image format $\rightarrow$ stacked vector X', initialize parameters $\{\theta_0, \theta'_0\}$, and learning rate η.
Input: Randomly selected input x' and parameters $\{\theta_{t-1}, \theta'_{t-1}\}$ from step $t-1$.
Forward Pass:
Select randomly drawn mask m: $m \sim \text{Bernoulli}(p)$
Compute hidden features: $v = a(w(x' \circ m) + b)$
Compute output: $y = o(w'v + b')$
Back-propagation Gradients:
Compute the loss function of decoding layer and encoding layer and, respectively, expressed as $L'_{W'}$ and L'_W.
Update weights of decoding layer: $W'_t = W'_{t-1} - \eta L'_{W'}$
Update weights of encoding layer: $W_t = W_{t-1} - \eta(M' \circ L'_W)$

Decoder. The decoder mapping $g_{\theta'}$ is used to reconstruct the output Y. It can be regarded as a reverse process of encoder. Thus, the decoder can be expressed as follows with its appropriately sized parameters $\theta' = \{W', b'\}$.

$$Y = g_{\theta'}(V) = o(W'V + b')$$

According to informax principle put forward by Linsker [15], a good representation is to retain a significant amount of information from the input, which means to learn parameters $\{\theta, \theta'\}$ that minimize the overall distortion function expressed as follows:

$$\min E(X, Y) = \min_{\theta, \theta', M} \sum_{i=1}^{n} \Delta(X, Y; \theta, \theta', M') = \min_{\theta, \theta'} \sum_{M} (p(M') \sum_{i=1}^{n} \Delta(X, Y; \theta, \theta'))$$

Once the randomly drawn mask m is chosen, it is applied to train the parameters $\{\theta, \theta'\}$ via stochastic gradient descent (SGD) by back-propagation gradients of the loss function. Specific calculation steps of SGD training with input dropout are provided in Table 10.1.

10.3 Experiments on TSV: Classification for Time-Series Data

In this section, we experimentally evaluate the performance of TSV on time-series classification. Four benchmarks are applied to perform comparisons for classification accuracy with some other measures such as ED, DTW, EDR, and LCSS, which are used in references [3, 16, 17]. These four datasets contain curve features,

including the popular CBF dataset, ECG200 dataset, synthetic control dataset, and trace dataset.

For a fair comparison, we keep all parameters of TSV invariable for four datasets, which is 900 input nodes with 70 % dropouts, 100 hidden nodes, learning rate $\eta = 1$, and we trained SAE 3000 epochs and classifier 1000 epochs, respectively. Note that, the performance of TSV may not be in the best situation for every datasets we tested; however, our focus is not the best performance of TSV on a specific dataset, but a robust performance on all datasets.

Since the number of input nodes cannot be changed, resizing each image patch into a certain size is necessary. Here we restrict the size of image patch to 30×30, which seems to be a draconian restriction to some datasets like trace dataset whose length is 275. However, we are surprised to see that results of trace are all correct. Because sometimes, the dimensionality of time series is very high and details is not the key for detecting the feature of time-series data, resizing provides a way to reduce the dimensionality of data but keep the most important visual characteristics. In this way, autoencoder can be trained with lower training error and faster running time. Furthermore, dimensionality of time series is reduced to 100 hidden layer output. It seems like that curves sharing the general characteristics of time-series data are separated by the autoencoder and each training sample can be expressed as a combination of these curves. Hence, if the number of hidden nodes is chosen properly, dimensionality of time-series data can be significantly reduced.

Training image pitches of time-series samples and visualization of weights learnt by SAE with input dropout is shown in Fig. 10.3. And we compared training error and running time under different dropout probabilities. Further, the error ratios of all methods based on four benchmarks are shown in Table 10.2.

In summary, SAE with 70 % input dropout has the lowest train error and it is one of the fastest algorithms according to the running time comparison as in Fig. 10.4. It improves 86 % training accuracy and reduces 16 % running time. And from

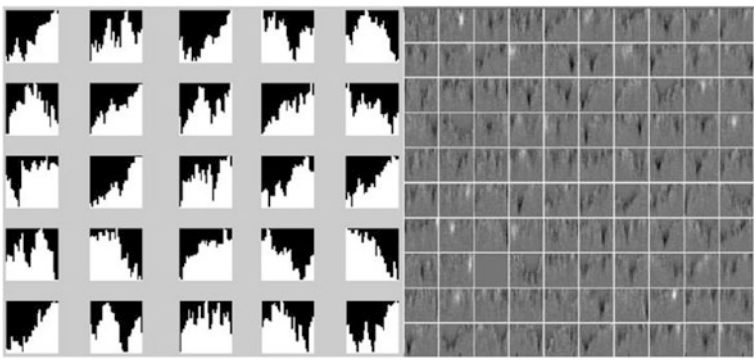


Fig. 10.3 Training image pitches of time-series samples and visualization of weights learnt by SAE with input dropout

Table 10.2 Error ratio of different methods

	CBF	Synthetic control	ECG200	Trace
1-NN Euclidean distance	0.087	0.143	0.16	0.36
1-NN DTW	0.003	**0.02**	0.23	0.02
1-NN EDR	0.013	0.117	0.21	0.15
1-NN ERP	**0**	0.037	0.21	0.08
1-NN LCSS	0.017	0.06	0.17	0.12
SVM Euclidean distance	0.123	0.0767	0.19	0.27
TSV	0.004	0.023	**0.15**	**0**

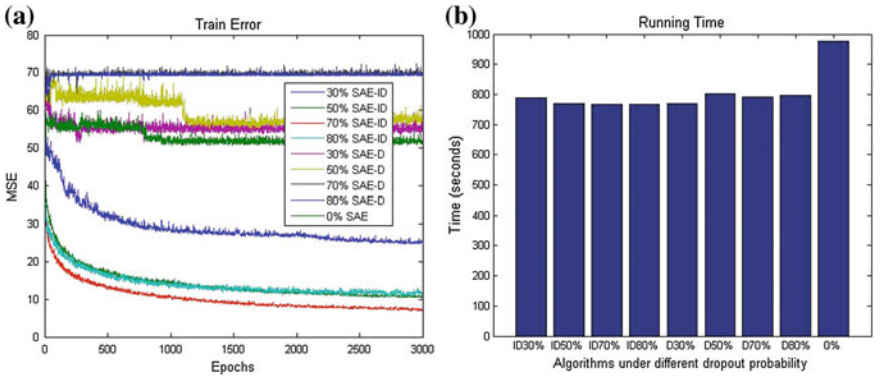


Fig. 10.4 Comparison of the training error and running time under different dropout probabilities

Table 10.2, we can see 1-NN ERP get the best effect on CBF data, but DTW and TSV have a very close performance. However, 1-NN DTW takes the first place on synthetic control dataset; in spite of this, these three methods are still quite close. TSV performs best on last two datasets, especially the trace dataset on which error ratio is zero, and 1-NN ERP and 1-NN DTW are inferior on the ECG200 dataset. Through experiments, we find that there is no clear evidence that one classification method tested is superior to others in all dataset tests in terms of accuracy. While TSV is a little bit more effective generally on the four datasets we used, some methods like 1-NN ERP and 1-NN DTW are superior on certain datasets but inferior on some other datasets. Hence, we believe that TSV is a more effective method compared to existing methods on the dataset with significant curve feature.

10.4 Conclusion

In this paper, we have presented a novel method called TSV to improve the performance of classification for time-series data on the datasets with a significant curve feature. We have clearly introduced the full architecture and algorithms of TSV, including SAE with input dropouts and a normal NN classifier. Then, we

evaluate the performance of TSV for time-series classification on four popular benchmarks. In this experiment, the results show that the performance of TSV is quite good in almost every datasets, which demonstrates that TSV is an effective method for time-series classification.

Acknowledgments This work was supported by Key Project of Chinese National Programs for Fundamental Research and Development (973 program) under Grant (2012CB720003), the National Natural Science Foundation of China (Nos. 91016004, 61004023, 61127007, 61174069) and the International S&T Cooperation Program of China (No. 2013DEE13040).

References

1. Faloutsos C, Ranganathan M, Manolopoulos Y (1994) Fast subsequence matching in time-series databases. ACM
2. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. KDD Workshop 10(16):359–370
3. Wang X, Mueen A, Ding H et al (2013) Experimental comparison of representation methods and distance measures for time series data. Data Min Knowl Disc 26(2):275–309
4. Yi BK, Jagadish HV, Faloutsos C (1998) Efficient retrieval of similar time sequences under time warping. In: IEEE 14th international conference on data engineering, pp 201–208
5. Kim SW, Park S, Chu WW (2001) An index-based approach for similarity search supporting time warping in large sequence databases. In: IEEE 17th international conference on data engineering, pp 607–614
6. Keogh E, Ratanamahatana CA (2005) Exact indexing of dynamic time warping. Knowl Inf Syst 7(3):358–386
7. André-Jönsson H, Badal DZ (1997) Using signature files for querying time-series data. Principles of data mining and knowledge discovery. Springer, Heidelberg, pp 211–220
8. Chen L, Özsu MT, Oria V (2005) Robust and fast similarity search for moving object trajectories. In: Proceedings of the 2005 ACM SIGMOD international conference on management of data, ACM, pp 491–502
9. Chen L, Ng R (2004) On the marriage of lp-norms and edit distance. In: Proceedings of the thirtieth international conference on very large data bases-vol 30, VLDB Endowment, pp 792–803
10. Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. ICML unsupervised and transfer learning, pp 37–50
11. Hinton G, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
12. Arel I, Rose DC, Karnowski TP (2010) Deep machine learning-a new frontier in artificial intelligence research. Comput Intell Mag IEEE 5(4):13–18
13. Hinton GE, Srivastava N, Krizhevsky A, et al (2012) Improving neural networks by preventing co-adaptation of feature detectors arXiv:1207.0580
14. Vincent P, Larochelle H, Lajoie I et al (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach Learn Res 11:3371–3408
15. Linsker R (1989) How to generate ordered maps by maximizing the mutual information between input and output signals. Neural Comput 1(3):402–411
16. Keogh E, Kasetty S (2003) On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Min Knowl Disc 7(4):349–371
17. Ding H, Trajcevski G, Scheuermann P et al (2008) Querying and mining of time series data: experimental comparison of representations and distance measures. Proc VLDB Endowment 1 (2):1542–1552

Chapter 11
Visualisation of the Early-Stage Vibration of the Automatic-Gauge-Control Hydraulic Cylinder Based on the Acoustic Emission Hits-Density Imaging

Hongzhi Chen, Yongli Zhao, Jie Huang and Chang Cheng

Abstract By taking into account the hits-density imaging technique that has been introduced by authors' previous publications, the acoustic emission bursts emitted from the automatic-gauge-control hydraulic cylinder under normal loading and early-stage vibration were investigated. First, six acoustic emission descriptors that showed significance regarding the cumulative probability distribution were visualised, followed by the reconstruction of the hits-density images using these features; Through the visual examination of the hits-density images regarding the cylinders' normal loading and early-stage vibration, apparent visual differences between normal and early-stage vibration under the same loading can be observed; By projecting the hits-density images onto the principal component space, a trajectory which represents the condition developments was observed, which shows a visual-based resolution for dealing with vast transient bursts data reduction, reconstruction and visualisation.

Keywords Visualisation · Hits-density imaging · Early-stage vibration · Acoustic emission

H. Chen (✉) · Y. Zhao · C. Cheng
State Key Laboratory of Hybrid Process Industry Automation System and Equipment Technology, Automation Research and Design Institute of Metallurgical Industry, Beijing 100081, China
e-mail: shadow_c3186@msn.com

J. Huang
Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing 100081, China

J. Huang
Fujian Institute of Education, Fuzhou 300025, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_11

11.1 Introduction

As the final stage to maintain the manufacturing tolerance under the acceptable level, even tricky requirements have been applied to the rolling processes according to the increasing quality requirements [1]. The automatic-gauge-control (AGC) hydraulic cylinder has been widely applied in the steel process according to a couple of advantages including fast response, high level of control accuracy, as well as the reliable pressure overload protection, etc. [2]. However, with the dysfunction for the AGC systems are unavoidable under the non-stop rolling process, previous studies showed a set of researches and contributions regarding the AGC system fault diagnostics, the works include obtaining fault knowledge [3]; pressure change forecasting based on enlarged neural networks [4]; the design of a decoupling subsystem based on differential geometric approach for robust diagnostic of load uncertainties [5]; as well as the innovative design of optimised BP neural works for dealing with one of major AGC system failure, namely multi-sensor faults [6, 7].

As the executing component that makes direct contact with the working rolls to justify the rolling space, the abnormal of the cylinder could directly influence the quality of products and brings underlying safety issues. Authors' previous publications investigated the condition changes of the hydraulic cylinder based on the acoustic emission (AE) and the hit-density imaging, with a development trajectory related to the normal and overloading conditions observed on the principal component space [8, 9]. Vibration is considered as one of major fault that could be resulted to the catastrophic consequences. However, the current fault detection process lacks diagnostic quality and time efficiency [10]. The manuscript presents the study of combination of hits-density imaging and AE for condition monitoring of the AGC cylinder early-stage vibration, thereby demonstrating the potential of using this method for visual-based health monitoring of the cylinder as well as providing the cross-validating option to detection cylinder vibration until failure (Fig. 11.1).

11.2 Statistic-Based Visualisation of Data Distributions

The cylinder under examination and the testing bed are the same as [8]. The normal loading is simulated by loading the cylinder to 5 (N_1), 10 (N_2), and 15 Mpa (N_3) and hold the pressure for 100 s, respectively. The cylinder vibration is simulated by adding a 20 Hz sinusoid signal to the piston while the cylinder loads were reached to the above-mentioned three conditions (denotes V_1, V_2, V_3 for 5, 10, and 15 Mpa vibrate, respectively), and also held for 100 s. In this case study, six features were found to show statistical significance among normal loading and early-stage vibration, namely, amplitude, counts, rise-time (RT), duration, average signal level over duration (ASLOD), and counts to peak (CTP), respectively. For detailed information of features, please refer to [11]. The detailed distributions of the

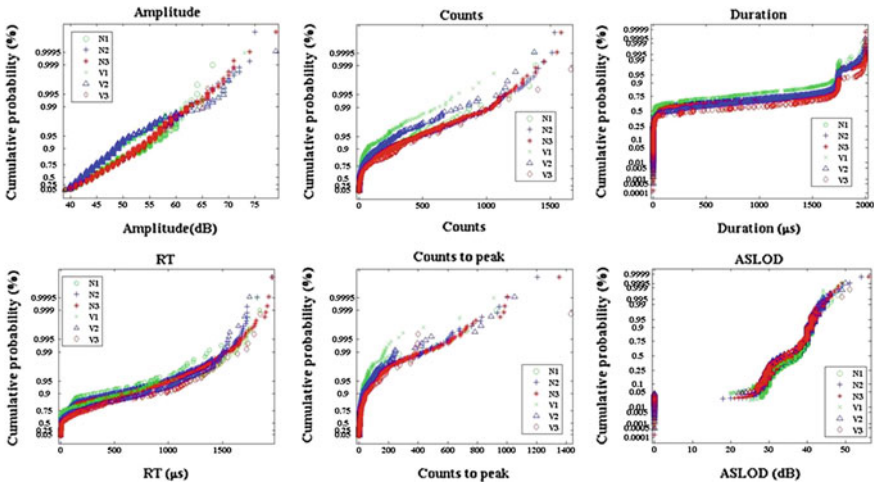


Fig. 11.1 Cumulative distributions of AE features

selected features can be visualised in Fig. 11.2, represents by using the 2D cumulative probability plot [12], with the horizontal and vertical axis denote the values and the occurrence probability of the selected features, respectively.

In terms of the AE amplitude distribution, compare with the normal and vibration of cylinder under the same load, over 95 % of amplitude values are overlapped; however, the statistical significance can be observed at the remaining 5 % of the curve, the maximum AE amplitude values are seen to be increased when the cylinder was developed from N_1 to N_3, with the maximum amplitude value over

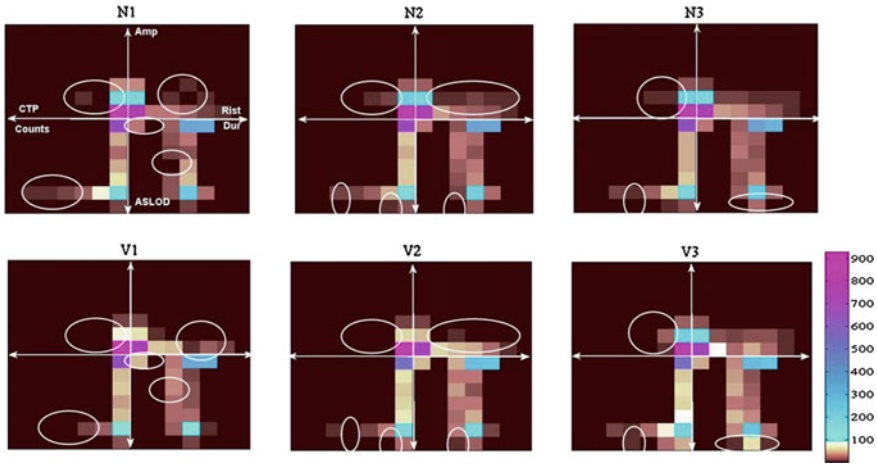


Fig. 11.2 Visualisation of cylinder conditions using hits density images

75 dB generated by N_3. V_1 and V_2 are seen to follow the similar rule as the normal condition; however, the maximum amplitude value is seen to decrease while the condition is developing to V_3, with the maximum value among vibrate groups >75 dB generated at V_2, this could be related to the underlying mechanism of the cylinder, as some relevant abnormal situations such as insufficient loading or leakage may happen while the cylinder is vibrating under high pressure (Fig. 11.1).

In terms of the cumulative probability distribution of the AE duration, it can be seen that the upper and lower band for the six conditions under examination are similar, which varies between 0 and 2000 μs; however, the distribution curves show separation at the band between 0 and 1500 μs. Similar scenario can be found in the distribution of signal ASLOD, with around 75 % of ASLOD between 20 and 40 dB are nonoverlapped.

Another three significant features are counts, CTP and RT. The counts is seen to be the most significant feature to recognise the cylinder vibrate, with around 10 % of curves are seen to be non-overlapped, and with vibration generated slightly higher value than the same pressure of normal loading. The maximum counts of AE waveform are observed in V_3, with some hits having over 1500 pulses exceeding the detection threshold recorded. By observing the distribution of CTP, although the maximum value does not show significant difference, at least 5 % of AE signals' CTP are nonoverlapped. A similar scenario can be observed in the cumulative distribution of RT, in terms of the curves representing the cylinder under the same loadings for both normal and vibration, with the maximum value of RT are increased corresponding to the increment of loading, as well as around 5 % of curves among six conditions shown non-overlapping.

11.3 Visualisation of Early-Stage Vibration of the Cylinder Using Acoustic Emission-Based Hits-Density Imaging

11.3.1 Construction of Hits-Density Image

Similar scenario as the previous literatures been published [8, 13, 14], the hits-density imaging is taken into account in order to reduce the amount of data to be processed, and to convert the non-continuous recorded signals into a uniform format [15]. The basic representation of the approach is determined by Eq. 11.1:

$$\mathbf{N}(F, \psi) = \boldsymbol{F}_{\psi_j}(\boldsymbol{V_1}, \boldsymbol{V_2}, \ldots \boldsymbol{V_i}) \tag{11.1}$$

where, **N** denotes the number of AE events constrained by the specific AE feature combinations, V_1, V_2,…, V_i are referred to as the AE descriptors that showing significant; F_{ψ_j} denotes the number of AE events satisfies the conditions determined by V.

The 2D image-based profile can be created based on the six selected features, with each individual image represents a certain cylinder condition, the image can be segmented into four quadrants by using two axis radiated via the horizontal and vertical directions from the centre of image, with each quadrant representing one of particular AE feature combination [8]. Pixel values inside each quadrant of image can be represented by using (11.2), which could be considered as the logic AND operation among two specific value ranges of the selected features:

$$F_{\psi_j}(V_1, V_2) = F_{\psi_j}(V_1) \cap F_{\psi_j}(V_2) \tag{11.2}$$

where, V_1, V_2 are denoted the specific value ranges of two selected features in each quadrant, respectively.

11.3.2 Visualisation Results of the Cylinder Early-Stage Vibration

By using the construction criteria of the hits-density imaging described in preceding sub-section, the hits-density images of the conditions being examined are illustrated in Fig. 11.2. In order to cross-visualise the relationships of AE features regarding the peaks and the whole shape of AE waveform above the detection threshold, the feature combinations are organised as the text labelled on the left top of the Fig. 11.2, such representation has been proved as useful for visualisation regarding to authors' previous publications [8, 13, 14, 16]. In Fig. 11.2, the vectors represent the value intervals of the selected features are as follows: $V_{\text{Amplitude}}$ = [40 50 60 70 75 80 85], V_{Counts} = [1 250 500 750 1000 1250 1500], V_{RT} = [0 300 600 900 1200 1500 1800], V_{ASLOD} = [0 10 20 30 40 50 60], V_{CTP} = [1 250 500 750 1000 1250 1500], and V_{Dur} = [0 300 600 900 1200 1500 1800]. If denoting the number of AE events satisfies a value range determined by one of individual AE feature as N, the number of AE signals that fall into the first six value intervals can be determined by using (11.3)

$$F_k \leq N_k < F_{k+1} \quad \text{for} \quad k = 1, 2, \ldots 6 \tag{11.3}$$

and the last value interval can be determined as $N_7 > F_7$.

Through the visual examination of the hits-density images, in the general view, the hits-density images that show the conditions of the cylinder are seen to have very high similarities, this is corresponding to the actual condition of the cylinder (i.e., the cylinder is still working under health condition even though it is vibrating), however, slight differences can still be observed for various conditions according to the intensity of each images.

As the differences circled on Fig. 11.2, N_1 generates the AE signals with {Amplitude ∩ CTP|(50–60) ∩ (500–750)}, as well as {Amplitude ∩ RIST|(70–80) ∩ (1200–1500)}, which is shown absence in V_1. To compare the differences

between N_2 with V_2, it is seen that in general the number of AE events emitted by N_2 are higher than V_2, and only N_2 generated the AE signals with {Amplitude ∩ CTP|(50–60) ∩ (250–750)}; The more significant differences could be observed on the sub-field constructed by using the Amplitude and RIST, with the AE signals having amplitude values between 50 and 60 dB are seen to increase through the direction of value of RIST increasing; Furthermore, a portion of subtle differences could be observed, with the AE events determined by {Counts ∩ ASLOD|(750–1000) ∩ (50–60), (250–500) ∩ (>60)}are emitted when the cylinder was at V_2.

A similar scenario could be observed while comparing N_3 with V_3, as shown in the right extreme column of Fig. 11.2, with slightly differences observed on the first, third, and fourth quadrants, as N_3 generated AE events with {Amplitude ∩ CTP|(50–60) ∩ (750–1000)} whereas there is an absence in the same zone of image created by V_3; Furthermore, V_3 generates the AE events with {ASLOD ∩ Counts|((>60) ∩ (600–900)),((>60) ∩ (1200–1500))}, however, it is unable to see number of AE events returned regarding to the same zone in N_3.

11.3.3 Trajectory of Development of Cylinder Statuses from Normal to Vibration

For further verification of the hits-density images as a potential tool for visual examination and monitoring of the cylinder early-stage vibration, the hits-density images were projected onto the PC space for further analysis using PCA [17], thereby verifying the ability of hits-density imaging as a tool for visualising the development trend among normal and early-stage vibration.

The projections of hits-density images onto the PC space using PCA are shown in Fig. 11.3, with the connection of each status demonstrating the development trajectory of hits-density images from N_1 to V_3. A clear trend from N_1 to V_3 can be observed on the axis system formed by the first three PC with more than 99 % of variances retained. The development trajectory is seen to decrease from N_1 to N_2 through the space formed by the first and third PC, and decrease through the space forming by the second PC when the loading were increased from N_2 to N_3. The trajectory is shown to increase mainly via the third PC and cross over the origin point formed by the first and second PC when the cylinder condition is developing from N_3 to V_1. The trend is again shown to decrease through the space formed by the second and third PC while the condition is changing from V_1 to V_2, and increase through the space formed by the first and third PC.

For the projection of hits-density images of normal and early-stage vibration onto the PCA space, two interesting scenarios are worth to mention. Three normal conditions are projected on the positive domain whereas three early-stage vibration conditions are projected on the negative domain while observing the space formed by the first and second PC. The development trajectory regrading to

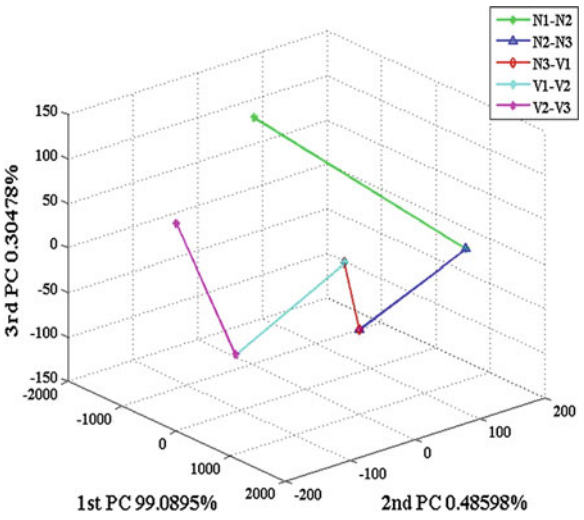


Fig. 11.3 PCA projection of hits density images

conditions N_2–N_3 and V_1–V_2, as well as the condition N_3–V_1 and V_2–V_3 are nearly parallel and symmetry to the origin. By comparing with other testing methodologies, as well as the simulation of other conditions, these two scenarios could trigger the further studies of the relationships of condition changes and the underlying mechanisms among these four cylinder conditions as the future works.

11.4 Conclusion

The paper presented an exploratory study in terms of visualisation of the early-stage AGC hydraulic cylinder vibration and normal loading by using the hits-density imaging. By visually examining the feature distributions of the cylinder via the cumulative probability plot, six AE descriptors appeared to have statistical significancy. By taking into account the hits-density imaging-based technique to construct the AE events emitted from the cylinder during normal loading and early-stage vibration conditions, significant differences between normal and early-stage vibration among the same loading could be observed via the image intensities. By projecting the hits-density images of the above-mentioned six conditions onto the PC space, a development trajectory for the cylinder condition from 5 Mpa normal loading to 15 Mpa early-stage vibration can be observed, with three normal loading conditions allocated on the positive side of the space and three early-stage vibration conditions allocated on the negative side of the space forming by the first and second PC. Also, another interesting scenario was observed on the PC space, as the trajectory that represents the condition development from N_2–N_3 to V_1–V_2 as well as N_3–V_1 & V_2–V_3 are parallel, which may lead to the further assessment of the deep links inside.

The result proposed a visual-based solution for monitoring the AGC cylinder early-stage vibration until failure, which is digging deeper into the subtle difference level, in order to deal with the classification and visualisation of early-stage vibration using the data-driven-based approach. With the condition-based maintenance of the metallurgical equipments currently rising up to the functionality-accuracy level, the hits-density imaging-based methodology has been proved as the potential tool for quick visual examination of cylinders' subtle condition changes. Further research activities will be concentrated on the optimisation, as well as on the remaining feature selection problem, thereby optimising the granularities as well as the feature combination that constructs the hits-density image based on the even detailed working conditions.

Acknowledgments The work is supported by the National Natural Science Foundation of China (Grant ref: 61401104).

Conflict of interest statement

None

References

1. Gajdzik B, Sitko J (2013) An analysis of the causes of complaints about steel sheets in metallurgical product quality management systems. Metalurgija 53:135–138
2. Liu H (2012) Application of predicted extrapolation control strategy in hot strip rolling mills gauge system. In: Proceedings world automat Cong (WAC), pp 1–4
3. Wang H, Rong Y, Cui J et al (2010) Study on knowledge processing of fault diagnosis for hydraulic AGC system. In: 2nd IEEE international conference Information management engineering (ICIME), pp 1–3
4. Wang H, Rong Y, Liu S et al (2010) Identification for hydraulic AGC system of strip mill based on neural networks. Intl Conf Comput Des Appl (ICCDA) 22:377–380
5. Dong M, Liu C, Li G (2010) Robust fault diagnosis based on nonlinear model of hydraulic gauge control system on rolling mill. IEEE Trans Control Syst Tech 18:510–515
6. Wang X, Zhang K (2012) Sensors fault diagnosis of hydraulic automatic gauge control system based on neural network optimized by genetic algorithm. In: IEEE international conference oxide materials electronic engineering (OMEE), pp 114–117
7. Wang X, Liu C, Li M (2010) Sensors fault diagnosis of hydraulic automatic gauge control system based on wavelet neural network. IEEE *International Conference Electronic Control Engineering* (ICECE), pp 3009–3012
8. Chen H, Wu C, Sun Y, Zhao H (2013) State identification of automatic gauge control hydraulic cylinder using acoustic emission. In: Proceedings Chinese *International Conference* Springer, Berlin, Heidelberg, vol 255, pp 585–594
9. Chen H, Lu Y, Zhao H et al (2013) Automatic gauge control hydraulic cylinder state identification using modified image based acoustic emission profile. In: Proceedings 2nd *International Conference Computer Science Electronic Engineering*, pp 2153–2159
10. Fan L, Li M, Shi L et al (2014) UIO-based fault diagnosis for hydraulic automatic gauge control system of magnesium sheet mill. Sens Transdec 65:22–28
11. Physical Acoustic Corporation (2003) PCI-2 based AE system user's manual. Phys Acoust Corp
12. Utts J (2014) Seeing through statistics. Thomson Brooks/Cole United States, Belmont

13. Shark LK, Chen H, Goodacre J (2011) Knee acoustic emission: a potential biomarker for quantitative assessment of joint ageing and degeneration. Med Eng Phy 33:534–545
14. Shark LK, Chen H, Goodacre J (2010) Discovering differences in acoustic emission between healthy and osteoarthritic knees using a four-phase model of sit-stand-sit movements. J Open Med Inf 4:116–125
15. Mascaro B, Prior J, Shark LK et al (2009) Exploratory study of a non-invasive method based on acoustic emission for assessing the dynamic integrity of knee joints. Med Eng Phy 31:1013–1122
16. Chen H (2011) Discovery of acoustic emission based biomarker for quantitative assessment of knee joint ageing and degeneration. Ph.D thesis, University of Central Lancashire, Preston
17. Jolliffe IT (1986) Principal component analysis. Springer-Verlag, New York

Chapter 12
Speaker Recognition Based on i-Vector and Improved Local Preserving Projection

Di Wu

Abstract In order to enhance the recognition performance of the i-vector speaker recognition system under unpredicted noise environment, an improved local preserve projection which was used for reduce dimension to i-vector is proposed on this paper. First, the nonzero eigenvalue is rejected when we solve the optimized objective function, only using the eigenvalue the value of which is greater than zero. A mapping matrix is obtained by solving a generalized eigenvalue problem which can settle the singular value problem which occurred in the traditional local preserve projection algorithm. The experiment result shows that the recognition performance of the method proposed in this paper is improved under several kinds of noise environments.

Keywords Computer application · i-vector · Local preserving projection · Manifold learning · Speaker recognition

12.1 Introduction

With the past decades, speech recognition has become a very popular area of research in pattern recognition, computer vision, and machine learning [1]. Mismatches between training and testing conditions are caused by some inevitable reasons such as channel distortion, different microphones, transmitting channels, or encoder. One of the main causes of the performance degradation is the additive noise that may appear in many practical applications. There are a large number of different solutions to alleviate this problem. We can identify three main classes of techniques for noise-robust ASR, namely feature enhancement method [2], model adaptation method [3], and score normalization method [4]. The feature enhancement method attempts to

D. Wu (✉)
College of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan 411004, China
e-mail: wudi6152007@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_12

normalize the distorted feature, or estimate undistorted feature form the distorted speech, and does not require any explicit knowledge about the noise. Some examples are the cepstral mean normalization (CMN), cepstral mean and variance normalization (CMVN), relative spectra (RASTA), and feature mapping. In contrast, the model adaptation methods work in the backend to compensate by modifying the acoustic models and carried out by using some type of knowledge about the noise. Some typical examples are maximum likelihood linear regression (MLLR), maximum a posterior (MAP), factor analyze (FA), and vector Taylor series (VTS) etc. The score normalization method tries to normalize the output score using various normalization methods, such as HNorm, TNorm, and ZNorm etc.

In the last years, the Gaussian Mixture Models based on Universal Background Model (GMM-UBM) [5] has become the most popular modeling approach in speech recognition; some generative models such as Eigenvoices, Eigenchannels, and the most powerful one, the Joint Factor Analysis, are built on the success of the GMM-UBM approach. Recently, a new method which is inspired from the joint factor analysis and consists in finding a low dimensional subspace of the GMM supervector space, named the total variability space that represents both speaker and channel variability, the vectors in the low dimensional space are called i-vectors [6]. The i-vector method has become the main stream in the speaker recognition system at home and aboard for the reason of its leading role in the NIST test.

Locality Preserving Projections (LPP) [7] is a manifold learning method widely used in pattern recognition and computer vision. LPP is also well known as a linear graph embedding method. But the traditional LPP method was unsupervised and was proposed for only vector samples, not being able to be directly applied to image samples; so there are been several types of improvements to conventional LPP [8]. The first type of the improvement is supervised LPP, which tries to exploit the class label information of samples in the training phase. The second type changes LPP into a nonlinear transform method by using the kernel trick. The third type of the improvement to LPP mainly focuses on directly implementing LPP for two dimensional rather than one-dimensional vectors and it has higher computational efficiency. And the last improvement seeks to obtain LPP solutions with different solution properties, such as orthogonal locality preserving method and uncorrelated LPP feature extraction method.

12.2 The Improved LPP Method

12.2.1 Description of LPP

LPP was proposed as a way to transform samples into a new space and to ensure that samples that were in close proximity in the original space remain so in the new space. Consider there have l training samples $X = \{x_i\}_{i=1}^{l}$, the goal of LPP is to minimize the following function [9, 10]:

$$\min\left(\sum_{i,j}\left(W^T x_i - W^T x_i\right)^2 S_{ij}\right) \tag{12.1}$$

The S_{ij} is a symmetric matrix and the element of the S_{ij} is defined as follows:

$$S_{ij} = \begin{cases} \frac{\exp\left\|x_i - x_j\right\|^2}{t} & \text{if } x_j \text{ is one of } K \text{ neighbors of } x_i \\ 0 & \text{else} \end{cases} \tag{12.2}$$

From the optimized function Eq. (12.1), we can see that the local structure of the feature space can preserved like in the original high dimension space after dimension reduction, which means close samples in the original space will still close in the new space, so the projection matrix W can be written as:

$$\begin{aligned} W &= \arg\min W^T \mathrm{XLX}^T W \\ &= \arg\min W^T X(D - S)X^T W \end{aligned} \tag{12.3}$$

In Eq. (12.3), D is diagonal matrix, $D_{ii} = \sum_j S_{ij}$, $L = D - S$, the solution of Eq. (12.3) can be obtained by finding the generalized eigenvalue of the following function:

$$\mathrm{XLX}^T W = \lambda \mathrm{XDX}^T W \tag{12.4}$$

12.2.2 New LPP Solution Scheme

In this section, we describe our new improvement to the conventional LPP solution scheme. First, we demonstrate the effective solution of the conventional LPP solution should be from a subspace XDX^T, for simplicity, we define matrix D_1, L_1 and S_1:

$$\begin{cases} D_1 = \mathrm{XDX}^T \\ L_1 = \mathrm{XLX}^T \\ S_1 = \mathrm{XSX}^T \end{cases} \tag{12.5}$$

Suppose that $\vec{\partial}_1, \vec{\partial}_2, \ldots, \vec{\partial}_n$ are the eigenvectors corresponding to the positive eigenvalues of D_1 while $\vec{\partial}_{n+1}, \vec{\partial}_{n+2}, \ldots, \vec{\partial}_N$ are the eigenvectors corresponding to the zero eigenvalues; in this paper, we regard eigenvalues that are less than 0.2×10^{-10} are zero eigenvalues. According to the nature of LPP, the ability of the preserving the neighbor relationship can be measured by $W^T L_1 W / W^T D_1 W$, that means the smaller $W^T L_1 W / W^T D_1 W$ value is, the better the local structure of samples is preserved, so the Eq. (12.4) can be rewrote as:

$$L_1 W = \lambda D_1 W \tag{12.6}$$

Then we design a matrix, $R = [\vec{\partial}_1, \vec{\partial}_2, \ldots, \vec{\partial}_n]$, using R, we respectively transform D_1, L_1, S_1 into the following matrices:

$$\begin{cases} \overline{D} = R^T D_1 R \\ \overline{L} = R^T L_1 R \\ \overline{S} = R^T S_1 R \end{cases} \tag{12.7}$$

Then we can directly solve the Eq. (12.7) since $\overline{D}$ is of full rank. Let $\vec{\beta}_1, \vec{\beta}_2, \ldots, \vec{\beta}_n$ denote the eigenvectors corresponding to eigenvalues $\vec{\lambda}_1, \vec{\lambda}_2, \ldots, \vec{\lambda}_n$ in the increasing order of Eq. (12.7). Using the matrix R, we produce $W = X^T R$, then we transform W into Y bu carrying out $Y = \mathrm{WG}$, where $G = [\vec{\beta}_1, \vec{\beta}_2, \ldots, \vec{\beta}_n]$, that is:

$$Y = \mathrm{WG} = (X^T R)G \tag{12.8}$$

12.3 The Improved i-Vector System

The main idea in traditional JFA is to find two subspace which represent the speaker and channel variabilities, respectively. The experiment shows that JFA is only partially successful in separating speaker and channel variabilities. While in the i-vector method proposed a single space that models the two variabilities and named it the total variability space [11, 12]:

$$M = m + T\omega \tag{12.9}$$

where M is the mean supervector which contain speaker and channel information, m is UBM supervector, T is a low rank matrix named total variability matrix, which represents a basis of the reduced total variability space and ω is a standard normal distributed vector, the components of ω are the factors and they represent the coordinates of the speaker in the reduced total variability space; these feature vectors are referred to as identity vectors or named i-vector for short.

The crucial step to the i-vector method is to compute total variability matrix T. At first, we train UBM using EM algorithm, and extract the Baum–Welch variables according to the trained UBM:

$$N_m = \sum_t \gamma_{m,t} \tag{12.10}$$

$$F_m = \sum_t \gamma_{m,t}(\xi_t - \mu_m) \tag{12.11}$$

The N_m and F_m represent zero-order and first-order statistic variable, respectively, t is the frame numbers, m represent the m-th hybrid vectors of UBM, $\gamma_{m,t}$ is the Gaussian sharing rate, that:

$$\gamma_{m,t} = \frac{N(\xi_{t;}\, \mu_m, \sum m)}{\sum_{i=1}^{M} N(\xi_{t;}\, \mu_i, \sum i)} \tag{12.12}$$

$N(\xi_{t;}\, \mu_m, \sum m)$ is the Gaussian component which the mean is μ_m and variance is $\sum m$, ξ_t is the random vector of the t frame, M is the mixed number of UBM. After calculate Baum–Welch variables, we can training matrix T using EM method as follows:

$$L = I + T^T \sum{}^{-1} NT \tag{12.13}$$

$$E(x) = L^{-1} T^T \sum{}^{-1} F \tag{12.14}$$

F is the vector arrangement of F_m, N, $\sum$ is the diagonal matrix of N_m, $\sum m$ respectively.

12.4 Experiment

The simulation experiments in this paper consist of two parts:

(1) In the clean background, we compare the performance of the conventional LPP method and the improved LPP method proposed in this paper utilizing in the i-vector system and GMM method. The results are shown in the Table 12.1.
(2) Under different noise environments, we explore the robustness of the new LPP method utilized in the i-vector system. The results are shown in the Table 12.2.

From the results shown in the Table 12.1, it is clear that the recognition performance of the i-vector system is better than the initial GMM recognition system whether under EER criterion or MinDCF criterion. We can see that the EER is

Table 12.1 Experiment result compared between initial LPP algorithm and improved LPP algorithm which is used for i-vector speaker recognition system

Method	EER(%)	MinDCF
LPP (i-vector)	4.72	0.19
Improved LPP (i-vector)	4.45	0.17
Conventional GMM	7.32	0.53

Table 12.2 Experiment result based on the improved LPP algorithm under different noise environments which is used for i-vector speaker recognition system

Voice environment	SNR (db)	EER (%)	MinDCF
Clean background	>40	4.45	0.17
White noise environment	0	7.04	0.335
	5	6.72	0.295
	10	5.91	0.276
	15	5.36	0.242
	20	4.93	0.204
Babble noise environment	0	6.89	0.314
	5	6.49	0.282
	10	5.71	0.255
	15	5.02	0.228
	20	4.76	0.189

reduced by 3 % and MinDCF is reduced by 0.35 % compared to the initial GMM system, so the experiment result confirms the superiority of the i-vector system powerfully. While further to see the result shown in the Table 12.1, the performance given by the improved LPP algorithm is better than the performance given by initial LPP algorithm, the EER is reduced 0.27 % and MinDCF is reduced 0.02 %. This improved method can enhance the recognition performance of the i-vector system for the reason that it can further discriminate the in-class samples and the near distance extra-class samples.

Form the experiment results shown in the Table 12.2, the performance given by the i-vector system based on the improved LPP scheme is better than the initial GMM method. The EER is 4.45 % and the MinDCF is 0.17 under clean background and it decreases 2.87 % and 0.36, respectively, compared to the initial GMM method.

The performance of the method proposed in this paper can reduce the EER and MinDCF certain degree under different signal noise rate (SNR) under white noise and babble noise environment. While the SNR is 20, the EER is 4.93 % and 0.17 under white noise environment and babble environment, and its decreases 2.29 % and 2.56 %, respectively compared to the initial GMM method.

12.5 Conclusion

In this paper, a new method of enhancing the speech recognition performance under i-vector system which its the most cutting edge recognition system in our knowledge is proposed, the new method is based on conventional LPP method and the motivation was that the conventional LPP method is always suffer from the SSS problem, and in this new scheme, We only using the eigenvectors corresponding positive eigenvalue when solving the optimized objective function and removing the zero eigenvalue.

Further work will concentrate on following two areas:

(1) Solving the small sample size (SSS) problem of the LPP method utilizing other mathematical method forever.
(2) The computational requirements for training the i-vector systems and estimating the i-vectors, however, are too high for certain types of applications. A simply method to the original i-vector extraction and training which would dramatically decrease their complexity while retaining the recognition performance is insistent demand.

Acknowledgments This work was supported in part by the National Science-technology Support Plan Project of China under contract 1214ZGA008, the Nature Science Foundation of China under contract 61263031, and the Science Foundation of Gansu Province of China under contract 1010RJZA046.

References

1. Kinnunen T, Li HZ (2010) An overview of text-independent speaker recognition: from features to supervectors. Speech Commun 52:12–40
2. Tohidypour HR, Seyyedsalehi SA, Behbood H, Roshandel H (2012) A new representation for speech frame recognition based on redundant wavelet filter banks. Speech Commun 54: 256–271
3. Perrachione TK, Del Tufo SN, Gabrieli JDE (2011) Human voice recognition depends on language ability. Science 333:595
4. Eskikanda PZ, Seyyedsalehia SA (2012) Robust speech recognition by extracting invariant features. Procedia Soc Behav Sci 32(3):230–237
5. Yang S, Zhaozhuang J, Deliang W (2009) An auditory based feature for robust speech recognition. ICASSP, Taibei, pp 4625–4628
6. Wu D, Cao J, Wang J, Li W (2012) Multi-feature fusion face recognition based on Kernel discriminate local preserve projection algorithm under smart environment. J Comput 7(10): 2479–2487
7. Du J, Huo Q (2011) A feature compensation approach using high-order vector taylor series approximation of an explicit distortion model for noisy speech recognition. IEEE Trans Audio Speech Lang Process 19(8):2285–2293
8. Jeong Y (2010) Speaker adaptation based on the multilinear decomposition of training speaker models. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, Dallas, USA pp 4870–4873
9. He Y, Han J (2011) Gaussian specific compensation for channel distortion in speech recognition. IEEE Signal Process Lett 18(10):599–602
10. Dehzangi O, Mab B, Chng ES, Li H (2012) Discriminative feature extraction for speech recognition using continuous output codes. Pattern Recogn Lett 33:1703–1709
11. Gu XH, Gong WG, Yang LP (2011) Supervised graph-optimized locality preserving projections. Opt Precis Eng 19(3):672–680
12. Dehak N, Kenny P, Dehak R, Dumouchel P, Ouel-let P (2010) Front-end factor analysis for speaker verification. IEEE Trans Audio Speech Lang Process 19(99)

Chapter 13
Giant Benthic HD Image Feature Extraction and Size Estimation Based on Canny Algorithm

Zhongjun Ding, Changcheng Wang and Pan Wang

Abstract Deep-sea benthic image features are difficult to extract because of its large amounts of information, auxiliary light imaging, and complex environmental background. To solve these problems, current study presents an approach to get texture information of sponge image captured by Jiao Long DSV. First, linear grayscale transformation is used to remove the seamount background and enhance contrast based on the image histogram analysis. The noise introduced by the suspended particles impurities is suppressed by median filter. Subsequently, compared with Prewitt and LoG algorithm, Canny operator is sure to get better edge extraction. Sponge texture information is most complete and noise is further reduced. Finally, mathematical morphology processing is carried out to perfect the texture by connecting intermittent textures, and the size estimation of the sponge based on hypothetical laser ruler is reliable and applicable.

Keywords Deep-sea image · Texture extraction · Size estimation · Jiao long DSV

13.1 Introduction

The development of human society is inseparable from the development and utilization of resources. In the situation of today's land resources drying up, people gradually set their sights on the deep ocean. As the submarine is rich in strategic metals, energy and biological resources, deep-sea research and utilization of resources have become a major national need for sustainable development. In recent

Z. Ding (✉)
Department of Technology, National Deep Sea Center, SOA,
No. 6 Xian Xia Ling Road, Qingdao, China
e-mail: greatwallwangcc@163.com

C. Wang · P. Wang
Department of Automation and Electrical Engineering, Qingdao University of Science & Technology, No. 53 Zhengzhou Road, Qingdao, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_13

years, with Jiao Long DSV put into use, China has the ability to be more systematic and comprehensive in deep-sea exploration and research. But in deep-sea environment, there is no natural light and all information is got through high-definition cameras with the auxiliary light source, resulting in an image uneven illumination and shadow interference. Also, undersea environment is complex with large background information. How to overcome these difficulties and extract the information of interesting, has attracted many researchers.

The image used in this paper is a HD sponge image captured by "Jiao Long" DSV in Seamounts Area of the South China Sea during its voyage of experimental application. Although target feature is obvious in the image, the complex background and large amounts of suspended particle impurities will be as noise causing a serious influence on texture extraction in later steps. Matlab software has powerful scientific computing capabilities with fast HD image processing speed and convenient methods for debugging. Simulation using MATLAB toolbox of image processing could achieve satisfactory results.

13.2 Image Analysis

Preliminary analysis of image is necessary, because it could provide reference for postprocessing. And accurate analysis result will reduce the workload of later simulation. As shown in Fig. 13.1 it is a frame RGB image captured by Jiao Long DSV (size: 1920 × 1080). Sponge feature information is obvious, but the background seamounts and huge amounts of suspended particulate impurities under strong light reflection in seawater would have a serious impact on texture extraction.

Histogram, shown in Fig. 13.2, is the most simple and effective tool in digital image processing, which is a function of the gray level. The horizontal axis is gray level, the vertical axis is the number of pixels in the corresponding gray levels.

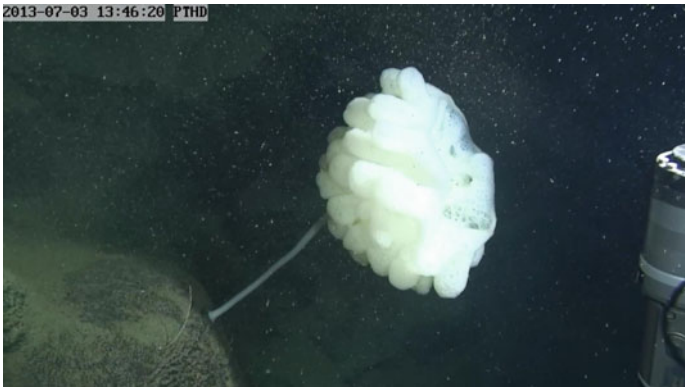


Fig. 13.1 The original sponge image

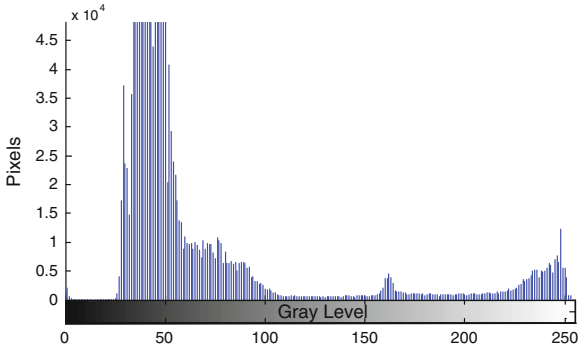


Fig. 13.2 Histogram of Fig. 13.1

Due to the gray-level histogram is quantitative analysis of an image, it tends to bring a lot of considerable information for researchers. As a result, it shows that the vast majority of pixels on gray-level range within [30 50], these low gray-level pixels constitute the background; and according to the two small waves in the gray value for [155 170] and [220 256], we could conclude that pixels within [220 256] constitute the sponge body information.

13.3 Gray-Level Transformation

Gray-level transformation enhancement is a kind of point operation. The number of pixels remains the same after the transformation, gray difference becomes larger, the contrast increases, and the image quality is optimized. The diagram of the linear gray-level transformation is shown in Fig. 13.3.

Pixels within [a b] of the original image $f(x, y)$ concentrated in [c d] of $g(x, y)$ after linear transformation.

$$g(x,y) = \begin{cases} c & f(x,y) < a \\ c + \frac{d-c}{b-a}(f-a) & a \le f(x,y) < b \\ d & f(x,y) \ge b \end{cases} \tag{13.1}$$

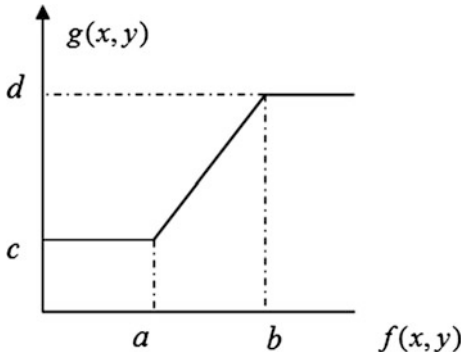


Fig. 13.3 Linear gray-level transformation

Fig. 13.4 Gray-level transformation of the sponge image. **a** contrast = 28.0047. **b** contrast = 40.0722. **c** contrast = 39.4223

Table 13.1 Comparison of the effect under different range of gray-level transformation

	(a)	(b)	(c)
Threshold transform [a b] → [0 256]	[180 256]	[52 256]	[30 256]
Contrast	28.0047	40.0722	39.4223
Sponge feature	incomplete	complete	complete

Note [a b] in the table corresponds to it in Fig. 13.3

Although the transformation compress pixels in gray levels within $< a$ and $\geq b$ into c and d, causing some loss of information, but it could precisely remove most unwanted background information, while increasing the contrast in deep-sea image processing [1].

According to the reference of histogram, a series of simulations are carried out by transforming pixels within [180 256], [52 256] and [30 256] into [0 256]. As is shown in a, b, and c of Fig. 13.4, sponge feature is gradually complete with the reduction of a, but noise increases more and more. Figure 13.4b has the strongest contrast, complete sponge feature, and little noise, shown in Table 13.1.

13.4 Median Filtering

Median filter is a kind of nonlinear filters. It achieves smooth effect by replacing the value of the target point with the middle value of its domain [2]. The simulation results show that median filter has quite a good filtering effect on isolated noise.

In one-dimensional case, median filter is a sliding window with odd number of pixels. Let $\{x_i, i \in I\}$ be the input sequence, I and m denote natural number set and window length. The output of the median filter is f_i.

$$f_i = Med\{x_{i-n}, \ldots, x_i, \ldots, x_{i+n}\} \tag{13.2}$$

where, $i \in I$; $n = \frac{(m-1)}{2}$.

Since the digital image can be seen as a sequence of two-dimensional, so its median filter is to extend one-dimensional case to a two-dimensional one. Filtering

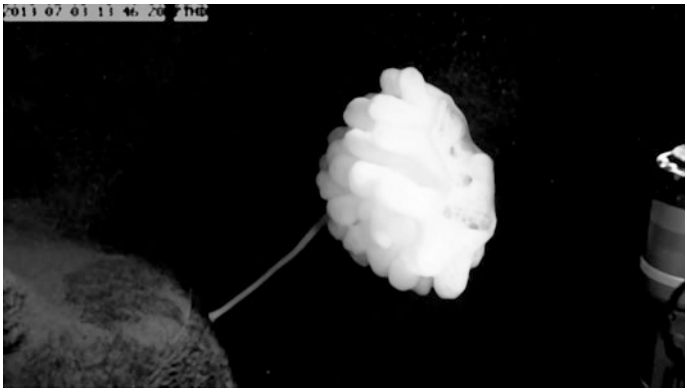


Fig. 13.5 Median filter for the sponge image in Fig. 13.4b

window is also two-dimensional, shape and size of the window depends on actual needs [3]. Therefore, two-dimensional median filter can be expressed as

$$f_{ij} = \sum_{A} \text{Med}\{x_{ij}\} \tag{13.3}$$

where, A is filtering window (Fig. 13.5).

As a result, the image inevitably appears fuzzy, but the sponge texture information is retained and the vast majority of image noise has been removed. Median filter reflects good filtering effect on such images, which is unattainable for Mean Filter and Wiener Filter.

13.5 Texture Feature Extraction

Texture feature extraction is based on edge information, which exists between target and background, caused by gray-level jump. Edge detection acquires texture information of the target image by analyzing gray jump of the image in the method of convolution or kind of convolution. Therefore, edge detection is an important means of computer vision identification. In general, edge detection algorithm is divided into four basic steps: Filter, enhancement, detection, and location. The commonly used edge detection operators are Roberts Operator, Sobel Operator, Prewitt Operator, LoG Operator, and Canny Operator.

Prewitt Operator: This is a kind of edge detection operator with direction, both gray-level jump and the jumping direction should be considered [4].

Prewitt operator has masks in eight directions, the two masks shown in Fig. 13.6 are in the horizontal direction and vertical direction with the entire weight of the mask is zero. That is to say, calculation of the mask is zero when all pixels have the

1	1	1
1	-2	1
-1	-1	-1

1	1	-1
1	-2	-1
1	1	-1

Fig. 13.6 Edge detection mask in the horizontal and vertical directions

same gray level. And greater difference of the gray level in the domain achieves greater calculation.

LoG (Laplacian-Gauss) Operator: This algorithm is the combination of the Gaussian filter and the Laplace edge detection algorithm. It gets results by convolution operation, taking the noise causing by differential algorithm or difference algorithm into consideration.

$$g(x,y) = \left[\left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \right] * f(x,y) \tag{13.4}$$

where $f(x,y)$ is the input image; σ is Gaussian radius.

Canny Operator: The calculating procedure of Canny Operator is as follows [5]:

1. To smooth sponge image with Gaussian filter.
2. To calculate amplitude and direction of the gradient with the finite difference of first-order partial derivative.
3. To process gradient magnitude with nonmaximum suppression.
4. To detect and connect edges with double-threshold method.

The key to edge detection is to suppress noise and accurately locate the position of the edge. Canny operator achieves optimization between locating accuracy and SNR [6] (Fig. 13.7).

As a result of simulation contrast, three operators achieve optimal texture feature extraction when the gradient threshold, respectively, is 0.01, 0.0008, and 0.04. Further comparison shows that Canny operator has the best results.

Fig. 13.7 Edge extraction with Prewitt operator, Log operator, and Canny operator. **a** Prewitt. **b** Log. **c** Canny

13.6 Mathematical Morphology Processing

Morphological processing of digital images is based on mathematical morphology, which includes four basic arithmetic operators: Dilation, erosion, opening, and closing operation [7]. Through combining these operators, researchers could analyze the structure of images to realize functions, such as image segmentation, edge detection, image enhancement, and restoration.

In this step, dilation operation connects areas with shorter distance in the image. But the disadvantage is that it also dilates small miscellaneous points (Fig. 13.8).

Each shaded area in the figure above, respectively, denotes object region $A(x, y)$, structure element $B(x, y)$ (Plus is the origin.) and dilation result C.

$$C = \{(x, y)|(x, y) \in A, B(x, y) \cap A \neq \phi\} \tag{13.5}$$

Structure element $B(x, y)$ traverses object image$A(x, y)$. Once the origin coincides with object point, save the point and the area corresponding with the rest of points in $B(x, y)$ [8].

Compared with texture image without dilation operation, the texture in Fig. 13.9 is more obvious. Discontinuous textures get connected and peripheral contour of the sponge is complete, which could be used as an effective basis for size estimation and image recognition.

13.7 Size Estimate

The ultimate goal of image processing is to obtain valuable information, providing condition and reference for further study. Figure 13.10 shows partial sponge image of Fig. 13.9 after inverse operation and size marking.

The parallel segment marked by blue arrow is the texture of hypothetical laser ruler, whose width is L, after processing through the above steps. This is possible in theory, because the laser ruler is continuous high-frequency signals, existing gray-level jump on the sponge. So the laser ruler is easy to be detected as edge

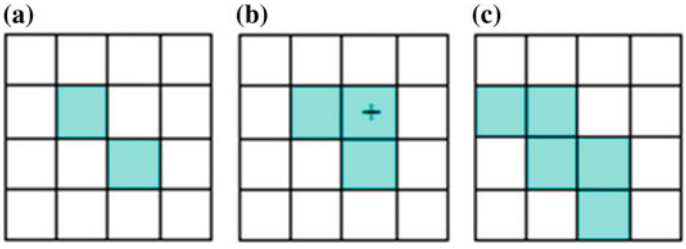


Fig. 13.8 Demonstration figure of dilation operation. **a** object region. **b** structure element. **c** dilation result

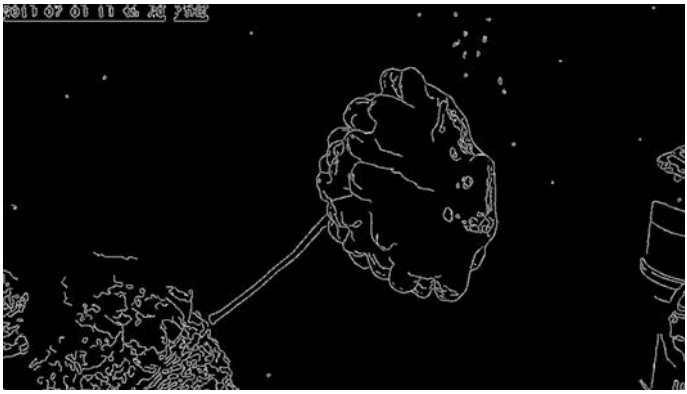


Fig. 13.9 Sponge image after dilation operation

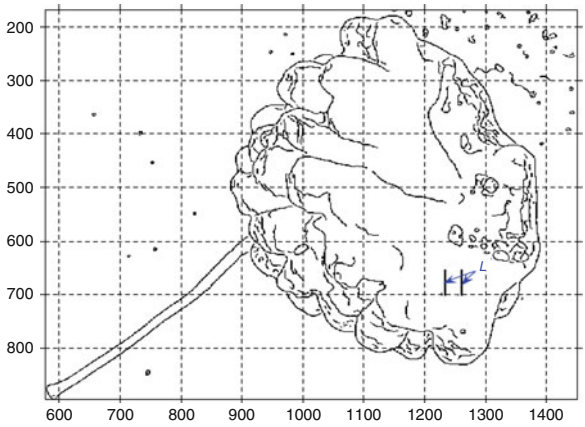


Fig. 13.10 Partial sponge image with hypothetical laser ruler

information by Canny operator. Based on the coordinate system of Fig. 13.10, scale factor between the actual object and its image, could be calculated through the laser ruler and its width in the image. Furthermore, some parameters could be calculated as shown in Table 13.2.

Table 13.2 The estimated size of some parts of the sponge

	Stem length	Stem width	Coronal size
Coordinate dimension	442.54	16.55	672.79
Estimated size	16.39*L*	0.61*L*	24.92*L*

13.8 Conclusions

With the exploration and research of the deep ocean, huge reserves and value accelerated the pace of development and utilization to the deep-sea resources in recent years. Image processing technology has become an effective means of data mining in this process. This paper has introduced histogram analysis, gray-level transformation, median filtering, edge detection, and morphological processing in the image processing. Amounts of simulation and comparison have been carried out according to characteristics of deep-sea image. From above analysis, it can be concluded that gray-level transformation has good effect on removing useless seamount backgrounds; the suspended particles impurities in seawater could be easily cleared by median filtering and texture features of the sponge extracted by Canny operator is applicable for further study. Although the size estimation is based on hypothetical laser ruler, it is reliable in theory. And the laser ruler will be installed on Jiao Long DSV soon.

References

1. Raji A, Thaibaoui A, Petit E et al (1998) Gray-level transformation-based method for image enhancement. Pattern Recogn Lett 19(13):1207–1212
2. Pan MS, Tang JT, Yang XL (2011) An adaptive median filter algorithm based on B-spline function. Int J Autom Comput 8(1):92–99
3. Demigny D (2002) On optimal linear filtering for edge detection. IEEE Trans Image Process 11 (7):728–737
4. Xishan T (2012) A novel image edge detection algorithm based on Prewitt operator and wavelet transform. Int J Advancements Comput Technol 4(19):73–82
5. William M (2011) The canny edge detector revisited. Int J Comput Vision 91(3):251–261
6. Zhao XM, Wang WX, Wang LP (2011) Parameter optimal determination for Canny edge detection. Imaging Sci J 59(6):332–341
7. Michielsen K, De Raedt H, Kawakatsu T (2000) Morphological image analysis. Comput Phys Commun 132(1):94–103
8. Isabelle B (2009) Duality versus adjunction for fuzzy mathematical morphology and general form of fuzzy erosions and dilations. Fuzzy Sets Syst 60(13):1858–1867

Chapter 14
An Improved GAFSA Based on Chaos Search and Modified Simplex Method

Pei-zhen Peng, Jie Yuan, Zhao-jia Wang, Yi Yu and Min Jiang

Abstract This paper combines the dynamically adjusting parameters, the chaos search (CS), and the modified simplex method (MS) with GAFSA, and the CS_MS_GAFSA is proposed. The algorithm speeds up the convergence by dynamically adjusting the parameters, and increases the probability of artificial fish escaping local extreme points by chaotic search for the current global optimum value. When the algorithm converges to the global optimum nearby, a simplex is constructed and the algorithm switches to MS which will continue to optimize until a certain stop condition is satisfied. Take the best point of simplex vertex at this time as the optimal value. The computational results on benchmark functions show that CS_MS_GAFSA does improve in optimizing accuracy and convergence speed.

Keywords Artificial fish swarm algorithm (GAFSA) · Global optimization · Dynamically adjusting parameters · Chaos search · Modified simplex method

14.1 Introduction

Artificial Fish Swarm Algorithm(AFSA) was proposed by Dr. Li Xiao-lei in 2002 [1]. By studying fish behavior characteristics, and applying the animal autonomous body model, the simulation of several typical fish behaviors (foraging behavior, swarming behavior, rear-end behavior, and random behavior) for optimization was proposed. AFSA is a new swarm intelligence optimization strategy, which has the advantages of strong robustness, global convergence, and not sensitivity to initial value. But it has also some obviously insufficient regards such as low optimizing accuracy and a slow convergence rate in the latter part of the algorithm and easy to

P. Peng (✉) · J. Yuan · Z. Wang · Y. Yu · M. Jiang
Key Laboratory of Measurement and Control of Complex Engineering of Ministry of Education, School of Automation, South East University, Nanjing, Jiangsu, China
e-mail: 213123683@seu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_14

fall into local minim [2]. In AFSA, the artificial fish only uses local optimal information to update location information without the use of global optimal. In order to improve the performance of AFSA [3], proposed a global AFSA (GAFSA) which adds the information of the global optimum artificial fish to the position update formula. The optimization efficiency of GAFSA improves indeed. But there are still problems of much invalid calculation in the latter part of the optimization process and not able to set a convergence accuracy [4] makes a disturbance to artificial fishes on flat areas in the process of artificial fish search to avoid falling into local optimum, but still cannot effectively solve the problem of low accuracy [5–7]. Hence combine AFSA with other intelligent methods to improve the accuracy of the algorithm. However, due to the presence of a random search, it cannot effectively improve the convergence speed.

14.2 A Brief Introduction to Basic AFSA and GAFSA

AFSA is the process of searching the optimal solution by using N artificial fish in n-dimensional space. As shown in Fig. 3.1, the current state of an artificial fish is X with the corresponding function value as $f(X)$, and X_v is a state within view field of this artificial fish. If $f(X_v)$ is smaller than $f(X)$ (i.e., the concentration of food at X_v is higher than X), make a step along this direction reaching the state X_{next}. On the contrary, continue to tour other locations within the current view field. The more the number of visits for, the more comprehensive understanding of the state within view field and it helps to make more accurate judgments and behavior decisions. For multi-state or infinite-state of the environment it does not have all traversal and allows the artificial fish making local optimization of some uncertainty which is helpful for looking the global optimum (Fig. 14.1).

Through the study of fish activity, apply its four acts (i.e., foraging behavior, swarming behavior, rear-end behavior, and random behavior) to the optimization

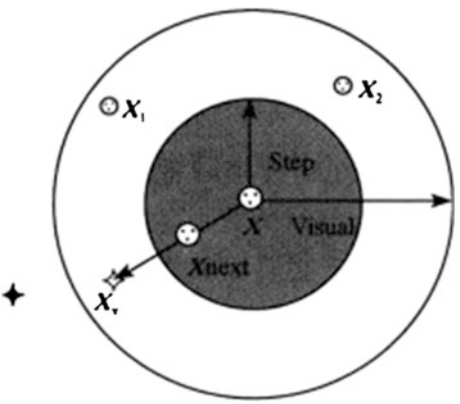


Fig. 14.1 The state of an artificial fish with field vision

process of the AFSA. These behaviors can be converted to each other under different conditions. By the evaluation of the four acts, fish selects the optimal behavior, in order to reach a location with higher concentration of food.

In order to improve the global search ability of AFSA and to overcome its shortcomings such as low precision and easy precocious, GAFSA adds the information of global optimum artificial fish to the update mode above. The parameters and artificial fish behavior in GAFSA are the same as in the basic algorithm. The only change is to add another new definition $X_{\text{best\_}af}$ that is location information of global optimum artificial fish. In GAFSA, artificial fish moves to the vector sum direction of the best artificial fish within the view field and the current global optimum artificial fish, thereby to accelerate the convergence rate of the algorithm. Simulation results show that GAFSA is superior to basic AFSA in optimization. The disadvantage of GAFSA is that the algorithm increases the amount of calculation, presents slow convergence in the latter part of the optimization process, carries too many invalid calculations, and the convergence precision cannot be prespecified.

14.3 An Improved Global Artificial Fish Swarm Algorithm

In GAFSA, parameters are fixed and the algorithm often presents a faster convergence in optimization of the early. However, the algorithm shows a slow convergence in the latter part. For this reason, this paper makes dynamical adjustment to the view field of artificial fish and crowding factor δ. In fact, the larger the Visual, the stronger the algorithm's global search capability is while the more probability of invalid calculation is. And when seeking a minimum value, the smaller the δ, the stronger of the algorithm's global convergence ability is. In the early to choose a larger view field and a smaller crowding factor will effectively improve the global search ability, convergence speed, and optimization accuracy of the algorithm. Specific adjustment rules are as follows:

$$\begin{cases} \text{Visual} = \text{Visual} \cdot \beta + 0.5, \ \text{if Visual} > 2 \\ \text{Visual} = 2, \ \text{if Visual} < 2 \\ \delta = \delta \cdot (1.5 - \beta) + 1, \ \text{if } \delta > 1 \\ \delta = 1, \ \text{if } \delta < 1 \\ \delta \leq 0.7 \cdot N \\ \beta = \exp(-30 \cdot (g/G)^s) \end{cases} \tag{14.1}$$

where g is the number of the current iteration and G the maximum number of iterations (according to the simulation experience here take as 100); s is an integer greater than 1 (here take a value from 1, 3, 5, and 10); Visual and δ can be set the initial value according to the problem which will be solved.

In the swarm intelligence optimization algorithm, due to the presence of random search, the algorithm may escape from local optima with a certain probability. Practice shows that because of ergodicity a chaotic search can be used as an effective

mechanism to prevent the optimization algorithm falling into local optima [8–10]. Its effect is often better than a simple random search. In the iterative process, CS_MS_GAFSA make a chaotic search for the current global optimum artificial fish to avoid artificial fish staying in the vicinity of a local minimum for long time. Thereby it will improve global convergence and optimization efficiency of GAFSA.

For the optimization problem as described in formula (2.1), each time the number of iterations in the corresponding chaotic search of CS_MS_GAFSA is for 20. Specific steps of the chaotic search described as follows:

a. Set $k := 0$. Take the best artificial fish on the bulletin board as $X^{(k)}$ and use the following formula to convert vector $X^{(k)}$ into chaotic one $cX^{(k)}$.

$$cx_j^{(k)} = \frac{x_j^{(k)} - x_{\min,j}}{x_{\max,j} - x_{\min,j}},\ j = 1, 2, \ldots, n \tag{14.2}$$

b. Produce chaotic vector of the next generation $cX^{(k+1)}$ by using chaos mapping.
c. Map chaotic vector $cX^{(k+1)}$ to decision one $X^{(k+1)}$ by using the following formula.

$$x_j^{(k+1)} = x_{\min,j} + cx_j^{(k+1)}(x_{\max,j} - x_{\min,j}),\ j = 1, 2, \ldots, n \tag{14.3}$$

d. Evaluate the merits of the new decision variable $X^{(k+1)}$.
e. If $X^{(k+1)}$ is better than $X^{(k)}$, output $X^{(k+1)}$ as the result of chaotic search. Otherwise, set k: = k + 1 and return to step b until $k \geq 20$.

Note: the logistic mapping is selected in step b. The iterative equation of logistic mapping is as Eq. (14.4).

$$Z(k+1) = \mu \cdot Z(k) \cdot (1 - Z(k)) \tag{14.4}$$

where μ is a control parameter. When $\mu = 4$, the formula is in a chaotic state.

Based on the dynamically adjusting parameters and chaos search strategy, CS_MS_GAFSA switches to MS to continue optimization when the search is close enough to the global optimum and after repeated iterations satisfies the formula (14.5).

$$\left| Y_{\text{best\_}af}^{(g)} - Y_{\text{best\_}af}^{(g-10)} \right| < \varepsilon_1 \tag{14.5}$$

where $Y_{\text{best\_}af}^{(g)}$ is the function value of the best artificial fish in the gth iteration and $Y_{\text{best\_}af}^{(g-10)}$ is one in the $(g-10)$th iteration, ε_1 is the precision prescribed.

When the difference of optimal solutions within 10 generations does not exceed ε_1, it can think that GAFSA has iterated to the global optimum nearby, and to continue iterating will make optimization speed slow down and optimization efficiency reduced. In order to accelerate the convergence speed, switch to MS [11] to continue optimization.

Make the best position X_{best_af} at this time as the starting point X to build a simplex as shown in Fig. 4.1. According to certain rules MS makes expansion, contraction, reflection, and other actions to build a new simplex. When termination condition (14.7) is satisfied, take the best point X^L of the simplex at this time as the optimal value. After testing, we found that the initial step length 0.05, the expansion factor 2, and the shrink factor 0.5 will make the algorithm perform better.

$$\bar{X} = \frac{1}{n}\sum_{j \neq H} X^{(j)} \tag{14.6}$$

$$\text{error} = \sqrt{\frac{1}{n+1}\sum_{j=0}^{n}\left[f(X^{(j)}) - f(\bar{X})\right]^2} < \varepsilon_2 \tag{14.7}$$

where, $\bar{X}$ is the centroid of $(n+1)$ vertices in current simplex after "worst point" X^H removed and $\varepsilon_2 > 0$ is the convergence precision prescribed ($\varepsilon_2 < \varepsilon_1$).

The process of CS_MS_GAFSA as follows:

1. Initialize the algorithm including locations of artificial fish, the number of artificial fish N, the max step *Step*, the view field *Visual*, the number of attempts *try_number*, the crowding factor δ, etc.
2. Calculate objective function values of all artificial fish and record the best artificial fish on the bulletin board.
3. Make artificial fish to perform foraging behavior, swarming behavior, rear-end, and random behavior, and evaluate the results obtained. If the state after execution is better than the current state, the artificial fish will move a step towards this direction.
4. Make chaotic search to the current best artificial fish for several times.
5. Adjust the parameters Visual and δ.
6. Update bulletin board.
7. If the difference of the optimal values in ten generations is less than ε_1, go to 8, otherwise go to 3.
8. Take the current optimal state to construct a simplex and continue optimization until the termination condition is satisfied. Output the best point on the simplex as the optimal solution.

14.4 Simulation and Results

14.4.1 Functions

In order to more directly reflect the merits of AFSA, GAFSA, and CS_MS_GAFSA, the three algorithms are tested on 34 benchmark functions [12]. Six typical examples are picked out to test the modified algorithm.

1. Sphere model function

$$f_1(X) = \sum_{i=1}^{n} x_i^2, \quad \text{with} -5.12 \le x_i \le 5.12$$
$$\min(f_1) = f_1(0, \ldots, 0) = 0. \tag{14.8}$$

2. Sum of different power function

$$f_2(X) = \sum_{i=1}^{n} |x_i|^{(i+1)}, \quad \text{with} -1 \le x_i \le 1$$
$$\min(f_2) = f_2(0, \ldots, 0) = 0. \tag{14.9}$$

3. Beale function

$$f_3(X) = [1.5 - x_1(1 - x_2)]^2 + \left[2.25 - x_1\left(1 - x_2^2\right)\right]^2 + \left[2.625 - x_1(1 - x_2^3)\right]^2,$$
$$\text{with} -4.5 \le x_i \le 4.5$$
$$\min(f_3) = f_3(3, 0, 5) = 0. \tag{14.10}$$

4. Hartmann function 1

$$f_4(X) = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{3} A_{ij}(x_j - P_{ij})^2\right), \quad \text{with } 0 \le x_i \le 1 \tag{14.11}$$

$$\min(f_4) = f_4(0.114614, 0.555649, 0.852547) = -3.86278$$

where:

$$a = [\,1 \quad 1.2 \quad 3 \quad 3.2\,]$$
$$A = \begin{bmatrix} 3 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix}$$
$$P = \begin{bmatrix} 0.36890 & 0.11700 & 0.26730 \\ 0.46990 & 0.43870 & 0.74700 \\ 0.10910 & 0.87320 & 0.55470 \\ 0.03815 & 0.57430 & 0.88280 \end{bmatrix}$$

5. Matyas function

$$f_5(X) = 0.26(x_1^2 + x_2^2) - 0.48x_1x_2, \quad \text{with} -10 \le x_i \le 10$$
$$\min(f_5) = f_5(0, 0) = 0. \tag{14.12}$$

6. De Jong's function 4(no noise)

$$f_6(X) = \sum_{i=1}^{n} ix_i^4, \text{ with } -1.28 \le x_i \le 1.28$$
$$\min(f_6) = f_6(0, \ldots, 0) = 0. \tag{14.13}$$

14.4.2 Data Analysis

The test results are shown in Table 14.1. Test software platform is MATLAB2013b with the operating system as Windows8.1.

Table 14.1 The result of test

F	D	P	FV	AFSA	GAFSA	CS_MS_GAFSA
f1	30	IT = 50 N = 20	TFV ASV BSV WSV NFC	0 35.3775 17.6139 65.5788 34800.9	0 0.00659085 0.000334334 0.0144631 25533.9	0 4.76E-06 1.27E-06 1.18E-05 26935.3
f2	20	IT = 50 N = 25	TFV ASV BSV WSV NFC	0 0.0235393 0.00534155 0.0423571 49429.6	0 5.80E-08 1.41E-09 2.95E-07 61113.8	0 2.99E-07 1.04E-07 1.29E-09 68172.9
f3	10	IT = 30 N = 20	TFV ASV BSV WSV NFC	0 0.0174986 0.0014787 0.0390852 30487.9	0 0.00659085 0.000334334 0.0144631 5533.9	0 4.76E-06 1.27E-06 1.18E-05 26935.3
f4	30	IT = 50 N = 26	TFV ASV BSV WSV NFC	-3.86278 -3.85178 -3.86222 -3.79823 71661.8	-3.86278 -3.82814 -3.8627 -3.75473 65765.1	−3.86278 −3.86278 −3.8627801 −3.8627679 71733.8
f5	12	IT = 30 N = 20	TFV ASV BSV WSV NFC	0 0.00657138 0.000277413 0.0177118 32234.4	0 0.0004272 1.81E-05 0.00150261 25030.4	0 3.99E-06 1.74E-06 1.08E-05 26697.6
f6	20	IT = 50 N = 25	TFV ASV BSV WSV NFC	0 2.29328 1.01664 4.29256 49436	0 1.30E-06 1.96E-07 3.28E-06 64309.9	0 1.22E-06 6.76E-08 4.01E-06 73466.5

Note Table 14.1, *f* function, *D* dimension, *P* parameter, *FV* function value, *N* number of artificial fish, *IT* number of iteration, *TFV* theoretical function solution, *ASV* average solution value(from the optimization of 50 times), *BSV* best solution value, *WSV* worst solution value, *NFC* number of calculation. And take ε_1 as 10e−4–10e−3 and ε_2 as 10e−5

As can be seen from Table 14.1, when the number of calculating the function is substantially the same, AFSA shows a slow convergence and low accuracy. After the global optimum artificial fish information is added to AFSA, convergence accuracy has been improved significantly. And for more prominent local extreme function, artificial fish in GAFSA can easily escape from local minima and quickly reach the global minimum. Due to the dynamically adjusting parameters, making chaotic search in the current optimal artificial fish and in the latter part of the optimization switching to MS, Global convergence of CS_MS_GAFSA is further enhanced, and its optimization convergence speed and accuracy have also been improved.

14.5 Conclusion

In order to overcome the flaws of GAFSA, such as much invalid calculation in the latter part and not able to set accuracy, the chaos search and modified simplex method are used to improve the algorithm. Based on GAFSA, CS_MS_GAFSA combines with dynamic adjustment parameters to make a chaotic search on the current best artificial fish. Thus enhance the ability of artificial fish to escape local minima, and improve the convergence speed and precision of optimization. When the iteration converges the vicinity of the global optimum, switch to the MS and continue to optimize until the precision condition is meted. 34 Benchmark Functions are used to simulate for GAFSA and CS_MS_GAFSA. Results show that CS_MS_GAFSA can improve the original algorithm on the convergence speed, accuracy, and global optimization convergence. It is effective and feasible.

Acknowledgments This work was part of the program on The National Natural Science Foundation of China funded under the National Natural Science Fund Committee, grant number 61374198.

References

1. Li XL, Qian JX (2003) Studies on artificial fish swarm optimization algorithm based on decomposition and coordination techniques. J Circuits Syst 2(1):1–6 (in Chinese)
2. Jiang MY, Yuan DF (2012) Artificial fish swarm algorithm and its application. Science Press, Beijing (in Chinese)
3. Jiang MY, Cheng YM, Yuan DF (2014) Improved artificial fish swarm algorithm, 5th edn
4. Ma XM, Liu N (2014) Improved artificial fish-swarm algorithm based on adaptive vision for solving the shortest path problem. J Commun 1, 35(1):1–6
5. Zhu KC, Jiang MY (2009) An improved artificial fish swarm algorithm based on chaotic search and feedback strategy. In: Proceedings of international conference on computational intelligence and software engineering (CISE), 2009, 1–4
6. Zhu KC, Jiang MY (2010) Quantum artificial fish swarm algorithm. In: Proceedings of the 8th world congress on intelligent control and automation, 2010, pp 1–5

7. Jiang MY, Cheng YM (2010) Simulated annealing artificial fish swarm algorithm. In: 8th world congress on intelligent control and automation, WCICA, 2010, pp 1590–1593
8. Zhang P, Wei P, Yu HY, Fei C (2012) Biogeography-based optimization algorithm by using chaotic search. J Univ Electron Sci Technol China 41(1):65–69
9. Liu H, Huang M, Huang D (2005) Survey on optimization method based on chaotic search and its developments. J Nanjing Univ Sci Technol 124–128
10. Bucolo M, Caponetto L Fortuna et al (2002) Does chaos work better than niose? IEEE Circuits Syst Mag 2(3):4–19
11. Yang XQ, Teo KL, Caccetta L (2001) Optimization methods and applications. Kluwer Academic Publishers, Boston
12. Rahnamayan S, Tizhoosh HR, Salama MMA (2008) Opposition-based differential evolution. IEEE Trans Evol Comput 12(1):64–79

Chapter 15
A Multi-modal Searching Algorithm in Computer Go Based on Test

Xiali Li and Licheng Wu

Abstract Pattern matching algorithms based on domain knowledge have a weak global sense. Monte Carlo algorithms have a weak tactical ability. This article proposed a multi-modal algorithm on the basis of tests. GNU Go software was used to make experiments. Through the analysis of the experiment data, we set the move thresholds of start, middle, and the end of a Go play game. In the start stage, pattern matching algorithm was used. In the middle stage, Monte Carlo tree search algorithm with pruning was used to search the optimal moves. In the end stage, pattern matching and crazy model were used. The multi-modal algorithm was applied to develop software. Experiments demonstrate that multi-modal algorithm can improve the performance of Go in 13 × 13 board and 19 × 19 board.

Keywords Go · Pattern matching · Monte Carlo · Pruning · Domain knowledge · Multi-modal

15.1 Introduction

The rules of Go are simple but the search space is very large, so that it is very difficult to improve performance of Go program [1]. The search algorithm is the key factor to evaluate the performance of Go program. The world's popular Go program mainly used [2, 3] the search algorithm that estimates liberty of each stone and the influence of its territory, the algorithm that uses patterns library and patterns representation and extraction to search the optimal moves, UCT (Upper Confidence Bound applied to Tree) algorithm, Monte Carlo Tree Search algorithm. UCT [4, 5] is UCB (Upper Confidence Bounds) algorithm applied to searching tree and is mainly to solve the multibandit problems. Monte Carlo Tree Search (MCTS) [6] is a method for finding optimal decisions in a given domain by taking random samples

X. Li (✉) · L. Wu
School of Information Engineering, Minzu University of China, Beijing 100081, China
e-mail: xiaer_li@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_15

in the decision space and building a search tree according to the results. The tree is built in an incremental and asymmetric manner. For each iteration of the algorithm, a tree policy is used to find the most urgent node of the current tree. The tree policy attempts to balance considerations of exploration and exploitation. A simulation is then run from the selected node and the search tree updated according to the result. MCTS is used by many Go programs in recent years but the algorithm costs much storage space and execution time.

15.2 Related Go Program

The current popular software of computer Go has Handtalk, Goemate, NearoGo, and Gnugo. Handtalk is written in assembly language and it has good execution speed besides strong attack ability. Goemate is an improved version of Handtalk with a new mode management system. NearoGo is a typical learning program with a neural network, relationship experts, characteristics experts, and external experts [7]. Gnugo is an open source project of software alliance FSF/GNU. It ever won the Olympic champion. The 9 board in the Gnugo version 3.8 used UCT algorithm and Monte Carlo method. After producing a move, Gnugo waits for the opponent to generate move position. Then according to the opponent's position, Gnugo uses UCT algorithm based on patterns to search the optimal node. Finally, it simulates the game by Monte Carlo method to find the highest winning rate move. GNU Go is an excellent program but it wastes the time in which the opponent generates moves.

Pattern recognition based on domain knowledge establish model of territory and its influence and thus divide the whole into subproblems, which break the whole in the Go game. Monte Carlo has a good global strategy but poor tactic [8, 9]. This paper puts forward a multi-modal search algorithm based on tests. We use Gnu Go software to do the test according to the process of playing. Through the analysis of test data, set the moves number threshold of the start, the middle, and the end of the game. Different algorithms are used in the three stages. Pattern matching algorithm is used in the start stage, UCT algorithm and Monte Carlo method are used in the middle of the game. Pattern matching algorithm and crazy model are used in the end of the game. Experiments show that the multi-modal algorithm can improve the chess.

15.3 Gnugo Test

15.3.1 Introduction of Gnugo

Monte Carlo simulation in Gnugo is based on patterns, for example, 3 × 3 adjacent sub, dutch act of the opponent, send eating state, raisins number. Gnugo has mc_montegnu_classic, mc_mogo_classic, and mc_uniform pattern libraries. Among

which mc_montegnu_classic is the initial random generation algorithm approximately, mc_mogo_classic is the approximate simulation strategy used in the earlier versions of Gnugo which was released in modification of UCT with patterns in Monte Carlo Go RR-6062 by Sylvain Gelly, Yizao Wang, Rémi Munos and Olivier Teytaud, mc_uniform is the so-called "light" simulation which select the moves in all legal hands randomly except filling unique eye. After producing a move, Gnugo waits for the opponent to generate move position. Then according to the opponent's position, Gnugo uses UCT algorithm based on patterns to search the optimal node. Finally, it simulate the game by Monte Carlo method to find the highest winning rate move.

15.3.2 Analysis of the Test

We use 9 × 9, 13 × 13, and 19 × 19 boards individually in Gnugo 3.9.1 version to do the test. The experimental data and analysis are as follows.

The average time of producing one step in 9 × 9 chess board is about 35s and the average time to complete a game is about 15–25 min. It is slower for the 9 × 9 board of go. If applying UCT algorithm and Monte Carlo method to 13 × 13 and 19 × 19 board of go, it requires much longer time. In addition, Monte Carlo method is based on the theory of probability and statistics and so the optimal method of generating moves is not absolutely optimal. So in the beginning and ending stage of the game, the moves produced only by the method of probability statistics is often unacceptable and is inferior to those of search methods based on pattern libraries.

We played 50 turns Go games between man and the machine using 13 × 13 and 19 × 19 board individually. Data is shown in Table 15.1.

Test results show that in the 13 × 13 board of Go the average number moves in the start process is about 15 moves. In the 19 × 19 Go board, the average moves in the start of the game is about 25.

In the front 50 moves of the games, using searching method based on pattern libraries can get good moves. From the 50 moves to the 65 moves, adopting UCT algorithm and Monte Carlo method can produce better moves. Using the method of pattern libraries is better after 65 moves in which the game basically enter the final stage. From 70 moves to 80 moves, UCT algorithm and Monte Carlo method are appropriate. After 80 moves, the search method based on pattern library can improve the chess.

Table 15.1 Move numbers of different algorithms in Gnugo

Go board size	Pattern libraries move numbers	UCT and Monte Carlo move numbers
13 × 13	1–50, 66–end	51–65
19 × 19	1–70, 81–end	71–80

15.4 Multi-modal Search Algorithm

According to the experimental results of Gnugo, this paper proposed multi-modal search algorithm. We divide the whole game play into the start, the middle and the end according to the process of playing chess. By analysis on the data acquired from test, we set the moves number threshold of the start, middle, and the end of the game. Different algorithms are used in the three stages. Pattern matching algorithm is used in the start. Monte Carlo tree search algorithm with pruning is used in the middle of game. Pattern matching algorithm and crazy model are used to get optimal moves in the end of the go. Experiments show that the multi-modal algorithm can improve performance of the chess effectively.

According to the test of Gnugo, we set the threshold of the number of moves and use different search algorithms in each stage. Let $\phi 1$, $\phi 2$, $\phi 3$ be the threshold of the first, the second, and the third stage separately, ϕ be the number of moves in Go game which satisfy the following:

$$20 \leq \phi 1 \leq 30 \tag{15.1}$$

$$70 \leq \phi 2 \leq 80 \tag{15.2}$$

$$81 \leq \phi 3 \tag{15.3}$$

If $\phi \leq \phi 1$, pattern matching algorithm is used. Analyze domain knowledge, specially pay much attention to block, eye, connection, etc., and use fuzzy clustering neural network for pattern recognition and matching.

Neural network pattern is very effective on Go pattern recognition and fuzzy attribute detection. BP network is used to establish pattern recognition classifier for domain knowledge. Hidden layer neurons nodes convert the original nonlinear sample into linear sample and the output layer neurons complete distinction of the category. Because the category of the sample is unknown on the beginning of Go type classification, we use fuzzy clustering to do automatic classification according to the similarity between samples.

If $\phi 1 \leq \phi \leq \phi 2$, according to the stone position, we use Monte Carlo tree search algorithm with pruning in Go. Basic Monte Carlo tree search algorithm is shown in the following [10–12] Table 15.2:

Because of large searching space and branch factor, we have to prune in Monte Carlo tree to improve the efficiency. According to the different degree of strong demand on domain knowledge, we divide the nodes in the game tree into root node and nonroot node [10]. The execution speed of the search algorithm is important to improve the efficiency pruning to obtain the optimal alternative of the move. To improve the speed of the algorithm, we design different move generator for root node and nonroot node separately. The ko fight, eyes, connection, and the others based on domain knowledge move generator are designed for root node. Light move generators are designed for nonroot node. We rank the moves generated by

Table 15.2 The basic algorithm of MCTS

Pseudocode
Use the current position of Go board as the root node of T
while (total number of simulations < limit) {
Simulate one game from the root of the tree to a final position, choosing moves as below:
Choice part:
For a situation in T, choose the move with maximal score according to the UCB1 formula unless simulation games through the node exceed the threshold set.
Simulation part:
For a situation out of T, choose the move according to a play-out policy.
Update win/loss and simulations statistics in all situations of T.
Add to a new descendant node T which is not yet in T.
}
Return the move simulated most often from the root of T.

different generator and select the optimal move as the branch of the tree. Thus cut the other branches.

If $\phi \geq \phi 3$, enter the closing phase and use pattern matching algorithm and crazy model. Crazy model is triggered manually when the contest time is almost consumed completely. Thus the algorithm will move totally randomly without any search.

15.5 Experiment Results

We use Gnugo 3.9.1 version to do test. Before using the multi-modal algorithm, Gnugo 3.9.1, silver star 10 and scala Go play with Gnugo 3.9.1 individually. Silver star 10 and scala Go are popular commercial Go programs, silver star Go had won the world chess championship in four consecutive years [1].

In the play games, we set the following parameters with the 32 K simulations per move.

$$\phi 1 = 25,\ \phi 2 = 80,\ \phi 3 = 81.$$

First, we set 1000 games between Gnugo 3.9.1 with Gnugo 3.9.1, silver star 10 and scala separately. They played as the offensive and defensive position in turn. The experiment were made in 13 × 13 board and 19 × 19 board.

Then we set 1000 games between the program using multimodal algorithm with Gungo 3.9.1, silver star 10 and scala separately. They played as the offensive and defensive position in turn. The experiments were made in 13 × 13 board and 19 × 19 board. The win-rate (%) contrast was showed in Figs. 15.1 and 15.2.

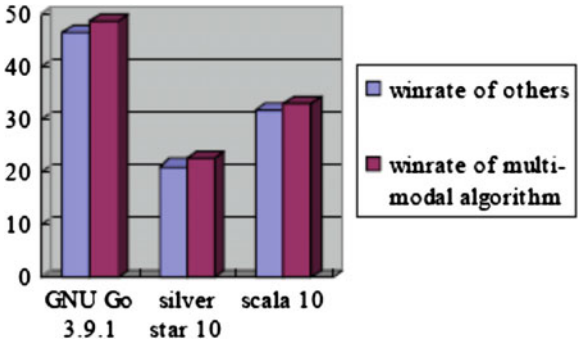


Fig. 15.1 Win-rate contrast in 13 × 13 board

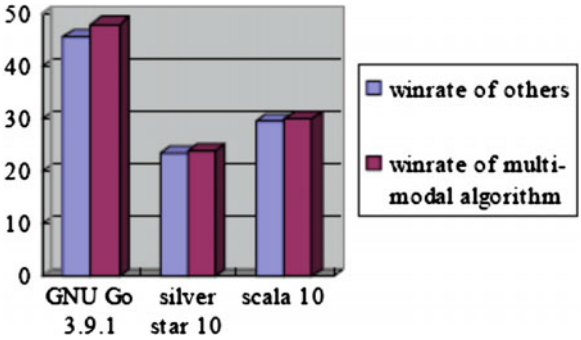


Fig. 15.2 Win-rate contrast in 19 × 19 board

We can see that the contrast of winning rate in 13 × 13 board from Fig. 15.1. The multi-modal algorithm can increase 2.4 % winning rate against GNU go, 1.7 % against silver star 10, and 1.2 % against scala Go in 13 × 13 board.

Figure 15.2 shows that the multi-modal algorithm can increase 2.5 % play against GNU Go 3.9.1 version, 0.4 % against silver star 10, and 0.7 % against scala Go in 19 × 19 board.

15.6 Conclusions

On the base of data acquired from Gnugo test, we divide the Go play into three stages and set different threshold for the moves. According to the threshold, we use multi-modal algorithm to search the tree and choose the good move sequences. This multi-modal algorithm can improve the performance of Go in 13 × 13 board and 19 × 19 board. The next future work is to study whether the optimal threshold exists.

Acknowledgments This work is supported in part by 2013 Beijing university youth talent plan No. YETP1294. Besides, it was supported in part by the NSFC project No.51375504 and the Program for New Century Excellent Talents in University.

References

1. Jun S (2005) Go and artificial intelligence. China Sport Sci Technol 41(6):135–138 (in Chinese)
2. Peng Y (2007) The studies of algorithms in computer go. doi:10.7666/d.y1075440 (in Chinese)
3. Lei Y (2010) Study on problems for computer go based on cognitive science. J East China Norm Univ. (in Chinese)
4. Kocsis L, Szepesvari Cs (2006) Bandit based Monte-Carlo planning. Mach Learn Lect Notes Comput Sci 4212:282–293
5. Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multiarmed bandit problem. J Mach Learn 47:235–256
6. Coulom R (2006) Efficient selectivity and backup operators in Monte-Carlo tree search. Comput Game Lect Notes Comput Sci 4630:72–83
7. Wei H (2010) Computer games and artificial intelligence. J Guide Sci Educ 27:106–107 (in Chinese)
8. Bouzy B (2005) Associating domain-dependent knowledge and Monte Carlo approaches within a go program. J Inf Sci 175:247–257
9. Li qun W (2007) Game of go-pattern storage and matching algorithm. J Northeast Dianli Univ 27(1):83–87 (in Chinese)
10. Chen K-H (2012) Dynamic randomization and domain knowledge in Monte Carlo tree search for go knowledge-based systems. J Knowl-Based Syst 34:21–25
11. Chalot G, Fiter C, Hoock J-B, Rimmel A, Teytaud O (2010) Adding expert knowledge and exploration in Monte-Carlo tree search. Adv Comput Game Lect Notes Comput Sci 6048:1–13
12. Coulom R (2007) Computing elo ratings of move patterns in the game of go. ICGA J 30 (4):198–208

Chapter 16
A New Smoke Detection Method of Forest Fire Video with Color and Flutter

Zhong Zhou and Ya-qin Zhao

Abstract Big region area, long distance, and fast motion are the main characteristics of smoke video. Traditional smoke video detection has a high false alarm rate. In order to improve the detection accuracy, we propose a novel method to detect smoke video based on color information and flutter analysis. The proposed method can effectively exclude the influence of those objects resembling the smoke motion and reduce false detection rate resulting from illumination change and smoke-color objects. In detail, a background updating model of forest fire smoke video is first built using the Kalman filtering method. Combining updating model with the smoke HSV and RGB color space, candidate smoke video region is segmented. Then, three flutter features, including flutter direction of the video smoke, changing rate of the smoke area, and background ambiguity, are extracted by sliding time window analysis of the candidate smoke. Finally, the characteristic values are used to judge whether it is smoke. Experiment results show that our method can detect the video smoke more accurately as well as faster than the state-of-the-art methods.

Keywords Smoke video detection · Kalman filtering · Color space · Flutter analysis · Wavelet transform

16.1 Introduction

Forests are valuable resources for human survival and development. The frequent forest fires cause severe damage to social economics, thus achieving automatic fire monitoring and analysis is an urgent need for forest fire prevention. Compared with traditional fire detection, the fire video detection technology can satisfy the current fire detection needs of a large space [1, 2], long distance, and large area.

Z. Zhou · Y. Zhao (✉)
College of Mechanical and Electronic Engineering, Nanjing Forestry University, No. 159 LongPan Road, Nanjing 210037, JiangSu, China
e-mail: yaqinzhao@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_16

Meanwhile, smoke occurs first, so smoke detection is directly related to the quality of the forest monitoring work.

There are three main smoke video detection methods. The first smoke detection method is based on the spatial and temporal features of smoke. Toreyin et al. [3] proposed a detection algorithm based on wavelet transform energy, but the wavelet calculation is complicated and the detection performance is also poor in the case of changing brightness. The second method is based on the movement of smoke.

Yuan et al. [4] used a sliding time window analysis of suspected smoke regions to detect the smoke by its cumulative amount and main movement direction. The third one is based on the static and dynamic features of smoke. Chen [5] detected the smoke based on its static feature of saturation and dynamic feature of diffusion. The methods are prone to false and missing alarms since the selected feature detection threshold greatly impacts on the detection performances.

Considering the following issues: the changing brightness of the light in the forest, the changing space background of the vegetation in different regions, as well as the situation of low smoke density in the early stage, we proposed a novel approach to smoke detection based on the color and flutter in forest smoke video. In detail, real-time background model is first built by Kalman filter, afterwards suspected smoke region is extracted by combining with the smoke color feature. Finally the sliding time window analysis is performed and the smoke is identified. Experimental results show that this method can effectively exclude the non-smoke interference and detect the smoke quickly and accurately.

16.2 Smoke Region Extraction Based on Color Information

16.2.1 Kalman Background Updating Model

Kalman filter is a linear minimum variance algorithm estimating a dynamic state sequence through the state equation and observation equation to describe a dynamic system [6, 7]. The smoke background can be considered as a stable system in the smoke video processing and some foreground objects can be considered as noise. We use Kalman filter method for the real updating model of the smoke background. Let $B(i, j, k)$ and $B(i, j, k + 1)$ denote the k-th and (k + 1)-th pixel value of the background image, respectively. $I(i, j, k)$ represents the pixel value of the k-th frame, then the recursive formula is:

$$B(i,j,k+1) = \begin{cases} B(i,j,k) + g(i,j,k) * (I(i,j,k) - B(i,j,k)) & k \neq 0 \\ B(i,j,k) & k = 0 \end{cases} \tag{16.1}$$

Default background B(i,j,0) image | Current frame I(i,j,k) | Kalman updated background B(i,j,k)

Fig. 16.1 Kalman background update

where $g(i,j,k)$ is the gain factor and can be computed by the following formula:

$$\begin{cases} g(i,j,k) = \beta * (1 - M(i,j,k)) + \alpha * M(i,j,k) \\ M_i(x,y) = \begin{cases} 1 & f\lvert I(i,j,k) - B(i,j,k-1)\rvert \geq Th \\ 0 & \text{others} \end{cases} \end{cases} \tag{16.2}$$

where $M_i(x, y)$ is the binary representation of the current moving target and Th is the threshold determined by Otsu method. The values α and β are both greater than 0 and less than 1 and the former is greater than the latter in order to separate the moving object from the background sequence and have adaptive characteristics. In this paper, we set $\alpha = 0.055$ and $\beta = 0.1$ so as to reduce the detection time of forest smoke video. The results of background updating model are shown in Fig. 16.1.

16.2.2 Suspected Smoke Segmentation

Traditional smoke segmentation model which use single color space is defective due to the complexity of the smoke color characteristic model. On the one hand, it is not easy to separate the smoke from the similar color background; on the other hand, it is not well to segment the early smoke out. Therefore, we segment the smoke in combination with HSV and RGB of the smoke color characteristic [8]. First we use adaptive characteristics to learn and update the background, subtracting the current frame from the real-time background to eliminate static interferences. Then we perform coarse segmentation of HSV and RGB color model and execute logical OR computation to exclude smoke-like objects and increase the valid range. Finally we get the whole smoke region with some necessary mathematical morphology [9, 10]. The main algorithm process is as follows (Fig. 16.2):

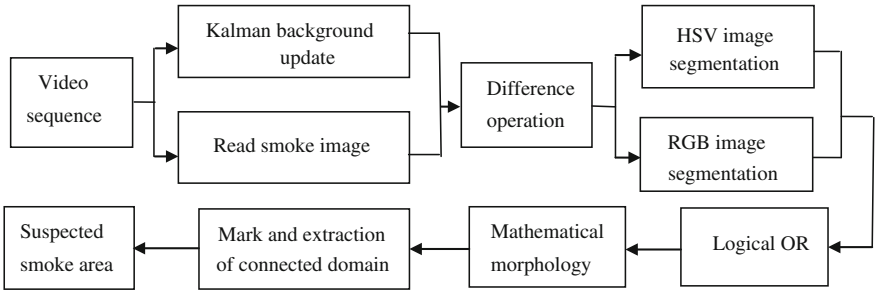


Fig. 16.2 Flowchart of suspected smoke extraction based on color

16.3 Flutter Analysis

16.3.1 Smoke Flutter Direction

Driven by the heat, Smoke moves from bottom to up. So we can detect the general moving trend during a period of time to represent the flowing direction of the smoke. In addition, the quality of massive images is concentrated on its center and therefore it can represent the movement of the whole motion of the smoke video. We calculate the center at the frame rate F_{rate} of the sliding time window, and use center coordinates sequence to compute the offset angle between two adjacent frames, and ultimately map to the direction coding [10, 11]. The $(p+q)$-th order moment of an image and the corresponding coordinates of the centroid can be expressed as the following formula:

$$m_{pq} = \sum_{i=w_s}^{w_e} \sum_{j=H_s}^{H_e} i^p j^q f(i,j) \tag{16.3}$$

$$\begin{cases} C_x(t,k) = \frac{m_{10}}{m00} \\ C_y(t,k) = \frac{m_{01}}{m00} \end{cases} \quad k = 1, 2, \ldots N \tag{16.4}$$

where, w_s and W_e represent the start and the end of the regions width respectively, and H_s and H_e denote the start and the end of the regions height, and $N = F_{\text{rate}}$ and $C_x(t,k)$ and $C_y(t,k)$ represent horizontal and vertical coordinates of the suspected smoke region centroid at the *t-th* time window and the *k*-th frame.

We use the average value of horizontal and vertical coordinates in a given time window to reduce interference caused by the air flow, which are $\overline{C_x(t)}$ and $\overline{C_y(t)}$.

$$\overline{C_x(t)} = \frac{1}{N} \sum_{k=1}^{N} C_x(t,k) \tag{16.5}$$

$$\overline{C_y(t)} = \frac{1}{N}\sum_{k=1}^{N} C_y(t,k) \tag{16.6}$$

$$\theta_t^* = \arccos \frac{\overline{C_x(t+1)} - \overline{C_x(t)}}{\sqrt{\left(\overline{C_x(t+1)} - \overline{C_x(t)}\right)^2 + \left(\overline{C_y(t+1)} - \overline{C_y(t)}\right)^2}} \tag{16.7}$$

where θ_t^* is the offset angle of centroids sequence. When the direction of vertical axis is negative, θ_t is adjusted by the following formula:

$$\theta_{\mathrm{t}} = \begin{cases} 2\pi - \theta_{\mathrm{t}}^* & \text{if} \quad \overline{C_y(t+1)} > \overline{C_y(t)} \\ \theta_t^* & \text{otherwise} \end{cases} \tag{16.8}$$

We get the value i,j of θ_{t} through its transformation shown in Fig. 16.3, for example, when $\theta_t = 35°$ then $i = 1, j = 1$. Finally, we convert it into the direction coding sequence DC_t by Eq. (16.9) and then get the direction of connected region.

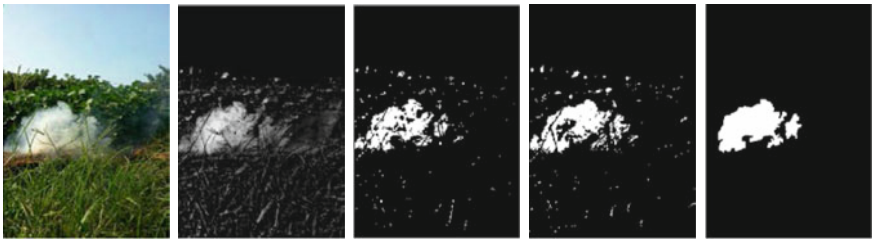
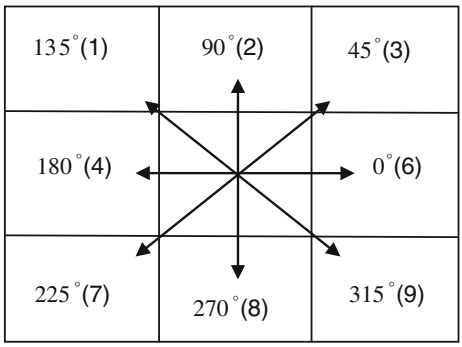


Fig. 16.3 Area extraction based on color

Fig. 16.4 Discrete encoding direction



We obtained the direction coding sequence aggregated on 1, 2, 3 because of the flow characteristics from bottom up (Fig. 16.4).

$$\mathrm{DC}_t = -3 * j + i + 5 \tag{16.9}$$

16.3.2 Changing Rate of the Flutter Smoke Area

In the early stage of the forest fire, the smoke moves slowly and expands growth with the fire and diffusion. We can decide the diffusion of the smoke by the area growth rate (ΔV_{ti}) of the suspected smoke, which is computed in the sliding time window [5]. In digital image processing, the smoke region can be represented by the number of pixels (P_i) and the time interval can be represented by the separated video frames. We have:

$$\Delta V_{\mathrm{ti}} = \frac{S_{i+k} - S_i}{t_1 - t_2} = \frac{P_{i+k} - P_i}{(i+k) - i} \tag{16.10}$$

where P_i and P_{i+k} represent the number of pixels in the i-th and $(i+k)$-th frame of the suspected smoke region.

We use the average value of the growth rate in a given time period to increase the detection accuracy and reduce the interference caused by the air flow; that is:

$$\overline{\Delta V_{\mathrm{ti}}} = \frac{1}{n} \sum \overline{\Delta V_{\mathrm{ti}}} \tag{16.11}$$

Here, $\overline{\Delta V_{\mathrm{ti}}}$ is the average growth rate. The region is smoke if it is greater than the predetermined threshold (STD) and vice versa.

16.3.3 Background Ambiguity

Smoke is sometimes partly transparent, and the texture and edge sharpness of the obscured scene region fall to be blurred, and then the high-frequency information will slowly decrease. In this paper, we use Harr wave in the gray space to extract the high-frequency energy through the energy analysis for the interest region ROI and the corresponding background region [12, 13]. The formula is:

$$E_k = \sum_{i,j \in \mathrm{ROI}} V_k^2(i,j) + H_k^2(i,j) + D_k^2(i,j) \tag{16.12}$$

Here, $V_k^2(i,j)$, $H_k^2(i,j)$ and $D_k^2(i,j)$ are the wavelet energy coefficient of horizontal, vertical, and diagonal decomposition, respectively, namely the high-frequency

energy in the k-th frame of the suspected smoke region and we can get the high-frequency energy E_b of the background area in the same way. It is as follows that the relative high-frequency energy decline ∂ of the k-th frame of the suspected smoke area:

$$\partial = \frac{E_b - E_k}{E_{\mathrm{Lb}}} \tag{16.13}$$

We use the average value, of the relative high-frequency energy decline within a time window computing ∂ every five frames, as the measurement of the smoke translucency.

16.4 Experimental Results and Analysis

16.4.1 Smoke Region Segmentation

Figure 16.5 shows the example results of smoke region segmentation and marked based on the proposed method. These images represent some challenging scenes, from left to right. They are the scenes far from camera, strong light, close to camera, and a strong wind with the sky which has similar color to smoke and a series of similar class interference. Experiment results show that smoke can be completely split in these complex environments. However, in case of strong illumination change some background pixels still are falsely detected as foreground pixels. Therefore, we combine flutter characteristics to more accurately make and locate at smoke regions.

16.4.2 Fluttering Feature Extraction

Smoke regions segmented and centroids marked of the first frame from five successive time windows of a video are shown in Fig. 16.6. As seen from Fig. 16.6, from right to left, the area of smoke region gradually increases and moving centroids shift up. Moreover, moving direction of centroids is from right to left due to wind direction.

Feature extraction results of a video clip in Fig. 16.5 are shown in Table 16.1. The features represent average value of 5 group flutter direction. We can see from Table 16.1 that the above video coding focus on direction 3, meeting the fluttering rule of smoke, and it is easy to see that the moving direction of the smoke is from left to the right. The diffusion characteristic of the smoke is verified by the average growth rate of the area, and the region is recognized as smoke one if its value exceeds the threshold 0.2. The wavelet high frequency energy of the smoke ∂ is

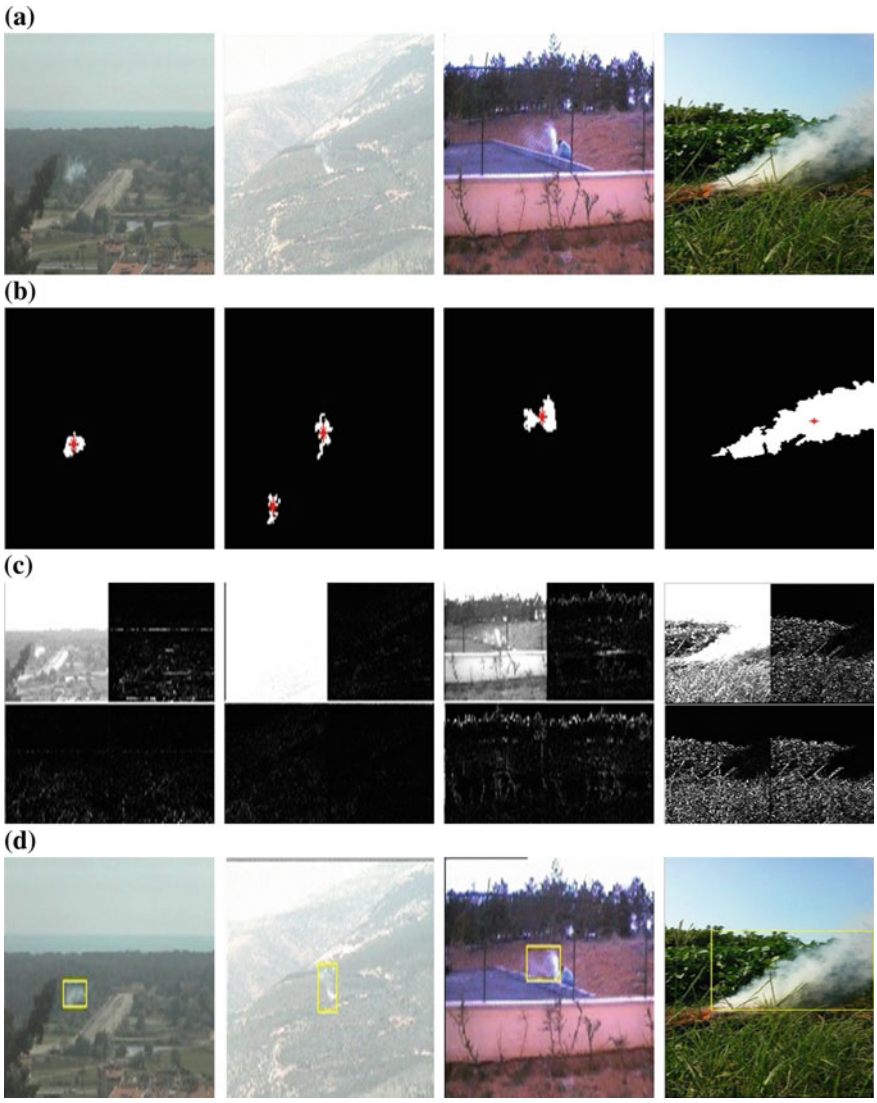


Fig. 16.5 Smoke segmentation and recognition. **a** Original video screenshot. **b** Extraction smoke suspected area and its centroid and size of the area. **c** A layer of wavelet transform edge subgraph. **d** The results of the identification and mark of smoke

decreased slowly due to the slow change of the smoke density in the open environment. Finally, we can correctly recognize the smoke by combining with the flutter feature, as shown in Fig. 16.5b–d.

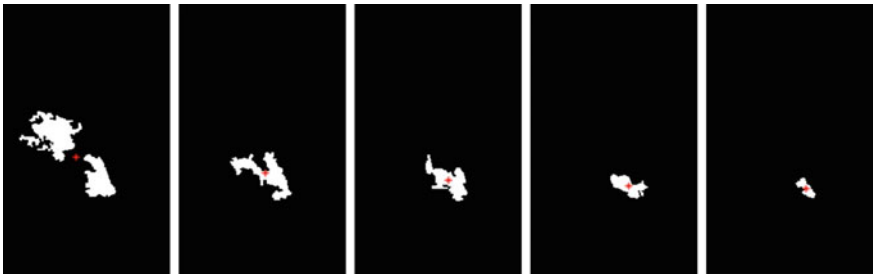


Fig. 16.6 Extraction of centroid and area of continuous video sequence

Table 16.1 Five groups flutter features of suspected smoke region

No.	Centroid coordinates	Angle and direction coding	Average rate of change in the area	Average rate of decline ∂
1	(32.27, 184.83)	61.28(3)	0.34153	0.032346
2	(36.84, 176.47)	42.55(3)	0.2449	0.054199
3	(48.51, 165.77)	18.11(3)	0.2514	0.090952
4	(71.51, 158.25)	21.78(3)	0.37229	0.12377
5	(92.09, 150.02)	7.71(3)	0.22798	0.15271

16.4.3 Smoke Recognition Results

The proposed method provides the possibility of smoke in the forest video automatically. The datasets we used are mostly from the public smoke video library. The experimental results in Table 16.2 show that our method has high true detection rate, especially in bad conditions of early smoke and complex environment.

Table 16.2 The experimental results of statistical comparison

This algorithm				Literature [8] the traditional algorithm		
Video capture	False alarm rate (%)	False negative rate (%)	Detection rate (%)	False alarm rate (%)	False negative rate (%)	Detection rate (%)
NO1	1.1	2.6	96.3	8.8	6.8	84.4
NO2	3.4	2.9	93.7	7	14.1	78.9
NO3	8.4	4.5	87.1	12.8	10.5	76.7
NO4	11.3	7.2	81.5	16.1	22.3	61.6
NO5	2.9	4.3	92.8	9.7	9.1	81.2

16.5 Conclusion

In this paper, we proposed a novel forest smoke video detection algorithm based on color and fluttering features. First, a real-time background model is built by Kalman filtering, which can effectively overcome illumination mutations and eliminate static interference, etc. Second, the suspected smoke move area is segmented through the smoke color feature. Lastly, fluttering motion analysis is performed on the suspected smoke region because of the smoke mobility and diffusion characteristics, extracting the flowing direction, change rate of the area, and background blur degree to constitute three-dimensional feature vectors, identifying and marking the smoke. Experiments demonstrate that the proposed algorithm is robust and high detection accuracy rate with a low false detection rate and false negative rate. Nevertheless, there are some limits, such as that non-smoke region sometimes identified falsely, and can be further improved by learning more representative features through pattern recognition and machine learning methods.

Acknowledgments Project supported by the National Natural Science Foundation of China (Grant No.31200496)

References

1. Hidenori M, Akihiro N, Fujio K (2012) Smoke detection in open areas with texture analysis and support vector machines. IEEE J Trans Electr Electron Eng 7(S1):S59–S70
2. Wu GL, Wu H, Zhang R et al (2013) Wildfire detection system based on image processing. Comput Technol Dev 10:227–231 (in Chinese)
3. Toreyin BU, Dedeoglu Y, Cetin AE (2005) Wavelet-based real-time smoke detection in video, pp 4–8
4. Yuan FN, Zhang YM et al (2008) Video smoke detection based on accumulation and main motion orientation. J Image Graph 13(4):808–813 (in Chinese)
5. Chen T-H(C-H), Yin Y-H, Huang S-F et al (2006) The smoke detection for early fire-alarming system based on video Processing. In: IEEE International Conference on intelligent information hiding and multimedia signal processing, Pasadena, California, USA, 18–20 Dec 2006, pp 427–430
6. Li H-b, Tang P-J, Wu Y (2009) Mixture gaussian background modeling improved by Kalman filtering. Comput Eng Appl 45(24):162–164 (in Chinese)
7. Yan Y, Huang XY, Liu AJ (2006) Application of Kalman filter in the moving image of background extraction and updating. Autom Instrum 2:28–30 (in Chinese)
8. Jiang D-H (2006) Research on fire smoke detection based on image processing. In: Master's degree thesis of Huazhong University of Science and Technology, pp 16–23 (in Chinese)
9. Horng WB, Peng JW, Chen CY (2005) A new image-based real-time flame detection method using color analysis. In: Proceedings of IEEE International Conference on networking, sensing and control. IEEE Press, Tucson, Arizona, pp 100–105
10. Zhao XC et al (2013) MATLAB Actual combat of digital image processing. China Machine Press, Beijing, pp 112–161
11. Wang T, Liu Y, Xie ZP (2011) Flutter analysis based on video smoke detection. J Electron Inf Technol 33(5):1024–1029 (in Chinese)

12. Calderara S, Piccinini P, Cucchiara R (2011) Vision based smoke detection system using image energy and color information. Mach Vis Appl 22:705–719
13. Yao T-W, Wang H-Q, Hu Y (2012) Fire smoke detection based on wavelet transform and sparse optical flow method. Comput Eng 38(6):204–206 (in Chinese)

Chapter 17
Research of the Subgroup Discovery Algorithm NMEEF-SD

Haichun Xie, Yong Zhang, Limin Jia and Yong Qin

Abstract Subgroup discovery (SD) is a data mining technique which could find the most interesting individual patterns from a population of individuals for the user. Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discover (NMEEF-SD) which is based on non-dominated sorting genetic algorithm II (NSGA-II) is a kind of algorithm for SD. First, the concept of subgroup discovery is introduced. Then NMEEF-SD algorithm and its main properties are researched. Finally, the algorithm is applied to analyze the concrete comprehensive strength dataset from UCI database, the result of experiment shows that the NMEEF-SD algorithm is able to extract fuzzy rules with interesting characteristics and is easy to understand.

Keywords Subgroup discovery · NMEEF-SD · Fuzzy rules · UCI

17.1 Introduction

As data volumes explode, it becomes important to find the useful knowledge from a large number of complex data. Knowledge Discovery in Database (KDD) is aimed at assisting humans in extracting useful information from the rapidly growing volumes of data [6]. Knowledge is usually expressed in the form of rules, descriptive and predictive are the two standards to measure the quality of the rules. Subgroup discovery (SD) can obtain descriptive and predictive rules that make it attracts a lot of attention from researchers.

H. Xie · Y. Zhang
School of Automation,
Nanjing University of Science and Technology, Nanjing 210094, China

L. Jia · Y. Qin (✉)
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University,
Beijing 100044, China
e-mail: qinyong2146@126.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_17

Subgroup discovery was initially proposed by Klogen [8] and Wrobel [10]. Lavrac and Kavsek proposed CN2-SD [9] which was developed by modifying parts of the CN2 [3] classification rule learner, CN2-SD obtains rules through its covering algorithm, search heuristic, probabilistic classification of instances, and evaluation measures. Herrera processed SDIGA [5] which is a genetic fuzzy system for data mining task of subgroup discovery, using fuzzy rules describing the inductive knowledge.

This paper will apply Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery (NMEEF-SD) to the strength datasets of concrete. The remainder of this paper is organized as follows. Section 17.2 shows the basic concept of subgroup discovery and NMEEF-SD. Section 17.3 discusses an experimentation of NMEEF-SD. Finally, some conclusions are offered in Sect. 17.4.

17.2 Analysis of NMEEF-SD

17.2.1 Subgroup Discovery

The task of subgroup discovery is to discovery groups that are statically most unusual from a given dataset and target property. Simple rules with highly significant and high support are used to describe those groups. Rules have the form: Cond > Class.

Class is the target property, which appears in the rule consequent. Cond (the rule antecedent) is a conjunction of features (attribute-values).

"The probability of coronary heart disease is higher in smokers who have a family disease" is a rule, and the rule can be defined as:

$$\begin{aligned} &\text{if } (\text{smoker} = \text{true and family history} = \text{positive}) \\ &\quad \text{then coronary} - \text{heart} - \text{disease} = \text{true} \end{aligned}$$

The subgroup of smokers who have a family disease is described in this rule; coronary heart disease is the target property. The population described by this rule has a higher probability for the target property.

Target attribute, description language of subgroup, quality measures, search strategy are the four main aspects of subgroup discovery algorithm. Target attributes may be binary, nominal, or continuous [7]; language is the representation of the subgroups which must be suitable for obtaining interesting rules; quality measures are used to measure the obtained rules. Coverage, significance, unusualness, and support are often chosen as the quality measurements to extract and evaluate the rules; search strategy is important for the dimension of the search space. Different strategies have been used in subgroup discovery, and top-down search strategy is the usual choice.

17.2.2 NMEEF-SD: Non-dominated Multi-objective Evolutionary Algorithm for Extracting Fuzzy Rules in Subgroup Discovery

Subgroup discovery algorithm selects a number of quality measures to measure the quality of rules obtained, so subgroup discovery can be considered as a multiobjective problem. Different quality measures in the evolutionary process of rules population can be regarded as different evolutionary goals of genetic algorithm, therefore multiobjective evolutionary algorithm (MOEAs) is suitable to solve multiobjective optimization problems in subgroup discovery.

Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery (NMEEF-SD) [2] is based on hybridization between fuzzy logic and genetic algorithms. NMEEF-SD uses NSGA-II [4] to exact interesting, novel, and interpretable fuzzy rules. In NMEEF-SD, each candidate solution is coded according to the "Chromosome = Rule" approach, where the antecedent is represented in the chromosome, and the consequent is prefixed to one of the possible values of the target variable in the evolution.

17.2.2.1 Objective Function

The objective is to obtain rules with high confidence, understandable, and generality in the process of rule discovery. To do so, support, confidence, and accuracy are selected as quality measures.

- *Support*: the frequency of correctly classified examples covered by rule.

$$\mathrm{Sup}(R) = \frac{n(\mathrm{Class} \cdot \mathrm{Cond})}{n(\mathrm{Class})}$$

  where *n(*Class•Cond*)* is the number of examples which satisfy the conditions for antecedent and *n(*class*)* is the number of examples for target variable indicated.
- *Confidence*: standard measure that determines the relative frequency of examples satisfying the complete rule among those satisfying only the antecedent.

$$\mathrm{Conf}(R) = \frac{\sum_{E_k \in E / E_k \in \mathrm{Class}} \mathrm{APC}(E_K, R)}{\sum_{E_k \in E} \mathrm{APC}(E_K, R)}$$

  where APC is the degree of compatibility between an example and the antecedent part of a fuzzy rule.

- *Accuracy*: the membership of the examples covered by the antecedent part of rule and satisfying consequent of the *rule*.

$$\mathrm{Accu}(R) = \frac{n(\mathrm{Cond} \cdot \mathrm{Cond}_i)}{n(\mathrm{Cond}) + k}$$

where K is the number of the objective variables.

17.2.2.2 Main Properties of the Algorithm

NMEEF-SD consists of initialization, genetic operators, fast non-dominated sort, re-initialization based on coverage and stop condition [2], a single operation scheme of the algorithm can be seen in Fig. 17.1.

The re-initialization based on coverage together with the crowding distance in the selection operator to enhance the diversity. On the other hand, the algorithm includes operators of biased initialization and biased mutation to promote generalization. In addition, only the final solutions which reach a predetermined confidence threshold are returned.

Fuzzy logic is used to process the continuous variables, by means of linguistic variable. This allows the use of numerical features without the need of a previous discretization [1].

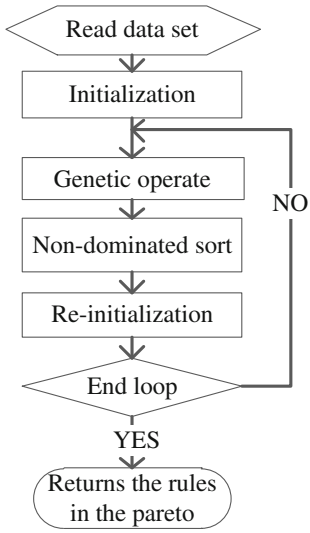


Fig. 17.1 The NMEEF-SD algorithm

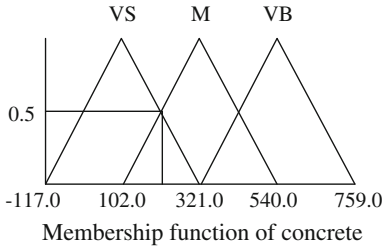


Fig. 17.2 Fuzzy partition for a numerical variable

17.3 Experimentation

Compressive strength of concrete dataset in UCI repository is selected as the experimental data. Experimental data consists of 1,030 samples, each sample includes 8 input variables and 1 output variable (compressive strength of concrete). Output variable is selected as the target variable in the dataset.

Compressive strength of concrete in the raw data is continuous variables, in order to apply to use subgroup discovery algorithm, compressive strength of concrete between 0 and 20 MPa is classified as low, 20–55 MPa as medium, and 55–80 MPa as high. Rearrange samples depend on target variables, where the compressive strength of 1–197 samples is low, the compressive strength of 198–883 samples is medium, and the compressive strength of 884–1030 samples is high.

All input variables are continuous variable, in order to apply the fuzzy rule, input variables are processed using fuzzy treatment. Triangular membership function is used to obtain rules with higher explanatory as seen in Fig. 17.2.

First, process the compressive strength of concrete dataset ten times by cross-validation, then execute algorithm. The algorithm is executed three times. The optimal rule set contains nine rules, which describe group of low, medium, and high, as can be seen in Table 17.1.

Results in Table 17.1 show that rules with simple structure contain less variables but have higher support, confidence, and accuracy.

Table 17.1 Best rules obtained by NMEEF-SD

Rule	Target property	Variable	Support	Confidence	Accuracy
R_1	High	4	0.578947	0.618307	0.500000
R_2	High	5	0.646617	0.811371	0.333333
R_3	High	4	0.691729	0.682891	0.333333
R_4	Medium	2	1.000000	0.747430	0.817204
R_5	Medium	4	0.941653	0.805049	0.940000
R_6	Medium	4	0.948136	0.801996	0.940000
R_7	Medium	6	0.478120	0.847051	0.875000
R_8	Medium	5	0.687196	0.840319	0.727273
R_9	Low	6	0.280899	0.614202	0.333333

Table 17.2 Rules of high compressive strength of concrete

Rule	Rule description
R_1	if (cement = VB and fly ash = VS and water = VS and coarse aggregate = M) then concrete compressive strength = high
R_2	if (cement = VB and fly ash = VS and superplasticizer = M and coarse aggregate = M and age = (N ORVL)) then concrete compressive strength = high
R_3	if(cement = (M OR VB) and fly ash = VS and water = VS and age = N) then concrete compressive strength = high

Select rules of high compressive strength of concrete from Table 17.1, description of the rules are shown in Table 17.2.

Analysis rules obtained indicate the following: The concrete that contains lots of cement, a small amount of fly ash, right quantity of coarse-aggregate and normal coagulation time performance for the high compressive strength of concrete.

Figure 17.3 shows target variables and Pareto sequence when algorithm was running, the red curve represents the average support, the blue curve represents the average confidence, and the green curve represents the average accuracy.

Figure 17.3 shows that during the process of extracting rules, the overall quality of the population tends to be stable with the evolution of the rules of population. There is still some fluctuation in shall scope mainly because NMEEF-SD algorithm uses random parent population to crossover operation. The diversity in the genetic population is increased with re-initialization based on coverage. The change curve of number of rules in Pareto shows there is a large fluctuation in the early evolution, with the increasing of number of rules, Pareto will maintain a stable number.

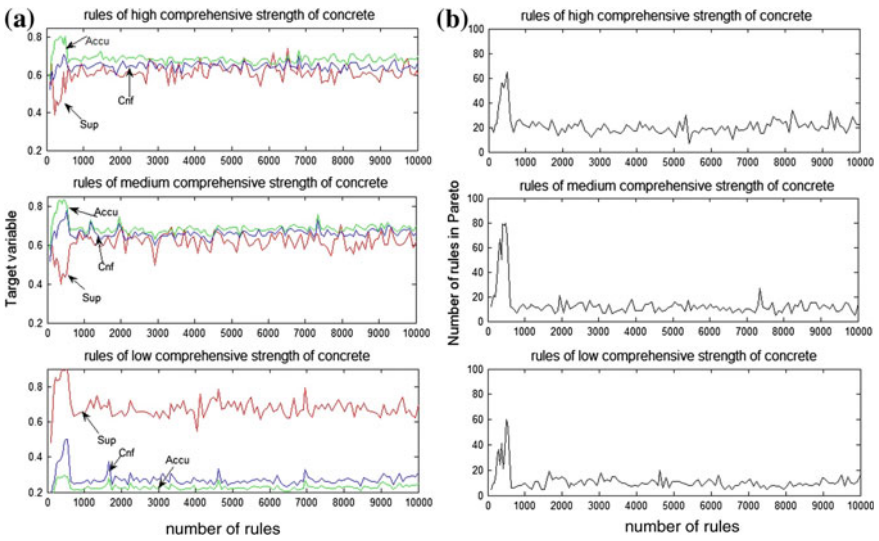


Fig. 17.3 Target variable during process of extracting and change curve of Pareto

The results show that the application of subgroup discovery algorithm in processing compressive strength of concrete dataset, on the one hand can extract simple, effective rules to describe different compressive strength of concrete. These rules can provide an effective reference for quality optimization of concrete in the manufacturing process; on the other hand the rules obtained with high classification accuracy can be used to predict the compressive strength of concrete and provide an effective tool for detecting compressive strength of concrete.

17.4 Conclusion

In this paper, Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery (NMEEF-SD) is researched. We realize the NMEEF-SD algorithm and apply to Compressive strength of concrete dataset in UCI repository. The experimental results indicate that subgroup discovery could efficiently extract rules that are interesting, understandable, and these rules can provide an effective basis for compressive strength of concrete detection.

Acknowledgments This research is supported by Independent Subject of Y. Qin (No. RCS2014ZT24) and Research Fund for the Doctoral Program (No. 20120009110035). The supports are gratefully acknowledged.

References

1. Carmona CJ, Chrysostomou C, Seker H, del Jesus MJ (2013) Fuzzy rules for describing subgroups from Influenza A virus using a multi-objective evolutionary algorithm. Appl Soft Comput 13(8):3439–3448
2. Carmona CJ, González P, del Jesus MJ, Herrera F (2009) Non-dominated multi-objective evolutionary algorithm based on fuzzy rules extraction for subgroup discovery. In: Proceeding of 4th international conference on hybrid artificial intelligence Systems, vol 5572. Springer, LNAI, pp 573–580
3. Clark P, Niblett T (1989) The CN2 induction algorithm. Mach Learn 3(4):261–283
4. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
5. Del Jesus MJ, González P, Herrera F, Mesonero M (2007) Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. IEEE Trans Fuzzy Syst 15 (4):578–592
6. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37
7. Herrera F, Carmona CJ, Gonzalez P, del Jesus MJ (2001) An overview on subgroup discovery: foundations and applications. Knowl Inf Syst 29:495–525
8. Klosgen W (1996) Explora: a multipattern and multistrategy discovery assistant. In: Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, pp 249–271

9. Lavrac N, Kavsek B, Flach PA, Todorovski L (2004) Subgroup Discovery with CN2-SD. J Mach Learn Res 5:153–188
10. Wrobel S (1997) An algorithm for multi-relational discovery of subgroup. In: Proceeding of the 1st European symposium on principles of data mining and knowledge discovery, vol 1263. Springer, LNAI, pp 78–87

Chapter 18
A Subgroup Discovery Algorithm Based on Genetic Fuzzy Systems

Shuo Dai, Yong Zhang, Limin Jia and Yong Qin

Abstract Subgroup discovery algorithm is a new data mining technique, which plays an important role in the induction of large data areas. First, the basic concepts of subgroup discovery algorithm and fuzzy system are introduced. Then subgroup discovery iterative genetic algorithm (SDIGA) is studied. Genetic fuzzy system is used in traditional subgroup discovery algorithm, the way that a weighted sum of multiple objective functions is taken in fitness function. After continuous crossover genetic, the best description of the rules is obtained. Finally, the proposed method is applied to the dataset of compressive strength of concrete in UCI database, and the experiment results show the effectiveness of SDIGA subgroup discovery algorithm.

Keywords Subgroup discovery algorithm · Genetic fuzzy systems · Rules

18.1 Introduction

In recent years, the technology of database has made great development. Large amounts of data need people to deal with. Subgroup discovery is a novel data mining technique aimed at extracting effective and useful information from target data, which provides people an efficient means for accessing knowledge.

At present, subgroup discovery algorithm is divided into three categories: classification algorithm for subgroup discovery, association algorithm for subgroup discovery, and evolutionary algorithms for subgroup discovery. EXPLORA [1] is the earliest subgroup discovery algorithm, which extracts rules with the decision

S. Dai · Y. Zhang
School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China

L. Jia · Y. Qin (✉)
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
e-mail: qinyong2146@126.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_18

tree, evaluate rules whether users need based on standard of universality, redundancy, and simplicity. APRIORI-SD [2], an improved classification algorithm by classification algorithm, processes rules and classifies probability of sample with weighted relative accuracy as a new measure of the quality. Subgroup-Miner [3] is an extended algorithm of EXPLORA and MIDOS, which inducts rules effectively with decision rules and interactive search.

This paper presents the SDIGA subgroup discovery algorithm based on genetic fuzzy systems. SDIGA describes interesting individuals in a more easily understood form. Simulations are carried out to verify the effectiveness of the proposed algorithm.

18.2 Subgroup Discovery

Subgroup discovery is a form of supervised learning, which is aimed at uncovering properties of a selected target population of individuals with the given property of interest. The subgroup is the rule used to describe the property, such as the following rule [4]:

Diabetics have a higher probability of cardio-cerebral-vascular disease; the corresponding regular expression is

$$\begin{aligned} &if\,(diabetic = true)\\ &then\,cardio - cerebral - vascular\,disease = true \end{aligned}$$

For the above rule, diabetics are individuals, while cardio-cerebral-vascular disease is the target property of interest. The individuals described by the rules have higher probability for the properties of interest.

In classification rule learning, an inducted subgroup description is shown in the following form:

$$Cond > Class \tag{18.1}$$

where Class is the target property of interest, which appears in the rule consequent. Cond (the rule antecedent) is a conjunction of features (attribute-values) [3].

18.3 SDIGA Subgroup Discovery Algorithm

Genetic fuzzy system is formed by fuzzy system and genetic algorithm. In fuzzy system, the capacity of learning and adaptation of genetic algorithm can contribute to application of rule induction.

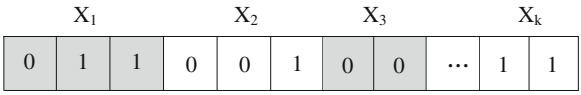


Fig. 18.1 Encoding model of a rule

The purpose of SDIGA allows us to obtain a set of understandable fuzzy rules with a flexible structure that describes different subgroups in data. In order to ensure this, the SDIGA must be run once by each value of the target property, i.e., each process of SDIGA can obtain a series of rules to describe aspecific value of the target property [5].

SDIGA uses fuzzy rules in DNF format. All the information relating to a rule is contained in a fixed-length chromosome with a binary representation in which, for each feature, a bit for each one of the possible values of the feature is stored. If the corresponding bit contains the value 0, it means that the value is not included in the rule; and if the bit contains the value 1, it indicates that the value is included. If a rule contains all the bits with the value 1, or all of them contain the value 0, this indicates that this feature has no relevance in the rule, and so the feature is ignored. In these cases, the feature is not included in the rule. This takes us to a binary representation model with as many genes by variable as possible values for the same one, as can be seen in Fig. 18.1.

The hybrid GA of SDIGA extracts a single DNF fuzzy rule in an attempt to optimize the confidence and support. We describe the elements of the hybrid GA: the chromosome representation, the fitness function, the reproduction model, and the post-processing phase of the hybrid GA [6].

In this process of rule discovery, the objective is to obtain understandable and general rules with high confidence. This means that the problem has several objectives: the support and the confidence of the rule. To achieve this, the weighted sum method that weights sets of objectives into a single objective is the simplest approach and lets us introduce the expert criteria related to the importance of the objectives for a specific problem in the rule generation process. The weight of one objective is chosen in proportion to the objective's relative importance in the problem. So, this proposal uses a weighted lineal combination in the following way:

$$\text{Fitness}(c) = \frac{w_1 \times \text{Sup}_3(c) + w_2 \times \text{Conf}(c) + w_3 \times \text{Accu}(c)}{w_1 + w_2 + w_3} \tag{18.2}$$

$$\text{Sup}_3(R) = \frac{\text{Ne}^+(R)}{\text{Ne}_{\text{NC}}} \tag{18.3}$$

$$\text{Conf}(R) = \frac{\sum_{E_k \in E / E_k \in \text{Class}} \text{APC}(E_k, R)}{\sum_{E_k \in E} \text{APC}(E_k, R)} \tag{18.4}$$

$$\mathrm{Accu}(R) = \frac{n(\mathrm{Class} \cdot \mathrm{Cond})}{n(\mathrm{Cond}) + k} \tag{18.5}$$

where Sup_3 represents confidence; Conf represents support; Accu represents accuracy; APC is antecedent part compatibility; $\mathrm{Ne}_{\mathrm{NC}}$ is the number of samples left uncovered by the previous rules; $\mathrm{Ne}^{+}(R)$ is the number of samples covered by the rule that is left uncovered by the previous rules; $n(\mathrm{Class} \cdot \mathrm{Cond})$ is the number of samples which meets antecedent and consequent; $n(\mathrm{Cond})$ is the number of sample which meets antecedent of rule R; k is the number of sample classification.

The scheme of the extraction model is shown as follows:

```
START
    Choose a target feature At_TAR
    Rule Set ← ∅
    REPEAT
      Execute the GA ( At_TAR ) obtaining rule R
      Local Search(R)
      If (confidence(R) >= minimum confidence and R represents new
examples)
          Rule Set ← Rule Set ∪ R
          Mark the set of examples covered by R
    WHILE (confidence(R) >= minimum confidence and R represents new
examples)
    END
```

18.4 Simulation Analysis

In this paper, a simulation of concrete compressive strength data Set of UCI database is conducted. It contains 1,030 samples. In each sample there are 9 variables, of which 8 are inputs and 1 output. We select output variable in the data set (concrete compressive strength) as the target variable. Distribution of samples in accordance with the target property in the data set is shown in Table 18.1.

In order to use subgroup discovery algorithm, low represents the corresponding compressive strength of concrete at 0–20 MPa; medium represents the corresponding compressive strength of concrete at 20–55 MPa; and high represents the corresponding compressive strength of concrete at 55–80 MPa. Rearranging the samples depends on target variables, where the compressive strength of 1–197

Table 18.1 The distribution of concrete sample

Target property	Sample	Proportion (%)
High	147	14.3
Medium	686	66.6
Low	197	19.1

samples is low, the compressive strength of 198–883 samples is medium, the compressive strength of 884–1,030 samples is high.

Then we obscure input variables and use standard triangle membership function as the divided form of membership function of the input variables, for getting higher interpretative rules.

The parameters used in the experiments are:

(1) Rule form: DNF;
(2) Population size: 100;
(3) Maximum number of evaluations of individuals in each SDIGA run: 10000;
(4) Crossover probability: 0.6;
(5) Mutation probability: 0.01;
(6) Number of linguistic labels for the numerical variables: 3;
(7) Quality measure weights for the fitness function: $w_1 = 0.3$, $w_2 = 0.4$, $w_3 = 0.3$;
(8) Whether to post-process: yes;

We load the compressive strength of concrete data set, for which cross-validation is ten times, producing ten datasets. To choose a set of rules that have the best quality, the ten data sets need to be inducted. The result is shown in Table 18.2.

The obtained rules above comprise four rules, which describe group of low, medium, and high. Table 18.2 shows that rules with simple structure contain less variables and higher support, confidence, and accuracy.

Each execution will return the highest rule. During the process of obtaining rules in Table 18.1, SDIGA totally performs hybrid genetic algorithm four times. In each processing, the changes in average fitness of the rules population is shown in Fig. 18.2.

Figure 18.2 shows that the obtained rules are increasing and the fitness of rule population is rising with the evolution of rule population. Because of the stable evolutionary strategy, the rule population can converge quickly. When the number of rules is up to 500, the population keeps stable.

SDIGA respectively have rule induction for the then dataset, after which we get 40 rules. Then, we sort these rules in accordance with the support, confidence, accuracy, and variables contained. The result is shown in Fig. 18.3.

This result shows that the support of rules obtained is higher than 0.5, which means that most of the target samples can be covered by the obtained rules. About 67 % rules have higher confidence and accuracy, in which there are less negative

Table 18.2 The quality of the best rules

Rule	Target property	Variable	Support	Confidence	Accuracy	Fitness
R1	High	4	0.765152	0.541844	0.666667	0.646283
R2	Medium	4	0.990291	0.764425	0.857143	0.860000
R3	Medium	2	1.000000	0.675653	0.674545	0.772625
R4	Low	5	0.882023	0.325789	0.442478	0.527666

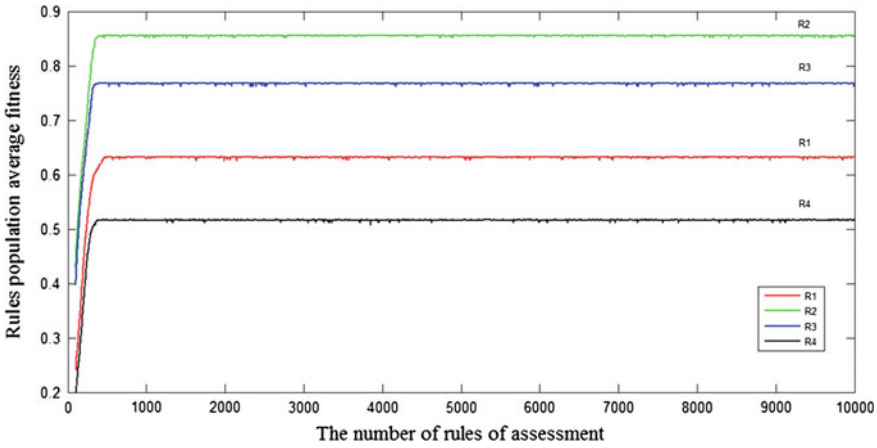


Fig. 18.2 The average fitness curve of the rule population of rule induction process

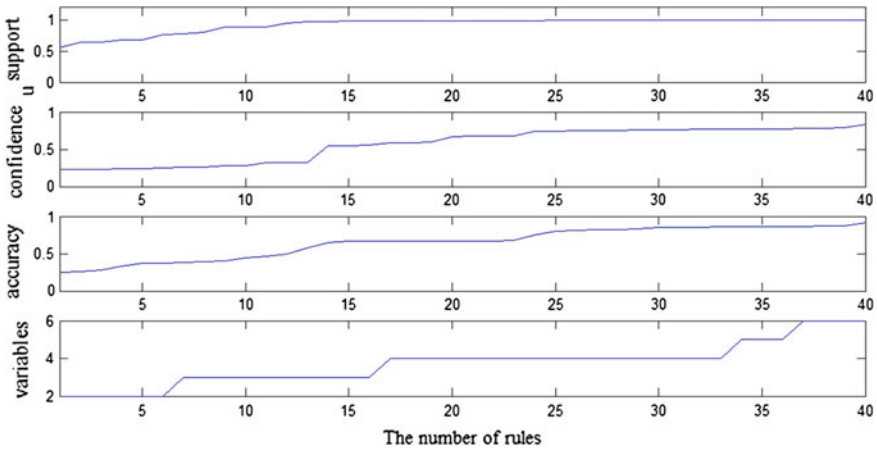


Fig. 18.3 Ordination diagram of rules according to the target function

samples. Most rules contain variables less than 5, which means that the rules have higher explanatory and simple structure. Therefore, SDIGA can obtain rules having better explanatory and accuracy to describe most samples of the target population.

18.5 Conclusion

This paper presents a study for SDIGA, which is targeted at characterizing population subgroups of a given target class. The fitness function is defined as the weighted summation of several target functions. Through the operation as crossover

and variation, rule population is constantly evolving and the best rules can be obtained. The data of UCI simulation show the effectiveness of SDIGA in rule induction.

Acknowledgments This research is supported by independent subject of Y. Qin(No. RCS2014ZT24) and research fund for the doctoral program (No. 20120009110035). The supports are gratefully acknowledged.

References

1. Novak PK, Lavrač N, Webb GI (2009) Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining. J Mach Learn Res 10:377–403
2. Kavsek B, Lavrac N (2006) APRIORI-SD: adapting association rule learning to subgroup discovery. Appl Artif Intell 20:543–583
3. Lavrač N, Cestnik B, Gamberger D, Flach P (2004) Decision support through subgroup discovery: three case studies and the lessons learned. Mach Learn 57(1–2):115–143
4. Kanagaraj N, Sivashanmugam P, Paramasivam S (2009) A fuzzy logic based supervisory hierarchical control scheme for real time pressure control. Int J Autom Comput 01:88–96
5. Kralj-Novak P, Lavrac N (2009) Webb GI supervised descriptive rule discovery: a unifying survey of constrast set, emerging pattern and subgroup mining. J Mach Learn Res 10:377–403
6. Del Jesus MJ, González P, Herrera F, Mesonero M (2007) Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. IEEE Trans Fuzzy Syst 15(4):578–592

Chapter 19
Secure Measuring and Controlling Methods Embedded SM4 Algorithm for Smart Home

Xiangdong Hu, Xiaopeng Qin and Haiming Mou

Abstract Smart home based on the Internet of things is gradually changing our daily life, while it faces such serious problems as secure measuring and controlling to appliances and protection of end user's privacy. To guarantee the safety and reliability of smart home system, a secure measuring and controlling method embedded the domestic SM4 cryptographic algorithm for smart home is proposed in this paper, which integrates such secure mechanisms as access control based on the physical addresses of smartphone terminal or sensor nodes used in measuring and controlling operations, authentication based on the keys used in the encrypted transmission of instructions of measuring, and controlling and abnormal detection based on analysis of data. On such basis, the comparison in performances is carried out by secure measuring and controlling methods embedded domestic SM4 algorithm or typical AES one. The results of test suggest that the proposed secure measuring and controlling methods for smart home is feasible and effective, and the delay time consumed in information processing of the proposed method is only 2.5 % more than the currently pervading ones without secure mechanism. The rate of delay based on the SM4 algorithm is about 4 % less on average than the original AES one embedded in nodes, and SM4 algorithm owns independent intellectual property right with more flexible in realization of system.

Keywords Internet of things · Smart home · Safety · SM4 algorithm · Measuring and controlling

X. Hu (✉) · X. Qin · H. Mou
College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China
e-mail: huxd@cqupt.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_19

19.1 Introduction

With the rapid development of the Internet of things, more and more smart home are connecting various kinds of household electrical appliances or sensing devices together through local family networks, and easily and efficiently controlling them by technologies such as computer, communication, measurement, and control for a more comfortable living environment [1]. Smart home brings people much convenience while it faces many potential risks of malicious attacks, such as injecting unauthorized control instructions into system or eavesdropping information of monitoring and so on, which will further threaten safe and reliable operations of smart home or users' privacy [2].

The current smart home industry in China is still in its infancy. There are no unified national standards or technical specifications, and its standards in safety are blank. Security problems of smart home based on the Internet of things technology need to be urgently solved to improve people's experience of living [3].

A building method of security-focused smart home embedded domestic SM4 cipher algorithm is proposed in this paper, which is helpful to deal with the potential risks in information security for smart home.

19.2 The System Model of Secure Smart Home

19.2.1 The Composition of Secure Measuring and Controlling System

The composition of secure measuring and controlling system in smart home is illustrated in Fig. 19.1.

The secure measuring and controlling system mainly includes a WiFi gateway, primary or secondary routing nodes, measuring and controlling nodes, and household appliances and devices. Measuring and controlling nodes are the fundamental part of smart home, their main tasks are to sense the environment of home, or to receive commands so as to control corresponding appliances or devices such as air conditioner, digital TV set, refrigerator, curtain, lighter, monitor, alarm, etc. The secure algorithm is stored in nodes, and all communications within the smart home systems are protected by SM4 block encryption algorithm. Moreover, when intrusion detection finds an illegal node or abnormal case, the alarm node will be activated and the host or hostess will receive a notice of alarm about the incident [4].

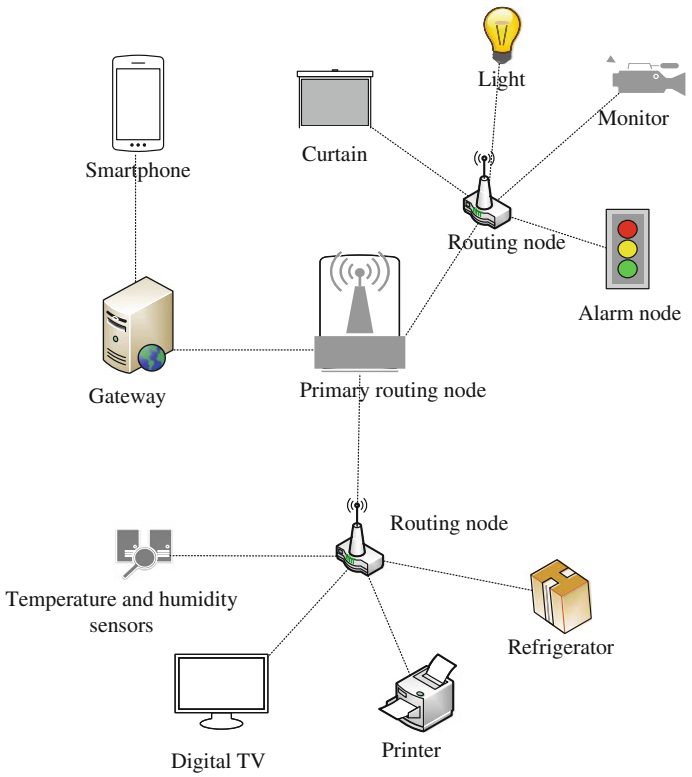


Fig. 19.1 The composition of secure measuring and controlling system in smart home

19.2.2 Security Mechanism

In order to ensure safe and reliable operation of smart home system and to protect users' privacy, here a secure measuring and controlling method embedded SM4 algorithm for smart home is proposed. The involved security mechanism is as follows: First of all, an access authentication is necessary to login system by the authorized user identity and password; Secondly, any node will be examined and verified based on the physical address and the preset key before they can access the network, and the unverified nodes will be excluded; Thirdly, each measuring and controlling instruction or message transmitted between nodes will be encrypted by SM4 cipher algorithm to improve the confidentiality of smart home system [5]; Finally, further security can be reached along with intrusion detection and alarm mechanism.

19.3 Implementation of Secure Measuring and Controlling System and Performance Evaluation

19.3.1 Hardware Implementation of the System

In order to verify the feasibility of smart home characterized in secure measuring and controlling, CC2530 chip is chosen as the core unit and a star topology structure is adopted to set up a simulation platform of secure smart home shown in Fig. 19.2 [6, 7].

As the human–machine interface of secure smart home system, the smartphone terminal based on Android4.0 is mainly responsible for sending measuring and controlling commands and receiving information from nodes including controlling and sensing ones. The WiFi gateway is responsible for two-way Zigbee signal conversion between the smartphone terminal and routing nodes. The routing node is responsible for establishing Zigbee network. The terminal node is used for connecting and controlling various kinds of household equipments. In addition, a simulated malicious terminal mainly acts as illegal invasion node to simulate intrusion behavior [8, 9].

19.3.2 Secure Measuring and Controlling Process

The secure measuring and controlling method is designed into APK (Android Package) software based on Android 4.0 smartphone terminal by dedicated Eclipse V22.3 development tools. The domestic SM4 cipher algorithm is embedded into the control software to compare with the original AES one. The secure measuring and controlling process of smart home is divided into two paths based on different

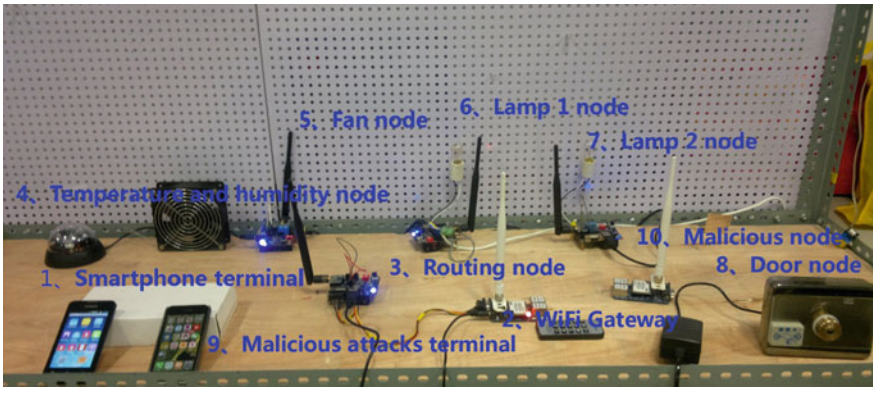


Fig. 19.2 Secure smart home system

functions of nodes: data acquisition and object controlling, as illustrated in Fig. 19.3.

To create a new SM4 function in Java file under Android platform, the statement of SM4 algorithm is: SMS4 sm4 = new SMS4 ().

An example of *key* used in SM4 is:

byte[] *key* = {*0x*01, *0x*23, *0x*45, *0x*67, *0x*89, *0xab*, *0xcd*,*0xef*, 0xfe, *0xdc*, *0xba*, *0x*98, *0x*76, *0x*54,*0x*32, *0x*10}.

An example of the plain text of instruction used in temperature measurement is as follows:

byte [] *temp* = {*0xF*, *0xC2*, *0x01, 0x01, 0xC4*, *0xFE*}.

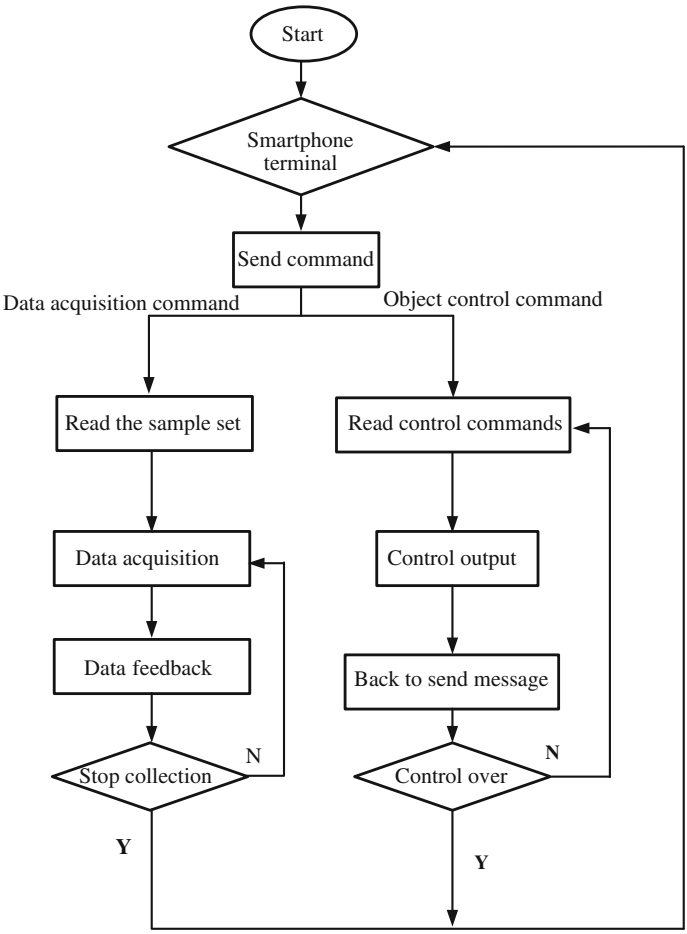


Fig. 19.3 Secure measuring and controlling process for smart home

The statements used in calling the SM4 cryptographic algorithm and sending commands are as follows:

*sm*4.*sms*4 (*temp*, *inLen*, *key*, *outTemp*, *ENCRYPT*);
SendCmd(*outTemp*);

19.3.3 The Performance Test of System

The performance of secure smart home system has been tested by adopting SM4 algorithm, and compared with the embedded AES one.

Secure measuring and controlling instructions in the smart home are similar to such format of data as "EF C1 02 01 C3 FE". The instruction means to open lamp 2. EF and FE are used as check digit. C1 is the serial number of equipment. 02 is No.2 lamp. 01 means that the lamp will be lighted on. C3 is checksum. The whole instruction is encrypted into 128-bit hexadecimal data by SM4 cipher algorithm. SM4 encryption function used in the smart home system is as follows:

*SM*4. *SMS*4 (*in, inLen, key, out*, *ENCRYPT*)

Correspondingly, SM4 decryption function is:

*SM*4. *SMS*4 *(out, inLen, key*, *in*, *DECRYPT*)

As a contrast, a smart home system embedded AES algorithm is realized in the same way. AES encryption functions: *out* = *Encrypt_Byte* (*in, key*), and AES decryption function is: *in* = *Decrypt_Byte*(*out, key*).

The above two algorithms are called to evaluate the performance of the proposed based on monitoring nodes and controlling ones. The results of test show each of them can protect the system and find abnormal behaviors from nodes, but the time to finish an instruction is different. A detailed evaluation is carried out as follows.

The test is divided into three cases: unencrypted, encrypted based on SM4, and encrypted based on AES. Smartphone terminal sends unencrypted sensing or controlling instruction to nodes every 4 s, and receives the feedback message from nodes which has finished the instruction, all records are stored in LogCat logs. Assuming the timestamp of sending unencrypted instruction as S_i, the timestamp of feedback is R_i, therefore, the delay between sending and receiving operations for an unencrypted instruction can be expressed as T_i:

$$T_i = R_i - S_i \tag{19.1}$$

Similarly, the encrypted test by SM4 or AES algorithm can be done. Let the timestamp of sending encrypted instruction be S_i, the timestamp of feedback is r_i, therefore, the delay between sending and receiving operations for an encrypted instruction can be expressed as t_i:

$$t_i = r_i - s_i \tag{19.2}$$

So one can get the ratio of time delay of encrypted instruction compared to unencrypted one:

$$d_i = \frac{t_i - T_i}{T_i} \tag{19.3}$$

The average ratio of time delay for m–times test can be obtained according to the following formula [10, 11]:

$$\overline{d} = \sum_{i=1}^{m} \frac{d_i}{m} \tag{19.4}$$

The impact of secure mechanism on time delay is plotted in Fig. 19.4 by sampling 50 times. Figure 19.4a, b show the impact of secure mechanism on controlling instruction or sensing one, respectively. They are classified into three kinds of situations: unencrypted case without security mechanism, encrypted case embedded SM4 algorithm, and encrypted case with AES algorithm.

According to Fig. 19.4, the average values of time delay of secure mechanisms is illustrated in Table 19.1 based on 50 sampling values. Table 19.1 shows that the average ratios of time delay are 2.61 and 6.25 % for sensing instructions with a secure mechanism based on SM4 and AES algorithms, respectively. The corresponding results are 2.28 and 7.91 % for controlling instructions. Although the secure mechanisms have little impact on time delay of system, while a secure guarantee is in prospect. Furthermore, the secure mechanism embedded SM4 algorithms have less time delay than one embedded AES algorithm, that is, one can get that the advantages of SM4 algorithm in time delay are 3.64 and 5.63 % for data acquisition and node controlling than the AES in Table 19.1.

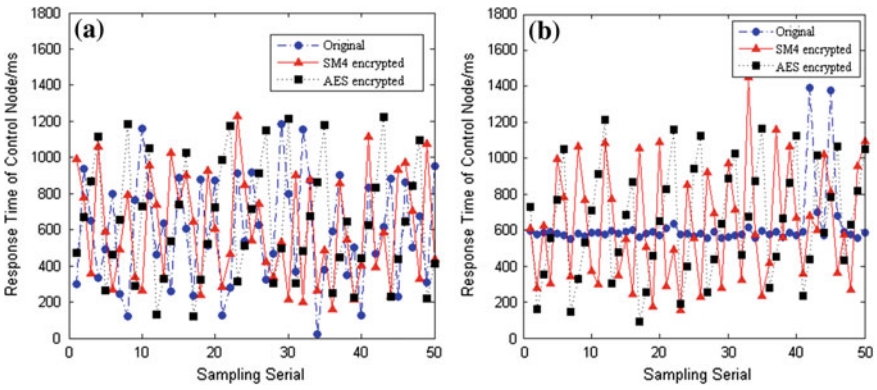


Fig. 19.4 The impact of secure mechanisms on time delay. **a** Control nodes, **b** sensing nodes

Table 19.1 The average impact of secure mechanisms on time delay

	Unencrypted	SM4 encryption		AES encryption	
	t/ms	t/ms	*d*/%	t/ms	*d*/%
Data acquisition	616.7	632.8	2.61	655.3	6.25
Node controlling	594	607.6	2.28	641	7.91

Although AES cryptographic algorithm is built in such Zigbee nodes as CC2530, which is convenient for secure communications between those nodes with the same configuration, it is hard for non-Zigbee nodes such as smartphone terminal, WiFi gateway, and so on. However, the proposed scheme embedded SM4 algorithm could solve this problem by distributing SM4 algorithm in all nodes of smart home, this is helpful to earn its flexibility in realizing the secure measuring and controlling system.

19.4 Conclusions

Smart home brings much convenience to people while it faces some potential risk from malicious attacks. Research on secure smart home is necessary for application and popularization. Secure measuring and controlling methods are proposed for smart home based on nodes embedded SM4 cryptographic algorithm. The secure mechanisms such as access control, encryption of instructions, authentication of devices, intrusion detection, and alarm are helpful to improve information security of smart home and to protect users' privacy. The results of simulation proved that the proposed method is effective and feasible to enhance the security of smart home system. The extra mechanism of security only brings about 2.5 % time delay in processing of instruction. The results show that the proposed methods embedded SM4 algorithm have about 4 % less time-consuming advantage in delay and flexibility of implementation compared to the typical AES scheme, and it does not involve the use of foreign intellectual property rights. This research provides a theoretical path and technical exploration for realization of secure smart home based on the Internet of things.

Acknowledgments This work is supported by the National Natural Science Foundation of China (61170219) and the Basic and Frontier Research Project of Chongqing Municipality (cstc 2013jcyjA40002).

References

1. Liu ZB (2004) X-10 protocol and its applications in smart houses. Microelectron Comput 21 (3):5–8 (in Chinese)
2. Hu XD, Wei QF, Tang H (2010) Model and simulation of creditability based data aggregation for the internet of thing. Chin J Sci Instrum 31(11):2636–2640 (in Chinese)

3. Wu CK (2010) A preliminary investigation on the security architecture of the internet of things. Bull Chin Acad Sci 25(04):411–419 (in Chinese)
4. Zhang XM, Wang GQ, Ding XN (2009) Development of an Internet home automation system. Chin J Sci Instrum 30(11):2423–2427 (in Chinese)
5. Wu J (2013) Research and implementation of hybrid cipher algorithm based on SM4 and Sm2. Softw Guide 12(8):127–130 (in Chinese)
6. Hou J, Wu CD (2009) Research of intelligent home security surveillance system based on Zigbee. Mech Electri Eng Mag 26(01):33–35 (in Chinese)
7. Liu YH, Zhang JX (2012) Smart home based on the Zigbee wireless. Intell Netw Intell Syst 05:122–125
8. Daehwan K, Daijin K (2006) An intelligent smart home control using body gestures. Hybrid Inf Technol 06:439–446
9. Ren XL, Yu HB (2007) Study on security of Zigbee wireless sensor network. Chin J Sci Instrum 28(12):2132–2137 (in Chinese)
10. Hu XD, Han KM, Xu HR (2014) Design and implementation of security focused intelligent household IOT. J Chongqing Univ Posts Telecommun 26(2):171–176 (in Chinese)
11. Jiang J, Liu T, Hu X (2008) Research and implementation of dynamic SMS4 algorithm. Netw Secur Technol Appl 9:92–93 (in Chinese)

Chapter 20
Multi-core Processor Simulation Vector Learning Optimization Based on S^2LS-SVM

Guanjun Wang, Ying Zhao and MinMing Tong

Abstract With the revolutionary progress of the EDA industry, the verification of microprocessor becomes more and more difficult. It is a big problem to optimize the huge verification stimuli. Verification stimuli efficiency problem is researched in our paper and multi-core processor verification vector learning method based on S^2LS-SVM is put forward. First, verification stimuli are generated according to coverage information, the simulation vectors feature selection and extraction is conducted by transition probability matrix. Initial S^2LS-SVM classifier is trained on the labeled training set, area labeling principle is used for unlabeled samples tagging, dynamic adjustment of centralized "inconsistent" semi-labeled samples; then, train a classifier with the label sample and semi-labeled samples, classifier predict the new stimuli vector is a redundancy or not, if it is redundant, it will not need to do the simulation. Effective label sample provides SMT Solver feedback to the classifier for incremental updates. Experimental results show that this method of training is fast, the simulation vectors can be reduced significantly and rapid verification closure is achieved. It also has important reference value for the future multi-core processors simulation.

Keywords Multi-core processor verification · S^2LS-SVM · Stimuli generation · Functional coverage · Transfer probability matrix

G. Wang (✉) · Y. Zhao
Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, People's Republic of China
e-mail: zywgj@cumt.edu.cn

M. Tong
School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, People's Republic of China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_20

20.1 Introduction

Based on the 2012 International Technology Roadmap for semi-conductor (ITRS) report [1], 10 nm high-speed and low-power CMOS technology will be widely used in 2020. The on-chip multi-core processors (CMP) has become the mainstream of the structure of high-performance microprocessors. The number of processor cores also has exponential growth similar to Moore's law, from 2, 4, 8 kernels to 128 kernels or more [2], commercial processor core focuses on 8–16 kernels level nowadays.

Rapid advances in technology also led to a substantial increase complexity in processor chip verification. Verification has become one of the major challenges in microprocessor design [3]. IC design verification process has become the most expensive work [4].

In this paper, we focus on the current research hotspot in simulation verification; in particular, the direction of machine learning-based simulation verification. Machine learning simulation verification focuses mainly on simulation vector generation and vector optimization, vector optimization techniques which have received the wide attention of researchers [5–8]. Previous studies have not been given sufficient attention to two issues: first, coverage indicators is single, second, classifier constructed using supervised learning method [6] and unsupervised learning methods [9], considering only pure labeled samples and the unlabeled samples, there is a long learning time and lack of learning precision. Based on the above work, in order to improve the efficiency of verification, we propose a method to verify the microprocessor vector optimization based on least squares support vector machine semi-supervised learning algorithm (S^2LS-SVM).

20.2 Processor Simulation Verification Vector Learning Based on S^2LS-SVM

In the traditional simulation process, new verification vectors are generated without classification or need a long learning process before input under verification processor. Defects in these two methods are that in first method simulation quality is not high, another method needs a long time to learn and the efficiency is low. The S^2LS-SVM-based microprocessor vector optimization method can overcome the above shortcomings by feature extraction and selection of training samples (including label samples and unlabeled samples) for rapid realization of semi-supervised learning algorithm to build a classifier, then Classifier conduct effective increments updates after simulation. The complete process is shown below (Fig. 20.1):

The difficulty in our method is classifier construct and intelligent feedback mechanism. To solve this problem least squares support vector machine semi-supervised learning methods is introduced in the construction of the classifier, the

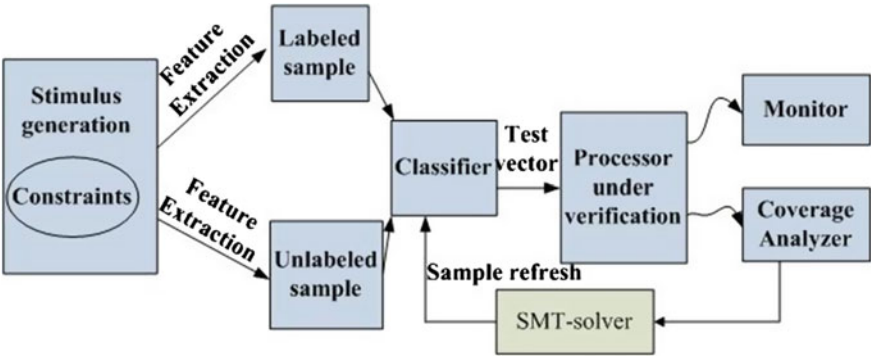


Fig. 20.1 Vector optimization framework based on S^2LS-SVM

use of learning classifier to validate vector filtering. A uniform coverage model coverage analysis and feedback mechanism is used in intelligent feedback stage.

20.2.1 Functional Coverage Model and Analysis

Coverage model defines a subset of the design of stimulus/response space, this subset to a certain degree (there is sufficient confidence) proved to be correct. Coverage metrics will be carried in order to assess test stimulus and ensure the expected functional to be fully verification. A coverage metrics can be a function of the design, structure, or source code. Coverage model is composed of the structure coverage model and functional coverage and assertion coverage and other targets defined by the components. Structural coverage is the code to achieve a design. Functional coverage model in SystemVerilog (SV) can be achieved by the two structures: coverage groups and coverage properties. Functional coverage models must obtain by hand independently. Also, functional coverage analysis in the process of simulation verification is still needed, so we can identify the area not yet covered by coverage model, and then find the demand in functional verification process. In the course of coverage analysis, we introduce intelligent feedback mechanism, feedback based on objective flaw method to accelerate verification closure.

In this work, we establish a unified multi-coverage model based on SV language. This model is easy to integrate into the current processor verification process. The specific framework considered in the literature [10] is shown in Fig. 20.2.

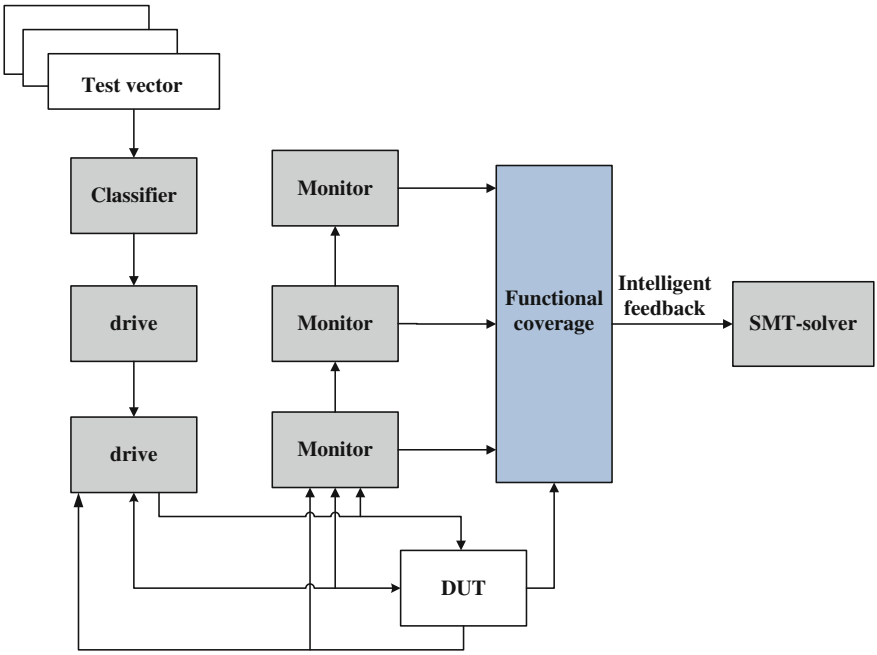


Fig. 20.2 Coverage analysis and intelligent feedback

20.2.2 Feature Extraction and Selection of Verification Vectors

Verification vectors of microprocessor are generally the instruction sequence verification vectors that can have multiple representations, such as assembly code, machine code, etc. It can be divided mainly into three types: (1) the structure verification vectors to verify some special functions in boundary conditions; (2) automatic generation vectors. The main part of verification vectors, including randomly generated and FSM-based generated vectors, aimed at finding errors may not be found in other vectors; (3) the existing benchmark test vectors mainly used to confirm the correctness of the design and test system performance. Conduct vector feature extraction and selection on the basis of the existing random test vectors. The involved definition in this method is given first:

Definitions 20.1 Transition probabilities $\xi_p(I_i, I_j)$

Defined instruction set I, suppose assembly instructions are n, build up the $n \times n$ matrix from the index in I. Transition probability is defined as $\xi_p(I_i, I_j)$, it represents at any given instruction set, the frequency of a combination of two instructions I_i, I_j. It is defined as:

$$\xi_p(I_i, I_j) = \frac{n(I_i, I_j)}{N(I)}$$

Transition probability matrix is composed of the transition probability elements, which is defined as follows.

Definitions 20.2 Transfer Probability Matrix, TPM:

$$\mathrm{TPM}_p(I, J) = \begin{vmatrix} x_p(I_1, I_1) & x_p(I_1, I_2) & L & x_p(I_1, I_n) \\ x_p(I_2, I_1) & x_p(I_2, I_2) & L & x_p(I_2, I_n) \\ M & & L & M \\ x_p(I_n, I_1) & x_p(I_n, I_2) & L & x_p(I_n, I_n) \end{vmatrix}$$

where $\xi_p(I_i, I_j)$ is the transition probability from instruction I_i to instruction I_j. The transition probability matrix is calculated and select k features from the matrix traversal are needed, test vectors are mapped onto selected k vectors, then verification vector feature extraction is complete.

20.2.3 Classifier Construction Based on S²LS-SVM

In the microprocessor verification process we want to get accurate classifier with small samples. Least squares semi-supervised learning algorithm is used in classifier construction. Its basic idea is: first, an initial classifier is constructed with labeled samples using S²LS-SVM, then unlabeled samples are marked iteratively based on area labeling approach, adjustment "inconsistent" sample labels in semi-marked samples set dynamically; finally, the labeled sample set and semi-labeled samples are in training together. The construction process is shown in Fig. 20.3:

The binary of classier problem of S²LS-SVM can be described as follows:

Given a mixed set of independent and identically distributed training samples with labeled and unlabeled samples $G = \{(x_1, y_1), \ldots, (x_n, y_n), x_{n+1}, \ldots, x_{n+m}\}$, $x_i \in R^d, y_i \in \{+1, -1\}$. To construct a classification decision function on mixed training dataset, to make the training samples separated by a maximum margin, and

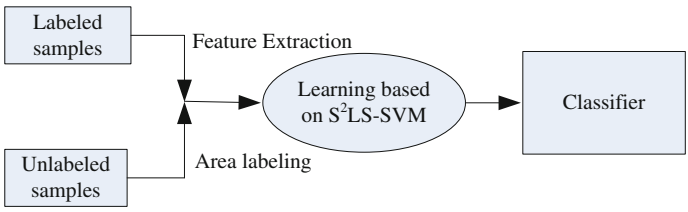


Fig. 20.3 Classifier model constriction

unlabeled samples as far as possible from the classification hyperplane. The decision function can be denoted as:

$$f(x) = \mathrm{sgn}\left[\sum_{i=1}^{n} y_i \alpha_i (x \cdot x_i) + b\right] \tag{20.1}$$

Solving decision function (20.1) needs to construct the following optimization problem:

$$\begin{cases} \min\ \Phi(\omega, e_i, e_j^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} e_i^2 + C^*\sum_{j=1}^{m} (e_j^*)^2 \\ s.t.\ \ y_i(w_i + b) = 1 - e_i, \quad e_i > 0,\ i = 1, \ldots, n \\ y_j^*\left(w \cdot x_j^* + b\right) = 1 - e_j^*, \quad e_j^* > 0,\ j = 1, \ldots, m \end{cases} \tag{20.2}$$

In the above computation the test set is supposed to be $X = \{x_1, x_2, \ldots, x_m\}$, every x_i is an n dimension vector, then $\overrightarrow{X} = \{\overrightarrow{x_1}, \ldots, \overrightarrow{x_n}\}$, kernel function $K(x_i, x_j)$ is used to measure the similarity between test vectors.

S^2LS-SVM raining process is the process of solving the above optimization problem. In this paper we propose a batch-mode labeling rule based on the value of discrimination function of unlabeled samples, selecting a batch most confident unlabeled sample as marked object, and labeling them iteratively. For convenient description, the unlabeled samples are labeled as "semi-label sample".

20.3 Experiment Results

The experimental environment settings are as follows: all verification is implemented on SUN Ultra 40 M2, the hardware configuration parameters: (CPU Type: AMD Opteron; CPU frequency: 2.2 GHz; operating systems: Solaris 10 Pre-Installed; Memory Capacity (GB): 4 GB; random hard drive capacity (GB): 250). Simulator Synopsys' VCS, VCS has a good performance on multi-core CPU verification. As used herein, the ALP module of VCS is used in OpenSPARC T1 processor core verification. This processor core is a highly scalable and high power efficiency multi-core processor. Unified coverage model (including functional, code and assertion coverage of user-defined attributes) and SV language is used in our implementation.

Prior to the processor test, classifier modeling experiments using two benchmark circuits are implemented first; the classifier generation time was 1.24 and 1.38 s. The required number of training vectors is shown in Fig. 20.4. Classifier learning success rate is a little lower for the learning process which has a certain amount of unlabeled sample participation and improving the success rate of our classifier is also the direction of future work.

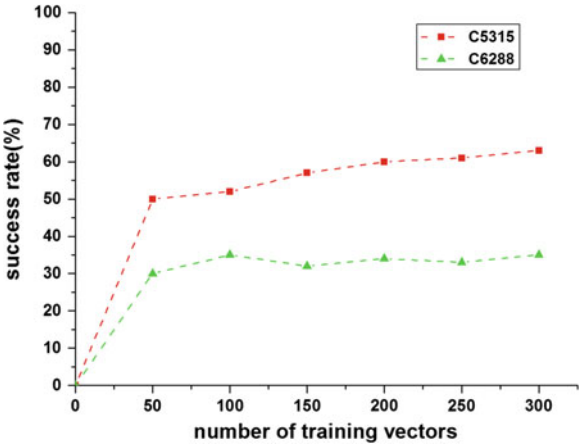


Fig. 20.4 The classifier model constriction with benchmarks

To get verification results for comparison, we generated test vector randomly and generate the appropriate classification model by our method. By comparison, a lot of redundant test vectors are filtered out by our classifier, this method uses fewer test vectors to achieve the same coverage. Experiments using fewer test vectors ($\approx 10k$) can achieve a coverage more than 96.35 %. Figure 20.5 illustrates the verification efficiency of this method.

To implement the experiment we selected 6 benchmarks from ISCAS89 in order to further analyze the precision of classifier. Compared to the Unsupervised support vector analysis approach in the literature [9] and [11] and Boolean Miner approach, our approach is better on learning accuracy than unsupervised analysis method and SVM method. The detailed result is shown in Fig. 20.6.

Finally, experimental analysis compares this method with respect to the validation of random simulation verification; the experiment results are shown in

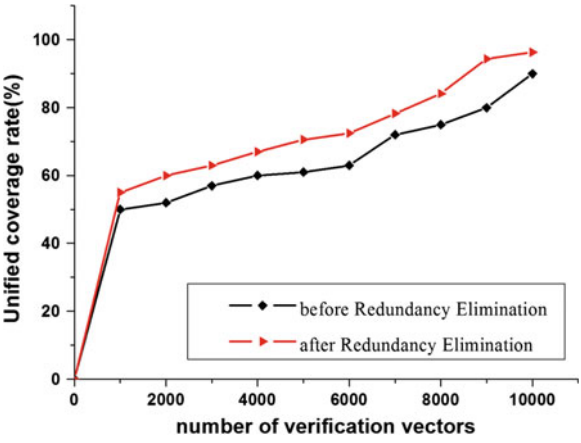


Fig. 20.5 Redundant vector reduction and coverage analysis

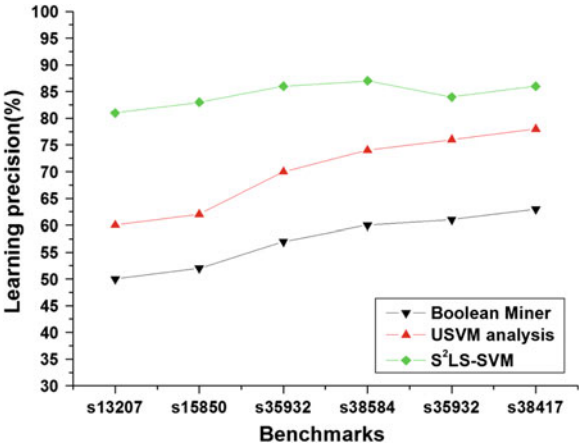


Fig. 20.6 Learning precision comparison of benchmarks

Table 20.1 Verification of OpenSPARC T1 Processor module

Simulation module	Random verification (%)	S^2LS-SVM verification (%)	Verification speedup
OpenSPARC ALU	92.45	97.68	1.73
OpenSPARC EXU	93.15	96.43	1.86

Table 20.1 through the simulation verification of several OpenSPARC T1 processor modules; the overall processor verification coverage of our method over the random verification is greatly improved.

20.4 Conclusion and Prospect

This paper discusses the application of the semi-supervised least squares support vector machine in simulation verification, the method can be used in a variety of test generation methods. It can be predicted that vector simulation optimization techniques based on machine learning will have great prospects. Research will focus on simulation vectors generation and labeled sample pretreatment in the next generation technology, a combination of active learning in classifier construct and optimization is also a worth development direction, and that is one of our future works.

Acknowledgments This work was supported by the National Natural Science Foundation of China (No.51104157). The fundamental research funds were from the Central Universities (2010NQA28), the National Postdoctoral fund (20070421041), and National college students' entrepreneurial practice project (201310290080)

References

1. The International Technology Roadmap For Semiconductors (2013) ITRS2012. http://www.itrs.net/Links/2012ITRS/Home2012.htm. Accessed 1 Feb 2013
2. Yang G, Sikun L, Wanxia Q (2012) Verification of on-chip multi-core processor: challenges, status, forecasts. J Comput Aided Des Comput Graph 24(12):1521–1532
3. Zhao Y, Lv T, Li HW, Li XW (2009) A novel circuit SAT solver in unbounded model checking. Chin J Comput 32(6):1110–1118
4. Zhao Y, Bian J, Deng S (2010) Constraints decomposition for RTL verification by SMT. J Comput Aided Des Comput Graph 22(2):234–239
5. Shen H, Wang P, Wei W et a1 (2009) A coverage directed test generation platform for microprocessors using genetic approach. J Comput Res Dev 46(10):1612–1625
6. Wang P, Guo Q, Shen H, Chen Y, Zhang H (2010) An approach to microprocessor simulation vector optimization using SVM. High Technol Lett 01:68–74
7. Guo Q, Chen T, Chen Y (2012) Model tree based multi-core design space exploration. J Comput Aided Des Comput Graph 24(06):710–720
8. Guo Q (2012) Online filtration of stimuli for microprocessor verification. J Comput Aided Des Comput Graph 24(05):690–698
9. Guzey O, Wang L, Levitt JR et al (2010) Increasing the efficiency of simulation-based functional verification through unsupervised support vector analysis. Comput Aided Des Integr Circuits Syst IEEE Trans 29(1):138–148
10. Bergeron J, Cerny E, Hunter A (2007) SystemVerilog verification methodology. BeiHang University Press, Beijing
11. Wen CHP, Guzey O, Wang L et al (2007) Simulation-based functional test justification using a boolean data miner. In: International Conference on Computer design, ICCD 2006. IEEE, pp 300–307

Chapter 21
Multiview Image Classification via Nonnegative Least Squares

Longfei Wu, Hao Sun, Kefeng Ji, Yaxiang Fan and Ying Zhang

Abstract Multiview object classification and recognition is of great importance in many applications such as computer vision and robotics by Kuo and Nevatia (Applications of computer vision (WACV), p. 18, 2009). This paper focuses on the specific case of multiview pedestrian image classification. The contributions of this paper are twofold. First, we collected a new multiview pedestrian dataset, which has been manually labeled with viewpoint, posture, and scene category tags. Second, a nonnegative least square (NNLS)-based multiview pedestrian image classification method is presented by Pang et al. (IEEE transactions on image processing, vol 20, pp. 1388–1400). Experimental results demonstrate that the proposed method is robust and effective.

Keywords Multiview pedestrian dataset · Object recognition · Nonnegative least square · Image classification

21.1 Introduction

Multiview object (such as face, human, and car) classification and recognition [1] in images and video frames is an important problem in many applications such as robotics, entertainment, surveillance, and pedestrian warning for driving assistance [2, 3]. Multiple observations of the same object at different viewpoints generate distinct even heterogeneous samples [4–6]. Object structural disparities among different viewpoints often lead to high risk of classification error. In this paper, we focus on a specific case: pedestrians. Multiview pedestrians have a large intraclass

L. Wu · H. Sun (✉) · K. Ji · Y. Fan · Y. Zhang
College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, People's Republic of China
e-mail: clhaosun@gmail.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_21

Fig. 21.1 The pedestrian of different scene in daily life

variation and the appearance changes dramatically as the viewpoint changes, as shown in Fig. 21.1. The problem is further complicated by occlusion, cluttered background, varying illumination and postures [7].

Pedestrian modeling and classifier design are the two major problems involved in pedestrian recognition systems. Various visual feature descriptors have been proposed for human and pedestrian recognition including Haar-like feature [8], HOG [9], Pyramid Local Binary Pattern [10]. Some recent research [11] suggest that a mixture of different kinds of visual features, motion, and depth information have better performance. Features on labeled samples are then fed into a classifier for pedestrian modeling. Linear SVM and the sparse representation-based classification (SRC) are among the most popular classifiers.

Methods for multiview pedestrian recognition generally fall into two categories, namely joint multiview modeling and divide and conquer modeling. Joint multiview modeling often shares features or parts between different views. Deformable part-based model (DPM) [12] is an elegant framework where view and posture variation are handled by the local deformation. Two ways are usually adopted in the divide and conquer strategy. One is to build parallel cascades, and the other consists in building a hierarchical structured classifier.

Due to the expensive cost of collecting enough samples of each pedestrian at each viewpoint, the number of multiview pedestrian samples is usually in the order of tens or hundreds while the number of features is in the order of thousands (such as 3780 dimensional HOG feature) or even tens of thousands. The low sample size and high dimensionality of multiview pedestrian data lead to inaccurate prediction results and slow classification performance. In this paper, we proposed to recognize multiview pedestrians based on nonnegative least squares (NNLS). In order to verify the effectiveness of the proposed method, we also collect a new database for multiview pedestrian detection and recognition. Experimental results under various viewpoints and postures have demonstrated that the proposed method is robust and efficient.

21.2 Multiview Pedestrian Dataset

The availability of public image repositories and recognition benchmarks has enabled rapid progress in visual object category and instance detection. For multiview pedestrian recognition, several datasets including the SDL dataset, the TUDBrussels dataset, and the INRIA dataset [7] can be used. In this paper, we contribute a new multiview pedestrian detection and recognition dataset. Compared to the existing datasets, our dataset is strictly divided into twelve distinctive continuous viewpoints from 0°–360°. Our dataset consists of 542 samples of three different persons performing three different actions (standing, walking, and running) in seven typical scenarios (square, garden, classroom, basketball court, buildings, road, and parking lot). All the samples have been resized to 100 × 32 pixels (Fig. 21.2).

The samples in our dataset have distinctive and continuous viewpoints ranging from 0° to 360°. Figure 21.3 shows an example of a subject pictured from 12 viewpoints at basketball court. In order to evaluate the impact of background clutter, we have also collect multiview pedestrian samples in various scenarios.

Fig. 21.2 Viewpoint ground truth in our dataset

Fig. 21.3 Several typical scene categories in our dataset

21.3 Nonnegative Least Squares

21.3.1 Nonnegative Matrix Factorization

Nonnegative matrix factorization (NMF) was originally proposed to model physical and chemical processes and has become increasingly popular in machine learning, signal processing, and computer vision in recent years. NMF codes naturally favor sparse, parts-based representation which would greatly benefit the multiview classification problem. Suppose the training data and the unlabeled but unknown data are represented by $X \in R^{m\times n}$ and $S \in R^{m\times p}$, respectively, where each column represents a sample and each row corresponds to a feature. The class labels of these n training samples are in the vector $c \in \{0, 1, \ldots, C-1\}^n$ where C is the number of total classes. The NMF is then formulated as

$$\min_{A,Y} \frac{1}{2}||X - AY||_F^2, \quad \text{s.t.} \quad A, Y \geq 0 \tag{21.1}$$

where X must be nonnegative and $\|\mathrm{g}\|_F^2$ is the Frobenius norm. Each sample (a column of X or S) is approximated by a nonnegative superposition of basis vectors (columns of A), that is $x_i \approx Ay_i$.

21.3.2 Nonnegative Least Squares

By coding a testing sample as a sparse linear combination of all training samples and then classifying it by evaluating which class leads to the minimal coding residual,

sparse representation-based classification (SRC) leads to interesting results for robust face recognition and a variety of tasks. Recent research has shown that the collaborative representation mechanism used in SRC is much more crucial to its success. Similar to the collaborative representation techniques, we can replace the basis matrix in NMF with the training data, and a testing sample is therefore approximated by a nonnegative and probably sparse combination of the training samples [13].

The NNLS classifier consists of two major steps [14, 15].

- Solve the NNLS optimization problem:

$$\min_{Y} \frac{1}{2}||S - XY||_F^2, \quad \text{s.t.} \quad Y \geq 0 \tag{21.2}$$

  where $X \in R^{m \times n}$ is the training set with m features and n samples (with label c) from C classes, $S \in R^{m \times p}$ is testing samples without labels. The optimization tries to find $s_i \approx Xy_i$.
- Predict the class label for the unknown samples:

$$p_i \leftarrow NS(y_i) \tag{21.3}$$

The predictor use the nearest subspace (NS) rule, which takes advantage of the discriminating property of sparse coefficients, and is generally more robust to noise. Suppose there are C classes with labels $0, 1, \ldots, C-1$, for a given new sample s, after obtaining its coefficient vector y, the regression residual corresponding to class i is computed as

$$r_i(s) = \frac{1}{2}||s - X\delta_i(y)||_2^2 \tag{21.4}$$

where $\delta_i(s) : R^n \rightarrow R^n$ returns the coefficients for class i. Its j-th element is defined by

$$(\delta_i(s))_j = \begin{cases} y_i & \text{if} \quad x_j \text{ in class } i \\ 0 & \text{otherwise} \end{cases} \tag{21.5}$$

Finally, a class label q is assigned to s, where

$$q = \min_{0 \leq i \leq C-1} r_i(s) \tag{21.6}$$

21.4 Experimental Results and Analysis

The effectiveness of the proposed method has been evaluated by multiview classification tasks. The 542 twelve-view pedestrian samples in our dataset have been converted to gray format and used for the classification. All the samples are resized

to 100 × 32 pixels and reshaped to a 3200-dimensional row vector. We use k-fold ($k = 3$ in our experiment) cross validation to randomly select 361 samples to form the training set and the rest 181 samples to form the test set. We repeated each classification experiment 20 times and reported the average classification accuracy. Table 21.1 shows the sample number and class label for each viewpoint.

We have conducted three groups of classification experiments on a machine with 3.00GHZ Inter CPU and 2.98 GB RAM based on matlab implementation.

First, we directly use the gray image vectors in the training set to form the basis matrix in NMF. Multiviewpoint classification is then performed using coefficients of testing samples based on nearest subspace rule. We compare NNLS methods with NMF methods (clusters 4, 8, 12, and 24) and the classification results are shown in Table 21.2. The classification time of NNLS and NMF with different clusters is reported in Table 21.3. It can be seen from Tables 21.2 and 21.3 that the NNLS method consistently outperforms NMF method with less computation time.

Second, the proposed NNLS-based classification method has been compared with several state-of-the-art methods including sparse representation-based classification (SRC), metasample sparse representation-based classification (MSRC), linear regression-based classification (LRC), and SVM with linear kernel. For the SRC method, the parameter lambda was set to 0.001. For the MSRC method, NMF with six clusters was used to generate the metasample and the lambda was set to 0.1. We also extract a 4788-dimensional HOG feature for each sample besides

Table 21.1 Sample number in each viewpoint

	Viewpoint											
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
Sample number	53	44	50	43	40	40	48	36	50	46	45	47
Label	0	1	2	3	4	5	6	7	8	9	10	11

Table 21.2 Performance of NNLS and NMF with different clusters

Methods	Classification accuracy												
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°	Balanced
NNLS	0.56	0.64	0.62	0.35	0.46	0.64	0.69	0.75	0.56	0.53	0.67	0.56	0.59
NMF(4)	0.47	0.43	0.71	0.40	0.38	0.69	0.63	0.42	0.65	0.47	0.53	0.67	0.54
NMF(8)	0.72	0.40	0.59	0.57	0.38	0.31	0.31	0.33	0.75	0.19	0.47	0.47	0.46
NMF (12)	0.61	0.73	0.44	0.36	0.38	0.21	0.56	0.75	0.41	0.60	0.60	0.20	0.49
NMF (24)	0.44	0.21	0.24	0.40	0.50	0.62	0.50	0.42	0.38	0.47	0.33	0.19	0.39

Table 21.3 Classification time of NNLS and NMF with different clusters

Classification time	Classification methods				
	NNLS	NMF(4)	NMF(8)	NMF(12)	NMF(24)
Seconds	2.23	8.22	9.75	15.19	31.04

Table 21.4 Twelve-view classification results using different methods

Methods	Classification accuracy												
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°	Balanced
NNLS	0.56	0.64	0.62	0.35	0.46	0.64	0.69	0.75	0.56	0.53	0.67	0.56	0.59
NNLS + HOG	0.83	0.60	0.47	0.67	0.46	0.54	0.75	0.42	0.65	0.80	0.53	0.73	0.62
SRC	0.67	0.60	0.76	0.57	0.50	0.42	0.75	0.33	0.82	0.60	0.47	0.56	0.60
SRC + HOG	0.78	0.53	0.82	0.71	0.50	0.43	0.88	0.58	0.82	0.53	0.40	0.69	0.65
MSRC + NMF(6)	0.72	0.53	0.59	0.36	0.50	0.29	0.50	0.17	0.48	0.40	0.40	0.31	0.45
MSRC + NMF(6) + HOG	0.78	0.60	0.71	0.60	0.38	0.43	0.69	0.50	0.81	0.27	0.40	0.53	0.57
LRC	0.83	0.40	0.71	0.43	0.64	0.36	0.75	0.25	0.82	0.53	0.40	0.56	0.57
LRC + HOG	0.72	0.53	0.71	0.53	0.46	0.57	0.88	0.50	0.81	0.40	0.67	0.60	0.62

Table 21.5 Four-view classification results using different methods

Methods	Classification accuracy				
	0°–90°	90°–180°	90°–270°	270°–360°	Balanced
NNLS	0.92	0.76	0.75	0.74	0.79
NNLS + HOG	0.82	0.88	0.82	0.83	0.83
SRC	0.84	0.71	0.82	0.87	0.81
SRC + HOG	0.84	0.88	0.93	0.78	0.86
MSRC + NMF (6)	0.67	0.80	0.52	0.70	0.67
MSRC + NMF (6) + HOG	0.71	0.76	0.77	0.80	0.76
LRC	0.96	0.71	0.77	0.74	0.79
LRC + HOG	0.82	0.85	0.84	0.78	0.82
SVM (linear kernel)	0.80	0.83	0.57	0.63	0.71
SVM (linear kernel) + HOG	0.82	0.85	0.82	0.80	0.82

using the original gray level vector. The classification accuracy is reported in Table 21.4. The linear kernel SVM cannot correctly classify the twelve viewpoints. It can be seen that the SRC + HOG method achieve the best classification accuracy, followed by the NNLS + HOG method and the LRC + HOG method. For multiview pedestrian classification, the HOG feature (4788-dimensional row vector in our case) consistently outperforms the original gray level pixel vector (3200-dimensional row vector in our case). The SRC + HOG method performs slightly better than NNLS + HOG method and it is computational expensive. The average classification time is over 1 h for SRC + HOG while NNLS + HOG can be accomplished in several minutes.

Third, we repartitioned the viewpoint to four intervals, 0°–90° (including 0°, 30°, 60°), 90°–180° (including 90°, 120°, 150°), 180°–270° (including 180°, 210°, 240°), and 270°–360° (including 270°, 300°, 330°). A four-viewpoint classification task is carried out using the same configuration as in the second group. The statistics of the results are shown in Tables 21.5 and 21.6. The SRC + HOG method still holds the best classification accuracy. NNLS + HOG perform slightly better than SVM with linear kernel. For the four-view classification experiments, the balanced accuracy has improved nearly 20% compared to the twelve-view case.

Table 21.6 Classification time of various methods for four-viewpoint

Classification time	Classification methods				
	NNLS + HOG	SRC + HOG	MSRC + NMF (6) + HOG	LRC + HOG	SVM (linear kernel) + HOG
Seconds	3.52	2860.40	3.37	0.09	0.06

21.5 Conclusion

Multiview pedestrian recognition finds many applications. In this paper, we contribute to the community a multiview pedestrian data. Motivated by the NMF and the collaborative representation mechanism in sparse presentation, we then present a novel multiview pedestrian classification method based on nonnegative least squares. Three groups of experiments have been carried on the multiview pedestrian dataset. The experimental results demonstrate that (1) HOG feature is better than the original gray pixel value vector for multiview pedestrian description. (2) NNLS classifier can achieve comparable classification results with SRC while using less computation time. For future work, we will try to adopt a hierarchical NNLS classification strategy.

Acknowledgments The authors would like to thank Mr. Guanliang Zhang for collecting and organizing the multiview pedestrian dataset. This work was supported in part by the National Natural Science Foundation of China under Grant 61303186, 61240058, and by the Ph.D. Programs Foundation of Ministry of Education of China under Grant 20124307120013.

References

1. Zheng J, Jiang LV, Phillips PJ, Chellappa R (2012) Crossview action recognition via a transferable dictionary pair. In: BMVC 2012, p 111
2. Kuo CH, Nevatia R (2009) Robust multiview car detection using unsupervised subcategorization. In: Applications of computer vision (WACV), p 18
3. Pang J, Huang Q, Yan S, Jiang S, Qin L (2011) Transferring boosted detectors towards viewpoint and scene adaptiveness. In: IEEE transactions on image processing, vol 20, pp. 1388–1400
4. Kan M, Shan S, Zhang H (2012) Multiview discriminant analysis. In: ECCV 2012, Part I, LNCS, vol 7572, pp 808–821
5. Zhang H, Nasrabadi NM, Huang TS, Zhang Y (2011) Multiview face recognition via joint dynamic sparse representation. In: International conference on image processing, pp 3025–3028
6. Timofte R, Zimmermann K, Gool LV (2009) Multiview traffic sign detection, recognition, and 3D localisation. In: Applications of computer vision (WACV), p 18
7. Ye Q, Han Z, Jiao J, Liu J (2013) Human detection in images via piecewise linear support vector machines. IEEE Trans Image Process 22:778789
8. Viola P, Jones M, Snow D (2005) Detecting pedestrians using patterns of motion and appearence. Int J Comput Vis 63:153161
9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE CVPR, pp 886–893
10. Sun H, Wang C, Wang BL, Elsheimy N (2011) Pyramid binary pattern features for realtime pedestrian detection from infrared video. Neurocomputing 74:707804
11. Enzweiler M, Gavrila DM (2011) Multilevel mixture of experts framework for pedestrian classification. IEEE Trans Image Process 20:2967–2979
12. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645

13. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the 2011 international conference on computer vision, pp 471–478
14. Heiler M, Schnorr C (2005) Learning nonnegative sparse image codes by convex programming. In: Proceedings of the 2005 international conference on computer vision, pp 1667–1674
15. Li Y, Ngom A (2013) Nonnegative least squares methods for the classification of high dimensional biological data. In: IEEE/ACM transactions on computational biology and bioinformatics, vol 10, pp 447–456

Chapter 22
An Automatic Method for Selecting Special and Rare Celestial Objects in Massive Spectra

Wenyu Wang and Bin Jiang

Abstract We proposed an automatic method for searching special and rare celestial objects in the massive spectra of the Sloan digital sky survey (SDSS). The data mining technique is employed and the massive SDSS spectra are identified quickly and efficiently. The high-dimensional spectra are mapped to feature space constructed by the principal component analysis (PCA), and dimensionality reduction is carried out accordingly. Massive SDSS spectra are classified by a well-trained support vector machine (SVM) and most of the noncandidates are excluded. Parameter optimization is also studied to guarantee the accuracy of PCA and SVM. Experiments show that this novel method can find rare celestial objects in an effective and efficient manner. We report the identification of six new white dwarf-main sequence (WDMS).

Keywords Data mining · Massive spectra · PCA · SVM

22.1 Introduction

The Sloan Digital Sky Survey [1] used a dedicated 2.5-m telescope at Apache Point Observatory, New Mexico, equipped with two powerful special-purpose instruments. The 120-megapixel camera imaged 1.5 square degrees of sky at a time. A pair of spectrographs fed by optical fibbers measured spectra of more than 600 galaxies and quasars in a single observation. The current phase of SDSS is SDSS-III [2] which is working to map the Milky Way, search for extra solar planets, and solve the mystery of dark energy.

The massive SDSS spectra are ideal for searching special and rare objects like white dwarf-main sequence (WDMS) binaries. WDMS is a binary star system

W. Wang · B. Jiang (✉)
School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China
e-mail: jiangbin@sdu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_22

containing a white dwarf primary and a low-mass main-sequence companion. WDMS is central to our understanding of binary evolution. The remaining ~1/4 of the WDMS binaries are close binaries which are progenitor candidates for cataclysmic variables and Ia supernovae. Research on WDMS has become more focused in recent years with the running of more sky survey telescopes. SDSS provides an invaluable observational dataset in which a large number of WDMS binaries have been efficiently identified. Raymond identified 109 white dwarf-M dwarf pairs with $g < 20$ m in SDSS DR1 [3]. Silvestri presented a spectroscopic sample of 747 detached close binary systems from the SDSS DR4, and the majority of these binaries are WD-M binaries [4]. Further, a catalog with 1253 WDMS binary systems from the SDSS DR5 was present. Heller identified 857 WDMS binaries from the DSS DR6 through a photometric selection method [5]. Rebassa-Mansergas provided a catalog of 2248 WDMS binaries identified from the SDSS DR7 [6].

However, the above methods they used require significant manual processing and cannot fulfill real-time processing requirements. Worse still, some candidates are left out. Thus, it is necessary to develop automatic methods like data mining technique for WDMS searching.

In this paper, we present a new application of principal component analysis (PCA) combined with support vector machine (SVM) to find WDMS in SDSS DR10 spectral data. We use the existing discovered WDMS spectra as templates to construct the feature space and train the classifier. Massive SDSS spectra are mapped to the feature space and then classified by SVM. Most of the non-WDMS are excluded and the greatly reduced final list of candidates can be identified manually and easily.

22.2 The Experimental Data

The experimental data are the DR10 optical spectra from SDSS. DR10 is the first release of the spectra from APOGEE, which uses infrared spectroscopy to study tens of thousands of stars in the Milky Way. DR10 also includes hundreds of thousands of new galaxy and quasar spectra from BOSS. In addition to the 57,454 infrared stellar spectra taken by du Pont, DR10 contains 1,848,851 galaxy spectra, 308,377 quasar spectra, and 736,484 stellar spectra selected over 14,555 square degrees. The 1D pipeline of SDSS classifies spectra based on the method of template matching. The spectral classification of every DR10 spectrum is designated with the keyword “subclass.” At present, the classification accuracy of special and rare objects like WDMS is still not satisfactory. The 896 discovered WDMS as mentioned above are selected as template spectra to construct the feature space and for training classifier. Most of them are WD-M binaries.

22.3 Method and Implementation

The experiment procedure consists of data preprocessing, feature extraction by PCA, sample classification by SVM, and parameter optimization which is discussed in detail in following.

22.3.1 Data Preprocessing

In our experiment, data preprocessing includes low signal-to-noise ratio (SNR) spectra reduction, wavelength unification, and raw counts rescaling, described as follows:

(1) Culling low SNR spectra. The information contained in SDSS spectra with SNR < 5 is usually drowned in noise. Most of these spectra are of poor quality with various problems and cannot be identified correctly even by manual identification.
(2) Unifying the wavelength to 3800–9200 Å with fixed step length.
(3) Normalizing the flux with formula (22.1):

$$x_i = x_i \Big/ \sqrt{\sum_{j=1}^{M} x_j^2} \tag{22.1}$$

where x_i is the count of each pixel of the spectrum, and M is the number of pixels per spectrum. Through the above steps, the data are trimmed to match the wavelength range with accepted S/N and all counts in a spectrum are rescaled to [0, +1].

22.3.2 Feature Extraction by PCA

The SDSS spectra are high dimensional. PCA is a dimensionality reduction method that reduces redundant information by approximating a large number of statistical variables with a smaller number of significant linear combinations. Besides, PCA can also be employed to extract the feature that is the key to classifications. PCA applies a mathematical procedure for transforming a number of correlated variables into a smaller number of uncorrelated variables called principal components (PC). The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA is widely used in astronomical data processing: Bertram used PCA to study the gas dynamics in numerical simulations of typical molecular clouds [7]. Ishida proposed the use of kernel PCA for supernovae

photometric classification [8]. Blustin applied PCA to study the variability of the X-ray continuum in the Seyfert 1 galaxy NGC 7469 [9].

With the method of PCA, a high-dimensional spectrum is transformed to a lower dimensional spectrum, and the sample projection is carried out accordingly. Some information will be lost, but this can be negligible if the eigenvalues are small. With most components left out, the final data will have fewer dimensions and the computation could be significantly reduced. The process of PCA is the last step of data preprocessing and described in detail as follows:

(1) Label SDSS WDMS template spectra as Pi($i \in [1, m]$), m is the number of WDMS template spectra;
(2) Construct WDMS template spectra matrix $P_{m\times n}$, where each row with n components (the number of pixels per spectrum) is normalized;
(3) Construct correlation matrix of $P_{m\times n}$: $C_{n\times n} = P^{T}_{n\times m} * P_{m\times n}$;
(4) Calculate and sort ascending the eigenvalue of $C_{n\times n}$.

Once eigenvectors are achieved from $C_{n\times n}$, order them according to decreasing eigenvalues. The principle components are the eigenvectors corresponding to the largest eigenvalues and the number of the final selected eigenvectors is discussed in detail in Sect. 22.3.4. We designate the value as k and thus the top k maximal eigenvectors are selected and others are abandoned.

Each maximal eigenvector is called an eigenspectrum. The first principle components are called the max feature spectrum and shown in Fig. 22.1. Obvious white dwarf combined with M-type star spectra can be found that is identical to the characteristic of templates.

The transformation matrix $E_{n\times k}$ consisting of k maximum eigenvectors is the eigenspectra matrix, and the corresponding k-dimensional space is the eigenspace (feature space). $E_{n\times k}$ can then be used to project any n-dimensional spectra to the k-dimensional feature space: $\bar{P}_{1\times k} = \tilde{P}_{1\times n} \times E_{n\times k}$, and a n-dimensional spectrum is then transformed to a k-dimensional spectrum ($n \gg k$).

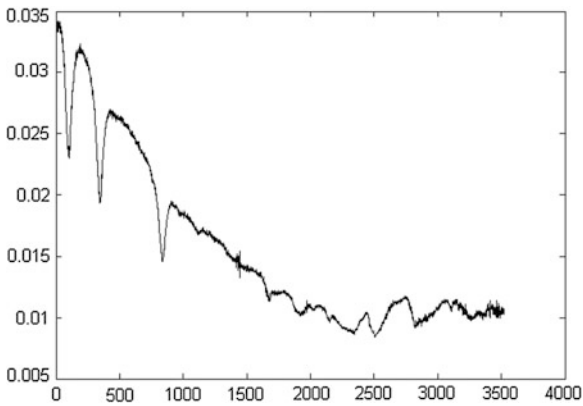


Fig. 22.1 WDMS template spectra from SDSS

22.3.3 Sample Classification by SVM

The key to classification in feature space is to achieve a decision boundary such that not only are the classes separated, but also that the separation is as large as possible. SVM is a supervised machine learning algorithm that can be used for predicting multivariate or structured outputs. SVM performs classification by constructing a hyperplane that optimally separates the data into two categories in a multidimensional space.

SVM has been applied to various astronomical applications such as the selection of Active Galactic Nuclei (AGN) candidates [10], the determination of photometric redshift, the classification of galaxies using synthetic galaxy spectra, and the morphological classification of galaxies using image data. Wozniak used SVM, *K*-means for automated classification of variable stars and compared their effectiveness to traditional methods [11]. Humphreys used decision trees, Knn (K-Nearest Neighbor), and SVM for classification of the morphological type of galaxy [12]. Their results show that SVM is especially efficient in isolating specific classes from the rest of the observed objects in high-dimensional space. One of the key problems of SVM is determining a kernel function that can transform the data into a higher dimensional space to make it possible to perform the separation. There are four kernel functions including linear, polynomial, radial basis function (RBF), and sigmoid. In our experiment, RBF kernel function is used for SVM. The definition of RBF function is: where σ is the parameter of RBF.

For the fixed functional form of the kernel, model selection amounts to tuning kernel parameters and the slack penalty coefficient C. The parameter C is the penalty for misclassification, a larger C corresponding to assigning a higher penalty to errors. If C is set large, the number of training errors will be reduced, but the classifier performs poorly on testing data. If C is too small, the number of training errors will increase such that the classifier performs poorly on the training set and also may not perform well on the testing data. Generally, one must set C to a fiducial value and check the classification accuracies, then adjust the parameter if needed to obtain a better model.

The performance of SVM also depends highly on a representative training set. In our experiment, the training samples consists of two parts. The first part is the 896 identified WDMS spectra that are used as positive samples and the second part is the 20,000 randomly selected SDSS spectra used as negative samples. The negative samples should be composed of most kinds of spectra except WDMS spectra. The spectral classification of every SDSS spectrum is provided by SDSS pipeline, according to which we adjust the negative samples to avoid the selection effect of the random selecting algorithm. We also make sure that there are no WDMS spectra in negative samples through manual review.

22.3.4 Parameter Optimization

The performance of SVM and PCA are both highly affected by their parameters and we need a way of constraining the final parameters objectively. The performance of SVM is highly affected by σ: The bigger of the σ, the more the trend of SVM to be a linear classifier; otherwise, SVM will be poorly trained or overtrained which depends on C. In our experiment, we try the combinations of C and σ with the range C = [0.001 0.01 0.1 1 10 100 1000] and σ = [0.001 0.01 0.1 0.5 1 10 20], and use grid search with cross-validation to select the best combination. The final accuracy can reach 99.5861 %.

The visualization of tuning C and σ is shown in Fig. 22.2.

The key parameter of PCA is the number of the feature space which is generally defined by experience. In our experiment, we first use formula (22.2) to calculate the variance contribution rate μ and construct the transformation matrix E. Results are shown in Table 22.1 and Fig. 22.3, respectively.

$$\mu = \sum_{i=1}^{L} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \qquad L < n \tag{22.2}$$

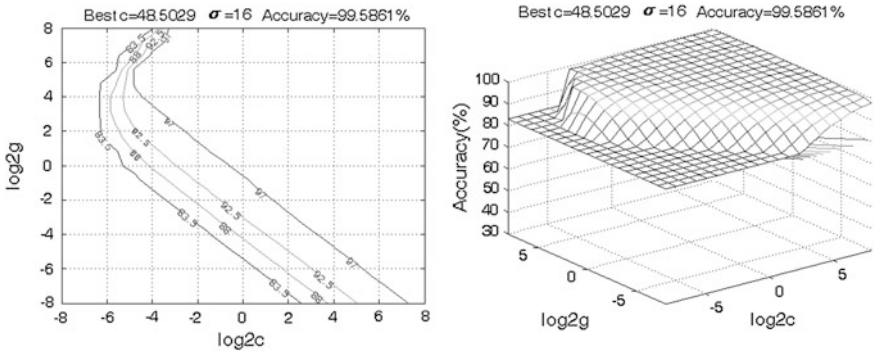


Fig. 22.2 Accuracy contour (*top*) and surface plot (*bottom*) of C and σ

Table 22.1 Variance contribution rate

Dim	μ (%)	Dim	μ (%)
1	87.31	30	98.53
2	91.13	35	98.75
3	93.33	40	98.92
4	94.00	42	98.98
5	94.49	43	99.01
10	96.24	44	99.04
20	97.84	45	99.07
25	98.26	50	99.20

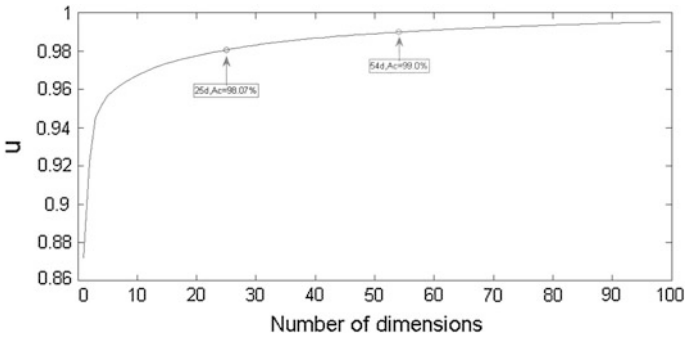


Fig. 22.3 Variance contribution rate

In the mathematical sense, the symbol μ represents the proportion of variance of PCA against all variance and capability of explaining most of the variance in the original data. The parameter μ is often used to constraint the final number of dimensions. The higher the μ value, the better the completeness and accuracy of PCA, but the computation will increase sharply accordingly. Thus, maximizing μ with least number of components is the key to PCA. In the experiment of Qin-DongMei et al. (2003), μ reached 98 % with the largest two eigenvalues selected. In our experiment, we found μ can reach only 89.24 % in the same situation. It can be seen from and Fig. 22.3 that the number of 54 is the point where μ already reaches 99 % and increases still slowly with increasing dimensions afterward. Given the accuracy and computation, the ideal number of dimensions here we can choose is 54, which yields a relevant variance contribution rate-μ of 99.00 %. The top 54 maximal eigenvectors are selected and others are abandoned. We also tried another way to constraint the parameter of PCA constrained by SVM. We first use PCA method to map the training set to 3–100 dimensional feature space, to be precise, we take the k of PCA to the value from 3 to 100, and then we use the k-dimensional ($k \in [3–100]$) training set to self-training and self-testing with the fixed SVM. During the process of SVM training, the kernel function and the range of parameters are fixed. The accuracy and time of self-testing of SVM are achieved which are shown in Fig. 22.4.

In Fig. 22.4 the accuracy of SVM self-testing are 99.9136 and 99.8173 % when the number of feature space is 25 and 54, respectively. The accuracy of 25 dimensional training set is better than that of 54 dimensional feature space testing by the same SVM. A number of calculation time is saved with better accuracy and less dimensions. Therefore, in our experiment with the specified training set, we determine the final number of the feature space as 25.

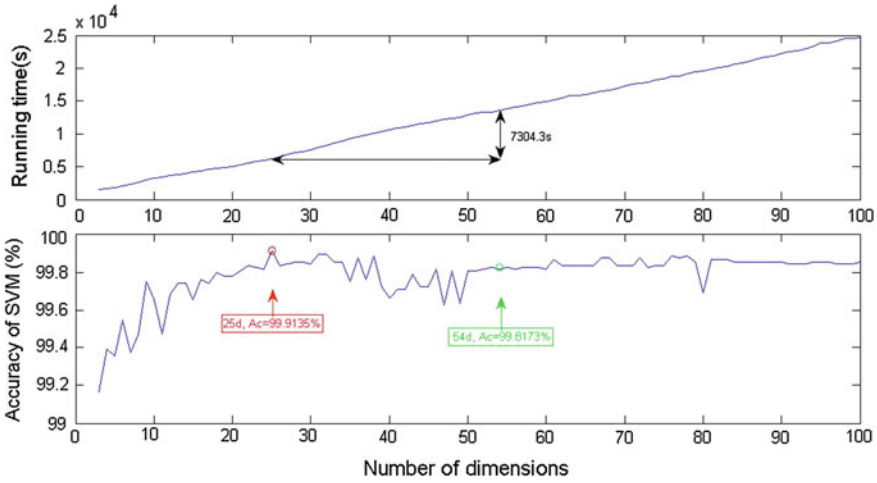


Fig. 22.4 Accuracy and running time of SVM

Table 22.2 Newly identified WDMS

RA	DEC	Plate	MJD	Fiber
240.7173	30.0478	53,496	1578	418
211.0498	50.5	53,446	1671	362
219.6209	46.5236	53,464	1674	106
241.303	8.3901	53,228	1728	617
355.9069	15.3772	53,242	1895	486
27.1771	14.2697	53,262	1899	519

22.4 Results

In the experiment, 3578 spectra are selected by SVM and 2723 of them are identified by template method as definite WDMS, of which 27 are new discoveries. After inspection, we found that 825 of them are poor quality spectra due to 2D pipeline and/or low SNR. The other 30 spectra are other objects with unusual spectra, mainly including cataclysmic variables (CVs) during outburst. In Table 22.2, six new WDMS are listed with RA, DEC, plate, MJD, and fiber.

22.5 Discussion and Conclusion

We have put forward a data mining method to search for WDMS systematically in massive spectra from the SDSS database. The proposed method is implemented on a typical quadcore desktop workstation. The overall processing time of the

experiment is within 12 h and mainly focuses on data reading and preprocessing. The experiment results demonstrate that the method runs efficiently and can be able to classify objects quickly. This method is also applicable to mining other specified celestial objects including variable objects in large sky survey telescope data. Once new celestial objects like CVs are found by the method, they can be reported to other telescopes for follow-up observation. If red shifts of objects can be calculated accurately with SDSS 1D pipeline, this method is especially applicable to supernova searching. As this is our first attempt at data mining the SDSS spectral data, we excluded only very low signal-to-noise data. As most of the templates show the feature of white dwarf combined with M-type star, we do not find galaxies and QSOs in the final results, but objects with similar binary spectral feature, and these byproducts can be easily distinguished from WDMS. That also suggests that the feature space constructed by template spectra from SDSS is not precise enough and needs improvement. SDSS is a much more systematic survey, it is expected that with the proposed method it will eventually provide strong constraints on the total number of WDMS. The SDSS data archive will be made publicly available in the standard data format for Virtual Observatories and a friendly user interface of this method will be embedded in it, enabling astronomers to explore the data interactively.

Acknowledgments This work is supported by the National Natural Science Foundation of China (11473019 and U1431102). This work has made use of the SIMBAD database, and also has made use of data products from the SDSS.

References

1. Kevork A, Zheng Z, Idit Z, David H et al (2005) Cosmology and the halo occupation distribution from small-scale galaxy clustering in the Sloan Digital Sky Survey. APJ 625: 613
2. Bertram E, Shetty R, Glover Simon CO et al (2014) Principal component analysis of molecular clouds: can CO reveal the dynamics. MNRAS 440:465B
3. Ignacio F, Anna P, de Carvalho RR et al (2006) A principal component analysis approach to the star formation history of elliptical galaxies in compact groups. MNRAS 370:828F
4. Frinchaboy PM, Thompson B, Jackson KM, Connell JO et al (2013) The open cluster chemical analysis and mapping survey: local galactic metallicity gradient with APOGEE using SDSS DR10. APJ 777:1
5. Galaz G, de Lapparent V (1998) The ESO-sculptor survey: spectral classification of galaxies with $Z < 0.5$. A&A 332:459
6. Company H et al (2009) A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. A&A 497:743
7. Ishida EEO, de Souza RS (2013) Kernel PCA for Type Ia supernovae photometric classification. MNRAS 430:509I
8. McGurk Rosalie C et al (2010) Principal component analysis of Sloan Digital Sky Survey stellar spectra. AJ 139:1261
9. Persson SE, West SC, Carr DM, Sivaramakrishnan A et al (1992) A near-infrared camera for Las Campanas Observatory. PASP 104:204

10. York Donald G, Adelman J, Anderson John, Anderson E Jr, Scott F et al (2000) The Sloan Digital Sky Survey: technical summary. AJ 120:1579
11. Wils P (2009) Data mining for Dwarf Novae in SDSS, GALEX and astrometric catalogues. Mon Notices R Astron Soc 402(1):436–446
12. Downes RA (2005) A catalog and atlas of cataclysmic variables: the final edition. J Astron Data 11:2

Chapter 23
Attitude Estimation of the Multi-rotor UAV Based on Simplified Adaptive Kalman Filter Algorithm

Xin Zhang, Yue Bai, Zhijun Xu and Rijun Wang

Abstract A simplified and improved attitude estimation algorithm is put forward, for the problems from the multi-rotor UAV in noise statistical characteristics unknown and time-varying, vibration as the main disturbance source, and high-frequent dynamic changes of the attitude angles. A simplified adaptive Kalman filter algorithm is designed for multi-sensor information fusion. With the attitude angle variance of the real-time dynamic calculation by the gyro and the accelerometer, the system noise variance and the measurement noise variance are estimated, so as to solve the unknown about the noise statistical characteristics, and to prevent the filter divergence. The flight experiment shows that the algorithm can guarantee the accuracy and the stability of the attitude angle estimation. The root-mean-square errors of the pitch and roll angles which are estimated by the simplified algorithm are 2.922° and 1.713°. The accuracy of the algorithm can meet the demands for the autonomous flight of the multi-rotor UAV.

Keywords Multi-rotor UAV · Attitude estimation · Adaptive Kalman filter · Quaternion method

X. Zhang (✉) · Y. Bai · Z. Xu · R. Wang
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
e-mail: liuguosong68@163.com

Y. Bai
e-mail: baiy@ciomp.ac.cn

Z. Xu
e-mail: xuzj538@ciomp.ac.cn

R. Wang
e-mail: wangrijun1982@126.com

X. Zhang · R. Wang
University of Chinese Academy of Sciences, Beijing 100039, China

X. Zhang
School of Electrical and Information Technology, Changchun Institute of Technology, Changchun 130012, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_23

23.1 Introduction

The multi-rotor UAV (Unmanned Aerial Vehicle) has some advantages, such as simple structure, small volume, low cost, high mobility, and is widely used in military and civil fields [1]. It is a prerequisite condition to obtain the flight attitude of high accuracy and high reliability for achieving the autonomous flight of the multi-rotor UAV [2].

MEMS (Micro Electro Mechanical Systems) gyro calculating process of attitude angle is also one on the strap-down matrix immediate updating. The typical algorithm of strap-down matrix immediate updating has three kinds [3]: the Euler angle method, the quaternion method and direction cosine method. The quaternion method is used for calculation of the attitude angle in this paper, with less computation, without orthogonal process, and with the smallest drift errors of the three methods. The attitude angle short-term precision is high used alone MEMS gyro calculation, but the calculation has serious integrator drift, and is sensitive to the vibration of the carrier. The MEMS accelerometer has better static properties, no integration to calculate attitude angle, but the short-term accuracy is poor. The separate use of one of the two will reduce the estimation precision of the attitude angle [4].

Kalman filter is currently recognized as the most suitable algorithm for the information fusion of UAV multi-sensor dynamic navigation, but the application of conventional Kalman filter algorithm is demanded to learn the statistical characteristics of the system noise and the measurement noise [5, 6]. However, in the practical applications of multi-sensor information fusion of the multi-rotor UAV, the statistical properties of the system noise and the measurement noise are unknown and time-varying, which makes the conventional Kalman filter algorithm lose its optimality, the estimation accuracy is greatly reduced, even will cause the filter divergence. This paper presents a simplified adaptive Kalman filter (SAKF) algorithm by simplifying and improving the Sage-Husa adaptive Kalman filter (SHAKF) algorithm [7]. Using the algorithm, the filter accuracy is improved, and the system noise variance and the measurement noise variance can be real-time tracked and estimated.

23.2 Quaternion Method for Immediate Updating of Strap-Down Matrix

In the multi-rotor UAV, the origin O of body coordinate system $Ox_by_bz_b$ is chosen to be located in the center of gravity of UAV, axis x_b is forward along the longitudinal axis, axis y_b is left along the horizontal axis, and axis z_b is perpendicular to the plane Ox_by_b along the vertical axis up. The navigation coordinate system $Ox_ny_nz_n$ chooses the local geographic coordinate system, namely the NEU (Northeast and Up) coordinate system. Solving the attitude of the UAV is the solution of relative relationship between the body coordinate system and the navigation coordinate system.

A quaternion can easily describe the angular motion of a rigid body, its attitude expression is the one with the four parameters, and the train of thought based on is the transformation one coordinate system to another coordinate system can be achieved by a single rotation around a defined vector in the reference coordinate system [8, 9]. Set the rotation quaternion of the body coordinate system relative to the geographical coordinates uses the symbol $\boldsymbol{Q}$ to represent for.

$$\boldsymbol{Q} = q_0 + q_1 i_b + q_2 j_b + q_3 k_b \tag{23.1}$$

The base i_b, j_b, k_b of the quaternion is consistent with the one $\bar{x}_b, \bar{y}_b, \bar{z}_b$ of the body coordinate system. Real-time updating of the quaternion in computer is completed by the numerical integration method, and the truncation error can be reduced using a high order method. The method commonly used to update the quaternion is fourth-order Runge–Kutta method [8]. The real-time updating strap-down matrix is:

$$\boldsymbol{C}_b^n = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{23.2}$$

In the multi-rotor UAV, the domain of definition about the pitch angle is $(-90°, 90°)$, the roll angle is $(-180°, 180°)$, the heading angle is $(0°, 360°)$. The principal values of the attitude angles can be determined as:

$$\begin{aligned} \theta_{\text{Main}} &= -\sin^{-1}(C_{b31}^n) \\ \gamma_{\text{Main}} &= \tan^{-1}\left(\frac{C_{b32}^n}{C_{b33}^n}\right) \\ \psi_{\text{Main}} &= \tan^{-1}\left[\frac{C_{b21}^n}{C_{b11}^n}\right] \end{aligned} \tag{23.3}$$

The truth values of the attitude angles can be expressed as:

$$\begin{aligned} \theta &= \theta_{\text{Main}} \\ \gamma &= \begin{cases} \gamma_{\text{Main}} & C_{b33}^n > 0 \\ \gamma_{\text{Main}} + 180° & C_{b33}^n < 0, C_{b32}^n > 0 \\ \gamma_{\text{Main}} - 180° & C_{b33}^n < 0, C_{b32}^n < 0 \end{cases} \\ \psi &= \begin{cases} \psi_{\text{Main}} & C_{b11}^n > 0, C_{b21}^n > 0 \\ \psi_{\text{Main}} + 360° & C_{b11}^n > 0, C_{b21}^n < 0 \\ \psi_{\text{Main}} + 180° & C_{b11}^n < 0 \end{cases} \end{aligned} \tag{23.4}$$

23.3 Nonlinear Attitude Estimation System Modeling

Based on the measurement characteristics of the good short-time dynamic performance of the gyro, and the good long-term static performance of the accelerometer, the two is fused with Kalman filter algorithm, the attitude calculation of the gyro as state prediction information, the attitude calculation of the accelerometer as measurement information. The nonlinear state space model of discrete time of the multi-rotor UAV attitude estimation is:

$$\begin{cases} \boldsymbol{X}_k = \boldsymbol{f}(\boldsymbol{X}_{k-1}) + \boldsymbol{W}_{k-1} \\ \boldsymbol{Z}_k = \boldsymbol{H}_k \boldsymbol{X}_k + \boldsymbol{V}_k \end{cases} \tag{23.5}$$

where

$\boldsymbol{X}_k = [\gamma_k \quad \theta_k \quad \psi_k]^T$ is 3 × 1-dimensional state vector,

$\boldsymbol{W}_k = [w_{\gamma k} \quad w_{\theta k} \quad w_{\psi k}]^T$ is 3 × 1-dimensional system noise vector,

$\boldsymbol{Z}_k = [z_{\gamma k} \quad z_{\theta k} \quad z_{\psi k}]^T$ is 3 × 1-dimensional measurement vector,

$\boldsymbol{V}_k = [v_{\gamma k} \quad v_{\theta k} \quad v_{\psi k}]^T$ is 3 × 1-dimensional measurement noise vector

The variance statistical characteristics of $\boldsymbol{W}_k$ and $\boldsymbol{V}_k$ are unknown and time-varying. In state equation, $\boldsymbol{f}(\cdot)$ adopts Formula (23.3) and (23.4) to solve attitude angle. The measurement equation is written in matrix form:

$$\begin{bmatrix} z_{\gamma k} \\ z_{\theta k} \\ z_{\psi k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_k \\ \theta_k \\ \psi_k \end{bmatrix} + \begin{bmatrix} v_{\gamma k} \\ v_{\theta k} \\ v_{\psi k} \end{bmatrix} \tag{23.6}$$

where

$$\begin{aligned} z_{\theta k} &= \arcsin(-\frac{a_{xk}}{g}), \\ z_{\gamma k} &= \arctan(\frac{a_{yk}}{a_{zk}}), \end{aligned} \tag{23.7}$$

where g is the acceleration of gravity, $\boldsymbol{a}_{bk} = [a_{xk} \quad a_{yk} \quad a_{zk}]^T$ is three-axis accelerated speed which is the output of MEMS accelerometer. The output is independent of the heading angle, so the heading angle estimation is not discussed in this paper.

23.4 Design of Simplified Adaptive Kalman Filter Algorithm

The conventional Sage-Husa adaptive Kalman filter (SHAKF) algorithm can realize the online adaptive estimation of the system noise and the measurement noise in theory. But in multi-rotor UAV attitude estimation, it will be found that the algorithm does not separate and estimate accurately the system noise and the measurement noise when their statistical properties are all unknown. Therefore, a simplified adaptive Kalman filter (SAKF) algorithm is designed in this paper.

For the SAKF algorithm, in a single information fusion cycle, the complete estimation steps are:

- State single-step predicting:

$$\hat{\boldsymbol{X}}_{k/k-1} = \boldsymbol{f}(\hat{\boldsymbol{X}}_{k-1}) \tag{23.8}$$

- Innovation sequence updating:

$$\boldsymbol{v}_k = \boldsymbol{Z}_k - \boldsymbol{H}_k \hat{\boldsymbol{X}}_{k/k-1} \tag{23.9}$$

- System noise estimating:

$$\hat{\boldsymbol{Q}}_k = \text{attitude angle variances calculated by gyro} \tag{23.10}$$

- Measurement noise estimating:

$$\hat{\boldsymbol{R}}_k = \text{attitude angle variances calculated by accelerometer} \tag{23.11}$$

- Filter gain updating:

$$\boldsymbol{K}_k = \frac{\hat{\boldsymbol{Q}}_k}{\hat{\boldsymbol{Q}}_k + \hat{\boldsymbol{R}}_k}. \tag{23.12}$$

- Attitude angle state estimating:

$$\hat{\boldsymbol{X}}_k = \hat{\boldsymbol{X}}_{k/k-1} + \boldsymbol{K}_k \boldsymbol{v}_k \tag{23.13}$$

The algorithm only needs a given initial value $\boldsymbol{X}_0$, just it can get the attitude angle state estimation $\hat{\boldsymbol{X}}_k (k = 1, 2\ldots)$ at time k by the recursive calculation.

The SAKF algorithm first should solve the problems of separating the system noise and measurement noise when their statistical properties are unknown, and of estimating the high-precision variances of the two. For the multi-rotor UAV attitude estimation system, the system noise is mainly determined by the gyro, the measurement noise is mainly determined by the accelerometer, the noise parameters are

relatively stable. Therefore, the system noise variance $\boldsymbol{Q}_k$ and the measurement noise variance $\boldsymbol{R}_k$ can be estimated online by the attitude angle variance real-time dynamic calculation of the gyro and the accelerometer. The simplified algorithm can not only solve the problems of noise statistical properties unknown, but also solve the problems that the system noise variance estimation may lose the semi-positive definiteness, the measurement noise variance estimation may lose positive definiteness which may lead to filter divergence.

23.5 Flight Experiment

The experimental data are acquired from the multi-rotor UAV carrying the ADIS16365 Inertial Measurement Unit (IMU). The real-time data are collected in a special flight state of the roll motion. The sampling frequency is $f_{\mathrm{IMU}} = 165$ Hz, and experimental time is about 120 *s*.

- Raw data is sampled by the airborne IMU on the multi-rotor UAV once every other TI = 0.006 s. Every other TG = 0.012 s, the attitude information of the gyro is calculated using the quaternion method. The attitude information of the accelerometer is calculated once every other TI = 0.006 s.
- Every other TK = 0.12 s, the information fusion is done using the SAKF algorithm. When the navigation time t is the integral multiples of the TK, the gyro calculates the attitude angle as the state single-step prediction of the SAKF algorithm, the accelerometer calculates the attitude angle, processed by the average filter and median filter, as measurement information.
- The attitude angles calculated by the gyro are the inserted values within each TK cycle, which renders the update cycle of the attitude angle up to TG = 0.012 s.
- Evaluation method for the effectiveness of the algorithm: Select the integrated miniature attitude measurement system *MTi* as the standard reference system, which evaluates the effectiveness of the SAKF algorithm. The *MTi* is produced by Xsens Company in Holland, which can directly output the high-precision attitude angle. The pitch angle and the roll angle measurement accuracy of the *MTi* is $\pm 0.5^{\circ}$. In the flight experiments, the sampling frequency of the *MTi* is $f_{MTi} = 10\,\mathrm{Hz}$. The heading angles needed by the quaternion method are also provided by *MTi* in the experiment. Algorithm evaluation index are the root-mean-square error, the calculation formula is:

$$\tilde{X} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_{AKF,i} - X_{MTi,i}\right)^2} \tag{23.14}$$

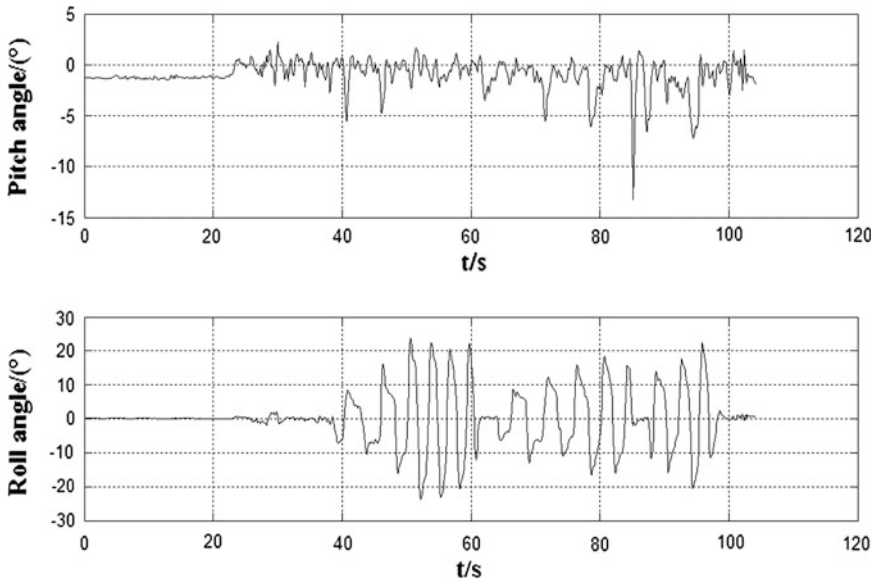


Fig. 23.1 Standard reference curves of the *MTi* pitch, and roll angles

where $X_{MTi,i}$ is the i-th compared point of the *MTi* values, $\hat{X}_{AKF,i}$ is the i-th attitude angle which is estimated by the SAKF algorithm, n is the total points compared.

The output reference curve of the *MTi* is shown in Fig. 23.1. The attitude angles curves, adopting conventional SHAKF algorithm, is shown in Fig. 23.2. The experimental results curve, which adopts the attitude angle calculation of the gyro using the quaternion method, and the estimation using the SAKF algorithm, is shown in Fig. 23.3.

The statistical properties of the system noise and the measurement noise, estimated by the conventional Sage-Husa adaptive Kalman filter (SHAKF) algorithm, have bigger error with actual values. It can make the misjudgement for the weight between the gyro and accelerometer, so the short-term accuracy of the attitude angle calculation is extremely low, as shown in Fig. 23.2. And the simplified adaptive Kalman filter (SAKF) algorithm can solve this problem very well, as shown in Fig. 23.3.

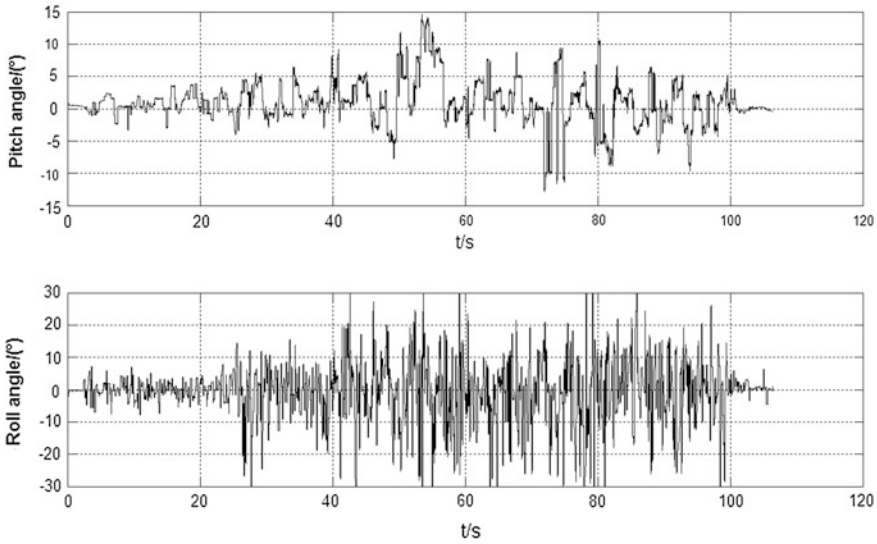


Fig. 23.2 Pitch and roll angles curves of the conventional SHAKF algorithm

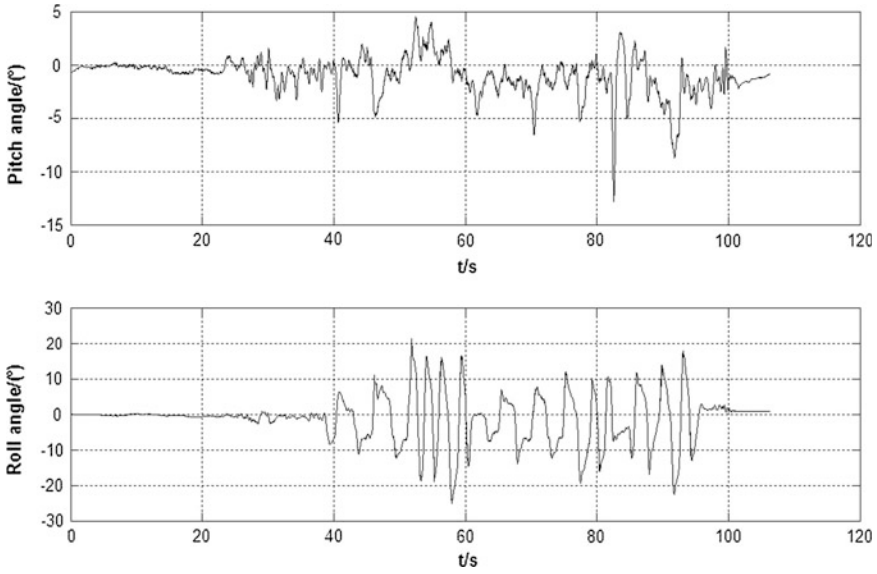


Fig. 23.3 Pitch and roll angles curves of the SAKF algorithm

23.6 Conclusion

- The SAKF algorithm can correctly judge short-term high-precision of the gyro and long-term better static performance of the accelerometer, so as to adjust adaptively the weighting coefficient of the inertial sensors, better ensure the estimation accuracy of the attitude angle.
- Comparing the attitude angle estimation by SAKF algorithm and the *MTi* standard reference system, the root-mean-square errors of the pitch angle and the roll angle are respectively: $\tilde{X}_{\text{pitch angle}} = 2.922^{\circ}, \tilde{X}_{\text{roll angle}} = 1.713^{\circ}, n = 36$.
- The SAKF algorithm can make the update frequency of the attitude angle achieving $f_{\text{SHAEKF}} = 82.5$ Hz, the attitude angle estimation is more suitable for the motion features of the high frequently dynamic change of the multi-rotor UAV.

To sum up, the SAKF algorithm is used to information fusion for the airborne MEMS gyro and the accelerometers, the estimation of attitude angle with high precision, good real-time performance, the strong stability of the filter, can meet the autonomous flight demands of the multi-rotor UAV.

Acknowledgments In this paper, the research was sponsored by the National Natural Science Foundation of China [No. 11372309 and No. 61304017], Science and Technology Development Plan Key Project of Jilin Province [No.20150204074GX], Science and Technology Special Fund Project of Provincial Academy Cooperation [No. 2014SYHZ0004].

References

1. Paw YC, Balas GJ (2011) Development and application of an integrated framework for small UAV flight control development. Mechatronics 21:789–802
2. Li D, Cheng XJ, Xu ZJ (2013) Gain adaptive sliding mode controller for flight attitude control of MAV. Opt Precis Eng 21(5):1183–1191 (in Chinese)
3. Titterton DH, Weston JL (2010) Strapdown inertial navigation technology. 2rd edn. National Defense Industry Press, Beijing (in Chinese)
4. Feng ZY, Zeng H, Zhang L et al (2011) Angle measurement based on gyroscope and accelerometer signal fusion. J SE China Normal Univ (Nat Sci Ed) 36(4):137–141 (in Chinese)
5. Gao W, Li JC (2013) Adaptive kalman filtering for the integrated SINS/DVL system. J Comput Inf Syst 9(16):6443–6450 (in Chinese)
6. Hajiyev C, Soken HE (2013) Robust adaptive kalman filter for estimation of UAV dynamics in the presence of sensor/actuator faults. Aerosp Sci Technol 28:376–383
7. Sage AP, Husa GW (1969) Adaptive filtering with unknown prior statistics. In: Proceeding of the joint automatic control conference, Tokyo, Japan, pp 760–769
8. Zhang RH, Jia HG, Chen T et al (2008) Attitude solution for strapdown inertial navigation system based on quaternion algorithm. Opt Precis Eng 16(10):1963–1970 (in Chinese)
9. Kong XY (2004) INS algorithm using quaternion model for low cost IMU. Robot Auton Syst 46:221–246 (in Chinese)

Chapter 24
Combination of Particle Swarm Optimization with LSSVM for Pipeline Defect Reconstruction

Huixuan Fu, Yuchao Wang and Sheng Liu

Abstract The nuclear function parameter and penalty parameter are pivotal factors which decide performance of Least Squares Support Vector Machines (LSSVM). Usually, most users select parameters for an LSSVM by rule of thumb, so they frequently fail to generate the optimal approaching effect for the function. In order to get optimal parameters automatically, a new approach based on particle swarm optimization and LSSVM was proposed, which automatically adjusts the parameters for LSSVM, ensuring the accuracy of parameter selection. This method was applied to pipeline 2D defect reconstruction; simulation results showed the method can overcome the difficulty of magnetic flux leakage signals, described defect geometrical characteristics, improving the reconstruction accuracy and practical value.

Keywords Pipeline · Magnetic flux leakage · 2D defect reconstruction · PSO · Least squares support vector machines

24.1 Introduction

With the rapid development of the oil and gas industry, pipeline transportation has become the main way for onshore oil and gas transport. But with the growth of pipe age, construction defects, vandalism and corrosion, and other reasons, pipeline accidents not only cause significant economic losses, but also serious environmental pollution and even endanger the lives and safety of production personnel.

Magnetic flux leakage methods are commonly used in pipeline defect detection. It uses ultrasound, magnetic flux leakage, and radiation detection principles to detect the various defects in the pipeline, without affecting the normal production conditions. Avoiding blind overhaul or repair that is not timely will save a lot of

H. Fu (✉) · Y. Wang · S. Liu
Harbin Engineering University, Nantong Street 145, Harbin, China
e-mail: fuhuixuan@hrbeu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_24

maintenance costs, resulting in significant economic and social benefits. The signal inverse problem is an important problem in magnetic flux leakage detection, which is determined from the defect length, width, depth, and other parameters of the measured signal. Inverse problem is very complex and iterative methods are commonly used approaches to solve the inverse problem. However, the iterative method has a large amount of calculation. To reduce the calculation amount, intelligent algorithms are proposed, such as Neural Networks [1, 2], improved optimization algorithm [3, 4] for solving the inverse problem.

Support vector machine (SVM) was developed by Vapnik [5]. SVM are established on the unique theory of the structural risk minimization principle. LSSVM is reformulations to the standard SVM. LSSVM was introduced by Suykens [6]. LSSVM simplifies the solution process of standard SVM to a great extent by substituting the inequality constraints by equality counterparts. Consequently, the decision function can be obtained by solving a group of linear equalities rather than quadratic programming. Particle swarm optimization (PSO) algorithm [7] proposed by Kennedy and Eberhart in 1995, has been applied to pattern recognition, signal processing and multi-objective optimization, etc. [8, 9].

A new approach based on particle swarm optimization and LSSVM was proposed. Using PSO to optimize the nuclear function parameter and penalty parameter in LSSVM ensures good performance of LSSVM. This method was applied to pipeline 2D defect reconstruction and the simulation results showed the validity for improving the reconstruction accuracy.

24.2 LSSVM Parameters to Network Performance Influence

24.2.1 Kernel Function Parameters Influence

The parameters in kernel function reflect the characteristics of training data, and greatly impact the generalization ability of the system. Taking RBF Kernel function as an example, the parameter σ^2 is the bandwidth of the RBF kernel. σ^2 and has great impact on the accuracy of the prediction. The value of σ^2 is too small ($\sigma^2 \to 0$) or too big ($\sigma^2 \to \infty$) will lead to very poor generalization of LSSVM.

24.2.2 Penalty Parameter γ Influence

The parameter γ determines the trade-offs between the minimization of the fitting error and the minimization of the model complexity. However, determining their values exactly is difficult for LSSVM.

24.3 PSO-LSSVM Algorithm

PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles.

Particle swarm optimization algorithm uses an information sharing mechanism. A group of n particles fly at a certain speed in the D-dimensional search space. Each particle in the search process takes into account its own search history and the best point within the group of other particles, and position changes to be based on this. Particle position and velocity changes according to the following equation in the PSO algorithm:

$$\begin{cases} v_i(t+1) = wv_i(t) + c_1 \,\text{rand}\,(p_i(t) - x_i(t)) + c_2 \,\text{rand}(\text{g}(t) - x_i(t)) \\ x_i(t+1) = x_i(t) + v_i(t+1) \end{cases} \tag{24.1}$$

where x_i is the position vector of the i-th particle, v_i is the speed vector. p_i is the best position of i-th particle during its search process, and g as the whole particle swarm's best position during the current search. c_1 and c_2 are called learning factor, which makes particles have the function of self-summary and learn to the best of the swarm, get close to the best position on its own as well as within the swarm. rand is the random number distributed in [0, 1]. Velocity of each particle is limited to the maximum speed range $v_{\max}$. w is the inertial coefficient, which determines the current speed of the particle inherited. It directly affects the global and local search ability of particle. A number of experiments have shown that the value of w in the range of [0.9, 0.2] in the algorithm will have better performance, and the value of w linear decrease is better than use fixed values.

24.3.1 Algorithm Design

24.3.1.1 Penalty Particle Dimension

Using particle swarm algorithm to optimize parameter σ^2 and γ for LSSVM improves the performance of LSSVM, and therefore particle vector dimension is 2.

24.3.1.2 Fitness Function

In order to make the square of the error minimum between LSSVM output and objective function, define $F(\sigma, \gamma)$ as the fitness function. The calculation of $F(\sigma, \gamma)$ takes the following form:

$$F(\sigma, \gamma) = \frac{1}{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 + e} \tag{24.2}$$

where σ^2 is the bandwidth of the radial basis function kernel, γ is penalty parameter, e is the small number of avoiding the denominator gets zero, x_i and $\hat{x}_i$, respectively, represent the actual and forecast values, n is the number of sampling.

24.3.2 Algorithm Realization

Particle swarm optimization is applied to the LSSVM optimization, the basic steps of the algorithm are as follows:

Step 1: Initialization parameters set the population size, inertia weight w, learning factor c_1c_2, the number of iterations, determine the particle dimension, random initialization of the initial position and velocity of particles, set inertia weight w, iterative times $t = 1$,

Step 2: For each particle in the population, according to formula (24.2) calculate the particle fitness function value F, each particle current fitness value compare with its history optimal value p_{best}, if the current value is better than p_{best}, update p_{best}, otherwise, p_{best} remains unchanged,

Step 3: Take each particle p_{best} compare with the value of the global optimal particle swarm $g_{\underline{\text{best}}}$, if the current value is better than g_{best}, update g_{best}, otherwise, g_{best} remains unchanged,

Step 4: According to formula (24.1), update the speed of evolution and location of each particle,

Step 5: $t = t + 1$,

Step 6: Termination judgments, if the number of generation equals a threshold or $\frac{1}{F(\sigma,\gamma)}$ is less than or equal to an positive less number, go to step 7, otherwise go back to Step 2.

Step 7: According to the above procedure, the optimal combination of kernels function parameter σ^2 and penalty constant γ in LSSVM is obtained.

24.4 Experiment Result

Take the data obtained from actual pipeline defects measured values as training samples. Take the Magnetic flux leakage signal as input, the defect geometric parameters(length and depth) as output, use the LSSVM and PSO-LSSVM, make the network realize given mapping relation. Sampling points for each set of data are 60. Select 90 sets of data as sample data, the foregoing 80 sets of data are used as the training data, the latter 10 sets of data are used as the test data.

Particle swarm parameters settings are as follows: particle vector dimension is 2, the number of individuals in population is 20, learning factor $c_1 = c_2 = 2$, the inertia weight w with iterative linear decrease from 0.9 to 0.2, maximum number of iterations is 100. The σ^2 and γ separately in [0, 32] and [1, 2048] initialization.

Mean absolute error (MAE) is used as the evaluation indicator to measure the performance, MAE is defined as follows:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left| \mathrm{x_i} - \mathrm{x_i'} \right| \quad (\mathrm{i} = 1, 2, \ldots, \mathrm{n}) \tag{24.3}$$

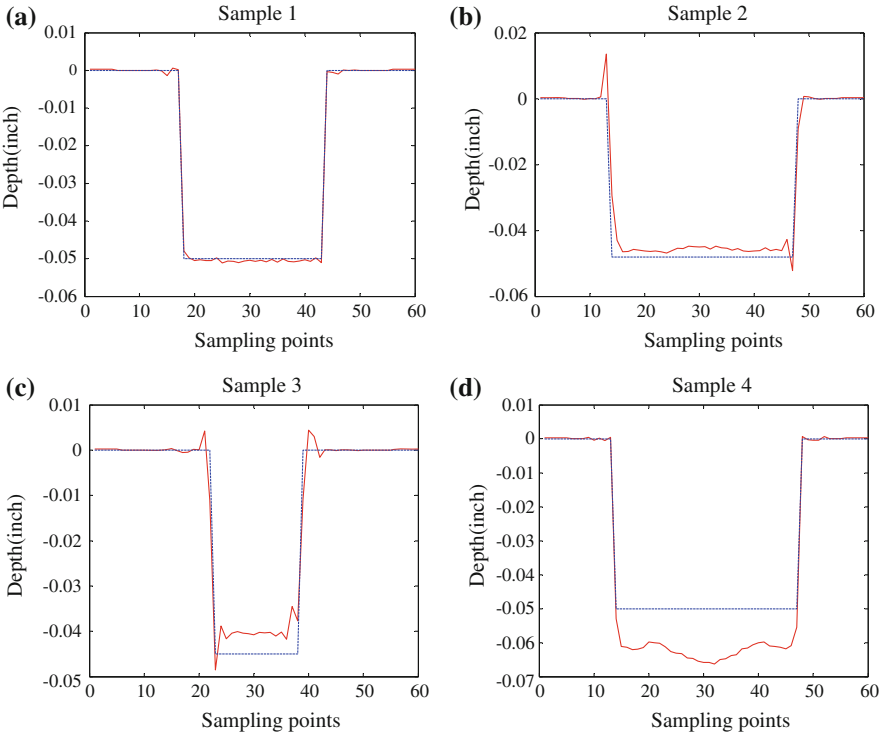


Fig. 24.1 Pipeline 2D defect reconstruction based on LSSVM

where x_i was real value of test data, $x_i^{'}$ was the reconstruction value of test data, n is the number of samples.

Due to limited space, this paper gives only four sets of test data reconstruction results. The LSSVM simulation of reconstruction results are shown in Fig. 24.1.

Figure 24.1 shows the four groups' sample actual pipeline defect value and reconstruction value after LSSVM training comparison. The blue dashed line stands for the real value of the pipeline defect, and the red solid line stands for the LSSVM reconstruction value of the pipeline defect. X axis is the sampling points; Y axis represents the defect depth, in inches. From Fig. 24.1 we can see that LSSVM cannot accurately reflect the geometric characteristics of defects, with big errors.

The PSO-LSSVM simulation of reconstruction results are shown in Fig. 24.2.

Figure 24.2 shows the results simulated by PSO-LSSVM algorithm with the same four sets of samples. The blue dashed line stands for the real value of the pipeline defect, and the red solid line stands for the LSSVM reconstruction value of the pipeline defect. X axis is the sampling points; Y axis represents the defect depth, in inches (inch). As can be seen from Figs. 24.1 and 24.2, PSO-LSSVM method can get more accurate fitting and reconstruction value than the LSSVM method. When the pipeline defect value changed frequently, PSO-LSSVM can get better

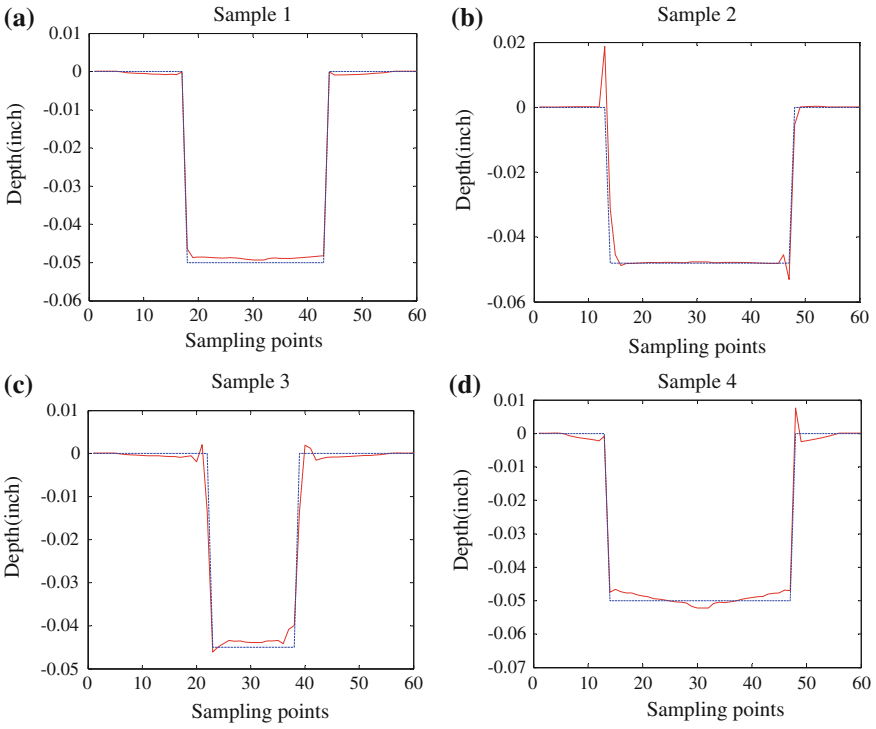


Fig. 24.2 Pipeline 2D defect reconstruction based on PSO-LSSVM

Table 24.1 Two methods errors and parameters selection results

Methods	MAE	γ	σ_2
LSSVM	0.2248	500	5
PSO-LSSVM	0.1285	1466	30

reconstruction results. This is due to the fact that the nuclear function parameter and penalty parameter can decide performance of LSSVM. By particle swarm optimization, LSSVM obtain the best parameters. Therefore, achieve a more satisfactory reconstruction results.

The MAE errors and LSSVM parameters selection results are given in Table 24.1.

From Table 24.1, σ^2 is kernel function parameter, γ is penalty parameter, error (MAE) of reconstruction based on LSSVM is 0.2248, error(MAE) of reconstruction based on PSO-LSSVM is 0.1285. As can be seen from Table 24.1, PSO-LSSVM method has high reconstruction accuracy than LSSVM method.

24.5 Conclusion

A method for pipeline 2D defect reconstruction based on PSO-LSSVM was presented. The results reported in this paper show clearly that PSO-LSSVM has better reconstruction properties than LSSVM. When the sample data have great oscillation, the errors of PSO-LSSVM reconstruction results are smaller than the LSSVM. The PSO-LSSVM is able to establish reconstruction of Pipeline 2D defect effectively with simple structure and good generalization ability. Therefore, this approach could be an alternative method for Pipeline 2D defect reconstruction.

Acknowledgments This work is supported by the National Natural Science Foundation of China (Grant No. 51409062, 51279036), and the Fundamental Research Funds for the Central Universities (HEUCF041530).

References

1. Tikaria M, Nema S (2010) Wavelet neural network based intelligent system for oil pipeline defect characterization. In: 3rd international conference on emerging trends in engineering and technology, pp 42–47
2. Chen J, Li L, Xu B (2008) Magnetic flux leakage testing method for well casing based on gaussian kernel RBF neural network. Adv Comput Theor Eng 228–231
3. Han W, Que P (2005) 2-D defect reconstruction from MFL signals based on genetic optimization algorithm. In: IEEE international conference on industrial technology, pp 208–513
4. Han W, Yang P, Ren H (2011) Application of damping-boundary-based PSO to MFL signal inversion. In: 2011 third international conference on measuring technology and mechatronics automation, vol 1, pp 532–525
5. Vapnik V (1995) The nature of statistical learning theory. Springer, New York

6. Suykens JAK, Vandewalle J (1999) Least-squares support vector machines classifiers. Neural Process Lett 9:293–300
7. Eberhart R, Kennedy J (1995) A new optimizer using particles swarm theory. In: Proceedings of the international symposium on micromechatronics and human science, pp 39–43
8. Zeng Yanyang, Kang Fengju, Hao Xu (2013) Improved PSO algorithm for multi-objective optimization based on grey relational degree. ICIC Express Lett 7:601–607
9. Zhu Y, Huang C (2012) Image denoising algorithm based on PSO optimizing structuring element. Control Decis Conf 2404–2408

Chapter 25
A Face Replacement System Based on 3D Face Model

Hong Song, Jie Lv, He Liu and Qingjie Zhao

Abstract In this paper, an efficient face replacement system based on 3D model is developed, which consists of four parts. First, given an input video containing a source face rotating from frontal to profile, after two manual clicks on a profile to tell the system where the eye corner and bottom of the chin are, a realistic looking source face 3D model is generated. Secondly, Random Forests Algorithm (RFA) is used to estimate target face pose. Thirdly, a source face image is rendered based on the result of pose estimation and color transfer method. Finally, Poisson image fusion algorithm for boundary blurring is used to seamlessly synthesize the source face into the target image. The effectiveness of our system is demonstrated with a set of experiments.

Keywords Face replacement · Face modeling · Pose estimation · Color transfer · Poisson image fusion

25.1 Introduction

Face replacement has extensive application prospects in filmmaking, games entertainment, medical field, etc. Many researchers have obtained important results in recent decades. However, most digital processing systems perform face replacement under similar pose, expression, and illumination.

Bitouk et al. proposed a system for automatic face replacement in images [1], which needed the candidate face image that is similar to the input face in appearance and pose. Liang et al. presented a practical system for face replacement in videos [2]. Facial expressions of each detected face is analyzed to select candidate

H. Song (✉) · H. Liu
School of Software, Beijing Institute of Technology, Beijing 100081, China
e-mail: anniesun@bit.edu.cn

J. Lv · Q. Zhao
School of Computer Science, Beijing Institute of Technology, Beijing 100081, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_25

face images from source video that are most similar to the target face in pose and expression. Then candidate replacements are blended to target video.

Cheng et al. presented a system for face replacement in videos [3], which used the source image to reconstruct the 3D face model and replaced the faces by using face alignment algorithm, a lighting and pose estimator, and a composition procedure. Their system can perform face replacement under dissimilar pose, but the face modeling method is affected in uncontrolled illumination. Lin et al. also proposed a system for face replacement based on personalized 3D face models [4]. A personalized 3D face model is used to render a facial image to the same pose as the target image. Color transfer and multi-resolution methods for boundary blurring are used to seamlessly synthesize the face into the image. However, in the process of 3D face modeling, the feature points on the profile are detected automatically using Local Maximum-Curvature Tracking (LMCT), which works very well only for Mongoloid looking people.

A face replacement system based on 3D face model is proposed in this paper, which can perform face replacement under dissimilar pose and color efficiently. A source face 3D model is generated by the proposed method that consists of extracting 2D facial feature points, deforming generic face model, $\sqrt{3}$-Subdivision and texture mapping. Then the 3D face model is adjusted according to the target face pose estimated by using Random Forests Algorithm (RFA). The source face color is adjusted based on target face using color transfer method. Finally, Poisson image fusion algorithm is used to synthesize the source face into the target image. The face replacement system framework mainly consists of three parts, as shown in Fig. 25.1, including generating source face 3D model, rendering source face image based on the result of pose estimation and color transfer method, and Poisson Fusion. There are two key contributions of this paper based on the proposed overall technique: (1) $\sqrt{3}$-Subdivision algorithm is used to refine the individual 3D geometric face model. (2) The source face is synthesized into the target image using Poisson image fusion algorithm.

25.2 Generating Source Face 3D Model

In this paper, the face modeling method consists of four components. First, given an input video, frontal and profile facial feature points are extracted. Secondly, based on the facial feature points, an individual 3D geometric face model is generated by deforming the generic model with improved Radial basis function. Then the model is refined by using improved $\sqrt{3}$-Subdivision. Finally, source face 3D model is built by mapping the texture to the 3D geometric face model.

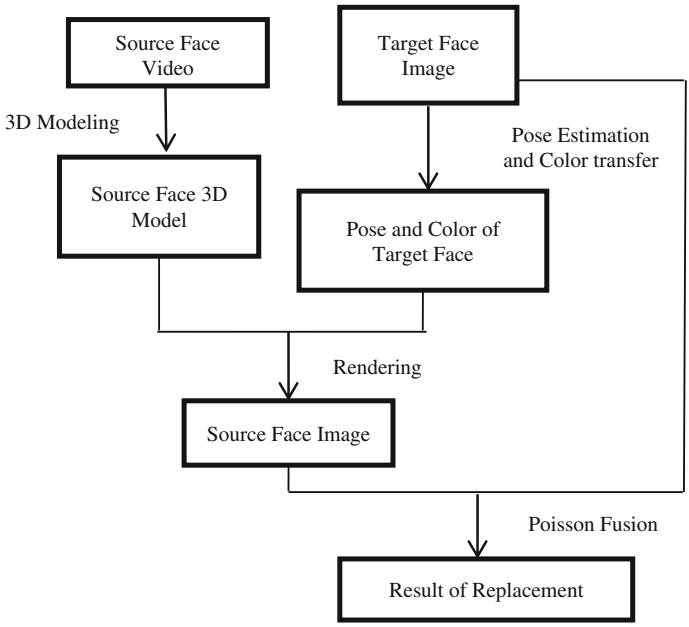


Fig. 25.1 Overview of the face replacement system

25.2.1 Extracting 2D Facial Feature Points

In this paper, Active Shape Model (ASM) is used to extract 68 facial feature points on the frontal image [4]. The training data sets are provided by Song et al. [5]. Firstly, the model [4] is built based on the training data sets. Secondly, for a new image, the detection result is fitted by solving the model parameters. The extracted feature points are shown in Fig. 25.2a.

The profile facial feature points are important for the subsequent steps. First, improved frame difference method that fuses the last few frame difference results is used to detect profile face region [6]. Then, canny algorithm is applied to extract a facial silhouette on the profile face region. Finally, the other 5 points on the profile face are extracted based on two manually labeled points and profile facial silhouette. The result is shown in Fig. 25.2b.

25.2.2 Deforming Generic Face Model

In this paper, Candide-3 model is used as a generic model to decrease the complexity of modeling. The points on Candide-3 model are compared with frontal facial feature points shown in Fig. 25.2a using orthogonal projection method; there

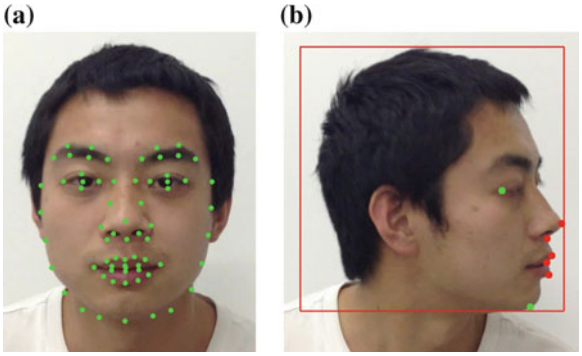


Fig. 25.2 **a** Extracted 68 facial feature points. **b** Profile facial feature points (two manually labeled points indicated by *green dots*, five automatically detected points indicated by *red dots*, face region indicated by *red rectangle*) (color online)

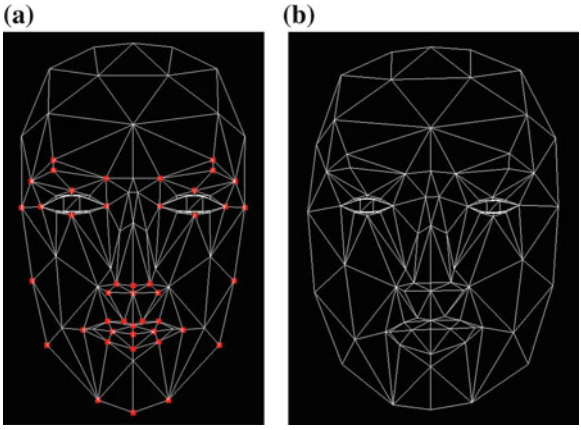


Fig. 25.3 **a** Candide-3 face model (feature points on Candide-3 model indicated by *red dots*). **b** Result of deformation with RBF (color online)

are 44 points as their correspondences (eyes, nose, mouth, and facial silhouette etc.), which is defined by comparing their semantic points. The result is shown in Fig. 25.3a.

Radial Basis Function (RBF) (see (25.1)) is a multivariate interpolation function of scattered data [7]. The points on the face model are distributed unevenly. Furthermore, the number of points on the frontal face is more than that on the profile face. In this paper, Sub-regional Multi-step RBF method is applied to deform the generic model [8]. The result of deformation is shown in Fig. 25.3b.

$$f(x) = \sum_{i=1}^{n} c_i g(\|x - x_i\|) + Mx + t, x \in R^d \tag{25.1}$$

where c_i is weight, $g(\|.\|)$ is a kernel function, x is a scattered dataset, x_i is center of kernel function, M is a coefficient matrix, and t is a constant matrix.

25.2.3 Subdivision

Due to sparse facial structure, the slower increase in the mesh complexity and the suitability for adaptive refinement, $\sqrt{3}$-Subdivision is applied to refine the individual 3D geometric face model (see Fig. 25.3b) in this paper [9]. $\sqrt{3}$-Subdivision consists of two smoothing rules shown in (25.2) and (25.3), one for the placement of the newly inserted vertices and the other for the relaxation of the old ones.

$$V_F = \frac{1}{3}(V_0 + V_1 + V_2) \tag{25.2}$$

where V_0, V_1 and V_2 are three vertices of a triangle, and the new vertex V_F is simply inserted at the center of the triangle.

$$V_v = (1 - \alpha_n)V + \frac{\alpha_n}{n}\sum_{i=0}^{i-1} V_i \tag{25.3}$$

where $\alpha_n = \frac{1}{9}(4 - 2\cos\frac{2\pi}{n})$, $V_0, V_1, \ldots, V_{n-1}$ are directly adjacent neighbors of vertex V in the unrefined mesh.

Because the original $\sqrt{3}$-Subdivision makes model volume become smaller, the improved algorithm for $\sqrt{3}$-Subdivision proposed in this paper is inspired by [9]. The vertex V_v shown in (25.3) after several subdivision iterations is adjusted to original position. The vertex V_F shown in (25.2) is also adjusted to the new position (see (25.4)), which makes the mesh keep its original shape. It is assumed that the three vertices of a triangle for original model are denoted by (V_0, V_1, V_2) respectively, they are denoted by (V_0', V_1', V_2') after 1 subdivision iteration respectively. The results of using improved $\sqrt{3}$-Subdivision method are shown in Fig. 25.4.

$$V_F^{'} = V_F + \frac{d_1}{d_1 + d_2 + d_3}S_1 + \frac{d_2}{d_1 + d_2 + d_3}S_2 + \frac{d_3}{d_1 + d_2 + d_3}S_3 \tag{25.4}$$

where V_F' is new position of V_F, d_1 is the distance between V_F and V_1, d_2 is the distance between V_F and V_2, d_3 is the distance between V_F and V_3, S_1 is the offset between V_1 and V_1', S_2 is the offset between V_2 and V_2', and S_3 is the offset between V_3 and V_3'.

25.2.4 Texture Mapping

The frontal and the profile face images are used to synthesize the individual facial texture image by combining the multi-resolution technique with weighted smoothing algorithm [10].

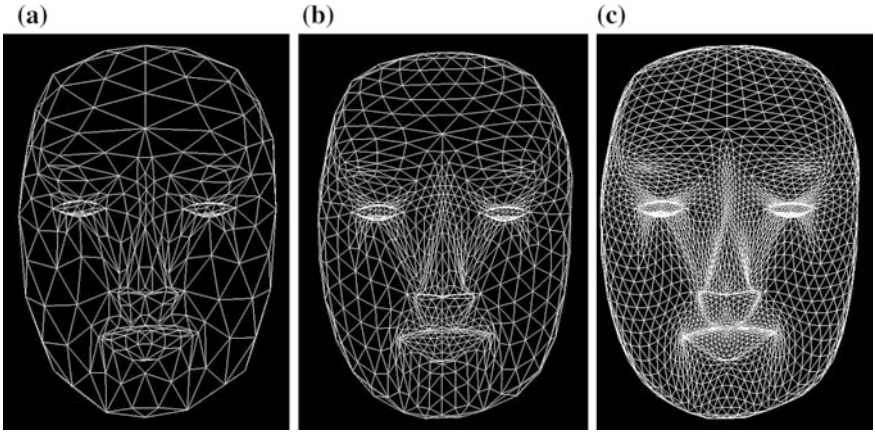


Fig. 25.4 Individual 3D geometric face model using improved $\sqrt{3}$-subdivision method. **a** After 1 subdivision iteration. **b** After 2 subdivision iterations. **c** After 3 subdivision iterations

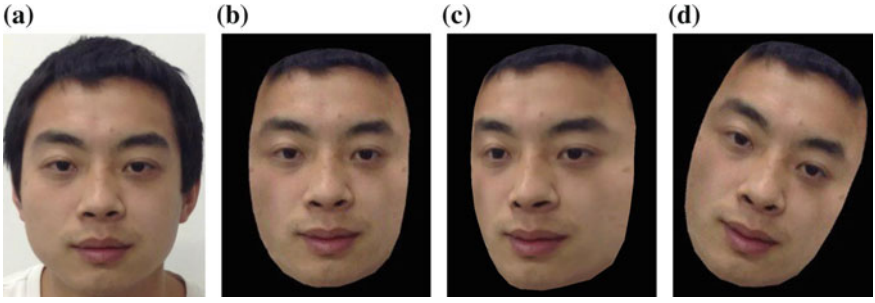


Fig. 25.5 Frontal face image and 3D face model in different views

In Sect. 25.2, some correspondences (eyes, nose, mouth, and facial silhouette) are established, while other correspondences are established using linear interpolation method. The result of texture mapping is shown in Fig. 25.5.

25.3 Rendering Source Face Image

The method of rendering source face image mainly comprises two steps, specific method is illustrated by the following parts:

Step 1: Estimating target face pose:
The target face pose is estimated by manual method or the method in [11]. In [11], Haar-like features are first extracted and tree-structured cascaded Adaboost classifiers are trained and applied to localize the face.

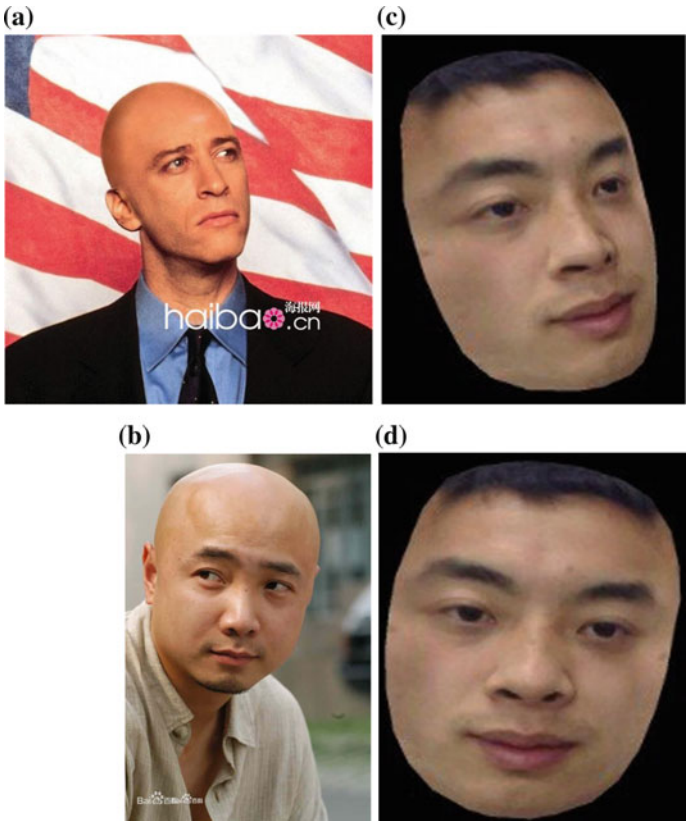


Fig. 25.6 Source face 3D model and target face image with similar poses. **a**, **b** Target face image. **c**, **d** Source face 3D model

Then the positions and scales of the whole face region are estimated from the detected results. Finally, based on this, randomized regression trees are learned to obtain the estimation results of head orientation.

Step 2: Rendering source face image

First, a source face 3D model is adjusted to generate source face image under similar pose with target face pose, as shown in Fig. 25.6. Secondly, the source face image is scaled to fit the target face image based on three feature points (two centers of eyes and center of mouth). Finally, colors between source and target faces become similar by using color transfer method based on Lαβ space [12], as shown in Fig. 25.7.

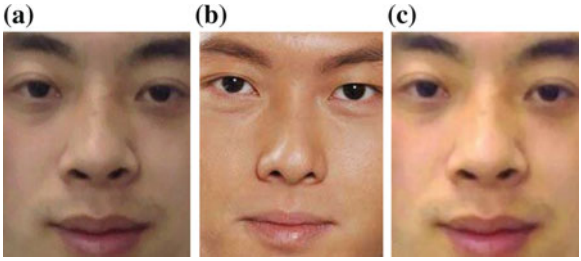


Fig. 25.7 **a** Source face image. **b** Target face image. **c** Result of color transfer method

25.4 Poisson Fusion

Poisson fusion for boundary blurring is applied to seamlessly synthesize the source face into the target image [13], specific algorithm is illustrated by the following parts:

First, replace regions of the source and target face image are selected. Then the region of the source face image is copied into the target directly. Secondly, the divergence value of guidance vector field is calculated. Thirdly, Poisson equation is created and solved by putting into the value of image channel. Finally, Laplace interpolation is used in interpolation region.

25.5 Experiments and Evaluation

The source videos and target image with 1024 × 768 were captured as our input data. All the experiments were performed on a quad 3.40 GHz core i7-2600 CPU and 2 GB RAM machine, some experimental results are shown in Fig. 25.8.

Candide-3 model has only 113 points, which can decrease deforming running time, but the final 3D face model will lose a number of details especially in boundaries [14, 15]. Improved $\sqrt{3}$-Subdivision is used to make the model become dense (see Fig. 25.4). As shown in Fig. 25.8, it indicates that the process of face replacement can run successfully by using our method, when the source and target images have dissimilar pose and color. However, in [1, 3, 2], their systems perform face replacement only when the poses for the source and target faces are similar.

Twenty people were selected to evaluate the face replacement system. The scores range on a scale from one to ten in this paper. The one point represents very badly, while ten points represents very well. The specific evaluation criteria mainly contain the boundary of replacement area, the whole face pixels mutation situation, and the level of the whole image fluency. Everyone gives high scores (more than seven) in respect of the whole face pixels mutation, while they give lower scores (less than or equal to seven) in other respects. Because some parts (eye corner, etc.) of the 3D models at large angles become distorted, so that source face image cannot be rendered very well.

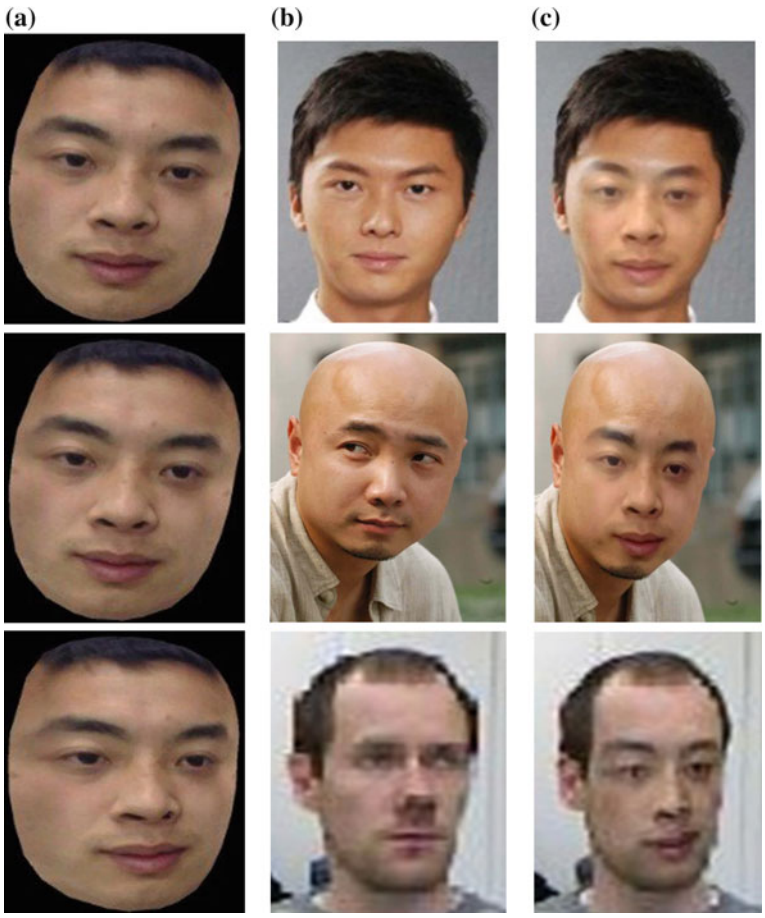


Fig. 25.8 **a** 3D source face models which have similar pose with target face images. **b** Target face images. **c** Results of face replacement

25.6 Conclusions

In this paper, an efficient face replacement system based on 3D face model is proposed, which is easy to use with minimal user intervention. A set of experiments were implemented to show the effectiveness of the system. That can be used for personal entertainment and medical beauty etc. In the future study, we will try to synthesize individual facial texture image by using face image sequences in different views, and establish illumination model to improve the face replacement.

Acknowledgments Project supported by the National Natural Science Foundation of China (Grant No.61240010 No.61175096).

References

1. Bitouk D, Kumar N, Dhillon S et al (2008) Face swapping: automatically replacing faces in photographs. ACM Trans Graph 27(3):39
2. Liang Y, Chen BY, Chuang YY et al (2009) Image based face replacement in video. Master's thesis, CSEI Department, National Taiwan University
3. Cheng YT, Tzeng V, Chuang YY et al (2009) 3D morphable model based face replacement in video. In: Proceedings of ACM SIGGRAPH
4. Lin Y, Gui LY, Wang SJ et al (2012) Face swapping based on 3D photo realistic head reconstruction. Tsinghua Sci Technol 52(5):602–607
5. Song H, Huang X, Wang S (2013) Automatic generation of portraits with multiple expressions. Acta Electronica Sin 41(8):1494–1499
6. Yin L, Basu A (1999) Integrating active face tracking with model based coding. Pattern Recogn Lett 20(6):651–657
7. Liu S, Wang CCL (2012) Quasi-interpolation for surface reconstruction from scattered data with radial basis function. Comput Aided Geom Des 29(7):435–447
8. Du P, Xu D, Liu C (2003) Research of individual 3D face model and its application. J Shanghai Jiaotong Univ 37(3):435–439
9. Huang J, Schroder P (2012) $\sqrt{3}$-based 1-form subdivison. Springer, Berlin, pp 351–368
10. Zhang C, Burt PJ, van der Wal GS (2013) Multi-scale multi-camera adaptive fusion with constrast normalization. US Patent 8,411,938
11. Li Y, Wang S, Ding X (2012) Person-independent head pose estimation based on random forest regression In: 2010 17th IEEE international conference on image processing (ICIP), IEEE, pp 1521–1524
12. Reinhard E, Ashikhmin M, Gooch B et al (2001) Color transfer between images. IEEE Comput Graph Appl 21(5):34–41
13. Pérez P, Gangnet M, Blake A (2003) Poisson image editing. ACM Trans Graph 22(3):313–318
14. Fan X, Peng Q, Zhong M (2010) 3D face reconstruction from single 2D image based on robust facial feature points extraction and generic wire frame model. Int Conf Commun Mob Comput 3:396–400
15. Huang J, Su Z, Wang R (2012) 3D Face Reconstruction based on improved CANDIDE-3 model. In: 2012 fourth international conference on digital home (ICDH), IEEE, pp 438–442

Chapter 26
Head Detection Using Extreme Learning Machine

Changliang Sun and Yuanlong Yu

Abstract It is difficult for most object detection systems to deal with the nonrigid objects since the descriptions of these objects are diverse. This paper proposes a new vision-based head detection method by using extreme learning machine (ELM). ELM is an efficient learning algorithm for generalized single hidden layer feedforward networks. This proposed method employs the histograms of oriented gradients (HOG) as features to describe the head objects. In order to improve the accuracy, HSV color features are also included. This proposed method is tested on PASCAL datasets. Experimental results have demonstrated the detection performance and efficiency of this proposed ELM-based head detection method.

Keywords Human head detection · Extreme learning machine · Histograms of oriented gradients · PASCAL database

26.1 Introduction

Object detection is a fundamental task of computer vision. Because of its wide applications in industry, video surveillance, and intelligent transportation, many researchers are engaged in this area. Object detection involves a series of disciplines including pattern recognition, image processing, and artificial intelligence.

This paper is supported by National Natural Science Foundation of China (NSFC) under grants 61105102 and 61473089.

C. Sun (✉) · Y. Yu (✉)
College of Mathematics and Computer Science, Fuzhou University,
Fuzhou 350116, Fujian, China
e-mail: suncl12345@sina.com

Y. Yu
e-mail: yu.yuanlong@fzu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_26

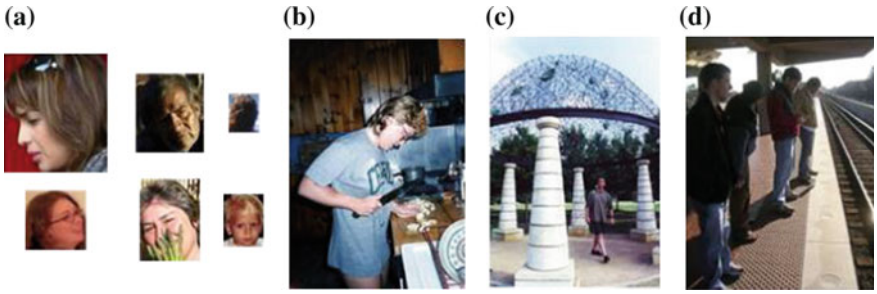


Fig. 26.1 **a** are some pictures in training set. **b**, **c**, **d** are selected from the testing set

This paper focuses on human head detection. Head detection can also be viewed as a classification problem in a two-class case (i.e., targets and background). There are many factors which can affect the detection performance; e.g., lighting conditions, background clutters, scaling, and so on. But the most challenging issue is about the variations of head appearance. Changes of viewing angles, target posture, or occlusion would lead to various appearances of heads, as shown in Fig. 26.1. These characteristics undoubtedly increase the difficulty of head detection.

Existing object detection methods can be divided into two categories: Classification based and template match based. The classification-based methods identify the target from background by training a classifier based on training samples. Support Vector Machine (SVM), neural network, and cascade of these basic classifiers are mostly used for this type of methods. But most of these methods are computationally expensive for training. Unlike classification-based methods, the template match-based method first obtain templates for the target from the training samples and then calculate the similarity between each region of interest (ROI) and templates by using some operations such as dot product. But it is difficult to obtain satisfactory templates to accommodate various appearance of the target.

This paper proposes a new method for head detection by using extreme learning machine (ELM) as basic classifiers. This method includes three steps: Feature extraction, ROI search, determining whether the window is a head area. This proposed method uses HOG as feature, which is widely used in object detection. It is invariant to geometrical and optical changes. ELM works for the "generalized" single hidden layer feedforward networks (SLFNs) while the hidden layer (or called feature mapping) in ELM need not be tuned. ELM has shown not only fast training speed but also better classification performance. In order to improve accuracy, the HSV color histogram is also integrated into feature representations.

The reminder of this paper is organized as follows. Section 26.2 presents the related work. Section 26.3 introduces the structure of this proposed system. The experimental results are shown in Sect. 26.4.

26.2 Related Work

As mentioned, classifiers play an important role in object detection. Currently, the main classifiers used in object detection are as follows: SVM, neural network-based, cascade of boosted classifiers, ELM, etc. SVM method is a very famous classifier in a two-class case. It is widely used in some areas and the high accuracy has been achieved. The most famous one is pedestrian detection using HOG features [1, 2]. Neural network-based generally uses error back propagation (BP) approach. This algorithm is trained through multiple iterations. [3–5] use this algorithm as the classifier. The character of cascade of boosted classifiers [6] is that use multiple samples to train classifiers from weak to strong. In recent years, the weak classifiers mainly include: simple binary decisions [7], single perceptron [8], and seminaive Bayesian classifier [9]. The most famous strong classifier is AdaBoost, this method is used to detect face in [10]. ELM works for the "generalized" SLFNs but the hidden layer (or called feature mapping) in ELM need not be tuned. This algorithm is proposed by Guang-Bin Huang [11, 12]. It has been successfully used in many applications such as image processing (image quality assessment [13], image super-resolution [14]), human action recognition [15], face recognition [16], and many other fields.

Feature extraction is also very important in computer vision. HOG is typically invariant in illumination changes and small deformations. The most famous one is pedestrian detection [1]. The original Haar-like feature has 3- and 4-rectangle features, it is always used for face recognition [17]. SRF is a Haar-like feature but includes 2-rectangle features [10].

26.3 System Architecture

26.3.1 Structure

Our system mainly consists of two parts: training and detection. For the training part, we extract HOG features from the training set and put the feature vectors into ELM classifier. Moreover, our positive samples were selected from the training data and the negative samples were captured from any source except the validation data. Figure 26.2 shows the training part.

The processes of our system include searching the ROI and judging whether the ROI is an object area. The sliding window approach is used to search over the whole image. ROI can be defined as windows and sliding from the initial position to the final position of the test image. The sliding steps are $(\Delta x, \Delta y)$. Moreover, the scale of head area in test image is variable, so scaling the image is needed. The scale step is Δs. The sliding window searches over the image pyramid. In our system, the sliding steps are (4, 4) and the scale step is 1.2.

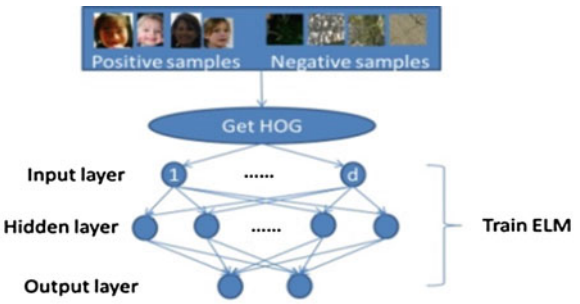


Fig. 26.2 The structure of training ELM classifier

To judge whether the window is a head area, we use ELM classifier which has been trained by training data. Because of the complicated background, we need to train multiple classifiers to filter the background. In the first part, we extract HOG features of the ROI and judged them by HOG classifiers. In the second part, we extract color features and judged by HSV classifiers. Figure 26.3 shows the system structure.

26.3.2 Details of the System

26.3.2.1 Parameters of the Features

The parameters of HOG feature are set as follows: window size is 30-by-35; block size is 10-by-10; block stride is 5-by-5; cell size is 5-by-5; orientation is quantified to 8 bins. The dimensions of HOG are 960.

In the HSV space, the range of hue ("H") is between 0 and 360°. The range of saturation ("S") is between 0 and 1. The range of value ("V") is between 0 and 255. We get the histograms from each block in hue and saturation (The value is greatly influenced by the light). There are 8 bins in each histogram. The dimensions of HSV histogram are 480.

26.3.2.2 Training Classifiers

The key parameter of ELM is the number of hidden nodes. Through the multiple experiment, we find that the closer between the number of training samples and the hidden nodes, the lower test accuracy. In Fig. 26.4, the number of training samples is 1200.

In our system, the number of the training samples is between 1000 and 1500. We choose the best hidden nodes from 100 to 2500.

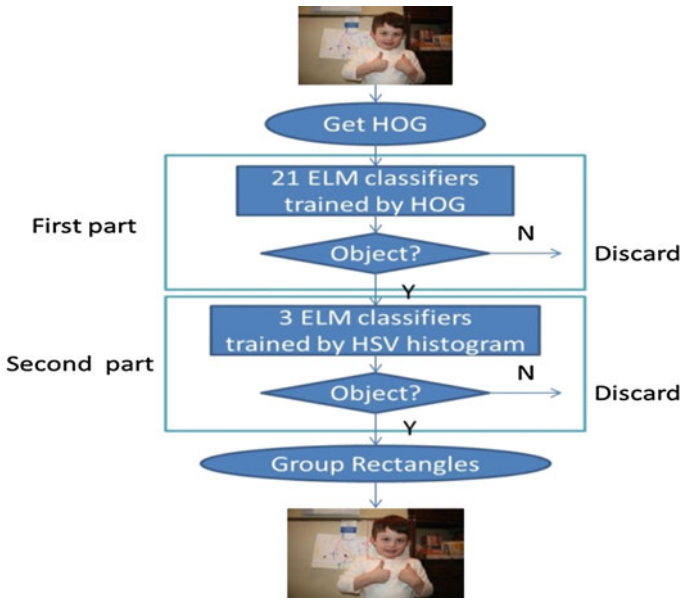


Fig. 26.3 In the first part, classifiers are trained by HOG. The rest of the candidate windows are judged by the second part. In this part, the classifiers are trained by HSV histograms

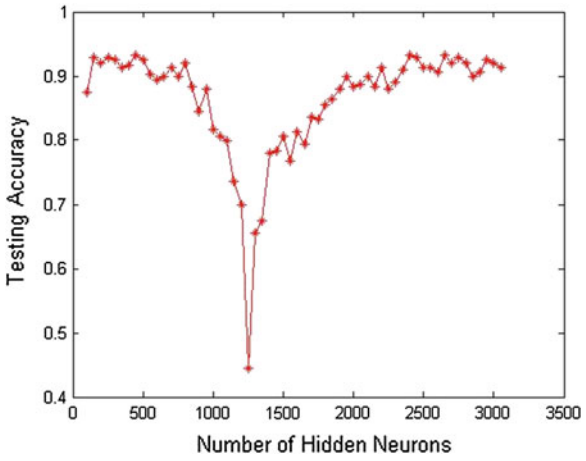


Fig. 26.4 The number of the hidden neurons is from 50 to 3050. The lowest test accuracy is 0.4355 when the number of hidden neurons is 1250

26.3.2.3 Optimization of Training Data

Multiple classifiers are needed in our system. We selected 25 groups of training data from the original datasets and picked up 21 best groups of it. These groups were used to train HOG classifiers. For color features, we picked up 3 best from 4 groups.

26.4 Experimental Results

26.4.1 Database

We evaluated our system on the PASCAL VOC challenge 2012 dataset. The goal of this challenge is to recognize objects from a number of visual object classes in realistic scenes [18]. There are 20 object classes which include: person, bird, cat, etc. but not head for object detection. Our dataset comes from the person layout challenge. In each object class, there are three sets of images provided: train, val, and trainval. In the training data, there are 324 head samples. To enrich the training set, we intercepted 506 head images from the person class. In the validation data, there are 296 test images and each image has one or more head area.

26.4.2 Evaluation

Detections are considered true or false positives based on the area of overlap with ground truth bounding boxes. To be considered a correct detection, the area of overlap a_0 between the predicted bounding box Bp and ground truth bounding box Bgt must exceed 50 % by the formula [18]:

$$a_0 = \frac{\text{area}(Bp \cap B\text{gt})}{\text{area}(Bp \cup B\text{gt})}$$

In order to evaluate our system intuitively, precision and recall are calculated to obtain a precision–recall curve in the test set.

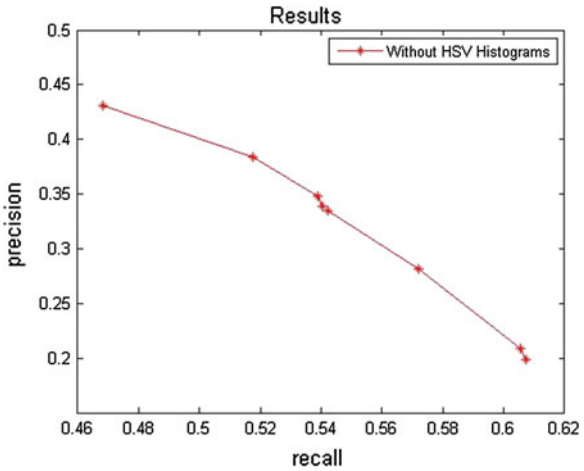


Fig. 26.5 The precision–recall curve after 21 classifiers detect

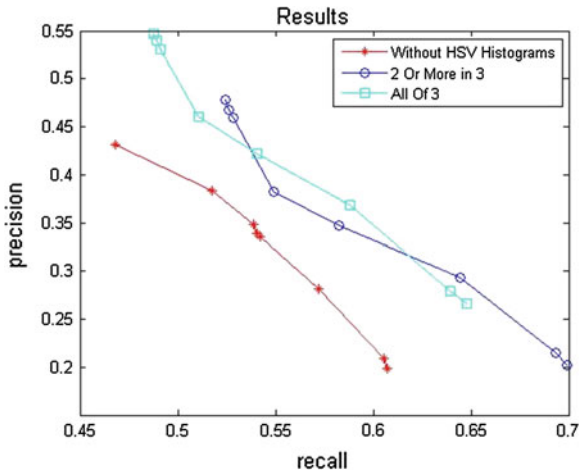


Fig. 26.6 "Without HSV Histogram" is the *curve* that uses HOG features only. "2 or more in 3" is that more than two *color* classifiers determined the window is head area. "All of 3" has the similar meaning

Fig. 26.7 Different scene of bounding prediction of test image

26.4.3 Results

To make the best experimental results, we conducted many different attempts and experiments. On the initial stage, we used HOG features only. After 21 classifiers detect, the precision–recall curve is shown in Fig. 26.5.

Then, we extracted color features to train classifiers. One strategy is to merge the HOG and HSV histograms into one feature vector. But the effect of this strategy is not ideal. The other is to use HOG classifiers in the former part and the rest of the candidate windows are judged by color classifiers. The test result shows that this model can improve the precision and recall rates. The contrast fig is shown in Fig. 26.6.

Figure 26.7 illustrates an example of bounding prediction of test images. The top of three images are the first part results. The bottom of three images are the final results.

26.5 Conclusion

This paper proposes an efficient method for head detection. This method uses 21 ELM classifiers trained by HOG and 3 ELM classifiers trained by HSV histograms. The experimental results have shown that when the head area is occluded or lateral in the test image, the proposed approach can still detect the position of the head area.

References

1. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of IEEE CS conference computer vision and pattern recognition, vol 1, pp 886–893
2. Alonso IP, Llorca DF, Sotelo MÁ et al (2007) Combination of feature extraction methods for SVM pedestrian detection. IEEE Trans Intell Transp Syst 8(2):292–307
3. Rowley HA, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Trans Pattern Anal Mach Intell 20(1):23–38
4. Rowley HA, Baluja S, Kanade T (1998) Rotation invariant neural network-based face detection. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 38–44
5. Zhao L, Thorpe CE (2000) Stereo-and neural network-based pedestrian detection. IEEE Trans Intell Transp Syst 1(3):148–154
6. Brubaker S, Wu J, Sun J, Mullin M, Rehg J (2008) On the design of cascades of boosted ensembles for face detection. Int J Comput Vis 77:65–86
7. Paisitkriangkrai S, Shen C, Zhang J (2008) Fast pedestrian detection using a cascade of boosted covariance features. IEEE Trans Circuits Syst Video Technol 18(8):1140–1151
8. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vision 57:137–154

9. Lehmann A, Leibe B, Gool LV (2009) Feature-centric efficient subwindow search. In: Proceedings of IEEE international conference on computer vision, Oct. 2009
10. Gualdi G, Prati A, Cucchiara R (2012) Multistage particle windows for fast and accurate object detection. IEEE Trans Pattern Anal Mach Intell 34(8):1589–1604
11. Huang GB (2014) An insight into extreme learning machines: random neurons, random features and kernels. Cogn Comput 6:1–15
12. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybernet 2(2):107–122
13. Decherchi S et al (2012) Circular-ELM for the reduced-reference assessment of perceived image quality. Neurocomputing 102
14. An L, Bhanu B (2012) Image super-resolution by extreme learning machine. In: IEEE international conference on image processing
15. Minhas R et al (2012) Incremental learning in human action recognition based on snippets. IEEE Trans Circuits Syst Video Technol 22(11):1529–1541
16. Choi Y et al (2012) Incremental face recognition for large-scale social network services. Pattern Recogn 45(8):2868–2883
17. Mita T, Kaneko T, Hori O (2005) Joint haar-like features for face detection. In: Proceedings of 10th IEEE international conference on computer vision, vol 2, pp 1619–1626
18. The PASCAL Visual Object Classes Challenge (VOC2012) http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2012/index.html

Chapter 27
Automatic Depression Discrimination on FNIRS by Using FastICA/WPD and SVM

Hong Song, Weilong Du and Qingjie Zhao

Abstract A method is proposed for distinguishing patients with depression from normal controls based on data measured by FNIRS during a cognitive task. First, Fast Independent Component Analysis (FastICA) and Wavelet Package Decomposition (WPD) are used to extract features from 52-channel Functional Near-Infrared Spectroscopy (FNIRS) data of patients with depression and normal healthy persons. Then a classifier based on Support Vector Machine (SVM) is designed for classification. The experimental results indicate that the proposed method can achieve a satisfactory classification with the accuracy 86.7647 % for total and 90.74 % for patients. Also, the results suggested that FNIRS may be a promising clinical tool in the diagnosis and treatment of psychiatric disorders.

Keywords Depression discrimination · FNIRS · FastICA · WPD · SVM

27.1 Introduction

Functional Near-Infrared Spectroscopy (FNIRS) works primarily by sensing the oxygen variation in the blood through the absorption of near-infrared light [1–3]. In recent years, FNIRS is applied in the diagnosis of psychiatric disorders like depression. But after the FNIRS signals of patients are detected, doctors can only discriminate depression by their clinical experience. In order to provide objective basis for the diagnosis, it is necessary to do more analysis on the FNIRS data and find the significant difference between healthy persons and patients with depression.

Depression discrimination has been done based on many kinds of signal, the most commonly used in the clinical setting are magnetic resonance imaging (MRI)

H. Song (✉)
School of Software, Beijing Institute of Technology, Beijing 100081, China
e-mail: anniesun@bit.edu.cn

W. Du · Q. Zhao
School of Computer Science, Beijing Institute of Technology, Beijing 100081, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_27

and Electroencephalograph (EEG). For example, bipolar disorder and schizophrenia were classified by combining functional magnetic resonance imaging (fMRI) and fractional anisotropy (FA) data using a multimodal canonical correlation analysis (mCCA) and joint ICA model [4]. Liu and Lu used ICA to extract features from functional magnetic resonance imaging (fMRI) and classified the patients with depression and normal healthy person, but the accuracy need to be improved [5]. Zhao and Luo improved the current diagnosis of depression using fMRI and diffusion tensor imaging (DTI) data by wavelet based fusion scheme, depression recognition is achieved using a leave-one-out cross-validation test [6]. Zhang and Wang used Common Spatial Subspace Decomposition (CSSD) to extract features from 16-channels Electroencephalograph (EEG) of melancholia patients and normal healthy persons, respectively [7]. Lou and Zhang realized the automatic classification between patients with depression and healthy persons by extracting the disease features from the melancholic's EEG signals [8]. The application of FNIRS in the depression discrimination is still in its infancy. Ahmad Chaddad improved the signal-to-noise ratio of FNIRS using FastICA, PCA, and wavelet [9], but did not do further work like classification. Zhu classified the patients with depression and normal healthy person using General Linear Model (GLM) [10] on FNIRS, but this method is based on time series instead of the frequency domain that represents the essence of signal.

FNIRS is noninvasive, and the patients can be tested safely and repeatedly. Its cost is lower than other techniques. In addition, the testing environment is comfortable and acceptable for patients. Based on the above advantages of FNIRS, a method based on FastICA + WPD + SVM is proposed in this paper to discriminate the patients with depression and healthy persons.

27.2 Dataset Description

The original data was provided by Peking University Institute of Mental Health, Experiment device is the machine that can produce 52-channel NIRS signal (ETG-4000; Hitachi Medical Corporation).

In this study, NIRS machine continuously monitored the prefrontal activation effect of subjects during one-back memory task. The NIRS test consists three periods: preperiod (prescan 5 s and waiting time 25 s), subjects need to sit still and straight ahead; stimulus period (70 s), subjects get some figures (0–9), if the current figure is the same as the previous figure, then right index figure is acquired to press, figure rendering time is 0.5 s, the interval between the figures is 2 s, and a total of 29 figures were applied; relax time (50 s), this period also need subjects to sit still and straight ahead.

The experiment dataset includes original NIRS data of 108 patients with depression and 30 healthy volunteers from communities. And the patients with depressions are the patients of Peking University Sixth Hospital inpatients from January 2012 to December 2013 period.

27.3 Depression Discrimination Method

The process of depression discrimination includes preprocess of FNIRS data, feature extraction, classifier training, and classification. The flowchart of depression discrimination is shown in Fig. 27.1.

27.3.1 Preprocessing

In order to reduce the effects of baseline fluctuation and high frequency noise, Band-pass filtering is used to do data preprocessing. Before band-pass filtering, Fourier transform is used to do the spectrum analysis to find which band the FNIRS exists in. In Fig. 27.3, the blue line plots the original oxyhemoglobin concentration variation of healthy person and depression, Fig. 27.2 plots the Fourier transform

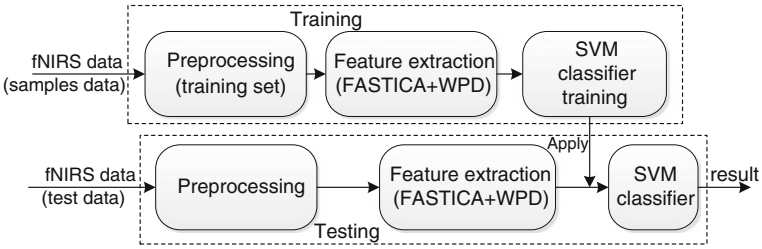
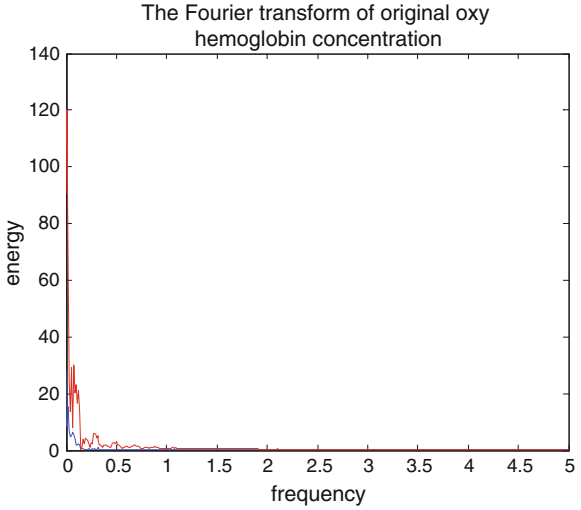


Fig. 27.1 The flowchart of depression discrimination

Fig. 27.2 The Fourier transform result of original oxyhemoglobin concentration, the blue line represents the healthy person and red one represents the depression



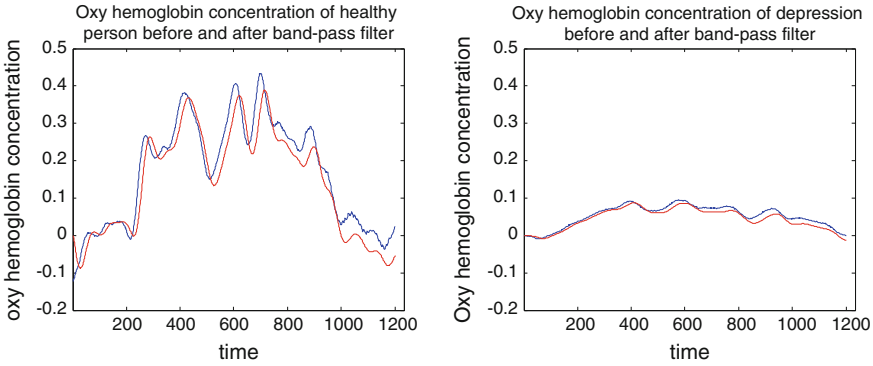


Fig. 27.3 Oxyhemoglobin concentration before and after band-pass filter

result of original oxyhemoglobin concentration variation of healthy person and depression, the blue line plots the Fourier transform result of healthy person and the red line plots the result of depression.

It is obviously shown in Fig. 27.2 that the frequency of FNIRS signal mainly exists in 0.01–0.5 Hz, so a band-pass filter range from 0.01 to 0.5 Hz can be designed to do the filtering. The filtering result of healthy person and depression using the band-pass filter is shown in Fig. 27.3, the blue line represents the original data and the red line represents the data after band-pass filtering. The curve after band-pass filter is smoother than before, some high frequency noise is reduced. This reduces the effects of high frequency noise and improved the classification accuracy.

27.3.2 Feature Extraction

27.3.2.1 FastICA-Based Dimension Reduction

In this paper, FastICA algorithm is used to reduce dimension. FastICA is invented by Aapo Hyvarinen at Helsinki University of Technology, among the independent component analysis algorithm, FastICA is very efficient and popular. It is based on a fixed-point iteration scheme maximizing non-Gaussianity as a measure of statistical independence. It can also be derived as an approximative Newton iteration. FastICA can reduce the FNIRS data dimension without negatively affecting the appropriate information in each signal.

27.3.2.2 WPD-Based Energy Feature Extraction

WPD can be regarded as the spread of Wavelet Decomposition (WD). Wavelet Package Transform is based on Rigorous mathematical theory and numerical methods, both the low and high frequency parts are decomposed recursively. As both low and high frequency parts are decomposed recursively in WPD, this can get higher frequency resolution, so we can extract effective discrimination feature [10, 11].

The essence of WPD is to decompose the signal according to different frequency components, so we can get the energy contained by each frequency component; and frequency is the most essential feature of signal, so energy contained by different frequency can be regarded as the feature from one kind of signal to another.

For example with four-stage wavelet packet decomposition, the structure diagram of four-stage wavelet package composition using db4 is shown in Fig. 27.4.

It is shown in Fig. 27.4 that the signal is decomposed into four stages and the final result includes 16 components, (4, 0), (4, 1), ……, (4,15) are the 16 components and every component has the same frequency.

The steps of energy feature extraction are as follows:

Step 1: Make the four-stage wavelet package composition with db4 on the data after reducing dimension.

Step 2: Reconstruct signal using the coefficient of wavelet package composition. For example, with four-stage wavelet packet decomposition

$$S = S_{40} + S_{41} + \ldots + S_{415} \tag{27.1}$$

$S_{4i}(i = 1, 2, \ldots, 15)$ refers to the coefficient of wavelet package composition.

Step 3: Compute the energy of every reconstructive signal $E_j(j = 0, 1, \ldots, 15)$. Where

$$E_j = \sqrt{\sum_{k=1}^{N} \left| S_{4j}(k) \right|^2} \tag{27.2}$$

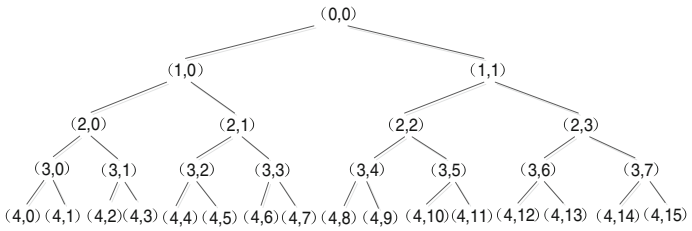


Fig. 27.4 The structure diagram of four-stage wavelet package composition

$S_{4j}(k)$ refers to the magnitude of reconstructive signal S_{4j} discrete points, $k = 1, 2, \ldots, N$, N is the point count of reconstructive signal.

Step 4: Construct the eigenvector. E_j refers to the energy of reconstructive signal, total energy

$$E = \sqrt{\sum_{j=0}^{15} E_j^2} \tag{27.3}$$

In this paper, because energy of the first reconstructive signal S_{40} accounts for up to 90 % of the total energy, the first reconstructive signal energy E_0 is selected to construct the eigenvector. So the eigenvector is

$$F = [E_{01}, E_{02}, E_{03}, E_{04}, E_{05}, E_{06}, E_{07}, E_{08}] \tag{27.4}$$

where E_{0d} (d = 1, 2, …, 8) refers to the energy of each dimension of the 8 dimensions data.

Figure 27.5 shows the eigenvector of 12 healthy persons and 9 patients with depression, red circle represents the eigenvalue of healthy person and blue triangle presents the eigenvector of patient with depression. According to Fig. 27.5, there exists significant difference between them, so the wavelet energy feature can be regarded as the feature of different kinds of FNIRS signal. The wavelet energy feature is available in discriminating depression.

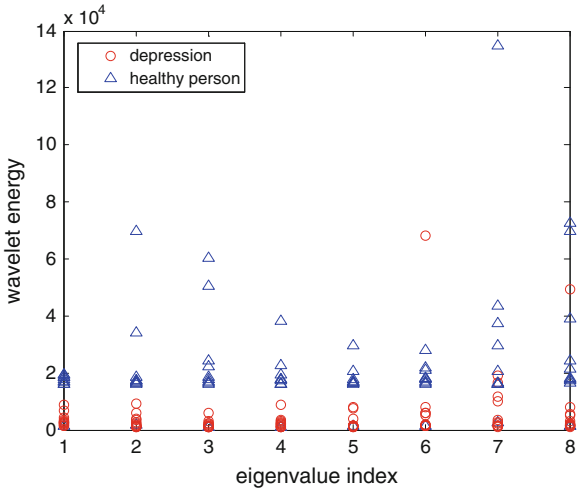


Fig. 27.5 Eigenvector of patient with depression and healthy person

27.3.3 SVM-Based Classifier Training

The earliest SVM algorithm was created by Vladimir N. Vapnik, and the present standard exemplification published in 1995 was proposed by Cortes and Vapnik in 1993 [12]. It has shown great advantages in the realm of machine learning. At the same time, SVM also has great advantage in solving high-dimension problems and the problems with small sample.

SVM is a classifier using maximum margin of classes [13]. It does pattern discrimination of two classes by searching a decision plane, the decision plane has the maximum removing to the nearest points in the training set, this decision plane is called the support vector. Starting with the training set that has some points $x_i \in R^n$, $i = 1, 2, 3, \ldots, n$, where every x_i is a part of the two types discriminated with the symbol $y_i \in \{-1, 1\}$. Supposing data is linear and separable, the task of maximum margin classification is to discriminate the two classes using a hyperplane, if not do this, the distance to the support vectors will not be the maximum and classifier will not be available. This hyperplane is called the optimal separating hyperplane (OSH). The goal of classifier training is to find the OSH between depression and healthy person.

27.4 Testing Result

In this paper, FastICA and LIBSVM software package are used during the training and testing process. The dimension of preprocessed data can be reduced from 52 to 8 using FastICA algorithm according to the contribution of eigenvalue. The LIBSVM software package is developed by Chih-Jen Lin of National Taiwan University [14].

Like the training set, the eigenvector of testing set also should be extracted. Using the same method with training set, the eigenvector of testing set can be extracted. Then the eigenvector will be input into the classifier, the output of classifier is the discrimination result and accuracy. In this paper, data of 54 depressions and 14 healthy persons are selected as the testing set.

The final testing result of depression and healthy is shown in Table 27.1. The method based on FastICA/WPD and SVM successfully discriminate 68 (54 depressions and 14 healthy persons) signals with an overall accuracy of 86.7647 % for the FNIRS classification. This accuracy is especially satisfactory for the depression discrimination.

Table 27.1 The result of discrimination

FastICA + WPD + SVM			Testing set	
			Depression	Healthy
			54	14
Classified result	Depression	53	49	4
	Healthy	15	5	10
Accuracy of depression(sensitivity)		49/54 = 90.74 %		
Accuracy of healthy (specificity)		10/14 = 71.43 %		
Overall discrimination accuracy		86.7647 %		

27.5 Conclusion and Discussion

In this paper, an effective classifier is trained by using FastICA/WPD and SVM. This classifier provided an objective basis for the clinical diagnosis and greatly improved the diagnostic accuracy of clinical depression. It can be an assist for doctors during the diagnosis of depression. This is also a big step forward in the diagnosis of depression on FNIRS.

Acknowledgments Project supported by the National Natural Science Foundation of China (Grant No. 61240010, No. 61175096).

References

1. Boas DA, Strangman G, Culver JP et al (2003) Can the cerebral metabolic rate of oxygen be estimated with near-infrared spectroscopy. Institute of Physics Publishing. Phys Med Biol 48 (2003):2405–2418
2. Strangman G, Boas DA, Sutton JP (2002) Non-invasive neuroimaging using near-infrared light. Biol Psychiatry 52:679–693
3. Jöbsis FF (1977) Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198:1264–1267
4. Sui J, Pearlson G, Caprihan A et al (2011) Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. NeuroImage 57(3):839–855
5. Zhu Y (2011) Methodology and applications of functional near-infrared spectroscopy. Dissertation Submitted to Institute of Automation, Chinese Academy of Sciences in partial fulfillment of the requirements for the degree of Doctor of Engineering (in Chinese)
6. Zhao J, Luo G, Yao Z et al (2012) Depression discrimination using fMRI and DTI data by wavelet based fusion scheme. J Southeast Univ (Engl Ed) 28(1):25–28
7. Zhang S, Wang W (2008) Application of CSSD and SVM for EEG signal classification. Chin J Biomed Eng 27(6):827–830 (in Chinese)
8. Lou E, Zhang S (2009) Application of Eigenvector methods and SVM for EEG signals classification. Chin J Med Phys 26(5):1415–1418 (in Chinese)
9. Chaddad A (2014) Brain function diagnosis enhanced using denoised FNIRS raw signals. J Biomed Sci Eng 7:218–227

10. Azimi-Sadjadi MR, Yao D, Huang Q, Dobeck GJ (2000) Underwater target classification using wavelet packet and neural networks. IEEE Trans Neural Netw 11(3):784–794)
11. Li X-X, Yang S-E, Yu M (2008) Feature extraction from underwater signals using wavelet packet transform. In: 2008 IEEE international conference neural networks and signal processing, ICNNSP, pp 400–405
12. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273
13. Heisele B, Ho P, Poggio T (2001) Face recognition with support vector machines: global versus component-based approach. In: Proceedings of eighth IEEE international conference on computer vision, ICCV 2001, vol 2
14. Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classification (Technical report). Department of Computer Science and Information Engineering, National Taiwan University

Chapter 28
Random Forests for Object Detection

Mingming Zhu, Lang Ye, Siyu Xia and Hong Pan

Abstract We present a method for object detection based on random forests. It is accomplished through the generalized Hough transform paradigm, where object centers are voted from small and local parts in images. Some previous works such as the well-known implicit shape model (ISM) take unsupervised clustering method during the training stage and it often lead to false-positive detections due to random constellations of parts. Thus, we employ a random forest to learn a more discriminative model. We use the KAZE local features to construct a random forest classifier and all leaf nodes in each tree constitute a discriminative codebook model. The codebook model is used to estimate object locations via probabilistic Hough voting. Furthermore, before the test stage, we also adopt a salient region detection step to reduce false-positive detections. Experiments show that our method provides good detection results in complicated environments.

Keywords Generalized hough transform · Random forests · Salient region detection

28.1 Introduction

Object detection and localization in cluttered scenes is still one of the main challenges in computer vision domain. In general, most methods in this field learn appearance and spatial relation models of the categories from labeled training images and use the obtained models to localize previously unseen instances in test images. Currently, two different approaches can be distinguished, namely sliding window and part-based methods. Sliding window-based methods like [11, 15] evaluate classifiers at all possible image locations and analyzing descriptors with fixed layouts like the histogram-of-gradients [6]. Although these approaches have

M. Zhu · L. Ye · S. Xia (✉) · H. Pan
School of Automation, Southeast University, Nanjing 210096, China
e-mail: xsy@seu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_28

shown to provide excellent results for rigid objects with fixed geometric properties, they yield limited performance for deformable objects.

Thus, recently, part-based models have become more popular. They mainly differ in the way the spatial relations between the individual parts are defined. One of the most simple structures is the BOW (bag-of-words) [14] model which calculates the histogram of local features without any spatial relation. A more complex method is the constellation model where each part is connected to all other parts. The deformable part model [8] is a tree-shaped model which is one of the most successful algorithms on the PASCAL VOC challenge. Another different type is the star-shaped model, where each part is only connected to a centroid part. This is the underlying spatial representation of the implicit shape model (ISM) [12].

The ISM is a generative codebook of interest point descriptors for a given class. It represents objects as a collection of a large number of prototype patches, with specific appearance and a defined location relative to the object center. At runtime, the interest point descriptors are matched against the codebook and the matches cast probabilistic votes about possible positions of the object. These votes are summed up into a Hough image and the peaks are considered as detection hypotheses. The whole detection process can be described as a generalized Hough transform [4]. However, it constructs the codebook by unsupervised clustering of local part appearances which will lead to many unreliable or redundant votes. These votes not only slow down the detection, but also reduce the precision of the result. Recently, Hough Forests have been presented in [9] as variant of an ISM and inspired a series of applications and extensions like [7, 13, 16]. Unlike previously presented ISMs, it extracts local features densely instead of just on interest points at test time, resulting in the aggregation of more evidence and increased robustness.

In this paper, we develop a new detection method based on random forests [5]. Given a dataset of images with bounding box annotated samples of the class instances, we learn a random forest classifier that is able to map an image patch to a probabilistic vote about the position of an object centroid. We use the KAZE features [2], which is more stable than SIFT and SURF, to represent image patches. At runtime, such a random forest is applied to patches extracted from test images and the resulting votes are accumulated in Hough images where the maxima are sought. In addition, we adopt a salient region detection step before test to make the detection more precise and efficient.

28.2 Training of Random Forests Classifier

Random forests are trained in a supervised way. A typical random forest consists of a set of binary decision trees. Each tree in the forest is an individual classifier and the final result is the integration of all the tree predictions. It has been shown [3, 5] that assembling together several trees trained in a randomized way achieves superior generalization and stability than a single deterministic decision tree. The randomization is achieved first by training each tree on a random subset of the

training data; and second by considering a random subset of possible binary tests at each nonleaf node. Among this random subset, the training procedure picks the binary test which splits the training samples in the optimal way.

28.2.1 Random Forests Construction with KAZE Features

KAZE features is a novel multiscale 2D feature detection and description algorithm in nonlinear scale spaces. Unlike other approaches, it does not detect and describe features at different scale levels by building or approximating the Gaussian scale space of an image. In contrast, it detects and describes 2D features in a nonlinear scale space by means of nonlinear diffusion filtering. The nonlinear scale space is built using efficient additive operator splitting (AOS) techniques and variable conductance diffusion. In this way, it makes the blurring locally adaptive to the image data and obtains superior localization accuracy and distinctiveness.

The first step of construction is to extract KAZE local features from training images. We describe the assemble of all local features extracted form training data as $F = \{f_i\}, i \in [1, N]$. Each description $f_i = (\mathbf{d}_i, c_i, \mathbf{r}_i)$ is composed by the KAZE descriptor $\mathbf{d}_i$, the class label c_i and the offset $\mathbf{r}_i$ of the feature location relative to the object center. Features sampled from inside the object bounding boxes are assigned $c_i = 1$ and others are assigned $c_i = 0$.

A random forest is defined as $T = \{T_i\}_{i=1}^{N_t}$, each tree is composed by the root node n_{root} and a series of ordered nonleaf nodes n_{split} and leaf nodes n_{leaf}. In general, every tree in a random forest is constructed recursively starting from the root. During construction, each node receives a set of training samples. If the depth of the node is equal to the maximum or the number of samples is small enough, the constructed node is declared a leaf and its vote information is accumulated and stored. Otherwise, a nonleaf node is created and an optimal binary test is chosen from a large pool of randomly generated binary tests. The training sample set that has arrived to the node is then split according to the chosen test into two subsets that are passed to the two newly created child nodes, after that the recursion proceeds.

28.2.2 Binary Test on Nonleaf Nodes

Given a nonleaf node n_{split}, we describe all the training samples arriving it as $\{f_k\}, k \in \left[1, N_{\text{split}}\right]$. These samples are tested by a subset of m feature components which are randomly selected from the entire feature descriptor space $\mathbf{R}^M$. The subset is described as $A_m = \left\{a_k, a_k \in \mathbf{R}^M\right\}, k \in [1, m]$. We use information entropy to estimate the impurity of class labels in a node

$$\text{Entropy}(I) = -\sum_{i=1}^{K} p(c_i, I) \log p(c_i, I) \tag{28.1}$$

where I represents the assemble of samples in the node, K is the number of class labels, and $p(c_i, I)$ equals to the proportion of samples labeled as c_i.

We define the information gain, which is caused by a binary split with feature component a_k, as:

$$G(I, a_k) = \text{Entropy}(I) - p(I_{l_k})\text{Entropy}(I_{l_k}, a_k) - p(I_{r_k})\text{Entropy}(I_{r_k}, a_k) \tag{28.2}$$

where I_{l_k} represents samples going to the left child node and I_{r_k} is samples going to the right child node. $p(I_{l_k})$ is the proportion of samples going to left and $p(I_{r_k})$ is the proportion going to right and they are calculated as

$$\text{Entropy}\left(I_j, a_k\right) = -\sum_{i=1}^{K} p(c_i|a_k, I_j) \log p(c_i|a_k, I_j), j = l_k, r_k \tag{28.3}$$

$$p\left(I_j\right) = \frac{|I_j|}{|I_{l_k}| + |I_{r_k}|}, j = l_k, r_k \tag{28.4}$$

The goal of binary test is to reduce the uncertainty of class labels in tree nodes. It means to reduce the impurity of class labels in a node and the best feature component chosen to split a node is the one which makes the information gain reach the maximum:

$$G(I, a^*) = \underset{a_k \in A_m, k \in [1, m]}{\arg\max} \{G(I, a_k)\} \tag{28.5}$$

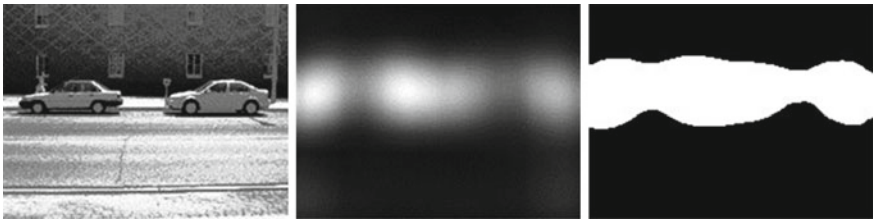
where $a^* \in A_m$ equals to the best candidate feature component.

28.3 Object Detection with Random Forests Classifier

28.3.1 Salient Region Detection

Given a test image, if we make the detection directly without any processing, there may be many false-positives due to richly textured backgrounds on which a large number of spurious object parts will be found. However, we find that objects are always the salient regions of an image where we focus our attention. Thus by detecting objects mainly in such salient image regions, most false-positives of backgrounds can be excluded.

We make the salient region detection through the approach proposed in [10], which is independent of features, categories, or other forms of prior knowledge of



Input image Saliency map Object map

Fig. 28.1 An example of salient region detection result

the objects. It is a fast method to construct the corresponding saliency map in spatial domain by analyzing the log-spectrum of an input image and extracting the spectral residual of an image in spectral domain. Figure 28.1 shows an example of salient region detection.

28.3.2 Probabilistic Hough Voting

After salient region detection, we have learned the probable regions where an object will occur. Then we increase the weight coefficient of features inside the salient regions and decrease the weight coefficient of those outside the salient regions during the Hough voting stage.

Given a test image, we need to predict the class labels of KAZE local features extracted from it. For a test feature f_{test}, the probability of its label being judged as $k(k = 0, 1)$ by the votes from all the trees in the forest is

$$p(k|f_{\text{test}}) = \frac{1}{N_t}\sum_{n=1}^{N_t} p_n\left(C_j = k\right) \tag{28.6}$$

where $p_n(C_j = k)$ is the probability of being judged as label k in leaf node j of a tree. Finally, the feature's class label C_{test} can be computed by

$$C_{\text{test}} = \arg\max_{k\in\{0,1\}}\{P(k|f_{\text{test}})\} \tag{28.7}$$

If a test sample is judged as a part of object, which means its class label is judged as $C_{\text{test}} = 1$ by the random forest classifier, then this feature will be used to make a probabilistic vote to compute the object center. Consider a test sample $f_i^p = (d_i^p, c_i^p, l_i^p)$ predicted as an object part $(c_i^p = 1)$ with its location and scale described as $l_i^p = (x_i^p, y_i^p, \lambda_i^p)$, we assume it reaches leaf node L with class label C_L in tree T_t. The offset assemble R_L and class label assemble c_L of training samples arriving the node are described as $R_L = \left\{\left(x_i^L, y_i^L, \lambda_i^L\right)\right\}$ and $c_L = \{c_i\}, i \in [1, R_L|]|$. The Hough

vote $l^{\text{vote}} = \left\{\left(x_i^{\text{vote}}, y_i^{\text{vote}}, s_i^{\text{vote}}\right)\right\}, i \in [1, n_{\text{vote}}^p]$ can be calculated from expression (28.8) to (28.10). Here n_{vote}^p represents the vote number coming from feature f_i^p and all the votes form a discrete probabilistic Hough voting space.

$$s_i^{\text{vote}} = \lambda_i^p / \lambda_i^L \tag{28.8}$$

$$x_i^{\text{vote}} = x_i^p + x_i^L \times s_i^{\text{vote}} \tag{28.9}$$

$$y_i^{\text{vote}} = y_i^p + y_i^L \times s_i^{\text{vote}} \tag{28.10}$$

When casting votes for the object center, we describe a hypothesis as $h = (O_n, x)$ where x is the center of object O_n. Thus the probabilistic votes contributed by feature f_i^p is:

$$p(O_n, x|f_i^p, l_i) = \theta_i \sum_{t=1}^{Nt} p(O_n, x|L, T_t, l_i) p(L|T_t, f_i^p) \tag{28.11}$$

where θ_i is the weight coefficient of votes from feature f_i^p (we set $\theta_i = 1$ for features inside salient regions and $\theta_i = 0.5$ for those outside salient regions), $p(O_n, x|L, T_t, l_i)$ equals to the probabilistic distribution of training samples and $p(L|T_t, f_i^p)$ is the probability of the feature reaching leaf node L. They satisfy

$$p(O_n, x|L, T_t, l_i) = \begin{cases} \frac{1}{|R_L|}, & C_L = 1 \\ 0, & \text{otherwise} \end{cases} \tag{28.12}$$

$$p(L|T_t, f_i^p) = \frac{1}{|L|} \tag{28.13}$$

where $|R_L|$ is the number of training samples stored in leaf node L and $|L|$ is the total number of leaf nodes in tree T_t. The votes are summed up into a Hough image and the peaks of it are considered as detection hypotheses.

28.4 Experiments

We evaluated our approach on the UIUC cars dataset where we provide a performance comparison with the related detection methods. The experiment was performed using the evaluation scheme and detection tolerances from [1] based on bounding box overlap.

The UIUC cars dataset contains images of side views of cars. The test data are split into the set of 170 images with 201 cars of, approximately, same scale (UIUC-Single) and the set of 108 images containing 139 cars at multiple scales (UIUC-Multi). We trained our random forest classifier on the available 550 positive and

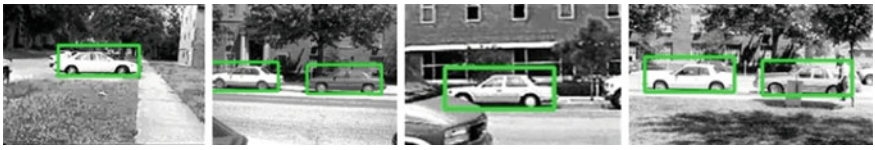


Fig. 28.2 Example detections on difficult images from the test set

Table 28.1 Comparison of different methods on the UIUC-Single database

Methods	Precision (%)
Implicit shape model [12]	91.0
ISM + MDL [12]	97.5
Hough forests [9]	98.5
Our method	99.0

Fig. 28.3 All missing detections (*left two*) and false-positives (*right two*)

500 negative training images. The minimum number of training samples in each node was set as 20 and the maximal depth of each tree was set as 15. We totally constructed 15 decision trees. Each tree was constructed by two-thirds of the whole samples and the number of feature components for random binary test was chosen as $\sqrt{D} = \sqrt{64} = 8$.

Figure 28.2 shows some example detections in difficult environments. Table 28.1 contains a comparison of detection precision with some other approaches. With a precision of 99.0 %, our method presents an improvement over the implicit shape model and Hough Forests on the UIUC-Single dataset. At the EER point, our method correctly finds 199 of the 201 test cases with only 2 false-positives. All of these cases are displayed in Fig. 28.3.

28.5 Conclusion

We have introduced an approach for object detection based on random forests which can be trained on large datasets without significant over-fitting. We use the KAZE local features to construct a random forest classifier and all leaf nodes in each decision tree constitute a discriminative codebook model. Such a forest can be efficiently used to detect instances of classes in natural images by probabilistic

Hough voting. In addition, we also adopt a salient region detection step before the test stage in order to reduce false-positive detections. Experimental results show that our method provides good detection results in complicated environments.

Acknowledgments This work has been funded, in part, by NSFC under Grant No. 61374038, Natural Science Foundation of Jiangsu Province under Grant No. BK20130639 and No. BK20131296.

References

1. Agarwal Shivani, Awan Aatif, Roth Dan (2004) Learning to detect objects in images via a sparse, part-based representation. IEEE Trans Pattern Anal Mach Intell 26(11):1475–1490
2. Alcantarilla PF, Bartoli A, Davison AJ (2012) KAZE features. In: Computer vision–ECCV 2012, Springer, Berlin, pp 214–227
3. Amit Yali, Geman Donald (1997) Shape quantization and recognition with randomized trees. Neural Comput 9(7):1545–1588
4. Ballard Dana H (1981) Generalizing the hough transform to detect arbitrary shapes. Pattern Recogn 13(2):111–122
5. Breiman Leo (2001) Random forests. Mach Learn 45(1):5–32
6. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, IEEE, CVPR 2005, vol 1
7. Dantone M et al (2012) Real-time facial feature detection using conditional regression forests. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR), IEEE
8. Felzenszwalb PF et al (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
9. Gall J, Lempitsky V (2009) Class-specific hough forests for object detection. In: 2009 IEEE conference on computer vision and pattern recognition (CVPR), IEEE
10. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition, 2007, CVPR'07, IEEE
11. Lampert Christoph H, Blaschko Matthew B, Hofmann Thomas (2009) Efficient subwindow search: A branch and bound framework for object localization. IEEE Trans Pattern Anal Mach Intell 31(12):2129–2142
12. Leibe B, Leonardis A, Schiele B (2004) Combined object categorization and segmentation with an implicit shape model. In: Workshop on statistical learning in computer vision, ECCV, vol 2, issue 5
13. Schulter S et al (2013) Alternating regression forests for object detection and pose estimation. In: 2013 IEEE international conference on computer vision (ICCV), IEEE
14. Sivic J, Zisserman A (2003) Video google: a text retrieval approach to object matching in videos. In: 2003. Proceedings ninth IEEE international conference on computer vision, IEEE
15. Viola Paul, Jones Michael J (2004) Robust real-time face detection. Int J Comput Vision 57 (2):137–154
16. Wohlhart P et al (2012) Discriminative hough forests for object detection. *BMVC*

Chapter 29
Dynamic Simulation of Tank Car Derailment and Structure Optimization Based on Solidworks, ADAMS, and ANSYS

Lanxia Zhang, Yong Qin, Dapeng Zhu, Li Wang and Jianghua Gao

Abstract This paper establishes a three-dimensional model of tank car derailment and collision to analyze gross motions and forces in this process with the GQ70 tank car as the research object. To simplify the complexity of the model, it is implemented in ADAMS, thus the author summarizes each of the three most dangerous conditions of tank car and the corresponding force spectrum files. Based on the above results, this collision is examined using dynamic, nonlinear finite element analysis. The results indicate that it is essential for optimization design of tank car, including optimization of the tank car itself and coupler. This paper aims to optimize tank car to be capable of surviving impact speed as twice as now. The contents of this paper provide the ideas of optimization design and the research results lay a solid foundation for railway vehicle safety management and vehicle maintenance, being of theoretical significance and practical value.

Keywords Transportation safety · Derailment and collision model · Finite element analysis · Optimization design · Computer simulation

L. Zhang · L. Wang (✉) · J. Gao
School of Traffic and Transportation, Beijing Jiaotong University,
No. 3 Shangyuancun, Haidian District, Beijing 100044, China
e-mail: wangli@bjtu.edu.cn

L. Zhang · Y. Qin · L. Wang · J. Gao
State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District,
Beijing 100044, China

D. Zhu
School of Traffic and Transportation, Lanzhou Jiaotong University,
Lanzhou 730070, China

Y. Qin · L. Wang
Beijing Research Center of Urban Traffic Information Sensing
and Service Technologies, Beijing Jiaotong University,
No. 3 Shangyuancun, Haidian District, Beijing 100044, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_29

29.1 Introduction

The whole society increasingly pays attention to railway transportation security, especially hazardous materials transportation safety. Currently, over 85 % dangerous goods in the national road network are transported by tank car. Although tank car derailment and collision are not frequent, when that happens it would bring unimaginable loss. For instance, on May 9, 2013, a train made up of 50 oil tankers almost entirely derailed, hence 10 carriages caught fire and no less than 52 people were injured in the derailment accident. Given that, optimization of the tank car structure plays a very important role to improve the level of tank car transportation security.

In recent years, scholars of home and abroad studied a lot to improve the level of tank car transportation safety. In Ref. [1], the author combined with the motion state of tank car derailment and established the equations of train derailment motion, which provided a theoretical basis to study tank car derailment. In Ref. [2], after analysis of the derailment environment, a reasonable improved design schemes were proposed based on full-scale impact tests. In Ref. [3], the author creatively designed sacrificial structure that could shield tank car and absorb energy, so that the sacrificial structure ensured the integrity of railroad tank cars. Considering train derailment, the domestic research about optimization design of tank car is relatively little. In Ref. [4], based on finite element and multibody dynamics technique, the paper presented a joint simulation on train anticollision system and verified its feasibility.

In this paper, on the basis of research achievements of domestic and foreign, the author will use the most common tank car (GQ70) in railway as the research object. The goal is that the crashworthy speed of optimization tank car is two times higher than unimproved one. Using three softwares, the paper improves the existing rail tank car, which provides constructive comments for future tank car design.

29.2 Baseline Tank Car Three-Dimensional Model Based on Solidworks

According to the real size of the tank car GQ70, the tank car three-dimensional solid model was established in Solidworks [4]. This paper keeps bogie, tank car body, coupler, and so on. To a large degree, these key parts influence on the outputs of simulation analysis. Some minor parts of the tank car are simplified and finally assembled into a whole. Each component model of tank car is shown in Fig. 29.1.

Figure 29.1a shows the bogie model. Every tank car consists of two bogies. Figure 29.1b shows a tank car model, which includes a tank car body, the bogies and the addition of the outtriggers attached to the draft gear. Figure 29.1c shows the train solid model in Solidworks. The model provides the foundation for the joint simulation.

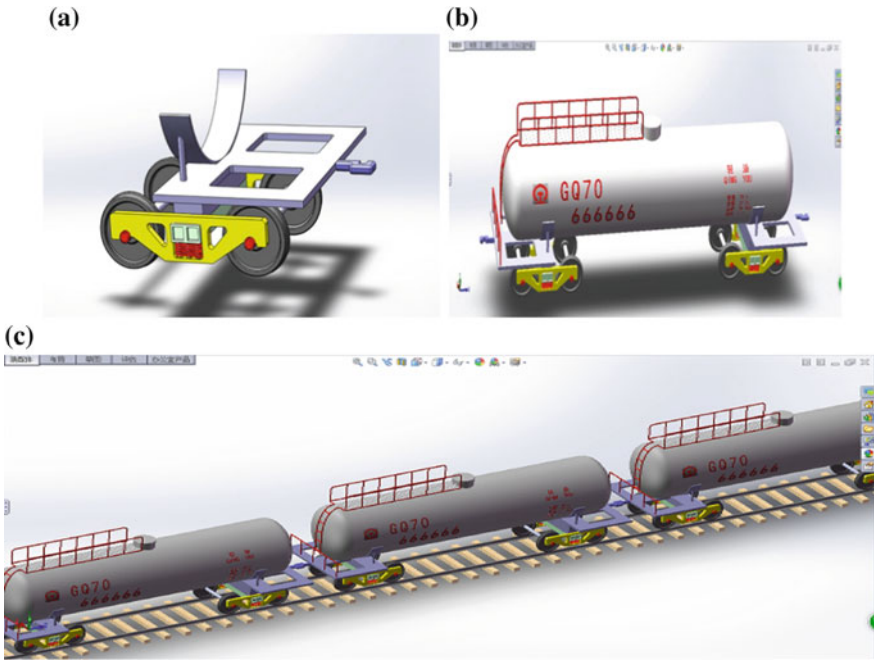


Fig. 29.1 Solidworks model of train, **a** bogie model, **b** tank car model, **c** train derailment model

29.3 Analysis of the Tank Car Derailment and Collision Dynamic Simulation Based on ADAMS

The whole model assembled in Solidworks is imported into ADAMS through public file interface. Then constraints are added in bogie internals between various parts. Contact model is established between the tank car and coupling based on impact function. Finally, tank car train simulation model is completed.

29.3.1 Basic Mathematical Model of Tank Car Derailment

Considering the different components of a tank car individually, they are applied to create a completed tank car model. Once the complete free body diagram for a car has been developed, the equations of motion can be derived. A generalized application of these forces and moments to the free bodies is shown in Fig. 29.2 [1]:

The forces in the x- and z-directions, and the moments are summed below for a basic car. The forces and moments do not apply to the lead and last cars in the train [5].

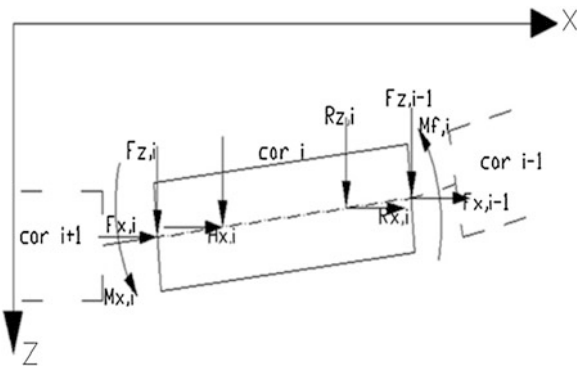


Fig. 29.2 Forces and torques applied to car *i*

The equation of force on *x*-axis:

$$\sum F_{x,i} = R_{x,i} + H_{x,i} - F_{x,i-1} + F_{x,i} - E_{i-1}\sin(\beta_{i-1}) + E_i \sin(\beta_i) \tag{29.1}$$

The equation of force on *y*-axis:

$$\sum F_{z,i} = R_{z,i} + H_{z,i} - F_{z,i-1} + F_{z,i} - E_{i-1}\cos(\beta_{i-1}) + E_i \cos(\beta_i) \tag{29.2}$$

The equation of moment:

$$\begin{aligned}\sum M_i =& 1/2 D_i \left[\left(R_{x,i}^2 + R_{z,i}^2\right)^{1/2} + \left(H_{x,i}^2 + H_{z,i}^2\right)^{1/2}\right]\left[\sin\left(\theta_i - \alpha_{f,i}\right) - \sin\left(\theta_i - \alpha_{r,i}\right)\right] \\ &+ M_{r,i} + E_{i-1}\cos(\theta_i - \beta_{i-1}) + 1/2 L_i\left(F_{x,i-1}^2 + F_{z,i-1}^2\right)^{1/2}\sin(\theta_i - \beta_{i-1}) \\ &+ 1/2 L_i\left(F_{x,i}^2 + F_{z,i}^2\right)^{1/2}\sin(\theta_i - \beta_{i-1}) + M_{f,i} + E_i\cos(\theta_\mathrm{i} - \beta_\mathrm{i})\end{aligned} \tag{29.3}$$

The summed forces and moments are then set equal to the inertial terms to complete the equations of motion for the *i*th car.

$$\sum F_{x,i} = M_i \bar{x}_i \tag{29.4}$$

$$\sum F_{z,i} = M_i \bar{z}_i \tag{29.5}$$

$$\sum M_i = I_i \ddot{\theta}_i \tag{29.6}$$

29.3.2 Dynamic Simulation of Tank Car Derailment and Collision

Based on the above two points, the mathematical model is combined with the three-dimensional solid model established in ADAMS. The ADAMS is used to carry out the kinematics and dynamics simulations, getting the simulation output results, as shown in Figs. 29.3 and 29.4.

The Fig. 29.3 shows the three most dangerous collision conditions in the process of train derailment. The first condition is the head impact that the coupler of one tank car impacts the head of a second tank car. The second condition is the shell impact that the coupler of one tank car impacts the shell of a second tank car. The third condition is the head-shell impact that the head of one tank car impacts the shell of a second tank car. Meanwhile Fig. 29.4 shows that the maximum relative impact velocity of tank car is about 10 m/s. The maximum relative impact speed is between car 7 and car 8.

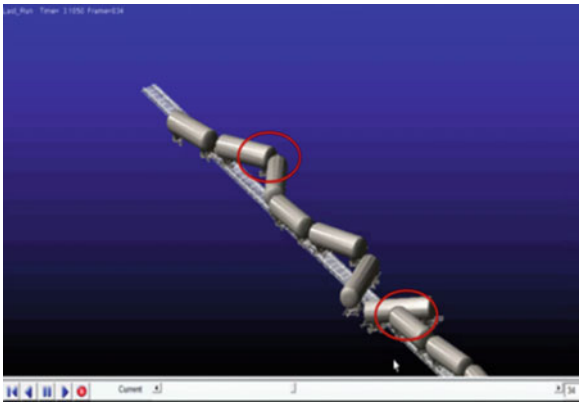


Fig. 29.3 The condition after tank car derailment and collision

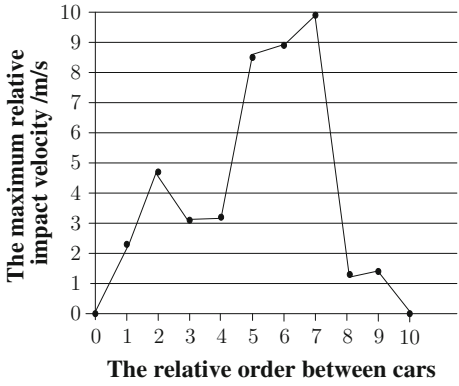


Fig. 29.4 The maximum relative impact velocity

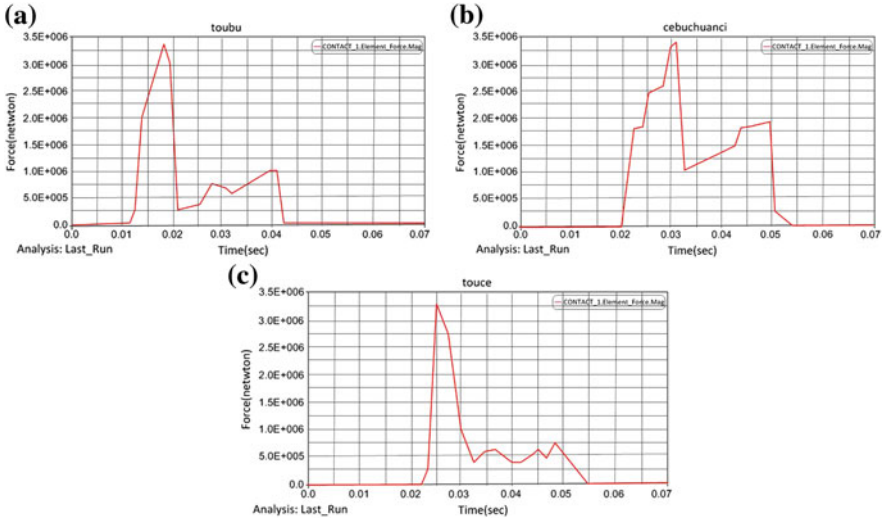


Fig. 29.5 The collision force–time curves in three collision conditions **a** the first condition **b** the second condition **c** the third condition

In view of the above-listed three conditions, this paper, respectively, established three kinds of tank car collision three-dimensional dynamic model in ADAMS [6]. According to simulation results of train derailment and the maximum relative impact velocity of tank car (10 m/s), the experimental simulation is carried out. The simulation outputs force spectrum file in the process of tank car collision. The final results provide the curve of collision force changing in time under the three most dangerous collision conditions as shown in Fig. 29.5.

Figure 29.5 are the collision force–time curves in three collision conditions. The trend of the change of collision force with time is similar in spite of different collision conditions, but there are also differences in maximum impact force. According to Fig. 29.5a, when relative impact velocity of tank car is 10 m/s at time 0.018 s in the first condition, impact force is measured to 3380 kN. Figure 29.5b shows that at time 0.027 s in the second condition, impact force is measured to 3428 kN. Figure 29.5c shows that at time 0.024 s in the second condition, impact force is measured to 3345 kN.

29.4 The Finite Element Analysis

First, tank car solid model built up in ADAMS is imported into the finite element analysis software ANSYS by file interface between them. The second step is to plot the grid, enact the parameters, load force spectrum file. Then, the transient dynamic analysis module of the ANSYS software is adopted to calculate the dynamic stress

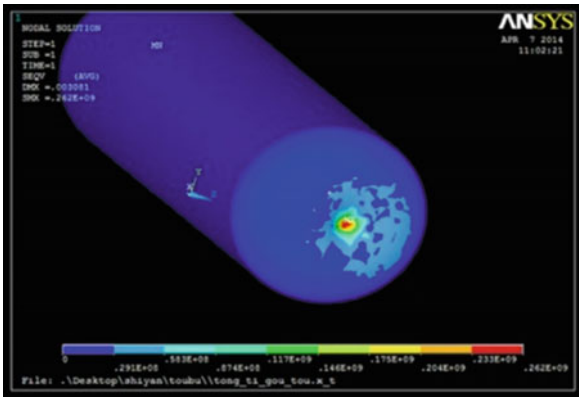


Fig. 29.6 Head stress nephogram before optimization

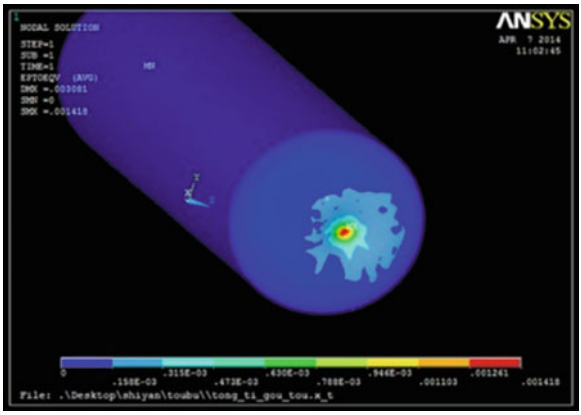


Fig. 29.7 Head strain nephogram before optimization

of the tank car framework [7, 8]. After computing and postprocessing, the ANSYS outputs the analytic result of the stress and strain nephogram. Figures 29.6 and 29.7 show the head stress and strain nephogram in the first condition.

According to the analysis of the stress and strain nephogram, the maximum stress value of tank car head is up to 262 Mpa and the maximum strain value of tank car head is up to 1.418. Tables 29.1 and 29.2 list the stress value in the second and third condition in Chap. 4. Thus the maximum stress value of tank car head and shell,

Table 29.1 Strain table before and after optimization

Condition	The place of maximum strain	The maximum strain before optimization	The maximum strain after optimization		
		V = 10 m/s	V = 10 m/s	Drop (%)	V = 20 m/s
1	Head	1.418	0.945	33	1.537
2	Shell	1.785	0.972	37	1.871
3	Head	1.357	0.702	48	1.611
	Shell	1.212	0.815	33	1.817

Table 29.2 Stress table before and after optimization

Condition	The place of maximum stress	The maximum stress before optimization/Mpa	The maximum stress after optimization/Mpa			Allowable stress/Mpa
		10 m/s	10 m/s	Drop (%)	20 m/s	
1	Head	262	147	44	267	295
2	Shell	294	179	39	331	345
3	Head	249	139	44	271	295
	Shell	230	149	35	293	345

respectively, surpass admissible stress and strain. What is more, the higher the strain value is, the more serious the deformation of tank car. The above analyses indicate that a tank protection system developed using only traditional tank car designs will have difficulty reaching the two times protection goal. In conclusion, it is necessary to optimize the existing tank car design in derailment and collision conditions.

29.5 Structural Optimization

As illustrated in the previous section, the capacity of tank car body crashworthy need be enhanced. Various means of increasing the strength of the tank car are provided in this section. What is more, an effective measure reducing the collision energy is proposed. Then the section analyzes feasibility and practicality of the optimization design.

29.5.1 Optimization Design

This paper designs two kinds of scheme to improve tank car.

(1) This paper has carried on the target-oriented protection about the tank car, which refer to the results of finite element analysis. It can disclose the weak part of tank car structure to make for the optimization design. That is, weak section has a series of formed ribs, which increase the tank car strength to resist deformation. Figures 29.8 and 29.9 are the tank car models with added formed ribs.

(2) At present, MT-2 draft gear is applied in railway wagon in our country. The paper uses HM-1 draft gear, which provides superior performance compared to MT-2. HM-1 draft gear's impedance property is matched with the impact velocity, and the buffer capacity is increased by 60–100 %. The internal structural diagram is as shown in Fig. 29.10. Figure 29.11 shows the impedance property of both the HM-1 and MT-2 draft gears.

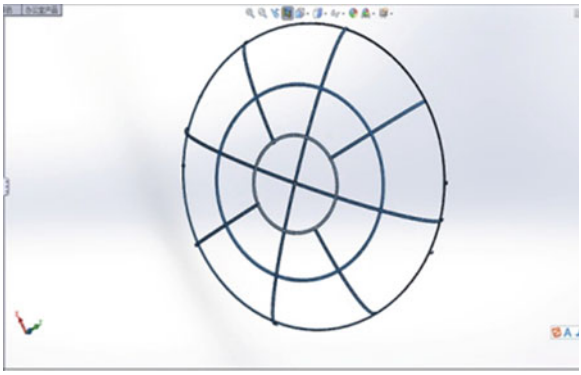


Fig. 29.8 The head formed ribs at the tank car

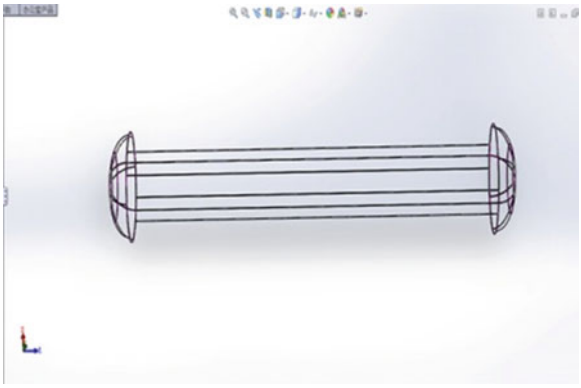


Fig. 29.9 The shell formed ribs at the tank car

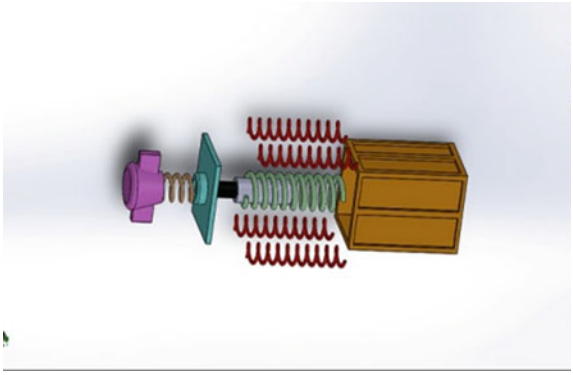


Fig. 29.10 The HM-1 internal structural diagram

29.5.2 Analysis After Optimization Design

In order to analyze the rationality after optimization, simulation experiment is carried out still with the velocity of 10 m/s. At the same time, the velocity is set to

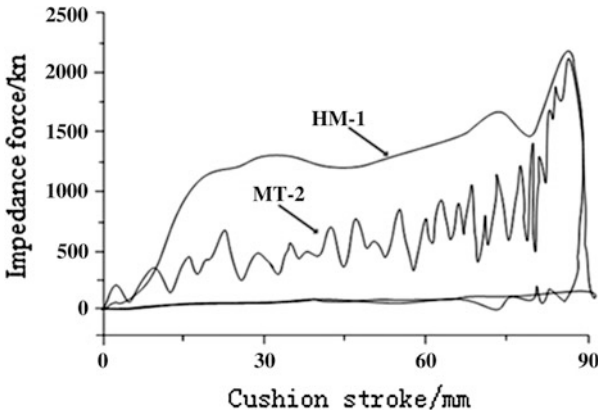


Fig. 29.11 The HM-1 and MT-2 impedance property

20 m/s to validate whether the structure can resist twice the velocity before or not. Once again, finite element analysis of the experiment is performed. Summarizing research results, the strain and stress nephogram are shown in Tables 29.1 and 29.2 in the above three conditions.

By the contrast of the stress and strain before and after optimization, the paper concludes as follows:

(1) When the relative impact velocity is 10 m/s, the maximum stress value at least have dropped by 35 % compared with before optimization;
(2) When the relative impact velocity is 20 m/s, the maximum stress value is all less than the allowable stress after optimization. The results indicate that the GQ70 body strength does satisfy the requirement of TB/T1335-1996. It turns out that the structure can maintain tank integrity for an impact speed with two times as the baseline tank car.

29.6 Summary

Based on the joint simulation of three softwares, the paper optimizes the design of rail tank car structure. Thus significant conclusions in this dissertation are mainly as follows:(1) According to the mathematical models founded in ADAMS, this paper imitates the movement of tank car during derailment, proving that three conditions are the most dangerous ones under the condition of derailment. Then the corresponding collision force spectra files are output. (2) The paper establishes finite element analysis model and get stress and strain nephogram under three collision conditions in ANSYS. (3) On these bases, the paper proposes two optimization ideas to ensure the tank car safety. The simulation results show that the proposed optimization strategy makes the collision speed of tank car increased by 2 times.

In summary, the co-simulation test can provide constructive comments for tank car optimization design, save development costs, and shorten development cycle. The method also supplies theoretical basis and technical support for the future vehicle operation and maintenance. What is more, the results can be used as a reference for tank car transportation safety studies. However, due to lack of author's time, some other interference such as bad climate have not been considered. In future more research needs to be done in the direction.

Acknowledgments This paper was supported by National High Technology Research and Development Program 863 of China (2012AA112001), Technology Research and Development Program of the Ministry of Railways (2013F022), and Fundamental Research Funds for the Central Universities of China (T14JB00380).

References

1. Paetsch CR (2005) Dynamic simulation of train derailments. Tufts Univ, Massachusetts
2. Tyrel D, Jeong D, Jacobsen K et al (2007) Improved tank car safety research. In: Proceedings of the 2007 ASME rail transportation division fall technical conference
3. Tyrel D, Jacobsen K, Talamini B, Carolan M (2007) Developing strategies for maintaining tank car integrity during train accidents. In: Proceedings of the 1st rail transportation division fall technical conference
4. Wang W et al (2011) Joint simulation of crashworthy train set based on finite element and multibody dynamic. J Tongji Univ (Nat Sci) 39(10): 1152–1156 (in chinese)
5. Jeong DY, Lyons ML et al (2007) Equation of motion for train derailment dynamics. In: Proceedings of the 2007 ASME rail transportation division fall technical conference
6. Gao X (2008) Study on modeling and simulation of crash problems of single railway vehicle based on ansys/ls-dyna. Beijing Jiaotong University, Beijing (in chinese)
7. Tang YH, Yu H et al (2008) Finite element analyses of railroad tank car head impacts. In: Proceedings of the 2008 ASME rail transportation division fall technical conference
8. Chen B (2013) Body finite element analysis of diesel tank car exported to Australia. Rolling Stock 51(4):5–8 (in chinese)

Chapter 30
Review of Knowledge Guidance in Intelligent Optimization Approaches

Lining Xing

Abstract The interaction between learning and evolution has recently become a popular research direction. Many scholars make use of knowledge to strengthen the guidance process in intelligent optimization methods. We review knowledge guidance in intelligent optimization approaches, which is normally carried out through artificial intelligence approaches and special knowledge models. Some researchers have also proposed algorithms with a double-layer evolution mechanism. These improved methods are able to discover some knowledge from previous iterations and to use the discovered knowledge to guide subsequent iterations.

Keywords Artificial intelligence · Intelligent optimization approaches · Knowledge guidance

30.1 Introduction

At present, more and more real-world optimization problems are formulated as large-scale global optimization (LSGO) problems. For example, in mission planning of multiple satellites, the number of tasks rises to about ten thousand, involving several hundred satellites. The large numbers of resources and tasks increases the difficulty of problem solving.

In the past few decades, many intelligent optimization approaches, i.e., genetic algorithms, simulated annealing, particle swarm optimization, and ant colony optimization, have been developed to solve LSGO problems effectively. Although these intelligent optimization approaches have demonstrated excellent optimization performance in solving small or medium-sized optimization problems, they still

L. Xing (✉)
College of Information System and Management, National University of Defense Technology, Changsha 410073, People's Republic of China
e-mail: xinglining@gmail.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_30

encounter serious challenges when applied to LSGO problems, including issues related to using several hundreds to thousands of decision variables.

How well intelligent optimization approaches handle LSGO problems remains an open issue. In recent years, research on employing intelligent optimization methods to solve LSGO problems has gathered momentum in both theoretical and empirical studies. For example, the interaction between learning and evolution has become a popular research direction. Many scholars have employed the knowledge to strengthen the guidance process in intelligent optimization methods. For this reason, we review knowledge guidance for intelligent optimization approaches.

30.2 Knowledge Guidance Through Artificial Intelligence Methods

Knowledge guidance for intelligent optimization approaches incorporating artificial intelligence is reviewed in this section. The artificial intelligence methods include tabu search [1], the culture algorithm [2] and machine learning [3], amongst others. In a specific implementation, some common important rules are extracted from the optimization process and studied using taboo search, the cultural algorithm, and machine learning. Individual evolution is controlled based on these rules obtained through learning.

To balance the positive effects of the selection operation and the destructive effects of crossover and mutation operations in the evolutionary process, Fan et al. adopted induction learning to guide the evolutionary process [4]. In other words, they extracted rules that reflect the errors and success of past evolution from its historic process through induction learning and then used these to guide subsequent evolution. This can speed up evolution while avoiding repetitive errors. Experiments based on function and layout optimizations have verified the effectiveness of this method.

Cao et al. learnt from the life cycle of individual evolution and proposed an ecological evolution model guided by this life cycle [5]. The algorithm based on the model sets up corresponding guided operators in each phase of an individual life cycle, thereby guiding the evolution of an individual to evolve based on its ecological features throughout its lifetime. Experimental results have verified the superiority of this approach.

Cen et al. proposed the grouping particle swarm algorithm based on the knowledge space to improve search accuracy [6]. The algorithm uses the K-means method to group particle swarms, utilizes a smaller maximum flight speed to strengthen the local search ability of particles in the group, incorporates the concept of a knowledge space into the grouping, and guides particles to search in a better solution space using particles in the knowledge space.

Li et al. adopted expert knowledge to assist in searching for an optimal solution using genetic algorithms and applied an improved genetic algorithm to optimization

of reactive power planning [7]. According to the expert knowledge, the ability to search for a local optimal solution of the genetic algorithm can be improved by artificial regulation of the local active control variable set, which is formed dynamically by a few select individuals, to control reactive power/voltage of a local substation and plant. Application in a real-life system shows that the genetic algorithm combined with expert knowledge is able to find a global optimal solution more effectively.

Chai added domain knowledge to ant colony optimization by means of a form of the taboo connection set, ensuring that the mutex information is calculated only once because it can be excluded from the main search cycle [8]. A case analysis demonstrated the effectiveness of this strategy.

The disadvantage of guiding intelligent optimization approaches using artificial intelligence methods is that it is very difficult to find common important rules. Besides, since these rules are local, the approaches do not consider corresponding rules that change along with the individual's environment [9].

30.3 Knowledge Guidance Using Specific Knowledge Models

When guiding intelligent optimization algorithms by means of specific knowledge models, the basic form of knowledge is determined at the first operation of the intelligent optimization algorithm, relevant knowledge is constantly adjusted together with the evolution of the algorithm according to the fixed rules, and then the evolution of subsequent individuals is guided by the obtained knowledge. Guidance of intelligent optimization algorithms using specific knowledge models can be interpreted as the interaction between evolution and learning [10–15]. Many studies suggest that the interaction between evolution and learning can greatly improve the performance of intelligent optimization approaches [16]. Scholars have independently adopted version space [17], case-based memory [18], Q-learning system [19, 20], and AQ learning system [21] to realize interaction between evolution and learning.

30.3.1 Knowledge Guidance Using Memory

Many scholars have tried to employ memories to realize the interaction between evolution and learning. Chung et al. defined the good features of individuals as beliefs and kept them in memories so as to continuously improve subsequent individuals [22]. Branke kept some better individuals in memories to improve subsequent individuals [23].

Gantovnik et al. used memories to realize the interaction between evolution and learning in genetic algorithms and applied an improved genetic algorithm to solve the optimum design of mixed variables [24, 25]. Louis et al. adopted a genetic algorithm with memories to solve the traveling salesman problem [26], while Yang employed the immigration method based on memories to realize interaction between evolution and learning in genetic algorithms and applied an improved genetic algorithm to solve dynamic optimization problems [27, 28].

Su et al. employed an immunological memory method to realize interaction between evolution and learning in the ant colony algorithm and applied an improved ant colony algorithm to solve the weapon target assignment problem [29]. Acan utilized external memory to realize interaction between evolution and learning in the ant colony algorithm [30], and by adding the function of partial permutations into the external memories, further improved the efficiency of the ant colony algorithm [31]. Shamsipur et al. also employed external memories to realize interaction between evolution and learning in the ant colony algorithm [32].

30.3.2 Knowledge Guidance Using Cases

Many scholars have used cases to realize interaction between evolution and learning [33, 34]. Louis et al. employed a case-reasoning method to select better features from case memories to improve subsequent individuals [18]. Louis et al. also employed a case-injected method to realize interaction between evolution and learning in genetic algorithms and applied an improved genetic algorithm to solve the traveling salesman problem [33]. Rasheed et al. employed case-based learning to realize interaction between evolution and learning [34], while Babbar-Sebens et al. proposed a case-based micro interactive genetic algorithm, which realizes interaction between evolution and learning through case memories and case reasoning [35].

30.3.3 Knowledge Guidance via the Learnable Evolution Model

The learnable evolution model mainly employs a machine learning method to guide the evolutionary process. Coletti et al. carried out a preliminary study on the learnable evolution model [36], while Wojtusiak studied constrained optimization problems in this model [37]. Kaufman et al. [38], Jourdan et al. [39], and Domanski et al. [40] applied the learnable evolution model, respectively, to the design of heat exchangers, the design of multiobjective water systems, and optimized design problems. Currently, the latest version of the learnable evolution model is LEM3, which was employed by Wojtusiak et al. to solve complex function optimization problems [41].

In recent years, more and more scholars have begun to use the learnable evolution model to realize interaction between evolution and learning in intelligent optimization approaches [16–21]. Michalski et al. adopted the learnable evolution model to realize interaction between evolution and learning in intelligent optimization approaches by summarizing the latest developments in the model [42, 43]. In the learnable evolution model constructed by Michalski [21], the interaction between evolution and learning is mainly implemented by a machine learning method. Ho et al. used a learnable genetic architecture to realize interaction between evolution and learning [16]. Wojtusiak designed an LEM3 system that can be used for various intelligent optimization approaches [44].

30.3.4 Knowledge Guidance Using Neural Networks

In an attempt to address the lack of clear guidance in genetic algorithms' individual evolution, Gu et al. proposed a knowledge model-based genetic algorithm, in which certain knowledge from the information of two current generations of individuals is acquired through the learning function of a neural network to control the evolution of certain individuals in the next generation [9]. The algorithm retains genetic operators as well as utilizing a neural network to construct the corresponding knowledge model used to guide the evolution of individuals. This not only allows the improved genetic algorithm to retain the strong global random search ability of genetic algorithms, but also gives it the self-learning ability and strong robustness of neural networks.

When guiding intelligent optimization approaches using knowledge models, existing research is limited to strengthening the guidance of a specific method using a specific model and aimed at a specific problem. The research results show a lack of systematicness, generality and expandability, and thus this approach has not been studied as a general method.

30.3.5 Knowledge Guidance Using a Knowledge Model

Based on existing intelligent optimization approaches, Xing et al. designed and realized various knowledge-based intelligent optimization approaches [45], making use of integrated modeling in which an intelligent optimization model is combined with a knowledge model. The intelligent optimization model searches the feasible space of the problem to be optimized using a neighborhood search strategy, while the knowledge model discovers useful knowledge from earlier stages of the optimization process and then uses the knowledge to guide the subsequent optimization process of the intelligent optimization model. The effective combination of an intelligent optimization model and knowledge model greatly improves the optimization performance of knowledge-based intelligent optimization approaches.

The results of solving the function optimization problem [46], asymmetric traveling salesman problem [47], double-layer capacitated arc routing problem [48, 49], and flexible job shop scheduling problem [50, 51] using the knowledge-based intelligent optimization approach are all relatively satisfactory.

30.4 Conclusions

In this paper we reviewed knowledge-based guidance methods for intelligent optimization approaches. Such knowledge-based guidance can be realized using either artificial intelligence methods or a specific knowledge model. Based on this review, we summarize further research directions as follows:

(1) Extend the types of knowledge. Combined with practical engineering problems, we should try to discover additional types of knowledge that could effectively improve the performance of the optimization process. Also, by abstracting the experiential knowledge of experts to practical problems, we could employ this knowledge to guide the optimization process effectively so as to enhance optimization efficiency as much as possible.
(2) Adopt new knowledge mining techniques. We should consider employing advanced knowledge mining through machine learning or data mining to obtain useful knowledge from the optimization process. By adopting machine learning or data mining to discover knowledge from the optimization process, it may, on the one hand, be possible to extract some very useful knowledge in an efficient manner. On the other hand, this may add to the computational complexity of the optimization process. This trade-off needs to be further investigated.

Acknowledgments This research is supported by the National Natural Science Foundation of China (No. 71331008 and 71101150), the Program for New Century Excellent Talents in University, Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201492), the Youth Training Program for Innovation and Entrepreneurship Platform of Science and Technology at Hunan Province, the Outstanding Youth Fund Project of Hunan Provincial Natural Science Foundation, the Top-notch Innovative Talents Training Plan of National University of Defense Technology and the Outstanding Youth Fund Project of National University of Defense Technology.

References

1. Holland JH (1975) Adaptation in natural and artificial systems. MIT Press, Cambridge
2. Cavaretta MJ (1994) Using a culture algorithm to control genetic operators. In: Proceedings of the third annual conference on evolutionary programming, World Scientific Pub. Co. Inc., San Diego, California, pp 24–26

3. Sebag M (1996) Controlling evolution by means of machine learning. In: Proceedings of the fifth annual conference on evolutionary programming, MIT Press, San Diego, California, pp 57–66
4. Fan L, Ruan HC, Jiao Y (2001) Conduct evolution using induction learning. J China Univ Sci Technol 31(5):565–634 (in Chinese)
5. Cao XB, Xu K, Zhang J (2000) Ecological evolution model guided by life period. J Software 11(6):823–828 (in Chinese)
6. Cen YS, Xiong FM, Zeng BQ (2010) Grouping particle swarm optimization algorithms based on knowledge space. Comput Eng Des 31(7):1562–1565 (in Chinese)
7. Li YN, Zhang LZ, Shu J et al (2001) Application of expert knowledge adopted genetic algorithm to optimization of reactive power planning. Power Syst Technol 25(7):14–17
8. Chai XL (2010) Ant swarm planning algorithm optimized by domain knowledge. Comput Eng Appl 46(14):17–19 (in Chinese)
9. Gu H, Gong YC, Zhao ZX (2000) A knowledge model based genetic algorithm. Comput Eng 26(5):19–20 (in Chinese)
10. Chi HM, Ersoy OK, Moskowitz H et al (2007) Modelling and optimizing a vendor managed replenishment system using machine learning and genetic algorithms. Eur J Oper Res 180 (1):174–193
11. Chou FD (2009) An experienced learning genetic algorithm to solve the single machine total weighted tardiness scheduling problem. Expert Syst Appl 36(2):3857–3865
12. Ho WH, Tsai JT, Lin BT et al (2009) Adaptive network-based fuzzy inference system for prediction of surface roughness in end milling process using hybrid taguchi-genetic learning algorithm. Expert Syst Appl 36(2):3216–3222
13. Hong Y, Kwong S (2008) To combine steady-state genetic algorithm and ensemble learning for data clustering. Pattern Recogn Lett 29(9):1416–1423
14. Liu F, Zeng GZ (2009) Study of genetic algorithm with reinforcement learning to solve the tsp. Expert Syst Appl 36(2):6995–7001
15. Sim KM, Guo YY, Shi BY (2009) Bayesian learning and genetic algorithm for supporting negotiation with incomplete information. IEEE Trans Syst Man Cybern B Cybern 39(1): 198–211
16. Ho NB, Tay JC, Lai EMK (2007) An effective architecture for learning and evolving flexible job-shop schedules. Eur J Oper Res 179(2):316–333
17. Reynolds RG (1994) An introduction to cultural algorithms. In: Proceedings of the third annual conference on evolutionary programming, World Scientific, River Edge, NJ, pp 131–139
18. Louis SJ, McDonnell J (2004) Learning with case-injected genetic algorithms. IEEE Trans Evol Comput 8(4):316–328
19. Juang GF, Lu CM (2009) Ant colony optimization incorporated with fuzzy Q-learning for reinforcement fuzzy control. IEEE Trans Syst Man Cybern Part A Syst Hum 39(3):597–608
20. Kamall K, Jiang LJ, Yen J et al (2007) Using Q-learning and genetic algorithms to improve the efficiency of weight adjustments for optimal control and design problems. J Comput Inf Sci Eng 7(4):302–308
21. Michalski RS (2000) Learnable evolution model: evolution process guided by machine learning. Mach Learn 38(1):9–40
22. Chung CJ, Reynolds RG (1996) A testbed for solving optimization problems using cultural algorithm. In: Proceedings of the fifth annual conference on evolutionary programming, MIT Press, Cambridge, pp 225–236
23. Branke J (1999) Memory-enhanced evolutionary algorithms for dynamic optimization problems. In: Proceeding of congress on evolutionary computation, IEEE Press, Piscataway, pp 1875–1882
24. Gantovnik VB, Anderson-Cook CM, Gurdal Z et al (2000) A genetic algorithm with memory for mixed discrete-continuous design optimization. Comput Struct 81(20):2003–2009
25. Gantovnik VB, Gurdal Z, Watson LT (2002) A genetic algorithm with memory for optimal design of laminated sandwich composite panels. Compos Struct 58(4):513–520

26. Louis S, Li G (1997) Augmenting genetic algorithms with memory to solve traveling salesman problems. In: Proceedings of the joint conference on information sciences, Duke University Press, Durham, North Carolina, pp 108–111
27. Yang SX (2008) Genetic algorithms with memory and elitism-based immigrants in dynamic environments. Evol Comput 16(3):385–416
28. Yang SX (1997) Memory-based immigrants for genetic algorithms in dynamic environments. In: Proceedings of genetic and evolutionary computation, IEEE Press, Washington DC, pp 1115–1122
29. Su M, Qian H, Wang XF (2004) Immune memory-based ant colony algorithm for weapon-target assignment solution. Comput Eng 34(4):215–217 (in Chinese)
30. Acan A (2004) An external memory implementation in ant colony optimization. In: Proceedings of the forth international workshop on ant colony optimization and swarm intelligence (lecture notes in computer science), Springer Press, Brussels, Belgium, vol 3172, pp 73–82
31. Acan A (2005) An external partial permutations memory for ant colony optimization. In: Proceedings of the fifth European conference on evolutionary computation in combinatorial optimization (lecture notes in computer science), Springer Press, Lausanne, Switzerland, vol 3448, pp 1–11
32. Shamsipur M, Zare-Shahabadi V, Hemmateenejad B et al (2009) An efficient variable selection method based on the use of external memory in ant colony optimization, application to QSAR/QSPR studies. Anal Chim Acta 27(1):39–46
33. Louis SJ, Li G (2000) Case injected genetic algorithms for traveling salesman problems. Inf Sci 122(2–4):201–225
34. Rasheed K, Hirsh H (1997) Using case-based learning to improve genetic-algorithm-based design optimization. In: Proceedings of the seventh international conference on genetic algorithms, IEEE Press, East Lansing, MI, pp 513–520
35. Babbar-Sebens M, Minsker B (2010) A case-based micro interactive genetic algorithm for interactive learning and search: methodology and application to groundwater monitoring design. Environ Model Softw 25(10):1176–1187
36. Coletti M (2002) A preliminary study of learnable evolution methodology implemented with C4.5. In: Proceedings of the congress on evolutionary computation, IEEE Press, Honolulu, HI, pp 588–593
37. Wojtusiak J (2006) Initial study on handling constrained optimization problems in learnable evolution model. In: Proceedings of the graduate student workshop at genetic and evolutionary computation conference, IEEE Press, Seattle, WA, pp 1–7
38. Kaufman KA, Michalski RS (2000) Applying learnable evolution model to heat exchanger design. In: Proceedings of the seventeenth national conference on artificial intelligence and twelfth conference on innovative applications of artificial intelligence, MIT Press, Austin, TX, pp 1014–1019
39. Jourdan L, Corne D, Savic D, et al (2005) Preliminary investigation of the learnable evolution model for faster/better multi-objective water systems design. Proceedings of the third international conference on evolutionary multi-criterion optimization. Lecture notes in computer science, vol 3410. Springer, Guanajuato, Mexico, pp 841–855
40. Domanski PA, Yashar D, Kaufman K et al (2004) An optimized design of finned-tube evaporators using the learnable evolution model. Int J Heat Vent Air Cond Refrig Res 10 (2):201–211
41. Wojtusiak J, Michalski RS (2006) The LEM3 implementation of learnable evolution model and its testing on complex function optimization problems. In: Proceedings of genetic and evolutionary computation conference, IEEE Press, Seattle, WA, pp 1281–1288
42. Michalski RS, Wojtusiak J, Kaufman KA (2006) Intelligent optimization via learnable evolution model. In: Proceedings of the eighteenth IEEE international conference on tools with artificial intelligence, IEEE Press, Arlington, VA, pp 332–335
43. Michalski RS, Wojtusiak J, Kaufman KA (2007) Progress report on learnable evolution model. George Mason Univ, Fairfax

44. Wojtusiak J (2009) The LEM3 system for multi-type evolutionary optimization. Comput Inf 28(2):225–236
45. Xing LN, Chen YW (2010) Research on the knowledge-based intelligent approaches. National University of Defence Technology Press, Changsha
46. Xing LN, Chen YW, Yang KW (2009) A novel mutation operator based on the immunity operation. Eur J Oper Res 197(2):830–833
47. Xing LN, Chen YW, Yang KW et al (2008) A hybrid approach combining an improved genetic algorithm and optimization strategies for the asymmetric traveling salesman problem. Eng Appl Artif Intell 21(8):1370–1380
48. Xing LN, Rohlfshagen P, Chen YW et al (2010) An evolutionary approach to the multi-depot capacitated arc routing problem. IEEE Trans Evol Comput 14(3):356–374
49. Xing LN, Rohlfshagen P, Chen YW et al (2011) A hybrid ant colony optimization algorithm for the extended capacitated arc routing problems. IEEE Trans Syst Man Cybern B Cybern 41 (4):1110–1123
50. Xing LN, Chen YW, Wang P et al (2010) A knowledge-based ant colony optimization for flexible job shop scheduling problems. Appl Soft Comput 10(3):888–896
51. Xing LN, Chen YW, Yang KW (2011) Multi-population interactive coevolutionary algorithm for flexible job shop scheduling problems. Comput Optim Appl 48(1):139–155

Chapter 31
A Hybrid Optimization Algorithm for Extreme Learning Machine

Bin Li, Yibin Li and Xuewen Rong

Abstract In this paper, a learning algorithm based on particle swarm optimization method (PSO) and a novel heuristic optimization method of gravitational search algorithm (GSA) for extreme learning machine (ELM) is proposed in terms of improving the generalization performance of single hidden-layer feed-forward neural networks, which is called as PSOGSA-ELM learning algorithm. The proposed learning algorithm uses a hybrid approach of PSO and GSA to select the optimal hidden biases and input weights of ELM, and then the output weights of ELM is analytically determined by the Moore-Penrose generalized inverse. The performance of the proposed algorithm is verified by regression and classification benchmark problems and is compared with PSO–ELM, GSA–ELM, and the original ELM learning algorithms, simulation results show that the proposed algorithm performs equal to or better than the other algorithms in terms of generalization performance and has good convergence speed.

Keywords Particle swarm optimization · Gravitational search algorithm · Extreme learning machine

31.1 Introduction

In recent years, feed-forward neural networks are commonly used for regression (function approximation) and classification problems. There are many learning algorithms for training the architecture of those neural networks, such as, gradient descent-based algorithms (e.g., Back Propagation (BP) method), standard optimization method-based algorithms (e.g., Support Vector Machine, SVM). For improving

B. Li (✉)
School of Science, Qilu University of Technology, Jinan 250353, China
e-mail: ribbenlee@126.com

Y. Li · X. Rong
School of Control Science and Engineering, Shandong University, Jinan 250061, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_31

the learning speed and generalization performance of the above traditional methods, extreme learning machine (ELM) was originally developed by Huang et al. can learn distinct observations with fast learning speed and good universal approximation performance [1], which is used in many applications.

However, In the ELM learning algorithm, the hidden biases and input weights are set randomly, those parameters are not optimal for the ELM, and the algorithm tends to require more hidden nodes (neurons) for improving the generalization performance than gradient descent-based algorithms [2]. Therefore, many improved algorithms were proposed based on heuristic principles and evolutionary computation methods [3, 4].

Recently, a new heuristic optimization algorithm named gravitational search algorithm (GSA) was proposed in 2009 by Rashedi et al. [5]. Based on the Newtonian gravity, this algorithm has the ability to find the most optimal results of problems solving. The convergence rate of GSA can be speeded up based on the initial parameters modulated suitably, which can fast arrive at the global optimal solution. However, the GSA has the disadvantage of easily getting into a local optimal solution. In order to overcome the drawbacks of GSA, some researcher proposed improved GSA algorithms based on evolutionary approaches, such as hybrid GSA and particle swarm optimization (PSO) method [6], hybrid GSA–genetic algorithm [7], and so on.

Compared with genetic algorithm, the PSO method is becoming very popular due to its simple evolutionary operators and ability to quickly converge to a reasonably good solution [8]. On the other hand, the basic idea of PSO and GSA is similar. Therefore, the hybrid PSO and GSA named PSOGSA algorithm of integrating the ability of exploration of PSO with the ability of exploitation of GSA are used commonly for improving the performance of optimization algorithms [9, 10].

In this paper, in order to improve the learning performance of the ELM, a hybrid approach based on PSO and GSA is used to optimize the parameter of hidden biases and input weights of ELM. Once the optimal training parameters are obtained, then the output weight can be determined analytically. The proposed algorithm uses the global approximation ability of ELM and the initial parameters optimized by means of PSOGSA algorithm. As the newly proposed learning algorithm is implemented with respect to some benchmark problems, the algorithm can be used to obtain better generalization performance with good convergence speed, although it takes little longer time for training process than ELM learning algorithm.

The rest of the paper is organized as follows. In Sect. 31.2, we introduce the basic idea of ELM learning algorithm. Section 31.3 proposes the improved algorithm of ELM based on the PSOGSA method; the efficiency and effectiveness of the proposed algorithm is demonstrated based on some regression and classification benchmark problems in Sect. 31.4. Finally, the concluding results are offered in Sect. 31.5.

31.2 Brief of Extreme Learning Machine

ELM was a batch learning algorithm for single hidden-layer feed-forward neural networks (SLFNs), which randomly selects the hidden parameters and input weights; and analytically determines the output weights, which is easily implemented and obtains the smallest norm of output weights and good generalization performance. The output function of SLFNs with $\tilde{N}$ hidden neurons can be represented by

$$f_{\tilde{N}}(\mathbf{x}) = \sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i, b_i, \mathbf{x}), \mathbf{x} \in R^n, \mathbf{w}_i \in R^n, \beta_i \in R^m \tag{31.1}$$

where $g(\mathbf{w}_i, b_i, \mathbf{x})$ is the output of the ith hidden neurons corresponding to the input $\mathbf{x}$, and $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ represents the weight vector connecting the ith hidden neuron and the output neurons [11].

For a set of N arbitrary samples $(\mathbf{x}_j, \mathbf{t}_j)$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]$ and $\mathbf{t}_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]$. The ELM learning algorithm with $\tilde{N}$ hidden neurons and activation function $g(\mathbf{w}_i, b_i, \mathbf{x})$ can be computed as

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i, b_i, \mathbf{x}_j) = \mathbf{t}_j \quad j = 1, \cdots, N \tag{31.2}$$

The above N equation of formula (31.2) can be written in matrix format compactly as

$$\mathbf{H}\beta = \mathbf{T} \tag{31.3}$$

where $\mathbf{H}$ is the output matrix of the hidden layer, and the ith column of $\mathbf{H}$ is the ith hidden neuron's output vector with respect to inputs $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and the j th row of matrix $\mathbf{H}$ is the output vector of the hidden layer with respect to input $\mathbf{x}_j$. Usually, the ELM learning algorithm with $\tilde{N}$ hidden neurons (such $\tilde{N} \leq N$) and any infinitely differential activation function $g(x)$, can approximate the N distinct arbitrary samples with zero errors. Then, the smallest norm of the least squares solution of the ELM learning algorithm for $\mathbf{H}\beta = \mathbf{T}$ is $\hat{\beta} = \mathbf{H}^{+}\mathbf{T}$, where $\mathbf{H}^{+}$ is the Moore-Penrose generalized inverse of the hidden-layer output matrix $\mathbf{H}$.

Thus, the simple learning approach for SLFNs with good generalization performance and fast learning speed called ELM can be summarized as follows:

Given a training set $\left\{(\mathbf{x}_j, \mathbf{t}_j) \middle| \mathbf{x}_j \in R^n, \mathbf{x}_j \in R^m, j = 1, \ldots, N\right\}$, activation function $g(x)$, and the number of hidden node $\tilde{N}$,

Step 1 Generate the values for parameters $\mathbf{w}_i$ and b_i of the hidden neurons randomly.
Step 2 Calculate the hidden-layer output matrix $\mathbf{H}$.
Step 3 Calculate the output weight $\beta : \beta = \mathbf{H}^{+}\mathbf{T}$.

Many papers based on ELM have been appeared since its introduction by Huang, and the ELM algorithm with fast learning speed and good generalization performance has been widely used for many applications [12], such as time-series prediction, pattern recognition, power systems and data analytics, image processing, etc.

31.3 The PSOGSA-ELM Algorithm

In the ELM algorithm, the input weights and hidden biases are tuned randomly. Therefore, much time of learning is saved. However, the tuned parameters of the neural networks are not optimal for the problems solving; and besides, the ELM algorithm may require more hidden neurons for improving the performance of SLFNs. Unfortunately, high structural complexity of the neural networks may reduce the generalization ability and also make it respond slowly to the training and testing data [13]. Therefore, to eliminate possible nonoptimal parameters of input weights and hidden neurons and to create more compact neural networks for improving the generalization performance, a hybrid PSO and GSA is used to select the optimal input weights and hidden biases of SLFNs.

GSA is a heuristic optimization method based on the Newtonian gravity. In GSA, the searching agents are a collection of masses, agents are considered as objects, and their performance is measured by their masses and all this agents (masses) attract each other by the gravity force, and this force causes a global movement of all agents toward the agents with heavier masses (optimal solution) [5].

Although GSA is a memory-less algorithm and works efficiently with good convergence rate, GSA suffers from slow searching speed due to the presence of heavier masses in the last iterations. For avoiding the above disadvantage of GSA algorithm, some researchers have provided a lot of modifications for the GSA algorithms [14]. In order to speed up the iteration process, Chen et al. used random key encoding scheme to optimize GSA [15]. In [6, 9], Mrijalili et al. proposed an algorithm using low-level co-evolutionary heterogeneous based on the PSO and the GSA algorithms. This hybrid PSOGSA algorithm combines the ability of social thinking in PSO with the local searching ability in GSA. In the PSOGSA algorithm, the velocity of PSO algorithm with the self-adaptive inertia weight has been modified as

$$v_i^k(t+1) = w \cdot v_i^k(t) + c_1 \cdot \text{rand}() a_i^k(t) + c_2 \cdot \text{rand}() \left(g_{\text{best}}(t) - x_i^k(t)\right) \qquad (31.4)$$

where w is the weighting function, c_1 and c_2 are the weighting factors, rand() is a random number in the interval 0 and 1, $a_i^k(t)$ is the acceleration of agent i at iteration t and g_{best} is the best solution until now.

In the PSOGSA algorithm, the masses near good solutions try to attract the other masses of exploring different parts of the space, the masses nearing the good solution move very slowly, and in this case, g_{best} helps to exploit the global best solution and save the best solution found so far, which is accessible at any time. The above ability of PSOGSA makes the algorithm powerful enough to solve a wide range of optimization problems [6, 9].

Based on the idea of PSOGSA algorithm and ELM learning algorithm, a novel single hidden feed-forward neural networks learning algorithm called PSOGSA-ELM is proposed. In this algorithm, the PSOGSA algorithm is used to optimize the parameters of input weights and hidden biases of neural networks, when the optimal parameters are found, then the output weights of the neural networks are determined analytically based on the least squares solution.

Therefore, the individual (agent or mass) in the search space of the proposed algorithm is composed of a set of input weights, hidden biases, and is defined as

$$\theta \in [w_{11}, w_{12}, \ldots w_{1\tilde{N}}, w_{21}, w_{22}, \ldots w_{2\tilde{N}}, \ldots w_{n1}, w_{n2}, \ldots w_{n\tilde{N}}, \ldots, b_1, b_2, \ldots b_{\tilde{N}}] \tag{31.5}$$

Based on the optimization ability of PSOGSA and the universal approximation ability of ELM learning algorithm, the detailed steps of the PSOGSA-ELM algorithm are as follows:

Step 1 the initial generation (swarm) of the proposed algorithm is randomly generated. Each mass in the generation is composed of a set of parameters as is shown in Eq. (31.5). All components in the mass are initialized randomly in the range of $[-1, 1]$.

Step 2 for each mass, the output weights are computed based on the equation $\beta = \mathbf{H}^{+}\mathbf{T}$, and the fitness of each mass is evaluated according to the approximation error (the root mean squared (RMSE) error in regression problems, and the classification accuracy in pattern recognition problems) of output in neural networks.

Step 3 with the fitness value of all masses, the acceleration a_i of a mass and the best solution until now g_{best} for the generation are computed according to the equations of PSO and GSA algorithms, respectively.

Step 4 each mass updates its position based on the Eq. (31.4), and the new swarm θ is generated.

Step 5 the above optimization process is iterated repetitively until the maximal iteration number is met.

Thus, the optimal parameters of input weights and hidden biases of the ELM learning algorithm are obtained; and then, the optimal ELM learning algorithm with good generation performance is applied to the testing samples.

31.4 Simulation and Result Analysis

In this section, the performance of the proposed PSOGSA-ELM learning algorithm is compared with PSO–ELM, GSA–ELM, and the original ELM on some benchmarks. Three regression and three classification benchmark problems are chosen to evaluate the performance of the different algorithms and the parameters specification of the benchmark problems are shown in Table 31.1. The attributes of the datasets are normalized into $[-1, 1]$ and the training and testing data of the datasets are reshuffled at each trial of simulation.

The simulation results have been conducted in MATLAB R2011a platform running in a Pentium 5, 2.5 GHZ CPU and 12 GB RAM. For each heuristic optimization approach in the regression benchmarks, the maximum iteration number is set to 20 and the population number is set to 50. However, in the classification problems, the maximum iteration number and the population number is set different combination for better comparing the performance of different algorithms.

31.4.1 Performance Comparison on Regression Problems

In this section, the performance comparison in terms of CPU time, Testing error on regression problems with different initial parameter G_0 is shown in Table 31.2. As shown in Table 31.2, the ELM learning algorithms based on optimization approaches have better generalization performance than the original ELM. From the Table, we can also conclude that the PSOGSA-ELM learning algorithm has better generalization ability with proper initial parameters of G_0 than the PSO–ELM and GSA–ELM in most cases.

As observed from Table 31.2, it is obvious that the testing RMSE is decreasing with the decreasing of the value of G_0, which means that the PSOGSA learning algorithm can improve the performance with little value of G_0. Therefore, the gravitational constant initial value G_0 is important for improving the performance of GSA learning algorithm in regression problems and should be set properly based on the problems solving.

Table 31.1 Parameters specification of the benchmark problems

Datasets	Names	Attributes	Class	Training data	Testing data
Regression	Auto-Mpg	7	1	300	98
	Abalone	8	1	3500	617
	Mackey Glass	4	1	4000	500
Classification	Image segmentation	19	7	1900	410
	Satellite image	36	7	3217	3218
	Diabetes	8	2	576	192

Table 31.2 Performance comparison of different algorithms on regression problems

Algorithms (swarm = 50, iteration = 20)	Auto-mpg (hidden nodes 15)		Abalone (hidden nodes 20)		Mackey Glass (hidden nodes 40)	
	G_0, C	Testing error	G_0, C	Testing error	G_0, C	Testing error
ELM	–	0.0770	–	0.0770	–	0.0552
PSO–ELM	–	0.0670	–	0.0746	–	0.0395
GSA–ELM	(200, 0)	0.0712	(200, 0)	0.0763	(200, 0)	0.0447
PSOGSA-ELM	(200, 0)	0.0700	(200, 0)	0.0757	(200, 0)	0.0446
GSA–ELM	(1, 0)	0.0699	(1, 0)	0.0749	(1, 0)	0.0394
PSOGSA-ELM	(1, 0)	0.0626	(1, 0)	0.0736	(1, 0)	0.0384
PSOGSA-ELM	(0.01, 0)	0.0624	(0.01, 0)	0.0728	(0.01, 0)	0.0387

31.4.2 Performance Comparison on Classification Problems

In this section, three classical benchmark problems are used for comparison of different algorithms. The performance of pattern classification problems is justified by classification rate (testing accuracy). Different values of the parameter G_0 are selected for comparing the performance of different algorithms in the section.

However, less value of the initial parameter G_0 may cause the ill-conditioned of matrix of output weights of the SLFNs. Therefore, for avoiding the singular of the hidden-layer output matrix $\mathbf{H}$ of the single hidden-layer feed-forward neural networks, it is clear that one can choose a big initial value G_0. This selecting strategy will decrease the generalization performance of the PSOGSA-ELM learning algorithm. On the other hand, one can add a regularization term as shown in reference [16]. In this situation, the output weight is calculated as

$$\beta = (1/C + \mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H} \tag{31.6}$$

Therefore, the PSOGSA-ELM learning algorithm with the combination of regularization term C and initial value G_0 is simulated in this paper. For simplicity, the regularization term C chooses 0.1 in order to avoid the singular of matrix $\mathbf{H}$.

It is known that the performance of PSO and GSA usually depend closely on the maximum iteration number max*iter* and the population number $\hat{N}$. Table 31.3 shows the simulation results with the combination of parameters max*iter* and $\hat{N}$. It is easy to find that the PSOGSA-ELM can achieve better generalization performance than other algorithms in most case. As observed from the Table, it can be also concluded that the classification rate tends to increasing when the initial gravitational value G_0 is decreasing.

Table 31.3 Performance comparison of different algorithms on classification problems with max *iter* = 100 and $\hat{N} = 50$

Algorithms (swarm = 50, iteration = 100)	Image segmentation (hidden nodes 40)		Satellite image (hidden nodes 60)	
	Testing accuracy	G_0, C	Testing accuracy	G_0, C
PSO–ELM	0.9373	–	0.8773	–
GSA–ELM	0.9097	(200, 0)	0.8619	(200, 0)
PSOGSA-ELM	0.9220	(200, 0)	**0.8777**	(200, 0)
GSA–ELM	0.9189	(1, 0)	0.8703	(1, 0)
PSOGSA-ELM	**0.9440**	(1, 0)	**0.8825**	(1, 0)
PSOGSA-ELM	0.9560	(0.01, 0)	0.8871	(0.01, 0)

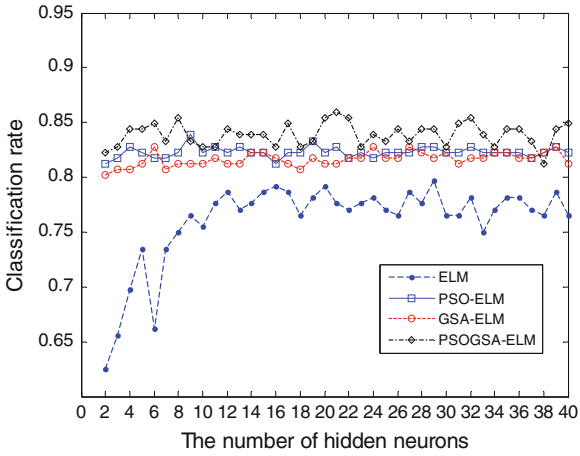


Fig. 31.1 The convergence *curves* of different algorithms based on different hidden nodes

The convergence curves of different algorithms in terms of the Diabetes classification benchmark problem based on different hidden nodes and iterative times with $G_0 = 1$ are shown in Figs. 31.1 and 31.2. In Fig. 31.1, the swarm number and iterative number are set 100 and 50, respectively. And in Fig. 31.2, the swarm number is set to 100. As shown in Fig. 31.1, for all hidden neurons, the PSOGSA-ELM learning algorithm can improve the classification rate with little value of G_0.

Moreover, as shown in Fig. 31.2, the classification rate of PSOGSA-ELM learning algorithm is also better than all of other algorithms with the increasing of the number of iterative time. The more the number of iterative time, the better the generalization performance, which mean that the convergence speed of the proposed algorithm is better than the performance of PSO–ELM and GSA–ELM algorithms.

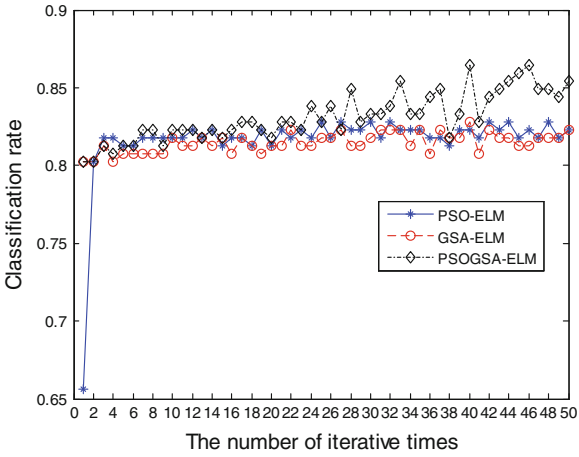


Fig. 31.2 The convergence *curves* of different algorithms based on different iterative times

31.5 Conclusion

In this work, a novel training algorithm, PSOGSA-ELM, has been developed based on the ELM and PSOGSA optimization technique. In this algorithm, the parameters of input weights, hidden biases are adjusted by the PSOGSA algorithm.

In conclusion, it is clear from the results that the generalization performance in terms of PSOGSA learning algorithm was found to be significantly improved in terms of the benchmark problems. In particular, for the PSOGSA learning algorithm, with too large G_0, the algorithm cannot achieve the refinement of the input weights and hidden biases, while with too small G_0, the output of the hidden weights may be ill-conditioned and cannot be solved by the Moore-Penrose generalized inverse. Therefore, how to choose a suitable initial value is still an unresolved theoretic problem for improving the performance of the proposed algorithm.

Acknowledgments This work is supported by National Natural Science Foundation of China (61233014, 61305130), Shandong Provincial Natural Science Foundation (ZR2013FQ003, ZR2013EEM027), and China Postdoctoral Science Foundation (2013M541912).

References

1. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501
2. Han F, Yao HF, Ling QH (2013) An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing 116:87–93
3. Zhu QY, Qin AK, Suganthan PN, Huang GB (2005) Evolutionary extreme learning machine. Pattern Recogn 38:1759–1763
4. Cao JW, Lin ZP, Huang GB (2012) Self-adaptive evolutionary extreme learning machine. Neural Process Lett 36:285–305

5. Rashedi Esmat, Nezamabadi-pour Hossein, Saryazdi Saeid (2009) GSA: a gravitational search algorithm. Inf Sci 179:2232–2248
6. Mirjalili S, Hashim SZM (2010) A new hybrid PSOGSA algorithm for function optimization. In: 2010 International conference on computer and information application (ICCIA), Tianjin, pp 374–377
7. Sun GY, Zhang AZ (2013) A hybrid genetic algorithm and gravitational search algorithm for image segmentation using multilevel thresholding. Pattern Recogn Image Anal Lect Notes Comput Sci 7887:707–714
8. Ratnaweera A, Halgamuge S, Watson HC (2004) Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Trans Evol Comput 8(3):240–255
9. Mirjalili SeyedAli, Hashim Siti Zaiton Mohd, Sardroudi Hossein Moradian (2012) Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. Appl Math Comput 218(22):11125–11137
10. Li CL, Dai J, Pan F (2012) Analysis on improvement of particle memory in gravitational search algorithm. J Comput Appl 32(10):2732–2735
11. Li B, Wang JM, Li YB, Song Y (2007) An improved on-line sequential learning algorithm for extreme learning machine. Lect Notes Comput Sci 4491:1087–1093
12. Rajesh R, Siva Prakash J (2011) Extreme learning machines—a review and state-of-the-art. Int J Wisdom Based Comput 1(1):35–49
13. Li B, Li YB, Rong XW (2013) The extreme learning machine learning algorithm with tunable activation function. Neural Comput Appl 22:531–539
14. Sourav Mallick SP, Ghoshal PA, Thakur SS (2013) Optimal static state estimation using improved particle swarm optimization and gravitational search algorithm. Int J Electr Power Energy Syst 52:254–265
15. Chen H, Li S, Tang Z (2011) Hybrid gravitational search algorithm with random-key encoding scheme combined with simulated annealing. IIJCSNS Int J Comput Sci Netw Secur 11(6): 208–217
16. Huang GB, Zhu HM, Ding XJ, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern 42(2):513–529

Chapter 32
P2P Network Trust Strategy Based on New Evaluation Criterion

Xiali Li and Licheng Wu

Abstract In this paper we proposed a new comprehensive trust strategy which uses Malicious Percentage (MP), Feedback Consistency Percentage (FCP) and Initial Trust Vector (IPV) as the new trust evaluation criterion to help filter malicious peers in selection of peers. Simulations demonstrate that the new trust strategy can speed the execution time of the algorithm and enhance the successful transaction rate when resisting cooperative-cheating peers.

Keywords P2P network · Trust strategy · Malicious percentage (MP) · Feedback consistency percentage (FCP) · Initial trust vector (IPV)

32.1 Introduction

P2P network has the characteristics of dynamics and anonymity, which has brought both opportunities and threat [1]. Thus trust model and strategy have been proposed to resist the attacks of malicious nodes. Trust Model in eBay System (TMBS) uses a central server to store and manages the evaluation information [2], the disadvantage is the overburdened central server. Reference [3] proposed EigenTrust algorithm and gave its implementation in the distributed network. EigenTrust get the global trust value of all the peers by finite number of trust iterations in the network. But it does not consider penalty factors and the overhead of the whole network. PeerTrust [4] is a typical adaptive local trust model which calculates the trust value using feedback evaluation and the number of transactions. In our before work, EnhanEigen model [5] was proposed. It used malicious percentage (MP), feedback consistency percentage (FCP) to filter malicious peers. The model is effective and simple .But is does not consider initial trust value of the peers.

X. Li (✉) · L. Wu
School of Information Engineering, Minzu University of China, Beijing 100081, China
e-mail: xiaer_li@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_32

In this paper, we deepen EnhanEigen [5] to a new comprehensive trust strategy and made simulations to verify its performance. We use MP, FCP, and Initial Trust Vector (ITV) parameters as new evaluation criterion. The strategy can distinguish false feedback to resist malicious attack. Compared with EnhanEigen, simulation results show that new strategy has a higher success rate in resistance to the feedback node and cooperative-cheating. Besides, new strategy can speed the execution time of the algorithm.

This paper is organized into four parts. The first is the introduction of some popular trust model and algorithm. The second describe new strategy with three evaluation criterion. The third shows experiments and analyzes data. The last is the conclusion and the promising work in the future.

32.2 Trust Strategy with New Evaluation Criterion

Our strategy adopts the basic thinking of EnhanTrust algorithm which is presented in [3]. The definition of MP and FCP was presented in our recent researches [5, 6].

32.2.1 Initial Trust Vector

To prevent cooperative-cheating, the strategy introduces pre-trusted nodes which will not harm the network and have a certain level trust value. This value is called IPV. We use IPV to enhance the performance of the resisting cooperative-cheating peers.

32.2.2 Comprehensive Trust Value

Assume peer *i*(consumer) selects peer *j*(server) as the download source and have got the file which was provided by peer *j*. Then peer *i* has to submit the evaluation on the peer *j* providing this service. After the end of one transaction, we collect the evaluations on the service from both the server and consumer. Let e_s be the evaluation value from the server and e_r be the evaluation from consumer. The value of e_s and e_r is one of the two discrete values $\{1,-1\}$. 1 means that the evaluation is positive and -1 means that the evaluation is negative. The comprehensive trust value C_i for peer i is calculated by the following:

$$C_i = \alpha g_i + \beta \mathrm{FCP}_i - \lambda \mathrm{MP}_i + \delta \mathrm{IPV}_i \tag{32.1}$$

In Eq. (32.1), $\alpha, \beta, \lambda, \delta$ are tuning parameters and satisfies the following equation:

$$\alpha + \beta + \lambda + \delta = 1 \tag{32.2}$$

g_i is the peer's global trust value calculated by the algorithm proposed in [5], FCP_i is the peer's FCP and MP_i is the peer's MP calculated by Reference [5, 6].

32.2.3 Analysis of the New Strategy

Filtering policy against malicious nodes make those nodes providing malicious service excluded from the choice of download sources. Even if feedback-cheating nodes raised the malicious peers successfully, the malicious nodes will be filtered out because of its high MP. So the new strategy can resist cooperative-cheating to some extent. In the next section, we will validate the strategy by the simulations.

32.3 Simulation Experiments and Analysis

In order to carry out the simulation experiments, we construct three different types of peers.

Good peers (g): This kind of peers provide authentic files with a probability of more than 85 %. They submit authentic evaluations to the services provided by themselves and by other peers.

Malicious and malign peers (mm): Such peers provide inauthentic files with a probability of more than 85 %. They always give a higher evaluation to their own services than the authentic situation. They also commit false feedbacks to the authentic services.

Feedback-cheating peers (fc): Providing authentic files with a probability of more than 85 % (less than 15 % probability of providing inauthentic files). But as servers, they always give praise to their own services. When these peers build up malicious collectives with mm peers, they will raise mm peers and slander other non-malicious peers.

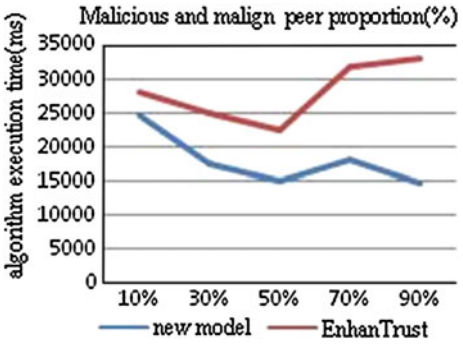
To prevent the possible impacts on the simulation results for unreasonable files allocation, we guarantee each file at least owned by one good peer. Each peer issues a file download request with equal probability. The threshold of MP is 0.15, because good peers will provide inauthentic files with a probability of less than 15 %.

Simulation experiments are divided into two parts: MM peers simulation to validate the execution time of algorithm and Cooperative-cheating simulation to validate the performance of resisting cooperative-cheating.

Table 32.1 Environment parameter in MM simulation

Parameters	Values
Number of peer	100
Number of file	500
Number of transaction	20,000
Pre-trusted peers	10
φ	0.15
α	0.15
β	0.45
λ	0.25
δ	0.15

Fig. 32.1 Execution time contrast



32.3.1 MM Peers Simulation

Simulation environment settings are shown in Table 32.1. We set the number of MM peers at different proportion of malicious collectives. We implement the new strategy and choose the EnhanTrust [5] as the benchmark. Then we get the contrast of algorithm execution time as shown in Fig. 32.1.

From Fig. 32.1 we can see that the new strategy costs less time. With increase in the fraction of malicious peers, our new strategy filters more and more malicious peers, at the same time the number of peers which can provide services is decreasing.

32.3.2 Cooperative-Cheating Simulation

Simulation environment settings are shown in Table 32.2. We set the number of cooperative-cheating peers at different proportion of all the peers. We implement the new strategy and choose the EnhanTrust [5] as the benchmark. Then we get the contrast of successful transactions rate as shown in Fig. 32.2.

Table 32.2 Environment parameter in cooperative-cheating simulation

Parameters	Values
Number of peer	100
Number of file	500
Number of transaction	20,000
Pre-trusted peers	10
g	50
mm	23
fc	27

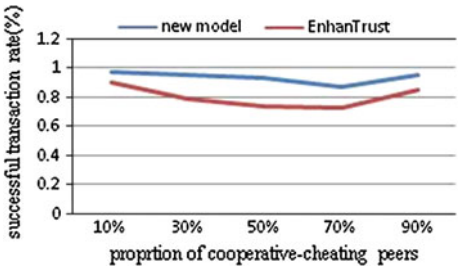


Fig. 32.2 Successful transaction rate contrast

Figure 32.2 shows the successful transactions rate contrast between the new model and EnhanTrust [5] when the cooperative-cheating peers is in different proportions of the whole peers. It can be seen that the successful transaction rate curves of two models roughly are the same. There is a turning point at 70 % in two curves. New strategy has a higher successful transaction rate than that of Enhan-Trust [5] in general. It demonstrates that our new strategy performs better in resisting cooperative-cheating peers than EnhanTrust.

32.4 Conclusions

The main contribution of this paper is proposing a new strategy with MP, FCP, and IPV evaluation criterion to help filter malicious peers. Experiments demonstrate that the new strategy is effective in resisting feedback-cheating peer behaviors and collective attack.

Acknowledgments This work is supported in part by 2013 Beijing University youth talent plan No. YETP1294. Besides, it is supported in part by the NSFC project No. 51375504 and the Program for New Century Excellent Talents in University.

References

1. Guang-sheng M, Deng-guo F, Pu-rui. S (2009) Colluding clique detector based on activity similarity in P2P trust model. J Commun 30(8):9–20 (in Chinese)
2. Yong-jun L, Ya-fei D (2010) Research on trust mechanism for peer-to-peer network. Chin J Comput 33(3):390–405 (in Chinese)
3. Kamvar SD, Schlosser MT, Garcia-Molina H (2003) The EigenTrust algorithm for reputation management in P2P networks. Stanford University, Stanford
4. Xiong L, Liu L (2004) PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities. IEEE Trans Knowl Data Eng 16(7):843–857
5. Xiali L, Jingjing X, Sujuan Z Xiaohua Y (2013) An EnhanEigen algorithm trust model in P2P network. In: International conference on mechanical intelligence and automatic control, Xi'an, pp 1177–1180
6. Xiali L, Licheng W, JingJing X, Huilin M Fang L (2014) A new trust model in P2P network. In: International conference on machine tool technology, mechatronics and information engineering, Guilin, Guangxi, China, pp 3024–3027

Chapter 33
A Novel Performance Evaluation Method for Visual Tracking Methods

Hui Teng, Lianzhi Yu, Huaping Liu, Fuchun Sun and Xiaojuan Liu

Abstract The evaluation for visual tracking is of utmost importance in computer vision. During the last two decades, several evaluation frameworks have been proposed and applied by a large amount of researchers, such as average central error, success rate, or Pascal score. It turns out that it is not easy to compare trackers fairly, precisely, and comprehensively with just only one evaluation index. Given the fact, a novel method for the evaluation of visual tracking algorithms is proposed in this paper, which combines both central error and Pascal score. Using this metric, an experimental validation was conducted finally and may serve as a reference for related research.

Keywords Visual tracking · Evaluation · Pascal score · Central error

33.1 Introduction

Visual tracking is of considerable importance in the field of computer vision and robotics, and therefore attracts a great amount of research effort. Given the initialized state (e.g., position of the target, weight and height of the target), the goal of the visual object tracking is to estimate the states of the object in image sequences. It plays very important role in large numbers of vision applications such as motion analysis, activity recognition, video surveillance, and traffic monitoring [1–3]. While much progress has been made in recent years [4–10], it is still a great challenge to develop a robust tracking algorithm that can successfully handle all

H. Teng (✉) · H. Liu · F. Sun · X. Liu
Department of Computer Science and Technology, State Key Laboratory of Intelligent Technology and Systems, TNLIST, Tsinghua University, Beijing, China
e-mail: yueguangth5@126.com

H. Teng · L. Yu
School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_33

scenarios such as occlusion, translation, rotation, scaling, as well as background clutters.

Beside the tracking method itself, the performance valuation of visual tracking is also important and challenging. A reasonable performance evaluation method can be used to demonstrate the advantage and disadvantage of each method, and therefore help to design better trackers in the future work.

To the best of our knowledge, there exist several kinds of evaluation metrics already, which are widely used in measuring the performance of trackers. Among related works (e.g., CT [11], ASLA [12]), the first metric is the success rate in the PASCAL VOC [13]. The other one is the center location error which can be used to measure the distance between two points (the centers of tracking bounding box and ground-truth bounding box). While in other papers, Pascal score is widely used. Pascal score is defined as the overlapping rate of the tracking bounding box and ground-truth bounding box.

Given the fact, a new framework for the evaluation of visual tracking algorithms is proposed in this paper, which presents a totally new definition of "central error," and what is more important is that our approach takes both central error and Pascal score into consideration. In that way, this method can describe the quality of the trackers in frame level more comprehensively and reasonably, and overcome the limitations of both Pascal score and central error without presenting two or more indices to demonstrate the performance of every tracking method. Using this approach, an experimental validation was conducted finally and may serve as a reference for related research.

This paper is organized as follows: Section. 33.1 gives a brief introduction about visual and existing measurement. A detailed description of the limitation of two existing evaluating method is discussed in Sect. 33.2. Section 33.3 presents the detailed introduction of criteria for evaluation. Experimental validation and discussion are successfully conducted in Sect. 33.4, followed by the conclusion in Sect. 33.5.

33.2 Limitation Analysis

Using these series of measuring metrics mentioned above, namely Pascal score or success rate and central error, the evaluation work can be successfully conducted to some extent. However, there are still some limitations with these two metrics.

33.2.1 Limitation of Pascal Score

For instance, we just cannot evaluate one tracker with only one particular evaluation metric. Especially, for Pascal score, it is not able to describe the tracking performance accurately under some common circumstances, as shown in Fig. 33.1:

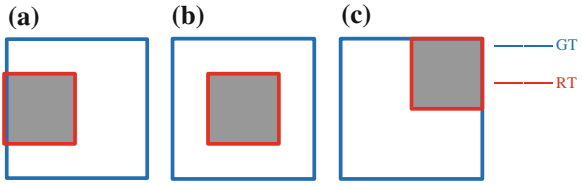


Fig. 33.1 Three different conditions of tracking results that the overlapping rate cannot handle

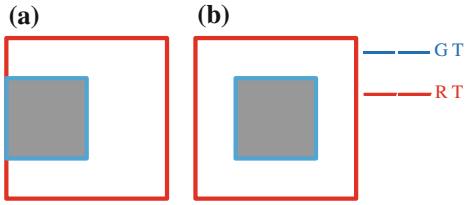


Fig. 33.2 $R_T \cup R_G = R_T$ and $R_T \cap R_G = R_G$

In Fig. 33.1, the blue box represents the bounding box of ground truth, while the red one represents the bounding box of the tracking result. And the shadow area means the overlap of these two boxes. What is more, (a), (b), and (c) are three different conditions of tracking results. Based on this, we can easily calculate the overlap rate (Pascal score) of each tracking results, (a), (b), and (c), which are all the same value that equals to 25 %. Although the value of Pascal score is totally the same, we are so confident to say that the tracking result (b) is much better than both (a) and (c); and what is more, (c) performs a little bit worse than (a). Due to this limitation, researchers have to use central error to demonstrate the differences between the two kinds of situations.

Moreover, when $R_T \cup R_G = R_T$ and $R_T \cap R_G = R_G$, as demonstrated in Fig. 33.2, Pascal score still cannot make the accurate evaluation.

33.2.2 Limitation of Central Error

Central error describes the distance error between two related points. As is shown in Fig. 33.3, the central errors of (a) and (b) are both zeros, which cannot reflect the fact that (a) is better than (b).

What is more, under some conditions, this metric may fail to work. As Fig. 33.4 shows, there is no overlapping area of ground-truth bounding box and the tracking result bounding box, which indicates that the tracker loses the target. However, central error still gives the calculation of the distance error. In this case, researchers have to use Pascal score to make up this kind of disadvantage; because, Pascal score

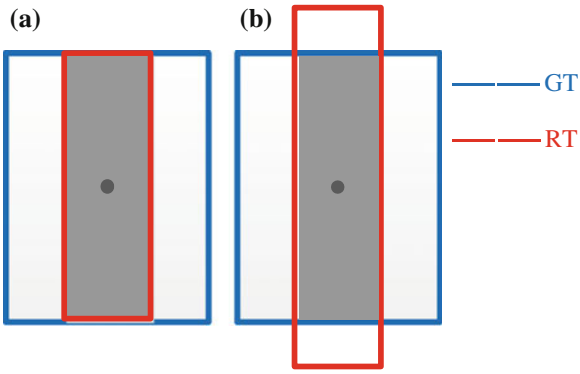


Fig. 33.3 Two different conditions with same central error

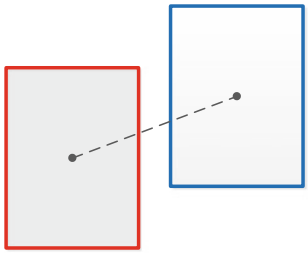


Fig. 33.4 No overlapping area

equals to zero under this condition and it proves to be a failure when the value of Pascal score is 0.

In addition, this metric is strongly influenced by the original size of the frame image. Under some circumstances, if the size of the frame image is relatively large, the great value of evaluation does not mean that the tracking performance is terribly bad.

Considering this limitation, researchers usually illustrate at least two or more indices to describe the performance of each single tracking algorithm.

33.3 Criteria for Evaluation

The method of evaluation proposed in this paper is based on two kinds of criterion: Position only (point-based technique) or region information as well (region-based technique).

The point-based technique takes into account the relative position errors (in pixels) between the center of the ground-truth tracks [14, 15] and the tracking results, which is named as "central error." As is shown in Fig. 33.5, central error is defined as the length of vector $\overrightarrow{O_1O_2}$:

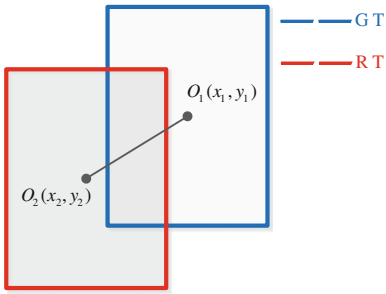


Fig. 33.5 Illustration of central error

$$\left|\overrightarrow{O_1O_2}\right| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{33.1}$$

where (x_1, y_1) is the coordinate of the center of the ground-truth bounding box, while (x_2, y_2) is the coordinate of the center of the tracking bounding box.

The region-based technique supplements the point-based one, which has been introduced in Sect. 33.1 and defined as follows:

$$\text{score} = \frac{\text{aera}(R_T \cap R_G)}{\text{aera}(R_T \cup R_G)} \tag{33.2}$$

Using Pascal score, researchers usually define a metric as follows, named Success Rate,

$$\text{score} = \begin{cases} \geq \tau, & \text{success} \\ < \tau, & \text{failure} \end{cases} \tag{33.3}$$

$$S(\tau) = \frac{\text{NUM}\{t|\text{score}(t) \geq \tau\}}{\text{Length}} \tag{33.4}$$

where τ is a threshold, the tracking result is considered as a success when score $\geq \tau$. Otherwise, it fails to track the target. $S(\tau)$ is the definition of success rate which is a ratio of the number of "success" and the length of whole sequence.

To combine the two techniques into one criterion, we introduce a new variable e shown in Fig. 33.6:

$$e_1 = \left|\overrightarrow{OA}\right| \quad e_2 = \left|\overrightarrow{OB}\right| \tag{33.5}$$

where A is the center of the ground-truth bounding box and O is the center of the overlapping part of the two boxes.

In addition, L_1 is defined as half of the length of the diagonal of the ground-truth bounding box; and L_2 is half of the length of the diagonal of the tracking result bounding box.

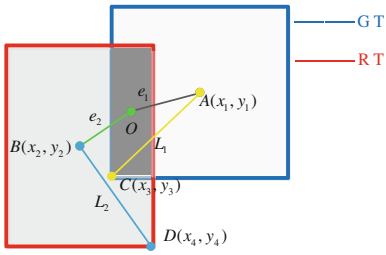


Fig. 33.6 Illustration of our proposed method

$$L_1 = \left|\overrightarrow{AC}\right| \quad L_2 = \left|\overrightarrow{BD}\right| \tag{33.6}$$

Thus we give a new "central error," which is

$$\text{Err} = \frac{e_1/L_1 + e_2/L_1}{2} \tag{33.7}$$

Figure 33.7 shows two kinds of limit conditions of the positional relationship between the two boxes. Figure 33.7a shows $e_1 = L_1$ and $e_2 = L_2$, and Fig. 33.7b shows that e_1, e_2 are getting increasingly smaller, and finally reach 0.

It is easy to understand that the maximum of Err is 1. In that case, $e_1 = L_1$ and $e_2 = L_2$, which indicates that there is no overlap with the ground-truth bounding box and the tracking bounding box. Err reaches the minimum when centers of R_T and R_G are completely the same value.

$$\frac{e_1/L_1 + e_2/L_2}{2} = \begin{cases} 1, & e_1 = L_1 \text{ and } e_2 = L_2 \\ 0, & e_1 = 0 \text{ and } e_2 = 0 \end{cases} \tag{33.8}$$

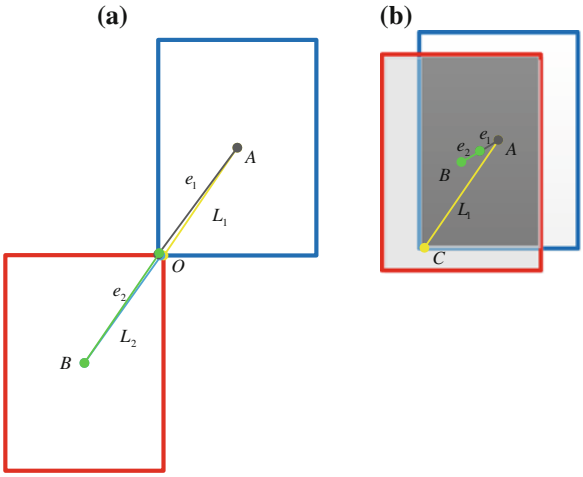


Fig. 33.7 Two kinds of limit conditions. **a** Err = 1. **b** Err = 0

However, we hope that the greater value of the evaluation represents the better performance; so Err can be redefined as follows:

$$\mathrm{Err} = \left| \frac{e_1/L_1 + e_2/L_2}{2} - 1 \right| = \begin{cases} 0, & e_1 = L_1 \text{ and } e_2 = L_2 \\ 1, & e_1 = 0 \text{ and } e_2 = 0 \end{cases} \tag{33.9}$$

In that way, the greater the value is, the better the performance can be obtained. Finally, the framework used to evaluate the trackers can be expressed as follows:

$$\mathrm{SCORE} = \frac{\mathrm{aera}(R_T \cap R_G)}{\mathrm{aera}(R_T \cup R_G)} * \left| \frac{e_1/L_1 + e_2/L_2}{2} - 1 \right| \tag{33.10}$$

Our method proposed in this paper is a kind of method integrating two important parts as is clearly shown in the expression (33.10). $\frac{\mathrm{aera}(R_T \cap R_G)}{\mathrm{aera}(R_T \cup R_G)}$ is the definition of Pascal score and $\left| \frac{e_1/L_1 + e_2/L_2}{2} - 1 \right|$ describes the distance error between relative points. And what is more important is that we normalized the central error with the help of e_1/L_1 and e_2/L_2. In that way, the influence of the image size is totally and perfectly eliminated. Based on the two parts in the expression together, our method would be an effective tool in evaluation equipped with the advantages of Pascal score and central error; and as a result, it overcomes the one-sidedness of either Pascal score or central error.

33.4 Experimental Validation

The proposed method is implemented in MATLAB, and we use five real-world datasets to validate the performance of our method. The datasets are available in the website (http://cvlab.hanyang.ac.kr/tracker_benchmark_v10.html). The challenges of these videos include illumination variation, partial occlusion, pose variation, background clutter, and scale change.

We use two state-of-the-art trackers (MIL [16], STRUCK [17]) to produce the tracking results. Using the tracking result-frames, the evaluation work can be conducted effectively. For fair evaluation, we evaluate the proposed tracker against those methods using the source codes provided by the authors. Each tracker is run with adjusted parameters.

33.4.1 Overcome the Limitation of Pascal Score

First, couple was chosen as one input dataset, which has 140 frames with translation, partial occlusion, background clutter, and scale change. Figure 33.8 shows

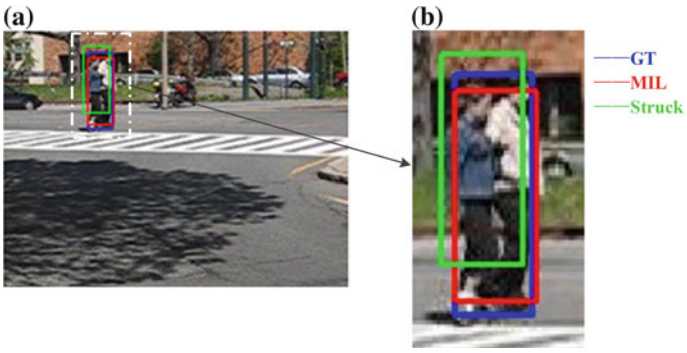


Fig. 33.8 Tracking results of MIL and STRUCK in one frame image (#53) of couple sequences. By zooming in the area of *white dotted* rectangle shown in (**a**), the tracking result boxes are illustrated more clearly in (**b**)

the tracking results of MIL which is represented in red box, and STRUCK in green. The blue box is the ground-truth bounding box. As illustrated in the below Fig. 33.8b, the overlap rate of MIL is 61.82 % and the result of STRUCK is exactly the same. However, it is easily to find out that the result of the red box is better than the green one. In that case, we cannot distinguish which tracking algorithm performs better just with one criterion. Our evaluation criterion solves this problem perfectly with our new redefinition of Err which taking the distance error into account. By calculating, the SCORE of MIL is 54.54 %, 10.27 % higher than STRUCK, which strongly indicates that MIL tracks so much better than STRUCK.

Second, when $R_T \cup R_G = R_T$ and $R_T \cap R_G = R_G$, we conduct the test using David sequence. Note that Pascal score cannot tell which tracker works better under this condition, while our metric show that STRUCK is 3 % superior to MIL on 53th frame in David.

33.4.2 Overcome the One-Sidedness of Pascal Score

In Fig. 33.9, frames 57 to 62 demonstrate the same situation which strongly indicates that our method can handle it more confidentially. Tables 33.1 and 33.2 show the resulting values of MIL and STRUCK of each frame. From frames 57 to 58, it can be easily seen from Table 33.1 that the difference of overlapping rate between MIL and STRUCK is slightly small. However, the results of our method shows that MIL is superior to STRUCK with a big difference as it is illustrated in Table 33.2, which means that our method can amplitude the difference to some extent. Frames 60 to 62 show that the results of the two tracking algorithms with the original evaluating criterion are totally the same. However, our method is able to detect the difference while the original one is not that competent.

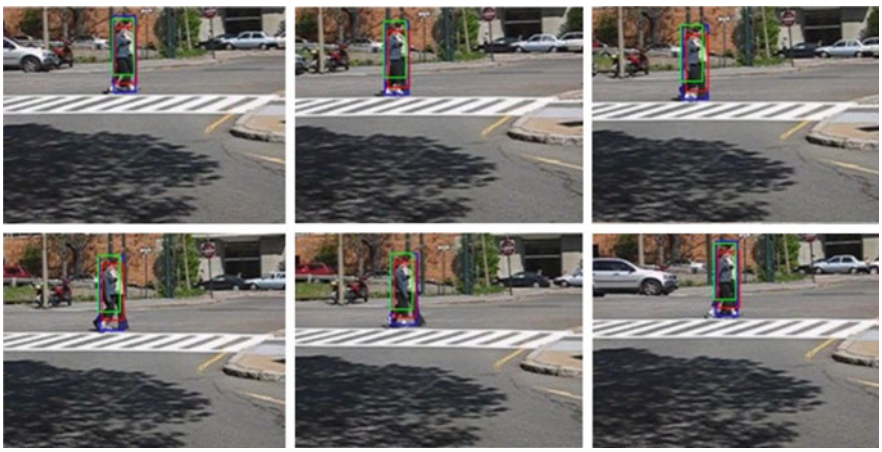


Fig. 33.9 Tracking results of MIL and STRUCK in several frames (#57 to #62)

Table 33.1 The evaluating results of couple using the pascal score

Frame no	57	58	59	60	61	62
MIL (%)	67.27	67.21	64.81	65.77	65.96	65.20
STRUCK (%)	62.71	62.65	56.36	65.77	65.96	65.20

Table 33.2 The evaluating results of couple using our method

Frame No.	57	58	59	60	61	62
MIL (%)	65.37	64.04	61.15	61.01	62.35	60.56
STRUCK (%)	47.75	45.00	39.84	47.60	49.95	47.79

As it is depicted in Figs. 33.10 and 33.11, with the frame number getting larger, the overlapping rates of evaluation (both the original method and ours) are dropping off gradually. Figure 33.10 shows that the tracking performance of MIL is better than that of STRUCK on most of the images, while from frames 47 to 90, the green line, which represents the tracking results of STRUCK with Pascal score, shows an obvious growing trend and obviously STRUCK works better than MIL on some images. However, compared with the tracking results with Pascal score in Fig. 33.10, the green line in Fig. 33.11 which represents the tracking result of STRUCK with our method shows a totally different performance. It shows an obvious gap between MIL and STRUCK after 40 frames passed, which indicates that our method correct the one-sidedness of Pascal score.

Finally, another two video sequences (Subway, Singer1) are tested using both Pascal score and our proposed metric. The main values of the tracking results of each video sequence are calculated which indicates our metric is able to capture the slight difference among many trackers.

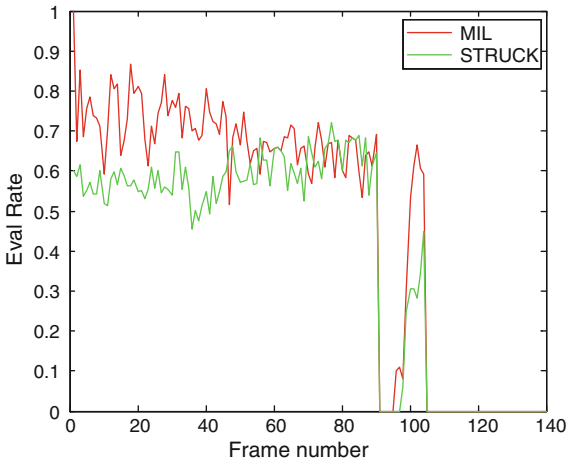


Fig. 33.10 Tracking results of couple with pascal score

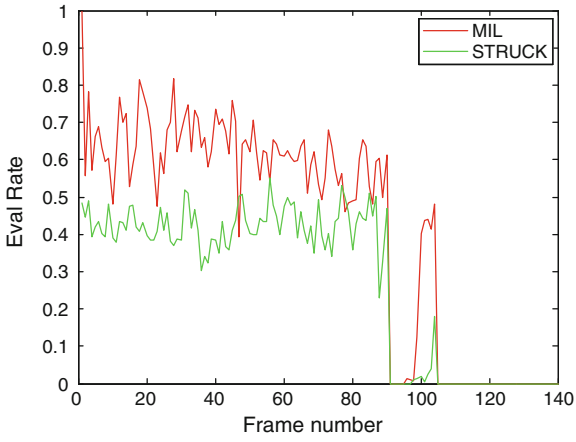


Fig. 33.11 Tracking results of couple with our method

33.4.3 Overcome the Limitation of Central Error

For Liquor sequence with heavily partial or even total occlusion and scaling change, we conduct the test using Pascal score, central error, and our proposed method. The number of the images in Liquor sequence is 1741. The target keeps still in the first 200 frames and both MIL and STRUCK show a reliable tracking performance. So we just pay attention on the tracking results when the target begins to move and they are shown in Fig. 33.12. STRUCK loses the target when frame number reaches 364, which is successfully evaluated by our proposed method and Pascal score as Fig. 33.12a, b demonstrate. However, central error fails to give a proper evaluating result. As can be seen from Fig. 33.12c, it is different to distinguish whether STUCK succeeds in tracking the target or not by central error,

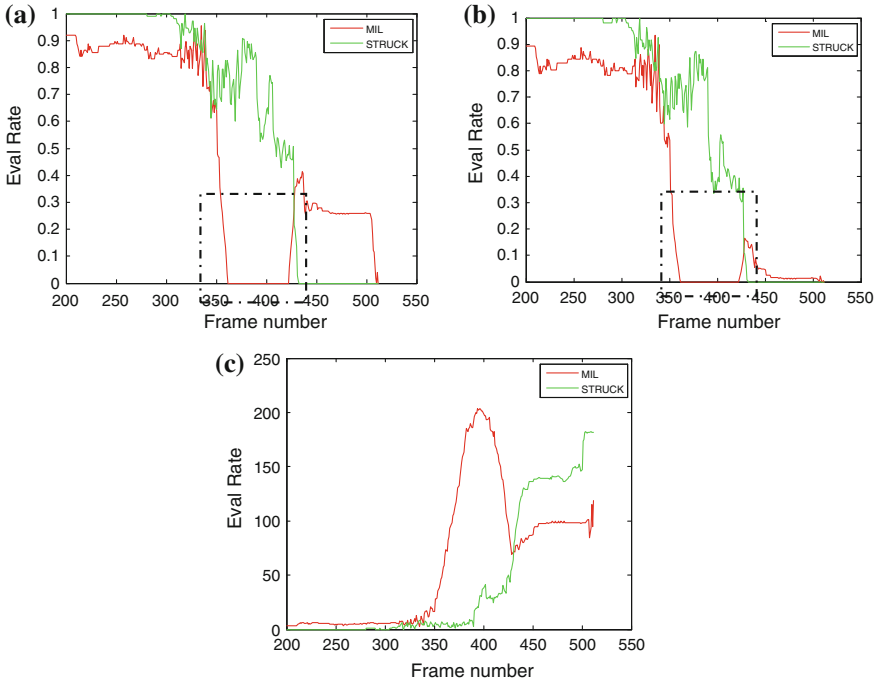


Fig. 33.12 **a** Tracking results of liquor with pascal score. **b** Tracking results of liquor with our method. **c** Tracking results of liquor with central error

because central error just calculate the distance error between two points; and when the tracker loses the target, it still continues computing the central distance error.

Besides, the dashed frames in Fig. 33.12a show that after 420 frames passed, MIL seemed to succeed in tracking. As a matter of fact, MIL does not track the target again but the target moves into the tracking result bounding box. Although the overlapping rate is almost 30 % or even higher under this condition, the central error is relatively great. Our method shows an accurate evaluation as shown in Fig. 33.12b.

33.5 Concluding Remarks

Pascal score and central error are the main methods used for evaluation of visual tracking. In this paper, we propose a novel metric for evaluating the performance of visual trackers. This metric takes both central error and Pascal score into consideration at the same time, which presents a new reference for the evaluation by redefining a new concept of distance error. Using this method, effective evaluation has been obtained by experimental validation. Our method shows a more precise

and comprehensive performance in evaluating, and overcomes the limitations and one-sidedness of either Pascal score or central error. It obtains more reliable evaluation results under some circumstances where Pascal score and central error cannot handle.

References

1. Irani M, Peleg S (1993) Motion analysis for image enhancement: resolution, occlusion, and transparency. J Vis Commun Image Represent 4(4):324–335
2. Maurer U, Smailagic A, Siewiorek DP et al (2006) Activity recognition and monitoring using multiple sensors on different body positions. In: IEEE international workshop on wearable and implantable body sensor networks, BSN 2006, vol 4, p 116
3. Cohen I, Medioni G (1999) Detecting and tracking moving objects for video surveillance. In: IEEE computer society conference on computer vision and pattern recognition, p 2
4. Isard M, Blake A (1998) CONDENSATION–conditional density propagation for visual tracking. IJCV 29(1):5–28
5. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. PAMI 25(5):564–577
6. Ross D, Lim J, Lin R-S, Yang M-H (2008) Incremental learning for robust visual tracking. IJCV 77(1):125–141
7. Babenko B, Yang M-H, Belongie S (2009) Visual tracking with online multiple instance learning. In: CVPR
8. Mei X, Ling H (2009) Robust visual tracking using L1 minimization. In ICCV
9. Hare S, Saffari A, Torr PHS (2011) STRUCK: structured output tracking with kernels. In: ICCV
10. Fan J, Shen X, Wu Y (2012) Scribble tracker: a matting-based approach for robust tracking. PAMI 34(8):1633–1644
11. Zhang K, Zhang L, Yang M-H (2012) Real-time compressive tracking. In: ECCV
12. Jia X, Lu H, Yang M-H (2012) Visual tracking via adaptive structural local sparse appearance model. In: CVPR
13. Everingham M, Van Gool L, Williams C, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. IJCV 88(2):303–338
14. Ristic B, Vo B-N, Clark D, Vo B-T (2011) A metric for performance evaluation of multi-target tracking algorithms. IEEE Trans Signal Process 59(7):3452–3457
15. Black J, Ellis T, Rosin P (2003) A novel method for video tracking performance evaluation. In: Proceedings of WPETS, pp 125–132
16. Babenko B, Yang M-H, Belongie S (2009) Visual tracking with online multiple instance learning. In: CVPR
17. Hare S, Saffari A, Torr PHS (2011) STRUCK: structured output tracking with kernels. In: ICCV

Chapter 34
Spatial Target Vision Measurement and Precision Compensation Based on Soft Computing

Kai Li and Feng Yuan

Abstract On the basis of geometrical relationship between three-dimensional spatial target system and double-theodolites, the measurement models of observing angle and spatial coordinates together with distance are established according to the principles of the space rendezvous and docking technology. In fact, due to the angles' trigonometric functions' nonlinearities and theodolite's inner characteristics, observing angle readings of double-theodolites produce errors at certain position. The reasons of such errors are analyzed, and the idea is provided of taking errors as nonlinear components, building up the Neural Network (NN) to simulate the nonlinear mapping between observing angles and distances, optimizing its weights, and regarding the outputs of NN as compensated term. Simulation curves imply weights of NN intermediate layer influence the final compensation precision. In experiment, NN with optimized weights is applied in the processing of measured data; the results of which certificate such idea adaptively compensate system errors' influences in binocular stereo vision.

Keywords Spatial target · Vision measurement · Precision compensation · Soft computing

34.1 Introduction

A kind of target simulator with three degrees of freedom, to some extent, is able to simulate real target's motion; during the course of which, the theories and techniques of Binocular stereo vision [1–5] are widely employed on the occasions of spatial geometric variables' measuring and precision analysis. Besides, double-theodolites system are commonly regarded as testing instruments to acquire the

K. Li · F. Yuan (✉)
1016 Room, Mechanical Building, Harbin Institute of Technology, No. 92, West Da Zhi Street, Harbin, China
e-mail: likaihit@hit.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_34

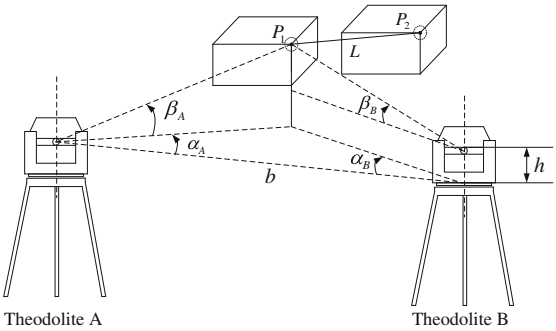


Fig. 34.1 Spatial three-dimensional measurement system by double-theodolites

vertical and horizontal observing angles relative to spatial target simulator in the theodolite frame of axes [6]. Coordinates axes' calibration between theodolites and spatial target simulator is implemented by characteristic points measurement; during the experiment of which, a characteristic point is fixed on the spatial target simulator, the horizon axis of X-line and the vertical axis of Y-line on double-theodolites' are matched with the ones of spatial target simulator's own frame axes, and the coordinates of the characteristic point and the moving distance between two characteristic points under global frame of axes can be calculated by mathematical calculation [7–11]. In Fig. 34.1, points on spatial target simulator are P_1 and P_2, and the actual measuring records of P_1 by double-theodolites are the observing angle of α_A, β_A, α_B, and β_B. Suppose if the horizon distance of b and the vertical height of h are known, the coordinates of P_1 are determined and the coordinates of P_2 are known, so as the spatial distance of L between P_1 and P_2.

In measurement, the horizon observing angle's scale is 0°–180°, and the vertical observing angle's scale is −90°–90° as the result of the field coverage's limitation of theodolite [12–14]. The zero standards are generally selected on the point of $\beta_A = \beta_B = 0°$. In experiment, it is found that if the zero standard points lie in the neighborhood of 90°, the measuring distance is likely to produce error of ΔL, the essence of which is a nonlinear component and belongs to a term of system error. In order to solve such problems, nonlinear mapping reflecting observing angle and measured distance is established according to the method of coordinates reverse calculation, the Neural Network (NN), is adopted to reflect the nonlinear relationship; the primary reason of generating system error is discussed, and the idea is proposed of compensating system error by means of optimizing the weights of NN and regarding the output of NN as compensation results.

34.2 System Analysis

The simplification of measuring system has been made in two aspects according to Fig. 34.1 [15–16]: (1). The position of theodolite's len is equivalent as a single point, which means theodolite's position is regarded as a characteristic point under

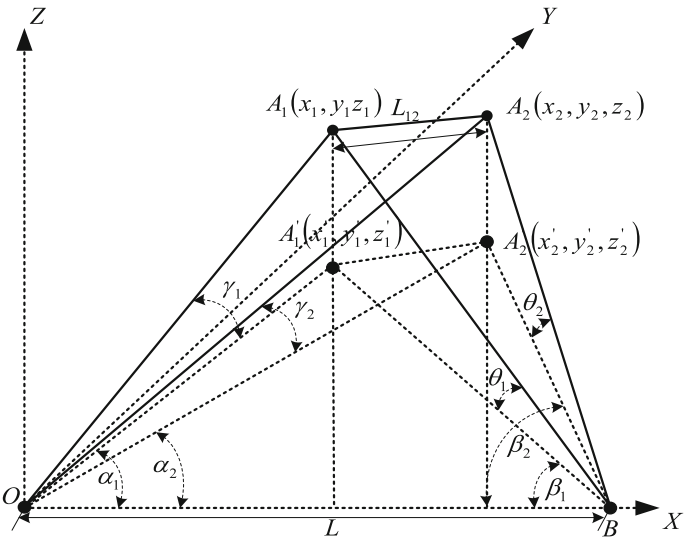


Fig. 34.2 The spatial measurement coordinates of double-theodolites

three-dimensional measurement frame of axes. (2). Height differences of h are neglected; the main purpose of which is to delete the influences of height differences in the mathematical expression and is easy to obtained by horizontal calibration.

In the spatial 3D frame of axes, supposing that theodolite A is located at zero point of O, and theodolite B is located at the point of B. Spatial 3D frame of axes come into being in Fig. 34.2. The coordinates' values are expressed in (x_1, y_1, z_1) and (x_2, y_2, z_2) under the frame axes of O-XYZ. A_1 and A_2's projection on the plain of O-XY are A'_1 and A'_2, the coordinates' value of which are $(x'_1,\ y'_1,\ z'_1)$ and $(x'_2,\ y'_2,\ z'_2)$. In Fig. 34.2, all of the spatial target characteristics points and the two theodolites are simplified to the geometric points in the spatial 3D frame of axes.

The actual records of theodolite are the horizon observing angle of α_1, α_2, β_1, and β_2, and vertical observing angle of γ_1, γ_2, θ_1, θ_2. Based on triangle formula deduction, the spatial distance of L_{12} and single point's coordinates' values of A_1 and A_2 can be attributed to the nonlinear function with the self-variables of observing angles as shown in the expressions of (34.1) followed by sine theorem.

$$\begin{aligned}
&\frac{OB}{\sin(180^\circ - \alpha_1 - \beta_1)} = \frac{OA'_1}{\sin\beta_1} \Rightarrow OA'_1 = \frac{OB\sin\beta_1}{\sin(\alpha_1 + \beta_1)} \\
&\Rightarrow \begin{cases} x'_1 = x_1 = OA'_1\cos\alpha_1 = \dfrac{L\sin\beta_1\cos\alpha_1}{\sin(\alpha_1+\beta_1)} \\ y'_1 = y_1 = OA'_1\sin\alpha_1 = \dfrac{L\sin\beta_1\sin\alpha_1}{\sin(\alpha_1+\beta_1)} \\ z'_1 = 0, z_1 = OA'_1\tan\gamma_1 = \dfrac{L\sin\beta_1\tan\gamma_1}{\sin(\alpha_1+\beta_1)} \end{cases}
\end{aligned} \tag{34.1}$$

On the basis of distance formula of spatial points, the mathematical model of measuring system is built as shown in the expressions of (34.2) and (34.3) [16–22].

$$L_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} = L\sqrt{\left(\frac{\sin\beta_1\cos\alpha_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\cos\alpha_2}{\sin(\alpha_2+\beta_2)}\right)^2 + \left(\frac{\sin\beta_1\sin\alpha_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\sin\alpha_2}{\sin(\alpha_2+\beta_2)}\right)^2 + \left(\frac{\sin\beta_1\tan\gamma_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\tan\gamma_2}{\sin(\alpha_2+\beta_2)}\right)^2} \tag{34.2}$$

$$L = \frac{L_{12}}{\sqrt{\left(\frac{\sin\beta_1\cos\alpha_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\cos\alpha_2}{\sin(\alpha_2+\beta_2)}\right)^2 + \left(\frac{\sin\beta_1\sin\alpha_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\sin\alpha_2}{\sin(\alpha_2+\beta_2)}\right)^2 + \left(\frac{\sin\beta_1\tan\gamma_1}{\sin(\alpha_1+\beta_1)} - \frac{\sin\beta_2\tan\gamma_2}{\sin(\alpha_2+\beta_2)}\right)^2}} \tag{34.3}$$

Expressions (34.2) and (34.3), simply the tangent values of vertical observing angle of γ_1 and γ_2 tend to be infinite when γ_1 and γ_2 approach the position of 90°, the results of which means the numerical errors of L and L_{12} may occur around the position of 90°. When γ_1 and γ_2 are apart from the position of 90°, such influence decreases. The mathematical model of measuring system is reformed in expression (34.4).

$$L_{12}(X) = \begin{cases} f(X) + g(X)\Delta\delta & \forall\varepsilon > 0, \gamma_1\gamma_2 \in 90^\circ \pm \varepsilon \\ f(X) & \forall\varepsilon > 0, \gamma_1\gamma_2 \notin 90^\circ \pm \varepsilon \end{cases} \tag{34.4}$$

In the expression (34.4), X and $\Delta\delta$ are in forms of vector as X = [α_1 α_2 β_1 β_2 γ_1 γ_2], $\Delta\delta$ = [$\Delta\gamma_1$ $\Delta\gamma_2$], besides, $f(X)$ and $g(X)$ are shown in expression (34.4). $g(X)\Delta\delta = \Delta L_{12}$ is the system error of measuring model. The system error exists in the condition of the nonlinear mapping expressions with $\forall\varepsilon > 0$, γ_1 $\gamma_2 \in 90^\circ \pm \varepsilon$.

34.3 Neural Network Approximating of Nonlinear Component

In experiment, training samples comes from the coordinates of target points, and 3D target simulator is used to set 20 points on the XY plain as showed in Fig. 34.3. In Fig. 34.3, the original zero point is A_5, and real length between any two points is set to 480 mm under precise measurement.

Testing data is expressed below: L = 800 mm, α_1 = 58°42′45″, α_2 = 45°45′38″, β_1 = 34°0′47″, β_2 = 42°5′28″, γ_1 = 14°11′27″, γ_2 = 11°57′20″, θ_1 = 9°23′36″, θ_2 = 11°10′27″. According to expression of (34.3), the precise distance between double-theodolites is $L = 4.553492 \times 10^3$ mm. NN is adopted with 1 input layer with 8 neurons, 1 hidden layer with 2 neurons, and 1 output layer with 1 neuron. The excitation functions are used in Tansig form, NN's architecture is employed in

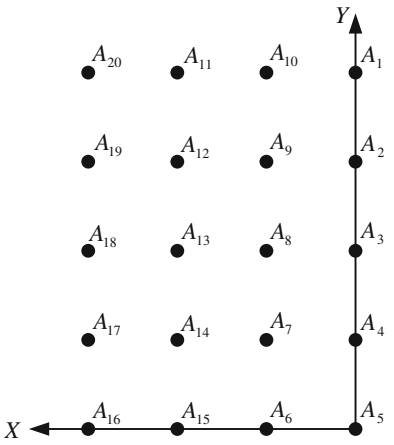


Fig. 34.3 Test points' distribution of target simulator

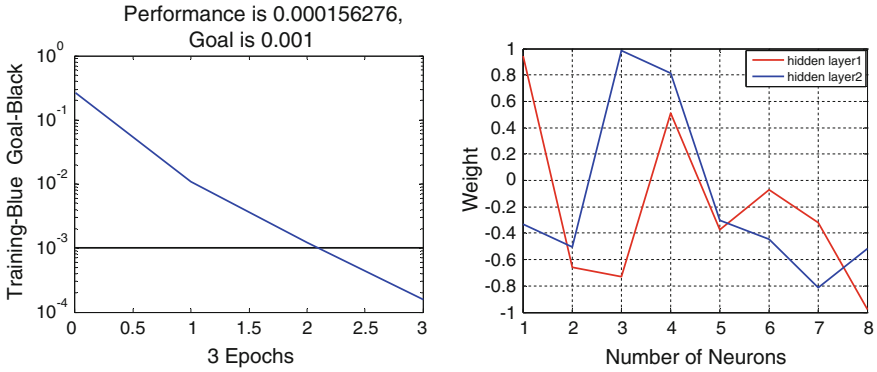


Fig. 34.4 Training error and variation of weights with the precision of 0.001

BP network, and training algorithm is selected with LSM or gradient decrease (GD). Curves with different training errors and weights' variation meeting different precise demands are shown in Figs. 34.4 and 34.5.

Figures 34.4 and 34.5 imply: (1) Under the condition of 100 times of training, higher precision demands more times of training. (2) All hidden layers' weights' variations are −1 to 1. (3) Taking error functions as performance index, weights' optimization relies on the training algorithm. (4) Comparing with the algorithms, BP has the merits of fast convergence and prevent from local optimal points.

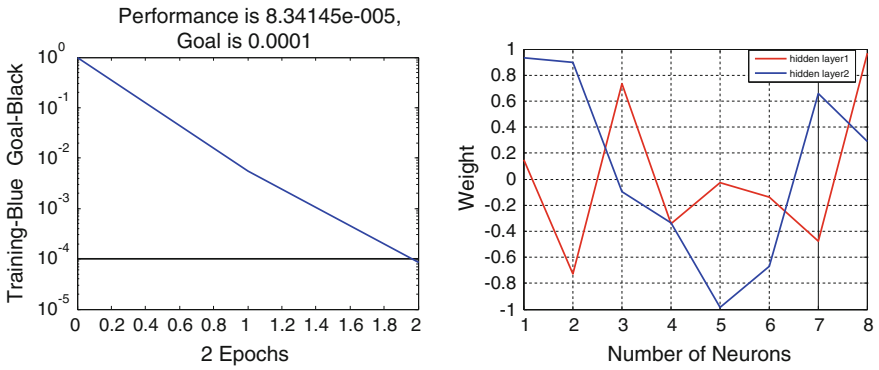


Fig. 34.5 Training error and variation of weights with the precision of 0.0001

34.4 Experiment Results

In order to validate such kind of compensation method, the actual 3D spatial target simulator is employed to simulate spatial targets as well as double-theodolites constituting 3D spatial coordinates of frame as shown in Fig. 34.6.

In experiment, target simulator is driven to point single dots as in Fig. 34.3. Points of A_1–A_5 are appointed to train NN, and points of A_6–A_{10} are taken as to simulate NN, both of which guarantee system error's existence conditions. The extreme positions of top area and bottom area on simulator are measured, and the coordinate points' distribution is expressed in Fig. 34.7.

Figure 34.7 explains that no matter at the top area or on the bottom area, compensated points' coordinates varies within a certain range, which means that system errors' compensation takes obvious effects on a wide-range vision field.

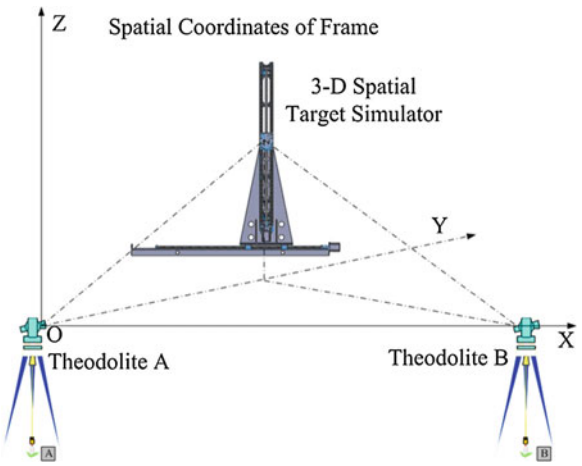


Fig. 34.6 The actual experiment system

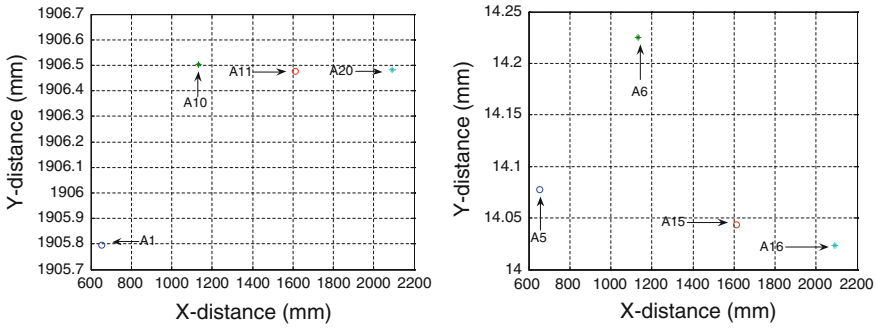


Fig. 34.7 Test dot coordinates at the *top* area and *bottom* area

34.5 Precision Analysis

Distance's uncertainties synthesis equations are shown in expression of (34.5):

$$\begin{cases} u_{x_1} = \sqrt{u_{D1}^2 + u_{D2}^2 + u_{D3}^2} = \sqrt{a_1^2 u_L^2 + a_2^2 u_{\beta_1}^2 + a_3^2 u_{\alpha_1}^2} \\ u_{y_1} = \sqrt{u_{D1}'^2 + u_{D2}'^2 + u_{D3}'^2} = \sqrt{b_1^2 u_L^2 + b_2^2 u_{\beta_1}^2 + b_3^2 u_{\alpha_2}^2} \end{cases} \tag{34.5}$$

According to the principles of equal effect, if the total uncertainties are u_D = 3 mm, expression of (34.6) is given

$$\begin{cases} u_{Dx1} = u_{Dx2} = u_{Dx3} = \frac{1}{\sqrt{3}} u_D = \frac{3\text{ mm}}{\sqrt{3}} = 1.73\text{ mm} \\ u_{Dy1} = u_{Dy2} = u_{Dy3} = \frac{1}{\sqrt{3}} u_D = \frac{3\text{ mm}}{\sqrt{3}} = 1.73\text{ mm} \end{cases} \tag{34.6}$$

34.6 Result

The method of establishing measuring model of three-dimensional spatial coordinates of frame, taking system errors as nonlinear components, training NN with data containing system error, and exciting NN to compute compensation terms are the main job of this paper. The idea of avoiding system errors' influence by means of soft computing is provided, and the idea and method are also validated in actual experiment. The compensation results of distance and coordinates do not absolutely match well, which interprets the randomness of samples' selection, mean arithmetical value of compensated data can reduce random errors to some degree. The procedure of weights optimization reflects NN is capable of adjusting the nonlinear mapping relations, on the special characteristics of which the system errors can be minimized to a certain degree.

Acknowledgments This project is financially supported by Natural Scientific Research Innovation Foundation in Harbin Institute of Technology (No. HIT.NSRIF.2010100) and China Postdoctoral Science Foundation (No. 2013M531025).

References

1. Zili Z, Jigui Z, Na G et al (2010) The design of double-theodolite 3D coordinate measurement system. Chin J Sens Actuators 23(5):660–664 (in Chinese)
2. Fuqiang Z, Guangjun Z, Jie J et al (2004) Three-dimensional coordinate measuring system with bino-theodolites on site. Chin J Mech Eng 40(1):165–169 (in Chinese)
3. Wei L, Jin G (2010) The research on servo control technology of O-E theodolite based on angular acceleration sensor. In: International conference on computer, mechatronics, control and electric engineering (CMCE), pp 629–632
4. Hu Z, Jigui Z, Shenghua Y et al (2008) Resolution improvement of the electronic theodolite in automatic guided laser theodolite system by subdivided locating method of image. In: Proceedings of SPIE the international society for optical engineering, Beijing, China, pp 716001–716009
5. Yanbin L, Guang J, Huiyang H (2004) Research on control error of an target simulator. Acta Simulata Systematica Sin 16(4):820–824 (in Chinese)
6. Luna CA, Mazo M, Lazaro JL et al (2005) Reduction of vibration-induced errors in 3D measurement system with vision and structured light. In: IEEE ISIE pp 1291–1296, 20–23 June 2005
7. Handel H (2008) Compensation of thermal error in vision based measurement systems using a system identification approach. In: ICSP2008 Proceedings, pp 1329–1333
8. Nagatomo S, Hayashi J, Hata S (2009) Proposal of calibration error-tolerant 3D measurement method using stereo vision for robot vision. In: Proceedings of the 2009, IEEE international conference on mechatronics and automation, pp 3263–3268
9. Huang J, Kosaka S, Imamura Y et al (2007) Measurement of distance error in reaching movement of the human hand without using vision. In: Proceedings of the 29th annual international conference of the IEEE EMBS pp 4866–4869
10. Chao W, Mingrong L, Zengtai Z (2007) Application of a pair of theodolites in the height dimensional precision structural framework measurement. Electro-Mech Eng 23(2):1–5
11. Netterville DDJ, Djurfors SG (1979) Controlling inherent uncertainties in double theodolite measurements. Am Meteorol Soc 18(2):1371–1375
12. Schaefer Joseph T, Charles A, Doswell lll (1978) The inherent position errors in double-theodolite pibal measurement. J Appl Meteorol 17(3):911–915
13. Lenz James, Edelstein Alan S (2006) Magnetic sensor and their applications. IEEE Sens J 6 (3):631–649
14. McElvain J, Campbell SP, Miller J et al (2010) Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems. Proc SPIE Digit Photogr 7537(1):1–11
15. Elizabeth Petrik. "Theodolite Instructions" *ACME Experiment*, 2011
16. Jacqueline M, de Faria L, Katsumi O, Arai M et al (1998) Objective measurement of contrast sensitivity function using contrast sweep visual evoked responses. Br J Ophthalmol 82:168–173
17. Min Z, Zongming Q, Jiamin Q et al (2008) Guideless spatial cooperate measurement technology based on coding pole Chinese. J Mech Eng 21(1):87–91 (in Chinese)
18. Schneck ME, Haegerstom-Portnoy G, Lott LA et al (2010) Monocular vs. binocular measurement of spatial vision in elders. Optom Vis Sci 87(8):526–530
19. Zheng X, Li C, Zhao X et al (2010) Depth measurement using monocular stereo vision syste-aspect of spacial discretization. Proc SPIE Optoelectron Imaging Multimedia Technol 7850(1): 1–9

20. Xie H, Jiang A, Wurdemann HA et al (2014) Magetic resonanace-compatible tactile force sensor using fiber optics and vision sensor. IEEE Sens J 14(3):829–838
21. Yan L, Yan-feng Q, Wan-xin S (2005) Double-theodolite measurement system used in the image calibration of space photographic instrument. Optoelectron Lett 3(15):0213–0216
22. Bin W, Wang B (2013) Automatic measurement in large-scale space with the laser theodolite and vision guiding technology. Adv Mech Eng 2013:1–9

Chapter 35
Hopfield Neural Network with Chaotic Positive Feedback and Its Application in Binary Signal Blind Detection

Guangyin Wu, Shujuan Yu, Rusong Huan, Yun Zhang and Kuiming Ji

Abstract This paper presents a blind signal detection algorithm based on linear Chaotic Positive Feedback Hopfield Neural Network (CPFHNN). The algorithm uses sequence with chaos initialization as the transmitting signal and utilizes the HNN with positive feedback to solve the quadratic programming performance function of blind detection and to achieve BPSK signal blind detection. This paper constructs a new energy function of CPFHNN and proves the stability of CPFHNN through simulation by configuring network parameters under asynchronous update mode and synchronous update mode. Compared with the literature without chaotic positive feedback Hopfield neural network blind signal detection algorithm, CPFHNN requires shorter receive data to reach the real global balance point, and reduces the calculation difficulty greatly and has a good quickness.

Keywords Chaotic positive feedback · Blind detection · BPSK signal

35.1 Introduction

In order to overcome the shortcomings of the second-order statistics literature blind algorithms and higher order statistics blind algorithm, which needs sending a lot of data belongs character set information and can not only be used in common zeroes channel [1], but also the use of Hopfield neural network (HNN) for blind detection algorithm has a certain research. However, HNN for blind detection only uses the negative feedback signal through active function iteration [2–4]. By studying the structure of HNN, we find dynamic neuron unit can construct complex dynamic neural feedback dynamic neural networks and neural networks with positive feedback can be constructed with a number of stable equilibrium points of the dynamic

G. Wu (✉) · S. Yu · R. Huan · Y. Zhang · K. Ji
Circuits and Systems Laboratory, Nanjing University of Posts and Telecommunications, Nanjing 210003, Jiangsu Province, People's Republic of China
e-mail: yyiner108@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_35

nervous system [5, 6]. Chaotic signal has the sensitivity to initial conditions, non-cyclical and continuous broadband width and other characteristics. Generally speaking, the effect of chaotic positive feedback of HNN is stable, the equivalent of hysteresis. However, if the feedback is negative, the network will generate oscillation or chaos. And the oscillation depends on the size of the connection weights from the feedback entirely [6–9].

Owing to these drawbacks, the model of Chaotic Positive Feedback Hopfield Neural Networks (CPFHNN) is proposed in this paper to solve the problem of blind detection. Depending on the practical problems, this paper constructs a new energy function optimization problem and blind detection model. By analyzing the given energy function, CPFHNN model is proved to be stable. Finally, we confirm CPFHNN with exceptional properties based on the transmitting signal.

The rest of this paper is structured as follows. Section 35.2 outlines the model of Chaotic Positive Feedback Hopfield neural networks and expressions, followed by Stability Analysis of energy function in Sect. 35.3. Section 35.4 gives the configuration of the weight matrix of blind recovery signal. And simulation results are presented in Sect. 35.5. Finally, Sect. 35.6 concludes the paper.

35.2 Construction of Chaotic Positive Feedback Hopfield Neural Network

Through different internal link structure, Dynamic neurons can construct a variety of complex dynamic neural networks. This paper adds a positive self-feedback for each, and then links these neurons to construct a new neural network model [10–12]. The CPFHNN architecture is shown on Fig. 35.1.

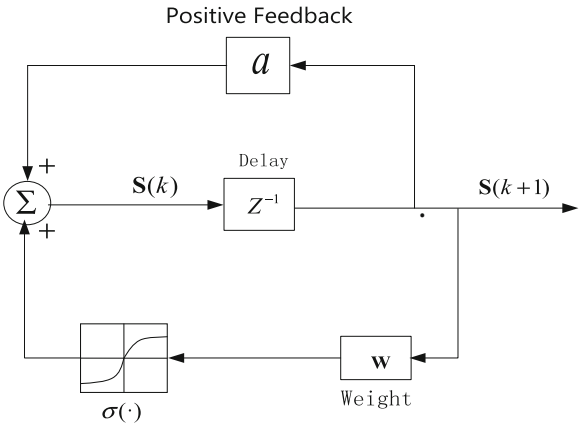


Fig. 35.1 Architecture of the CPFHNN

This model is expressed as:

$$\mathbf{S}(k+1) = \alpha \mathbf{S}(k) + \delta(\mathbf{W}\mathbf{S}(k)) \tag{35.1}$$

where $\mathbf{S} = [s_1, s_2, \ldots, s_N]$ is the vector of Neural state, α is the parameters of Linear self-feedback of CPFHNN, $\mathbf{W}$ is the weights matrix of Network connection and $\mathbf{W} \in C^{N\times N}$, $\mathbf{W}^T = \mathbf{W}$.

35.3 Stability Analysis of the Architecture of CPFHNN

There are various methods to determine the stability of neural networks. As used herein, the method is constructing Lyapunov energy function based on the networks, and determining the network stability according to Lyapunov second theorem. This paper constructs a new Lyapunov energy function [7, 8, 10].

$$\mathrm{E}(\mathrm{S}(k)) = -\frac{1}{2}\mathbf{S}^T(k)\mathbf{W}\mathbf{S}(k) + (1-\alpha)\sum_{i=1}^{N}\int_0^{S_i(k)} \sigma^{-1}(\tau)d\tau \tag{35.2}$$

In this paper, we will analyze the stability of the networks in asynchronous and synchronous update mode. So, we have the following theorem.

Theorem 35.1 *CPFHNN is composition of N neurons, weight matrix* $\mathbf{W}^T = \mathbf{W}$, *diagonal elements* $\omega_{ii} > 0$, *and positive feedback coefficient is* $0 \le \alpha \le 1$, *Lyapunov energy function in asynchronous update mode is defined in* (35.2).

Proof Let us consider $\Delta E(k) = E(k+1) - E(k)$, and only one neuron state is updated in each feedback vector of Asynchronous update mode. Without loss of generality, suppose that only s_i(k) is updated, so $\Delta s_i = s_i(k+1) - s_i(k) \neq 0$ and$\Delta s_j = s_j(k+1) - s_j(k) = 0, \mathrm{j} \neq \mathrm{i}$.

$$\begin{aligned}\Delta E(k) &= E(k+1) - E(k)\\ &= -\frac{1}{2}\Delta \mathrm{s}_\mathrm{i}^2\omega_{\mathrm{ii}} - \Delta \mathrm{s}_\mathrm{i}\cdot\sigma^{-1}(\mathrm{s}_\mathrm{i}(\mathrm{k}+1) - \mathrm{as}_\mathrm{i}(\mathrm{k})) + (1-\mathrm{a})\int_{\mathrm{s_i(k)}}^{\mathrm{s_i(k)+\Delta s_i}} \sigma^{-1}(\tau)\mathrm{d}\tau\end{aligned} \tag{35.3}$$

According to $\int_{s_i(k)}^{s_i(k+1)} \sigma^{-1}(x)\mathrm{d}x = \Delta s_i \cdot \Delta\sigma^{-1}(s_i(k+1)) - \frac{1}{2}(s_i)^2[\sigma^{-1}(\xi_i)]'$, while $s_i(k) \le \xi_i \le s_i(k+1)$. So, (35.3) can be rewritten as:

$$\Delta E(k) = -\frac{1}{2}\Delta s_i^2\left\{\omega_{ii} + (1-a)\left[\sigma^{-1}(\xi_i)\right]'\right\} + \Delta s_i \cdot \left[(1-a)\sigma^{-1}(s_i(k+1)) - \sigma^{-1}(s_i(k+1) - as_i(k))\right] \tag{35.4}$$

In the paper, this activation function selected as $\sigma(\mathrm{x}) = \tanh(\mathrm{x})$ is a monotonically increasing function in $(-\infty, +\infty)$ [10, 12]. According to the nature of the inverse function, since $\sigma^{-1}(\mathrm{x})$ is also a monotonically increasing function for $(-1, +1)$, so $[\sigma^{-1}(x)]' \geq 0$. As long as $(1-\alpha) \geq 0$ and $\omega_{ii} > 0$, we can guarantee $-\frac{1}{2}\Delta s_i^2\left\{\omega_{ii} + (1-\alpha)[\sigma^{-1}(\xi_i)]'\right\} \leq 0$. Since the detection signal is BPSK, then the detection signal just has two polarities.

If $s_i(k) < s_i(k+1)$, then there must be $s_i(k) < 0$ and $s_i(k+1) > 0$. And $\alpha \geq 0$ can make $s_i(k+1) \leq s_i(k+1) - \alpha s_i(k)$ come true. The feature of $y = \sigma^{-1}(x), \alpha \leq 1$ makes $0 < (1-\alpha)\sigma^{-1}(s_i(k+1)) \leq \sigma^{-1}(s_i(k+1) - \alpha\, s_i(k))$ come true. Similarly $0 < (1-\alpha)\sigma^{-1}(s_i(k+1)) \leq \sigma^{-1}(s_i(k+1) - \alpha\, s_i(k))$, when $s_i(k) < s_i(k+1)$.

Depending on the above analysis, we obtained that Δs_i and $(1-\alpha)\sigma^{-1}(s_i(k+1)) - \sigma^{-1}(s_i(k+1)) - \alpha\, s_i(k)$ are of the opposite sign. Therefore, there is always $\Delta s_i[(1-\alpha)\sigma^{-1}(s_i(k+1)) - \sigma^{-1}(s_i(k+1) - \alpha\, s_i(k))] \leq 0$.

In summary, as long as there are $0 \leq \alpha \leq 1$ and $\omega_{ii} > 0$, we can guarantee $\Delta E(k) \leq 0$. If $\mathbf{S}(k) = \mathbf{S}(k+1)$, $\Delta E(k)$ is equal to zero and the network reaches to the equilibrium point. It means that the network energy decreases during operation and converges to a stable equilibrium state finally. This completes the proof.

Theorem 35.2 *CPFHNN is composition of N neurons, suppose* $\mathbf{W}^T = \mathbf{W}$,$\mathbf{W}$ *is Positive definite matrix and Positive feedback coefficient is* $0 \leq \alpha \leq 1$, *Lyapunov energy function in synchronous update mode is defined in* (35.2).

Proof Let us consider $\Delta E(k) = E(k+1) - E(k)$, and the whole of the neuron states is updated in all feedback vectors of synchronous update mode. Suppose that $G_i(s_i(k)) = \int_0^{s_i(k)} \sigma^{-1}(\tau)\mathrm{d}\tau$, we rearrange the Eq. (35.2) as follow: $E(\mathrm{k}) = -\frac{1}{2}\mathbf{S}^T(\mathrm{k})\mathbf{WS}(\mathrm{k}) + (1-\alpha)\sum_1^N G_i(s_i(k))$. Consider the change E in between two discrete time steps under the assumption of synchronously updating:

$$\begin{aligned}\Delta E(k) &= E(k+1) - E(k) \\ &= -\frac{1}{2}\Delta\mathbf{S}^{\mathrm{T}}(k)\mathbf{W}\Delta\mathbf{S}(k) - \Delta\mathbf{S}^{\mathrm{T}}(k)\mathbf{WS}(k) + (1-a)\sum_{i=1}^{N}\Delta G_i(s_i(k))\end{aligned} \tag{35.5}$$

Besides, we can get expression $\Delta G_i(s_i(k)) = \int_{s_i(k)}^{s_i(k+1)} \sigma^{-1}(\tau)\mathrm{d}\tau = \Delta s_i(k)\sigma^{-1}(\xi_i)$ by Integral mean value theorem. Where, $s_i(k) \leq \xi_i \leq s_i(k+1)$ or $s_i(k+1) \leq \xi_i \leq s_i(k)$. Hence the following inequality holds.

$$\Delta E(k) = -\frac{1}{2}\Delta \mathbf{S}^{\mathrm{T}}(k)\mathbf{W}\Delta \mathbf{S}(k) - \sum_{i=1}^{N} \Delta s_i(k)\left[\sigma^{-1}(s_i(k+1) - as_i(k)) - (1-a)\sigma^{-1}(\xi_i)\right] \tag{35.6}$$

If $s_i(k+1) > s_i(k)$, then there must be $s_i(k+1) > 0$, $s_i(k) < 0$. And $\alpha \geq 0$ can make $s_i(k+1) - as_i(k) > \xi_i$ come true. Since $[\sigma^{-1}(x)]' > 0$, according to the feature of $y = \sigma^{-1}(x)$, we can get $\sigma^{-1}(s_i(k+1) - as_i(k)) > (1-a)\sigma^{-1}(\xi_i), a \leq 1$. Similarly, when $s_i(k+1) < s_i(k)$, $\sigma^{-1}(s_i(k+1) - as_i(k)) > (1-a)\sigma^{-1}(\xi_i)$.

From the analysis, $\Delta s_i(k)[\sigma^{-1}(s_i(k+1) - as_i(k)) - (1-a)\sigma^{-1}(\xi_i)] \geq 0$.

In summary, positive definite matrix $\mathbf{W}$ and $0 \leq a \leq 1$ can guarantee $\Delta E(k) \leq 0$ in synchronous update mode. At last, we can draw out that this network is stable under the Lyapunov second theorem.

35.4 The Configuration of Weight Matrix of Blind Recovery Signal

In case of no noise, SIMO (Single-Input Multi-Output) receiving signals and blind treatment equation in a digital communication system are as follows [13]:

$$(x(k))_{q\times 1} = [h_0, \ldots, h_M](s(k))_{(M+L+1)\times 1} = \sum_{j=0}^{M} (h_j)_{q\times 1} S(k-j) \tag{35.7}$$

$$\mathbf{X}_N = \mathbf{S}\Gamma^{\mathrm{T}} \tag{35.8}$$

where, q is over-sampling factor, M is channel order, L is equalizer coefficients, $[\mathbf{h}_0, \ldots, \mathbf{h}_M]$ is impulse response of communication channel, $\mathbf{X}_N$ is receiving data matrix and $(\mathbf{S})_{N\times(L+M+1)}$ is transmitting signal matrix. Depending on (35.8), when Γ is full column rank, there must be $\mathbf{Q} = \mathbf{U} \cdot \mathbf{U}_c^{\mathrm{T}}$ to make $\mathbf{Q}\mathbf{S}_N(k-d) = 0$ come true. Then, $\{\mathbf{S}_N(k-d) | d = 0, 1, \ldots, M+L\}$ and $\mathbf{U} \in R^{N\times(N-(L+M+1))}$ is unitary matrix which is decomposed from $\mathbf{X}_N = [\mathbf{U}, \mathbf{U}^{\mathrm{T}}] \cdot \begin{bmatrix} \mathbf{D} \\ 0 \end{bmatrix} \cdot [\mathbf{V}^{\mathrm{T}}]$. Therefore, construct optimization problems function of performance function as follows:

$$J_0 = \mathbf{S}_N^T(k-d)\mathbf{Q}\mathbf{S}_N(k-d) = \mathbf{S}^T\mathbf{Q}\mathbf{S} \tag{35.9}$$

$$\hat{\mathrm{s}} = \mathrm{argmin}\{\mathrm{J}_0\} \tag{35.10}$$

In order to use CPFHNN solving formula (35.8) for blind signal detection problem, so we make a network connection matrix configuration as form:

$$\mathbf{W} = \mathbf{I} - \mathbf{Q} \tag{35.11}$$

When $\mathbf{S}(k) = \mathbf{S}(k-1)$, CPFHNN can reach the equilibrium point. So the extreme point of the energy function $E(k)$ is the point of the optimization problem. And the point which we get is also transmitting signal needed detecting.

35.5 Simulation Result

This experiment uses BPSK signal as the transmitting signal sequence. Channel noise is Gaussian white noise. Each simulation result comes out of 100 Monte Carlo experiments. Meanwhile, in order to facilitate mapping, simulation experiments regard BER less than or equal to 10^{-5} as Zero.

Experiment 1: When the length of data is fixed, we make BER comparison of CPFHNN, HNN and second order statistics algorithm of literature for SSA and TXK [14, 15] by Monte Carlo experiments. The transmitting signal is BPSK. The amount of data of input signals is 100. Where, $\alpha = 0.2$ and $\alpha = 0.7$ mean that positive feedback coefficient of neural network are 0.2–0.7. Figure 35.2 shows arithmetic average BER.

From Fig. 35.2, we know that CPFHNN algorithm outperforms HNN algorithm, second-order statistical algorithm SSA algorithm and TXK algorithm in the BER. And by comparing, the CPFHN algorithm has the strongest anti-interference ability. And the CPFHNN algorithm feedback coefficient is greater, the BER performance

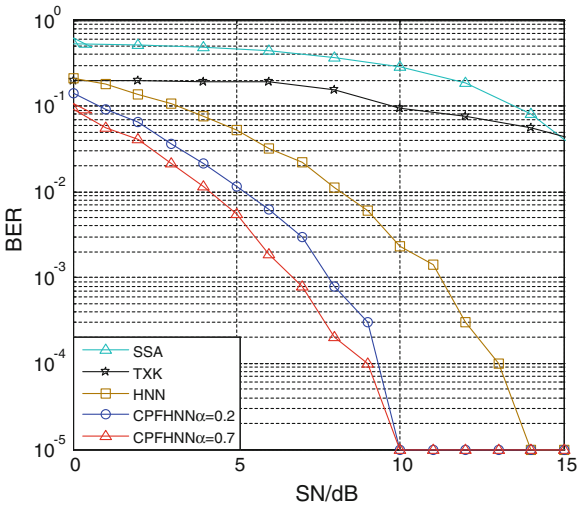


Fig. 35.2 Different algorithms' BER comparison

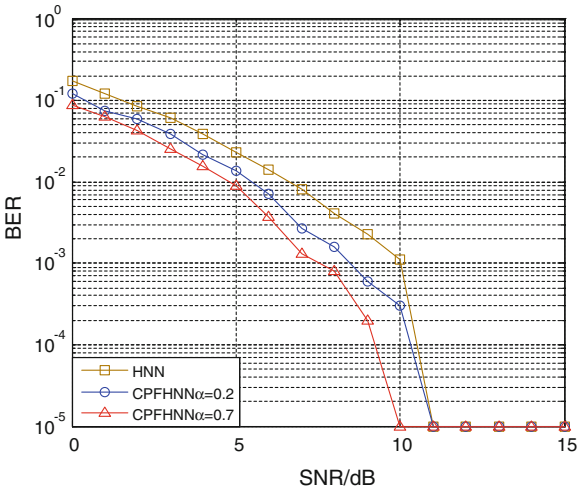


Fig. 35.3 In one common zero channel, different algorithms' BER comparison

will be better. Actually, it is worth noting that the time of CPFHNN algorithm will also be a corresponding increase if the feedback coefficient is greater.

Experiment 2: When the length of data is fixed, we carry out Monte Carlo algorithms experiment for CPFHNN and HNN in the channel with common zeroes.

The channel selected by Fig. 35.3 is a multipath of random real number channel containing a common zero. Where, set a common zero (−0.5) with random weights and random delay factor. Figure 35.4 is simulated in a multipath random real number channel containing two common zeroes, which are (3.0000) and (−0.5) with random weights and random delay factor. Where, $\alpha = 0.2$ and $\alpha = 0.7$ mean that positive feedback coefficient of neural network are 0.2–0.7.

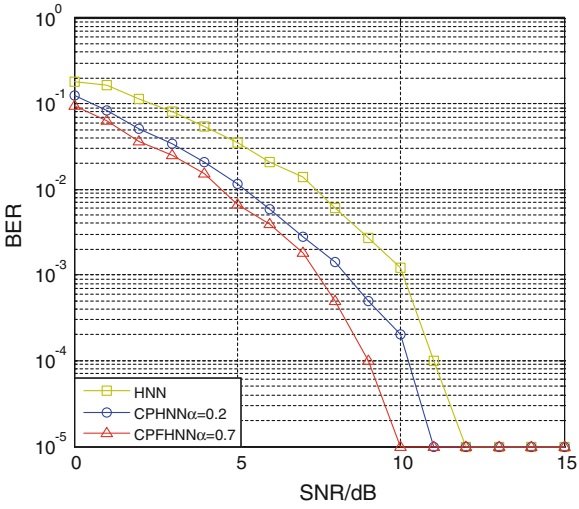


Fig. 35.4 In two common zeroes channel, different algorithms' BER comparison

From Figs. 35.3 and 35.4, the CPFHNN algorithm can resolve the problem of channel with common zeroes. Compared with the traditional HNN algorithm, CPFHNN algorithm has better anti-noise performance. Also, the CPFHNN feedback coefficient is greater, and the BER will be better in the channel with common zeroes.

35.6 Conclusions

We introduce positive feedback neural network into the blind detection algorithm, and present a Lyapunov function corresponding to this algorithm. The proposed algorithm has better performance of BER than the traditional network under the same conditions. Simulation shows noise resistance ability of CPFHNN is stronger than the traditional HNN, and the CPFHNN can solve the issue about the channel with common zeroes. Furthermore, performance of CPFHNN of BER is superior to the traditional HNN network. Besides these, CPFHNN exhibits strong robustness in the unusual difference of transmitting data.

It remains future work to put the network structure and the function of Lya-punov in this paper for multi-valued state.

Acknowledgement Project supported by National Natural Science Foundation of China under Grants No. 61104103, 61274080, 61302155. Jiangsu Province Science and Technology Department General Program No. BK2011753. Natural Science Foundation of the Jiangsu Higher Education Institutions of China (12KJB510014).

References

1. Zhang Y (2012) Blind hopfield neural network based on complex signal detection. Nanjing University of Posts and Telecommunications Library (in Chinese)
2. Yi S, Jun W (2012) Robustness analysis of global exponential stability of recurrent neural networks in the presence of time delays and random disturbances. IEEE Trans Neural Netw Learn Syst 23:87–96
3. Ruan XK, Zhang Z (2011) QAM signal blind detection based on continuous hopfield neural networks. J Electron Inf Technol 33:1600–1605 (in Chinese)
4. Yun Z, Zhi-Yong Z (2011) 64QAM signal detection based on complex discrete hopfield network blind. J Electron Inf Technol 33:315–320 (in Chinese)
5. Noriyasu H, Madan MG, Jin L (2003) static and dynamic neural networks. Wiley, New York
6. Horne BG, Hush DR (1993) Progress in supervised neural networks. IEEE Signal Process Mag 10:8–39
7. Smith K, Wang LP (1998) On chaotic simulated annealing. IEEE Trans Neural Netw 9:716–718
8. Zhang L, Zhang DZ (2010) A self-feedback stability analysis of cellular neural network and its application in function optimization. Control Decis 25:53–56 (in Chinese)
9. Nong J (2012) Global exponential stability of delayed hopfield neural networks. In: International conference on computer science and information processing (CSIP), Xi'an, China

10. Jiang Y (2011) A wavelet chaotic simulated annealing neural network and its application to optimization problems. Network Computing and Information Security (NCIS). Guilin, China
11. Chen LN, Aihara K (1997) Chaos and asymptotical stability in discrete-time neural networks. Physica 104:286–325
12. Aurèle B, Christopher JR, Justin R (2012) Convergence and rate analysis of neural networks for sparse approximation. IEEE Trans Neural Netw Learn Syst 23:1377–1389
13. Bai ER, Yong ZZ (2005) Direct blind sequence detection SIMO channels including public zero. Acta Electronica Sinica 33:671–675
14. Guanghan Xu, Tong Lang, Kailath T (1994) Blind identification and equalization based on second-order statistics:a time domain approach. IEEE Trans Inf Theory 40:340–349
15. Moulines E, Cardoso JF, Duhamel P, Mayrargue S (1995) Subspace methods for the blind identification of muitichannel FIR filters. IEEE Trans Signal Process 43:516–525

Chapter 36
Image Annotation with Nearest Neighbor Based on Semantic Information

Wei Wu and Guanglai Gao

Abstract Most of the Nearest Neighbor (NN)-based image annotation methods do not achieve desired performances. The main reason is that much valuable information is lost when extracting visual features from image. In this paper, we propose a novel weighted NN-based method. Instead of using Euclidean distance, we learn a new distance metric with image semantic information to calculate the distance between the two images. Meanwhile, we utilize textual information of each image tagged by users to form weights of NN-based model. When introducing the semantic information, our method can minimize the semantic gap for intraclass variations and interclass similarities, and improve the annotation performance. Experiments on image annotation dataset of ImageCLEF2012 show that our method outperforms the traditional classifiers. Moreover, our method is simple, efficient, and competitive compared with state-of-the-art learning-based models.

Keywords Image annotation · Nearest neighbor · Distance metric learning · Entropy weight

36.1 Introduction

Image annotation and retrieval have drawn considerable attention in both research and practical areas. The goal of image annotation is to automatically recognize visual concepts from image semantic concepts set, and turns out to be extremely challenging due to the large intraclass variations and interclass similarities. Recently, there have been many research communities engaged in this work, such as ImageCLEF [1], TRECVID, Pascal VOC, etc., which confirm the challenges in this field.

W. Wu (✉) · G. Gao
Department of Computer Science, Inner Mongolia University,
No.235 West College Road, Hohhot, China
e-mail: cswuwei@imu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_36

The image annotation methods often use learning-based classifiers, and rarely use Nearest Neighbor (NN)-based classifiers, because they provide inferior performance relative to learning-based methods. But we may underestimate the effectiveness of NN-based method. Boiman et al. [2] claim that the main reason resulting in the low performance of NN-based algorithms is the information loss when extracting image visual features, particularly when constructing bag of visual words (BoVW)-based features. BoVW-based features are harmful in the case of NN-based model, which has no training phase to compensate for this loss of information. The method proposed by [2] does not use BoVW model, but directly uses local features for NN-based classifier, and achieves better performance than learning-based models.

In this paper we propose a novel NN-based method which can greatly reduce the semantic information loss, thereby improving the performance of large scale image annotation. We still use BoVW features, but introduce the image semantic information for computing distance between images. In our model, we first utilize image semantic information for distance metric learning (DML) [3, 4], and get a new distance measure. Then we generate multiple clusters for each image category using *k*-means algorithm based on this new distance. Each cluster contains a set of images and some user textual tags (We also regard them as image semantic information). We then assign a weight to each cluster according to the importance of these textual tags, and construct a semantic weighted NN-based classifier.

There are some existing works related to NN-based model. Blitzer et al. [5] learn a Mahanalobis distance metric for the traditional *k*NN model. Wang et al. [6] propose image-to-class-based NN model. Wang et al. [7] introduce the semantic relations based on WordNet for distance metric learning. Verma and Jawahar [8] present a two-step variant of the classical *k*NN algorithm. Our method is different from all the above-mentioned methods, we use a different distance metric optimization strategy, and meanwhile, we introduce the user tag information for calculating weights, and propose a novel weighted NN-based framework. Experiments on dataset of ImageCLEF2012 [1] image annotation task confirm the effectiveness of our method, the result of our model outperforms the traditional classifiers and a new baseline of NN model [9], and is competitive compared with state-of-the-art models.

The paper is organized as follows: Sect. 36.2 describes the DML using semantic information, and Sect. 36.3 introduces our NN-based model. Section 36.4 describes the experiments and results. Finally, we conclude our work and shed light on the future work in Sect. 36.5.

36.2 Distance Metric Learning

The objective of DML is to find an optimal Mahalanobis metric A from training data. In our method, we extract the pairwise constraints from training images for distance metric learning. We formalize the representation of the pairwise features

constraints set as $\{(f_{i1}, f_{i2}, y_i)\}_{i=1}^{N}$, where f_{i1} and f_{i2} are two image features. And if both f_{i1} and f_{i2} belong to the same image category, then $y_i = 1$, otherwise $y_i = -1$. It is worth noting that how to select pairwise constraints can greatly affect the annotation performance. For the image semantic annotation task, there are the large intraclass variations and interclass similarities, so we comply with such selection criterion: one is that the features are of the same image category but with large variation, the other is that the features are of different image category but with large similarity.

Specifically, we firstly extract features of all the training images and use the *k*-means algorithm in Euclidean distance space to cluster the image features for each image category, with the result that *k* centers are formed for each image category. Then we regard these centers as visually different "images" in the same semantic category (namely, the images with a large intraclass variation), and for each pair of these images, we construct pairwise constraints ($f_{i1}, f_{i2}, y_i = 1$). Last, for each center of an image category, we search for the closest image in Euclidean distance in any other image category (namely, the images with a high interclass similarity), and construct pairwise constraints ($f_{i1}, f_{i2}, y_i = -1$).

Given the feature pairwise constraints information, the goal of our task is to learn a distance metric A to effectively measure distance between any two image features f_{i1} and f_{i2}, which can be represented as formula (36.1):

$$d(f_{i1}, f_{i2}) = \sqrt{(f_{i1} - f_{i2})^T A (f_{i1} - f_{i2})} \tag{36.1}$$

To find an optimal metric A, the distances between visual features of the same semantic category should be minimized, and meanwhile distances between features of different semantic category should be maximized. Based on this principle, we formulate this distance metric learning problem into the following optimization:

$$\begin{aligned} \min_{A,b} \; f(A,b) = \sum_i y_i(||f_{i1} - f_{i2}||_A^2 - b) + \frac{\lambda}{2}\mathrm{tr}(A^T A) \\ s.t. \quad y_i(||f_{i1} - f_{i2}||_A - b) \le 1 \\ A \ge 0, ||A|| = 1/\sqrt{\lambda} \end{aligned} \tag{36.2}$$

where $\| \bullet \|_A$ is the Mahalanobis distance between two features under metric A. With the first inequality constraints, minimizing this term will make the distance between two semantically identical image features closer. The second term of the objective function is the regularization term, which prevents the overfitting by minimizing this model. The second constraint is introduced to prevent the trivial solution by shrinking metric A into a zero matrix. Parameter λ is a constant, b is a threshold. We use a stochastic gradient search algorithm to solve this optimization problem [5]. The algorithm is an iterative process, and empirically, this iterative algorithm converges quickly with no more than five iterations.

36.3 Nearest Neighbor-Based Model

We first use k-means clustering method to construct clusters for each training image category. Instead of using Euclidean distance, we use our trained distance metric when running the clustering algorithm, and this is the main difference with [6]. Thus we get k cluster center features for each image category: $f_1, f_2, \ldots, f_k$. Now our work is to search out the image class C which minimizes the sum $\sum_{i=1}^{k} d(f_{\text{test}}, f_i^C)$, where the distance function $d(\cdot)$ is based on the new distance, shown by formula (36.1), f_{test} is the feature of test image, and C denotes the image category, f_i^C denotes the ith cluster feature of image class C, and k takes the same value for all the image class. We also consider that each cluster contains a set of images and some textual terms. So we can utilize this semantic information to assign a weight for each cluster. The major idea is that, the higher the frequency of a term in a cluster is, the more representative this cluster will be. In contrast, if a large number of different terms occur in a cluster, this cluster would be not well representative for related image class [10]. We can calculate entropy according to the terms in a cluster, and this entropy can be viewed as a weight for a cluster. The higher the weight of a cluster is, the greater the distance will be to this cluster. So our NN-based classifier can be changed to this form: minimizing the sum $\sum_{i=1}^{k} w_i^C d(f_{\text{test}}, f_i^C)$, where w_i^C is the entropy weight of the ith cluster for image class C. The entropy can be calculated as follows:

Considering training images for class C are divided into k clusters, $\{C_1, C_2, \ldots, C_k\}$, and there are y unique textual terms $\{t_1, t_2, \ldots, t_y\}$. Assume that a cluster contains several images and each image is assigned several textual terms (It is a truth for dataset of ImaegCLEF [1]). Hence a cluster can be viewed as a collection of terms, $C_i = \cup\, t_j$. The entropy of the ith cluster can be defined as:

$$\text{Entropy}^i = \sum_{t_j \in C_i} \left(\frac{tf_{t_j}^i}{\sum_{t_v \in C_i} tf_{t_v}^i} \log \left(\frac{\sum_{t_v \in C_i} tf_{t_v}^i}{tf_{t_j}^i} \right) \right) \tag{36.3}$$

where $tf_{t_j}^i$ denotes the frequency of term t_j in the ith cluster. Our classifier can therefore be summarized as follows:

1. Constructing k clustering centers for each image category C: $(f_1^C, f_2^C, \ldots, f_k^C)$.
2. Calculating the k entropy weights for each image category C: $(w_1^C, w_2^C, \ldots, w_k^C)$.
3. Computing the visual feature f_{test} of test image.
4. Classification result:

$$\hat{C} = \arg\ \min_C \sum_{i=1}^{k} w_i^C d(f_{\text{test}}, f_i^C) \tag{36.4}$$

When applying to the multilabel image annotation problem, we need only to compute the sum $\sum_{i=1}^{k} w_i^C d(f_{\text{test}}, f_i^C)$ for each image class, and then sort the class labels in ascending order according to these sums.

36.4 Experiments and Results

Experimental images are from the image annotation dataset of ImageCLEF2012 [1]. There are a total of 94 concept categories for annotation. The range of these concepts is fairly wide, including natural elements, environments, people, impression, transportation, etc. There are 15,000 images for training and 10,000 images for testing, within the range of 94 concepts, and each of these images has several user tags provided by organization. We need to allocate each test image with multiple concept labels, and then sort these labels according to the similarities between the image and labels. The evaluation measurement is the MiAP (Mean interpolated Average Precision) which is widely used in the field of image annotation and retrieval.

36.4.1 Experiments with Different Features

We select three features for experiments. They are Color Histograms, Fuzzy Color and Texture Histogram (FCTH), and BoVW based on SIFT local features. The size of BoVW is fixed at 500 considering the balance between the annotation performance and computational cost [11].

First, we test the above three features using the traditional *k*-NN classifier. The best result we obtained is using SIFT features, and the value of MiAP is 0.2702 when parameter *k* takes 50. Then we test our method using the same features. The experimental results of ours are plotted in Fig. 36.1. In Fig. 36.1, we use formula (36.1) to calculate distances for all the features except SIFT-JSD. SIFT-JSD denotes JSD distance for SIFT local features. The parameter *k* in Fig. 36.1 is the number of clusters for each image category.

We can see that the SIFT local features get the best result in our method, MiAP reaches 0.3143 without entropy weight, which is higher than traditional *k*-NN

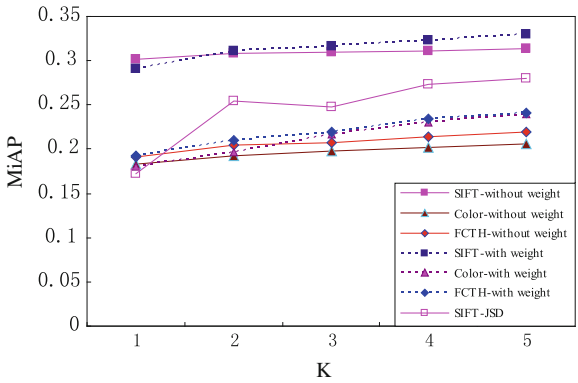


Fig. 36.1 Results of our method (*k* is the parameter of *k*-means algorithm)

method, the MiAP is 0.2702. And MiAP achieves 0.3297 with entropy weight, 1.5 % higher than without entropy weight, which confirms that our weighted strategy is effective.

And we also learn that the curve is relatively flat in Fig. 36.1, which means that the parameter k has not much effect on performance. From the point of view of computational cost, the value of k of our method is far less than the traditional method. Actually, when the value of k is 1, the performance is much better than the traditional k-NN. In our experiments, we test only the value of parameter k to 5, and it is shared by all the image classes.

Finally, we also learn that the use of semantic distance indeed increases the performance. We can see that the result using SIFT local features with semantic distance is better than JSD distance. This shows that the introduction of semantic distance is effective.

36.4.2 Experiments with Other Methods

We also compare our method with other classifiers, as shown in Fig. 36.2. The methods for comparison are traditional k-NN, distance weighted k-NN (dw-kNN), Naive Bayesian (NB), NBNN method proposed by Boiman et al. [2], Baseline [9], and SVM model. The results obtained by these models use the same SIFT local features as our method, and the kernel function of SVM we used is Histogram Intersection Kernel (HIK). The model of ImageCLEF means the method which achieved the best result published by ImageCLEF2012 [1] using multiple visual features.

In Fig. 36.2, we learn that the performance of k-NN is close to NB, and the performance of NBNN is very close to SVM. We can see that the result of our method outperforms all the other methods except the best result by Image CLEF2012. This best result using multiple visual features by ImageCLEF2012 achieved 0.3481, slightly better than ours, that is, our method is competitive.

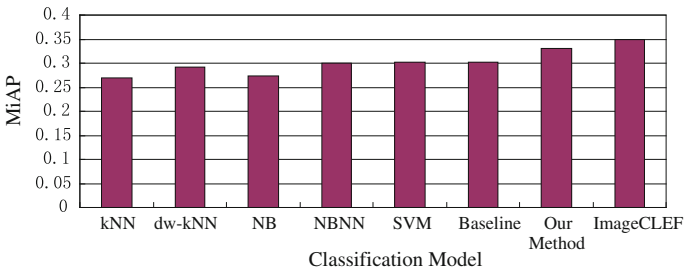


Fig. 36.2 The results comparison with different models

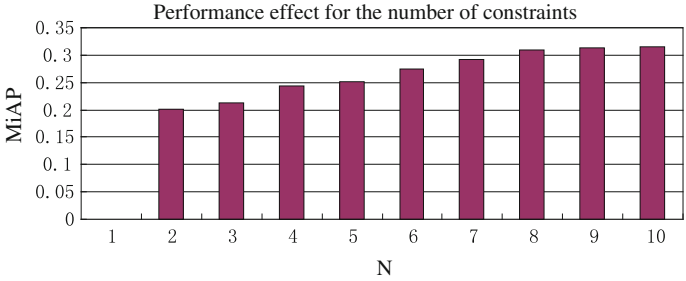


Fig. 36.3 Performance effect of the number of pairwise constraints

36.4.3 Experiments for Distance Metric Learning

Finally, we test the impact on annotation accuracy of the number of pairwise constraints on distance metric learning. We first build N visually different "images" using a clustering method (actually, N denotes the number of clustering centers) for each category. And we extract a pair of features from each of $N \times (N\text{-}1)/2$ pairwise "images" for each category. Then for each of N "images" of each category, we select five different categories of images to construct five pairwise features. Thus for all 94 image categories, we totally have $94 \times (N \times (N - 1)/2 + 5 \times N)$ pairwise feature constraints. We let N take values from 2 to 10, with the result that the number of pairwise constraints varies from 1,034 to 8,930. We use SIFT local features to carry out experiment (see Fig. 36.3).

We can see from Fig. 36.3 that the performance gets better with the increase of the number of pairwise constraints. And in our experiment, we find that the trend of improvement increases slowly when N value exceeds 8. So when considering the trade-off between the computational cost and performance, we take $N = 10$ for the above experiments. And we use the same number of constraints as other visual features. When efficiency is not cared, we think that if we take greater N, we can achieve better performance.

36.5 Conclusion

In this paper we described a novel semantic weighted NN-based classifier based on semantic distance. Our experiments on the ImageCLEF2012 image dataset achieved good results. This confirmed that our method is suitable to large-scale image classification task with high intraclass variations and interclass similarities. In the future, if we can explore more efficient and automatic selecting method of pairwise constraints for DML, it would be more effective for performance.

Acknowledgments This work is supported by Inner Mongolia NSF 2014MS0606.

References

1. Thomee B, Popescu A (2012) Overview of the ImageCLEF 2012 flickr photo annotation and retrieval task. In: CLEF 2012 working notes, Rome, Italy
2. Boiman O, Shechtman E, Irani M (2008) In defense of nearest-neighbor based image classification. In: Proceedings of CVPR, pp 1–8
3. Wang S, Jiang S, Huang Q, Tian Q (2012) Multi-feature metric learning with knowledge transfer among semantics and social tagging. In: Proceedings of CVPR, pp 2240–2247
4. Grauman K, Sha F, Hwang SJ (2011) Learning a tree of metrics with disjoint visual features. In: Advances in neural information processing systems, pp 621–629
5. Weinberger KQ, Saul LK (2009) Distance metric learning for large margin nearest neighbor classification. J Mach Learn Res 10:207–244
6. Wang Z, Hu Y, Chia LT (2010) Image-to-class distance metric learning for image classification, computer vision–ECCV. Springer, Heidelberg, pp 706–719
7. Wang F, Jiang S, Herranz L et al (2012) Improving image distance metric learning by embedding semantic relations, advances in multimedia information processing–PCM. Springer, Heidelberg, pp 424–434
8. Verma Y, Jawahar CV (2012) Image annotation using metric learning in semantic neighbourhoods, computer vision–ECCV. Springer, Heidelberg, pp 836–849
9. Makadia A, Pavlovic V, Kumar S (2008) A new baseline for image annotation, computer vision–ECCV 2008. Springer, Heidelberg, pp 316–329
10. Su JH, Chou CL, Lin CY, Tseng VS (2011) Effective semantic annotation by image-to-concept distribution model. Multimedia IEEE Trans 13(3):530–538
11. Jia Y, Huang C, Darrell T (2012) Beyond spatial pyramids: receptive field learning for pooled image features. In: Proceedings of CVPR, pp 3370–3377

Chapter 37
Tea Leaves Classification Based on Texture Analysis

Zhe Tang, Fang Qi, Yi Zhou, Fangfang Pan and Jianyong Zhou

Abstract An SVM with texture analysis-based feature extraction classification method is presented for identification of fresh tea leaves in this paper. This method is proved to be very efficient and effective in the identification of fresh tea leaves through real experiments. First, the texture characteristic parameters of tea leave images are obtained by texture feature extraction. After that, different categories of fresh tea leaves are identified through SVM training. These texture parameters for texture classification include energy, correlation, and contrast obtained from gray-level co-occurrence matrix (GLCM). Experimental results show that the use of SVM for classification of tea leaves can achieve very good results, and the successful classification rate can be as high as 83 %.

Keywords Texture analysis · GLCM · SVM · Tea leaves classification

37.1 Introduction

Nowadays, fresh tea leaves can be categorized roughly by a conventional machine of fresh-leaf grader, but the classification accuracy is not good enough for tea process. For instance, one bud with one tea leaf may be mixed with two leaves one, while one bud with two leaves normally mixes with multiple leaves even after using a fresh-leaf grader. In addition, the traditional fresh-leaf grader neither classifies tea leaves automatically nor meets the continuous production requirements of tea automation production line. According to the quality requirement of tea processing,

Z. Tang · F. Qi (✉) · F. Pan
School of Information Science and Engineering, Central South University, Changsha, China
e-mail: csqifang@csu.edu.cn

Y. Zhou
Singapore Institute of Technology, 10 Dover Drive, Singapore, Singapore

J. Zhou
Changsha Xiangfeng Tea Machinery Manufacturing Co., Ltd, Changsha, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_37

different process control parameters are crucial for different types of fresh tea leaves. For instance, temperature in water-removing process for one bud with one leaf is higher than that of one bud with two leaves. Moreover, the time and temperature for wind power of winnowing, spreading for cooling, and rolling roasting are also different for different leaves types. In general, the automatic production line of fresh tea leaves should meet the requirements of shape, moisture content, and other key indications. Therefore, it is necessary to classify the proportion of different types of fresh tea leaves correctly.

With the development of computing intelligence technology, computing vision technology has been widely applied in assortment and detection of food and agriculture production. Some researchers used visual technologies for automatic gross tea leaves categorization. The application in [1] used multispectral imaging machine vision and combined features (including shape, texture, and feature) to categorize tea leaves. However, it can only distinguish just single leaf, and the classification is inefficient. Chen et al. provided a method which employ near-infrared spectroscopy based on support vector machine for the rapid identification of green, black, and Oolong teas [2]. And it achieved a good result on the method of different kernel functions (RBF, polynomial) to analyze those finished tea leaves. Wu et al. did a research on multispectral imaging and SVM technology which showed that when SVM kernel function is the RBF, the sorting can be more accurate, efficient for tea leaves [3]. Although their studies obtained good results, they were only sorting the tea finished product and not applicable for fresh leaves. Chen et al. applied artificial neural network technology to the fresh tea leaves online classification [4]. However, that method requires too many characteristic parameters which affects the efficiency of classifying. In this paper, a texture analysis-based SVM approach has been adopted and it can effectively avoid the curse of dimensionality and requires fewer samples with good fault tolerance, flexibility, and adaptability in dealing with high-dimensional sample space.

The paper is organized as follows: In the next section, gray-level co-occurrence matrix (GLCM) for tea leave feature extraction is presented with the analysis of the experiment results. In the third section, several typical texture classification methods are compared and an SVM based on texture features for training and assorting is introduced. In that section, experimental results are also presented and discussed. Lastly, a summary and conclusion is presented.

37.2 Texture Features Extraction Based on Gray Level Co-occurrence Matrix

Feature extraction is an important step for image classification. Texture features can represent important information of an image and differentiate image regions. There are different methods for texture analysis including statistical, geometrical, model-based, etc. For the statistical approaches, GLCM has been proved to be a very

efficient texture descriptor used in image analysis. It is an approach to extract texture features regarding gray-level transition between two pixels via a co-occurrence matrix. It gives a spatial relationship which is defined among pixels in a texture. The matrix represents the joint distribution of gray-level pairs of neighboring pixels. Therefore, these matrices provide different information which is obtained by modifying the spatial relationship (different orientation or distance between pixels). GLCM describes the occurrence probability of i and j gray value in direction, d pixel distance between two pixels. It described a technique to discriminate between four different grades of made black tea using textural features based on gray tone spatial dependencies [5]. The statistical features were computed from tea images and wavelet decomposed images. The multilayer perception technique has been used for data classification.

Haralick has proposed 14 statistics calculated by GLCM, although the 14 statistics are able to express some specific the information of texture; it encounters some problems, such as information jumbled, presentation repeated [6], and effect from rate of classification. Usually, the three textural parameters (energy, correlation, and contrast) need to be calculated. The following Eqs. (37.2.1)–(37.2.3) define the three textural descriptors used in the experiment.

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \tag{37.2.1}$$

$$\text{Contrast} = \sum\nolimits_{i,j} (i-j)^2 p(i,j) \tag{37.2.2}$$

$$\text{Correlation} = \frac{\sum_{i,j} ijp(i,j) - u_x u_y}{\sigma_x^2 \sigma_y^2}, \tag{37.2.3}$$

where u_x, u_y, σ_x, σ_y are defined as follows:

$$u_x = \sum_{i,j} ip(i,j)$$

$$u_y = \sum_{i,j} jp(i,j)$$

$$\sigma_x^2 = \sum\nolimits_{i,j} (i-u_x)^2 p(i,j)$$

$$\sigma_y^2 = \sum\nolimits_{i,j} (i-u_y)^2 p(i,j)$$

In order to identify the type of different fresh tea leaves, three image databases of different fresh tea leaves are considered; the typical sample of fresh tea leaves are given in Fig. 37.1.

For the GLCM calculation for texture parameters, the interpixel angle is kept constant at 0°; the distance between pixels during the Co-occurrence matrix is set at [0 2; 0 3; 0 4; 0 5; 0 6; 0 7; 0 8; 0 9; 0 10; 0 11; 0 12; 0 13; 0 14; 0 15; 0 16;].

Fig. 37.1 Three samples of different types of tea leaves: **a** one bud with one leaf tea, **b** one bud with two leaves tea, and **c** one bud with multileaves tea

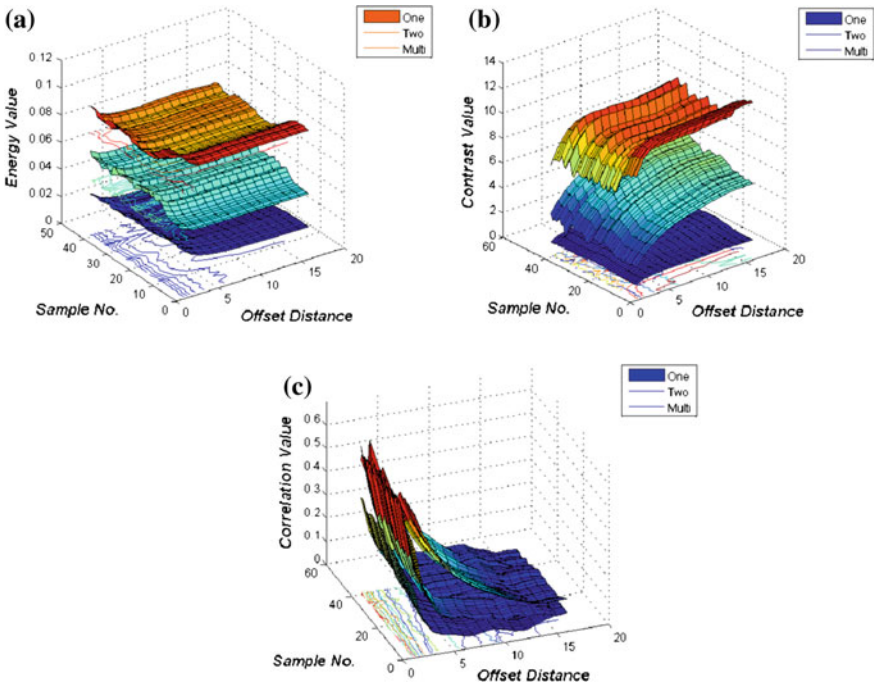


Fig. 37.2 Three texture value of different tea leaves: **a** energy value, **b** contrast value, and **c** correlation value

The three values of three types of tea leaves are given from Fig. 37.2. Experimental results show that the energy and contrast of texture can effectively differentiate tea leaves type.

37.3 Classification of Fresh Tea Leaves Based on SVM

Texture classifiers include supervised and unsupervised texture classifier. The design of supervised classifier is based on the structure of discrimination function and the decision-making rules, and then set the parameters of discrimination function according to training samples. To extract the texture features of unknown categories samples and classify the samples through the discrimination function and the decision-making rules, the methods include Bayes decision for minimum errors [7], extracting classifiers based on Fisher linear discriminant analysis [8, 9], and classification criteria in the nearest neighbor [10] are followed.

SVM is a method in which nonlinear problem in low-dimensional space is mapped to a high-dimensional linear space [11, 12]; it can ensure no increase of the computational complexity and also effectively overcome the over-fitting problems in neural network learning problems; thus, it provides high generalization ability. Kernel methods play an important role in SVM to deal with nonlinear problem, which provides a powerful way of detecting nonlinear relations using well-understood linear algorithms in an appropriate feature space. There have several well-know kernels, such as Eqs. (37.3.1)–(37.3.4):

1. The linear kenerl:
$$K(x_i, x_j) = (x_i \cdot x_j) \tag{37.3.1}$$
2. Polynomial:
$$K(x_i, x_j) = (x_i \cdot x_y + c)^d \tag{37.3.2}$$
3. Radial basis:
$$K(x_i, x_j) = \exp(-g\left\|x_i - x_j\right\|^2) \tag{37.3.3}$$
4. Sigmoid:
$$K(x_i, x_j) = \tanh\left[k(x_i \cdot x_y) + c\right], \tag{37.3.4}$$

where c and d are kernel parameters.

To achieve successful tea leaves classification, the energy value of texture features was first set as the input parameter of SVM. Figure 37.3a shows that the result of SVM training with different proportion of fresh tea leaves which consist of mainly one bud with one leaf and one bud with two leaves. It is shown that the boundary value of energy is about 0.055, when the value of energy is between 0.026 and 0.055, the tea leave type will be categorized into one bud with two leaves; otherwise, it will be categorized to the other. Figure 37.3b shows that the result of SVM training with different proportion of fresh tea leaves for the classification between one bud with two leaves and one bud with multiple leaves. It is shown that that it has an obvious boundary which can distinguish the type of fresh tea leaves.

Figure 37.4 presents the results of SVM classification by using the contrast value of texture features. It is shown that the clear boundary value between one bud with one leaf and one bud with two leaves is about 3 in Fig. 37.4a. Figure 37.4b shows that the SVM classification between one bud with two leaves and multileaves by

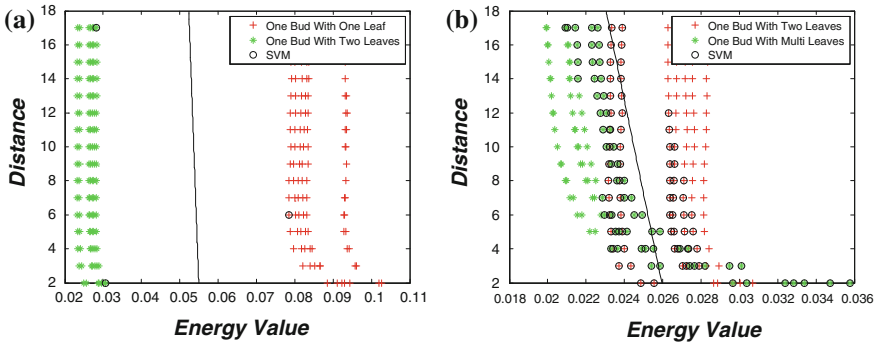


Fig. 37.3 The results of SVM training for data with energy value: **a** between one bud with one leaf and one bud with two leaves, **b** between one bud with two leaves and one bud with multileaves

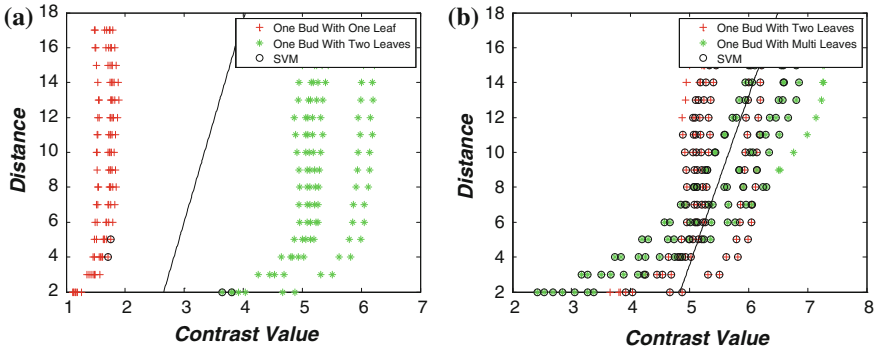


Fig. 37.4 The results of SVM classification for data with Contrast value: **a** between one bud with one leaf and one bud with two leaves, **b** between one bud with two leaves and one bud with multileaves

using contrast value, it is also shown that the boundary is not very clear by using a straight line.

To achieve successful classification rate, the SVM has been applied in different GLCM pixel distances from 3 to 6, from 11 to 14. In Fig. 37.5, the contrast value of texture features is set as the input parameter of SVM, and the distance between the pixels during the co-occurrence matrix is set differently in SVM classification. It is show that the successful classification rate is much better than that using only one straight line.

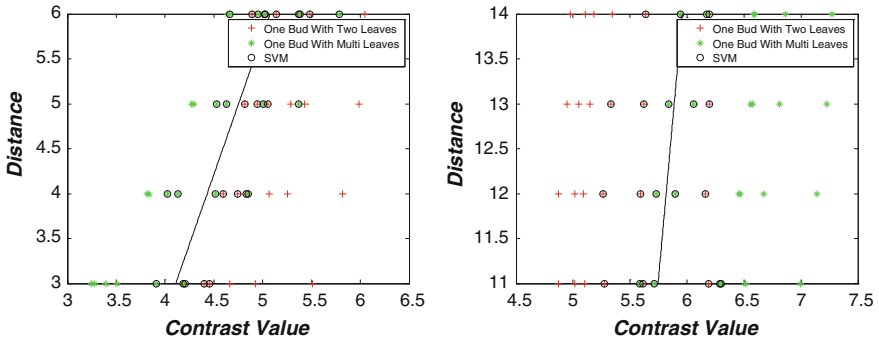


Fig. 37.5 The results of SVM classification with different pixel distance for data with contrast value

37.4 Experiments Evaluation

To verify the effectiveness of the proposed method for tea leaves classification, 60 tea images are sampled from video camera for training. There are three types for these tea leaf images, one bud with one leaf, one bud with two leaves, and one bud with multi-images. There are 20 images for each tea leave type. After training, the SVM is able to sort the tea leave images efficiently. There are 90 samples for testing the classification rate. These samples are divided into three groups, each group is consists of 30 samples, the number of different types of fresh tea leaves is 100. Experimental results indicate that using SVM for classification of fresh tea leaves is very effective. The result of successful classification rate is given in Table 37.1.

37.5 Summaries

The classification of fresh tea leaves based on texture analysis and SVM method has been proposed in this paper. The proposed method is the first time to realize automatic assorting of fresh tea leaves through computing vision technology, and an important guarantee for improving the quality of tea processing is provided.

Table 37.1 The accuracy of fresh tea leaves sorting

Sample	One bud with one leaf (%)	One bud with two leaves (%)	One bud with multileaves (%)
A	84	80	94
B	80	75	90
C	78	82	93
Average	80	78	92

The experiment results indicated that the energy and contrast value of GLCM for tea leaf image are suitable indications for fresh tea leaves type. When the energy value is less than 0.026, it means that the main fresh tea leaves in that picture are one bud with multiple leaves, between 0.026 and 0.05 is one bud with two leaves, otherwise is one bud with one leaf. While using the contrast value for SVM classification, the classification between one bud with two leaves and one bud with multileaves need to be considered in separate pixel distances. This tea type information can be used to select corresponding process parameters for tea production line to ensure high-quality tea manufacturing. Experimental results show that the SVM method for texture classification of fresh tea leaves can receive the successful classification rate at 83 %.

Acknowledgments This work is supported by the Natural Science Foundation of China (31470028, 61103035), Funds of Science and Technology Plan in Hunan province (2014GK3029), Funds of Science and Technology Plan in Changsha city (K1306035-11), the Project of Hunan Strategic Emerging Industries (2014GK1020). Authors are grateful for the anonymous reviewers who made constructive comments.

References

1. Li XL, He Y (2009) Classification of tea grades by multi-spectral images and combined feature. Trans Chin Soc Agric Mach S1:113–118
2. Chen Q, Zhao J, Fang CH et al (2007) Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). Spectrochim Acta Part A Mol Biomol Spectrosc 66(3):568–574
3. Wu D, Yang H, Chen X et al (2008) Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. J Food Eng 88 (4):474–483
4. Chen Y, Chang C, Xiao H et al (2010) Artificial neural networks technology in the fresh tea sorting. J Agric Netw Inform 7:013
5. Gill GS, Kumar A, Agarwal R (2013) Nondestructive grading of black tea based on physical parameters by texture analysis. Biosyst Eng 116(2):198–204
6. Du CJ, Sun DW (2004) Recent developments in the applications of image processing techniques for food quality evaluation. Trends Food Sci Technol 15(5):230–249
7. Chen Y, Wang JZ (2003) Support vector learning for fuzzy rule-based classification systems. Fuzzy Syst IEEE Trans 11(6):716–728
8. Li WB, Chen YY (2010) Using fisher linear discriminate analysis to extracting classifiers. Comput Eng Appl 46(14):132–134
9. Zhang X, Yan W, Zhao X et al (2007) Nonlinear biological batch process monitoring and fault identification based on kernel fisher discriminant analysis. Process Biochem 42(8):1200–1210
10. Hwang WJ, Wen KW (1998) Fast kNN classification algorithm based on partial distance search. Electron Lett 34(21):2062–2063
11. Wei CT, Wang N, Zhang LH et al (2013) Remote sensing image classification based on texture features. J Guilin Univ Technol 1:016
12. Zhong J, Ming H (2009) Bayes decision for minimum errors applied in recognition of handwritten english letters. J Liaoning Univ Technol (Nat Sci Ed) 29(2):98–101

Chapter 38
A Pedestrian Detection Method Based on MB_LBP Features and Intersection Kernel SVM

Xuejie Nian, Ke Xie, Wankou Yang and Changyin Sun

Abstract Pedestrian detection is a hot research topic in pattern recognition and computer vision. We combine MB_LBP (Multiscale Block Local Binary Patterns) feature and Histogram Intersection Kernel SVM and apply them to pedestrian detection. MB_LBP features, which make up for the lack of LBP (Local Binary Patterns) features in robustness, is a kind of effective texture description operator. Histogram Intersection Kernel Support Vector Machine has the advantage of fast classification and high accuracy in object recognition. It can be used for further enhancing the system's real-time performance. The experiments show that the proposed approach has higher precision than the classical algorithm HOG+ LinearSVM and the HOG_LBP Features Fusion tested on the established benchmarking datasets—INRIA.

Keywords Pedestrian detection · MB_LBP features · HIKSVM

38.1 Introduction

With the development of computer vision technology, pedestrian detection technology has become a hot issue in recent years. It widely is used in visual surveillance, behavior analysis, and automated personal assistance field. But detecting

X. Nian · K. Xie · W. Yang · C. Sun (✉)
School of Automation, Southeast University, Nanjing 210096, China
e-mail: cysun@seu.edu.cn

X. Nian · K. Xie · W. Yang · C. Sun
Key Lab of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China

X. Nian · K. Xie · W. Yang
Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_38

humans in images is a challenging task owing to human deformation, illumination change, occlusion, and wide range of poses.

The pedestrian detection algorithms can be basically divided into three categories currently: (1) Shape matching method based on the module [1], (2) Characterization method based on the apparent [2], (3) the method of combining the shape and texture [3]. Pedestrian Detection can be divided into partial detection [4] and overall detection [5] according to the way the body is detected. The shape matching method is a conventional human detection based on image processing technology, which could not solve the difficulties caused by a wide range of body shape and variable appearance. In this paper, a pedestrian detection method based on improved texture description is proposed. Felzenszwalb et al. [4] proposed part-based models which achieved good detection results, but the models are complex and difficult to compute. Based on this, the model based on the overall body is adopted in this paper.

When it comes to the selection of features descriptors, HOG (Histograms of Oriented Gradient) descriptors [5] stand out for its unique advantages for human detection. HOG descriptors provide excellent performance compared with other existing feature sets including wavelets [6, 7]. But HOG descriptors for pedestrian detection exhibit poor performance under severe occlusion and complex background. Then the fusion of HOG features and multiple features is recommended, as described in the papers [3, 8]. The features descriptors in these papers were connected in the series directly, the detection rate has increased a lot. However, the higher dimensionality of this features descriptor is not conducive to real-time systems. LBP [9] feature was first proposed as an effective texture description for face detection. However, the LBP feature has some limitations, while it is susceptible to noise, and lack of information on the overall coarse-grained grasp. Therefore, we propose to use MB_LBP (Multiscale Block LBP) [10] to extract pedestrians' features, which overcomes the limitations of LBP to improve the robustness of the descriptors and achieved good results.

Support Vector Machine (SVM) [11] and AdaBoost [12] classifiers are two mainstream classifications. LinearSVM obtains higher classification speed, but quite a long training time, and training complexity that increases with the number of support vectors exponential growth. The traditional LinearSVM is not suitable to the histogram-based features. In this paper, a new kernel—Histogram Intersection Kernel (HIKSVM) [13] is adopted to replace LinearSVM as classifier. The experimental results show that this method improves the detection rate and reduces the running time, which has a great advantage.

38.2 MB_LBP Features

Basic LBP algorithm uses typically a fixed-size block of 3 * 3 rectangular. Then the image corresponding of local texture T can be assumed that the joint distribution of the density distribution of gray pixels within the local area as Fig. 38.1 shows.

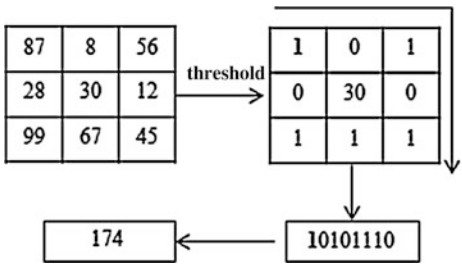


Fig. 38.1 The basic LBP operator

Selecting a point g_c, in the image as the center relative to the surrounding 8 points g_0, g_1, …, g_7 the distribution of the texture feature T can be represented as:

$$T \sim (g_0 - g_c, g_1 - g_c, \ldots, g_7 - g_c) \tag{38.1}$$

The binarization of the 8 pixels neighborhood takes the gray value of the central point as reference. The formula as:

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), \ldots, s(g_7 - g_c)) \tag{38.2}$$

where $s(x)$ is a function calculating the binarization of x and t is used for binary representation. According to the (38.3–38.4), the LBP values of the window.

$$\mathrm{LBP} = \sum_{t=0}^{7} S(g_i - g_c)\, 2^i \tag{38.3}$$

$$S(x) = \begin{cases} 1(x > 0) \\ 0(x \le 0) \end{cases} \tag{38.4}$$

The application of a center pixel point in the LBP algorithm is expanded to a rectangular region including a lot of pixel points in MB_LBP algorithm. The mean value of the rectangular region is regard as the threshold, and the comparison between pixels in the traditional LBP algorithm was replaced by the comparison of the average gradation between pixel blocks. In this paper, we use MBs_LBP represented the LBP operator of the s * s pixel blocks. Figure 38.2 shows the calculation process.

The MB_LBP values calculated as:

$$\mathrm{MB\_LBP} = \sum_{i=0}^{7} S(g_i - g_c)\, 2^i \tag{38.5}$$

As can be seen from Fig. 38.3, the texture of the response gradually thickens and turns to be stable as the size of the block is increasing. But too coarse texture representation would make the loss of image detail information. In this paper, the 3 * 3 pixel blocks perform better.

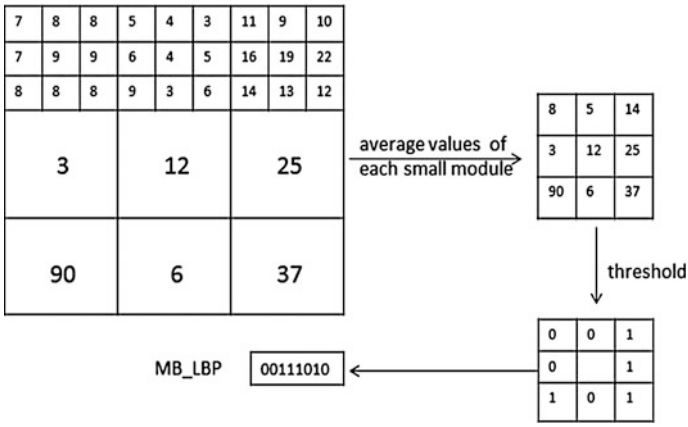


Fig. 38.2 The image shows the calculation process of the MB_LBP algorithm

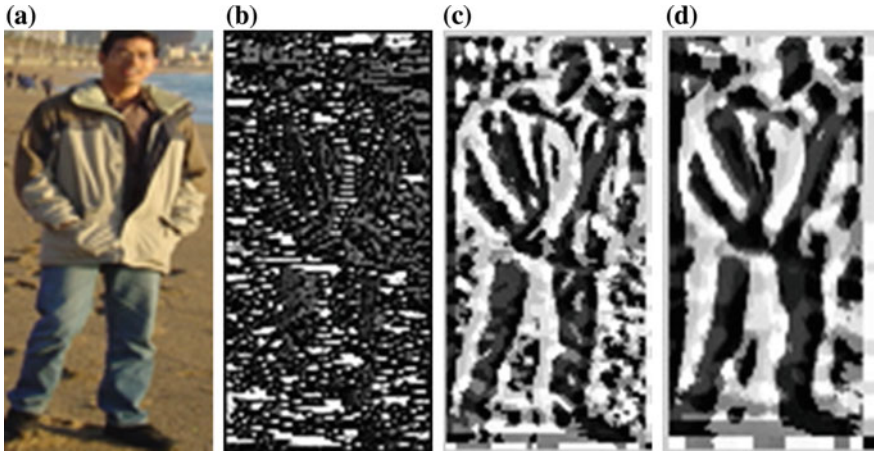


Fig. 38.3 Response after filtering. **a** Original image, **b** filtering by LBP, **c** filtering by MB3_LBP, **d** filtering

38.3 Classifiers

Vapnik proposed SVM which is a learning method based on structural risk minimization criteria. SVM can be divided into "linearSVM" and "non-linearSVM," depending on the kernel functions. The kernel function we used is as follows:

$$K_{\mathrm{GHI}} = \sum_{i=1}^{N} \min(|X(i)|^{\beta}, |Z(i)|^{\beta}), \beta > 0 \tag{38.6}$$

From the C-SVM, the decision function is shown as below:

$$h(x) = \sum_{l}^{m} \alpha_l \beta_l k(x, X_l) + b, W = \sum_{l=1}^{m} \alpha_l y_l X_l \tag{38.7}$$

where x is an n-dimensional feature vector, $X_l : l \in \{1, 2, \ldots, m\}$ are the support vectors. The intersection kernel function can be rewritten as the form:

$$k(x, z) = \sum_{i=1}^{n} \min(x(i), z(i)) \tag{38.8}$$

The decision function is transformed into the following form:

$$\begin{cases} h(x) = \sum_{l=1}^{m} \alpha_l \beta_l k(x, X_l) + b = \sum_{l=1}^{m} \alpha_l \beta_l (\sum_{i=1}^{n} \min(x(i), X_l(i))) + b \\ \quad = \sum_{i=1}^{n} (\sum_{l=1}^{m} \alpha_l \beta_l \min(x(i), X_l(i))) + b = \sum_{i=1}^{n} {}_i(x(i)) + b \\ h_i(s) = \sum_{l=1}^{m} \alpha_l \beta_l \min(s, X_l(i)) \end{cases} \tag{38.9}$$

On consideration of (38.9), it is observed that $h_i(s)$ is the summation of the m product, which produced after s compared with the value of the i-th dimension of the support vectors. It was shown that if $X_l(i)$ is sorted in the first place, and then compared with s, the calculation would be simplified. The $\bar{X}_l(i)$ indicates $\{x_l(i)_{(l=1)}^{m}$ in ascending order of size, and with the corresponding weight $\bar{\alpha}_l$ and label $\bar{y}_l$. Let r be the largest integer that satisfies $\bar{x}_r(i) \leq s$, then:

$$h_i(s) = \sum_{l=1}^{m} \bar{\alpha}_l \bar{y}_l \min(s, \bar{X}_l(i)) = \sum_{1 \leq l \leq r} \bar{\alpha}_l \bar{y}_l \bar{X}_l(i) + s \sum_{r \leq l \leq m} \bar{\alpha}_l \bar{y}_l \tag{38.10}$$

where,

$$A_i(r) = \sum_{1 \leq l \leq r} \bar{\alpha}_l \bar{y}_l \bar{X}_l(i), B_i(r) = \sum_{r \leq l \leq m} \bar{\alpha}_l \bar{y}_l \tag{38.11}$$

A_i and B_i are independent of the input sample, only concerned with the support vector and α. It was visible that $h_i(s)$ is a piecewise linear function. If $h_i(\bar{x}_r)$ is calculated in advance and stored in table form, then $h_i(s)$ can be obtained through the look-up table after searching s position in $\bar{X}_l(i)$. Although the space complexity of the calculation of $h(x)$ has increased, the time complexity is reduced to $O(\text{nlogm})$. Figure 38.4 shows the flowchart of the proposed methods.

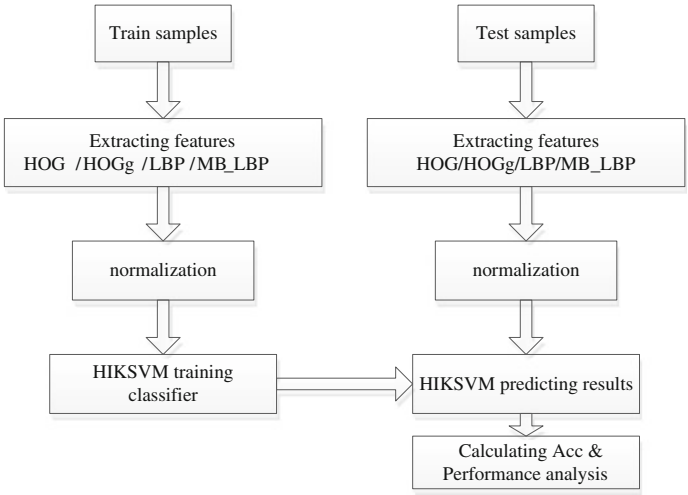


Fig. 38.4 The flowchart of this system

38.4 Experiments

38.4.1 Dataset

We tested our detector on INRIA datasets [5], providing the original picture and the corresponding standard file. In this paper, we obtained 2,416 pedestrian images cut from 614 positive training images, and resized these pedestrian images to 64 * 128 pixels used as positive training samples. Figure 38.3 shows some samples. Pedestrian occupies the entire image substantially, less redundant information in the image. A fixed set of 12,180 patches sampled randomly from 1,218 person-free training photos provide the negative training set, which have the same size of positive samples. 1,126 pedestrian samples are used as positive testing set, which are cut from 288 positive testing images. The negative testing set contains 9060 samples, which are the mirror image of a fixed set of 4,530 patches sampled randomly from 453 person-free testing images.

38.4.2 Experimental Results and Analysis

We have completed multiple sets of experiments based on INRIA dataset. Where HOG_LBP descriptors were the direct combination of series HOG and LBP, HOGg were described by paper [4], HOGg_LBP descriptors were also the direct tandem like HOG_LBP. Meanwhile sets of experiments using HIKSVM classifier were finished. In order to validate the effectiveness of our MB_LBP + HIKSVM system,

Table 38.1 Accuracy comparison

Feature	HOG	HOGg	HOG_LBP	HOGg_LBP	MB_LBP
Acc					
Classifier					
LinearSVM	88.2887	89.119	88.9432	89.0506	90.9846
HIKSVM	88.9432	89.4315	89.3339	89.4706	92.2739

we compare our system with other algorithms with Acc (Accuracy) (38.13) and FPPW-missrate curve, the result is shown in Table 38.1, Figs. 38.5 and 38.6.

$$\text{Acc} = (TP + TN)/(P + N) \tag{38.13}$$

TP (True Positive) is defined as the number of positive samples detected correctly, TN (True Negative) as the number of negative samples detected correctly.

Here, we describe the 3,780-dimensional HOG features from [5], the 3,255-dimensional HOGg features from [4], the 1,888-dimensional LBP of uniform $LBP^2_{8,1}$ from [14], while setting the detection window size to 64 * 128. In this paper, the MB_LBP features are calculated by letting s in MBs_LBP be 3 and generating histograms based on dividing the image window into 32 * 16 pixels block, total of 16 blocks. The dimension of the MB_LBP descriptor is 4,096. Experimental results show MB_LBP feature outperforms other features mentioned above, and the Acc rate reached 92.2739 % as the italics.

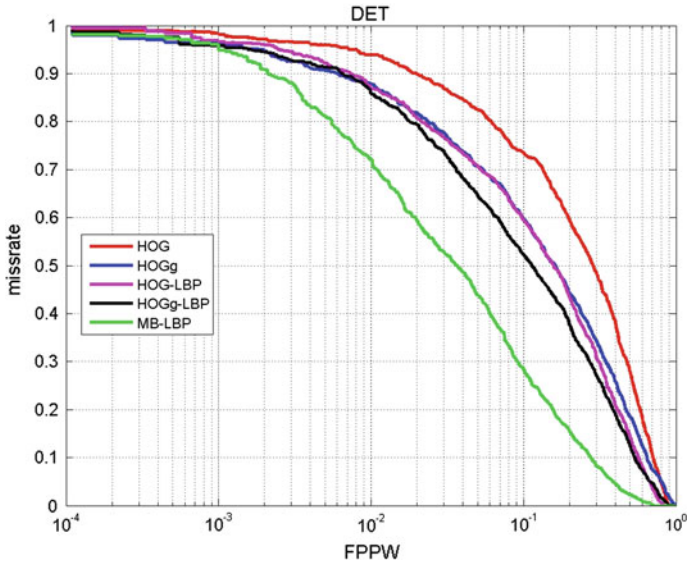


Fig. 38.5 MB_LBP versus other features

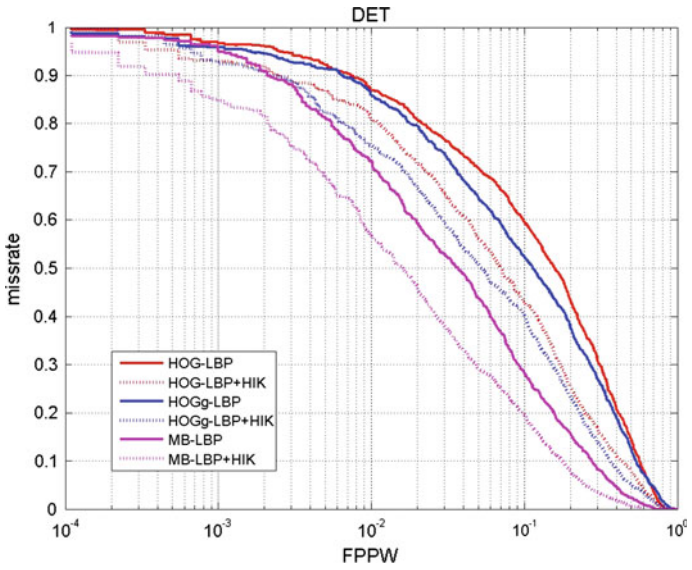


Fig. 38.6 LinearSVM classifier versus HIKSVM classifier

Figure 38.5 shows the performance comparison between different features. Only LinearSVM classifier is used in this set of experiments. Figure 38.6 compares LinearSVM with HIKSVM.

From Fig. 38.5, MB_LBP feature shows the best performance, superior to other features, when LinearSVM is used. In Fig. 38.6, conclusions could be drawn: the classification performance of HIKSVM is better than LinearSVM; our system, MB_LBP + HIKSVM, achieves the optimum performance.

38.5 Conclusions

The proposed pedestrian detection method based on MB_LBP features and HIK-SVM achieves good detection results on the established benchmarking datasets—INRIA.

MB_LBP feature in this paper compared with LBP feature, has the stronger robustness, and describes the image texture features effectively. It has lower dimensionality than the fusion features. Combining HIKSVM classifier, the feature can make the overall performance of pedestrian system has larger ascension. However, the singleness of the description of the images features in this paper makes the images information cannot be expressed completely. Therefore, the integrity of the description of the images information and the limitation of the dimension of the images features will be the major research direction in the future, thereby further to improve the performance of pedestrian detection system.

Acknowldgments This project is partly supported by NSF of China (61375001), partly supported by the open fund of Key Laboratory of Measurement and partly supported by Control of Complex Systems of Engineering, Ministry of Education (No. MCCSE2013B01), and partly supported by the open project program of Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University (No. CDLS-2014-04), and partly supported by China Postdoctoral Science Foundation (2013M540404),and partly supported by the Ph.D. Programs Foundation of Ministry of Education of China (No. 20120092110024).

References

1. Gavrila DM (2007) A bayesian, exemplar-based approach to hierarchical shape matching. IEEE Trans Pattern Anal Mach Intell 29(8)1–8
2. Prisacariu VA, Reid ID (2009) Fast HOG-A real-time GPU implementation of HOG. Technical Report, 2310/09, OUEL
3. Songzhi Sun (2011) Pedestrian detection of several key technology research. Xiamen university, Ph.D. thesis, Fujian Xiamen (in Chinese)
4. Felzenszwalb P, Girshick R, McAllester D et al (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
5. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 1, pp 886–893
6. Mohan A, Papageorgiou C, Poggio T (2001) Example-based object detection in images by components. IEEE Trans Pattern Anal Mach Intell 23(4):349–361
7. Viola P, Jones MJ, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. In: Proceedings of the 9th international conference on computer vision. Nice, France, vol, 2 pp 734–741
8. Bin H, Shengjin W, Xiaoqing D (2010) Multi features combination for pedestrian detection. University of Rochester (USA). J Multimedia 5(1):79–84
9. Ahonen T, Hadid A, Pietikainen M (2004) Face recognition with local binary patterns. In: Proceedings of the 8th european conference on computer vision. Berlin, Springer, vol 3021 pp 469–481
10. Lun Z, Ru-feng C, Shi-ming X et al (2007) Face detection based on multi-block LBP representation. Lecture Notes in Computer Science. Berlin, Springer, vol 4642, pp 11–18
11. Cortes C, Vapnik V (1995) Support vector network. Mach Learn 20(3):273–297
12. Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In computational Learning Theory: Eurocolt 95, Springer, 904: 23–27
13. Barla A, Odone F, Verri A. (2003). Histogram intersection kernel for image classification. In: Proceedings of the 2003 international conference on image processing, vol 3, pp 513–516
14. Ojala T, Pietikainen M (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

Chapter 39
Moving Vehicle Detection Using Oriented Histograms of Differential Flow and Extreme Learning Machine

Li Kang and Yuanlong Yu

Abstract Detecting vehicle in video is a challenging problem owing to the motion of vehicle, camera and background, and to variations of speed. This paper proposes a classifier-based supervised method to detect moving vehicle from a moving camera. Motion descriptors based on oriented histograms of differential optical flow be able to describe the motion pattern of the object from changing background. And Extreme Learning Machine provides excellent generalization performance at fast speed. Our sample images taken by a camera in helicopter include 2000 images. Experiment results shown that this proposed method has not only overall discrimination performance but also low computational cost.

Keywords Differential optical flow · Optical flow · Histogram of oriented gradient · Extreme learning machine

39.1 Introduction

Effective and accurate detection and tracking of moving vehicles in video streams is one of the key components for intelligent traffic monitoring systems. This component can be applied in video surveillance, intelligent surveillance for the drone, etc. Detecting moving cars from moving camera is a challenging problem owing to the background clutters and moving cameras or backgrounds, variations of speed, illumination, and foreground objects.

Optical flow, temporal difference, and background subtraction are the three main approaches developed for motion detection [1]. The approach using optical flow

L. Kang (✉) · Y. Yu (✉)
College of Mathematics and Computer Science, Fuzhou University, 2 Xue Yuan Road, University Town, Fuzhou, China
e-mail: canny0222@sina.com

Y. Yu
e-mail: yu.yuanglong@fzu.edu.zu

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_39

[2–5] works well only when the cameras and backgrounds are static. That is, using optical flow directly fails to eliminate the effects of camera motion. Temporal difference [6–8] is a common way for extracting moving objects. It depends heavily on the visual homogeneity and the speeds of the moving objects. Another widely used approach to motion detection is background subtraction [9–11]. But this method is effective only when a fixed reference background image exists. In order to meet the real-time requirements of mobile devices with camera, not only accuracy but also computational speed, the dependency of scene and various motion modes should be concerned for real-time applications.

This paper proposes a classifier-based supervised method to detect moving vehicles from a moving camera. The first contribution of this proposed method is to use histogram of oriented differential flow (HOF) [10]. HOF is a kind of motion-based descriptors which is combined with Histogram of Oriented Gradient (HOG) [12] descriptors. The descriptor is especially fit for the implement of describing complex and various motion modes. The second contribution is to set up a classifier based on extreme learning machine (ELM) [13]. ELM offers good generalization performance. The computational cost of ELM is also very cheap for both training and testing procedures such that the system can be put into real-time applications. Its advantage is that the detection performance is less dependent on parameter adjusting unlike other popular algorithms.

39.2 Architecting of the Moving Vehicle Detector

In the proposed method, the feature vectors of HOF are extracted from each serial frame and fed into Extreme Learning Machine to classify foreground and background.

The feature extraction of HOF combines motion descriptors with a process similar to HOG process shown in Fig. 39.1. The method firstly computing optical flows and differential flows from consecutive image frames. Differentials of optical flow are fed to a oriented voting process based on oriented spatial gradients of flow components. The descriptor is normalized in overlapping blocks consists of cells, and the resulting normalized histograms are concatenated to generate window descriptor vector.

Classification of ELM process contains training and classifying process. In ELM training process, a set of training images (here consecutive image pairs that flow can be got) containing vehicles have been manually marked as positive samples. Another set of negative training images is selected randomly from the background images. All of the training images are fixed-sized as a window. Figure 39.2 shows the ELM detecting process of moving vehicle detection.

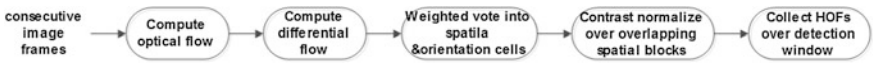


Fig. 39.1 The feature extraction process for moving vehicle detection

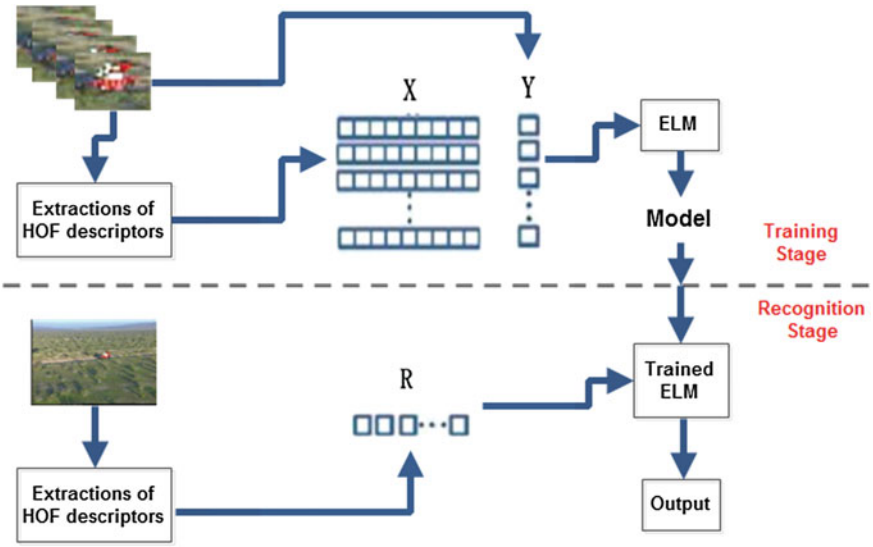


Fig. 39.2 The ELM detection process of moving vehicle detection

39.3 Oriented Histograms of Differential Optical Flow

39.3.1 Histograms of Oriented Gradients

Dalal and Triggs first proposed the algorithm of HOG for pedestrian detection in [12]. By dividing the image window into small spatial regions (cells), accumulating a local 1-D histogram of gradient directions over the pixels of the cell. Normalizing these histograms over overlapping blocks consist of cells, and the resulting normalized histograms are concatenated to generate window descriptor vector.

Gaussian smoothing followed by different discrete derivative mask can be used to get the gradient. The spatial derivatives [1, 0, −1] mask without any form of smoothing is the most common method. For color images, calculating gradients for each color channel, and gradient vector of the pixel is the largest norm. Let G_x, G_y denote the x and y components of gradient for one pixel. The gradient magnitude is $\sqrt{G_x^2, G_y^2}$ and the gradient orientation is $\tan^{-1}(G_x/G_y)$. Calculating orientation histogram on orientation of gradient established for each cell. The orientation bins are evenly divided into 0°−180° or 0°−360°. The votes are interpolated bilinearly between the neighboring bin centers in both orientation and position [12]. The weights of voting is a function of the gradient magnitude, and the votes are accumulated into orientation bins over local cell. Effective local contrast normalization is essential for good performance to reduce the impact on the local variations in illumination and foreground–background contrast. Let v is the unnormalized descriptor vector, ε be a small constant. There are many different normalization schemes, such as L2-norm, L1-norm, L1-sqrt, etc.

39.3.2 Histograms of Oriented Differential Flow

The HOF descriptors use differential flow to establish oriented histogram voting like HOG descriptors to cancel out most of the effects of camera motion. The approach is that the two flow components I^x, I^y take their local gradients (I_x^x, I_x^y) and (x_y^x, x_y^y) separately, calculate the corresponding gradient magnitudes and orientations, and weighted vote into local orientation histograms in the same way like HOG process. A separate histogram also can be built for each component, or combining them using the winner-takes-all voting method. The separate histograms show better performance in many experiments.

39.4 Extreme Learning Machine Classification

Guang-Bin Huang of Nanyang Technological University first proposed the algorithm of ELM in 2004. Different from other traditional feedforward networks, ELM be able to learn fast and produces good generalization performance. After the input weights and the hidden layer biases are chosen randomly, this algorithm is equal to a linear system.

39.4.1 The Brief of SLFNs

A single-hidden layer feedforward neural networks (SLFNs) shown in Fig. 39.3. The feature vectors are fed into the input layer. The hidden layer contains $\widetilde{N}$ hidden

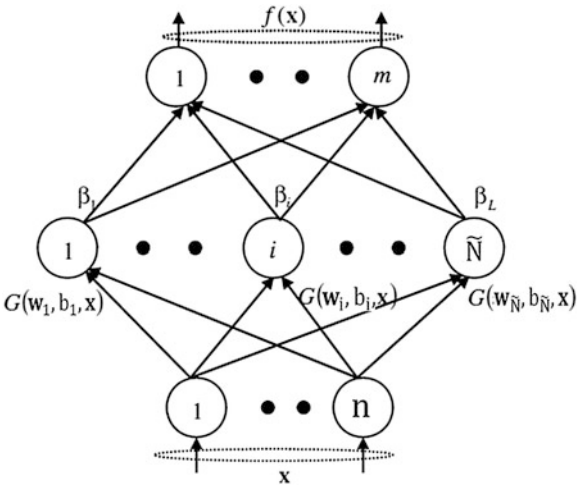


Fig. 39.3 A single-hidden layer feedforward neural networks

nodes. All of the input vectors are fed to each hidden node with the input weight w_i for the ith node. By the influence of biases b_i for the ith node and activation function $G(x)$, the output vectors received by accumulating all of the output vectors of the hidden layer with the output weights β_i of the ith node. The output of output layer is a multidimensional vector indicates which class the input vector belongs to.

39.4.2 Extreme Learning Machine

For N samples (x_j, t_j), where $x_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T \in R^n$ and $t_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in R^m$, the output of the jth sample through the neural networks mathematically modeled as

$$\sum_{i=1}^{\widetilde{N}} \beta_i g_i(\mathbf{x}_i) = \sum_{i=1}^{\widetilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_i + b_i) = \mathbf{o}_j,\ j = 1, \ldots N \tag{39.1}$$

where $g(x)$ is activation function, $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the input vector with the ith hidden node. $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden node with the output nodes, and b_i is the biases of the ith hidden node. $\mathbf{w}_i \cdot \mathbf{x}_j$ denotes the dot product of $\mathbf{w}_i$ and $\mathbf{x}_j$. And the learning process of ELM with $\widetilde{N}$ hidden nodes aim to minimize the error of outputs $\sum_{j=1}^{\widetilde{N}} \|\mathbf{o}_j - \mathbf{t}_j\|$.

For N samples, above equation can be written as a linear system.

$$\mathbf{H}\beta = T \tag{39.2}$$

where

$$\begin{aligned} &\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_{\widetilde{N}}, b_1, \ldots b_{\widetilde{N}}, \mathbf{x}_1, \ldots, \mathbf{x}_N) \\ &= \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}} \cdot \mathbf{x}_1 + b_{\widetilde{N}}) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}} \cdot \mathbf{x}_N + b_{\widetilde{N}}) \end{bmatrix}_{N \times \widetilde{N}} \end{aligned} \tag{39.3}$$

$\mathbf{H}$ is the hidden layer output matrix, the ith column of $\mathbf{H}$ is ith hidden node output with respect to inputs $x_1, x_2, \ldots, x_N$.

For fixed input weights $\mathbf{w}_i$ and the hidden layer biases b_i, training process aim to finding a least squares solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$. And the solution is shown in the following:

$$\beta = \mathbf{H}^{+}\mathbf{T} \tag{39.4}$$

where $\mathbf{H}^{+}$ is the Moore–Penrose generalize inverse of matrix $\mathbf{H}$. A solution of the Moore–Penrose generalized inverse of $\mathbf{H}$ is $\mathbf{H}^{+} = (\mathbf{H}^{T}\mathbf{H})^{-1}\mathbf{H}^{T}$ through the orthogonal project when $\mathbf{H}^{T}\mathbf{H}$ is nonsingular5.

39.5 Experiments

To train our ELM detector, sample images obtained from a general video taken by a moving camera in helicopter and the size of the images is 352 × 240. The sample set includes 900 training images and 2000 testing images. The set of training images includes positive windows and negative training windows.

We have conducted our experiments on a personal computer with an Intel Pentium 2.90 GHz CPU using Microsoft Visual Studio 2010. The computation time of training time is about 14 s, and the computation time of detecting time for per frame is about 0.6 s, depending on the image quality and the number of vehicles. Our default detector following below properties: [–1, 0, 1] radient filter with no smoothing; linear gradient voting into nine orientation bins in 0° – 180°; 10 × 10 pixel blocks of four 4 × 4 pixel cells; L1-Hys; 40 × 30 detection window; ELM classifier; 2000 hidden nodes; C parameter of ELM is 1.2.

To quantify the performance of the detector we calculate the Precision $(N_{\text{TruePos}}/(N_{\text{TruePos}} + N_{\text{FalsePos}}))$ and Recall $(N_{\text{TruePos}}/(N_{\text{TruePos}} + N_{\text{FalseNeg}}))$ to describe the accuracy and completeness for our experiments result. Given the default pattern, the Precision in this experiment is average 50 % and the Recall is average 50 % for about 90 m. Figure 39.4 shows some detection results of HOF and ELM.

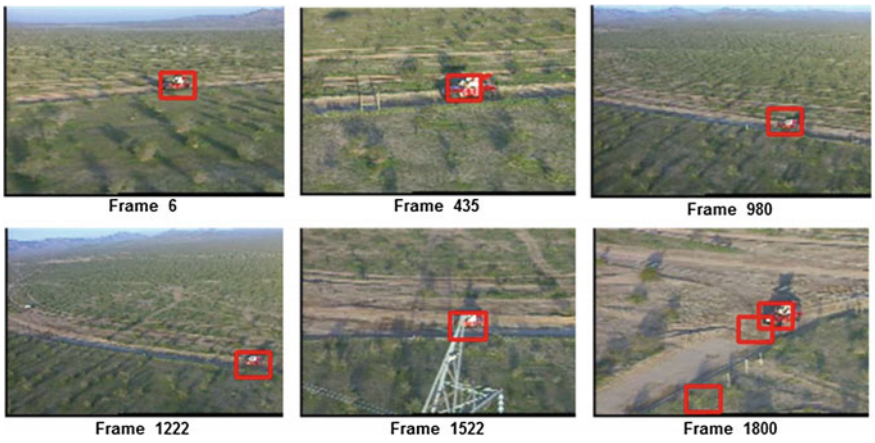


Fig. 39.4 Some detection results of HOF and ELM

Table 39.1 A comparison of detection results of the different motion descriptors

	(I_x^x, I_x^y)	(I_y^x, I_y^y)	(I_x^x, I_y^y)	$(I_x^x, I_x^y) + (I_y^x, I_y^y)$	IMHdiff (I_x^x, I_x^y)	HOG + (I_x^x, I_x^y)
Precision (%)	47.25	35.49	40.24	16.35	23.89	11.99
Recall (%)	39.78	50.91	54.00	13.41	53.78	49.78

Table 39.2 Recognition precision and recall of different threshold

$\hat{a}$	0.32	0.36	0.38	0.4	0.42	0.45	0.5
Precision (%)	61.13	48.90	45.7	40.24	33.64	27.51	18.90
Recall (%)	34.39	44.56	49.33	54.00	58.17	63.67	72.39

We list the result of different motion descriptors with the default pattern in Table 39.1. $(I_x^x, I_x^y) + (I_y^x, I_y^y)$ combining orientation histograms for the x- and y-derivatives. IMHdiff use larger spatial displacements ([–1, 0, 0, 0, 1] mask). HOG + (I_y^x, I_y^y) includes the HOG appearance descriptors and differential flow motion descriptors.

The threshold $\hat{a}$ of ELM is to describe similar extent for ELM binary classification. Table 39.2 illustrates recognition precision and recognition recall with different threshold in default pattern. The following experiments use (I_x^x, I_y^y) to build orientation histograms.

39.6 Conclusion

In this paper, we proposed a method to detect moving vehicles in videos with moving cameras. This method combines motion descriptors based on differential optical flow and Extreme Learning Machine classifier. We also studied different schemes of motion descriptors. A separate histogram for each flow component provides a more robust result instead of combining two histograms. We find a larger spatial displacements are in less precisions. Combining of static appearance descriptors and motion descriptors may not be much of an advantage. But this method provide an overall discrimination performance to extract moving vehicles at fast speed. And the new method could eliminate most of the effects of motion camera.

Acknowledgments This paper is supported by National Natural Science Foundation of China (NSFC) under Grants 61105102 and 61473089.

References

1. Ji X Effective vehicle detection technique for traffic surveillance systems
2. Sasa G, Loncaric S (2000) Spatio-temporal image segmentation using optical flow and clustering algorithm. In: First international workshop on image and signal processing and analysis, Pula, Croatia pp 63–68
3. You-shan Qu, Wei-jian Tian, Ying-cai Li (2003) The moving target detecting based on the parallel label discontinuous frame difference optical field integrated with gray intensity analysis. Acta Photon Sin 32(2):182–186
4. Smith SM, Brady JM (1995) ASSET-2; real-time motion segmentation and shape tracking. IEEE Trans PAMI 8(17):814–820
5. Thompson WB, Pong TC (1990) Detecting moving object. Int J Comput Vis 4:39–57
6. Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. Eur Conf Comput Vis 751–767
7. Needham CJ, Boyle RD (2001) Tracking multiple sports players through occlusion. Congestion Scale Br Mach Vis Conf 1 93–102
8. Li X, Liu ZQ, Leung KM (2002) Detection of vehicles from traffic scenes using fuzzy integrals. Pattern Recognit 35:967–980
9. Murray D, Basu A (1994) Motion tracking with tracking with an active camera. IEEE Trans Pattern Anal Mach Intell 16(5):449–459
10. Kim JB, Kim HJ (2003) Efficient region-based motion segmentation for a video monitoring system. Pattern Recognit Lett 24:113–128
11. Cai Q, Aggarwal JK (1999) Tracking human motion in structured environments using a distributed-camera system. IEEE Trans Pattern Anal Mach Intell 21(12):1241–1247
12. Dalai N, Triggs B (2005) Histograms of oriented gradients for human detection. In: computer vision and pattern recognition, CVPR 2005. In: IEEE computer society conference on vol 1, pp 886–893
13. Huang GB, Zhu QY, Siew CK et al (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1/3):489–501. doi:10.1016/j.neucom.2005.12.126

Chapter 40
K-Means Clustering Based on Density for Scene Image Classification

Ke Xie, Jin Wu, Wankou Yang and Changyin Sun

Abstract K-means clustering has been extremely popular in scene image classification. However, due to the random selection of initial cluster centers, the algorithm cannot always provide the most optimal results. In this paper, we develop a density-based k-means clustering. First, we calculate the density and distance for each feature vector. Then choose those features with high density and large distance as initial cluster centers. The remaining steps are the same with k-means. In order to evaluate our proposed algorithm, we have conducted several experiments on two-scene image datasets: Fifteen Scene Categories dataset and UIUC Sports Event dataset. The results show that our proposed method has good repeatability. Compared with the traditional k-means clustering, it can achieve higher classification accuracy when applied in multiclass scene image classification.

Keywords K-means · Density · Scene image classification

40.1 Introduction

Scene image classification is widely applied in many fields, like content-based image and video retrieval, remote sensing image analysis, biometrics recognition, and intelligent video surveillance [1].

K. Xie · J. Wu · W. Yang · C. Sun (✉)
School of Autmation, Southeast University, Nanjing 210096, China
e-mail: cysun@seu.edu.cn

K. Xie · J. Wu · W. Yang · C. Sun
Key Lab of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China

K. Xie · J. Wu · W. Yang
Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_40

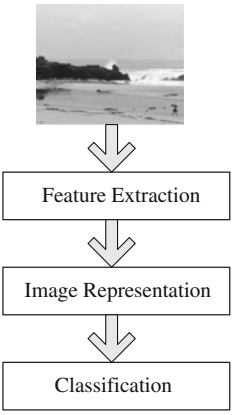


Fig. 40.1 The flow chart of scene image classification

The existing methods of scene image classification can be divided into three categories. (1) Methods based on the global low-level features of an image. Oliva and Torralba [2] presented a low-dimensional image representation, spatial envelope. (2) Methods based on image segmentation. Vogel [3] used image segmentation to get some meaningful target area. (3) Methods based on the visual vocabulary of image blocks. Fei-fei Li [4] proposed BOW (bag-of-words) model which has been very popular in scene image classification. This model treats image as a document consisting of many "visual words," then computes a statistical histogram for classification. Figure 40.1. gives the flow chart of BOW model for scene image classification.

The first step is to extract features from local patches. Generally, an image will generate a large amount of features. How to organize these features to represent the image becomes very important. BOW used clustering algorithm to classify the features into groups (namely clusters).

K-means algorithm [5] is widely used in clustering problem. It assumes the closer the distance between two objects, the greater the similarity. However, the k-means clustering cannot always find the optimal configurations due to the randomness of the initial cluster centers and the uncertainty of the number of cluster centers. The density-based clustering method is used to select initial cluster centers according to the density distribution of data points [6–8]. Redmond [9] got the initial cluster centers by building the Kd-tree. Recently, Rodriguez and Laio [10] proposed a new clustering algorithm by fast research and to find density peaks. The clusters were defined as local maxima in the density of data points. There is no need to iterate for many times and the clustering assignment can be finished in only one step. However, the density will be computationally costly when there are a lot of data points.

Here, we propose an improved clustering algorithm which combines the k-means clustering and density-based clustering. It can achieve higher classification accuracy rate when applied in the scene image classification.

The remainder of the paper is organized as follows. Section. 40.2 presents our proposed clustering algorithm in detail. Section. 40.3 gives some experimental results and compares our method with previous algorithms on scene image classification. Section. 40.4 concludes this paper and provides some ideas for the future work.

40.2 Density-Based K-Means Clustering

K-means is a very simple way to solve the clustering problems. However, the algorithm is sensitive to the selection of the initial cluster centers. Moreover, the number of cluster center (namely k) is uncertain. When applied in real-world problem, we must try many times to find the optimal value. In this paper, we propose an improved k-means clustering and we call it density-based k-means clustering.

Compared with the traditional k-means algorithm, the initial cluster center is determined by density and distance. As mentioned in [10], for each point x_i, the density ρ_i is defined as:

$$\begin{aligned} \rho_i &= \sum_j \chi(d_{ij} - d_c) \\ \chi(x) &= \begin{cases} 1, & x<0 \\ 0, & x \geq 0 \end{cases} \end{aligned} \tag{40.1}$$

d_c is a cutoff distance. It is not a fixed value. ρ_i is equal to the number of points that are closer than d_c to point x_i. δ_i is the minimum distance between x_i and any other point with higher density :

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}) \tag{40.2}$$

For the point with highest density, take $\delta_i = \max_j(d_{ij})$. Only those points with relative high ρ_i and high δ_i are considered as cluster centers. Therefore, we can define $\gamma_i = \rho_i \delta_i$ and choose those points with large γ_i as cluster centers.

Suppose we have n feature vectors $x_1, x_2, \ldots, x_n$ obtained from one training image, and k (k < n) is the number of clusters. The procedure of density-based k-means for the scene image classification can be described in the following steps:

(1) Compute ρ_i, δ_i and γ_i for each feature vector $x_i (i = 1, \ldots, \mathrm{n})$, and sort γ in decreasing order.
(2) Choose k feature vectors with large γ as initial cluster centers $c_1, c_2, \ldots, c_k$.
(3) Assign each remaining feature vector to the nearest cluster center.
(4) Update the cluster centers by calculating the average of the vectors belonging to the same cluster.
(5) Repeat step 3 and step 4 until k cluster centers no longer change.

40.3 Experiments

The scene image classification is implemented as follows. First, we use SIFT algorithm [11] to extract features for all the training and testing images. Then we use different clustering methods to organize the features and build the BOW model for each image. Finally, we use the LIBSVM package [12] to implement the SVM classification. Barla A et al. has found, compared with other standard kernel function, the histogram intersection can achieve very promising results on image classification [13]. Therefore, we construct a histogram intersection kernel as the kernel function of SVM. In our experiment, we choose one-against-all classification.

40.3.1 Dataset Used

In order to evaluate our proposed clustering algorithm, we have implemented three different experiments on two well-known scene image datasets. For each class in the dataset, we take 50 images for training and 20 images for testing.

40.3.1.1 Fifteen Scene Categories Dataset

The Fifteen Scene Categories dataset is built by Lazebnik [4]. The dataset contains 15 classes in JPG format, 8 of which are from MIT Scene dataset [2]. They are mostly grayscale images in 300 * 300 pixels. We randomly select eight classes of scene images as our experimental data. They are CALsuburb, coast, forest, highway, inside city, mountain, street, and tall building. Figure 40.2 gives some examples of eight scene classes.

40.3.1.2 UIUC Sports Event Dataset

The UIUC Sports Event dataset is collected by Li and Fei-Fei [14]. It contains eight sports event categories: Rowing (250 images), badminton (200 images), polo (182 images), bocce (137 images), snowboarding (190 images), croquet (236 images),

Fig. 40.2 Examples of eight scene classes

Fig. 40.3 Samples of eight sports categories

sailing (190 images), and rock climbing (194 images). Figure 40.3. gives some samples of the dataset.

The images in this dataset are color. Therefore before computing features, we convert each color image to gray image. And in order to reduce computational complexity, we scaled all images to within 300 * 300 pixels.

40.3.2 K-Means Clustering

Table 40.1 gives the experimental results of k-means clustering for scene image classification. Since the number of cluster center k is uncertain, we choose k from 50 to 450 to get the optimal parameter.

From Table 40.1, we can see that when the number of cluster centers increase, the iteration times increase at the same time. For the Fifteen Scene Categories dataset, although the accuracy has reached the highest at k = 450 the iteration times is very large. In order to achieve a balance between iteration times and accuracy, we take k = 300. The accuracy rate is 92.5 %. For the UIUC Sports Event dataset, the optimal k is also 300. The corresponding accuracy rate is 66.25 %.

Table 40.1 Results of k-means clustering

k (The number of centers)	Fifteen Scene Categories dataset		UIUC Sports Event dataset	
	Iteration times	Accuracy (%)	Iteration times	Accuracy (%)
50	18	89.375	8	57.5
100	13	90	13	58.75
150	16	89.375	10	60.625
200	19	91.25	13	62.5
250	18	91.25	15	65
300	15	92. 5	17	66.25
350	18	92.5	18	65.625
400	20	92.5	19	65
450	25	93.75	24	65.625

40.3.3 Density-Based Clustering

We carried out the density-based clustering proposed by Rodriguez and Laio. Tables 40.2 and 40.3 give the results according to different dc.

From Table 40.2, it is easy to find the optimal dc = 0.5 for the Fifteen Scene Categories dataset. The corresponding accuracy is 93.125 % which is higher than that in k-means clustering. For the UIUC Sports Event dataset, the optimal dc = 0.9. The corresponding accuracy is 67.5 % which is also higher than that in k-means clustering.

40.3.4 Density-Based K-Means Clustering

This time, we tried our proposed clustering algorithm for scene image classification. According to Table 40.1, we take k = 300 (the number of centers). The parameter dc is varied. Tables 40.4 and 40.5 give the final results.

Apparently, compared with k-means and density-based clustering, the density-based k-means we proposed in this paper achieves higher classification accuracy. For the Fifteen Scene Categories dataset, the most optimal dc = 0.6. The accuracy

Table 40.2 Results of density-based clustering on 15 scene categories dataset

dc	Accuracy (%)
0.1	86.875
0.2	88.75
0.3	90.625
0.4	90.625
0.5	93.125
0.6	90.625
0.7	89.375
0.8	91.875

Table 40.3 Results of density-based clustering on UIUC sports event dataset

dc	Accuracy (%)
0.5	58.125
0.6	61.25
0.7	62.5
0.8	63.75
0.9	67.5
1.0	63.75
1.1	58.75
1.2	58.125

Table 40.4 Results of density-based k-means clustering on 15 scene categories dataset

dc	Iteration times	Accuracy (%)
0.1	17	92.5
0.2	20	91.875
0.3	18	91.875
0.4	20	92.5
0.5	18	92.5
0.6	14	94.375
0.7	13	90.625
0.8	14	90.625

Table 40.5 Results of density-based k-means clustering on UIUC sports event dataset

dc	Iteration times	Accuracy (%)
0.4	15	64.375
0.5	21	69.375
0.6	22	68.75
0.7	19	68.125
0.8	21	69.375
0.9	29	68.75
1.0	22	73.75
1.1	21	69.375
1.2	18	68.125

has reached 94.375 %. As can be seen from Table 40.5, for the UIUC Sports Event Dataset, the most optimal dc = 1.0, the corresponding accuracy has reached 73.75 %.

40.3.5 Discussion

The above experiments showed that our proposed density-based k-means clustering can always achieve better performance. But for different database, the optimal parameters are not the same. In the Fifteen Scene Categories dataset, the optimal cutoff distance dc is 0.6. In the UIUC Sports Event dataset, the best dc = 1.0.

40.4 Conclusion

In this paper, we proposed a density-based k-means clustering for scene image classification. Compared with the traditional k-means algorithm, the initial cluster centers are not randomly selected. For each feature vector, we calculate two quantities: The density and the distance. Then we can find those points with high

density and high distance, which we call cluster centers. After we have determined the cluster centers, the remaining steps are the same with k-means clustering.

We have done several experiments to compare different clustering algorithm for scene image classification on two well-known scene image datasets. Experimental results show that our proposed method achieves better performance. When applied in multiclass scene image classification, it has good repeatability and high classification accuracy rate.

However, the selection of the cutoff distance dc is still a problem. How to find the most optimal parameter for the clustering problem of scene images is our next work.

Acknowledgments This project is partly supported by NSF of China (61375001), partly supported by the open project program of Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University (No.CDLS-2014-04), partly supported by China Postdoctoral Science Foundation (2013M540404) and partly supported by the Ph.D. Programs Foundation of Ministry of Education of China (No. 20120092110024).

References

1. Zhou L (2012) Research on key technologies of scene classification and object recognition. Graduate School of National University of Defense Technology (in Chinese)
2. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vision 42(3):145–175
3. Vogel J, Schiele B (2004) Natural scene retrieval based on a semantic modeling step. In: ACM international conference image video retrieval, New York, pp 207–215
4. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid classification for recognizing natural scene categories. In: IEEE computer society conference on computer vision and pattern recognition. New York, pp 2169–2178
5. Hartigan J, Wang M (1979) A k-means clustering algorithm. Appl Stat 28:100–108
6. Ester M, Kriegel HP, Sander J et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd vol 96, pp 226–231
7. Ankerst M, Breunig MM, Kriegel HP et al (1999) Optics: ordering points to identify the clustering structure ACM sigmod record. ACM 28(2):49–60
8. Hinneburg A, Keim DA (1998) An efficient approach to clustering in large multimedia databases with noise. In: KDD vol.98, pp 58–65
9. Redmond SJ, Heneghan C (2007) A method for initialising the k-means clustering algorithm using Kd-trees. Pattern Recogn Lett 28(8):965–973
10. Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science 344 (6191):1492–1496
11. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Inter J Comput Vis 2(60):91–110
12. Chang CC, LIN CJ (2001) LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/∼cjlin/libsvm
13. Barla A, Odone F, Verri A (2003) Histogram intersection kernel for image classification. In: International Conference on Image Processing vol.3(2):p III-513–16
14. Li LJ, Fei-Fei L (2007) What, where and who? Classifying events by scene and object recognition. In: IEEE 11th international conference on IEEE. computer vision ICCV, pp 1–8

Chapter 41
Study on Fusion Methods for Motion Evaluation Based on Somatosensory Rehabilitation System

Xin Guo, Yanwen Li and Longtao Su

Abstract Traditional rehabilitation training generally needs to be implemented in a specific lab environment through professional equipment and physician assists, which leads to a relatively high cost for the patients' rehabilitation. Because of the limitation of mechanical structure, it is always excessive or insufficient in training intensity, which will affect physicians to evaluate the training effect of patients scientifically and accurately. This paper builds a somatosensory rehabilitation system to assist patients with motor dysfunction with their training activities, and provides a relatively scientific assessment. By analyzing the data acquired via Kinect, the paper presents a fusion algorithm based on Hausdorff Distance algorithm and Dynamic Time Warping algorithm. The related experiments prove that the method is suitable for evaluating the matching degree of rehabilitation training action flows.

Keywords Kinect · Hausdorff distance · Dynamic time wrapping · Rehabilitation

41.1 Introduction

The target of rehabilitation training is to restore the impaired function of the body, and reach the full recovery of motor function [1]. The training methods medical institutions widely used are joint training activities, transfer and walking trainings, compensatory and common muscle trainings, etc., which always need to be implemented in specific lab environments with the assistant of physicians; thus makes it impracticable for users' training activities in community and promote family conditions. More importantly, the heavy and complicated data collection work will always influence physicians to make accurate judgments for patients'

X. Guo · Y. Li (✉) · L. Su
School of Control Science and Engineering, Hebei University of Technology, Tianjin 300130, China
e-mail: leeyw19837@hotmail.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_41

training effect. We use Kinect to complete the motor-related biological information capture and detection process in this paper, and presents an HD-DTW algorithm to evaluate patients' training performances. It provides a reference for patients and physicians to evaluate the effect of rehabilitation and reduces the burden of physicians to a certain extent.

41.2 Extraction of Kinect Dataflow

The evaluation methods in this paper are based on the mathematical algorithm processing of the selected joint points. The simplified model of human body skeleton abstracted by Kinect consists of 20 major joints [2], based on which the specific movement trajectory and certain joint angle value can be determined.

As shown in Fig. 41.1, shoulder center joint acts as the leading key point for upper limb joints, and its space coordinate determines the rotation, horizontal tilt, and other motions of upper body.

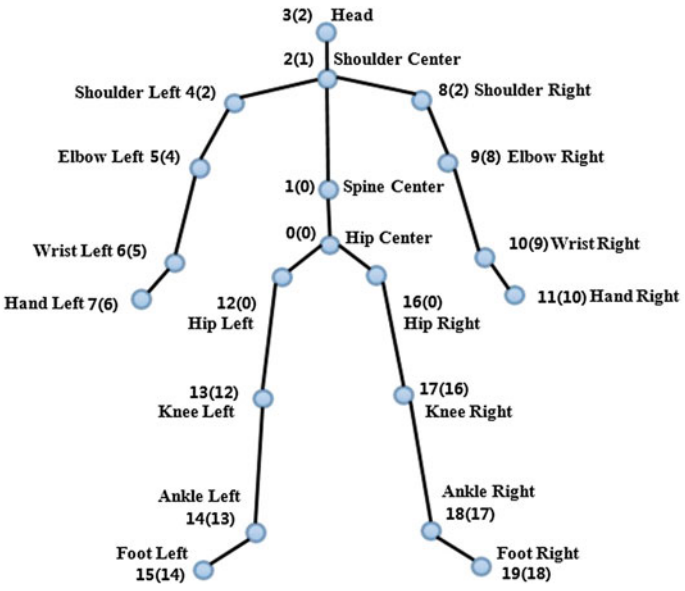


Fig. 41.1 Simplified model of human body skeleton. *Digit labels* without *bracket* are automatically assigned numbers of joints by Kinect according to the obtained skeleton dataflow; *digit labels* in *brackets* indicates the upper bearing node of the joint, i.e., the parent node

41.3 Motion Similarity Assessment Method

41.3.1 Summery

Using the fusion algorithm, we can calculate the distance of training point set and template dataset captured by Kinect, and take it as a movement similarity evaluation standard. This paper uses the Hausdorff Distance algorithm and Dynamic Time Warping algorithm to calculate the matching degree of training and template point sets, respectively [3], and evaluates the stability of these two algorithms to determine their respective weight while fusing them together as the final matching degree reference value. The specific mathematical expression is shown as formula (41.1).

$$G = \lfloor \alpha G_{HD} + (1 - \alpha) G_{DTW} \rfloor, \tag{41.1}$$

where G_{HD} and G_{DTW} represents the matching degree of training point set and template point set calculated by HD algorithm and DTW algorithm, respectively. We use a level 5 classification to represent the degree in this paper. α represents the degree of confidence of HD algorithm in the assessment and $1 - \alpha$ means that of DTW algorithm in contrast. G, the final result, is a round down value of the weighted result of the two algorithms. Due to the lower coefficient of variation of HD distance, which indicates that the algorithm has lower volatility, we tend to improve the confidence level of DTW algorithm.

41.3.2 Hausdorff Distance Algorithm

Hausdorff Distance (HD), is a mathematical definition for the distance of two finite closed set [4]. The smaller the distance is, the higher similarity the datasets have [5]. Limited to the precision of Kinect, noises in the process of data collection might be introduced. We adopt an improved form of HD algorithm—Least Trimmed Square HD (LTS-HD) in this paper [6]. It is a rewritten form of traditional HD algorithm which takes the mean value of part of the testing sequences as one-way HD distance and considers the average effect of all feature points with obvious properties. The improved HD algorithm reduces the influences of noise and isolated points on the measuring results, thus improves the accuracy of matching result.

41.3.3 Dynamic Time Warping Algorithm

Dynamic Time Warping (DTW) is a mature pattern recognition algorithm that is widely used in speech recognition field [7]. As there exists similar problems as speech recognition in motion matching, the data sequences captured by Kinect are

time-varying datasets with different length and speed, thus it is preferable to use DTW algorithm as a motion matching evaluating method [8]. As a kind of elastic matching algorithm, when comparing two groups of time correlation sequences' similarity using DTW, the lengths of the two are not necessarily identical, while the monotonicity with the time change of the two sequences is always required the same [9]. As we set specified limits to the training operation process in this paper, the training gesture is carried out according to the time sequence, and certain repeatability is presented. Therefore, it is well said that the DTW method is used to compute the similarity of template and training motions.

41.4 Experiments and Result Analyses

41.4.1 Data Normalization

In this paper, we select the abduction action of left shoulder joint to do the following instructions. The sketch of the gesture is shown as below (Fig. 41.2):

We use the somatosensory rehabilitation training platform as a medium of data acquisition in this paper. The template dataset and training dataset are stored in a text file format separately, in preparation for the subsequent MatLAB component calls. As shown in Fig. 41.3, it is the aspect of XOY projection of three-dimensional space coordinates point cloud of template motion trajectory and training motion trajectory.

Considering that each trainer's space position cannot be exactly the same as that of the template every time in data acquisition process, in order to reduce the

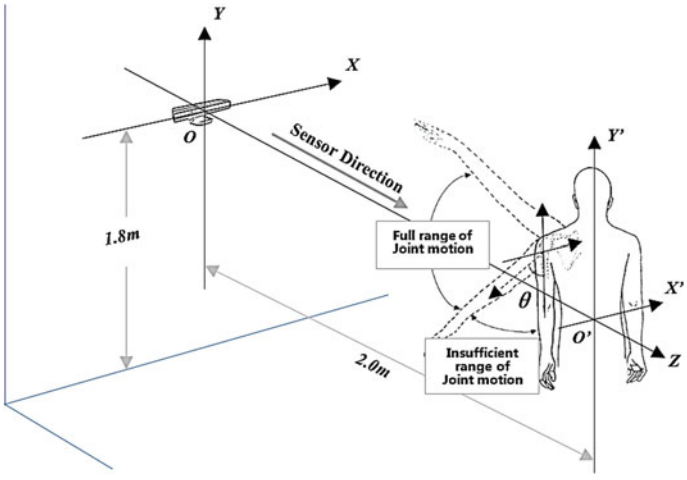


Fig. 41.2 The sketch of the abduction action of *left* shoulder in Kinect coordinate view. θ represents for the angle of the moving limb and specific coordinate axis, i.e., the joint angle

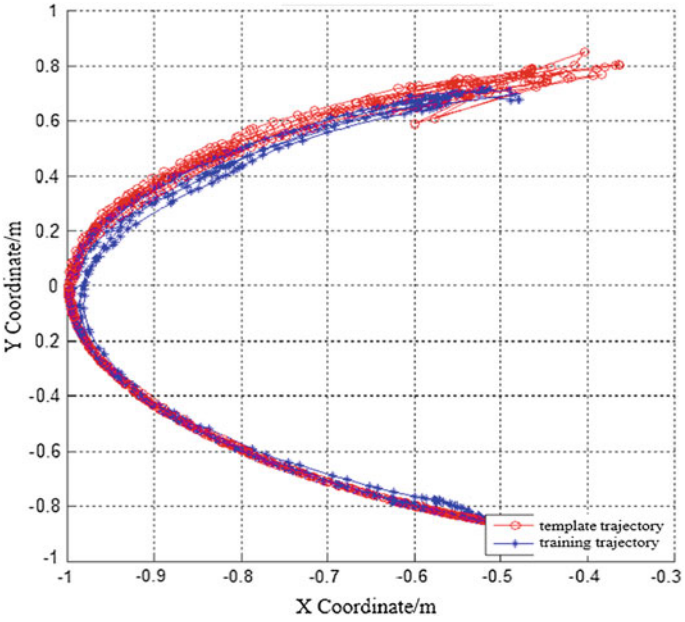


Fig. 41.3 The aspect of *XOY* projection of three-dimensional space coordinates point cloud of template motion trajectory and training motion trajectory

influences on subsequent matching results caused by trainers' own factors such as physical height, arm exhibition, and other individual differences, the captured point's coordinates have been dealt with normalization method in the application. Formula (41.2) shows the specific computational method:

$$P_N(X,Y,Z) = \frac{P_o(X,Y,Z) - P_{\mathrm{SC}}(X,Y,Z)}{|P_o(X,Y,Z) - P_{\mathrm{SC}}(X,Y,Z)|}, \tag{41.2}$$

where $P_o(X,Y,Z)$ stands for the real-time joint coordinates calculated by Kinect, $P_{SC}(X,Y,Z)$ means relative coordinate datum. Especially in this experiment, it is more preferable to select shoulder center as the joint node.

41.4.2 Algorithm Validation

Totally, 50 groups of experiments are conducted in this paper, and 10 training groups for each level from Level 1 to Level 5. Table 41.1 shows the basis of level classification of shoulder joint-abduction gesture.

DTW distance and HD distance is computed via MatLAB programmes, respectively [10]. Take DTW distance for instance. The results and analyses are shown as below.

Table 41.1 Basis of level classification of shoulder joint-abduction gesture

Gesture	Completion degree	Range of motion (%)	Specific values	Level	Instruction
Shoulder abduction	Insufficient	0–50	0°–70°	1	The horizontal flexion degree should be limited below 20°
		50–65	70°–85°	2	
	Sufficient	65–75	85°–125°	3	
		75–90	125°–145°	4	
		90–100	145°–160°	5	

It can be inferred from Figs. 41.4 and 41.5 that the matching result is acceptable while the DTW distance lies in the range of 20–70; for HD algorithm, the scope is 0.01–0.03. Further experiments are carried out to verify the feasibility of the fusing algorithm and Table 41.2 shows the average range of motions (ROM) values of six typical training gestures of left shoulder Joint, as well as the rating results.

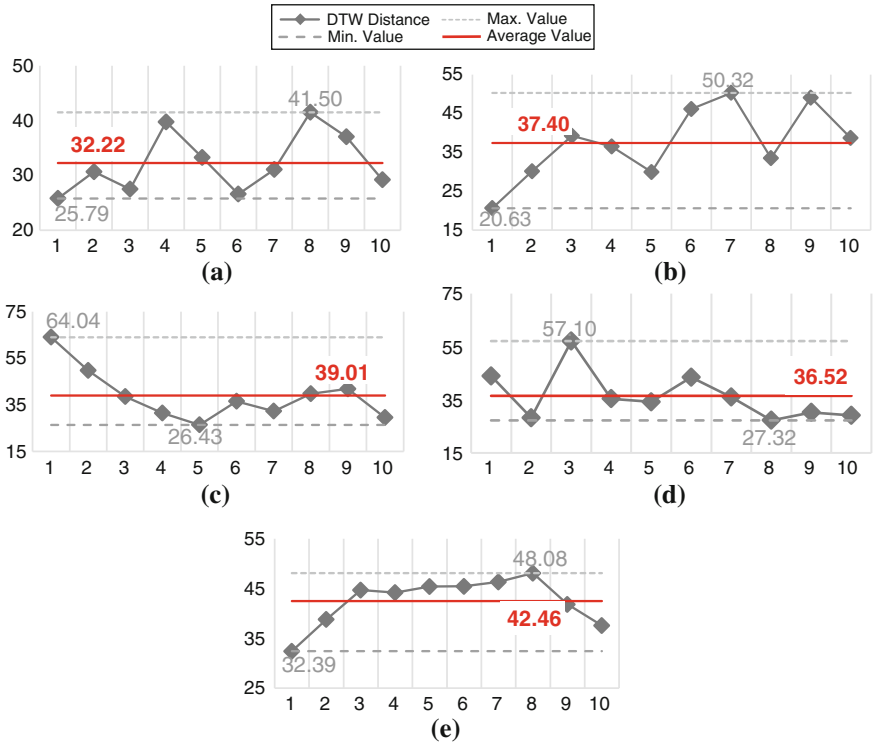


Fig. 41.4 Figure (**a–e**) illustrate the DTW distance of training dataset and template dataset of corresponding level from Level 1 to Level 5, with the maximum, minimum, and average values marked, respectively, and the abscissa represents for testing group number

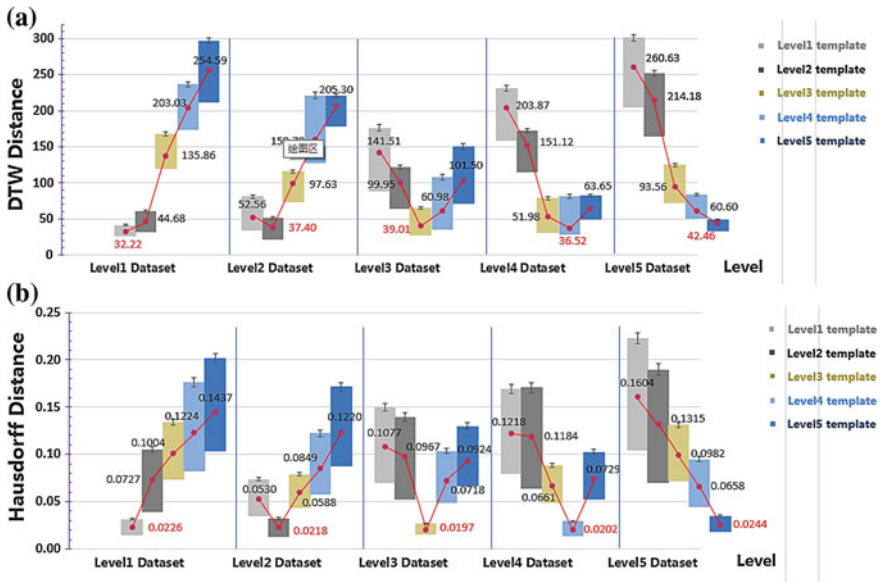


Fig. 41.5 Contrast of algorithm distances under different and corresponding levels. The *colored blocks* represent the range of algorithm distances while the *red points* indicate the average distances between the training dataset of specific level and template dataset of each level. **a** The result of DTW algorithm. **b** The result of HD algorithm

According to the experiment results above, we see that the HD-DTW fusion algorithm introduced in this paper is suitable to evaluate the matching degree of the training motion and its corresponding template motion.

Table 41.2 Average ROM values of six typical training gestures of Left Shoulder Joint and the corresponding rating results

Training gesture	ROM of template	ROM of training (Avg. value)	Ratings			
			Group 1	Group 2	Group 3	Avg. value
Flexion	0°–168.23°	0°–164.31°	5	5	5	5
Extension	0°–46.33°	0°–37.95°	5	5	4	5
Abduction	0°–161.49°	0°–158.72°	5	5	5	5
Adduction	0°–42.58°	0°–38.10°	4	5	5	5
Horizontal abduction	0°–95.65°	0°–82.65°	4	4	5	4
Horizontal adduction	0°–43.89°	0°–41.77°	5	5	5	5

The confidence level assigned to DTW algorithm is 0.6 and 0.4 is to HD algorithm

41.5 Conclusion

In this paper, we studied human motion evaluation methods based on the dataset acquired by Somatosensory Rehabilitation System, and proposed a motion similarity evaluation method based on the fusion of HD algorithm and Dynamic Time Warping algorithm. The paper verified the feasibility of these two kinds of algorithms by experiments, respectively, and analyzed the judging effect of the fusion algorithm. We will focus on increasing the operational efficiency of fusion algorithm and improving its applicability in future work.

References

1. Dai AB et al (2014) Somatic sense interactive technology and its application in motor rehabilitation. J China Rehabil Theor Pract 20(1):41–45 (in Chinese)
2. Kang J, Seo D, Jung D (2011) A study on the control method of 3-dimensional space application using kinect system. Int J Comput Sci Netw Secur 11(9):55–59
3. Shang FH, Sun DC, Lv HX (2010) Improvement algorithm of improved DTW efficiency. Comput Eng Des 31(15):3518–3520 (in Chinese)
4. Chen LZ, Wang BB, Dong JG (2009) A modified algorithm for model-based matching using hausdorff distance. Comput Technol Dev 19(5):82–85 (in Chinese)
5. Huttenlocher DP, Klanderman GA, Rucklidge WJ (1993) Comparing images using the hausdorff distance. IEEE Trans Pattern Anal Mach Intell 15(9):850–863
6. Lu S et al (2011) Image registration method based on key point feature and improved hausdorff distance. Syst Eng Electron 33(7):1664–1667 (in Chinese)
7. Barczewska, Drozd K (2013) Comparison of methods for hand gesture recognition based on dynamic time warping algorithm. Comput Sci Inf Syst 9(8):207–210
8. Liu XP, Xia XY, Lu JT (2013) Normalization of motion sequences based on DTW and hermite interpolation. J Syst Simul 25(7):1598–1603 (in Chinese)
9. Liu XM, Zhao D, Hao AM (2012) Human motion data retrieval based on dynamic time warping optimization algorithm. Pattern Recog Artif Intell 25(2):352–360 (in Chinese)
10. Wang L, Ye H, Xia LZ (2009) Activity recognition method for combining shape sequence and DTW. J Data Acquis Process 24(5):615–620 (in Chinese)

Chapter 42
The Active Above-Knee Prosthesis Gait Control Based on the EMPC

Yan Zhang, Yongchang Zhang, Lingling Chen and Bokai Xuan

Abstract On-line optimization control has been recognized as a superior algorithm for constrained systems. To deal with various disturbances and serious nonlinearity, active above-knee prosthesis (AKP) is treated as constrained switching system in this paper. The kinetic characters of AKP were analyzed when an amputee was walking on level-ground. A piecewise affine model can be built, and the AKP system was divided into 15 subsystems. Explicit model predictive control algorithm was applied to switch between subsystems. State-space partitions and explicit control laws are calculated off-line. On-line control signal can be obtained by looking up table. This algorithm can make the prosthesis to track the flexion curve of health limb accurately. Finally, results are shown that this method can tune the control impedance for powered knee prostheses with less energy consumption and high operation speed.

Keywords Active above-knee prosthesis · Piecewise affine model · Explicit model predictive control · Prosthesis control

42.1 Introduction

In recent years, with the improvements of amputees' needs, active above-knee prosthesis (AKP) is studied by more and more scientists. Active AKP can help amputee to recover lots of daily living activities and should be more reasonable on structure, lower energy consumption on motion, and more reliable on controlling.

Motion signals, pressure value and electromyographic signal collected from residual limb or health limb of amputee are used to judge the motion intention of amputee and generate control signals to drive powered knee of prostheses. Finite

Y. Zhang · Y. Zhang (✉) · L. Chen · B. Xuan
School of Control Science and Engineering, Hebei University of Technology, 300130 TJ, no. 8 Guangrong Road, Hong Qiao, Tianjin, China
e-mail: zhangyongchang.ok@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_42

state machine (FSM) control and adaptive control are common control methods for active AKP. FSM controller [1] was used for locomotion mode prediction of powered lower limb prostheses. There are only 4 states to be built by pressure value of pads. Constrained conditions and disturbances can not be considered in this method. Iteration control [2] and ANN adaptive control [3] are also common used. Tracking performance is verified well under common walk under a settled gait. But, when gait changes, more time is needed to modify parameter to adapt new gait. For active AKP, an algorithm with less energy consumption, high operation speed and ability of overcoming system disturbance is necessary.

Explicit model predictive control (EMPC) [4] used multi-parametric programming theory to solve constrained quadratic programming (QP) problem. And then, the control signal can be calculated by looking up table methods, which means its calculating speed can meet the requirements [5]. A linear time-invariant model or piecewise linear model is needed in this process.

Force analysis is absolute for active AKP control with powered knee. On the base of the kinetic model of human lower limb, the curve of the knee joint torque can be obtained in this paper, which can be regarded as the input of active AKP system. A piecewise affine (PWA) model is founded by division-combination strategy. Active AKP control problem is concluded to a constrained infinite time optimal control problem. EMPC control method is applied to solve this problem off-line. State-space partition and the corresponding explicit control law are used to establish the EMPC controller. At last, simulation results verified the effectiveness of the proposed control method.

42.2 Human Kinetic Equation

One of the most repetitive daily activities of human is walking on level ground. Sagittal axis is located in a plane that is parallel to the central plane of the sagittal suture. In general, walking along the sagittal axis is discussed. Two legs swing alternately, smoothly, and coordinately. The stance phase and swing phase are defined in normal gait cycle [6]. The gait of human is periodic repetitions of the stance phase in which a foot is on the ground and the swing phase in which the lower lamb moves after toe-off. In normal circumstances, the gait cycle time is between 1.0 and 1.6 s and the swing phase accounts for about 40 % of the whole gait cycle. The swing phase of the overall time is short relatively, but a wide range of knee angle change is happened in this phase. As the healthy limb, the function of prosthetic is to ensure the body's stability on the stance phase, and is able to adjust walking speed to make sure the smoothness during the change process from swing to stance phase on the swing phase.

Active AKP system can be described as a natural instability, strong coupling, and nonlinear problem in real-time. It is a trajectory tracking problem, the goal of tracking is swing angle curve of health limb of amputees. In this study, what to be concerned is the swing curve along the sagittal axis of uniaxial active AKP knee

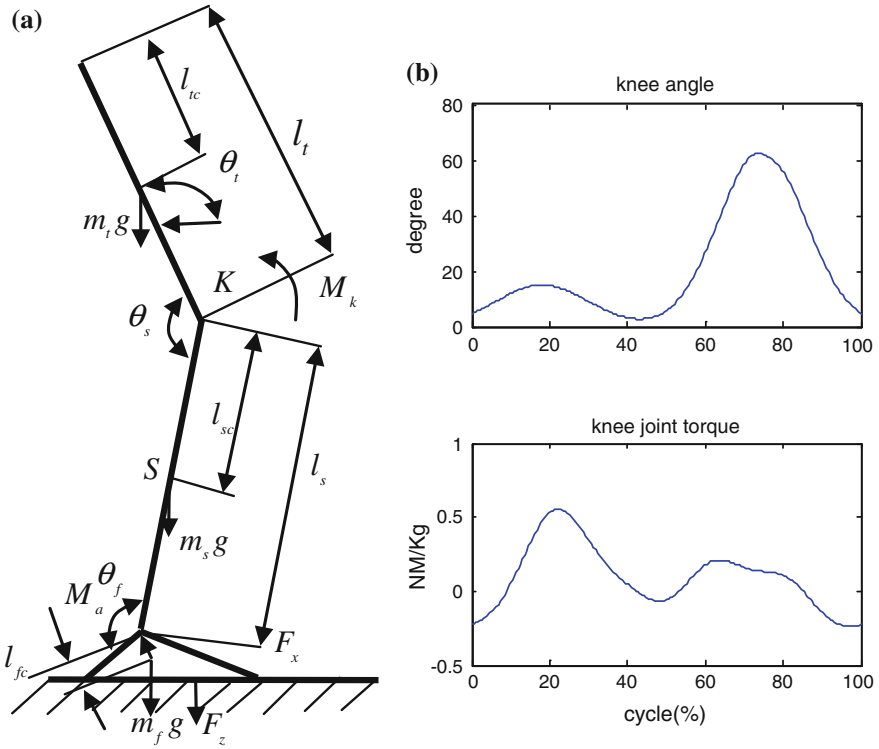


Fig. 42.1 **a** Model of human lower limb. **b** Knee angle measured by experiment and knee joint torque calculated by human lower limb model

joint. Swing on other direction is ignored. For simplicity, Lagrangian dynamics model [7] of uniaxial active AKP is shown as follow (Fig. 42.1a):

$$M_a = J_f\ddot{\theta}_f + m_f\ddot{x}_f l_{fc}\sin\theta_f - m_f(\ddot{z}_f + g)l_{fc}\cos\theta_f + F_z D_x - F_x D_z \tag{42.1}$$

$$\begin{aligned} M_k = {} & J_s\ddot{\theta}_s + m_s\ddot{x}_s l_{sc}\sin\theta_s - m_s(\ddot{z}_s + g)l_{sc}\cos\theta_s + M_a \\ & + (m_f\ddot{x}_f - F_x)l_s\sin\theta_s - (m_f g + m_f\ddot{y}_f - F_z)l_s\cos\theta_s \end{aligned} \tag{42.2}$$

where M_a, M_k are the torque of ankle and knee. J_f, J_s are the inertia moment of rotation around mass center of the shank and the foot respectively. D_x, D_z are the projection along X, Z direction of the distance from the ground reaction force acting point to the ankle joint point A. F_z, F_x are component of ground reaction force along X, Z direction. $\ddot{x}_s/\ddot{x}_f$, $\ddot{z}_s/\ddot{z}_f$ is acceleration of mass center of the shank/foot along X, Z direction.

The knee joint torque is the key problem in this paper. Key model parameters [8] have been listed in Table 42.1. Other human motion parameters, such as angle,

Table 42.1 Key inertial parameters of adult human body

Parameter	Variable	Value	Parameter	Variable	Value
Foot mass	m_f	0.93	Shank length	l_s	0.385
Shank mass	m_s	3.086	Shank inertia	J_f	0.0043
Foot mass center[a]	l_{sc}	0.224	Foot inertia	J_s	0.026
Shank mass center	l_{fc}	0.03898			

[a]Parameter mass center of foot/shank is the length form mass center of foot/shank to ankle/knee joint. All the data is obtained form 1) GB/T 17245-1998 Centroid of audlts. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Beijing. 2) GB/T 17245-2004 Inertial parameter of audlts. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Beijing

angular velocity, angular acceleration and ground reaction force can be tested by experiment. The knee angle and the knee joint torque are shown as Fig. 42.1b.

42.3 Modeling and Control Strategy

42.3.1 PWA Model Building

Obviously, what is shown in Eq. (42.2) is a serious nonlinear system. The knee joint torque is the control input signal of the active AKP, and the output of the system is the knee angle.

As we know, angular velocity is differential of angle:

$$\dot{\theta}_s(k) = (\theta_s(k+1) - \theta_s(k))/\Delta t \tag{42.3}$$

Then a PWA model [8] is built by following steps:

1. The knee angle curve is divided into four parts by extreme points. So it is done on the knee joint torque curve.
2. To keep resemblance to the knee angle curve, the knee angular velocity is also divided by extreme points.
3. All break points are applied in the knee angle, the knee angular velocity and the knee joint torque curve. The equal number of parts is ensured.
4. In each part, a mapping feather as following can be found:

$$\dot{\theta}_s(k) = a_1 t + k_1 \tag{42.4}$$

$$M_k(k) = a_2 t + k_2 \tag{42.5}$$

The relation between $\dot{\theta}_s(k)$ and $M_k(k)$ can be found:

$$\dot{\theta}_s(k) = bM_k(k) + f \tag{42.6}$$

At last, Eq. (42.7) model is obtained by substituting Eq. (42.3) into Eq. (42.6):

$$\theta_s(k+1) = A_i\theta_s(k) + B_iM_k(k) + F_i, \quad i = 1, 2, \ldots, N \tag{42.7}$$

A piecewise linear model with affine constant term is established. Here, linearization on knee angular velocity will change primitive knee angle. Many uncertainties exist in the process of human motion, whichwill lead to the collected data float within a certain range. Therefore, in order to obtain ideal curve, the curve fitting process should be done in prior of modeling. An acceptable error is shown between knee angle curve before and after linearzed in Fig. 42.2.

In this paper, knee ankle is divided into 4 parts, what means 4 PWA models are built. The number of subsystem of each PWA model (N in Eq. 42.7) is 3-4-4-4. In order to ensure the accuracy of model, more divided points than the number of extreme points are introduced. There is only 11 parts can be seen in Fig. 42.2 but there are 15 parts in fact. N represents the number of segment-model, the larger N represents by the more the number of the segment-model, and the more accurate of

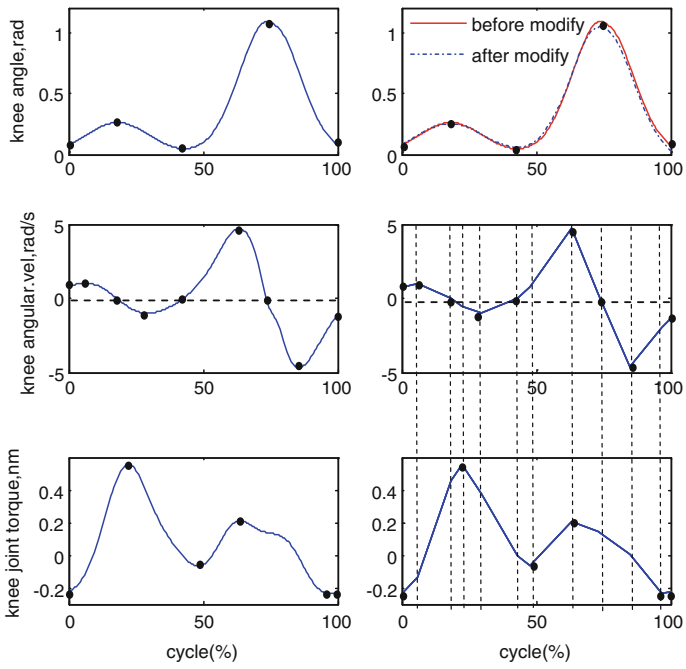


Fig. 42.2 Linearization and modeling process

established model. But too big N will be make the corresponding control strategy increased, where the work intensity of the controller is increased.

42.3.2 EMPC Control Theory

One of the outstanding advantages of EMPC is its high speed [9]. Here method in solving multi-parametric QP problem is represented to replace the traditional MPC strategy. Consider the LTI system

$$x(t+1) = \tilde{A}x(t) + \tilde{B}u(t) \tag{42.8}$$

$$y(t) = \tilde{C}x(t) \tag{42.9}$$

where $x(t) \in \Re^n$ is state variable, $u(t) \in \Re^m$ is the input variable, $A_i \in R^{n\times n}$, $B_i \in R^{n\times m}$. For the current state $x(t)$, MPC solves the optimization problem

$$\min_U J(U_N, x(0)) = \|Px_N\|_2 + \sum_{k=0}^{N-1} \left(\|Qx_k\|_2 + \|Ru_k\|_2\right) \tag{42.10}$$

$$s.t.\ Ex(t) + Lu(t) \le M \quad t \ge 0 \tag{42.11}$$

with respect to $U = \left[u_t', \ldots, u_{t+M-1}'\right]'$, where input weights $R = R' > 0$, state weights $Q = Q' \ge 0$. N is prediction horizon and M is control horizon. The final cost matrix $P(P \ge 0)$ is calculated from the algebraic Riccati equation. This problem can by some algebraic manipulation be reformulated as

$$V_z(x) = \min_z \frac{1}{2} z'Hz \tag{42.12}$$

$$s.t.\ \ Gz \le W + Sx(0) \tag{42.13}$$

where $z = U + H^{-1}F'x(0)$. Note that $H > 0$.

The above mp-QP problem can be solved by applying the Karush-Kuhn-Tucker (KKT) conditions under any initial state $x(0)$. The solution to this problem is shown as follows:

$$U = H^{-1}\tilde{G}'(\tilde{G}H^{-1}\tilde{G}')^{-1}(\tilde{W} + \tilde{S}x) - H^{-1}F'x(0),\ x \in CR_0 \tag{42.14}$$

with CR_0 is a polytope (a state-space partition) that all the system states x met the KKT conditions. When x is not in CR_0, the process above will be executed again. At last, a finite number of polytopes can be obtained. For simplicity, we describe this solution as

$$u_i = f_i x + gi, x \in CR_i, i = 1, \ldots, n \tag{42.15}$$

The first element of U is used as the control signal by rolling optimization theory.

In this paper, a trajectory tracking problem is discussed. Some augmentations should be done on the basis mentioned above. We define the knee angle trajectory as ref which is abbreviation of reference.

We redefine state variable x, control signal u in this paper as follows:

$$x(k) = [\theta_s(k), M_k(k), \mathrm{ref}(k)]' \tag{42.16}$$

$$u(k) = \Delta M_k(k) \tag{42.17}$$

Here we rediscribe the system as:

$$\begin{bmatrix} \theta_s(k+1) \\ M_k(k) \\ ref(k) \end{bmatrix} = \begin{bmatrix} A_i & B_i & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \theta_s(k) \\ M_k(k-1) \\ ref(k) \end{bmatrix} + \begin{bmatrix} B \\ I \\ 0 \end{bmatrix} \Delta M_k(k) + \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} F_i \tag{42.18}$$

The system output is as follow:

$$y(k) = \theta_s(k) \tag{42.19}$$

The system input can be obtained as

$$M_k(k) = M_k(k-1) + \Delta M_k(k) \tag{42.20}$$

42.4 Simulation

This simulation is carried out under MATLAB using Multi-Parametric Toolbox 3.0 [10]. On the based of the established of the prosthetic model, system constraints are defined as:

PWA model 1 $-0.3 \le M_k \le 0.6,\ 0.025 \le \theta_s \le 0.255;$
PWA model 2 $-0.3 \le M_k \le 0.6,\ 0.05 \le \theta_s \le 0.255;$
PWA model 3 $-0.3 \le M_k \le 0.6,\ 0.05 \le \theta_s \le 1.1;$
PWA model 4 $-0.3 \le M_k \le 0.6,\ 0.025 \le \theta_s \le 1.1.$

The switch rule from model to model is defined by out decision: $[\mathrm{ref}(k-1) - \mathrm{ref}(k-2)][\mathrm{ref}(k) - \mathrm{ref}(k-1)] < 0$. The EMPC control system of active AKP is shown as Fig. 42.3. All of the current states are collected by EMPC controller.

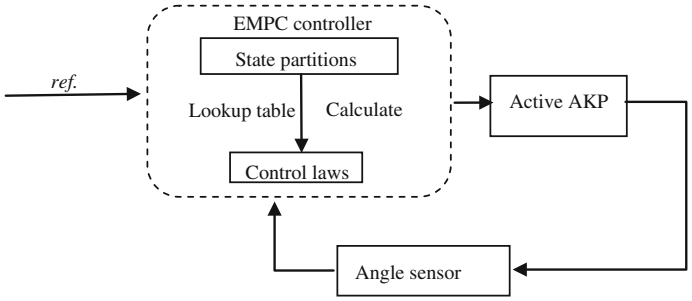


Fig. 42.3 EMPC control system

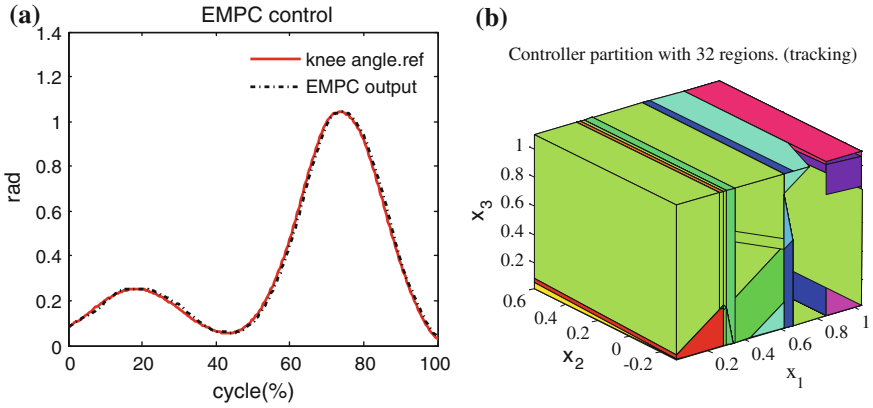


Fig. 42.4 **a** System response of EMPC controller. **b** State-space partitions of PWA model 4 (32 regions in shown)

The outputs $\theta_s(x_1)$, the current control signals $M_k(x_2)$ and the reference trajectory ref. (x_3) should be treated as the states of this system. Off-line control laws can be calculated by solving this multi-parameter constrained QP problem. Control frequency is the same as sample frequency, 148 Hz. Prediction horizon is set for 3-3-2-2. The number of state-space partition after solving off-line is 108-169-34-32. In other words, there are 343 explicit control laws in this system. The gait tracking result during swing phase is shown in Fig. 42.4a. And one of the state-space partitions is shown as Fig. 42.4b.

42.5 Conclusions

This smooth and steady control for AKP which can be treated as a time-varying nonlinear system with constraints and disturbances is the target of lots of scientists. To solve this problem, human gait planning and three-link-rigid body chain with

two freedom degree model are introduced for analyzing the model of human lower limb in this paper. EMPC solves the nonlinear problem of active AKP system based on the PWA model. As explicit control laws can be calculated off-line and the calculation burden is quiet small by looking up table on-line, there is a great advantage of timeliness in using EMPC in the chips installed in active AKP. The merits of this approach lie in the good control accuracy and robustness. The simulation results have shown these conclusions.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant to: 61203323) and Scientific Research Foundation of the Higher Education Institutions of Hebei Province, China (Grant to: Q2012079).

References

1. Young AJ, Simon AM, Hargrove LJ (2014) A training method for locomotion mode prediction using powered lower limb prostheses. IEEE Trans Neural Syst Rehabil Eng 22 (3):671–677
2. Liu ZJ, Zhao WZ, Geng YL, Yang P, Wang RC (2011) Intelligent prosthesis system based on flexible ILC. J Central South Univ (Sci Technol) 42(S1):524–529 (in Chinese)
3. Yu HL, Shen L, Zhao Z, Xu ZH, Qian XS (2010) CMAC control and case simulation for intelligent prosthetic leg. Control Eng China 17(2):182–189 (in Chinese)
4. Bemporad A, Morari M, Dua V, Pistikopoulos EN (2002) The explicit linear quadratic regulator for constrained systems. Automatica 38(1):3–20
5. Zhang J, Cai WB, Xie BF (2011) Realization of explicit model predictive control system based on ARM-Linux platform. J Southeast Univ (Nat Sci Edition). doi:10.3969/j.issn.1001-505.2011.s1.005 (in Chinese)
6. Geng YL, Yang P, Xu XY, Chen LL (2011) Design and simulation of active transfemoral prosthesis. J Hebei Univ Technol 40(5):1–4 (in Chinese)
7. Inoue K, Wada T, Harada R (2013) Novel knee joint mechanism of transfemoral prosthesis for stair ascent. In: IEEE international conference on rehabilitation robotics. doi:10.1109/ICORR.2013.6650436
8. Rodrigues L, How JP (2001) Automated control design for a piecewise affine approximation of a class of nonlinear systems. In: American control conference. doi:10.1109/ACC.2001.946412
9. TøNdel P, Johansen TA, Bemporad A (2003) An algorithm for multi-parametric quadratic programming and explicit MPC solutions. Automatica 39(3):489–497
10. Herceg M, Kvasnica M, Jones CN, Morari M (2013) Multi-parametric toolbox 3.0. In: Control conference (ECC), Zurich, Switzerland, July, 17–19, pp 502–510

Chapter 43
Application of Second-Order Cone Programming Theory to Robust Adaptive Beamforming

Rong-Yi Zhang and Hai-Yan Song

Abstract Due to the array steering vector errors and small-sample errors and so on, the performance of the Standard Capon Beamformer (SCB) may become worse than that of the Conventional Beamformer (CBF) in practical engineering applications. Nowadays, almost all existing algorithms to improve the robustness of SCB utilize the steepest descent method, conjugate gradient method, least mean squares (LMS) method and so on. However, most of them have the slow convergence and inefficient computation. Aiming at the above problems, this paper presents a unified process to solve them by the second-order cone programming (SOCP) theory, which can not only enhance the convergence speed but also improve the calculation accuracy so as to overcome the shortcomings of the previous optimization solutions effectively. It turns out that most adaptive beamforming algorithms involve a non-convex problem which is minimization of a quadratic function subject to infinitely many non-convex quadratic constraints. In this paper, it is shown that the proposed algorithm can be reformulated in a convex form as the so-called SCOP and solved efficiently (in polynomial time) using the well-established interior point method. Finally, computer simulations show the algorithm's excellent performance as compared with existing adaptive beamforming algorithms via the computer simulation examples.

Keywords Second-order cone programming · Robust adaptive beamforming · Sedumi

R.-Y. Zhang
Mechitrone Engineering College, Heilongjiang Institute of Technology, Harbin, China
e-mail: zhangrongyi11@126.com

H.-Y. Song (✉)
School of Electrical and Information Engineering, Heilongjiang Institute of Technology, Harbin, China
e-mail: songhaiyan0508@hrbeu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_43

43.1 Introduction

With the rapid development of electronic information science and technology, the adaptive beamforming becomes an important branch of the array signal processing research, which is widely used in radar, sonar, medical ultrasound imaging, wireless communication, and other areas. Without loss of generality, adaptive beamforming plays an important role in modern national defense and economic construction [1, 2].

The earliest adaptive beamformer is Multiple Sidelobe Canceller (MSC), which consists of a high gain main channel and several low gain auxiliary channels [3]. In 1969, Standard Capon Beamformer (SCB) was proposed, which minimizes the array output power subject to the constraint that the signal of interest (SOI) does not suffer from any distortion. In ideal conditions that the array steering vector is accurately known, SCB has a better resolution and much better interference rejection capability than the conventional beamformer (CBF). However, in practice, the array steering vector is often imprecise due to some factors such as snapshots errors, array response errors and so on. This may cause a substantial degradation of the performance of SCB [4].

In the past 30 years, robust approaches to SCB were urgently required [5]. Until now, there have been several existing approaches such as the so-called linearly constrained minimum variance (LCMV) beamformer [6], diagonal loading (DL) [7], subspace-based adaptive beamforming methods and so on [8]. However, these approaches have their own limitation or drawbacks. For examples, LCMV loses one degree of freedom for interference suppression. DL has difficulty in choosing the diagonal loading factors. Subspace-based adaptive beamforming is sensitive to the imprecise knowledge of the noise covariance matrix.

Recently, several methods which have a clear physical background and make explicit use of an uncertainty set of the array steering vector are proposed, such as Worst-Case Robust Beamforming (WCRB), Norm Constraint Robust Beamforming (NCRB), and so on [9, 10]. However, the approaches to solve these Robust Capon Beamforming (RCB) methods are different, for examples, the steepest descent method, conjugate gradient method, least mean squares (LMS) method and so on. And most of them have the slow convergence and inefficient computation.

In this paper, we show how to efficiently compute RCB methods by a unified theory, that is, the second-order cone programming (SOCP) theory. It turns out that most adaptive beamforming algorithms involve a non-convex problem which is minimization of a quadratic function subject to infinitely many non-convex quadratic constraints. In fact, we prove that despite the apparent differences between each RCB approach, they can be reformulated in a unified convex form as the so-called SCOP and solved efficiently (in polynomial time) using the well-established interior point method, such as Sedumi, which can not only enhance the computation convergence speed but also improve the calculation accuracy so as to overcome the shortcomings of the previous optimization solutions effectively.

Moreover, we also discuss the performance of the proposed methods in detail. Numerical examples illustrate the effectiveness of our RCB for SOI power estimation and output SINR.

43.2 SOCP Theory

43.2.1 Convex Optimization [11]

Convex optimization is a kind of special mathematic optimization problems, which include Least Square Method (LSM), linear programming, Second-Order Cone Programming (SOCP), semi-definite programming, and so on. With the developing of the optimization algorithms, it is found that convex optimization problems are prevalent in automatic control systems, signal detection and estimation, statistical data analysis, communication networks, electronic circuit design and other practical engineering applications. The stable and efficient optimal solution can be obtained through the interior point method which is an effective tool.

Definition 43.1 A set S is convex if the line segment between any two points in S lies in S, that is, if for any $x^{(1)}$, $x^{(2)} \in S$ and any λ with $0 \le \lambda \le 1$, we have $\lambda x^{(1)} + (1-\lambda)x^{(2)} \in S$.

Definition 43.2 A function $f : \boldsymbol{R}^n \to \boldsymbol{R}$ is convex if ***domf*** is a convex set and if for all $x^{(1)}$, $x^{(2)} \in \boldsymbol{domf}$, and λ with $0 \le \lambda \le 1$, we have

$$f(\lambda x^{(1)} + (1-\lambda)x^{(2)}) \le \lambda f(x^{(1)}) + (1-\lambda) f(x^{(2)}) \tag{43.1}$$

Consider the following problem:

$$\begin{cases} \min\limits_{x} & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & a_i^{\mathrm{T}} x = b_i, \quad i = 1, \ldots, p \end{cases} \tag{43.2}$$

where $f_0, \ldots, f_m$ are convex functions, and the feasible set of a convex optimization problem is convex:

$$S = \left\{ x \middle| \begin{matrix} f_i(x) \le 0, i = 1, \ldots, m; \\ a_i^{\mathrm{T}} x = b_i, i = 1, \ldots p \end{matrix} \right\} \tag{43.3}$$

So Eq. (43.2) is called a convex optimization problem.

43.2.2 SOCP Definition [12]

The second-order cone program (SOCP) can be represented as:

$$\begin{cases} \min\limits_{x} & f^{\mathrm{T}}x \\ \text{s.t.} & \|A_i x + b_i\|_2 \le c_i^{\mathrm{T}} x + d_i, \quad i = 1, \ldots, m \\ & Fx = g \end{cases} \tag{43.4}$$

where $x \in R^n$ is the optimization variable, $A_i \in R^{n_i \times n}$, $F \in R^{p \times n}$. We call a constraint of the form $\|Ax + b\|_2 \le c^{\mathrm{T}}x + d$, $A \in R^{k \times n}$, a second-order cone constraint.

43.3 SOCP Form and Solution to Robust Adaptive Beamforming

Norm Constraint Robust Capon Beamformer (NCRB) imposes an additional constraint on the Euclidean norm of the weight vector to improve the robustness of the Capon beamformer [10]. So it can be formulated as follows:

$$\begin{cases} \min\limits_{\boldsymbol{w}} & \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{w} \\ \text{s.t.} & \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta) = 1, \ \|\boldsymbol{w}\|^2 \le \zeta \end{cases} \tag{43.5}$$

This optimization problem can be solved by Lagrange multiplier method, so the optimization weight vector is:

$$\boldsymbol{w}_{\mathrm{NCRB}} = \frac{(\boldsymbol{R} + \lambda \boldsymbol{I})^{-1}\boldsymbol{a}(\theta)}{\boldsymbol{a}(\theta)^{\mathrm{H}}(\boldsymbol{R} + \lambda \boldsymbol{I})^{-1}\boldsymbol{a}(\theta)} \tag{43.6}$$

From the structure of $\boldsymbol{w}_{\mathrm{NCRB}}$, we can know that NCRB belongs to the diagonal loading method. However, the solution process by Lagrange multiplier method is really complex.

NCRB can also be solved by SOCP. First, compute the Cholesky decomposition of the sample covariance matrix $\boldsymbol{R}$:

$$\boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{w} = (\boldsymbol{U}\boldsymbol{w})^{\mathrm{H}}(\boldsymbol{U}\boldsymbol{w}) = \|\boldsymbol{U}\boldsymbol{w}\|^2 \tag{43.7}$$

let $\zeta = \sigma^2$, and then

$$\|\boldsymbol{w}\|^2 \le \zeta \Rightarrow \|\boldsymbol{w}\| \le \sigma \tag{43.8}$$

Equation (43.5) can be simplified as:

$$\begin{cases} \min\limits_{\boldsymbol{w}} \|\boldsymbol{U}\boldsymbol{w}\| \\ \text{s.t.} \quad \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta) = 1, \ \|\boldsymbol{w}\| \le \sigma \end{cases} \tag{43.9}$$

Equation (43.9) can be further solved by SOCP. Introduce a non-negative y_1, and let $\boldsymbol{y} = [\,y1 \quad \boldsymbol{w}^{\mathrm{T}}\,]^{\mathrm{T}}$, $\boldsymbol{b} = [\,-1 \quad \boldsymbol{0}_{1\times N}\,]^{\mathrm{T}}$, hence $-y_1 = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$. So Eq. (43.9) can be formulated as:

$$\begin{cases} \min\limits_{\boldsymbol{y}} (-y_1) \\ \text{s.t.} \quad \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta) = 1, \ \|\boldsymbol{U}\boldsymbol{w}\| \le y_1, \ \|\boldsymbol{w}\| \le \sigma \end{cases} \tag{43.10}$$

Through the zero cone and second-order cone transformation, we can obtain:

$$\begin{aligned} 1 - \boldsymbol{a}(\theta)^{\mathrm{H}}\boldsymbol{w} &= 1 - [\,0 \quad \boldsymbol{a}(\theta)^{\mathrm{H}}\,]\boldsymbol{y} \\ &= \boldsymbol{c}_1 - \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{y} \in \{0\} \end{aligned} \tag{43.11}$$

$$\begin{bmatrix} y_1 \\ \boldsymbol{U}\boldsymbol{w} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{0}_{N\times 1} \end{bmatrix} - \begin{bmatrix} -1 & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{N\times 1} & -\boldsymbol{U} \end{bmatrix}$$

$$= \boldsymbol{c}_2 - \boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{y} \in SOC_1^{1+N} \tag{43.12}$$

$$\begin{bmatrix} \sigma \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} \sigma \\ \boldsymbol{0}_{N\times 1} \end{bmatrix} - \begin{bmatrix} 0 & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{N\times 1} & -\boldsymbol{I} \end{bmatrix}$$

$$= \boldsymbol{c}_3 - \boldsymbol{A}_3^{\mathrm{T}}\boldsymbol{y} \in SOC_2^{1+N} \tag{43.13}$$

Let $\boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \boldsymbol{c}_3 \end{bmatrix}$, $\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}_1^{\mathrm{T}} \\ \boldsymbol{A}_2^{\mathrm{T}} \\ \boldsymbol{A}_3^{\mathrm{T}} \end{bmatrix}$, so:

$$\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 0 & \boldsymbol{a}^{\mathrm{H}}(\theta) \\ -1 & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{N\times 1} & -\boldsymbol{U} \\ 0 & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{N\times 1} & -\boldsymbol{I}_{N\times N} \end{bmatrix} \in C^{(2N+1)\times(N+1)} \tag{43.14}$$

$$\boldsymbol{b} = [\,-1 \quad \boldsymbol{0}_{1\times N}\,]^{\mathrm{T}} \in \mathbb{R}^{(N+1)\times 1} \tag{43.15}$$

$$\boldsymbol{c} = [\,1 \quad 0 \quad \boldsymbol{0}_{1\times N} \quad \sigma \quad \boldsymbol{0}_{1\times N}\,]^{\mathrm{T}} \in \mathbb{R}^{(2N+1)\times 1} \tag{43.16}$$

Hence the optimization weight vector $\boldsymbol{w}$ which is obtained by SOCP satisfies the following equation:

$$\begin{aligned} &\max_{\mathbf{y}} \quad \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \quad \text{subject} \quad \text{to} \\ &\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \in \{0\} \times SOC_1^{1+N} \times SOC_2^{1+N} \end{aligned} \tag{43.17}$$

Then, by applying the software tool Sedumi to Eq. (43.17), we can obtain the optimization weight vector $\boldsymbol{w}_{\mathrm{NCRB}}$.

43.4 SOCP Form and Solution to Robust Adaptive Beamforming

Simulation condition: The number of the array elements is 35 (there are five rows along the coordinate axis y and seven columns along the coordinate axis z), the source frequency is 200 kHz, the element spacing is half-wavelength, the incidence angle and the elevation angle are respectively $\theta = -1^\circ$ and $\varphi = 1^\circ$, SNR is 20 dB, the array steering vector error is −5 dB, which is defined as:

$$10\log_{10}\left(\frac{\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2}{\|\tilde{\boldsymbol{A}}(\theta,\varphi)\|_{\mathrm{F}}^2}\right) \tag{43.18}$$

Figures 43.1 and 43.2, respectively, show the spatial spectrums by CBF and NCRB. Note that under the condition of the noise and the array steering vector error, the NCRB has better performance than CBF, such as the narrower main lobe

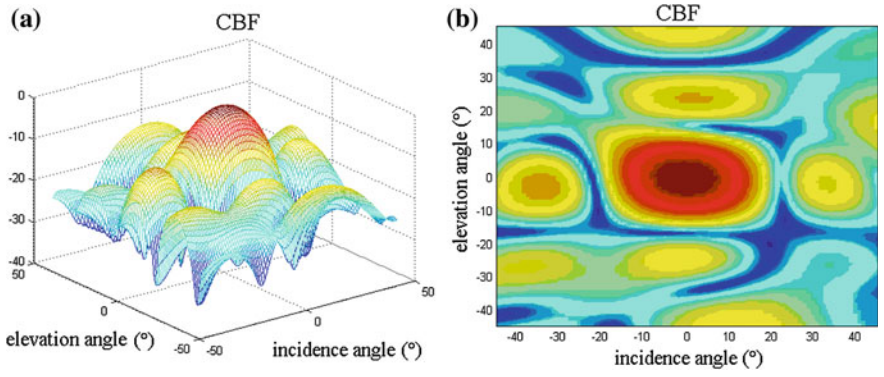


Fig. 43.1 The spatial spectrum for CBF. **a** Three-dimension graph, **b** Top-view graph

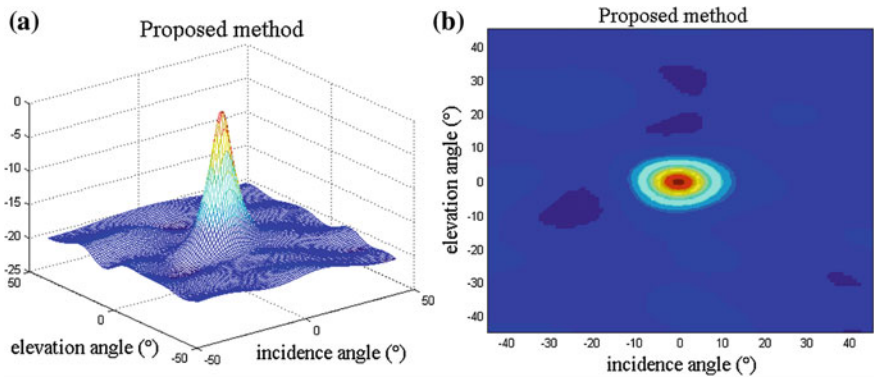


Fig. 43.2 The spatial spectrum for NCRB. **a** Three-dimension graph, **b** Top-view graph

and the lower side lobe. It is obvious from the simulation figures that NCRB can suppress the side lobe to −20 dB, so NCRB can highly improve the robustness against the noise and the array model error.

43.5 Conclusion

This paper has presented a unified process to improve the robustness of SCB by the second-order cone programming theory. The simulated results demonstrate that the proposed method provides excellent performance as compared with existing adaptive beamforming algorithms.

Acknowledgments This paper is supported by the Scientific Research Foundation of the Education Department of Heilongjiang Province, China (Grant No. 12541657).

References

1. HAMID KRIM and MATS VIBERG (1996) Two decades of array signal processing research. IEEE Signal Process Mag
2. Vincent F, Besson O (2004) Steering vector errors and diagonal loading. IEEE Proc-Radar Sonar Navig 151(6):337–343P
3. Applebaum SP, Chapman DJ (1976) Adaptive arrays with main beam constraints. IEEE Trans Antennas Propag 24(5):650–662P
4. Capon J (1969) High-resolution frequency-wave number spectrum analysis. Process IEEE 57 (8):1408–1418P
5. Jisung O, Kim S-J, K-L Hsiung (2005) A computationally efficient method for robust minimum variance beamforming. IEEE
6. Gershman AB, Nickel U, Bohme JF (1997) Adaptive beamforming algorithms with robustness against jammer motion. IEEE Trans Signal Process 45:1878–1885P

7. Carlson BD (1988) Covariance matrix estimation errors and diagonal loading in adaptive arrays. IEEE Trans Aerosp Electron Syst 24:397–401P
8. Yu JL, Yeh CC (1995) Generalized eigenspace-based beamformers. IEEE Trans Signal Process 43:2453–2461P
9. Vorobyov SA, Gershman AB, Luo ZQ (2003) Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. IEEE Trans Signal Process 51(2):313–324
10. Li Jian, Stoica Petre, Wan Zhisong (2004) Doubly constrained robust capon beamformer. IEEE Trans Signal Process 52(9):2407–2423P
11. Sergiy A, Vorobyov YR, Alex B. Gershman (2005) Robust adaptive beamforming using probability-constrained optimization. IEEE, pp 934–939
12. Sturm JF (1999) Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones[EB/OL]. Optim Meth Softw 11:625–653P

Chapter 44
Facial Geometric Feature for Cascade Eye Detection

Chang Guo, Xianye Ben, Xikai Fu, Fei Liu and Zhenqing Zhang

Abstract An innovative method for cascade eye detection using facial geometric features is proposed. The cascade detector is used to detect pairs of eye of candidates. The size similarity, the horizontal symmetry, the horizontal sextant angle, and the scale coefficients extracted from facial geometric feature are integrated to improve eye detection. Both the BioID database and real-time video stream were used in the experiments. Experimental results show that the proposed eye detection method is significantly more accurate than the cascade detector alone without sacrificing overmuch time-efficiency.

Keywords Cascade eye detection · Facial geometric feature · Face recognition

44.1 Introduction

Face recognition has become one of the most important topics in computer vision [1]. Eye detection is commonly considered to be a major step in the face recognition system [2] because the image coordinates of face images can be aligned based on the positions of eyes. Therefore, the accuracy of the eye detection greatly affects the face recognition performance. However, it is very difficult to detect the eyes accurately because of various conditions such as complex backgrounds, occlusion, illumination, and closed eyes.

To tackle this problem, many researchers have introduced the cascade eye detection methods where the eyes are detected through a coarse-to-fine process. Park et al. [3] proposed a method where an eye filter was used to extract eye candidates, and the nonnegative matrix factorization (NMF) algorithm is used to minimize image reconstruction error to detect the eye positions. Song et al. [4] also

C. Guo · X. Ben (✉) · X. Fu · F. Liu · Z. Zhang
School of Information Science and Engineering, Shandong University, No.27, Shanda Nanlu, Jinan 250100, China
e-mail: benxianyeye@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_44

presented a two-step method where the eyes are first modeled by the proposed visual-context pattern algorithm, and then the semisupervised boosting algorithm is used to detect eyes precisely. Choi et al. [5] included the generalized binary pattern feature in the cascade detector to improve detection accuracy, among the encoded binary patterns; they chose an effective set of them using the AdaBoost (Adaptive Boosting) feature selection method. Although their proposed generalized binary pattern provides higher eye detection accuracy according to experimental results, the method is time-consuming. Wang et al. [6] introduced a real-time automatic technique for eye detection. However, a few images with poorly detected eyes could often decrease the average performance of the automatic eye detection. Kawato et al. [7] introduced a feature-based tracking approaches for eye detection, they proposed to detect the point between the eyes which they think is more common and stable for most people. But experiment showed their method failed when eyes were occluded and the eye-brows or other hair parts were often taken as eye-like regions. Thus, their method needed a high contrast eye image. Choi et al. [8] proposed a new confidence measure to evaluate the eye detection results and combined two different eye detectors. They constructed a hybrid detector by combining two different and complementary detectors. Experimental results showed that the proposed method improved face recognition performance under various circumstances such as different illumination and the presence of glasses; however, the algorithm was computationally expensive. Zhu et al. [9] combined the Adaboost algorithm and a hybrid matching method to form a novel eye detection technique. First, facial part in the image was located using Adaboost algorithm and the area of eyes was positioned through the hybrid feature extraction method. Edge density, chrominance, HSV properties, and skin color cues were applied in extraction process. This algorithm effectively reduced the eye-detection candidate area and improved the detection accuracy; however, the method also fails when one or both eyes are closed. Therefore, the problem of eye detection is still far from being fully solved owing to factors including occlusion, lighting conditions and closed eyes, all affect the performance of eye detection algorithms. There are potentials existing in algorithm improvement.

In this paper, we propose a novel cascade eye detection method using facial geometric feature. In the first step, the cascade detector is used to locate pairs of eye candidates. In the second step, the size similarity coefficient, the horizontal symmetry coefficient, the horizontal sextant angle coefficient, and the scale coefficient extracted from facial geometric feature are integrated for precise eye detection.

This paper is organized as follows. The details of the AdaBoost algorithm are explained in Sect. 43.2. In Sect. 43.3, we introduce four facial geometric features for cascade eye detection. Some experimental results are given in Sect. 43.4 which shows the validity of our method by comparison with the classic method, and we finally conclude in Sect. 43.5.

44.2 AdaBoost

The AdaBoost is an iterative algorithm which has been widely used for many applications including face detection, image retrieval, texture classification, and object detection. AdaBoost [10] aims at training different weak classifiers according to one training set. AdaBoost classifiers can exclude unnecessary training data features, and concentrating on the essential training data features.

Unlike neural networks and SVM (Support Vector Machine), the AdaBoost training process selects only those features known to improve the predictive power of the model, reducing dimensionality and potentially improving execution time as irrelevant features do not need to be computed.

Viola et al. [10] proposed a robust cascade face detection method with Haar-like features, providing high performance in locating the face region. Therefore, this method [10] is used to detect the face region followed by the eye cascade detector to determine the eye candidates within the upper two-thirds of the detected face region. An input patch is classified as a positive sample only if it passes tests in all the nodes in the eye cascade detector, and most negative patches would be quickly rejected by the early nodes. Suppose N region $\{e1, e2, \ldots, eN\}$ are detected, $\frac{N(N-1)}{2}$ pairs of eye candidates are created by the different permutation and combination.

44.3 Facial Geometric Feature for Cascade Eye Detection

To improve the eye detection accuracy, four facial geometric features are proposed in this study. These features for each eye candidate include the size similarity coefficient, s, the horizontal symmetry coefficient, h, the horizontal sextant angle coefficient, k, and the scale coefficient, c:

$$s = \sqrt{(h_1 - h_2)^2 + (w_1 - w_2)^2} \tag{44.1}$$

where h_1 and h_2 are the height of two eyes, w_1 and w_2 are the width of two eyes. h represents how close the distance between d_1 and d_2 is among candidates

$$h = |d_1 - d_2| \tag{44.2}$$

where d_1 and d_2 are horizontal distances between the perpendicular bisector of the face region and two eyes, respectively.

k represents the angle between the connection of the centers of two eyes and the horizontal line.

$$k = \tan^{-1} \frac{|y_1 - y_2|}{|x_1 - x_2|} \tag{44.3}$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of the centers of two eyes.

c represents a coefficient according to the geometric distribution of facial features.

$$c = \sqrt{\left(\frac{w_1}{w_0} - \alpha\right)^2 + \left(\frac{h_1}{h_0} - \beta\right)^2} + \sqrt{\left(\frac{w_2}{w_0} - \alpha\right)^2 + \left(\frac{h_2}{h_0} - \beta\right)^2} \tag{44.4}$$

where w_0, h_0 are the width and the height of the face region, respectively. α is the ratio of the width of an eye to the width of a face, and β is the ratio of the height of an eye to the height of a face. Based on the prior knowledge, $\alpha = \frac{1}{5}$ and $\beta = \frac{1}{10}$.

For each candidate, the decision metric d is defined as

$$d = \mu_1 \times s + \mu_2 \times h + \mu_3 \times a + \mu_4 \times c \tag{44.5}$$

where $\mu_1, \mu_2, \mu_3, \mu_4$ are weighted factors for these four facial geometric features. In this paper, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0.25$. The smaller d is, the higher the confidence level of eye detection is. Thus, the best location of the eyes can be determined by the decision metric. Figure 44.1 depicts the detection process.

44.4 Experimental Results

To evaluate the performance of the proposed method, we compared the proposed method with the classic cascade detector in the BioId database. This database consists of 1521 gray-scale images out of 23 different individuals with a resolution of 384 × 286 pixels. The classic cascade detector achieves 86.79 % accuracy, but also as in the case of the proposed method, there is a large improvement in terms of precision (12.62 %). The right column of Figures 44.2, 44.3, 44.4, 44.5, and 44.6 shows several typical sample images with false eye detections by using the classic cascade detector; the left column of Figs. 44.2, 44.3, 44.4, 44.5, and 44.6 shows successful face and eye detection using the proposed method. In the case of the

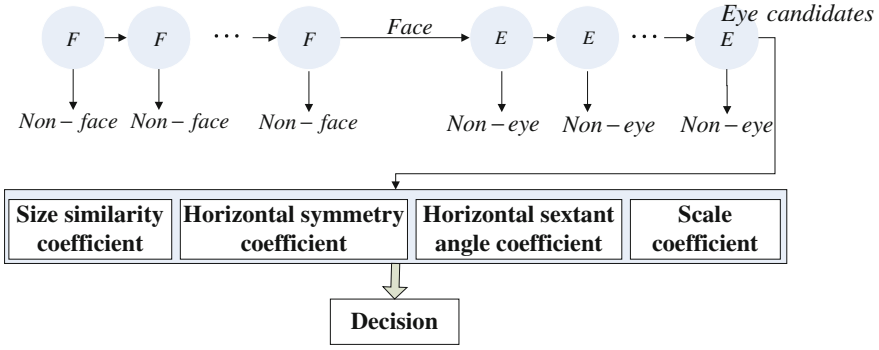


Fig. 44.1 Detection process

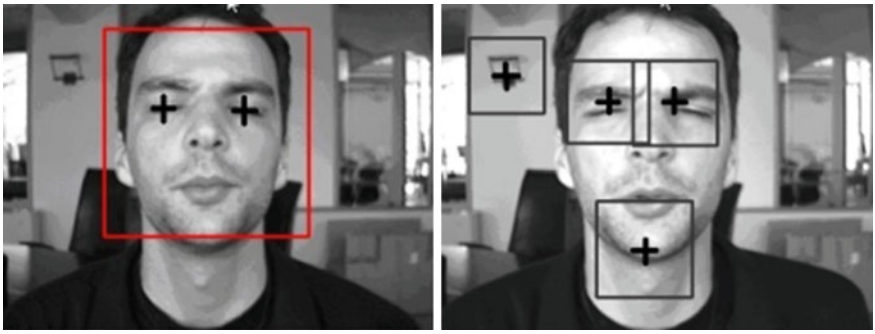


Fig. 44.2 Experimental result of the proposed method (*left* column) and the classic cascade detector (*right* column)

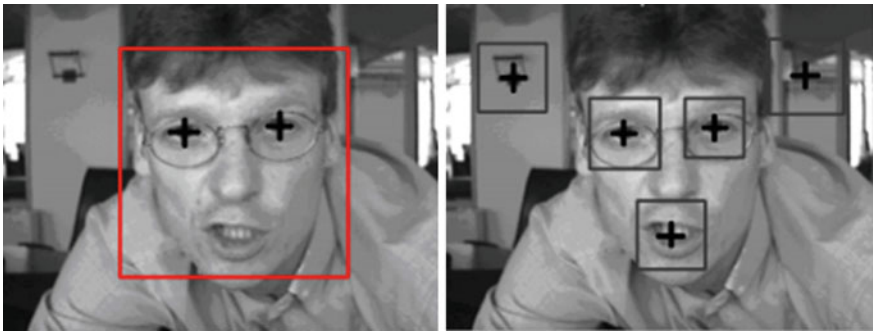


Fig. 44.3 Experimental result of the proposed method (*left* column) and the classic cascade detector (*right* column)

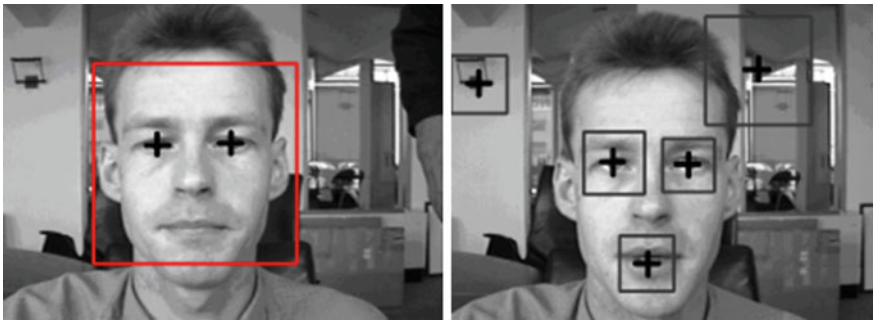


Fig. 44.4 Experimental result of the proposed method (*left* column) and the classic cascade detector (*right* column)

Fig. 44.5 Experimental result of the proposed method (*left* column) and the classic cascade detector (right column)

Fig. 44.6 Experimental result of the proposed method (*left* column) and the classic cascade detector (*right* column)

classic cascade detector, many factors, including complex backgrounds, bright or dark illumination, were found to be the potential reasons causing false positive detection.

Table 44.1 shows the time-consumption comparisons of the proposed method and the classic cascade detector. The results show that the proposed method can maintain real-time performance in spite of additional time consumption.

Figure 44.7 shows the results of the proposed method using real-time videos captured in this study. The results show high detection accuracy in cases of

Table 44.1 Time-consumption (*ms*) comparisons of the proposed method and the classic cascade detector (CPU: Pentium (R) Dual-core CPU T4500 2.30 GHz, RAM: 2G)

Method	Figure 44.1	Figure 44.2	Figure 44.3	Figure 44.4	Figure 44.5
Proposed	312	374	343	187	171
Classic cascade detector	172	234	203	140	141

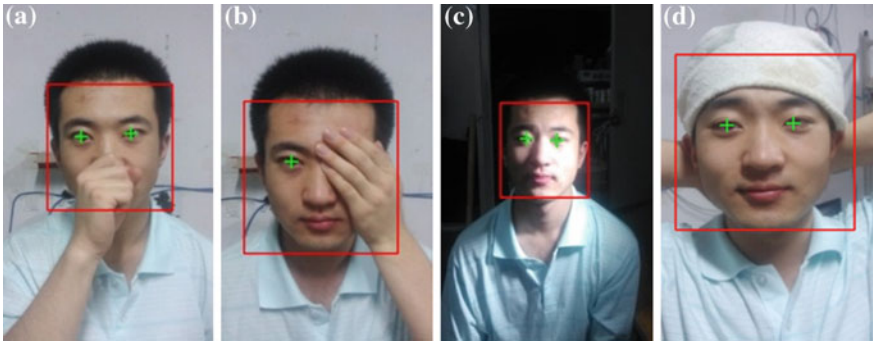


Fig. 44.7 Detected face boxes and detected eye crosses under different conditions. **a** Occlusion of nose and mouth, **b** Occlusion of one eye, **c** Intensity of illumination, **d** Wearing a hat

occlusions of the nose and the mouth or one eye, intensity of illumination, and wearing a hat. On average, the proposed method spends 439 ms on processing an image with the resolution of 338 × 566 pixels.

44.5 Conclusion

An innovative two-step method for eye detection has been proposed by using the combination of the cascade eye detector and the facial geometric features. The proposed method successfully uses the facial geometric features extracted from each pair of eye candidates to improve the detection accuracy. These features include the size similarity coefficient, the horizontal symmetry coefficient, the horizontal sextant angle coefficient, and the scale coefficient. The results of the experiments conducted on images in the BioID database and the real-time video stream show that the proposed method can effectively improve detection accuracy and achieve real-time performance.

Acknowledgments This project is supported by the Natural Science Foundation of China (Grant No. 61201370), the Independent Innovation Foundation of Shandong University (Grant No. 2012GN043), the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20120131120030), the Research Award Fund for Outstanding Middle-aged and Young Scientist of Shandong Province (Grant No. 2013BSE27058), the National Science Foundation for Post-doctoral Scientists of China (Grant No. 2013M530321), and the Independent Innovation Foundation for Post-doctoral Scientists of Shandong Province (Grant No. 201303100).

References

1. Marsico DM, Nappi M, Riccio D, Wechsler H (2013) Robust face recognition for uncontrolled pose and illumination changes. IEEE Trans Syst Man Cybern: Syst 43(1):149–163
2. Hassaballah M, Kanazawa T, Ido S (2010) Efficient eye detection method based on grey intensity variance and independent components analysis. IET Comput Vis 4(4):261–271
3. Park CW, Park KT, Moon YS (2010) Eye detection using eye filter and minimisation of NMF-based reconstruction Error in facial image. Electron Lett 46(2):130–132
4. Song ML, Tao DC, Sun Z, Li XL (2010) Visual-context boosting for eye detection. IEEE Trans Syst Man Cybern B Cybern 40(6):1460–1467
5. Choi I, Kim D (2013) Generalized binary pattern for eye detection. IEEE Signal Process Lett 20(4):343–346
6. Wang P, Green MB, Ji Q et al (2005) Automatic eye detection and its validation. In: IEEE Computer Society Conference on computer vision and pattern recognition—workshops. CVPR Workshops. IEEE, 2005:164
7. Kawato S, Tetsutani N. Real-time detection of between-the-eyes with a circle frequency filter. In: Proceedings of The 5th asian conference on computer vision (ACCV2002), vol 2, pp 442–447
8. Choi S, Lee Y, Kim C (2015) Confidence measure using composite features for eye detection in a face recognition system. IEEE Signal Process Lett 22(2):225–228
9. Zhu D, Xia S, Zhou X et al (2014) Hybrid method for human eye detection. In: Control and decision conference (2014 CCDC), The 26th Chinese, pp 5368–5373
10. Viola P, Jones M (2011) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE international conference computer vision and pattern recognition, pp 511–518

Chapter 45
MPCA on Gabor Tensor for Face Recognition

Jin Wu, Xuejie Nian, Wankou Yang and Changyin Sun

Abstract There is a growing interest in subspace learning techniques for face recognition. This paper proposes a novel face recognition method based on MPCA with Gabor tensor representation. Although the Gabor face representation has achieved great success in face recognition, the excessive dimension of the data space often brings the algorithms into the curse of dimensionality dilemma. In this paper, we propose a 3rd-order Gabor tensor representation derived from a complete response set of 40 Gabor filters. Then MPCA (Multi-linear Principal Component Analysis) is applied to each Gabor tensor to extract three discriminative subspaces. The dimension reduction is done in such a way that most useful information is retained. The subspaces are finally integrated for classification. Experimental results on ORL database show promising results of the proposed method.

Keywords Face recognition · MPCA · Gabor tensor · Representation

45.1 Introduction

Recently, face recognition has been a hot research topic in computer vision, since face recognition has potential application values as well as theoretical challenges. In the real world, face recognition is vulnerable to the various changes by expression, illumination, pose, etc.

J. Wu · X. Nian · W. Yang · C. Sun (✉)
School of Automation, Southeast University, Nanjing 210096, China
e-mail: cysun@seu.edu.cn

J. Wu · X. Nian · W. Yang · C. Sun
Key Lab of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China

J. Wu · X. Nian · W. Yang
Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_45

The Gabor wavelets [1, 2] exhibit desirable characteristics of spatial locality and orientation selectivity. It is robust to variations due to expression and illumination changes and is one of the most successful approaches for face recognition. However, while Gabor features are usually very high-dimensional data and there are redundancies among them, it is well known that face images lie in a manifold of intrinsically low dimension; therefore, the Gabor feature representations of faces could be analyzed further to extract the underlying manifold by some statistical approach such as subspace methods.

Principal Component Analysis (PCA) is a well-known scheme for dimension reduction [3, 4]. Kirby et al. [5] and Turk et al. [6] introduced the PCA method to face recognition for the first time and then, Ye et al. [7] proposed GPCA, Yang et al. [8, 9] proposed 2DPCA and Bi-2DPCA methods which retain a certain degree of spatial configuration information for the picture compared to PCA method [10]. However, in our paper, the Gabor tensor presentation is higher order tensors, if we apply the PCA (input an image object as a 1-D vector) or 2DPCA (input an image object as a matrix) to reduce dimensions, the spatial structure information from different Gabor features will be out of consideration. To avoid this, it is often helpful to process the data in its original form and order. In [11], Lu et al. proposed a scheme called MPCA to conduct dimensionality reduction.

In this paper, we apply the MPCA (Multi-linear Principal Component Analysis) for the dimension reduction of Gabor feature. We propose a 3rd-order Gabor tensor representation derived from a complete response set of 40 Gabor filters. Then MPCA is applied to each Gabor tensor to extract three discriminative subspaces. The dimensionality reduction is done in such a way that most useful information is retained. The subspaces are finally integrated for classification. Experiments on face databases show that the proposed MPCA on Gabor Tensor algorithm outperforms the traditional vector-based subspace learning algorithms.

The rest of the paper is organized as follows: Sect. 45.2 describes the Gabor tensor representation. Section 45.3 briefly reviews some basic multi-linear concepts and MPCA framework. The MPCA with Gabor tensor representation and the experimental results on ORL database are demonstrated in Sect. 45.4, and in Sect. 45.5, we conclude the paper.

45.2 Gabor Tensor Representation

The Gabor wavelets, whose kernels are similar to the two-dimensional (2D) receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity. The representation of faces using Gabor features has been extensively and successfully used in face recognition. The family of 2D Gabor kernel filters composed of five frequencies and eight orientations can be defined as follows [12]:

$$\psi_{\mu,v}(z) = \frac{\left\|k_{\mu,v}\right\|^2}{\sigma^2}\exp(-\frac{\left\|k_{\mu,v}\right\|^2\|z\|^2}{2\sigma^2}) \cdot [\exp(ik_{\mu,v}z) - \exp(-\frac{\sigma^2}{2})] \tag{45.1}$$

where μ and v define the orientation and scale of the Gabor kernels, respectively, $z = (x, y)$ denotes each pixel location of the input image. The wave vector $k_{\mu,v}$ is defined as follows:

$$k_{\mu,v} = k_v e^{i\phi_\mu} \tag{45.2}$$

where $k_v = k_{\max}/f^v$ denotes the frequency, $\phi_\mu = \pi\mu/8 \in [0, \pi)$ denotes the orientation.

In this paper, we use Gabor transforms of five different scales $v \in \{0, 1, \ldots, 4\}$ and eight orientations $\mu \in \{0, 1, 2, \ldots, 7\}$, then 40 Gabor bands with predefined parameters $\sigma = 2\pi$, $k_{\max} = \pi/2$, $f = \sqrt{2}$ are generated for preliminary feature representation.

Let $I(\mathrm{z})$ denote a grayscale facial image, and $\psi_{\mu,v}(\mathrm{z})$ be a filter corresponding to the orientation and scale, then the response of an image $I(\mathrm{z})$ to a wavelet $\psi_{\mu,v}(\mathrm{z})$ is obtained by the convolution:

$$J(\mathrm{z}) = I(\mathrm{z}) * \psi_{\mu,v}(\mathrm{z}) \tag{45.3}$$

The Gabor wavelet coefficient obtained for a given scale and orientation in the above equation is a complex number, which can be expressed as follows:

$$J(\mathrm{z}) = Ae^{i\varphi} = \mathrm{Re}(J(\mathrm{z})) + j\mathrm{Im}(J(\mathrm{z})) \tag{45.4}$$

It has been discovered that the magnitude varies slowly, while the phase information varies its rotation with the spatial position. For this reason, only the magnitude is usually used for face classification. Figure 45.1 shows the results of a face image with Gabor filters.

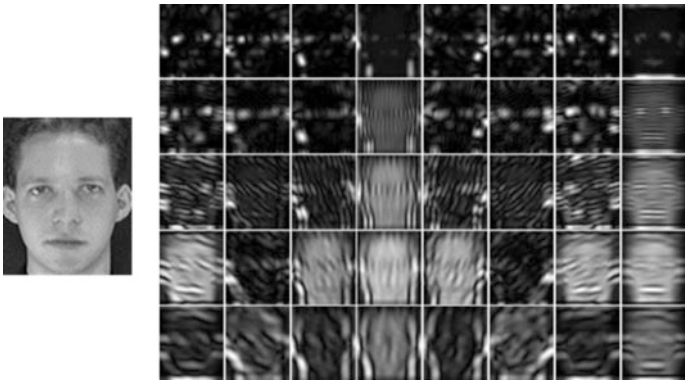


Fig. 45.1 A face image and its Gabor filter with magnitude

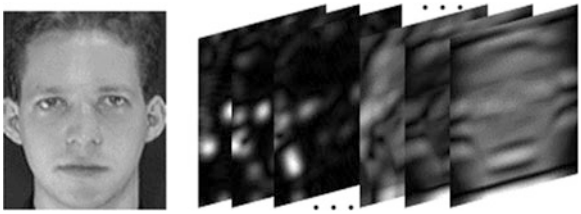


Fig. 45.2 A face image and its corresponding 3rd-order Gabor tensor

For every image pixel we have totally 40 Gabor magnitude coefficients which can be regarded as a Gabor feature vector of 40 dimensions. Therefore, a $h \times w$ 2D image can be encoded by 40 Gabor filters to form a $h \times w \times 40$ 3rd-order Gabor tensor. Figure 45.2 shows an example of a face image with its corresponding 3rd-order Gabor tensor.

45.3 Synopsis of the Multi-linear Approach

45.3.1 Notations and Basic Multi-linear Algebra

We use the following notational conventions. Indices are denoted by lowercase letters and span the range from 1 to the uppercase letter of the index, e.g. $n = 1, 2, \ldots, N$. We denote vectors by lowercase boldface letters, e.g., x; matrices by uppercase boldface, e.g., U; and tensors by calligraphic letters, e.g., χ.

A tensor is a multidimensional or N-way array of data [13]. An Nth-order tensor is denoted as $A \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, It is addressed by N indices i_n, $n = 1, 2, \ldots, N$ and each i_n addressed the n-mode of A.

The n-mode product of a tensor $A \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ by a matrix $\mathrm{U} \in \mathbb{R}^{J \times I_n}$, denoted by $A \times_n \mathrm{U}$, is a tensor $B \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \cdots \times I_N}$. Figure 45.3 shows an example of the 1-mode product of a 3rd-order tensor $A \in \mathbb{R}^{6 \times 7 \times 5}$ by a matrix $\mathrm{U} \in \mathbb{R}^{5 \times 6}$.

45.3.2 MPCA

In this section, an MPCA (Multi-linear Principal Component Analysis) solution to the problem of dimensionality reduction for tensor objects is introduced.

Let $\{A_m, m = 1, 2, \ldots, M\}$ be a set of M tensor samples in $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \cdots \otimes \mathbb{R}^{I_N}$. The total scatter of these tensors is defined as $\psi_A = \sum_{m=1}^{M} \left\| A_m - \overline{A} \right\|_F^2$, where $\overline{A}$ is the mean tensor calculated as $\overline{A} = (1/M) \sum_{m=1}^{M} A_m$. The n-mode total scatter matrix of these samples is then defined as $\mathbf{S}_A^{(n)} = \sum_{m=1}^{M} (A_{m(n)} - \overline{A}_{(n)})(A_{m(n)} - \overline{A}_{(n)})^T$,

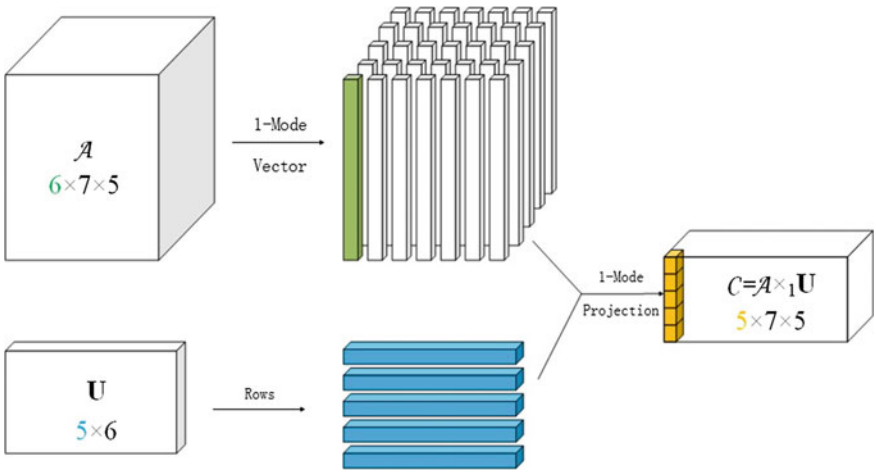


Fig. 45.3 An example of 1-mode product

where $A_{m(n)}$ is the n-mode unfolded matrix of A_m and $\otimes$ denotes the Kronecker product.

A set of M tensor objects $\{A_1, A_2, \ldots, A_M\}$ is available for training. Each tensor object $A_m \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ assumes values in a tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \cdots \otimes \mathbb{R}^{I_N}$, where the n-mode dimension of the tensor is I_n. The MPCA aims to compute a multi-linear transformation $\mathrm{U}^{(n)} \in \mathbb{R}^{P_n \times I_n}, n = 1, 2, \ldots, N$ that maps the original tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \cdots \otimes \mathbb{R}^{I_N}$ into a tensor subspace $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \cdots \otimes \mathbb{R}^{P_N}$ (with $P_n < I_n$, for $n = 1, 2, \ldots, N$): $B_m = A_m \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}, m = 1, 2, \cdots, M$, so that $B_m \in \mathbb{R}^{P_1 \times P_2 \times \cdots \times P_N}$, $m = 1, 2, \ldots, M$ captures most of the variations observed in the original tensor objects, assuming that these variations are measured by the total tensor scatter.

The pseudo-code for the MPCA algorithm could be found in reference [11].

45.4 Experimental Results

45.4.1 MPCA with Gabor Tensor Representation

In this part, we propose an idea which applies the MPCA to 3rd-order Gabor tensor. First, we compute the Gabor features $\mathbf{G}_i \in \mathbb{R}^{H \times W} (i = 1, 2, \ldots, 40)$ for each image sample which can form a 3rd-order Gabor tensor $A_i \in \mathbb{R}^{H \times W \times G}, (i = 1, 2, \ldots, N)$, where H donates the height, W donates the width of the image sample and $G = 40$. And then, we apply the MPCA method to the training set of M training samples and work out the projection matrix $\mathbf{U}^{\mathrm{n}} \in \mathbb{R}^{P_n \times I_n}, n = 1, 2, 3$ for the dimension reduction. Next, we utilize the projection matrix $\mathbf{U}^{\mathrm{n}} \in \mathbb{R}^{P_n \times I_n}, n = 1, 2, 3$ to reduce

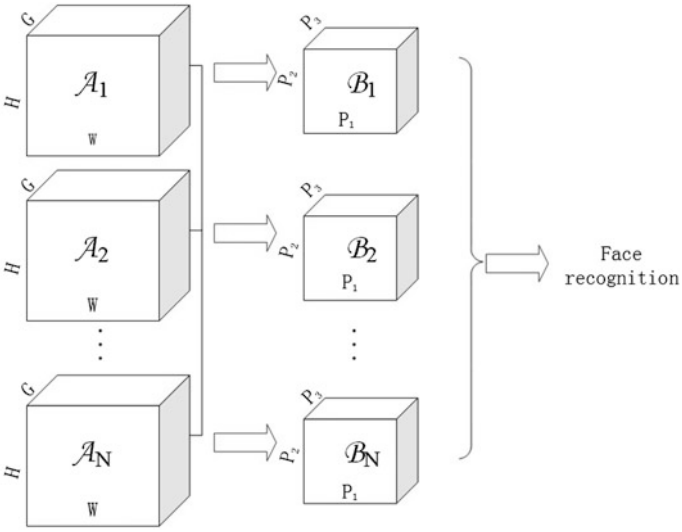


Fig. 45.4 Procedure of our method

Gabor tensor dimension of the training samples and test samples. At last, we compute the distance between the dimensions reduced tensor and decide which class the test samples belong to. The procedure is displayed in Fig. 45.4.

45.4.2 Experiments and Results

To verify the effectiveness of the MPCA with Gabor tensor algorithm, we apply our idea in the face recognition and conduct the experiments on the ORL face databases. The experimental results are compared with Gabor + PCA and PCA method.

In ORL, there are 40 persons, everyone has ten different images. There are variations in facial expression (open/closed eyes, smiling/non-smiling.), facial details (glasses/no glasses) and scale (variation of up to about 10 %). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to about 20 degrees. The images are grayscale with a resolution of 92 × 112. Figure 45.5 gives the ten different images of a person in the ORL database.

In our experiment, we select the first l ($l = 2, 3, \ldots, 9$) images of an individual for the gallery set and the rest for the probe set. We fix the feature dimension at 45 and apply the nearest neighbor classifier. The recognition rate was plotted in Fig. 45.6.

Then we choose the first five images of an individual for the gallery set and the rest for the probe set. We performed experiments with varying dimensions. The final results in Fig. 45.7 show that our method outperforms other across all feature dimensions.

Fig. 45.5 10 images of one person in the ORL database

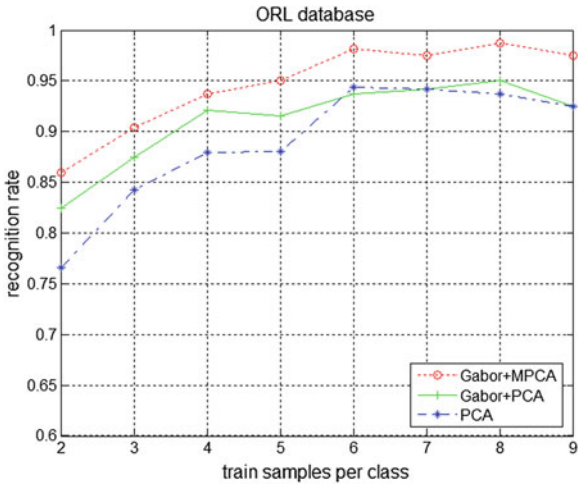


Fig. 45.6 Face recognition rate varying training samples

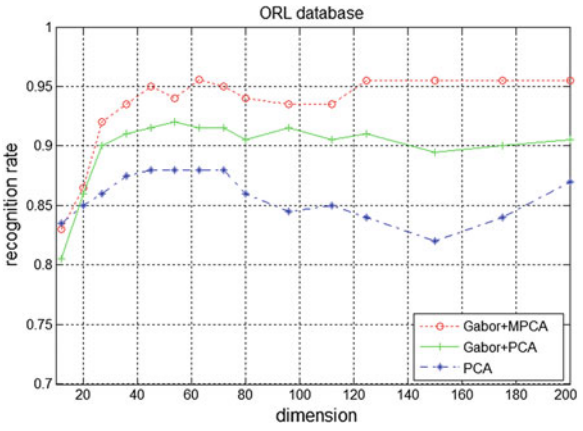


Fig. 45.7 Face recognition rate varying dimensions

45.5 Conclusion

In this paper, a novel algorithm, MPCA with Gabor tensor representation has been proposed for face recognition. A face image is first encoded as a 3rd-order Gabor tensor. After unfolding the Gabor tensor and applying MPCA analysis with them, we can effectively avoid the curse of dimensionality dilemma and overcome the small sample size problem to extract different discriminative subspaces. Experimental results on ORL database have shown the accuracy and robustness of the proposed method to the slight variation of expression and illumination. Compared with traditional algorithms, such as Gabor + PCA and PCA, our algorithm effectively avoids the curse of dimensionality dilemma and alleviates the small sample size problem.

Acknowledgments This project is partly supported by NSF of China (61375001), partly supported by the open fund of Key Laboratory of Measurement and partly supported by Control of Complex Systems of Engineering, Ministry of Education (No. MCCSE2013B01), and partly supported by the open project program of Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University (No. CDLS-2014-04), and partly supported by the China Postdoctoral Science Foundation (2013M540404), and partly supported by the Ph.D. Programs Foundation of Ministry of Education of China (No. 20120092110024).

References

1. Hamamoto Y, Uchimura S, Watanabe M, et al (1996) Recognition of handwritten numerals using Gabor features. In: International conference on pattern recognition. IEEE Computer Society, vol 3, pp 250–253
2. Wiskott L, Fellous JM, Kuiger N et al (1997) Face recognition by elastic bunch graph matching. IEEE Trans Pattern Anal Mach Intell 19(7):775–779
3. Belhumeur PN, Hespanha JP, Kriegman D (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
4. Chung KC, Kee SC, Kim SR (1999) Face recognition using principal component analysis of Gabor filter responses. IEEE international workshop on recognition, analysis, and tracking of faces and gestures in real-time systems, proceedings, pp 53–57
5. Kirby M, Sirovich L (1990) Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108
6. Turk M (1991) Pentland A. Eigenfaces for recognition. J Cogn Neurosci 3(1):71–86
7. Ye J, Janardan R, Li Q (2004) GPCA: an efficient dimension reduction scheme for image compression and retrieval. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 354–363
8. Yang J, Zhang D, Frangi AF et al (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 26 (1):131–137
9. Yang J, Xu Y, Yang JY (2010) Bi-2DPCA: a fast face coding method for recognition. Pattern Recognit Recent Adv, pp 313–340
10. Kong H, Li X, Wang L, et al (2005) Generalized 2D principal component analysis. In: IEEE international joint conference on neural networks, 2005. IJCNN'05, Proceedings. IEEE, vol 1, pp 108–113

11. Lu H, Plataniotis KN, Venetsanopoulos AN (2008) MPCA: Multilinear principal component analysis of tensor objects. IEEE Trans Neural Netw 19(1):18–39
12. Yan S, Xu D, Yang Q et al (2005) Discriminant analysis with tensor representation. IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005. IEEE, vol 1, pp 526–532
13. Kolda TG (2001) Orthogonal tensor decompositions. SIAM J Matrix Anal Appl 23(1):243–255

Chapter 46
Age Group Estimation on Single Face Image Using Blocking ULBP and SVM

Liang Hu, Zheyuan Li and Hong Liu

Abstract Since age implies essential individual information for human beings, age estimation has more and more applications in intelligent human–computer interactions and personalized recommendation in SNS, etc. However, precise age estimation based on single image is difficult due to diverse appearances among people, and irregular quality of sample acquisition. Based on general knowledge that wrinkles increase with age, Uniform Local Binary Patterns (ULBP) is always an effective texture descriptor, but it loses relative location information. In this paper, an age group estimation algorithm is proposed, where after efficient preprocessing, blocking ULBP is used to gain facial textures and a trained multi-class SVM is applied to fulfill age classification. The ages of subjects are divided into five groups: children (0–6), juveniles (7–18), youth (18–40), middle-aged (40–65), and old people (≥66). Experiments are implemented on FG-NET and Morph Aging Database and the estimation accuracy achieves 81.27 %.

Keywords Age group estimation · ULBP · PCA · SVM

46.1 Introduction

Age provides essential individual information for human being, for it can somewhat reflect the psychological and social states of certain age group. In a few decades, age group estimation based on face image has more and more applications in

L. Hu (✉)
School of Science, Minzu University of China, Beijing 100081, China
e-mail: jelly_can@163.com

Z. Li · H. Liu
Engineering Lab on Intelligent Perception for Internet of Things (ELIP),
Shenzhen Graduate School Peking University, Shenzhen 518055, China
e-mail: zheyuanli@cis.pku.edu.cn

H. Liu
e-mail: liuh@pkusz.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_46

intelligent human–computer interaction, personalized recommendation in Social Network Service, Juvenile protection system, and so on.

The earliest work of age group estimation based on single face image was published in 1993 by Young Ho Kwon and Niels da Vitoria Lobo, where a method of age estimation based on craniofacial development theory and wrinkle analysis was proposed [4]. In this paper, face images can only be classified into three age-groups: babies, adults, and senior adults. Recently, a number of in-depth studies have appeared, and according to the research methods, there are three kinds of age estimation algorithms: anthropometry model, aging pattern subspace [3], and regression model [5–7].

Since wrinkles increases with age, the facial texture is one of the important facial information. We use ULBP which is an effective texture descriptor to estimate age. However, not all textures at multiple different locations are equally important to age estimation. The location information is lost in the ULBP descriptors. So image regions are divided into N × N partition and blocking ULBP is adopted. PCA is utilized here to prove blocking ULBP is a better feature descriptor.

In this paper, an age group estimation algorithm is proposed, where after efficient preprocessing, blocking ULBP is used to gain facial textures and a trained multi-class SVM is applied to fulfill age classification. The ages of subjects are divided into five groups: children, juveniles, youth, middle-aged, and old people.

In the remainder of the paper, Sect. 46.2 describes the algorithm of age estimation based on face images, including preprocessing, feature extraction, and feature classification. Section 46.3 presents experimental l results and analysis. Conclusions are given in Sect. 46.4.

46.2 Algorithms

In this framework, face detection algorithm is first applied to obtain face regions, followed by preprocessing. Then features are extracted using ULBP; meanwhile, PCA is a global complementary descriptor for comparison experiments. Finally, feature vectors are put into SVM to train classifier for further age group estimation.

46.2.1 Preprocessing

First face images are converted into grayscale; then histogram equalization is applied to reduce the effect of illumination and complexion. As face may be tilted, we should revolve the face image to make the two eyes at the same level. The process and effect of face gesture normalization is shown in Fig. 46.1.

Locating eyes accurately is a pivotal step in the face gesture normalization. The coordinates of left eye and right eye are denoted as (x_1, y_1) and (x_r, y_r), respectively. The rotation angle θ at which the face image should be rotated is defined as formula (1):

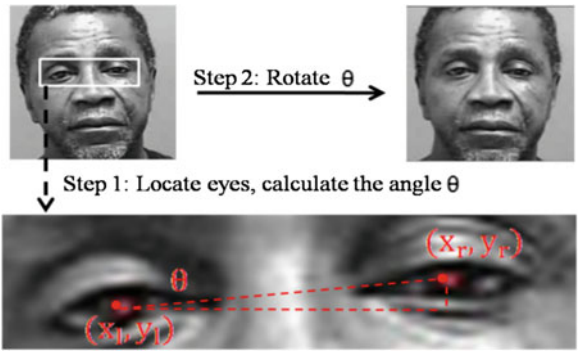


Fig. 46.1 Face gesture normalization

$$\theta = -\arctan\left(\frac{y_r - y_l}{x_r - x_l}\right) \tag{46.1}$$

P(x, y) is the final coordinate after $P(x_0, y_0)$ has rotated along θ. So the rotation matrix is shown as formula (2):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{46.2}$$

Through the matrix, the pose of face is standardized. Finally, the size of face image is normalized. The size of preprocessed face images is 200 × 200 pixels.

46.2.2 Feature Extraction Using Blocking ULBP

It is generally known that wrinkles increase with age, so the facial texture is one of the important information of age. ULBP [8] (Uniform Local Binary Patterns) is based on LBP [1] which is widely used in object recognition, image retrieval, and video analysis, is an effective operator in texture measure. First, the face image is extracted by ULBP, as shown in Fig. 46.2.

Then the ULBP image is divided into N × N partition. Then ULBP histograms of each block are obtained. Finally, the feature vector of face image is accumulated of ULBP histograms of each block. The process of 4 is shown in Fig. 46.3.

46.2.3 Feature Extraction Using PCA

PCA (Principal Component Analysis) is a holistic method of descriptors for face texture [9]. It is a multivariate statistical method for getting principal information

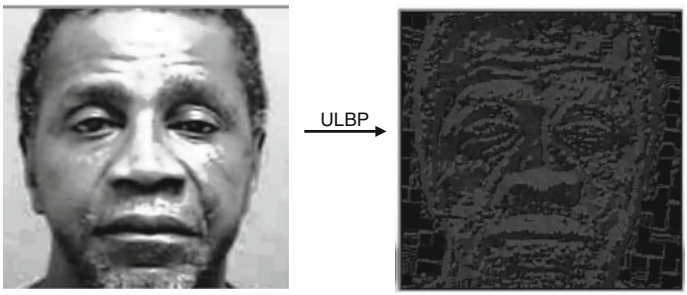


Fig. 46.2 The textual feature extracted by ULBP

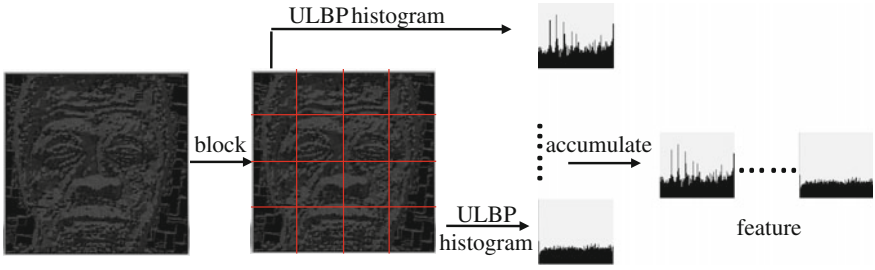


Fig. 46.3 The blocking ULBP

from observational data. So we use PCA to reduce the face image dimensions and extract the face image feature to estimate age.

46.2.4 Feature Classification

In this paper, age estimation is considered as age classification. SVM (Support Vector Machine) is a popular classifier in pattern recognition. Here a multi-class SVM is applied to fulfill the age group classification. At the same time, we take KNN (k-Nearest Neighbor) to the classification for comparing the effect of SVM.

46.3 Experimental Results and Analyses

Experiments are implemented on the platform of C++ and LIBSVM [2] from C. In the part of face detector and eyes detector, we use the free source code which is provided by mplab.ucsd.edu, which is an improved version and has been shown to perform rather well.

46.3.1 Databases

The aging datasets consist of 2,000 training images and 2,000 testing images selected randomly from the FG-NET Aging Data base and Morph Aging Database. The range in ages is from 0 to 69. Since the facial expressions are diverse and light intensity is various, the environments of datasets are close to the real world. Typical images from datasets are shown in Fig. 46. 4.

Depending on the psychological and social needs, age is divided into: children (0–6-years old), juveniles (7–18-years old), youth (18–40-years old), middle-aged (40–65-years old), and old people (older than 66-years old).

46.3.2 Effects of Different Preprocessing Phases

In the part of image preprocessing, histogram equalization and face gesture normalization are adopted. In order to verify the effects of these pretreatment methods, experiments are designed and the result is shown in Fig. 46.5.

As shown above, according to histogram equalization without face gesture normalization (deep blue bars) compared with face gesture histogram equalization and face gesture normalization (middle blue bars), the accuracies have very big promotion after face gesture normalization, especially blocking ULBP. This is because face may be tilted, without face gesture normalization, the texture in the same area of the face is divided into different blocks.

When histogram equalization and face gesture normalization (middle blue bars) is compared with face gesture normalization without histogram equalization (light blue bars), the accuracies of the former one have strange slight declines. This is because ULBP compare the relative values among near pixels of image, it is

Fig. 46.4 Typical images from datasets

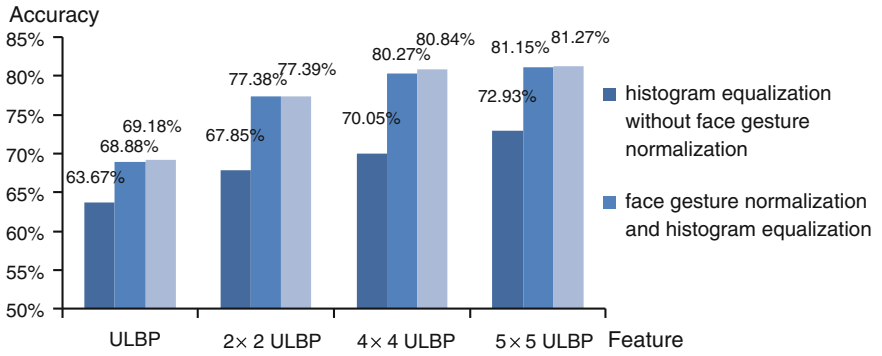


Fig. 46.5 Accuracies of different pretreatment method

insensitive to illumination conditions. Meanwhile, in the processing of histogram equalization, different grayscales are merged into the same grayscale, so some texture information is lost and accuracy declines.

In conclusion, histogram equalization is not necessary in preprocessing of face regions, as face gesture normalization does.

46.3.3 Comparison of PCA and Blocking ULBP

In this part, blocking ULBP is compared with global PCA. For ULBP, if features are extracted directly, only global texture property of face images is obtained, and the information of texture position is lost. Since not all textures at multiple different locations are equally important, image can be divided into N × N partition. The computation increases as the number of sub-blocks increases, such that complexity increases. Comparative experiments were made by using different numbers of blocks to selecting the appropriate one.

As Fig. 46.6 demonstrates, the accuracy of ULBP is higher than the accuracy of PCA, because ULBP is an effective operator in texture measure. Wrinkle is one of the important characteristics reflecting ages, so ULBP has a good effect for age estimation.

As the number of the blocks increase, the accuracy becomes higher, but accuracy will stay steady when the number of the block achieves some value. That is because when the number of the block reaches a certain value, it is enough to distinguish the different location of the texture.

The computation complexity increases as the number of blocks increases. That is because feature number increases, such that complexity increases.

According to the experiment results, age features extracted based on 4 × 4 ULBP are adopted, guaranteeing not only accuracy requirements, but also shorter time.

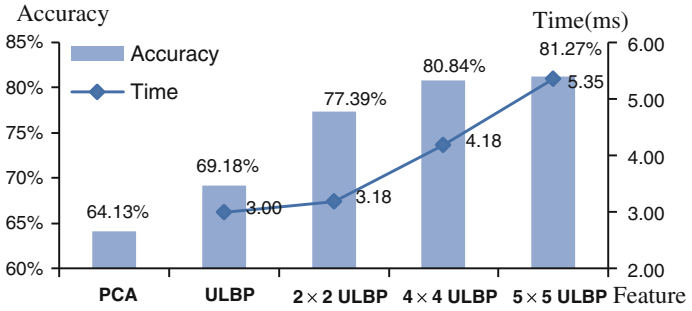


Fig. 46.6 Time/accuracies of PCA/different blocking ULBP

46.3.4 Comparison of Different Classifiers

In this part, as vector classifier, SVM is compared with traditional KNN classification.

As shown in Fig. 46.7, SVM performs better than KNNs. Because KNN is the clustering algorithm based on comparable measurement between data and face images are classified according to minimum-distance method. It is too simple to handle complex classification problems. SVM shows good performance in classification of high dimensionality. Some estimated results are listed in Fig. 46.8.

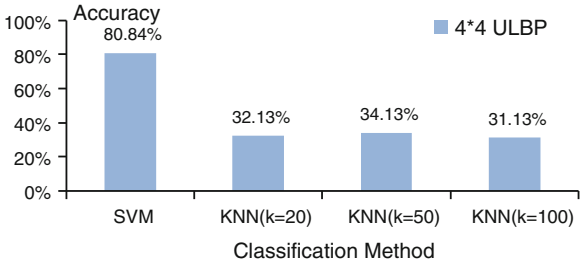


Fig. 46.7 Accuracies of different classifiers

Fig. 46.8 Some estimated results

46.4 Discussions

In detail, the accuracy reaches 80.27 % after face gesture normalization, compared to 70.05 % before, so it is very important for our algorithm. Meanwhile, ULBP is insensitive to illumination conditions, so histogram equalization is unnecessary. The age estimation rate reaches 80.84 % by 4 × 4 ULBP, with shorter feature extraction time. So it is adopted.

In this paper, ages of subjects are divided into five groups: children, juveniles, youth, middle-aged, and old people. An age group estimation algorithm is proposed, where after efficient preprocessing, blocking ULBP is used to gain facial textures and a trained multi-class SVM is applied to fulfill age classification. Besides, complementary experiments were conducted for the preprocessing effect, feature extraction, and classifier selection. The estimation rate of the algorithm achieved 81.27 %.

References

1. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. In: Computer vision-eccv, pp 469–481
2. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):27
3. Geng X, Zhou ZH et al (2006) Learning from facial aging patterns for automatic age estimation. In: Proceedings of the 14th annual ACM international conference on multimedia: ACM, pp 307–316
4. Kwon YH, da Vitoria Lobo N (1993) Locating facial features for age classification. In: Optical tools for manufacturing and advanced automation: International Society for optics and photonics, pp 62–72
5. Lanitis A, Taylor CJ (2000) Robust face recognition using automatic age normalization. In: Electrotechnical conference, 2000 MELECON 10th Mediterranean: IEEE, pp 478–481
6. Lanitis A, Taylor CJ, Cootes TF (1999) Modeling the process of ageing in face images. In: The proceedings of the seventh IEEE international conference on computer vision, 1999, pp 131–136
7. Lanitis A, Taylor CJ, Cootes TF (2010) Toward automatic simulation of aging effects on face images. IEEE Trans Pattern Anal Mach Intell 24(4):442–455
8. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7): 971–987
9. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cogn Neurosci 3(1):71–86

Chapter 47
Target Detection Algorithm Based on Gaussian Mixture Background Subtraction Model

Kejun Wang, Ying Liang, Xianglei Xing and Rongyi Zhang

Abstract Background subtraction method is an effective moving target detection method. The difficulty lies in looking for the ideal, reliable background model for complex scenes and being updated well. While Gaussian mixture model can quickly establish a good background model, process fast, and eliminate the impact of light well. So it becomes one of the commonly used methods in target detection. This paper presents a background subtraction algorithm based on Gaussian mixture. First, background can be obtained accurately using Gaussian mixture model. Then the video lost in the process of establishing is dealt with using background subtraction method. Lastly, detect the target.

Keywords Gaussian mixture model · Background subtraction · Moving target detection

47.1 Introduction

At present, there are many kinds of target detection methods, such as Background subtraction method [1], Inter-frame difference method [2], and Optical flow method [3]. Due to the dynamic changes in the practical application scenarios background, like stormy weather, Illumination changes, the impact of shadow and object in the Background, making background modeling and moving object detection becomes a very difficult task, among which, the background subtraction method is one of the most commonly used methods in moving target detection. It provides perfect information on moving targets, but is sensitive to dynamic scenes. The difficulty of background subtraction method is to find an ideal, reliable background model for complex scenes while being updated well.

K. Wang (✉) · Y. Liang · X. Xing · R. Zhang
Harbin Engineering University, College of Automation, 5th Floor, 61 Building, Harbin 150001, Heilongjiang, China
e-mail: wangkejun@hrbeu.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_47

Gaussian mixture model can quickly establish a good background model, process fast, and eliminate the impact of light well. So it becomes one of the commonly used methods in target detection. However, the background modeling method has a drawback that needs a video sequence to establish background model. Furthermore, it is unable to get accurate target information before the background of the model is obtained. In this paper, we proposed the background subtraction algorithm based on Gaussian mixture model to solve this problem. First, background can be got accurately using Gaussian mixture model. Then the video lost in the process of background establishing is dealt with using background subtraction method. And finally the target will be detected. The video sequence during the getting background process will have an impact on Gaussian mixture model parameters if only Gaussian mixture model is used. This effect will increase the probability of the emergence of regional targets and save the region as a target.

47.2 A Background Subtraction Method

Background subtraction method is an approach to use the differential of current image and background image to detect the movement regional, the basic idea of which is to build a background model similar to the background, then using differential operation between the motion sequence and the background model, and detecting the regional changes that are relative to the background model. Currently there are many background subtraction methods, the most commonly used among which are three differential method, adaptive background modeling method, w4 methods, codebook model, Hidden Markov method, and Gaussian background modeling method. In order to enhance the applicability of the background model, some researchers did a lot of research on mixed Gaussian background model, such as kernel density estimation, estimation methods based on mean replacement, covariance matrix, etc.

The most simple background subtraction is the time difference method, in which the first step is taking a video image as the current background, then doing differential operation with the next image and closing value to extract the motion area of the image. There are many improved algorithms derived from this idea, such as the three differential methods, adaptive differential method. Time differential method is not sensitive to light so that it is easy to produce cavities, and usually we cannot extract all the pixels of the moving target, that is to say, the target the detection is impossible without motion. Time differential method and filtering method based on the time axis have faster processing speed, but the algorithm is simple and rough, so it cannot meet the requirements of intelligent monitoring system for background modeling. Many researchers are working to develop a more practical background model to reduce the impact of dynamic scenes changes to the object extraction. Harltaoglu et al. used minimum, maximum intensity value, and

the maximum time differential value to statistically model each pixel in the scene image and update the background periodically [4]. Kentaro Toyama et al. proposed wallflower algorithm for image pixel level, regional level, frame-level processing [5], in which background modeling is made under the three levels respectively; Karmann and Brandt [6], Kilger [7] established adaptive background model based on Kalman filter that can adapt to different weathers and different time illumination changes. Elgammal used nonparametric model based on kernel density estimation to describe the background distribution [8], which can get the background model accurately, but the algorithm has a high demand on computer memory.

47.3 Gaussian Mixture Model

The main words in all headings (even run-in headings) begin with a capital letter. Articles, conjunctions, and prepositions are the only words which should begin with a lower case letter. Gaussian mixture model can detect moving targets under multimodal background effectively, and updated background constantly as the dynamic background changes, so it achieved good results in moving target detection. Stauffer et al. [9] modeled the distribution of each pixel with adaptive Gaussian mixture background model, which can update the model in real time, and eliminate the impact of lighting well and the interference caused by short activities of background. M. Harville et al. [10] made a theoretical derivation for Gaussian mixture algorithm and analyzed the process of parameter estimation in detail. Many researchers carried out in-depth analysis on Gaussian mixture model and proposed a variety of improved algorithms [11, 12].

Gaussian mixture model is commonly used in the probability density fitting target detection and clustering algorithms. The basic assumption is that the data to be processed follows Gaussian mixture distribution or can be produced by many Gaussian distribution. Since the Gaussian distribution has good computing features, Gaussian mixture model is widely used. In addition, it can approximate any continuous probability distribution infinitely by increasing the number of Gaussian mixture models, so the Gaussian mixture model can be arbitrarily complex. Hui et al. [13] used a single Gaussian statistical model to get the background and can be used in indoor scenes successfully, but cannot adapt to versatile outdoor scenes. Stauffer et al. [9] used a Gaussian mixture model for background modeling, and can adapt to the change of light and background confusion well.

When using Gaussian mixture model for background modeling, it is assumed that each pixel of the video sequence is independent of each other first, meanwhile, the value of each pixel can be produced by Mixed Gaussian distribution, and each pixel of the background image is represented by mixed model consists of k Gaussian distribution respectively. In [11] the probability of the current observation point discriminated as a background model is defined as

$$P(X_t) = \sum_{k=1}^{K} \omega_{k,t} \times \eta(X_t, \mu_{k,t}, \sigma_{k,t}^2) \tag{47.1}$$

where K is the number of distributions, $\omega_{k,t}$ is an estimate of the weight (the portion of the data accounted for by this Gaussian) of the kth Gaussian in the mixture at time t; $\mu_{k,t}$ and $\sigma_{k,t}^2$ are the mean value and covariance of the kth Gaussian in the mixture at time t and $\eta(X_t, \mu_{k,t}, \sigma_{k,t}^2)$ is a Gaussian probability density function, which is calculated as follows:

$$\eta(X_t, \mu_{k,t}, \sigma_{k,t}^2) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t-\mu_t)^T \Sigma_k^{-1}(X_t-\mu_t)} \tag{47.2}$$

Gaussian background modeling process can be divided into three steps, whose process is as follows:

Step 1. parameter update
The parameters of the model can change with the video sequence, and update after comparing each of the pixel and its corresponding k Gaussian distributions, and the formula are as follows:

$$\begin{cases} \omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \\ \mu_{k,t} = (1-\rho)\mu_{k,t-1} + \rho X_t \\ \sigma_{k,t}^2 = (1-\rho)\sigma_{k,t-1}^2 + \rho(X_t - \mu_{k,t})^T(X_t - \mu_{k,t}) \\ \rho = \alpha\eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \end{cases} \tag{47.3}$$

where α is the update rate of $\omega_{k,t}$; ρ is the update rate of $\mu_{k,t-1}$. $M_{k,t}$ is 1 when the size of pixel X_t match the Gaussian distribution its corresponds, otherwise it is 0. If X_t does not match the Gaussian distribution of current Gaussian mixture model, the Gaussian distribution of the lowest priority is removed, and replaced with a new Gaussian distribution to be replaced, meanwhile, Initialize the variance and weight, where the variance becomes larger, and the weight becomes smaller.

Step 2. background estimation
To describe the current context by using the Gaussian distribution of the best combination, after the Gaussian mixture model parameters are updated. And the relative value $\omega_{k,t}/\sigma_{k,t}$ acts as the reference standard, before sorting the Gaussian distribution model depending on the size ratio. The Gaussian distribution best matching background is replaced in front of the sequence. Gaussian distribution which interferes the environment will be placed at the end, and will be replaced with a new Gaussian distribution with the continuous update. Assuming the number of Gaussian distributions is N, the first B Gaussian distributions are used to describe the background.

$$B = \arg\min_{b}\left(\sum_{k=1}^{b} \omega_{k,t} > T\right) \tag{47.4}$$

where T is a set of background description threshold, and we can find the optimum Gaussian distribution to describe background by selecting T.

Step 3. foreground segmentation

Using background model of the first B described Gaussian distributions to process the current frame, Gaussian distributions are compared with each of the pixels of the current frame in accordance with the priority. And it is determined as foreground points if there is no match with a Gaussian distribution, otherwise it is determined as background points.

47.4 Background Subtraction Algorithm Based on Gaussian Mixture

For background modeling of short video sequence, background modeling will take some time at the expense of a video sequence, so the full mark cannot be made. To zsolve this problem, we use background subtraction algorithm of fusion Gaussian mixture background modeling to obtain the appropriate background by using Gaussian background modeling which is fast and good, and then make a differential operation for each frame by using background subtraction method before we get the target. For improved background subtraction model, the background image can be quickly obtained using Gaussian mixture model [14]. If background subtraction method is simply used for processing video, the desired goal can be got, but it cannot suppress noise well. Using video sequences processed by Gaussian mixture model can suppress noise to a large extent, which reduces a lot of trouble for subsequent processing.

When using a Gaussian mixture model to process the first N frames, the weights update rate α and the parameter update rate ρ are large at the beginning, the α in the update Eq. (47.3) is determined by the following formula:

$$\alpha = \begin{cases} 1/t & if\,(t < N) \\ \alpha' & else \end{cases} \tag{47.5}$$

where N is a fixed value, indicating the first N frames of the video sequence, and t is the number of frames of a video sequence processed, and α' is the learning rate of the original Gaussian mixture model. If the background model is expected to get a stable Gaussian distribution, it is better to get more gray value that needs to be added up as the Gaussian distribution consists of large numbers of gray values when N is small. When the model becomes a stable Gaussian distribution, a smaller learning rate is fine, which can increase the processing speed. But N must not be too small, as it is

difficult to achieve a stable Gaussian state, also N must not be too large, for video sequences are small, and in order to accelerate the establishment of modeling the first N frame model, the update rate α should be greater than α', namely $\alpha' \leq 1/N$, so $N \leq 1/\alpha'$. For the first t frame images, the number of the match of Gaussian mixture model for each Gaussian distribution is denoted as $M_i \quad 1, 2, \ldots, K$, and the value of update Eq. (47.3) is determined by the following formula:

$$\rho = \begin{cases} 1/M_i & if(t < N) \\ \dfrac{\alpha'}{M_i} & else \end{cases} \tag{47.6}$$

For the first N frames of the above formula, the larger the value of the Gaussian distribution, the greater the probability to be used as the background, and its mean value is closer to the true background and the parameter update rate is smaller. With the increase in the number of processed frames, parameter update rate becomes small tending to be stable, whose value is α'. So the background model becomes stable, and the background model can be established quickly and accurately.

In order to extract moving targets more accurately, and eliminate noise and flashing light and other factors, the background got from Gaussian mixture model can be taken as background, using differential operation between the current frame and the background, and using an "and" operation with moving targets detected by Gaussian mixture model after thresholding process, and then the test result is obtained.

$$B_r(x, y) = \mu_{N,t}(x, y) \tag{47.7}$$

$$D_c = \begin{cases} 1 & if|I_r(x, y) - B_r(x, y)| > T \\ 0 & else \end{cases} \tag{47.8}$$

where B_r represents the background detected by Gaussian mixture model, D_c is the moving target detected by background subtraction.

47.5 Experimental Results and Analysis

This paper conducted a comparative experiment between background subtraction algorithm based on Gaussian mixture and Gaussian mixture background modeling algorithm, the experiment of which is implemented using VC++, and Gaussian mixture model is from open-source code of opencv website. The background subtraction algorithm based on Gaussian mixture obtains background using first dozen image frames of Gaussian mixture model with background subtraction algorithm to achieve the process.

Figures 47.1, 47.2, and 47.3 show the effect of video processing, Fig. 47.1 is the original image, and Fig. 47.2 is a sequence from Gaussian mixture background

Fig. 47.1 Original sequences

Fig. 47.2 Sequences after mixture Gaussian background model

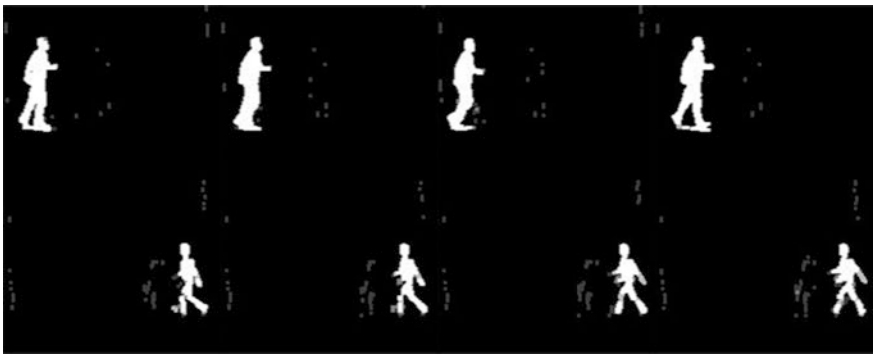


Fig. 47.3 Rendering of algorithm

modeling algorithm, it is clear that the model is at the expense of first dozen frames. And for the smaller video sequence, the ratio is relatively large for the whole video. Figure 47.3 is a rendering of algorithm, indicating that the algorithm can be a good solution to this problem and laid a good foundation for feature extraction.

The video processing result can be seen from Figs. 4.1, 4.2, and 4.3 that we selected a series from run and walk respectively. The result of detecting a target in this paper has been greatly improved relative to Gaussian mixture model, in which it can extract the complete contour figure of the human body, and does a good job of foreshadowing for further feature extraction.

47.6 Conclusion

This paper presents a background subtraction algorithm based on Gaussian mixture to solve video sacrifice problems in the background modeling, and a good result can be obtained combining the two methods and it can detect human targets well. By comparing with the Gaussian mixture background model, the experimental results show that this method is feasible, and achieves good results. Only a background subtraction algorithm combined with Gaussian mixture model algorithm is shown in this paper, and the future work is to try other background subtraction algorithms combined with Gaussian mixture model, in order to get the best result of human target detection.

References

1. Lipton AJ, Fujiyoshi H, Patil RS (eds) Moving target classification and tracking from real-time video. WACV'98 Proceedings, fourth IEEE workshop on applications of computer vision, 1998: IEEE
2. Collins RT, Lipton A, Kanade T, Fujiyoshi H, Duggins D, Tsin Y et al (2000) A system for video surveillance and monitoring: Carnegie Mellon University, the Robotics Institute, Pittsburg
3. Meyer D, Denzler J, Niemann H (eds) (1997) Model based extraction of articulated objects in image sequences for gait analysis. In: Proceedings, international conference on image processing, 1997, IEEE
4. Haritaoglu I, Harwood D, Davis LS (2000) W 4: real-time surveillance of people and their activities. IEEE Trans Pattern Anal Mach Intell 22(8):809–830
5. Toyama K, Krumm J, Brumitt B, Meyers B (eds) Wallflower: principles and practice of background maintenance. In: The proceedings of the seventh IEEE international conference on computer vision, 1999, IEEE
6. Karmann KP, Brandt A (eds) (1990) Moving object recognition using an adaptive background memory. In: Cappellini V (ed) Time-varying image processing and moving object recognition, vol 2. Elsevier Science Publishers B.V, Amsterdam
7. Kilger M (ed) (1992) A shadow handler in a video-based real-time traffic monitoring system. In: IEEE workshop on applications of computer vision, proceedings, 1992: IEEE

8. Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. Computer Vision—ECCV 2000. Springer, Heidelberg, pp 751–767
9. Stauffer C, Grimson WEL (eds) (1999) Adaptive background mixture models for real-time tracking. IEEE computer society conference on computer vision and pattern recognition, 1999, IEEE
10. Harville M (2002) A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models. In: Computer vision—ECCV 2002. Springer, Heidelberg, pp 543–560
11. Stauffer C, Grimson WEL (2000) Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intell 22(8):747–757
12. Lee D-S, Hull JJ, Erol B (eds) (2003) A Bayesian framework for Gaussian mixture background modeling. In: Proceedings 2003 international conference on image processing, 2003 ICIP 2003, IEEE
13. Hui Y, Songzhi S, Li W, Shaizi L (2008) Moving objects detection method based on the improved adaptive background mixture model
14. Wen H, Fenggang H, Han S (2005) Human gait recognition based on continuous HMM. Appl Sci Technol 32(2):50–52

Chapter 48
A Bran Specks Detection Method Based on PCNN

Tianfei Chen, Xiang Wu and Xiujuan Li

Abstract In order to achieve vision detection for tiny bran specks in flour, this paper proposes a new detection method based on pulse coupled neural network (PCNN). First, the flour image is mapped into gray entropy image using local gray entropy transformation, so the location of bran specks in flour can be enhanced in image. Then, the PCNN is utilized for the gray entropy image, and final target segmentation can be completed after iterative processing, while the optimal iteration number is determined according to the minimum cross entropy. The compared experimental results not only demonstrate the effectiveness but also show that the proposed method has a higher detection sensitivity.

Keywords Vision detection · Bran specks · Gray entropy transformation · PCNN

48.1 Introduction

Bran specks in flour are referred to as visible spots whose color is darker than wheat flour, and its main ingredient is wheat bran which is difficult to separate out completely in the process of flour milling [1]. Because bran specks will not only affect the purity and whiteness of wheat flour, but also influence other related indicators of flour production process [2]. Therefore, accurate detection for bran speck is an important part in the process of flour production.

Due to higher efficiency of vision detection, online detection for bran specks can be achieved with its help, and its core technology is to find bran specks position through image segmentation. At present, a number of image segmentation algorithms have been proposed, including algorithm using threshold [3], methods based on edge detection [4] and hybrid segmentation methods based on edge/region [5].

T. Chen (✉) · X. Wu · X. Li
School of Electrical Engineering, Henan University of Technology, No 100 LianHua Road, Zhengzhou City, China
e-mail: chen_tianfei@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_48

However, bran specks are relatively tiny targets in the captured image, for example, some bran specks are even less than 10 pixels, so the difficulty of image processing increases and most segmentation algorithm cannot get ideal results, some even fail.

A bran specks detection method based on PCNN is proposed. First, gray entropy transformation for flour image is used, and original image can be mapped into gray entropy image in which the location of bran specks are enhanced. In order to complete final image segmentation, the PCNN is employed to process gray entropy image by iteration, and the minimum cross entropy is used to determine the optimal iteration number. The experimental results demonstrate the effectiveness of this method, besides, the method avoids missing detection and has higher sensitivity.

48.2 Image Processing Algorithm

The core for bran specks detection is the image processing algorithm. The reasonable designed processing algorithm will not only improve the detection accuracy, but also avoid the situation of missing and false detection. Figure 48.1 shows the actual image of flour captured by camera. From the image, we can see that some black bran specks exist in flour, and our destination is to accurately find their locations.

48.2.1 Gray Entropy Transformation

In information theory, entropy is the measure of uncertainty for the probability of event that it can effectively reflect information which event contains. Because the gray value of pixels reflect the spatial distributions of image energy, we can combine it with the concept of entropy and extend it to the field of image processing, so the gray entropy concept can be obtained [6], and is taken as the

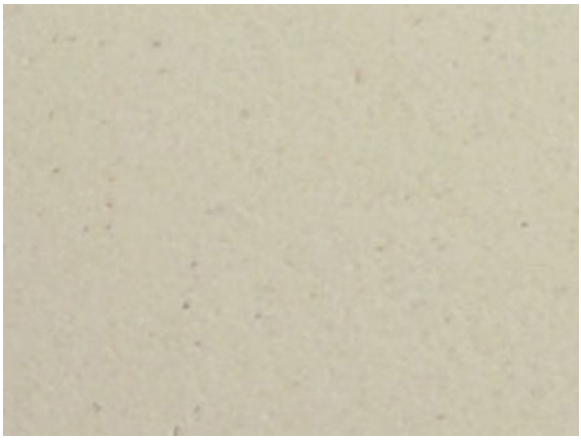


Fig. 48.1 Actual captured image for flour

measure of uncertainty for the spatial distribution of energy. For an image whose size is $M \times N$, f_{ij} represents the gray value of pixel (i, j), and the corresponding gray entropy S_f is defined as:

$$S_f = -\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij} \lg p_{ij} \tag{48.1}$$

In the formula, p_{ij} is the gray distribution of image, so $p_{ij} = f_{ij} \Big/ \sum_{i=1}^{M}\sum_{j=1}^{N} f_{ij}$. If $M \times N$ is the local window of image, and S_f is called local gray entropy of image. The gray entropy represents the macroscopic statistical characteristics of image energy distribution, and reflects difference degree of gray value. If the value of gray entropy is large, the gray value of this area is relatively uniform, and if the value of gray entropy is smaller, the difference degree of gray value of this area is larger, and the confusion level of image gray is stronger. Because the image background is uniform or slowly changed, the gray entropy of most part of the image is approximately equal, but for places where small targets appear, the gray value changes greatly, and so the corresponding gray entropy also changes greatly.

In addition, we can see that the gray entropy transformation is a kind of nonlinear spatial filtering from formula (1). When we carry on the gray entropy transformation for an image with size $W \times H$, a local window with size of $w \times h$ (typically $w = h$) is selected, and then the current pixel is taken as the center to compute its gray entropy value, the same operation is carried on for the other pixels. After that, the distribution of gray entropy value of each pixels is mapped into [0, 255] in proportion. Finally, the gray entropy image can be obtained with the processing of color reverse.

48.2.2 The Simplified PCNN Model and Parameter Setting

48.2.2.1 The Simplified PCNN Model

As the third generation of artificial neural network, PCNN [7] is founded under the inspiration of the biological visual cortex model, and it is a kind of feedback network composed of a number of interconnected neurons. Because the traditional PCNN model has a number of parameters to adjust, the simplified PCNN model is regularly chosen in practice [8], as shown in Fig. 48.2.

As can be seen from Fig. 48.2, the neurons of PCNN are an integrated dynamic nonlinear system, which includes the accepted domain, the internal activities domain (modulation part), and the pulse generator. In the field of image segmentation, a PCNN neuron is usually regarded as a pixel. When these PCNN neurons are connected in some way, a single pulse coupled neural network is formed.

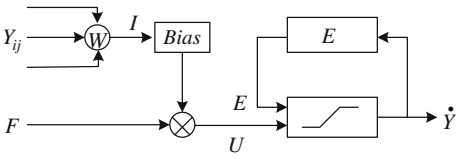


Fig. 48.2 The simplified PCNN model

When an image is input into the PCNN, if pulse happens for a PCNN neuron (or pixel), the surrounding neurons with similar gray value will also have pulse to occur, and thus a binary image is generated from the pulse sequence, which contains information such as edge, texture, area, and so on. The mathematical descriptions are as follows:

$$F_{ij}[n] = f_{ij}; \tag{48.2}$$

$$L_{ij}[n] = \sum_{kl} \omega_{ijkl} Y_{kl}[n-1]; \tag{48.3}$$

$$U_{ij}[n] = F_{ij}[n] \cdot \left\{1 + \beta_{ij} \cdot L_{ij}[n]\right\}; \tag{48.4}$$

$$\theta_{ij}[n] = \theta_{ij}[n-1] - \Delta + V_\theta \cdot Y_{ij}[n-1]; \tag{48.5}$$

$$Y_{ij}[n] = \text{step}(U_{ij}[n] - \theta_{ij}[n]); \tag{48.6}$$

In the formula, i, j is the label of PCNN neurons, n is interaction number, $F_{ij}[n]$ is the network input. Generally, $F_{ij}[n]$ is usually set to the normalized gray value of pixel. $L_{ij}[n]$, $U_{ij}[n]$, $\theta_{ij}[n]$ is, respectively, connecting input, internal activities, and dynamic threshold; ω_{ijkl} is connection weight coefficient, which reflects the influence from surrounding neurons. β_{ij} is connection strength coefficient, which mainly reflects the effect on internal activities from the connection item L_{ij}. When the item of internal activities U_{ij} is greater than the threshold θ_{ij}, the output of neuron is 1. Otherwise, the output is 0. In the interactive process, the threshold θ_{ij} decay according to certain rules.

In this paper, the minimum cross entropy [9] is used to determine the optimal number of interactions, and then we can get the final segmentation results. When the minimum cross entropy is applied to image segmentation, it generally is to determine the threshold which make the information difference minimum between original image and binary image after segmentation. The cross entropy between original image and the segmentation result is defined as:

$$D(t) = \sum_{f=0}^{t} [f \times h(f) \times \lg \frac{f}{\mu_1(t)} + \mu_1(t) \times h(f) \times \lg \frac{\mu_1(t)}{f}] + \sum_{f=t+1}^{G} [f \times h(f) \times \lg \frac{f}{\mu_2(t)} + \mu_2(t) \times h(f) \times \lg \frac{\mu_2(t)}{f}] \tag{48.7}$$

In the formula, f represents gray value of image, $h(f)$ is the histogram of image gray value, G represents the upper bound of gray value, and t is assumed threshold. $\mu_1(t)$, $\mu_2(t)$ is respectively the average gray value of target and background, shown as follows:

$$\mu_1(t) = \sum_{f=0}^{t} f \times h(f) \Big/ \sum_{f=0}^{t} h(f); \mu_2(t) = \sum_{f=t+1}^{G} f \times h(f) \Big/ \sum_{f=t+1}^{G} h(f); \tag{48.8}$$

48.2.2.2 Determination for PCNN Parameters

When the model structure of PCNN is determined, the network parameters will also have direct influence on the final segmentation results. In the simplified PCNN model, the network parameters include: connection weight coefficient ω_{ijkl}, connection strength coefficient β_{ij}, step size of threshold Δ and iteration number n. The optimal number of iteration can be determined by minimum cross entropy, so the remaining parameters are determined according to the following rules:

(1) Connection g weight coefficient ω_{ijkl}.
The connection weight coefficient reflects the effect from surrounding neurons. If gray value of central pixel is more similar to the gray value of surrounding pixels, the influence is greater. Thus, connection weight coefficient can be defined according to its gray value and surrounding spatial characteristics, as follows:

$$\omega_{ijkl} = \frac{1/(|d_{gray}(k,l) \times d_{space}(k,l)| + 1)}{\sum_{k,l} 1/(|d_{gray}(k,l) \times d_{space}(k,l)| + 1)} \tag{48.9}$$

In the formula, $d_{gray}(k, l)$ is the absolute difference of gray value between pixel (i, j) and pixel $(i + k, j + l)$; $d_{space}(k, l)$ is the Euclidean distance between two pixels. If the gray difference or euclidean distance is relatively bigger, the influence of connection weight coefficient is limited.

(2) Connection strength coefficient β_{ij}
This parameter is used to control the increased amplitude by F_{ij}, and is defined as follows:

$$\beta_{ij} = \sqrt{V_{ij}}/M_{ij} \tag{48.10}$$

V_{ij}, M_{ij} are respectively the mean and variance of gray value in local window. If the gray distribution of local window is not uniform, the value of parameter β is relatively large. Otherwise, the parameter β is small.

(3) Step size Δ
In order to traverse all possible gray values, the threshold is dynamically adjusted through linear approach. It is determined as follows:

$$\Delta = 1/n_{\max} \tag{48.11}$$

$n_{\max}$ is the maximum number of iterations.

48.2.3 Flowchart of Algorithm

According to the above content, the specific description for bran specks detection algorithm is as follows:

①. Input the captured image of flour;
②. Carry on the gray entropy transformation and proportion mapping, and the image of gray entropy is generated.
③. Set the PCNN parameters, including connection weight coefficient ω_{ijkl}, connection strength coefficient β_{ij}, and step size Δ.
④. According to the setted parameters, automatic segment the image of gray entropy.
⑤. Analyze the segmentation results to determine whether the flour is qualified.

48.3 Experiment and Analysis

Because the size of bran specks is usually very small, it is difficult to directly locate its positions in image. Therefore, the local gray entropy transformation should be applied, and thus the gray entropy image is obtained. Figure 48.3 shows the image of gray entropy. Compared with Fig. 48.1, the positions of bran specks are enhanced in the image of gray entropy.

Recently, Otsu algorithm [10] is commonly used in image segmentation to determine the optimal threshold by searching the maximum variance between target and background. The segmentation result of Ostu algorithm is shown in Fig. 48.4. According to the segmentation results, the locations of bran specks are marked on the original image, as shown in Fig. 48.5. We can see that missing detection obviously exists.

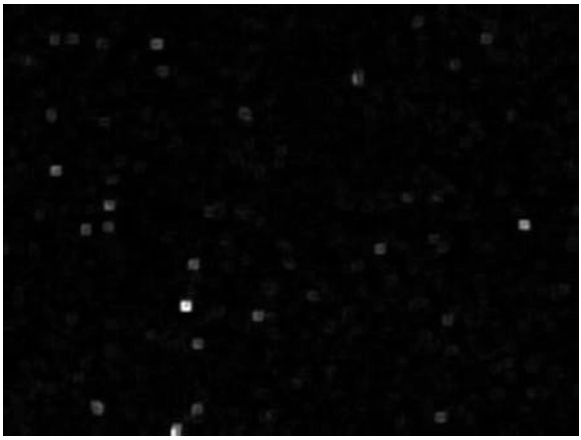


Fig. 48.3 The gray entropy image

Fig. 48.4 Segmentation result by Otsu

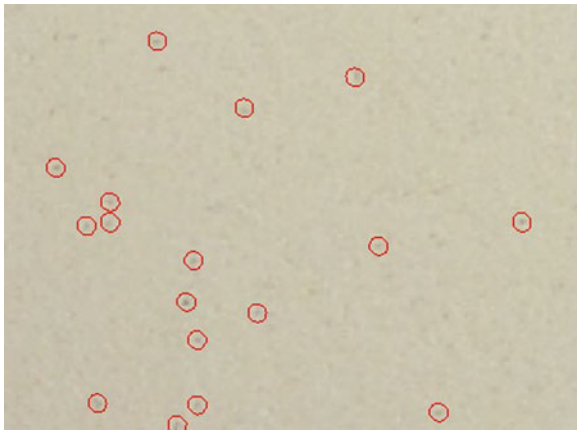


Fig. 48.5 Marked result for Otsu

Fig. 48.6 Segmentation result by PCNN

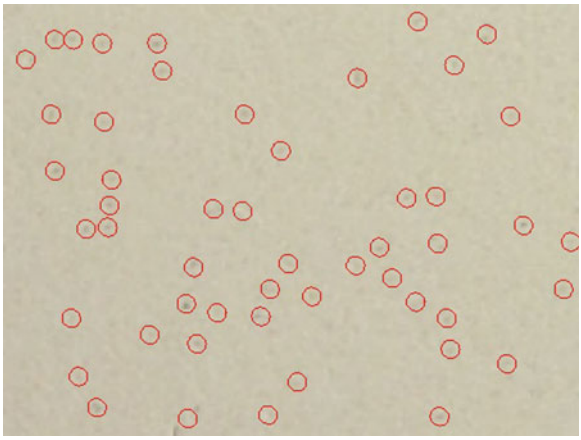


Fig. 48.7 Marked result for PCNN

In this paper, we use PCNN to segment the gray entropy image, and it is better to deal with the situation where the background and target have overlapping gray value, the adjacent pixels which have similar gray value can realize synchronous ignition for neurons, and the small gray difference in local region is ignored.

We use the principle of minimum cross entropy to determine the optimal iteration number. Therefore, once the minimum of cross entropy is obtained, the optimal iteration number can be determined at the same time. The remaining network parameters of PCNN are set according to Sect. 2.2.2.

Figure 48.6 shows the segmentation result of PCNN, and its marked result is shown in Fig. 48.7. Compared with Fig. 48.5, we can see that PCNN has higher sensitivity, the locations of bran specks that brightness is not very obvious in the gray entropy image can also be found. So the situation of missing detection is avoided, and the detection result of PCNN is better than Otsu algorithm.

48.4 Conclusion

In this paper, a bran specks detection method based on PCNN is proposed. Because size of bran specks is commonly small, the traditional image processing algorithm cannot get ideal results, some algorithm even fail. In order to ensure detection effect, the proposed method first carries on the gray entropy transformation, and the gray entropy image is formed. Then, the PCNN is employed to process the gray entropy image by iteration, and the minimum cross entropy is used to determine the optimal iteration number. The compared experimental results demonstrate the effectiveness of the proposed method, and has higher sensitivity and avoids missing detection.

Acknowledgments This work was partly supported by Science and Technology Research Project of The Education Department Henan Province (No. 14B413001) and High level talents fund of Henan University of Technology (No. 2014BS008).

References

1. Liu XX (2002) A brief overview on Chinese standards of wheat flour. J Chinese Cereals Oils Assoc 5(1):1–6 (In Chinese)
2. Li TQ (2005) Computer measurement of bran specks in wheat flour. J Chinese Cereals Oils Assoc 20(2):26–29 (In Chinese)
3. Wu YQ, Wu JM, Zhan BC (2011) An effective method of threshold selection for small object image. Acta Armamentarii 32(4):469–475 (In Chinese)
4. Amit C, Lawrence HS, James SD (1996) Deformable boundary finding in medical images by integrating gradient and region information. IEEE Trans Med Imaging 15(6):859–870
5. Zugaj D, Lattuati V (1998) A new approach of color images segmentation based on fusing region and edge segmentation output. Pattern Recognit 31(2):105–113
6. Zhang YL, Wang Y, Lu HZh (2008) Block objects detection based on entropy of brightness. Syst Eng Electron 30(2):201–204 (In Chinese)
7. Eckhorn R, Reitboeck HJ, Arndtetal M (1990) Feature linking via synchronization among distributed assemblies: simulation of results from cat cortex. Neural Comput 2(3):293–307
8. Bi YW, Qiu TSh (2005) An adaptive image segmentation method based on a simplified PCNN. Acta Electronica Sinica 33(4):647–650 (In Chinese)
9. Brink AD, Pendock NE (1996) Minimum cross-entropy threshold selection. Pattern Recognit 29(1):179–188
10. Otsu N (1979) A threshold selection method from grey level histograms. IEEE Trans Syst Man Cybern 9(1):62–66

Chapter 49
Sentiment Analysis Across Languages Based on Domain-Specific Emotional Dictionary

Shu-hui Huang, Xin Yan, Zheng-tao Yu, Xiao-hui Liu and Feng Zhou

Abstract With the rapid pace of globalization, various comments and communication platforms in different languages spring up online. Correspondingly, cross-language sentiment analysis becomes particularly important. In this paper, we describe our experience of participation in the sixth Chinese Opinion Analysis Evaluation task 2. For the deficient training corpus of specified target language, a classified method, which gets sentiment orientation across language relied on emotional dictionary, is proposed in this paper, with aid of translation tools and training corpus in Chinese. The special way goes as follows: First, Fisher criterion is used to identify benchmark words. Considering the correlation between words, words and sentences, words and documents, as well as impacts of denied words on different sentiment words, clustering gets completed by the improved information bottleneck algorithm. Subsequently, a Chinese emotional dictionary based on sentence structure will be built. Then, calculate the sentiment weights of those translated texts by referring to negative word dictionary and degree adverb dictionary, further, the trend of texts can be identified. Evaluation results have proven that the proposed method is feasible.

Keywords Across languages · Sentiment orientation analysis · Domain-specific emotional dictionary

49.1 Introduction

With the booming appearance of network technology, more and more people tend to express their views on hot events, goods, or service through various comments platforms. Research and analysis to sentiment tendencies of comments can offer better understanding to whether people are satisfied with the product or service,

S. Huang · X. Yan (✉) · Z. Yu · X. Liu · F. Zhou
School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan, China
e-mail: kg_yanxin@sina.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_49

obviously, it can help businessmen to get the poll on their product or service and improve it later. Because of worldwide rapid development of the network media, the comment texts contain multilanguage, so research on sentiment orientation across languages has a great significance.

Recently, experts are increasingly concerned about the cross-language sentiment analysis. Domestic and foreign scholars have proposed many strategies to work it out by taking comparable corpus, iterative learning, modeling, translation, and other factors into consideration. Wan [1] used the English corpus as training data and unlabeled Chinese data for co-training to study sentiment orientation of cross-language. Abbasi et al. [2] proposed a cross-language sentiment classification method by using sentences norms and syntactic features. Boyd-Graber and Resnik [3] offered a multilingual supervised LDA model, creating multilingual Dirichlet distribution based on multilingual synonym set, and then predicting the sentiment orientation of multilingual text based on topic regression method. Su [4] gave a sentiment classification method relied on bilingual feature extension, which can significantly reduce the dimension of feature vectors without loss of sentiment classification accuracy, and moreover, improve the efficiency of learning. Guilin et al. [5] presented a mixed CLOA model for cross-lingual opinion analysis, fusing self-training model and co-Training model, and obtaining good results. Liu [6] put forward a cross-language text classification algorithm dependent on active learning, which can effectively enhance the adaptability of classifier in the target language, and finally obtain excellent classification effects. Ghorbel [7] brought up a supervised learning way to train the classifier by using posts in French online forums, selecting different classification feature to obtain better classification results of cross-lingual sentiment.

When participating in the sixth session Chinese orientation analysis evaluation (COAE2014) task 2: analysis of cross-language sentiment orientation based on document level, due to lack of multilingual parallel or comparable corpus, the above methods are inoperative in analyzing orientation of multiple languages. By means of translation tools and training corpus in Chinese, this study proposes a cross-lingual sentiment analysis method based on domain-specific emotional dictionary and made a moderate success in all participating team of COAE2014. Considering the big impacts of domain knowledge on orientation selection, a Chinese domain emotional dictionary would be built, so that analysis about multiple languages finally turns to research on texts in Chinese through translating. Then, sentiment weights of those translated texts get calculated by referring to negative word dictionary and degree adverb dictionary, thus orientation of texts can be identified. Experimental results show that this strategy is feasible.

49.2 The Construction of Domain Emotional Dictionary

General sentiment dictionary is insufficient to analyze the sentiment trend of a text in specific domain. For example, a hotel evaluation text may appear the word “暗” (dim), it is an adjective without sentiment usually, which tends to express negative

sentiment there. Therefore, domain emotional dictionary [8–11] plays an important role in orientation classification in specific domain text.

In [12], the authors analyzed the effect of sentence structure to the polarity of sentiment word and proposed an approach based on sentence structure for constructing domain-oriented sentiment lexicon. The method of constructing emotional dictionary used in this study is an improved method based on [12]. Compared with [12], the method in this paper keeps an eye on the passive influence of spurious correlation between words, caused by the attached modified relation appearing between negative word and adjacent sentiment word in the same sentence. For the common use of negative word in hotel comment, elimination of spurious correlation accompanied with negative words can improve the accuracy of judging the orientation of domain sentiment words.

49.2.1 The Selection of Benchmark Word Set

The corpus we used was automatically collected from www.ctrip.com, which contains 10,000 documents and classified into two categories: positive and negative.[1] In this study corpus on hotel area is used as the foundation, and the method in [5], which exploits classification distinction and sentiment word table to choose benchmark words, is deployed to calculate the expectation and variance of the probability about the feature items' frequency in positive and negative corpus, respectively. Bring the result into the Fisher function, and acquire strong mutual distinguished characteristic words. Subsequently, sort those words from high to low by frequency in the corpus, and manually select intersection with the Chinese sentiment word set of HowNet[2] (beta version).

49.2.2 Construction of Domain Emotional Dictionary Based on Sentence Structure

Sentence structure has a certain impact on the sentiment orientation of a sentence, especially sentence with negative word or adversative word which would lead to differentiated or worse opposite sentiment orientation. Obviously, using domain emotional dictionary based on sentence structure can improve the accuracy. The construction of sentence pattern-based domain lexicon works as below: create the semantic correlation matrixes between candidate words and benchmark words, sentences, documents, and then use the information bottleneck algorithm to obtain orientation of candidate words: positive, negative, or neutral. Details are given below:

[1] http://www.searchforum.org.cn/tansongbo/senti_corpus.jsp.

[2] http://www.keenage.com/.

g_{c_i} is a candidate word, g_{s_j} is a benchmark word, and $G_c = \{g_{c_i} | 1 \leq i \leq m\}$ stands for candidate word set, $G_s = \{g_{s_j} | 1 \leq j \leq n\}$ presents benchmark word set. The improved mutual information method is used to calculate the semantic similarity between words [13], the formula is as follows:

$$I(g_{c_i}, g_{s_j}) = \log \frac{N \times p(g_{c_i}, g_{s_j})}{p(g_{c_i}) \times p(g_{s_j})} \tag{49.1}$$

where, N presents the number of corpus texts, $p(g_{c_i})$ presents the g_{c_i} 's probability; similarly, $p(g_{c_i}, g_{s_j})$ denotes the probability of g_{s_j} and g_{c_i} appearing at the same time. In particular, negation disambiguation method based on maximum entropy model [14] is used in negative sentences where the negative word has different affiliations for g_{c_i} and g_{s_j}. It means that negative words have different modified relation for g_{c_i} and g_{s_j}, taking the opposite of time number in order to eliminate the phenomenon of spurious correlation between two words for different negative words' modification. If modified relation of negative word with candidate word and benchmark word is not taken into account, there will be a semantic related false phenomenon. Then use the method mentioned in [7] to do Laplace smoothing.

$$\text{sim}(g_{c_i}, g_{s_j}) = \frac{\beta_1 n_s(g_{c_i}, g_{s_j}) + \alpha\beta_2 n_d(g_{c_i}, g_{s_j}) + \lambda}{\sum_{g_i \in G_c, g_j \in G_s} [n_s(g_i, g_j) + \alpha n_d(g_i, g_j) + \lambda]} \tag{49.2}$$

where, $n_s(g_{c_i}, g_{s_j})$ represents the number of g_{c_i} and g_{s_j} simultaneously in one sentence; $n_d(g_{c_i}, g_{s_j})$ represents the time of g_{c_i} and g_{s_j} appearing in the adjacent sentence; λ represents smoothing parameter; α represents sentiment weight ($0 < \alpha < 1$). Apparently, the use of negation elimination can be divided into following conditions:

1. Candidate word g_{c_i} and benchmark word g_{s_j} occur in the same sentence, the orientation tends to be inconsistent. For example: "房间通风不好且有异味" (this room is not well ventilated and has a nasty smell), this study finds that using the method of [14] concludes that the negative word "不" (not) modifies "好" (well), and does not modify "异味" (nasty smell), so words "好" (well) and "异味" (nasty smell) have an opposite orientation. Therefore, the parameter β_1 is added in the calculation of word relevancy, if the candidate word and benchmark word do not have a negative word modification in one sentence at the same time, $\beta_1 = -1$; else, $\beta_1 = 1$.
2. When candidate word g_{c_i} and benchmark word g_{s_j} occur in the adjacent sentences, the parameter β_2 gets added. If two sentences have negative words and these negative words have the same modified relation toward the candidate and benchmark word, $\beta_2 = 1$, else, $\beta_2 = -1$. If there is only one negative sentence and the negative word do have the modification relation toward the candidate or benchmark word, $\beta_2 = -1$; else, $\beta_2 = 1$.

Build matrix B of m rows and 2 columns $B = [B_{ij}]_{m \times 2}$ to represent the semantic relativity between benchmark words and m candidate words, B_{i1} refers to the sum of

the relativity between candidate word g_{c_i} and all benchmark words in positive word set, B_{i2} refers to the sum of the relativity between candidate word g_{c_i} and all benchmark words in negative word set.

Define candidate word set $G = \{g_i | 1 \le i \le m\}$, sentence set $S = \{s_j | 1 \le j \le n\}$, and to calculate the correlation aff(g_i, s_j) between candidate word g_i and sentence s_j. Here, if candidate word g_i appear in sentence s_j, the correlation is aff(g_i, s_j), else is 0.

Experiments show that negative sentence and turning sentence would cause inconsistent orientation between the sentence and the word. So, parameter γ is added to calculate the correlation sim(g_i, s_j) between candidate word g_i and sentence set s + (s-).

$$\mathrm{sim}(g_i, s) = \sum_{s_j \in s} \gamma \mathrm{sim}(g_i, s_j) \tag{49.3}$$

if s_j is negative sentence, the value of γ is -1; else if s_j is turning sentence, the value of γ is $\delta(0 < \delta < 1)$; else, γ is 1. Here:

1. If s_j is a negative sentence, judge which sentiment word is modified by negative word in one sentence by using the method of negation disambiguation. Sentiment word modified by negative words has the opposite orientation with its sentence, and rest sentiment words have the same orientation in the sentence.
2. If s_j is a turning sentence, its semantic focus is on the adversative clause. When the candidate word locates in adversative clause, the δ value is 1; else the δ is 0.

Therefore, δ values 0.5.

Likewise, build matrix C of m rows and 2 columns $C = [C_{ij}]_{m \times 2}$ to represent the semantic relativity between benchmark words and candidate words, C_{i1} refers to the semantic relativity between candidate word and all positive sentences, C_{i2} refers to the semantic relativity between candidate words and all negative sentences. $E = [E_{ij}]_{m \times 2}$ represents the semantic relativity between candidate words and documents, E_{i1} represents the semantic relativity between candidate words and positive documents, E_{i2} represents the semantic relativity between candidate words and negative documents. The correlation strength calculation is offered in [11]. Finally, get above three matrixes normalized.

Domain emotional dictionary can be obtained by using information bottleneck algorithm, clustering all candidate words into three piles: positive, negative, and neutral set.

49.3 Experiment and Result Analysis

The experiment steps are as below (Fig. 49.1):

1. Google interface is called to translate the original target texts (English, French, German, and Spanish four languages on hotel review) into Chinese, then Chinese target texts are obtained;

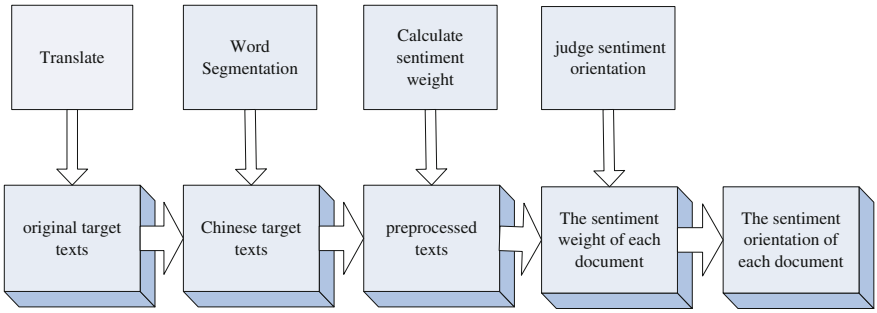


Fig. 49.1 Experimental flowchart

2. employ ICTCLAS to obtain the Chinese text segmentation, the results called preprocessed texts;
3. calculate the sentiment weights of each preprocessed documents in the text by using domain sentiment dictionary, degree adverb, and negative dictionary;
4. obtaining the sentiment orientation of each document according to the sentiment weight calculated above.

49.3.1 Experiment Data

The sentiment word set of HowNet (beta version) is chosen, in which positive Chinese evaluation words are combined with Chinese positive sentiment words to obtain positive foundation emotional dictionary, owning a total of 4566 words; Chinese negative evaluation words is combined with Chinese negative sentiment words, negative foundation emotional dictionary is obtained, which has a total of 4370 words. The obtained positive foundation emotional dictionary and negative foundation emotional dictionary merge into a whole foundation emotional dictionary. The benchmark words are formed from the intersection of the characteristic words chosen by Fisher criterion and the foundation emotional dictionary. Then extract words whose part of speech tag is a, an, ad, al, or vl as candidate words, employing the information bottleneck algorithm mentioned above and the evaluation corpus on hotel mentioned in Sect. 49.2.1 to construct the hotel area emotional dictionary, and get 86 positive benchmark words and 73 negative benchmark words; the final domain emotional dictionary can be obtained with positive words 1052, also negative words 864.

The object of this study is cross-language sentiment analysis in four languages on hotel review, which contains 2000 German texts, 2000 Spanish texts, 2000 French texts, and 4000 English texts, containing 4699, 3620, 5536, and 12,600 sentences separately.

49.3.2 Sentiment Weights Calculation

Constructed emotional dictionaries in this study contain positive word, negative word, negative word, and degree adverb word dictionary. Where, the negative word dictionary is manually selected and the degree adverb dictionary is constructed by using "Chinese degree words" in "word set for sentiment analysis (beta version)" of HowNet. Total sentiment weight of the sentence is calculated as follows:

$$O_b = \sum_{i=1}^{n} \sum_{j=1}^{k} (-1)^s Q_{mj} P_{m_a} \tag{49.4}$$

where, n stands for the sentence number in one document; k stands for the sentiment word number in one sentence; s stands for the number of the negative word which appear before the sentiment word m_j; Q_{m_i} is the value of the sentiment word m_j and P_{m_a} is the value of degree adverb words which exists before sentiment word m_j.

If the total sentiment weight of a document is bigger than 0, the sentiment orientation of this document is positive; If it is less than 0, this document sentiment orientation is negative; otherwise, the sentiment orientation of this document is neutral.

49.3.3 Evaluation and Analysis of Result

Corpus mentioned in Sect. 49.2.1 is used as the foundation, and accuracy is used to evaluate the performance of the algorithm of labeling the sentiment orientation of candidate words. The calculation formula of accuracy is described as formula (49.5), where W_a represents candidate word set, w represents candidate word, C stands for w polarity function (the standard sentiment orientation of these candidate words is obtained by manual annotation), F is a polarity function which is obtained by w using clustering algorithm.

$$\text{Accuracy}(w) = \frac{\{w | w \in W_a \wedge C(w) = F(w)\}}{|W_a|} \tag{49.5}$$

The accuracy evaluation result is shown in Table 49.1. It can show that features extracted by Fisher criterion mentioned above and the foundation emotional dictionary takes intersection for the construction of domain emotional dictionary. This method can enlarge the foundation sentiment words in scope and effectively expand

Table 49.1 Comparison of accuracy in sentiment orientation of candidate words

Domain dictionary construction algorithm	Accuracy
The improved sentence structure algorithm	87.61 %
Sentence structure algorithm in [12]	85.2 %

Table 49.2 Evaluation results

KMUST_LIIP	Hotel_de_2000	Hotel_en_4000	Hotel_es_2000	Hotel_fr_2000
Pos_P	0.777	0.792	0.729	0.777
Pos_R	0.896	0.942	0.865	0.835
Pos_F1	0.832	0.86	0.791	0.805
Neg_P	0.909	0.935	0.848	0.929
Neg_R	0.527	0.607	0.501	0.5
Neg_F1	0.667	0.736	0.63	0.65
Macro_P	0.843	0.864	0.789	0.853
Macro_R	0.712	0.775	0.683	0.668
Macro_F1	0.75	0.798	0.711	0.728
Micro_P	0.712	0.775	0.683	0.668
Micro_R	0.712	0.775	0.683	0.668
Micro_F1	0.712	0.775	0.683	0.668

domain sentiment vocabulary. It can also be seen from the table that the reason why the method of construction of domain emotional dictionary based on sentence pattern has a high accuracy is that sentence structures do have a certain impact on the sentiment orientation of a sentence. Compared with the construction methods mentioned in [12], the method using mutual information to calculate the semantic similarity between candidate words and benchmark words has a better improvement; the use of negation disambiguation realizes the correct judgment toward affiliated relation of negative words in a negative sentence, further, false semantic correlation is avoided and the judgment accuracy of semantic correlation has been improved.

By participating in COAE2014, we get the experimental data (see Table 49.2):

This evaluation task makes a moderate success in all participating team. Table 49.2 has shown that English texts own a better accuracy of sentiment orientation analysis, also a higher Recall Rate and *F* value than texts in another three languages. This is mainly caused by the high accuracy of translation into Chinese compared with the other three using Google Translation. Moreover, the negative text has a better accuracy than the positive when it comes to sentiment orientation analysis, accompanied with a lower Recall Rate and *F* value. It can be finally concluded that the negative text is easier to be mistaken as a positive one. The reason may lie in the less comparable vocabulary of the constructed domain negative word dictionary.

49.4 Conclusions

With the aid of translation tools and Chinese training corpus, this paper proposes a method to analyze cross-language sentiment orientation based on domain-specific emotional dictionary, constructing a Chinese domain-specific sentiment dictionary

dependent on sentence structure. Then combine the negative dictionary with degree adverb dictionary to calculate the sentiment weight of the Chinese text and estimate the sentiment orientation of each text using this sentiment weight. In order to acquire better results of cross-language sentiment orientation analysis, the next step is to combine comparable corpus, syntactic analysis, and semantic information to get better results.

Acknowledgments This work was supported by National Natural Science Foundation of China (NSFC) via grant 61462055 and 61175068.

References

1. Wan XJ (2009) Co-training for cross-lingual sentiment classification. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP, vol 1, Association for Computational Linguistics, pp 235–243
2. Abbasi A, Chen H, Salem A (2008) Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums. ACM Trans Inf Syst (TOIS) 26(3):12
3. Boyd-Graber J, Resnik P (2010) Holistic sentiment analysis across languages: multilingual supervised latent Dirichlet allocation. In: Proceedings of the 2010 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 45–55
4. Su Y (2013) Research on bilingual sentiment classification. Master's thesis, Suzhou University, Jiangsu (in Chinese)
5. Gui L, Xu R, Xu J, Yuan L, Yao Y, Zhou J, Cheung R (2013) A mixed model for cross lingual opinion analysis. In: Natural language processing and Chinese computing, Springer, Berlin, pp 93–104
6. Liu Y (2011) Research on cross-language text classification. Master's thesis, Beijing Institute of Technology (in Chinese)
7. Ghorbel H (2012) Experiments in cross-lingual sentiment analysis in discussion forums. In: Social informatics, Springer, Berlin, pp 138–151
8. Chen XD (2012) Research on sentiment dictionary based emotional tendency analysis of Chinese microblog. Master's thesis, Huazhong University of Science and Technology (in Chinese)
9. Wang SG, Li DY, Wei YJ, Song XL (2009) A synonyms based word sentiment orientation discriminating. J Chin Inf Process 5:009 (in Chinese)
10. Yang C, Feng S, Wang DL, Yang N, Yu G (2010) Analysis on web public opinion orientation based on extending sentiment lexicon. J Chin Comput Syst 4:691–695 (in Chinese)
11. Zhang XQ, Jiang XF (2010) Domain-oriented sentiment lexicon based on sentence-level corpus. In: Fifth national youth workshop on computational linguistics (in Chinese)
12. Liao XW, Zhang XQ (2011) Domain-oriented sentiment lexicon based on sentence structure. J Fuzhou Univ (Nat Sci Ed) 4:012 (in Chinese)
13. Wu Q, Tan S, Zhai H, Zhang G, Duan M, Cheng X (2009) SentiRank: cross-domain graph ranking for sentiment classification. In Proceedings of the 2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology, IEEE Computer Society, vol 01, pp 309–314
14. Zhang C, Fei X, Zhu J (2010) Negation disambiguation using the maximum entropy model. In: Proceedings of the 6th international conference on natural language processing and knowledge engineering (NLPKE-2010), pp 1–5

Chapter 50
An Adaptive Color Image Segmentation Algorithm Based on Gaussian Mixture Model Applied to Mobile Terminal

Jia-Qiang Wang, Han-Bing Qu, Wei Jin, Chao Hu and Hai-Jun Tao

Abstract This paper presents an adaptive color image segmentation algorithm based on Gaussian mixture model. Image edge posterior probability density is estimated with Gaussian mixture model. To derive image edge, we use variational approximation method to estimate Gaussian mixture model parameters. Finally, we give some image segmentation experiments to verify performance of our algorithm combined with priori position information. Experimental results show that our algorithm can be applied to color image segmentation on mobile terminal.

Keywords Gaussian mixture model · Posterior probability density · Variational approximation

50.1 Introduction

The penetration of mobile terminal with camera is increasing obviously, accompanying the demand for image processing on these mobile terminals, such as mobile phone, iPad. Image segmentation algorithm running on mobile terminal must face the difficulties of great difference in image quality, complex image background, and limited computing performance of mobile terminal. It is significant to develop image segmentation and recognition-based mobile terminal.

Currently, there are many complex environments image segmentation algorithms obtained, such as Level Set [1–3], Graph Cut [4–6], and Bag-of-Word [7], they have an excellent performance on image segmentation, especially when the background is complex. However, the character of high computational complexity, large

J.-Q. Wang (✉) · H.-B. Qu · W. Jin · C. Hu
Pattern Recognition Research Center, Beijing Institute of New Technology Applications, Beijing 100094, China
e-mail: jqwang.bj@gmail.com

H.-J. Tao
China Jiliang University, Hangzhou 310018, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_50

training samples of these algorithms cause the poor application on mobile terminal devices.

Recent years, along with in-depth theoretical study, the probability model algorithm has gradually penetrated into many areas of computer vision, and achieved great success. Compared with traditional image segmentation algorithm, probabilistic model algorithm assumes that the model parameters are random variables, priori probability distribution is assigned first, then the posterior probability distribution parameters are obtained by sampling or approximation methods. Probability model algorithm gains the just balance between applicability and computational complexity. Therefore, further research is needed about probabilistic model framework in field of image segmentation.

50.2 Gaussian Mixture Model Parameters Variational Approximation

Gaussian mixture model is a commonly used probability density estimation method, by combination of multiple weighted distribution, that can estimate effectively for double peak distribution or multipeak distribution of 1-D data, and bimodal or multimodal distribution of N-D data. For Gaussian mixture model, a variety of classic model parameter estimation methods already exist, this paper uses variational approximation method to estimate Gaussian mixture model parameters.

Gaussian mixture model can be expressed as follows:

$$p(\chi|\mathcal{Z},\mu,\Lambda)=\prod_{n=1}^{N}\prod_{k=1}^{K}\mathcal{N}\left(x_n|\mu_k,\Lambda_k^{-1}\right)^{z_{nk}} \tag{50.1}$$

where $\chi=\{x_1,\ldots,x_N\}$ is a collection of points, consisting of N points. $\mu=\{\mu_k\}$ is the mean of Gaussian model, $\mathcal{Z}=\{z_1,\ldots,z_N\}$ is corresponding indicator latent variable, z_n is a $1-\text{of}-K$ binary vector with elements $\{z_{n,k},k=1,\ldots,K\}$, $\{\Lambda_k\}_{k=1}^{K}$ is the precision matrix for each Gaussian component in the mixture model. The conditional distribution of latent variable $\mathcal{Z}$ given the mixing coefficients π is

$$p(\mathcal{Z}|\pi)=\prod_{n=1}^{N}\prod_{n=1}^{N}\pi_k^{z_{nk}} \tag{50.2}$$

The prior over mixing proportions π is Dirichlet distribution with hyperparameter α^0.

$$p\left(\pi|\alpha^0\right)=C\left(\alpha^0\right)\prod_{k=1}^{K}\pi_k\alpha_k^0-1 \tag{50.3}$$

$C(\alpha^0)$ is the normalization constant. The joint prior distribution of $\{\mu_k, \Lambda_k\}$ is Gaussian-Wishart with hyperparameter m_0, β_0 is the relative precision of mixture component, W_0 is the scale matrix of precision, and v_0 is degree of freedom of mixture precision.

After derivation, we can get a series of variational approximation formulas similar to EM algorithm, consisting of the following three steps:

Step 1. Each Gaussian component response calculation of each data element.

$$r_{nk} \propto \tilde{\pi}_k \tilde{\Lambda}_k^{\frac{1}{2}} \exp\left\{-\frac{D}{2\beta_k} - \frac{v_k}{2}(x_n - m_k)^T W_k (x_n - m_k)\right\}. \tag{50.4}$$

With

$$\ln \tilde{\Lambda}_k = \sum_{i=1}^{D} \psi\left(\frac{v_k + 1 + i}{2}\right) + D \ln 2 + \ln \|W_k\| \tag{50.5}$$

$$\ln \tilde{\pi}_k = \psi(\alpha_k) - \psi(\hat{\alpha}) \tag{50.6}$$

Where $D = 1$, r_{nk} is the responsibility matrix with regard to $\mathcal{Z}$, β_k is the relative precision of mixture component, υ_k is the degree of freedom of mixture precision, m_k is the mean of mixture components. α is prior sample size of mixing coefficients π. ψ is noise precision matrix.

Step 2. Statistics calculation of each data.

$$N_k = \sum_{n=1}^{N} r_{nk} \tag{50.7}$$

$$\bar{x}_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} x_n \tag{50.8}$$

$$S_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} (x_n - \bar{x}_k)(x_n - \bar{x}_k)^T \tag{50.9}$$

Step 3. Update posterior probability distribution parameters of each variable.

$$\alpha_k = \alpha_0 + N_k \tag{50.10}$$

$$\beta_k = \beta_0 + N_k \tag{50.11}$$

$$m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k \bar{x}_k) \tag{50.12}$$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^T \tag{50.13}$$

$$v_k = v_0 + N_k \tag{50.14}$$

Then we obtain the posterior probability of Gaussian mixture model parameters.

50.3 Image Edge Detection Based on Gaussian Mixture Model

Images captured by mobile terminal always contain complex background, take palm print recognition as an example, you need to capture palm image and pre-process the original image to get the ROI. But the original image has different environment, as Fig. 50.1 shows, image brightness and background vary widely. It is difficult to crop and identify palm region from image using common edge detection algorithm and binary method. Figure 50.2 shows the edge detection result with Sobel and binarization result with OTSU directly.

Considering the actual situation of computing performance of mobile terminal, we use hybrid probabilistic model estimate probability density of image edge information.

We assume color image as $I(x, y)$, then edge information can be written as

$$G(x, y) = \left|\frac{\partial I(x, y)}{\partial x}\right| + \left|\frac{\partial I(x, y)}{\partial y}\right| \tag{50.15}$$

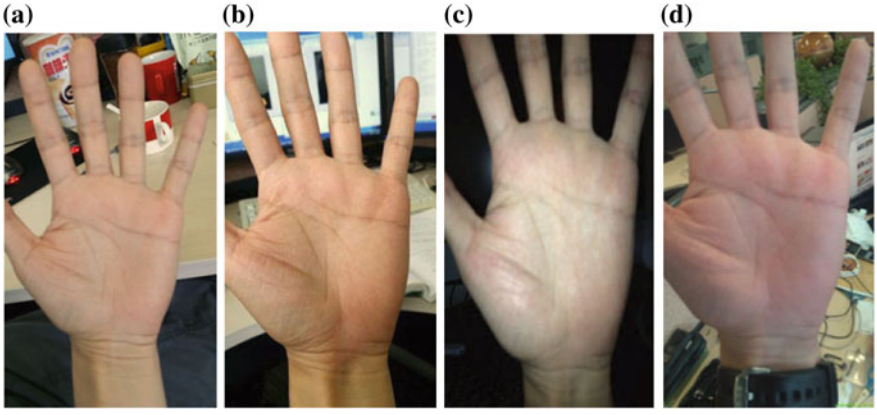


Fig. 50.1 Palm images captured in different environments. **a** is image 1. **b** is image 2. **c** is image 3. **d** is image 4

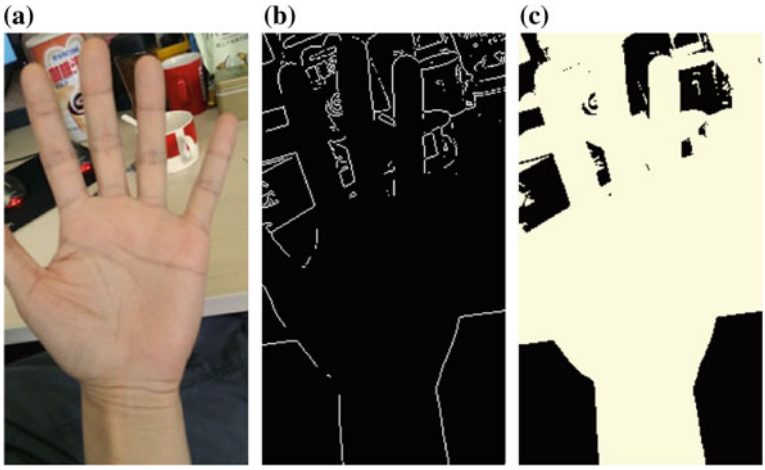


Fig. 50.2 Sobel edge detection and OTSU binarization directly; **a** orginal image; **b** sobel edge detection image; **c** otsu binarization image

Defining r, g, and b as unit vectors in the direction of R, G, and B, respectively, we have the vector

$$u = \frac{\partial R}{\partial x} r + \frac{\partial G}{\partial x} g + \frac{\partial B}{\partial x} b \tag{50.16}$$

$$\upsilon = \frac{\partial R}{\partial y} r + \frac{\partial G}{\partial y} g + \frac{\partial B}{\partial y} b \tag{50.17}$$

Defining g_{xx}, g_{yy}, and g_{xy} as dot products of u and υ, we get

$$g_{xx} = u^T u = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2 \tag{50.18}$$

$$g_{yy} = \upsilon^T \upsilon = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2 \tag{50.19}$$

$$g_{xy} = u^T \upsilon = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y} \tag{50.20}$$

Finally, we can obtain color image edge information by the following equation:

$$G(x, y) = \frac{1}{2}\sqrt{g_{xx} + g_{yy} + g_{xx}^2 + g_{yy}^2 - 2g_{xx}g_{yy} + 4g_{xy}^2} \tag{50.21}$$

Fig. 50.3 Edge magnitude

Processing hand image is shown in Fig. 50.1 in accordance with Eq. (50.20), the edge magnitude image is shown in Fig. 50.3.

Gaussian mixture model parameters estimation based on variational Bayesian framework makes estimation of collection points probability distribution possible. Using the above Eqs. (50.4–50.14) estimate image edge $G(x, y)$ probability density, then we can obtain the posterior probability density $p(G(x, y))$.

$$p(G(x, y)) = \sum_{k=1}^{2} \pi_k \mathcal{N}(G(x, y)|\mu_k, \Lambda_k) \tag{50.22}$$

The image segment result using the above calculation is shown in Fig. 50.4a.

If we combine connected domain and hand position prior knowledge, assume that connected domain contains center point of image as palm region, palm image segmentation can be obtained, as shown in Fig. 50.4b.

We can see that if the background color is relatively close to the hand, foreground by our algorithm still contains some background information, additional information should be added to segmentation accurately, such as hand color.

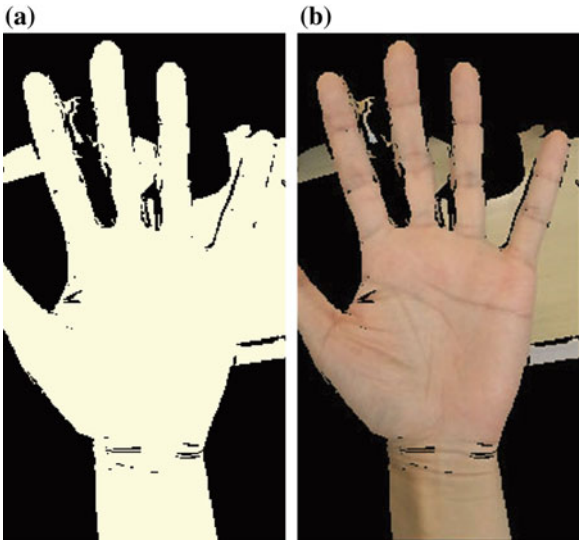


Fig. 50.4 Image segment using image edge Gaussian mixture model. **a** image segment using image edge Gaussian mixture model. **b** palm image segmentation

50.4 Experimental Results

We present an adaptive color image segmentation algorithm based on Gaussian mixture model. Image edge posterior probability density is estimated with Gaussian mixture model. To verify our algorithm, some image segmentation are presented, as shown in Fig. 50.5.

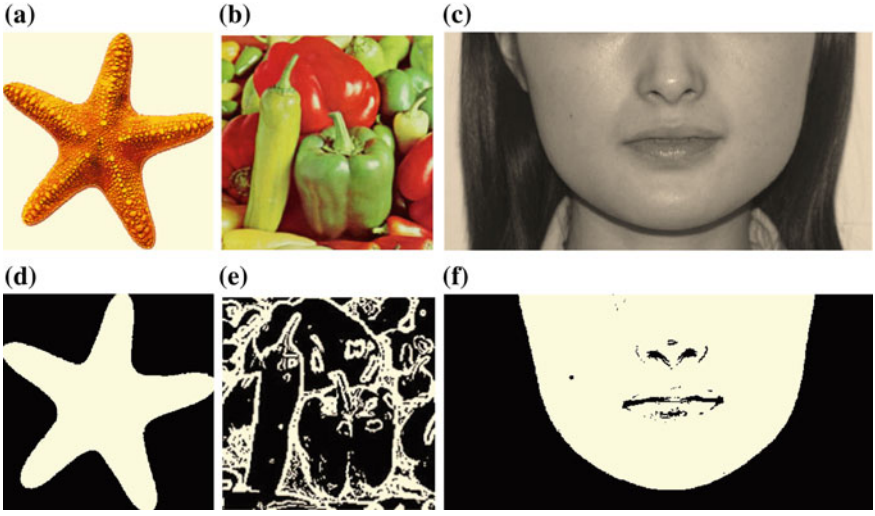


Fig. 50.5 Some image segmentation experiments using our algorithm;(**a**–**c**) are original images; (**d**–**f**) are segmentation images

50.5 Conclusion

This paper presents an adaptive color image segmentation algorithm based on Gaussian mixture model. In the process of image segmentation, we do not have a specified segmentation threshold *t*. Image edge posterior probability density is estimated with Gaussian mixture model from edge magnitude.

From the experimental results, we can take our algorithm results as a prior knowledge, combined with other information segment image accurately.

References

1. Osher S, Sethian JA (1988) Fronts propagating with curvature dependent speed: algorithms based on the Hamilton-Jacobi formulation. J Comput Phys 79:12–49
2. Malladi Ravikanth, Sethian James A, Vemuri Baba C (1995) Shape modeling with front propagation: A level set approach. IEEE Trans Pattern Anal Mach Intell 17:158–175
3. Vese Luminita A, Chan Tony F, Tony, Chan F (2002) A multiphase level set framework for image segmentation using the mumford and shah model. Int J Comput Vision 50:271–293
4. Boykov Yuri, Veksler Olga, Zabih Ramin (2001) Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell 23(11):1222–1239
5. Kolmogorov V (2004) Graph based algorithms for scene reconstruction from two or more views. PHD thesis, USA
6. Vladimir Kolmogorov and Ramin Zabih (2004) What energy functions can be minimized via graph cuts. IEEE Trans Pattern Anal Mach Intell 26:65–81
7. Li F (2005) A bayesian hierarchical model for learning natural scene categories. IEEE Comput Vision Pattern Recogn 2:524–531

Chapter 51
Study and Implementation of Accurate Retrieval System Based on Attractions Interest Model

Ruiqiang Fan, Junping Du and Yipeng Zhou

Abstract With the rapid development of Web 2.0 and social networks, travel has become more and more popular. But it becomes very difficult to filter out the travel information which meets the needs from the mass tourism information on the Internet. In this paper, we implement an information retrieval system based on topic model to solve this issue. First, we construct the attractions interest model by using topic model to analyze large attractions comments. Then we use a highly efficient algorithm to implement topic query expansion for realizing semantic query. On this basis, we propose the method of topic distribution sorting. The retrieval system can return more accurate retrieval results to users by speculating user query intent. Experimental results show that the ranking ability of our accurate retrieval system shows advantages beyond traditional retrieval system.

Keywords Information retrieval · Tourism topic model · Latent dirichlet allocation · Query expansion

R. Fan (✉) · J. Du
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: rqfan@bupt.edu.cn

J. Du
e-mail: junpingdu@126.com

Y. Zhou
School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China
e-mail: yipengzhou@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_51

51.1 Introduction

With the rapid development of Web 2.0 and social networks, travel has become more and more popular. Before traveling, tourists usually prefer to search for relevant tourism information and experiences to gain a general understanding of their planned destination. So it is necessary to study how to retrieve the required information accurately from the mass tourism information on the Internet. Traditional retrieval system cannot meet the retrieval demand of users and understand user query intent because natural language has "polysemy" and "semantic ambiguity" phenomenons [1]. The theory of topic model can effectively solve these problems. LDA can model "polysemy" and "semantic ambiguity" language phenomenons, which allows the retrieval system to get the retrieval results to match user query on the semantic level, rather than just appear as an intersection on a vocabulary level.

Since LDA topic model was proposed, there have been many improved algorithms of LDA for mining various types of information. In tourism domain, there are also some related work such as mining knowledge from travelogues [2], travel package recommendation by mining existing travel packages [3], visualizing textual travelogues by mining location-relevant images [4], etc. However, there are fewer researches on knowledge mining of attractions comments. As the carrier of tourist views, attractions comments express the most impressive sights of tourists. Attractions comments contain what aspects of the attractions are focused on, and represent the interest of attractions in the eyes of tourists. So from attractions comments, we can mine highly valuable knowledge. And attractions comments are highly refined and usually have a clear topic compared with travelogues which have more textual noise. This property is very helpful for constructing high-quality topic model.

In this paper, we construct the attractions interest model by using an improved algorithm of LDA to analyze large attractions comments. Then we use a highly efficient algorithm to implement topic query expansion for realizing semantic query. On this basis, we realize the method of topic distribution sorting. And the retrieval system will return more accurate retrieval results to users by speculating user query intent from the perspective of interest topic distribution. Through the above three steps, we implement our accurate retrieval system.

51.2 Construction of Attractions Interest Model

LDA can effectively mine the topic semantic of long text. But it may not work effectively for topic mining of attractions comments. Attractions comments are usually very short, often containing only one sentence. Besides, attractions

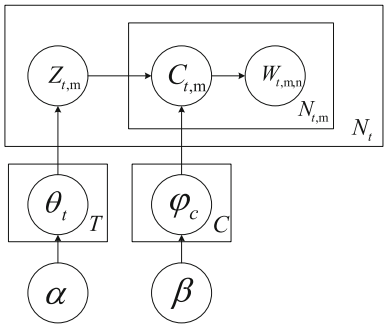


Fig. 51.1 The generation process of each comment

comments usually express only one topic. But in the traditional LDA, each word has an independent topic label, which does not conform to the characteristics of attractions comments. Therefore, LDA is not suitable for mining topic semantic of attractions comments and cannot effectively find out attractions interest in comments.

In this paper, we use Twitter-LDA topic model to deal with the knowledge mining of attractions comments [5]. We remove unnecessary processes of the original algorithm and adjust the topic generation model as to depict the generation process of attractions comments more properly and mine topic information of attractions comments more conveniently. Besides, we regard the comments set of each site as a long document. The generation process of each comment is illustrated in Fig. 51.1. θ_t denotes word distribution of topic and φ_c denotes topic distribution of the comments set.

And φ_c is the attractions interest model which consists of a series of attractions topics. Each topic consists of a series of interest words and each word has its own topic probability value.

After the model training is completed, we will get the interest dictionary, which is composed of large numbers of words, each of which is unique. It records all the interest words that appear in the attractions comments set, such as "culture," "history," "emperor," etc. The interest dictionary denotes what tourists are concerned about, namely attractions interest points.

51.3 Accurate Retrieval System

The overall structure of accurate retrieval system is shown in Fig. 51.2.

The overall retrieval process is as follows. First, the system will preprocess user query input and extract interest points of user query based on the interest dictionary. Second, the system will execute the process of Topic Query Expansion based on these interest points and get the intermediate output subsequently. Besides, based

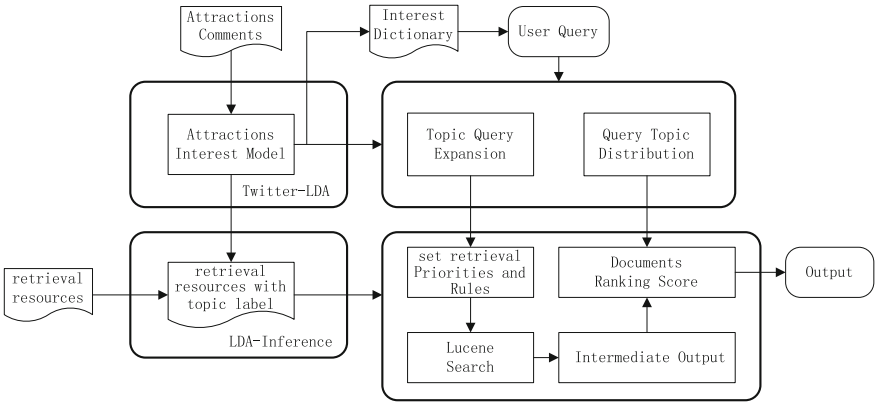


Fig. 51.2 System architecture

on attractions interest model and interest points, system will transform user interest points to the form of topic probability vector to speculate user query intent. Finally, the documents in intermediate output will be reordered through the method of Topic Distribution Sorting.

51.3.1 Topic Query Expansion

In this paper, we realize Query Expansion based on attractions interest model, namely Topic Query Expansion [6]. We can get the probability value of an interest word in each topic through attractions interest model. Words in the same interest topic have semantic correlation with one another. And we can evaluate the semantic correlative degree between interest words by calculating the probability that they appear at the same time. Through the above idea, we can get expansion words of user query, namely the words with highest semantic correlative degrees.

We use $rel_{a,b}$ to denote the semantic correlative degree between word a and word b. In general, the range of semantic correlative degree between words is [0, 1], and the correlative degree between an interest word and itself is defined as 1. Because the size of the interest dictionary is large, the time complexity of the algorithm will be too high if we take traditional multitraverse and sorting to calculate the semantic correlative degree of each word in the dictionary for getting the final set, and retrieval system requires a very high standard of query responding ability. In this paper, we take the strategy of divide and conquer (D&C) to solve the time problem. The description of the algorithm is as follows:

```
1. expSet = { }
2. Foreach z in Z do
3.       heapMin = new Heap[k]
4.       Foreach w in Wz do
5.             rela,w = Mz,a*Mz,w*p(z)
6.             If(rela,w > heapMin.top().rel)
7.                   heapMin.remove()
8.                   heapMin.insert(w)
9.                   heapMin.adjust()
10.            Foreach w in heapMin do
11.                  expSet.put(w, rela,w+preRela,w)
12. Return sorted(expSet, k)
```

Here, a denotes the seed word to be extended, and *expSet* denotes the final words expansion set. k denotes the size of expansion set. Z denotes the topics list. *heapMin* denotes min heap. Mz,q denotes the probability value of word q in topic z. $p(z)$ denotes the probability value of topic z. Because the value of k and the size of *expSet* are both much smaller compared with the size of the interest dictionary, our algorithm not only greatly reduces the time complexity, but also reduces the space complexity.

51.3.2 Topic Distribution Sorting

Each user query implies potential query intent. If the information retrieval system could speculate user query intent based on query words and let the documents which are in line with user query intent rank higher, this would undoubtedly improve system retrieval quality. Based on attractions interest model, interest points can be extracted from user query, and they indicate user query intent obviously.

And we can calculate the probability value of each topic under specific user query as $p(z|q) \propto M_{z,q}$. Here, M denotes attractions interest model and Mz,q means the probability value of word q in topic z. So user query vector can be expressed as $Q = (z_1, z_2, z_3, z_4 \ldots)$. Each retrieval document has its own semantic topic distribution and this distribution represents the content characteristics of this document. Because retrieval documents are usually long text and may express several different topics, we use traditional LDA to execute the process of topic distribution inference to calculate the topic distribution of retrieval documents [7]. The Gibbs Sampling formula of LDA is as follows:

$$p(z_i = k|\overrightarrow{z}_{\neg i}, \overrightarrow{w}) \propto \theta_{m,k} \times \varphi_{k,t} = \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\Sigma_{k=1}^{K}(n_{m,\neg i}^{(t)} + \alpha_k)} \times \frac{n_{k,\neg i}^{(t)} + \beta_t}{\Sigma_{t=1}^{V}(n_{k,\neg i}^{(t)} + \beta_t)} \quad (51.1)$$

We note that in our system, φ is attractions interest model and for retrieval documents, it is unchangeable. What we need to calculate is the topic distribution of retrieval documents. Thus, we modify LDA Gibbs Sampling formula as formula (51.2). Based on this formula, we can use LDA inference to calculate the topic distribution of retrieval documents, and θnew is used to denote it.

$$p(z_i = k|\overrightarrow{z}_{\neg i}, \overrightarrow{w}) \propto \theta_{m,k} \times M_{k,t} \quad (51.2)$$

TopicScore is used to denote the value of measuring the similarity between document topics and user query topics. Its value is $Q \times \theta_{\text{new}}$. Lucene is used to implement our system and Lucene supports setting flexible weights for each query word to influence documents correlation. We note that every expansion word has its own *rel* value which is significantly associated with query priority. Therefore, we make the following assumptions. The query weight of an expansion word is proportional to the *rel* value of the word. In order to adjust all of the words' *rel* value to the appropriate range, we execute the following preprocessing. maxRel denotes the highest *rel* value andσ denotes the expansion weight coefficient. Then the query weight of an expansion word is $(\sigma \times \text{Rel})/\text{maxRel}$, and the upper limit of query weight is set toσ. Besides, the query model of user query is set as "And-Query" and expansion words' are set as "OR-Query." This is to use the combined query model for avoiding query topic offset.

Through the above method, we get the intermediate retrieval results. And in the results, each document has LuceneScore and TopicScore. Because the two types of score have different value ranges, the normalization procedure is applied to them. λ denotes ranking weight coefficient and the final ranking score calculation formula is as formula (51.3). We will get the final retrieval results by reordering the intermediate output based on the value of DocScore.

$$\text{DocScore} = (1 - \lambda) \times \text{LuceneScore} + \lambda \times \text{TopicScore}\,, \ \lambda \in (0, 1) \quad (51.3)$$

51.4 Evaluation and Discussion

We programmed web spider to crawl approximate 110,000 Beijing attractions comments as the data set for constructing attractions interest model and 5000 travelogues related to attractions in Beijing as test retrieval resources from popular tourist web sites such as mafengwo, baidulvyou, xiecheng, etc.

After using *ICTCLAS* to perform word segmentation, we noted that the words which mainly reflected the attractions interest semantic in comments were usually

adjectives and nouns, such as palace, magnificent, etc. Instead, the words with other speech generated much textual noise. Therefore, we only extracted adjectives and nouns in comments and constructed the stop words list to filter meaningless words. Besides, we filtered the words with too high frequency or too low frequency.

We employed *nDCG* to measure the ranking performance of our system. *nDCG* is widely used to measure the ranking quality of IR system. And the relevance of documents was rated on a scale from 0 (lowest relevance) to 4 (highest relevance). We conducted the contrast test with three retrieval methods.The traditional text retrieval method based on Lucene was selected as the method 1 [8], and used L to denote it. Based on method 1, we added Topic Query Extends to the system, and used L + TQE to denote it. Based on method 2, we added Topic Distribution Sorting to the system, and used L + TQE + TDS to denote it.

The number of topics was set to 40, 50, 60, 70, 80, 90, 100 respectively to conduct our test. Under different topics, we set $\alpha = T/50$, $\beta = 0.01$, which was the default settings suggested in [9]. The number of Gibbs Sampling iterations was set to 400 and *BurnTime* was set to 200. The result was sampled every 10 iterations and averaged the 20 results as the final model. We set the size of expansion set to 4, namely $k = 4$. And we set $\sigma = 0.5$, $\lambda = 0.2$ after trying a range of values.

We designed 15 user query examples, and asked 7 graduate students to grade the retrieval results. We recorded the *DCG@10*, *DCG@5* values of each query examples and averaged all the test values as the final test result. Table 51.1 depicts the final average ranking performance of the three methods. Figure 51.3 shows the change curve of ranking performance under different topic numbers.

As shown in Table 51.1, it can be seen that the ranking ability of method 2 (L + TQE) increases significantly compared with method 1(L). This demonstrates

Table 51.1 Comparison on the ranking ability of different methods

	L	L + TQE	L + TQE + TDR
NDCG@5	0.487	0.659	0.738
NDCG@10	0.522	0.681	0.766

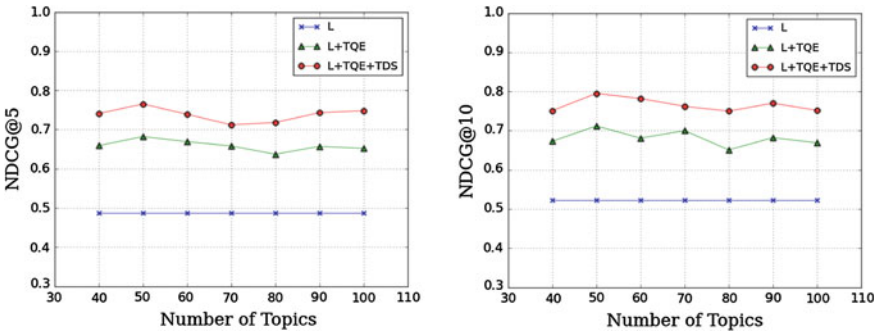


Fig. 51.3 The ranking ability of different methods under different topic numbers

the method of Topic Query Expansion is effective. And method 3(L + TQE + TDS) achieves the best performance, showing advantages beyond the other methods. This demonstrates the method of Topic Distribution Sorting improves the ranking ability of retrieval system effectively. All of the results indicate that the accurate retrieval system based on attractions interest model can improve the retrieval ranking quality significantly.

Besides, we can see that method 2(L + TQE + TDS) and method 3(L + TQE) both get the best performance, respectively, when the number of topics is set to 50, as shown in Fig. 51.3. The reason is that the quality of attractions interest model is the best when topic number is 50. And the system can conduct the process of Topic Query Expansion and Topic Distribution Sorting more effectively. We can see that the quality of attractions interest model is the important factor that affects ranking ability.

51.5 Conclusion

Attractions comments are valuable text resources which contain attractions interest in the eyes of tourists. In this paper, we construct the attractions interest model by analyzing large numbers of comments about the attractions in Beijing based on topic model. Then we use a highly efficient algorithm to realize topic query expansion. On this basis, we propose the method of topic distribution sorting to speculate user query intent. Experimental results show that the ranking ability of our accurate retrieval system shows advantages beyond traditional retrieval system. For the future work, we plan to study how to construct the attractions interest model with high quality more effectively and incorporate personalized user query to our system for understanding user query intent more accurately.

Acknowledgments This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), the National Natural Science Foundation of China (No. 61320106006), Beijing Excellent Talent Founding Project (2013D005003000009).

References

1. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Sci Am 284(5):28–37
2. Hao Q, Cai R, Wang C et al (2010) Equip tourists with knowledge mined from travelogues. In Proceedings of the 19th international conference on World wide web, ACM pp 401–410
3. Liu Q, Chen E, Xiong H et al (2014) A cocktail approach for travel package recommendation. IEEE Trans Knowl Data Eng 26(2):278–293
4. Lu X, Pang Y, Hao Q, et al (2009) Visualizing textual travelogue with location-relevant images. In: Proceedings of the 2009 international workshop on location based social networks, ACM, pp 65–68
5. Zhao X, Jiang J (2011) An empirical comparison of topics in twitter and traditional media. Singapore management university school of information systems technical paper series. 10 Nov 2011

6. Carpineto C, Romano G (2012) A survey of automatic query expansion in information retrieval. ACM Comput Surv (CSUR) 44(1):1
7. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
8. Zhang C, Zhan S (2013) Research and implementation of full-text retrieval system using compass based on lucene. In Proceedings of the 2012 international conference on communication, electronics and automation engineering, Springer, Berlin, pp 349–356
9. Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Acad Sci USA 101 (Suppl 1):5228–5235

Chapter 52
Study on Emergency Anomaly Detection for Tourism Activities

Xiaoyu Han, Junping Du and Yipeng Zhou

Abstract With the development of society and economy, tourism activities become an increasingly important part of people's life, at the same time the emergencies are more frequent than ever before. In this paper, we have studied the past system of emergency anomaly detection for tourism activities, eliminated the limitation of the system, and created many modules about emergency anomaly detection. This paper adopts the method with group of mental perception including the Internet, social networking microblogging, and video data using social force model, implements the functions and examines the effect of each module. Finally, we verify this paper by experimental results. This research provides reliable basis for emergency anomaly detection in the tourism activities, and provides security and convenience for tourists traveling activity.

Keywords Anomaly detection · Emergency · Social force model · Group of mental perception · Social networking

X. Han · J. Du (✉)
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: junpingdu@126.com

X. Han
e-mail: laohanzzz@163.com

Y. Zhou
School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China
e-mail: 2yipengzhou@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_52

52.1 Overview

Tourism is a sunrise industry in the world today, according to the United Nations world tourism organization's latest report; the global tourism industry still keeps growing [1].

With the growing number of tourists and tourism revenue, the sudden mass incidents are also increasing. Crowd is the main reason for abnormal emergency in tourism. Nowadays, the scenic area has taken a variety of methods for emergency anomaly detection, which uses analyzing the crowd density [2–5] monitoring video as the basis of the early warning. Scholar Lapage first put forward the concept of tourism carrying capacity [6]. Sidenbladh and Block Izq from Brown University presented a video monitoring system with human as the goal, using the characteristics of people moving orbit to judge whether an exception occurs [7]. Since the previous methods were restricted to the scene, the effect is not ideal in actual application.

In the design and implementation of this research, we use the crowd density detection of scenic surveillance video combining with abnormal scenario analysis, along with microblogging or other social network real-time state to do emotional analysis. According to the experiment result of the module test, research completed emergency anomaly detection in the scenic spots well.

52.2 Overall Design and Implementation

Several ways are combined in this paper for early warning. Some research modules are shown in Fig. 52.1. The research is mainly divided into four modules: A, Internet history-related event module: This module mainly classifies tourism activities into different types of emergencies based on the similarity, so as to divide new emergencies into the appropriate type according to the relevant description and provide correct decision support based on past experience. B, The scenic spot microblogging sentiment analysis module: This module mainly uses the scenic spot area interface provided by Sina microblogging, extracts microblogging released by visitors, reflects current emotions of the visitors, and thereby provides a basis for early warning. C, Monitor video anomaly detection module: Based on the optical flow method and social force model, analyzing and processing within the scenic area surveillance video, significant changes in a short period of time are able to be detected by this module and as the basis of providing early warning for possible exception. D, Monitor video crowd density detection module: Tourism activities are always along with massive crowds gathered together, and the crowd overcrowding is likely to lead to stampede, trample accidents. In order to inform the scenic area in advance to avoid the happening of abnormal events, the real-time monitor and control area population density is necessary.

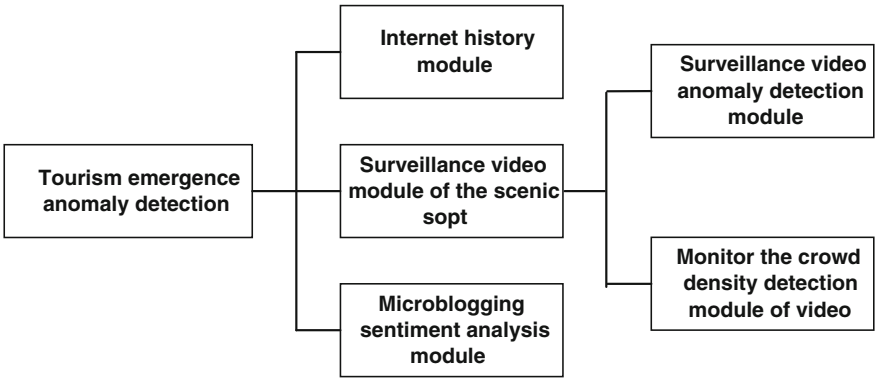


Fig. 52.1 Research modules

52.2.1 Internet History Module

The Internet records many representative travel emergencies. Web portals have described the process and subsequent processing method for all kinds of different types of travel emergencies, which coincides with tourism emergency anomaly detection.

We fetch a lot of emergencies, and extract keywords from description to represent the feature vector of each event type. For new travel incidents, module inputs the text description, then does a series of operations of automatic word segmentation, similarity computation, feature extracting, and finally incident type is decided by the most similar event type.

Crawl the experimental sample first. In the experiment, we crawl 10 typical sudden events in every type. Then we calculate the corresponding vectors of each text as training set, and do the classification experiments using KNN algorithm.

52.2.2 Social Network Microblogging Sentiment Analysis

Microblogging users have the feature of first time reporting hot issues spontaneous, and focusing on the spread trend of emergencies. Therefore, by using weibo location-based service interface and fetching published weibo in the range of target sites, the congestion of scenic spot can be judged according to the number of weibo. If it is above a certain degree of negative, the scenic spot is suggested not suitable for more tourists to enter, so as to achieve the purpose of warning.

Get microblogging data: According to API interface provided by Sina microblogging, this paper selects a specific Beijing scenic spot for test. Module gets the microblogging check-in points based on longitude and latitude of the scenic spot; and extracts weibo released within the scope of each check-in point.

Microblogging sentiment analysis: Module sets the microblogging emotional attributes classified as negative, neutral as positive and extracts keywords about emotions. Then module chooses the algorithms of native bayes and decision tree which perform well to automatically tag microblogging emotion.

Scenic spot warning: This paper gives the warning order to tourists based on degree of negative emotions at unit time period.

52.2.3 Anomaly Detection Under the Surveillance Video

The most popular attractions in scenic have monitor equipments which record the real-time status of tourists. We are able to judge whether an exception occurs based on the movement characteristics of tourist groups.

Data collection: In this module we choose UMN dataset of university of UCF. This dataset includes two parts, normal behavior and abnormal behavior.

This module processes the data set and calculates the flow vector between adjacent frames then turns the scenic video to sub-frames and calculates the changes between adjacent frame images using optical flow. Module selects feature points to calculate the matrix of speed change and transforms the matrix to optical flow vector.

Appling social force model on the optical flow vector: Helbing's social force model of pedestrian movement meets the laws of particle mechanics, uses the force vector to describe the real force pedestrians and intrinsic motivation. Here is Helbing's social force model:

$$m_i \frac{\mathrm{d}v_i}{\mathrm{d}t} = F_a = F_p + F_{\text{int}} \tag{52.1}$$

F_p is a personal spontaneous force, F_{int} is on behalf of the force between pedestrians, v_i represents the actual velocity vector.

$$v_i^q = (1 - p_i)O(x_i, y_i) + p_i O_{\text{ave}}(x_i, y_i) \tag{52.2}$$

$$F_{\text{int}} = \frac{1}{\tau}(v_i^q - v_i) - \frac{\mathrm{d}v_i}{\mathrm{d}_t} \tag{52.3}$$

v_i^q represents the desired forward speed vector, $O_{\text{ave}}(x_i, y_i)$ represents the average speed of (x_i, y_i).

Calculate the interaction between adjacent pedestrian. The method uses bag of words and *K*-means to turn original video frames into pressure diagram.

Analyze the final threshold of different scenarios with LDA to detect anomaly: This method extracts image feature patch using SIFT, clusters all patches using *K*-means and selects feature word. Module uses *K*-means unsupervised to turn the sequence of video frames into two categories, in order to achieve the purpose of anomaly detection warning.

52.2.4 Crowd Density Detection Under the Surveillance Video

In the design of this module, we determine whether to warn or not based on crowd density obtained by surveillance video using a density analysis method to dynamic judge whether the carrying capacity is more than scenic threshold.

Scene background extraction: the training module uses PETS pedestrian dataset offered by UCSD University. First we do binarization processing to frame sequences, average pixel values for each point, then obtain the background scene based on the theory that each point on the scene should be used as the background most of the time.

Moving foreground extraction: Module uses the method of background subtraction that subtract each frame with the background to obtain the foreground moving object. Module should deal with low-density and high-density differently.

Calculating eigenvectors of moving foreground: This module uses gray level co-occurrence matrix to further process the foreground images, respectively, selects angle of 0, 45, 90, 135, a total of four matrixes. On the co-occurrence matrix texture analysis, module calculates energy, entropy, moment of inertia, the correlation between the mean and standard deviation of each picture as the last eight dimensional feature vectors.

52.3 Experiment

52.3.1 Internet Historical Events Module

When conducting experiments to test the accuracy of the classification procedures, the paper selects randomly 80 % records as the training set, then trains the contact of classification between feature word frequency and variables of types using KNN. We put the remaining 20 % as a test set to examine the effectiveness of classification.

For KNN classification algorithm, the performance of algorithm accuracy rate varies when the K value differs. As can be seen from Fig. 52.2 which shows the relationship between the accuracy rate and K values, when K value takes 5, KNN algorithm can obtain the best performance.

52.3.2 Emotional Analysis Module

Sentiment analysis module, which has added microblogging and other social networks and is based on measure of comfort and current mood of scenic tourists, selects different algorithms with same experimental data for processing and reaches a general performance of 65 %. After several experiments, optimized algorithm which is combined by algorithm BFTree and NBTree can further enhance the performance that forecast accuracy reaches 73 %.

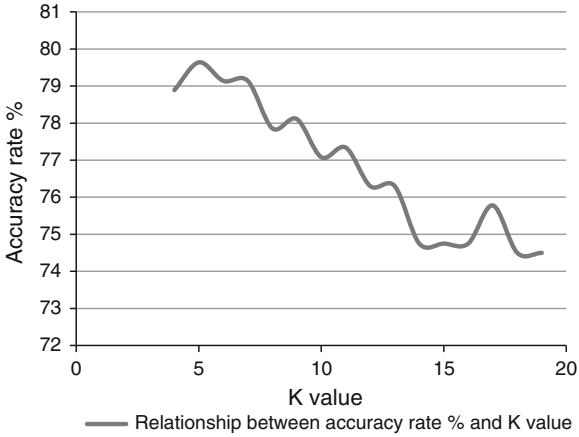


Fig. 52.2 Relationship between accuracy rate and *K* values

52.3.3 Anomaly Detection Module Under Video Surveillance

When dealing with the anomaly detection of target scenic surveillance video, we need to intercept the video frames. We Intercept video frames according to different rates and examine what influence the time interval between testing video frames does to the accuracy of abnormality detection. We set one second interception with one frame or five frames as comparative tests, in which the use of force matrix dimensions with social force model are 20 × 20, 24 × 24, 30 × 30, 32 × 32, 40 × 40, 48 × 48, 60 × 60, 80 × 80, 96 × 96, 120 × 120. Relations, which are affected by the interception of video frames, between force matrix dimensions with social force model and anomaly detection accuracy of tourism activities in the scenic area under video surveillance intervals are shown in Fig. 52.3.

Experimental results show that the time interval between the adjacent video frames is shorter; anomaly detection of the tourism activities in the scenic surveillance video is more accurate. In addition, the effect of the experiment is also affected by the time when abnormal behavior occurs in the test video. Under actual

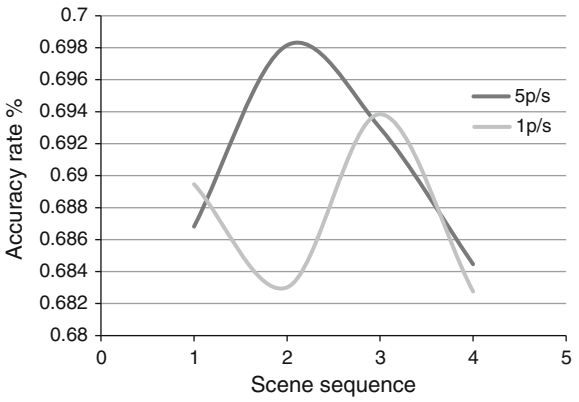


Fig. 52.3 Relationship between accuracy rate and frame interval

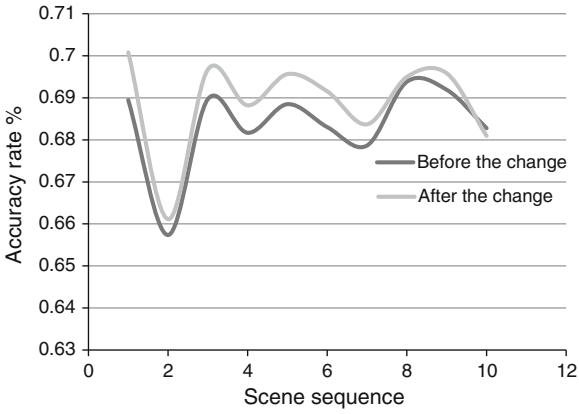


Fig. 52.4 Comparison of before and after modifying the abnormal nodes

conditions, the crowd will tend to calm at the end of video, while changes of mutual social force are not big, should be seen as normal. Therefore, we modify the abnormal node of video sequence, add a controlled trial, and thus get a more realistic test about the performance of anomaly detection function. Changes between before and after the amendments are shown in Fig. 52.4.

As can be seen from the figure, the performance of abnormal detection in the video sequence was significantly optimized when modified the abnormal nodes. The overall performance of anomaly detection of the scenic tourist activities surveillance video can reach accuracy of 70 %. Module can basically meet the needs of system applications; provide a basis for the scenic spot of emergency anomaly detection.

52.3.4 Population Density Detection Under the Surveillance Video

After the surveillance video of the tourist attractions is processed, research further gets texture characteristics vector of each frame and the label situation of population density according to the module function. To synthesize the advantages of various algorithms, experiment selects a number of algorithms to test on the same dataset. In this experiment, the video sequences are selected from two different spots scenarios as different sets of data, we choose five algorithm which are classic and have better performance to test the function. The performance of algorithms is shown in Table 52.1. Each value represents the algorithms' prediction accuracy under different scenarios.

The results can be seen from above table, performance of population density detection module can satisfy basic needs of practical application. Overall performance of each algorithm can achieve 75 % accuracy rate, and the algorithm for detection under different scenarios also maintains a high accuracy rate, indicating that the research modules for different scenarios have well portability.

Table 52.1 The accuracy of the different algorithms in different scenarios

Correct rate	SVM (%)	Logistic (%)	Native bayes (%)	IBk (%)	J48 (%)
Scene 1	75.3143	76.8214	71.5857	74.2429	79.1000
Scene 2	68.5965	80.3289	73.6623	82.6096	83.8596

52.4 Conclusion

This paper analyzes the current situation of tourism activities when there are emergencies occurring, and then completes the overall design and each submodule's detail design of emergence anomaly detection based on the system deficiencies in the past. Research mainly uses the social force model in anomaly detection module and combines the real-time and spread of social networking microblogging, completes the basic functions of each module. Finally, the results show the effect of each module functions and prove the system function is more optimized to supply convenience and security for tourism activities. The next step we will do will be further refinement to solve the jitter of crowd threshold in the video based on the thesis of the functionality of the system.

Acknowledgments This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), the National Natural Science Foundation of China (No. 61320106006). Beijing Excellent Talent Founding Project (2013D005003000009).

References

1. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: 2009 IEEE conference on computer vision and pattern recognition, IEEE, pp 935–942
2. Qin XH, Wang XF, Zhou X, Liu YF, Li YQ (2013) Counting the crowd under a variety of crowd density scenes. J Image Graphics 18(4):392–398 (in Chinese)
3. Yue XJ (2012) The algorithm of crowd density detection in intelligent monitoring system. Shenzhen Graduate School. Accessed Dec 2012 (in Chinese)
4. Xing G, Huang CC (2011) A model for evaluating the carrying capacity of tourism destinations based on grey relational analysis and harmony factors. In: Control decision conference (CCDC), Chinese, IEEE, pp 1768–1771
5. Lin Q, Zhang L (2014) Crowed abnormal detection based on GLCM and optical flow. Comput Mod. doi:10.3969/j.issn.1006-2475.2014.03.027 (in Chinese)
6. Zhou WB, Li R (2010) Pedestrian abnormal behaviour recognize based on trajectory features. Igence Ident Tech. doi:10.3969/j.issn.1006-4052.2010.12.043
7. Chong XY (2012) Crowd behaviour analysis based on visual surveillance data. Beijing Jiaotong University. Accessed June 2012 (in Chinese)

Chapter 53
Fault Diagnostic Method for Photovoltaic Grid Inverter Based on Online Extreme Learning Machine

Pu Yang, Xiao Li, Jiangfan Ni and Jing Zhao

Abstract As the key equipment in photovoltaic system, the operating reliability of photovoltaic grid inverter influences on the security and stability of photovoltaic system deeply. To diagnose the fault of photovoltaic grid inverter effectively, the paper proposes an online extreme learning machine algorithm for fault diagnosis of photovoltaic grid inverter, which uses wavelet analysis method to analyze the output current signal of the inverter. Using extreme learning machine algorithm for online learning, the fault feature of different fault status can be obtained. The proposed algorithm is applied to fault diagnosis of the actual photovoltaic grid inverter, the experimental results demonstrate the effectiveness of the algorithm.

Keywords Wavelet analysis · Online extreme learning machine · Inverter · Fault diagnosis

53.1 Introduction

The grid-connected inverter of the photovoltaic power station feed power to grid, any key component failure of the main circuit will not only make the whole photovoltaic power plant shutdown or damage the equipments, but also reduce the station generating benefit. Therefore, realizing online fault diagnosis of photovoltaic inverter, can not only ensure the photovoltaic power station to supply power normally and prevent serious consequences caused by faults, but also reduce manual maintenance costs and increase the power plant economic benefits.

In recent years, the artificial neural network method applied in fault diagnosis provides a new way [1]. Compared with the traditional method, the artificial neural network method without additional hardware and complex fault model can

P. Yang (✉) · X. Li · J. Ni · J. Zhao
College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China
e-mail: ppyang@nuaa.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_53

make the real-time judgment of fault status and fault type of inverter [2]. And it can put the extracted fault characteristic signal in neural network data, then judge whether fault occurs, thus can realize the fault diagnosis of inverter [3]. Neural network learning can be divided into two ways as offline learning and online learning. But in general, online learning is restricted to factors such as network complexity and slow convergence learning. Traditional learning methods are mostly used in offline mode. But usually, there are time-varying parameters in the actual system or the industrial system, making learning sample unable to meet the difference caused by time-varying system parameters. It will make great difference between the network output and the expected output obtained in offline learning. The main reason is that real-time changeable factors failed to be reflected in network.

This paper presents a fault diagnosis of photovoltaic grid inverter based on a kind of neural network with online extreme learning machine method (ELM) [4], and compares the online and offline diagnosis results. The experimental results show that the proposed method can diagnose fault online effectively for inverter, and has fast convergence speed and good antinoise ability.

53.2 Online ELM Learning Algorithm

Nowadays, there are two kinds of algorithms for feed-forward neural networks: Offline and online learning. Online learning algorithm can be applied in solving real-time problems and can be more suitable for normal industrial application. Compared with general online learning algorithm, learning algorithm of ELM has better performance [5]. We can obtain the output weight in the hidden layer with the premise of randomly selected input weight and neuron thresholds of the hidden layer. Besides, this algorithm has other excellent properties, such as fast learning speed and good generalization capability.

Extreme learning machine is a kind of new feed-forward neural networks with single hidden layer. The structure of ELM is a neural network model with 3 layers, including input layer, hidden layer, and output layer.

The output is expressed as follows:

$$f_N(X) = \sum_{i=1}^{N} \beta_i g(W_i, b_i, X), X \in R^n, W_i \in R^n, \beta_i \in R^m \tag{53.1}$$

where $g(W_i, b_i, X)$ means the output of X via ith neuron in hidden layer, $g(\cdot)$ is the excitation function of hidden layer(usually using Sigmoid function), $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]$ is the weight which connect ith neuron in hidden layer with output neuron.

To solve formula (53.1), we can consider the equation as a special kind of linear regression mode

$$t_j = \sum_{i=1}^{N} \beta_i g(W_i, b_i, X) + e_j = \beta \varphi_j + e_j \tag{53.2}$$

where t_j is the desired output, $\beta = [\beta_1, \beta_2, \ldots, \beta_N]$ is output weight vector, $\varphi_j = [g(W_1, b_1, X), \ldots, g(W_N, b_N, X)]^T$ is regression vector, e_j is set to 0 and its error signal and irrelevant to regression vector.

We use orthogonal least square Givens QR decomposition to modify the output weight of the neural network. Least square algorithm finds weight vector β . The Nth expression

$$V(N, \beta) = \sum_{j=1}^{N} \lambda^{N-j} e_j^2 = \sum_{j=1}^{N} \lambda^{N-j} \left[t_j - \beta \varphi_j\right]^2 \tag{53.3}$$

The weighted mean sum of output errors is the optimal criteria, where $0 < \lambda < 1$ is forgetting factor.

Equation (53.3) can be rewrited as

$$V(N, \beta) = [T(N) - \beta \Phi(N)]^T \Lambda(N) [T(N) - \beta \Phi(N)] \tag{53.4}$$

where $\Lambda(N) = \text{diag}\left[\lambda^{N-1}, \lambda^{N-2}, \ldots,\right] = \text{diag}[\lambda \Lambda(N-1)]$. Minimize β can make a square solution for $\beta(N)$, which satisfies

$$\Phi(N)^T \Lambda(N) \Phi(N) \beta(N) = \Phi(N)^T \Lambda(N) T(N) \tag{53.5}$$

Let $\Lambda^{\frac{1}{2}}(N)$ represents $\Lambda(N)$, then the Eq. (53.5) can be

$$\begin{aligned} &\left(\Lambda^{\frac{1}{2}}(N) \Phi(N)\right)^T \Lambda^{\frac{1}{2}}(N) \Phi(N) \beta(N) \\ &= \left(\Lambda^{\frac{1}{2}}(N) \Phi(N)\right)^T \Lambda^{\frac{1}{2}}(N) T(N) \end{aligned} \tag{53.6}$$

Using QR decomposition

$$\Lambda^{\frac{1}{2}}(N) \Phi(N) = Q(N) R(N) \tag{53.7}$$

where $Q(N)$ is an orthogonal matrix, $R(N)$ is an upper triangular matrix. We can have $R(N)\beta(N) = p(N)$ or

$$\beta(N) = R(N)^{-1} p(N) \tag{53.8}$$

where

$$p(N) = Q(N)^T \Lambda^{\frac{1}{2}}(N) T(N) \tag{53.9}$$

For the sample of neural network, update $\lambda(k) = \lambda_0 \lambda(k-1) - \lambda_0$, where $0 < \lambda(0) < 1$ and $0 < \lambda_0 < 1$. Establish the following matrix

$$\begin{bmatrix} \lambda(k)^{1/2} R(k-1) & \lambda(k)^{1/2} p(k-1) \\ \varphi^T(k) & t(k) \end{bmatrix} \tag{53.10}$$

It can be transformed as $\begin{bmatrix} R(k) & p(k) \\ 0 & \Delta \end{bmatrix}$ with Givens, then calculate p(k) with Eq. (53.).

In conclusion, online ELM using least square Givens QR decomposition can be as follows:

Step 1: Initialization. Set an initial training sample, the sample size is N_0, the excitation function is $g(x)$, and the number of hidden neuron is N. Initialize the learning algorithm as follows:

(1) Assign the weight ω_i and threshold b_i
(2) Calculate the initial output matrix of hidden layer $H_0 = \{h_1, \ldots, h_{N_0}\}$, where $h_i = [g(W_1, b_1, X_i), \ldots, g(W_N, b_N, X_{N_0})]^T, i = 1, \ldots, N_0$
(3) Estimate the initial output weight $\beta(0) = (H_0^T H_0)^{-1} H_0^T H_0$, where $T_0 = [t_1, \ldots, t_{N_0}]$
(4) Set k = 0.

Step 2: Online learning. For the subsequent input sample (x_i, t_i)

(1) Calculate the output vector of hidden layer $h_{k+1} = [g(W_1, b_1, X_i), \ldots, g(W_N, b_N, X_{N_0})]^T$
(2) Calculate the new output weight $\beta(k) = R(k)^{-1} p(k)$ with above algorithm.
(3) k = k+1, back to step 2.

53.3 Wavelet Analysis Theory

For an arbitrary function $f(t) \in L^2(R)$, the wavelet transform is a multiresolution time-frequency analysis method, the expression is

$$WT_f(a,b) = \langle f(t), \varphi_{a,t}(t) \rangle = \frac{1}{\sqrt{a}} \int_R f(t) \overline{\varphi} \left(\frac{t-b}{a} \right) \mathrm{d}t \tag{53.11}$$

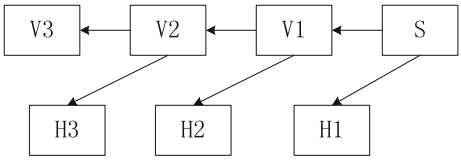


Fig. 53.1 Three layers multiresolution decomposition structure

According to the basic idea of multiresolution analysis, the original signal is decomposed by using the scaling function and wavelet function

$$f(t) = \sum_k c_{j,k} 2^{-j/2} \phi(2^{-j}t - k) + \sum_k d_{j,k} 2^{-j/2} \phi(2^{-j}t - k),\ j,\ k \in Z \tag{53.12}$$

Then the signal $s(t)$ is decomposed into low-frequency part (frequency less than $2j$) and high-frequency part (frequency between $2j$ and $2j + 1$) [6]. As shown in Fig. 53.1, the relationship of decomposition is

$$S = V3 + H3 + H2 + H1 \tag{53.13}$$

where $V3$ is the low-frequency coefficients of signal, $Hi, (i = 1\text{-}3)$ is the high-frequency coefficient of signal.

53.4 Fault Analysis and Diagnosis of Photovoltaic Grid Inverter

53.4.1 Fault Feature Extraction

The steps of signal feature extraction by using wavelet analysis are as follows:

(1) The output current signals are decomposed by wavelet, and then wavelet decomposition coefficient is reconstructed. Signals in different frequency bands are extracted, the reconstructed signal is shown in formula (53.13).
(2) Solution of the total energy of each band signal, as the following formula

$$E_j = \int |S_j(t)|^2 \mathrm{d}(t) = \sum_{k=1}^{n} |Cd_j(k)|^2, (j = 1, 2...4; k = 1, 2...n) \tag{53.14}$$

where $E_0 = \sum_{k=1}^{n} |Ca_3(k)|^2$.

(3) Construct the feature vector. Compared with the energy of extracted low frequency, the extracted high-frequency energy is very small. The high-frequency energy part of the three-phase output current after wavelet decomposition cannot represent fault, so it can be omitted. The low-frequency part retains, and then it will be normalized. The fault feature vector is obtained as follows:

$$T = [E_a, E_b, E_c] \tag{53.15}$$

53.4.2 Fault Classification

The main circuit structure of a three-phase grid-connected inverter is shown in Fig. 53.2. There is a great probability of failure since the inverter power switches work in the high-frequency state. We take the open-circuit fault in the power switches as an example for failure analysis. The main circuit of PV inverter contains six power switches. For this reason, there are a variety of fault combinations. Considering the actual operation of the inverter, we assume that there are two power switches failed simultaneously at most in order to simplify the process of failure analysis. So there are five types of fault in photovoltaic grid inverter as follows:

(1) The inverter is working properly, no fault;
(2) Only one power switch fails, VT1–VT6;
(3) Two power switches of the same bridge arm fail, which can be VT14, VT36, and VT25;
(4) Two power switches in the same half-bridge fail, which can be VT13, VT24, VT35, VT46, VT15, and VT26;
(5) Two crossing power switches between bridge arms fail, which can be VT12, VT23, VT16, VT34, VT56, and VT45;

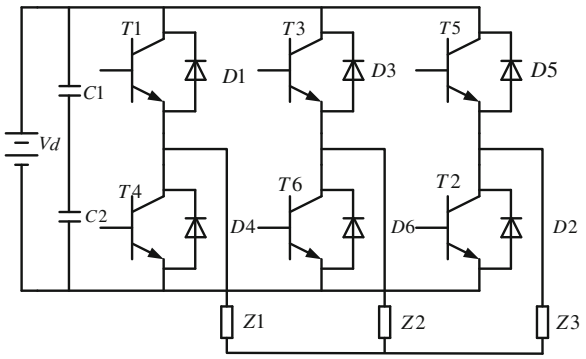


Fig. 53.2 The main circuit structure of a three-phase grid-connected inverter

Table 53.1 Inverter fault code table

Fault sequence	Fault code		Fault mode	Fault sequence	Fault code		Fault mode
	High	Low			High	Low	
1	001	000	Normal	12	100	011	VT46
2	010	010	VT1	13	100	100	VT35
3	010	011	VT4	14	100	101	VT26
4	010	100	VT3	15	100	110	VT15
5	010	101	VT6	16	100	111	VT24
6	010	110	VT5	17	101	010	VT16
7	010	111	VT2	18	101	011	VT34
8	011	010	VT14	19	101	100	VT45
9	011	100	VT36	20	101	101	VT12
10	011	110	VT25	21	101	110	VT23
11	100	010	VT13	22	101	111	VT56

53.4.3 Fault Code

Five categories of 22 kinds of inverter faults are considered in this paper, which can be encoded as the output of neural network classifier, so that neural networks can distinguish all the faults. Encoded with 6 bits as $X_6X_5X_4X_3X_2X_1$, including high-order $X_6X_5X_4$ and low-order $X_3X_2X_1$. High order represents the fault types, a total of five kinds, namely the normal (001), a power switch failure (010), two power switches of the same bridge arm fail (011), two power switches in the same half-bridge fail (100), and two crossing power switches between bridge arms fail (101). Low coding order consists of two parts. The first two bits represent the failed arm, can be 01,10, and 11. The last bit represents the failed original power switch tube, 0 means the top one and 1 means the bottom one. Table 53.1 shows the fault code.

53.5 Experiment Simulation and Result Analysis

The paper gathers a year of the failure data of photovoltaic grid inverter from a photovoltaic power plant in the country, and extracts 880 groups of feature data as the training data. These samples are divided into two parts: 660 groups of data are chosen as learning samples; the other 220 groups of sample data are chosen as the test sample.

To verify the validity of the method, the online learning algorithm based on ELM proposed in the paper and the offline learning algorithm based on wavelet neural network (WNN) are used in this simulation experiment. According to the number of fault features and the fault codes studied in the paper, the network has three input nodes and five output nodes. The number of hidden layer nodes selection is a complex problem, and there is no theoretical basis to follow. If the

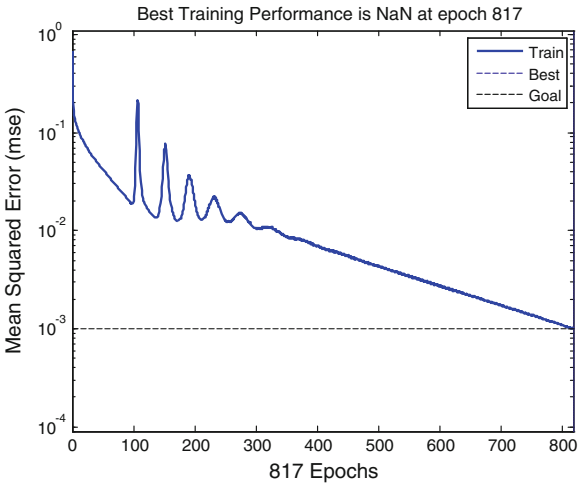


Fig. 53.3 Training error curve of WNN

number is too small, we may get no network from training. If the node number of hidden layer is too large, it will make the learning time too long and the error not necessarily minimal. Through a series of comparative form simulation experiment, ultimately we set the number of hidden layer nodes to 12 [7]. The largest number of network training is 1,000 times, the target mean error is 0.001, the learning rate is 0.5, the simulation uses MATLAB language. The output of two kinds of network and mean error curve of the actual fault classes are shown in Figs. 53.3 and 53.4. We can find that, when the two methods have the same network input and expected output, they are obviously different in the error and convergence time. The WNN meet the accuracy requirement of the error after 817 iterations, but the error curve oscillates. While after 608 iterations, the mean error of online ELM proposed in this

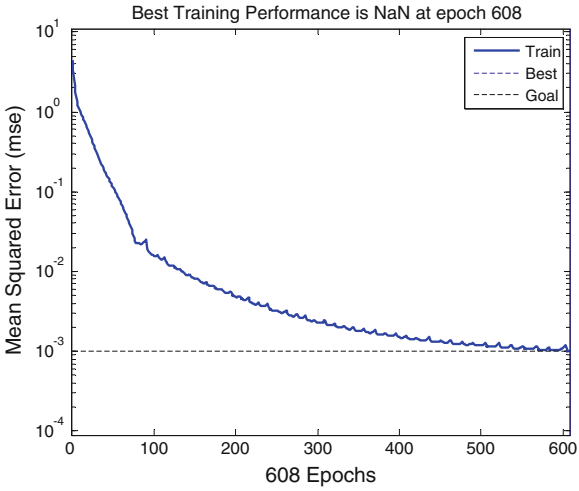


Fig. 53.4 Training error curve of online ELM

Table 53.2 Fault diagnosis results

Fault diagnosis algorithm	Training time	Test time	Accuracy (%)
WNN	5.8310	0.3120	79
Online ELM	1.2373	0.0026	91

paper declines from initial value 1.0010–0.001, the mean error curve descends fast and oscillation is eliminated, which shows a better ability of fault diagnosis.

The trained WNN and online ELM are used for fault diagnosis of the 220 testing samples in the above sample database, the results are shown in Table 53.2.

From Table 53.2, in the 220 samples of diagnosis results, the accuracy rate is generally 80 %, under the same condition, the accuracy of online ELM fault diagnosis (91 %) is higher than WNN (81 %). It illustrates that applying the online ELM to the photovoltaic grid inverter fault diagnosis is accurate and feasible.

53.6 Conclusion

This paper applied online ELM and wavelet neural network to the fault diagnosis of photovoltaic grid inverter. Through training, diagnosis, and analysis of 880 group samples, a conclusion was drawn under the same conditions of network input and expected output. We can see that the online ELM proposed in this paper has better performance in accuracy, convergence time, and stability.

Acknowledgments The work of this paper is supported by the National Natural Science Foundation of China (Grant No. 61203090, No. 61374130) and the Natural Science Foundation of Jiangsu Province of China (No.BK2012384).

References

1. Ruikun G, Liang M, Yanjun Z et al (2011) Fault diagnosis for transformer based on quantum neural network information fusion. Power Syst Prot Control 39(23):79–88 (in Chinese)
2. Xujin Z, Jianbo T, Jianghong H (2002) A method of fault diagnosis based on BP neural network. Syst Eng Theory Pract 6(6):61–66 (in Chinese)
3. Yuanzhang W, Zhihua L, Chunhua W (2013) Fault diagnosis technologies for photovoltaic system. Chin J Power Sources 37(9):1700–1705 (in Chinese)
4. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501 (in Chinese)
5. Li B, Wang J, Li Y, Song Y (2007) An improved on-line sequential learning algorithm for extreme learning machine. Lect Notes Comput Sci 1087–1093 (in Chinese)
6. Charfi F, Sellami F, Al-Haddad K (2006) Fault diagnostic in power system using wavelet transforms and neural networks. In: 2006 IEEE international symposium on industrial electronics, pp 1143–1148
7. Wang SH, Zhao-xia Gao, Yao C et al (2008) Determining the number of BP neural network hidden layer units. J Tianjin Univ Technol 24(5):13–15 (in Chinese)

Chapter 54
An Improved Particle Swarm Optimization Based on Adaptive Mutation and P Systems for Micro-grid Economic Operation

Zhang Sun, Tao Liu, Jun Wang, Juan Luo and Hong Li

Abstract A particle swarm optimization (PSO) algorithm based on adaptive mutation and P systems is proposed to overcome trapping in local optimum solution and low optimization precision in this paper. The algorithm combines the evolutionary rules of PSO, the strategy of adaptive mutation with the hierarchical membrane structure, and communication rules of P systems. At the same time, in order to achieve rapid economic operation and effectiveness of the micro-grid the proposed algorithm is investigated in experiments which are based on the function optimization of micro-grid's economic operation. Furthermore, the feasibility and effectiveness of the proposed algorithm are shown in the experimental results.

Keywords Particle swarm optimization algorithm · P systems · Function optimization · Economic operation

54.1 Introduction

PSO has high optimized efficiency and robustness as an optimum algorithm. It has achieved good results in dealing with nonlinear problems, but the drawback is easy to fall into local optimal solution and has slow convergence. For the drawbacks of PSO algorithm many improved methods [1–3] have been proposed in many literatures. In recent years, several membrane algorithms [4–7] have been proposed in which the basic idea is to take some evolutionary algorithms as a subalgorithm in the membrane, and to fast search out the optimum solutions using the parallel processing ability of membrane. Genetic algorithms, particle swarm optimization algorithms, artificial fish swarm algorithm, and quantum algorithms have been regarded as subalgorithm from the current research results.

Z. Sun (✉) · T. Liu · J. Wang · J. Luo · H. Li
School of Electrical and Information Engineering, Xihua University, No 999 Jinzhou Road, Chengdu, Sichuan, China
e-mail: sunzhang5431@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_54

An improved particle swarm optimization algorithm with adaptive mutation inspired by P systems is proposed. The mutation probability for the current best particle is determined by two factors, including the variance of the population's fitness and the current optimal solution. The diversity of population can be improved by adaptive mutation operation and the ability of PSO to break away from the local optimum is greatly improved by the mutation. Based on the framework of PSO and P systems, the algorithm combines the adaptive mutation's evolutionary rules of PSO with the hierarchical membrane structure and communication rules of P systems. In order to reflect the theoretical and practical value of the algorithm, the micro-grid economic operation experiment is carried out on function optimization problems. The results verify that whether this new algorithm is more effective than the general PSO.

54.2 Improved Particle Swarm Optimization Algorithm Based on P Systems

54.2.1 Basic Algorithm

(1) Particle swarm optimization algorithm
PSO algorithm was originally proposed by Eberhart and Kennedy in 1995. It is an evolutionary algorithm based on swarm intelligence technology, and it studies the movement behavior of birds and fish population.

(2) Adaptive mutation algorithm
Adaptive mutation particle swarm optimization algorithm (AMPSO) is introduced to adaptively adjust the mutation operation according to the particle's position state. Better position may be found by particles under the action of current local optimal value, so the mutation operation is designed as a random number.

(3) P systems
Currently, P systems can be classified into three types: cell-like P systems, tissue-like P systems, and neural-like P systems [8]. It is a hierarchical arrangement of membranes. The one in the outermost layer is called skin that separates P systems from environment. The membrane defines a region. Each region contains a multiple set of objects and transformation or communication rules.

54.2.2 The Improved Particle Swarm Optimization Algorithm

(1) The improvement of inertia weight.
Inertia weight ω plays an important role in the capability of searching local optimum value and global optimization value, as the nonlinear equations in micro-grid economic operation are difficult to be optimized, inertia weight

which is changed with index is introduced to balance the global and local optimal values:

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times (1 - \exp(-20 \times \frac{k^6}{I_1^6}) \tag{54.1}$$

Normally, $\omega_{\max}$ = 0.9, $\omega_{\min}$ = 0.4, k is iteration number, I_1 is the iteration number of each basic membrane.

(2) Adaptive mutation algorithm.

$$p_m = (p_{\min} - p_{\max})(\frac{\sigma^2}{N}) + p_{\max} \tag{54.2}$$

$$p_m = \begin{cases} k, & \sigma^2 < \sigma_d^2 \&\& f(g\text{best}) > f_d \\ 0, & \text{others} \end{cases} \tag{54.3}$$

where p_m is the group mutation probability of group global value. σ^2 is variance of group fitness value, the values of σ^2 associate with the actual problem, and generally far less than the maximum value of σ^2, f_d denotes the optimal value in theory, k is any value between 0.1 and 0.3, $p_{\max}$ and $p_{\min}$ denote the maximum and minimum value of mutation probability.

(3) The improvement of particle initialization.

The basic idea of the improved particle swarm optimization algorithm is to use the monolayer structure of the P system (PSOPS), and perform separately PSO algorithm with adaptive mutation in various membranes (AMPSOPS). And then the best individual of each membrane will be sent to skin membrane by transition rule of P system to guide the algorithm evolution. The updating of velocity and location in improved particle swarm optimization algorithm is shown in Eqs. (54.4)–(54.6). The algorithm determines whether to perform a mutation operation based on population variability capacity (P_{m}), thereby expands the scope of optimization.

$$p\text{bset}_{i,j+1}^{\mathrm{d}} = \begin{cases} p\text{best}_{i,j}^d, f(x_{i,j+1}^d) > f(p\text{best}_{i,j}^d) \\ x_{i,j}^d, f(x_{i,j+1}^d) < f(p\text{best}_{i,j}^d) \end{cases} \tag{54.4}$$

$$\begin{cases} v_i^{k+1} = wv_i^k + c_1 r_1 \left(P_{\text{best}i} - X_i^k\right) + c_2 r_2 \left(P_{\text{Gbest}} - X_i^k\right) \\ X_i^{k+1} = X_i^k + v_i^{k+1} \end{cases} \tag{54.5}$$

$$\begin{cases} v_i^{k+1} = wv_i^k + c_1 r_1 \left(P_{\text{best}i} - X_i^k\right) + c_2 r_2 \left(P_{\text{Gbest}} - X_i^k\right) \\ X_i^{k+1} = X_i^k + v_i^{k+1} \\ \omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \left(1 - \exp(-20 \times \frac{k^6}{I_1^6}\right) \end{cases} \tag{54.6}$$

where, V_i is velocity of the ith particle, X_i is the position the ith particle, P_{best} denotes the group optimal solution of particle, P_{Gbest} denotes the global

optimal value of particle, c_1 and c_2 are the acceleration factors, and the value of them are between 0 and 2, r_1 and r_2 are the random numbers in the range of [0, 1], the diversity of population is kept by these two parameters.

Single-layer membrane structure is applied in the algorithm. Each step is described in detail as follows:

(i) Initialize a one level membrane structure [0[1]1[2]2[3]3,…, [m]m]0, m elementary membrane regions are contained in skin membrane 0.
(ii) Produce a group that contains pop-size individuals, and assign randomly each individual to m elementary membranes where exist at least one individual. At the same time, skin membrane is empty.
(iii) Perform improved particle swarm optimization algorithm in each elementary membrane. (a) Setting the variable parameters, the iterations of basic membrane I_1, the minimum weight and maximum weights $\omega_{\min}$ and $\omega_{\max}$, acceleration factors c_1 and c_2, initialization of particle swarm position, speed, and local optimum value; (b) Calculating the fitness value of each particle; (c) If $f(qi) < f(pi\text{best})$, update the ith individual local optimum value; If $f(pi\text{best}) < f(gi\text{best})$, update the global optimal value of subpopulation; (d) Mutation probability is calculated according to Eqs. (54.3) and (54.4); (e) The position, velocity, and weight of each particle are updated, respectively, according to Eq. (54.5); (f) A random number r is generated in the range [0, 1]. If $r < p_m$, mutation is performed according to Eq. (54.4). Otherwise, operation (iv) is executed.
(iv) The particle with the highest fitness value is sent into skin membrane through transfer rules. Therefore, m particles are stayed in skin membrane. Then select the best individual from the m particles and denote as q. If $f(q) < f(G_{best})$, then update the global optimal value G_{best}.
(v) If the end condition of the algorithm is satisfied, stop running, and output skin membrane result. Otherwise G_{best} return basic membrane to affect the next generation update and return to (iii).

54.3 The Application for Micro-grid Economic Operation Optimization

In this paper, the Micro-grid includes: photovoltaic (PV), wind turbine (WT), micro turbine (MT), fuel cell (FC), and diesel engine (DE). PV and WT are easy influenced by weather, so their outputs are difficult to control. DE generation cost is low, but the pollution is large, etc. Visibly, these DG features are different and operation ways vary widely. So establishing energy management model of each DG is a necessary condition of micro-grid economic operation.

54.3.1 Objective Function

When the micro-grid is in the grid-connected and island mode, the micro-grid objective function of economic and environmental operation can be stated as follows:

$$\min C(P) = \sum_{t=1}^{T}\sum_{i=1}^{3}\{F_i[P(t)] + O_i[P(t)] + C_{\mathrm{DEP}} + \sum_{k=1}^{M}\alpha_k E_k^i[P(t)]\} + \sum_{t=1}^{T}\left[C_b(t)P_{\mathrm{buy}}(t) - C_s(t)P_{\mathrm{sell}}(t)\right] \tag{54.7}$$

$$\min C(P) = \sum_{t=1}^{T}\sum_{i=1}^{3}\{F_i[P(t)] + O_i[P(t)] + C_{\mathrm{DEP}} + \sum_{k=1}^{M}\alpha_k E_k^i[P(t)]\} \tag{54.8}$$

where $P(t)$ is the output power of the MT, FC, and DE groups in the period of t; F_i denotes the fuel costs of the MT, FC, and DE; O_i is the operation and maintenance cost of MT, FC, and DE: α_k denotes the external cost of emitting the kind of k gas; i_{ke} stands for the emission factor of the MT, FC, and DE when the emission type is k; M stands for the type of gas emissions, including CO_2, NO_x, or SO_2; $C_b(t)$ and $Cs(t)$, respectively, purchase price and sale price at time t; $P_{\mathrm{buy}}(t)$ stands for the number of purchasing power; $P_{\mathrm{sell}}(t)$ stands for the number of sale electricity; T stands for the total number of hours in optimizing the cycle. C_{DEP} is the depreciation cost of equipment.

54.3.2 Constraint Conditions

The constraint conditions are shown in formula (54.9). Where p_{it} is controlled power, the p_{bt} is purchase power, the p_{st} is sell power, and the p_{loadt} is load power. And also the power must be within a certain range.

$$\sum_{i=1}^{N}(P_{\mathrm{it}}) + P_{\mathrm{bt}} - P_{\mathrm{ts}} = P_{\mathrm{loadt}} \tag{54.9}$$

54.4 Simulation Experiments and Discussion

54.4.1 Test System

The related parameter values of the PSO are stated as follows: the size of the particle population is taken as 30, the maximum iterations number of the basic film

Table 54.1 The composition of DG subsystem in micro-grid

Type of DG	Coefficient of operation (RMB/kWh)	Lower limit (kW)	Upper limit (kW)
PV	0.009	0	5
WT	0.029	0	10
FC	0.042	4	30
DE	0.083	5	65
MT	0.081	0	30

Table 54.2 The demand of micro-grid load

t	1	2	3	4	5	6	7	8	9	10	11	12
Pl/kW	8.4	8.4	8.4	8.4	8.4	8.4	8.4	84.5	84.5	84.5	84.5	81.2
t	13	14	15	16	17	18	19	20	21	22	23	24
Pl/kW	68.4	60.6	81.2	81.2	84.5	38.2	8.4	8.4	8.4	8.4	8.4	8.4

Table 54.3 Real-time electricity price

t	1	2	3	4	5	6	7	8
RMB/kWh	0.2294	0.1692	0.1243	0.0926	0.0287	0.1626	0.259	0.3693
t	9	10	11	12	13	14	15	16
RMB/kWh	0.4932	0.5028	0.7742	0.9558	0.9462	1.4241	0.9462	0.7551
t	17	18	19	20	21	22	23	24
RMB/kWh	0.3823	0.3486	0.3427	0.3948	0.4251	0.3326	0.2867	0.2151

as 5, the maximum iterations number of the outer membrane as 50, c_1 as 2, c_2 as 2. The electricity price of purchase and sale in this paper is taken at the 0.83 RMB of the peak, 0.49 RMB as the usual, 0.17 RMB of the valley. The data of DG subsystems in micro-grid are shown in Table 54.1. The micro-grid load demand is shown in Table 54.2. The real-time electricity price is shown in Table 54.3.

54.4.2 Example Analysis

In this paper, generation power of DG, number of demand load, and electrical price at time $t = 10$ are considered as example.

After the simulation of three different algorithms, we can get the results as follows: the best individual fitness value (the final overall cost) is 66.5898 RMB, the power of the fuel cell is 10 kW, the power of the micro turbine is 5 kW, the power of diesel generator is 0 kW, and the power of main grid is 68.4 kW.

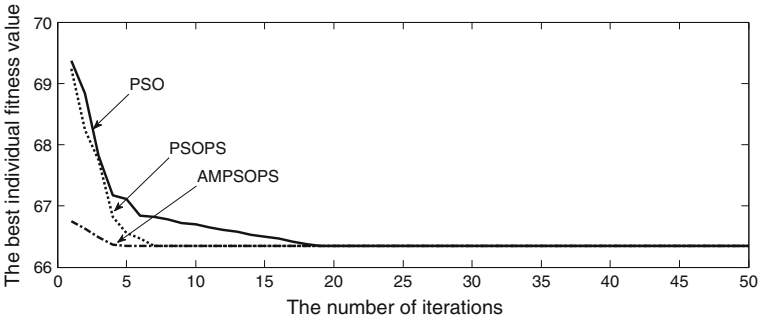


Fig. 54.1 Best individual fitness values corresponding to each algorithm

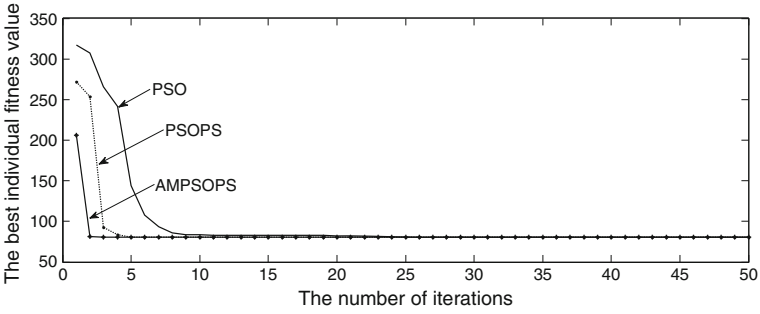


Fig. 54.2 Best individual fitness values corresponding to each algorithm

As can be seen from Fig. 54.1, the optimization capabilities of PSO, PSOPS, and AMPSOPS algorithms are the same. From the simulation curves, the algorithm which is based on P system has faster convergence relative to others, reflecting the ability of high-speed parallel computing of P system. The convergence of AMPSOPS is better than PSOPS through the comparison between PSOPS and AMPSOPS.

After the simulation of three different algorithms, we can get the results as follows: the best individual fitness value (the final overall cost) is 80.8928 RMB, the power of the fuel cell is 10 kW, the power of the micro turbine is 65 kW, the power of diesel generator is 8.4 kW, and the power of main grid is 0 kW.

As can be seen from Fig. 54.2, the optimization capabilities of PSO, PSOPS, and AMPSOPS algorithms are the same. The AMPSOPS algorithm which is based on P system and adaptive mutation operator has faster convergence than others, reflecting the rapidity of algorithm, so it can well satisfy the need of micro-grid real-time scheduling. Contrasting the Figs. 54.1 and 54.2, we can see that the final overall cost of micro-grid in the islanded mode is higher than it is in grid-connected mode.

54.5 Conclusions

In this paper, the objective function is determined by considering the characteristics of micro-grid system and requirements of environment and economic operation. Through comparing the simulation results, it can be seen that the optimization capability of AMPSOPS algorithm is stronger than that of PSO or PSOPS. Via comparing the simulation curves, the stability, optimization, and convergence are all improved after combining the P systems with AMPSO algorithm. Furthermore, it can also be seen that the shortcoming of PSO algorithm such as randomness, trapped in local optimum value are both improved in this algorithm in a certain extent. The optimization and improvement of the algorithm should be further studied in future works.

Acknowledgment This work was partially supported by the National Natural Science Foundation of China (Grant No. 61472328), and the Innovation Fund of Postgraduate, Xihua University (ycjj2014147).

References

1. Zhang TC, Geng GF (2012) Reactive power optimization for medium voltage distribution network based on improved particle swarm optimization. Power Syst Technol 36(2):158–162
2. Chen QY, Chen WH, Dai CH, Zhang XX (2014) Reactive power optimization based on modified particle swarm optimization algorithm for power system. Proc CSU-EPSA 26(2):8–13
3. Zhan ZH, Zhang J, Li Y, Chung HSH (2009) Adaptive particle swarm optimization. IEEE Trans Syst Man Cybern PART B-Cybern 39(6):1362–1381
4. Huang L, Wang N (2006) An optimization algorithm inspired by membrane computing. In: Proceedings of ICNC, vol 4222, pp 49–52
5. Wang T, Wang J, Peng H, Tu M (2012) Optimization of PID controller parameters based on PSOPS algorithm. ICIC Express Lett 6(1):273–280
6. Tu M, Wang J, Song XX, Yang F (2013) An artificial fish Swarm algorithm based on P Systems. ICIC Express Lett 4(3):747–753
7. Zhang GX, Gheorghe M, Wu CZ (2008) A Quantum-Inspired evolutionary algorithm based on P systems for a class of combinatorial optimization. Fundam Informaticae 87:93–116
8. Wang J, Li Z, Peng H, Zhang GX (2011) An extended spiking neural P system for fuzzy knowledge representation. Int J Innovative Comput Inform Control 7(7A):3709–3724

Chapter 55
Chinese Character Recognition Based on Energy Value of the Dual Peripheral Coordinates

Li Yuan, Tian Wang, Zhiyan Li and Wei Liu

Abstract Feature extraction is the most important part in the process of Chinese character recognition (CCR). When there are such factors as deformity, tilt, uneven illumination existing on the text images, conventional recognition methods such as cellular feature or the linear density characteristics show a degrade on recognition performance. In this paper, we propose a new feature extraction method for CCR based on the energy value of the dual peripheral coordinates. During registration, training samples of the Chinese characters are transformed into binary images. Then the character feature vector is extracted by calculating the square of the coordinates of the peripheral points. During recognition, test samples of Chinese characters are isolated from text images by the steps of image correction, layout analysis, and image segmentation, then the feature vector of dual energy value of the peripheral coordinates is extracted. Character recognition is achieved by searching the minimum Euclidean distance between the testing samples features and the registered training samples features. Experimental results on scanned images and photo images show the effectiveness and accuracy of the feature of dual peripheral coordinate energy value approached in this paper.

Keywords Chinese character recognition · Character feature extraction · Euclidean distance

55.1 Introduction

Optical character recognition of Chinese characters is a typical problem of pattern recognition. Chinese characters are of widely varying complexity, consisting of radical frequency, position-specific radical, one to more distinct strokes. The most

L. Yuan (✉) · T. Wang · Z. Li · W. Liu
School of Automation and Electrical Engineering, University of Science and Technology Beijing, 30 Xueyuan Road, Haidian District 100083, Beijing, China
e-mail: lyuan@ustb.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_55

important influencing factors are various similar radicals and stroke components [1]. And the quality of feature extraction directly affects the accuracy of classification and identification.

Generally, a Chinese character recognition (CCR) system contains the following four parts: layout analysis, image rectification, Chinese character segmentation, and character feature extraction and recognition. Layout analysis aims to divide the contents of the text images into specific blocks. Representative methods are categorized into top-down type or bottom-up type [2]. Image rectification applies tilt correction or distortion correction on text images. Representative methods are rectangle block-based method [3] or Hough transformation-based method [4]. Character segmentation aims to isolate individual Chinese characters to be recognized. Connected components analysis and projection analysis are commonly applied for the row segmentation and column segmentation. Chinese character feature extraction can be divided into two kinds: statistical feature extraction and structural feature extraction. Statistical feature extraction methods take Chinese characters as images and use subspace or kernel-based approaches, neural networks or Wavelet transformation tools for feature extraction [5–7]. The structural features are often represented by the spatial relationship between strokes or radicals, which can describe the essential characteristics of Chinese characters. Commonly used structural features are stroke numbers, stroke direction, stroke positions and sequence, basic skeleton, connected components, closed regions, coarse periphery, cellular feature, text feature, etc. [8]. A good recognition performance is based on a good quality of the scanned text images to a great extent. When the text source images are of high resolution and low level of noise pollution, the structural feature-based methods will get good performance. But in real applications, the text images taken under different circumstances will be contaminated by various factors like illumination variation, camera conditions, and parameters. The text images might posses tilt, distortion, or uneven illumination. In such case, the scanned image-based feature extraction methods will not work very well.

In order to obtain further improvement on the CCR efficiency and accuracy, a new Chinese character feature extraction method based on the energy value of dual peripheral coordinates is proposed in this paper. Dual peripheral coordinates mean the coordinates of the first point encountered from the four directions and the coordinates of the second point whose gray value changed from white to black. The energy value means the square root of the dual peripheral coordinates. Figure 55.1 shows a system diagram of the proposed CCR method. During registration,

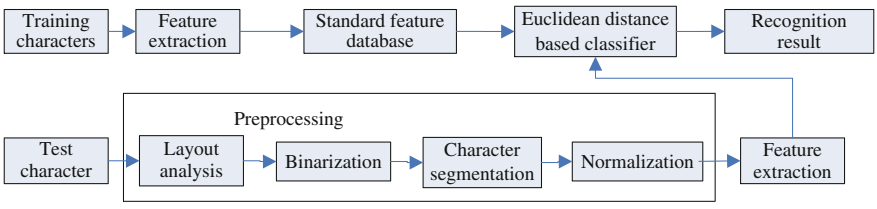


Fig. 55.1 System diagram of the CCR process

we extract character features of training samples of a standard Chinese character set, and store the feature vectors in a database. During recognition stage, we extract character features of unknown Chinese characters and then match them against those feature vectors of the standard CCs stored in the database one by one. The character corresponding to the minimum Euclidean distance in the standard CC database is the recognized character.

The rest of the paper is organized as follows: Sect. 55.2 introduces the principle of the proposed feature extraction method. Section 55.3 details the character recognition experimental results and comparison with other methods. Section 55.4 concludes the paper.

55.2 Feature Extraction Based on Energy Value of the Dual Peripheral Coordinates

All of training CCs are scanned or photographed and transformed into images. These CC images may contain some noise interference during image acquisition process which will reduce the quality and affect the recognition rate of the whole CCR system. Therefore, we need to preprocess them before recognition to enhance image quality. Generally, the preprocessing of printed CCR consists of image filtering and enhancement and binarization. The binarization process can transform all the CC samples into matrices that only contain 0 and 1 values. The CC is in black color and the background is in white color. The size of all CC images are normalized to 24 × 24. We define four directions to each character image from the surrounding including upper, lower, left, and right directions, denoted as U, Lo, L, and R, respectively.

Our proposed Chinese character feature extraction method is based on the energy value of dual peripheral coordinates. Here, dual peripheral coordinates mean the coordinates of the first two points encountered from the four directions whose gray value changed from white to black as shown in Fig. 55.2. In Fig. 55.2, $P1$ is the first peripheral point, $P2$ is the second peripheral point. The coordinates are denoted as (a_1, b_1) and (a_2, b_2), respectively. If the second peripheral point does not exist, then replace it with the final pixel in the same row or column.

The energy value is defined as the square root of the dual peripheral coordinates as shown in Eqs. 55.1 and 55.2.

$$\text{Ener } 1 = \sqrt{(a_1^2 + b_1^2)} \tag{55.1}$$

$$\text{Ener } 2 = \sqrt{(a_2^2 + b_2^2)} \tag{55.2}$$

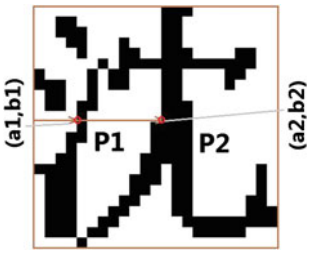


Fig. 55.2 The coordinate of the pixels changes from *white* to *black*

For the four directions U, Lo, L, and R, we locate the dual peripheral coordinates and calculate their corresponding energy values. The final feature vector of a Chinese character is defined as:

$$E = [E1_{ij}, E2_{ij}], \quad i = 1, \ldots, 4, \; j = 1, \ldots, 24 \tag{55.3}$$

in which $E1_{ij}(i = 1, \ldots, 4, \; j = 1, \ldots, 24)$ denotes the energy values of the first peripheral points, its dimension is 24×4, $E2_{ij}(i = 1, \ldots, 4, \; j = 1, \ldots, 24)$ denotes the energy values of the second peripheral points, its dimension is 24×4, i represents the four directions, j is the index on each direction. The total dimension of the feature vector is $24 \times 4 \times 2$.

In order to further reduce the feature dimension, we regroup the energy values on each direction by calculating the average of three adjacent energy values in the same direction as shown in formulas (55.4)–(55.6).

$$\begin{aligned}
E1_{ij} &\rightarrow E1_{i1}, E1_{i2}, E1_{i3}, \ldots, E1_{i22}, E1_{i23}, E1_{i24} \\
E1^*_{ij^*} &\rightarrow \quad E1^*_{i1^*} \quad \ldots \quad E1^*_{i8^*} \\
E2_{ij} &\rightarrow \underbrace{E2_{i1}, E2_{i2}, E2_{i3}}, \ldots, \underbrace{E2_{i22}, E2_{i23}, E2_{i24}} \\
E2^*_{ij^*} &\rightarrow \quad E2^*_{i1^*} \quad \ldots \quad E2^*_{i8^*} \\
E1^*_{ij^*} &= \sum_{j^*=1}^{8} \left(E1_{ij} + E1_{i(j+1)} + E1_{i(j+2)} \right) \\
E2^*_{ij^*} &= \sum_{j^*=1}^{8} \left(E2_{ij} + E2_{i(j+1)} + E2_{i(j+2)} \right)
\end{aligned} \tag{55.4}$$

$$i = 1, \ldots, 4, j^* = 1, \ldots, 8, j = 3 * (j^* - 1) + 1 \tag{55.5}$$

The final feature vector of each Chinese character is then refreshed as:

$$E^* = [E1^*_{ij^*}, E2^*_{ij^*}],\ i = 1, \ldots, 4, j^* = 1, \ldots, 8 \tag{55.6}$$

Here, $E1^*_{ij^*}(i = 1, \ldots, 4, j^* = 1, \ldots, 8)$ denotes the energy value of the first peripheral point, its dimension is 8 * 4. $E2^*_{ij^*}(i = 1, \ldots, 4, j^* = 1, \ldots, 8)$ denotes the energy value of the second peripheral point, its dimension is 8 * 4. The dimension of E^* will be $8 \times 4 \times 2$.
For character recognition, we apply the Euclidean distance-based classifier. The Euclidean distance D_l between a testing sample feature vector and a training sample feature vector is defined as:

$$D_l = \sum_{i=1,2,\ldots,n_i, j=1,2,\ldots,n_j} \left| ET_{i,j} - EL^l_{i,j} \right| \tag{55.7}$$

in which $ET_{i,j} = [ET\_1_{i,j}, ET\_2_{i,j}]$ is the feature vector of a testing sample; $EL_{i,j} = [EL\_1_{i,j}, EL\_2_{i,j}]$ is the feature vector of a training sample; l is the index of the training sample; i is the four directions; j is the feature dimension of each direction, and $n_i \times n_j$ is the feature dimension, here, $n_i = 4$, $n_j = 8$.

The classification function is defined as:

$$ET \in C_k \quad \text{if} \quad D_k = \min_{l=1,\cdots,M} \sum_{i=1,2,\ldots,n_i, j=1,2,\ldots,n_j} \left| ET_{i,j} - EL^l_{i,j} \right| \tag{55.8}$$

in which M is total number of training samples.

55.3 Experimental Analysis

55.3.1 Experiment Description

The main steps of the character recognition stage are: layout analysis, image binarization and tilt correction, character segmentation.

55.3.1.1 Layout Analysis

Layout analysis aims to divide the contents of the text images into specific blocks. Here, we use the method of connected component labeling. We apply image corrosion with a circle of radius 60 pixels, and locate and mark the connected components on the image. Figure 55.3a shows a simple example of the layout analysis.

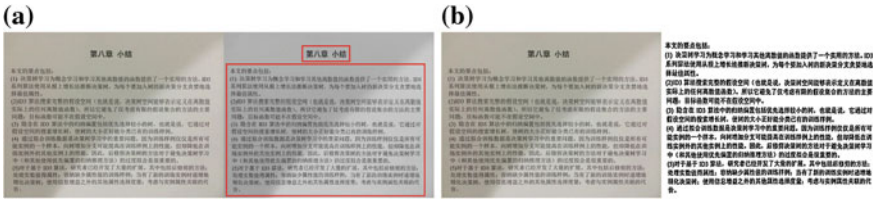


Fig. 55.3 Results of layout analysis and image binarization. **a** Result of layout analysis. **b** Result of image binarization

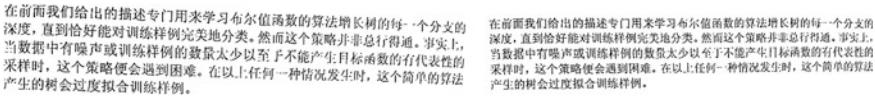


Fig. 55.4 Result of tilt correction

55.3.1.2 Image Binarization and Tilt Correction

Here, we apply adaptive Otsu [9] approach for image binarization. This approach is effective in the case of uneven illumination on the text images. Figure 55.3b shows the result of the binarization.

For tilt correction, we first figure out the tilt angle by locating the uppermost point of the leftmost CC and rightmost CC, and calculating the slope of the straight line determined by these two points. Then tilt the image according to the angle of inclination determined by the slope. Figure 55.4 shows an example of the tilt correction.

55.3.1.3 Character Segmentation

Character segmentation aims to isolate individual Chinese characters to be recognized. There are two steps involved: row segmentation and column segmentation. For row segmentation, project the binarized text image on the horizontal direction, and collect a histogram of the projection. Then segment each row according to the projection gap as shown in Fig. 55.5a. For column segmentation, project each row text image vertically, and collect a histogram of the projection. Then segment each CC according to the projection gap mark with red dots as shown in Fig. 55.5b.

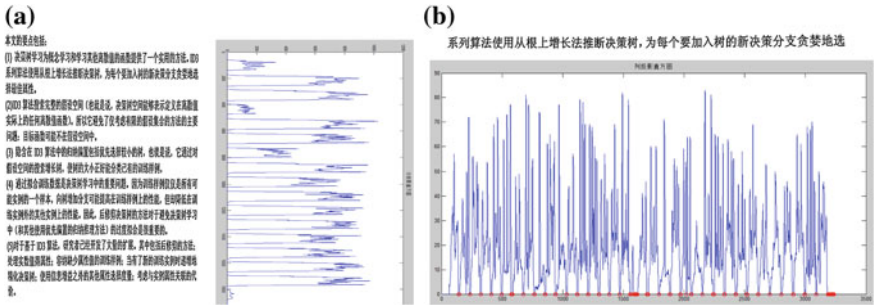


Fig. 55.5 Example of character segmentation. **a** Histogram of the *horizontal* projection. **b** Histogram of the *vertical* projection

Table 55.1 Comparison with other CCR methods

Method	Time (s/100 CCs)	Recognition rate (%)
Our method	8.18s	93.7
Cellular feature [10]	12.7s	65.0
Linear density [11]	16.84s	67.9

55.3.2 Experimental Results

Our experiments are based on a training dataset which contains 3500 Chinese characters of the GB level-1charset, 1 sample per character. Testing CCs were divided into two parts. The first one is the scanned image dataset which contains 3500 Chinese characters of the GB level-1charset, 1 sample per character. The second dataset comes from text images taken by cell phone which contains 1100 Chinese characters of the GB level-1charset, 1 sample per character. Every CC is transformed into binary image, and is normalized to the size of 24 × 24.

On the test dataset1, we use 3500 CC segmented from scanned images, 1 test sample per CC. The rank-1 accuracy is 98.1 %.

On the test dataset2, we use 1100 CC segmented from images taken by cell phone, 1 test sample per CC. The rank-1 accuracy of testing phone images is 93.7 %. Table 55.1 shows the comparison of our method with other two popular feature extraction methods. We can see that our method outperforms the other two methods.

55.4 Conclusion and Further Work

In this paper, we discussed the Chinese character recognition principle and process and proposed a new feature extraction method for Chinese character recognition. This feature is based on the energy value of the dual peripheral coordinates from the

four directions (including upper, lower, left, and right) of a Chinese character image. It is effective even when part of the character information is missed after image binarization. This kind of feature is a combination of structural feature and statistical feature. Experimental results on scanned images and photo images show the effectiveness and accuracy of the proposed feature extraction method. For future research, we will extend our experiments on different Chinese character fonts and optimize the threshold function in the image binarization process.

Acknowledgments This work was supported by the National Natural Science Foundation of China, Project Grant No.: 61300075.

References

1. Wu Y, Mo D, Tsang YK, Chen HC (2012) ERPs reveal sub-lexical processing in Chinese character recognition. Neurosci Lett 514:164–168
2. Chen M, Ding XQ (2001) Analysis, understanding and representation of Chinese newspapers with complex layout. J Tsinghua Univ 41(1):29–32
3. Kwag HK, Kim SH, Jeong SH, Lee GS (2002) Efficient skew estimation and correction algorithm for document images. Image Vis Comput 20:25–35
4. Yang QU, Yang LP (2001) Hough transform OCR image slant correction method. J Image Graph 6(2):178–182 (in Chinese)
5. Tao DP, Liang LY, Jin LW, Gao Y (2012). Similar handwritten Chinese character recognition by kernel discriminative locality alignment. Pattern Recogn Lett 186–194
6. Wei W, Liu M, Gao WN, Wang DD, Liu J (2012). A new kind of wavelet transform for handwritten chinese character recognition. In: Proceedings of the 2012 2nd international conference on instrumentation and measurement, computer, communication and control, IMCCC, pp 1223–1226
7. Du J, Huo Q (2014) An irrelevant variability normalization approach to discriminative training of multi-prototype based classifiers and its applications for online handwritten chinese character recognition. Pattern Recogn 47:3959–3966
8. Nie JX (2009) Research on feature extraction and matching recognition of printed chinese character recognition system. Master thesis, Dalian University of Technology (in Chinese)
9. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst 9 (1):62–66
10. Fu T (2010) Research based on high-order neutral network in character recognition. Northeast Normal University, China (in Chinese)
11. He YX (2011) Chinese character recognition of printed document images. Yanshan University, China (in Chinese)

Chapter 56
Quantum Particle Swarm Optimization Based on P Systems for Applications in the Economic Operation of Micro-grid

Tingting He, Jun Wang, Zhang Sun and Tao Liu

Abstract The strategies of economy micro-grid are made based on the mathematical model of wind turbines (WT), photovoltaic (PV), fuel cells (FC), micro gas turbine (MT), and other types of distributed power. And the optimization model is built by considering the costs of power generation, pollution control, operation and maintenance, and so on. The quantum particle swarm optimization which is based on P system (QPSOPS) is proposed in this paper according to the model and it is implemented using MATLAB programming. Not only the conversion operations from serial to parallel are achieved, but also the convergence speed and accuracy are improved in the algorithm. The results of QPSOPS and QPSO algorithm are compared in this paper through a micro-grid example. The results show that the QPSOPS algorithm has advantages of high accuracy, fast convergence, which can reduce the running costs of micro-grid and realize the economic operation of micro-grid.

Keywords Micro-grid · Economicm operation · P systems · QPSOPS (quantum particle swarm optimization)

56.1 Introduction

Micro-grid as an important future direction of development of the smart grid, its economy operation attracted the attention of many scholars in recent years. Currently, swarm intelligence optimization algorithms such as: artificial fish swarm algorithm (AFSA), particle swarm optimization (PSO), and so on are hotter, they are applied to problems in economic operation of micro-grid by some experts and have achieved a better optimization results. Although AFSA is easy to jump out of local extreme, its optimizing accuracy is not high and has long search time [1].

T. He (✉) · J. Wang · Z. Sun · T. Liu
School of Electrical and Information Engineering, Xihua University, No 999 JinZhou Road, Chengdu City, JinNiu, China
e-mail: 864662987@qq.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_56

And these shortcomings can be overcome by the PSO algorithm [2]. But in the practical application, the PSO algorithm is the same with other optimization algorithms which may produce premature convergence easily, has poor global search capability and slow convergence speed. In order to solve these problems, quantum particle swarm optimization (QPSO) is proposed by some experts who combine quantum searching mechanism in quantum algorithm with particle swarm [3]. And experimental results also prove that QPSO algorithm with a stronger ability to optimizing and better optimization efficiency, it also has some shortcomings [3]. For example, a long running time, optimal results are also need to be improved, etc.

However, there are several advantages in P systems which are suitable for solving optimization problems, for example, distribution, parallelism, easy program implementation, and so on [4]. In particular, the advantages of parallelism and easy programming if applied to the QPSO algorithms must be able to improve QPSO algorithm's drawbacks which are mentioned above to some extent. Based on this, an algorithm is proposed which is a combined P system with quantum particle swarm algorithm in this paper. The algorithm not only can make more rapid convergence, but can also make the convergence results more in line with the requirements.

56.2 Establishment of Objective Function of Micro-grid

56.2.1 Generation Cost of the Distributed Power

(1) The cost of PV, WT will not be considered, idealized:

$$C_{\mathrm{GER(PV,WT)}} = 0 \tag{56.1}$$

(2) The cost of FC [5]:

$$C_{\mathrm{GER(FC)}} = C_{\mathrm{nl}} \times \frac{1}{\mathrm{LHV}} \sum_{0}^{J} \frac{P_J}{\eta_j} \tag{56.2}$$

In the formula, the price of the natural gas is 2.05 yuan/m^3, LHV is 9.7 kWh/m^3, P_J stands for the output power in period J, η_J is the total efficiency of the battery within the time interval J.

(3) The cost of MT [6]:

Notes: Any fuel cell efficiency is the ratio of electrical power output and the input of fuel and both of them are in the same units. The cost of MT can be calculated as follows:

$$C_{\mathrm{GER(MT)}} = C_{\mathrm{nl}} \times \frac{1}{\mathrm{LHV}} \sum_{0}^{J} \frac{P_J}{\eta_j} \tag{56.3}$$

(4) The cost of DG [7]:

The cost of fuel power system can be expressed by the actual power output with a quadratic polynomial model as follows:

$$C_{\mathrm{GER(DE)}} = \sum_{i=1}^{N} (d_i + e_i P_{\mathrm{DE},i} + f_i P_{\mathrm{DE},i}^2) \tag{56.4}$$

In the formula, N stands for the number of DE, d_i, e_i, f_i, respectively, represents the coefficient of generators usually provided by the manufacturer. $P_{\mathrm{DE},i}$ stands for the output power of diesel generators.

56.2.2 Cost of Buying or Selling Electricity from Smart Grid

$$C_{\mathrm{ELE}} = \sum_{t=1}^{T} (A_{\mathrm{buy}} P_b - B_{\mathrm{sell}} P_s) \tag{56.5}$$

Formula (56.5), T is the total number of periods for optimization cycle, $A_{\mathrm{buy}}, B_{\mathrm{sell}}$, respectively, represents the number of buying or selling electricity, P_b, P_s, respectively, stands for the price of buying or selling electricity. Purchase price and sale price are taken as 0.83 yuan/kWh at peak, 0.49 yuan/m^3 as usual, 0.17 yuan/m^3 at valley [8].

56.2.3 Cost of Dealing with the Pollution Which is Caused by Distributed Power

$$C_{\mathrm{pollu}} = \sum_{n=1}^{N} \alpha_n E_n[P(n)] \tag{56.6}$$

In the formula, N stands for the number of polluting gas, $P(n)$ stands for the type of polluting gas, α_n stands for the number of polluting gas N, E_n stands for the cost of dealing the polluting gas (yuan/kg).

56.2.4 Cost of Operating and Maintenance of the Distributed Power

$$C_{r/m} = g_x \times P(t) \tag{56.7}$$

Formula (56.7), $P(t)$ stands for the number of electricity generation. g_x is used to reflect the relationship between the cost on experience. g_{PV} is 0.0096 yuan/kWh, g_{WT} is 0.0296 yuan/kWh, g_{FC} is 0.029 yuan/kWh, g_{DG} is 0.088 yuan/kWh, g_{MT} is 0.0356 yuan/kWh.

56.2.5 Cost of Depreciation

$$C_{DEP} = \sum_{i=1}^{N} \frac{C_{ins}^{i} \times f_{cr}^{i}}{P_{cr}^{i} \times \tau_{g}^{i}} P_{g}^{i} \times \Delta t \tag{56.8}$$

$$f_{cr}^{i} = \frac{s(1+s)^{n_i}}{(1+s)^{n_i} - 1} \tag{56.9}$$

In the formula, C_{ins}^{i} stands for the cost of buying and installing devices. f_{cr}^{i} stands for the coefficient of recycling, S stands for the rate of interest or recovery, τ_{g}^{i} stands for the depreciable or using years of equipments, n_i stands for the hours that the devices can use.

56.3 The Application of QPSOPS Algorithm in the Micro-grid Economic Operation

56.3.1 Quantum Algorithm

Currently, there are representatives of the basic quantum algorithms: Deutsch quantum algorithm (DQA), Shor quantum algorithm (SQA), Grover quantum algorithms (GQA), and so on. Although the principles of these three algorithms vary, all of them can reflect the accelerating effect on classical algorithm, and can truly reflect the quantum parallelism [9]. However, it must be noted that these three algorithms are pure quantum algorithms, its high computing performance can be achieved with the aid of quantum computers which has quantum hardware structure. But the quantum computer is still be at the testing and development stage. Therefore, parallel search mechanism based on quantum search algorithm cannot be achieved currently.

56.3.2 Quantum Particle Swarm Optimization

In practical application, basic PSO algorithms are the same with other optimization algorithms which may produce premature convergence easily, has poor global search capability and slow convergence speed. In order to solve these problems, quantum particle swarm optimization (QPSO) is proposed by some experts who combine quantum searching mechanism in quantum algorithm with particle swarm.

This algorithm can encode the current position of the particle by quantum bits, search the best position of the particle by quantum revolving door and make the variation of particle position by quantum nongate in order to avoid premature convergence. The experimental results also prove that the algorithm has a better ability to optimizing and better optimization efficiency than basic PSO algorithm [10].

56.3.3 Quantum Particle Swarm Algorithm Based on P Systems

The author found that the QPSO algorithm also has some shortcomings. For example, a long running time, optimal results are also need to be improved, etc.

However, there are several advantages in P systems which are suitable for solving optimization problems, for example: distribution, parallelism, easy program implementation, and so on. In particular, the advantages of parallelism and easy programming if applied to the QPSO algorithms must be able to improve QPSO algorithm's drawbacks which are mentioned above to some extent. Based on this, an algorithm is proposed which is a combined P system with quantum particle swarm algorithm in this paper. The algorithm not only can make more rapid convergence, but can also make the convergence results more in line with the requirements. The basic structure of P systems is expressed as follows [11] (Fig. 56.1):

In this paper, P systems in QPSOPS algorithm only used the concept of two membranes and not related to the application of the relevant rules about P systems.

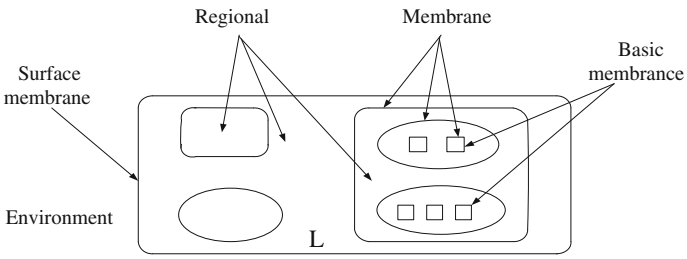


Fig. 56.1 The block diagram of P system

56.4 Case and Result Analysis

In this case, an actual load demand of a teaching building at a certain time in a winter day which supplied by the micro-grid completely is used. We can control the actual output of each distributed power part directly by using QPSOPS algorithm which optimizes the result got from the objective function. And it can not only meet the load demand, but also achieves economic operation of the micro-grid.

Different objective function can be got according to different operation strategies. Graphs which are about the economic operation of micro-grid can be got as follows by using different algorithms which are based on the same objective function for optimizing:

(1) Economy

$$\begin{aligned} C_{\min(\mathrm{ECO})} = C_{\mathrm{nl}} \times \frac{1}{\mathrm{LHV}} \sum_{0}^{J} \frac{P_J}{\eta_j} + \sum_{i=1}^{N} \left(d_i + e_i P_{\mathrm{DE},i} + f_i P_{\mathrm{DE},i}^2\right) \\ + \sum_{t=1}^{T} \left(A_{\mathrm{buy}} P_b - B_{\mathrm{sell}} P_s\right) + g_x \times P(t) \\ + \sum_{i=1}^{N} \frac{C_{\mathrm{ins}}^i \times f_{\mathrm{cr}}^i}{p_{\mathrm{cr}}^i \times \tau_g^i} P_g^i \times \Delta t \end{aligned} \tag{56.10}$$

The simulation results of the three algorithms are shown in Fig. 56.2, individual best fitness value (minimum total cost) for ¥65.9251. The output distribution of distributed power if got as follows: FC 10 kW, MT 5 kW, DG 0 kW.

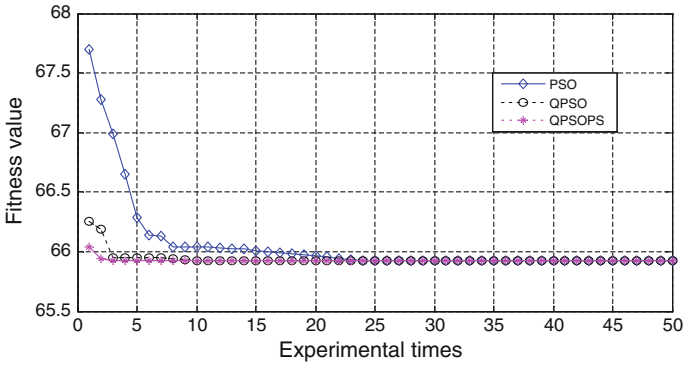


Fig. 56.2 Individual optimal distribution of the three algorithms

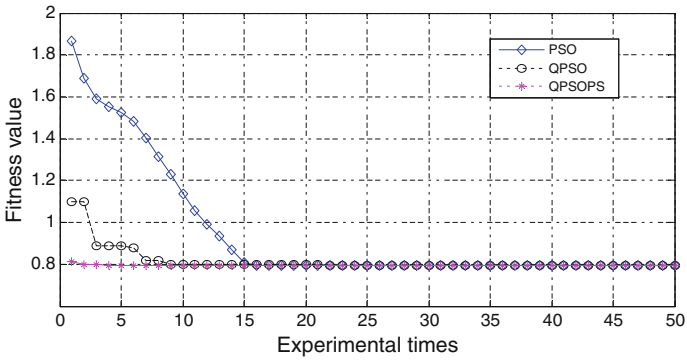


Fig. 56.3 Individual optimal distribution of the three algorithms

(2) Environmental

$$C_{\min(\mathrm{ENV})} = \sum_{n=1}^{N} \alpha_n E_n[P(n)] \tag{56.11}$$

The simulation results of the three algorithms are shown in Fig. 56.3, individual best fitness value (minimum total cost) for ¥0.7965. The output distribution of distributed power if got as follows: FC 4 kW, MT 5 kW, DG 0 kW.

(3) Economy and Environmental

$$\begin{aligned} C_{\mathrm{total}} = C_{\mathrm{nl}} \times \frac{1}{\mathrm{LHV}} \sum_{0}^{J} \frac{P_J}{\eta_j} + \sum_{i=1}^{N} (d_i + e_i P_{\mathrm{DE},i} + f_i P_{\mathrm{DE},i}^2) \\ + \sum_{t=1}^{T} (A_{\mathrm{buy}} P_b - B_{\mathrm{sell}} P_s) + \sum_{n=1}^{N} \alpha_n E_n[P(n)] \\ + g_x \times P(t) + \sum_{i=1}^{N} \frac{C_{\mathrm{ins}}^i \times f_{\mathrm{cr}}^i}{p_{\mathrm{cr}}^i \times \tau_g^i} P_g^i \times \Delta t \end{aligned} \tag{56.12}$$

The simulation results of the three algorithms are shown in Fig. 56.4, individual best fitness value (minimum total cost) for ¥67.1957. The output distribution of distributed power if got as follows: FC 10 kW, MT 5 kW, DG 0 kW.

The simulation results can be seen from Figs. 56.2, 56.3, and 56.4 and some conclusions can be got as follows: First, all the algorithms are able to converge; Second, the running time of program, which QPSO algorithm takes, is the longest but it can be reduced obviously after adding the P systems, while obtaining the QPSOPS algorithm. This is because of the effect of the parallel advantage of P systems. Third, it can be obtained by seeing the results of optimizing whether the

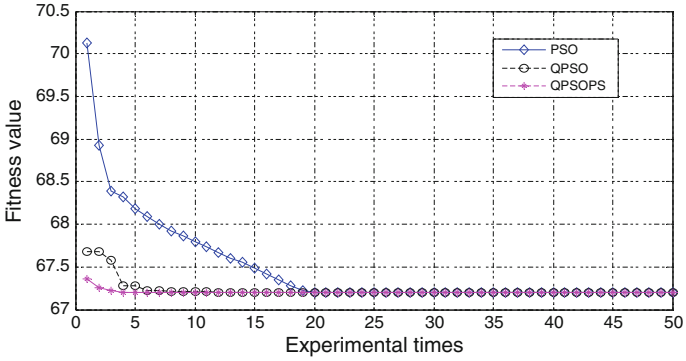


Fig. 56.4 Individual optimal distribution of the three algorithms

worst or average result got by QPSOPS algorithm are better than the other two algorithms. The finding ability and optimize efficiency are really enhanced when compared with the original algorithm after adding the P systems in QPSOPS algorithm.

In summary, it can be seen from the case of micro-grid that the faster convergence and better convergence results which meet the optimization required can be got in QPSOPS than in QPSO or PSO algorithm.

56.5 Conclusion

Based on the comprehensive consideration of power generation, pollution control, operation and maintenance, the optimization model of economic operation of micro-grid is established which give the objective function of the economic operation. The QPSOPS algorithm is proposed in this paper which combined P systems with quantum particle swarm algorithm for the optimization objective function. But actual case proves that the algorithm has fast convergence and high optimized accuracy advantages. It can also reduce the total running costs for micro-grid and achieve the economic operation of micro-grid. In future studies, we wish to consider using the algorithm applied to other fields and go further study about the versatility and practicality of the algorithm.

Acknowledgment This work was partially supported by National Natural Science Foundation of China (No: 61472328).

References

1. Huang YS, Li XZ (2013) Global composite adaptive artificial fish swarm of micro-network and network optimization [J]. Shanxi Electric Power 6:9–13 (in Chinese)
2. Miao YY, Lu JL, Zhu GD (2012) Based on improved multi-objective particle swarm optimization and network optimization scheduling micro-grid. Electric Power Sci Eng 28 (7):15–20 (in Chinese)
3. Xu CP, Hu XB (2011) SOC test scheduling based on quantum particle swarm optimization algorithm. J Sci Instrum 32(1):113–119 (in Chinese)
4. Zeng ZJ (2007) Fuzzy modeling and simulation of P system. Guangdong: Zhongshan University. (in Chinese)
5. Ming D, Jin YL, Qin MM, Wei Y (2009) Modeling and control of distributed generation in fuel cells. Grid Technol 33(9):8–3
6. Yong WS (2011) Based micro-distributed power network design and implementation. Electric Power Autom Equipment 31(4):120–123
7. Wei Y (2010) Distributed power optimization scheduling. Hefei University of Technology
8. Sun SJ (2012) Multi-energy micro-grid optimal allocation and economic operation model. Anhui: Hefei university (in Chinese)
9. Li SY, Li PC (2009) Quantum computing and quantum optimization algorithm. Harbin Institute of Technology Press, pp 1–117 (in Chinese)
10. Wang ZD, Liu LG, Liu ZF, Wang S, You ZA (2014) Configuring wind and fire baling capacity and DC placement quantum particle swarm optimization algorithm. China CSEE 34 (13):2055–2062 (in Chinese)
11. Liu C, Han M, Xing J (2013) A global intelligent optimization algorithm based on membrane systems. J Electron 2013(5):871–877 (in Chinese)

Chapter 57
Research of Mobile Vehicle Information Acquisition System Based on GPS Satellite Positioning

Jianhua Xie and Jian-hua Xiao

Abstract This paper discusses the implementation of information acquisition functions of mobile vehicle management system, in allusion to the disadvantage that the traditional vehicle management system cannot collect the data of vehicle in real time monitoring to vehicle trajectory of vehicle. The system combines with GPS satellite positioning and GPRS communication equipment, based on the principle of the analysis of terminal data processing, design of GPS terminal software and socket server communication program. Experiments proved that the system uses C++ and Socket components which can obtain the vehicle location information in real-time and upload it to the server, combined with GIS to realize the application service for mobile vehicle, such as provide location query, mobile trajectory query, safety monitoring, and information query of peripheral service.

Keywords Satellite positioning · Mobile vehicle management system · Communication · Information collection

57.1 Introduction

Traditional vehicle management systems can realize the vehicle management of common functions, such as vehicle information management, transport, and maintenance for examination and approval, the driver and the process management and so on. They cannot collect the data of the vehicle, and are unable to dynamically monitor the path of the running train track. The mobile vehicle management

J. Xie (✉)
Department of Electronics and Information Technology, JiangMen Polytechnic, Jiangmen 529000, China
e-mail: jmwisdom@163.com

J. Xiao
School of Economics and Management, Wuyi University, Guangdong Jiangmen 529020, China

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_57

system is based on GPS satellite positioning, combination of GPS and GPRS communication mode, using satellite positioning and base station, to data acquisition of the unit vehicle, positioning, monitoring, scheduling, security, and management. Its principle is: install a GPS/GPRS vehicle terminal in each vehicle, through the GPRS wireless mobile communication network connected to the rear management platform, to form a powerful real-time monitoring management system. Combined with GIS to realize 24 h real-time monitoring of subordinate vehicles, at the same time it can also reflect the driver's violation in time and the remote supervision of vehicles realized by the system.

57.2 System Network Topology

The GPS vehicle management system network structure is shown in Fig. 57.1. GPS vehicle terminal installed in vehicles by real-time communication with GPS satellite gets the current location and data such as vehicle real-time speed. GPS vehicle terminal can also install a mobile phone SIM card, start GPRS communication, connect to the Internet through GSM network, and thus connect to the network server. It uploads data such as the real-time vehicle location, vehicle real-time speed

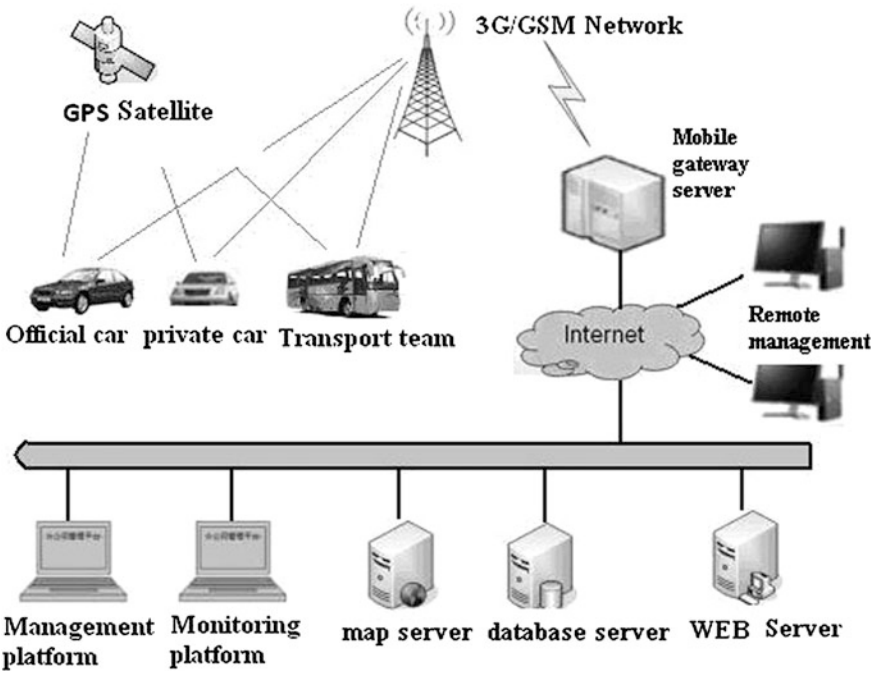


Fig. 57.1 GPS vehicle management system network structure

data, and so on to the server and saves them. WEB management system combined with GIS geographic information system software, vehicle position, and the other information in the form of maps, etc., are visually displayed in the browser interface and supply management monitoring and scheduling.

57.3 Terminal Communication Protocol and Platform

57.3.1 Terminal Communication Principle

Vehicle terminal of the GPS module after obtaining location data from GPS satellites, transmits them to wireless GPRS communication module and sends the GPRS packet data to the GSM base station after processing. Packet data after encapsulation by SGSN, SGSN through GPRS backbone and gateway support contacts GGSN. GGSN accordingly sends for integrated data processing, and then to the WEB database server by Internet [1, 2]. The principle of GPRS and Internet connection is shown in Fig. 57.2.

GPS vehicle-mounted unit of vehicle accepts GPS satellite signal real-time, calculates the onboard unit information such as latitude and longitude, angle, and altitude, speed through the GSM/GPRS wireless mobile communication network and Internet network system monitoring management center to send out related information [3]; GSM/GPRS wireless mobile communication network remote communication as the center of the onboard unit and monitor management system, realizes onboard unit location information, alarm information sent to the center and center of on-board unit on scheduling, control commands transmission [2]; Vehicle terminal with embedded industrial GSM wireless module, its data communication is based on GSM wireless GPRS communication network.

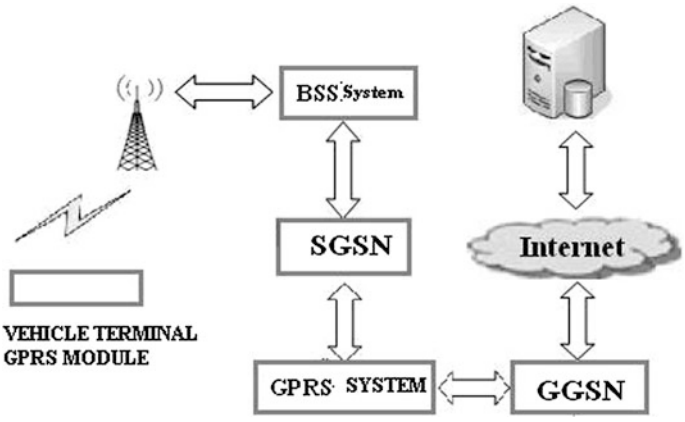


Fig. 57.2 Principle of GPRS and Internet connection

57.3.2 The Agreement in General

Each terminal has a unique ID number of 11 terminal after it delivery. Terminal connects to the platform by TCP. Terminal needs to set up the platform IP and port number first, the connect type is connecting to the platform for devices actively. After connected to the platform, the device will send you a terminal registration message; the message contains the registration device ID. Platform will reply terminals after receipt of the registration information and then terminal will send continuous returned message, the device ID will be contained continuous returned message. One of the most important of continuous returned message is GPS data message, the message contains the UTC time, latitude, latitude N (northern hemisphere) or S (southern hemisphere), longitude, longitude hemisphere E (east longitude) or W (west longitude), ground speed, magnetic declination and other data.

57.4 Mobile Vehicle Management System Function Module

This system includes four modules: the vehicle information management module and system management module of traditional vehicle management system; GIS map control module and vehicle positioning operation module.

The main function of each module:

(1) The GIS map operation module: electronic map browsing query, vehicles query, etc., to realize vehicle real-time display on the map. Administrators can additional recording, modify, and annotations the electronic map by MapXtreme [4].
(2) Vehicle positioning operation module: using GPS terminal to realize the vehicle positioning function, display vehicle position and trajectory on the electronic map of the center of the server.

The system of data acquisition and vehicle monitoring function mainly includes the following two aspects:

(1) Realizing the vehicle positioning information and running status, running speed, and the other information collection, have a two-way communication of information and data by the communication network (GPRS/GSM/CDMA) and the monitoring center.
(2) Sending information about vehicle position and running state to the monitoring center position, real-time monitoring the movement track of vehicles, realizing Far EasTone of location information of mobile vehicle, the purpose of the concentrated display and management [5].

57.5 Communication Function Design and Implementation

57.5.1 The Communication Software Design

According to the demand, this system takes vehicle terminal product GPS838 from Shenzhen Henderson keda ample technology co., LTD., it has the function of GPS + GPRS MODEM IP. Discussing the communication software design of this system in the following.

57.5.1.1 Insertion TCP/IP Protocol

In order to make the terminal device be able to facilitate the application of GPRS communication, the system of GPRS wireless module embedded the written program based on TCP/IP and the PPP protocol in the terminal single-chip microcomputer system.

TCP/IP protocol is a standard protocol, which can be described by the layered model. When the data package, each layer to add their own information content to a data head, this data head will be defined to the data volume by the protocol of the next layer data. After the server-side processing program receives the GPRS data, stripping the corresponding data head, and then treats the rest of the packet content as data volume.

System adopts the UDP protocol which spends less to implement the GPRS communication. Terminal transmits UDP data message to the server data acquisition software by GPRS wireless communication, server data acquisition software is responsible for analyzing and saving the datagram.

57.5.1.2 Terminal Data Processing

The transmission and use of data packets in the terminal and platform servers are based on IP grouping, that is, all the data packets are based on the IP packet. Generally, IP packet uses the PPP protocol to transmit. Wireless module sends the PPP message to the network platform server IP which has been set by terminal, when server receives data, it will analyze data and obtain the terminal IP. ACK frame of server will also be send to the GPRS module based on the terminal IP address, so as to realize collect data from the terminal and Internet network though the transparent data transmission of the GPRS module.

It should be pointed out that, the vehicle terminal without a static IP address, the server cannot request to establish a connection first, and the server must have a fixed IP, so that monitoring terminal cannot be found after landing in the GSM network through the IP server.

57.5.1.3 The Server-Side Data Acquisition Software

Design a socket server communication service program, specifically dedicated to communicate with all terminals, for data acquisition and send related commands, save the location information flow of each terminal, for GIS software processing to realize trajectory playback, display, and other business functions.

Socket interface is API of TCP/IP network; it defines many functions and routines. Acquisition software call the interface function, can be easily designed TCP/IP network communication program. This system adopts the socket of data message type; it is a kind of connectionless sockets, corresponding to the connectionless UDP service applications.

57.5.2 Communication Software Implementation

57.5.2.1 Communication Interaction Control Process

After the vehicle terminal after start up, GPRS connection will start automatically and connected to the Internet by mobile communication network, IP and port to establish a connection to the server platform. Terminal will send handshake signal (BP01) to platform actively at first, after platform receiving the handshake signal, and respond it, command number (AP01). And terminal will send signal register command number (BP04) to platform, platform will reply after it received the command (AP04). With this, the terminal had succeeded in connecting to the platform. Then they can send and receive information or command from each other. The server socket software and GPS terminal communication control process as shown in Fig. 57.3.

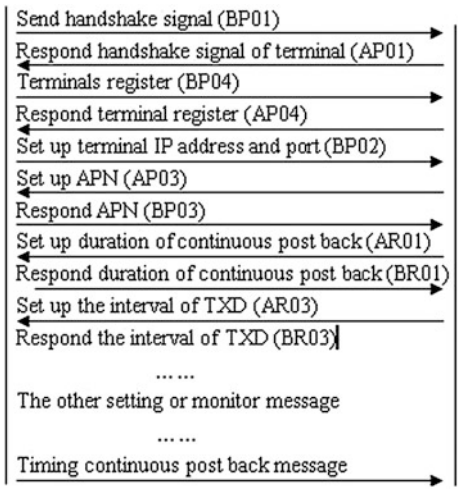


Fig. 57.3 Main control flow chart of communication

57.5.2.2 Socket Server Communication Processing Software Implementation

Data acquisition server software using Borland C++ Builder development. The main control process of communication software processing as shown in Fig. 57.4. The left part is Terminal software processing unit, the right is the server socket software processing unit. Server Socket communication process including initialization Socket and message data processing [6].

The message data processing source code as shown in Fig. 57.5. The function of OnGetClientData functions Program as follows: to deal with all kinds of data, analyze the request of the terminal and the uploaded location data, respond to the request of the terminal, save the location data to the database.

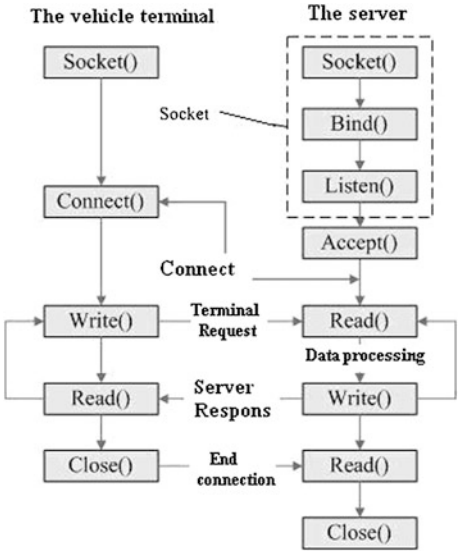


Fig. 57.4 Socket communication software flow chart

```
int __fastcall TIOCPServer::GetData(TCLIENT_IO_DATA *pClientIOData,TCLIENT_PER_IO_DATA *IOData)
{
  TClientMess *pClientMess;
  int i;
  ProceDataPoolCritical->Enter();
  pClientMess = ProceDataPool.Get();  //Message memory pool
  ProceDataPoolCritical->Leave();
  ......
  pClientMess->pClientIOData = pClientIOData;
  pClientMess->IOData = IOData;
  ......
  DataQueue.push(pClientMess);  //The message data is pushed into the queue
  ......
  OnGetClientData();      //Processing data
  return 1;
```

Fig. 57.5 Message data processing source code

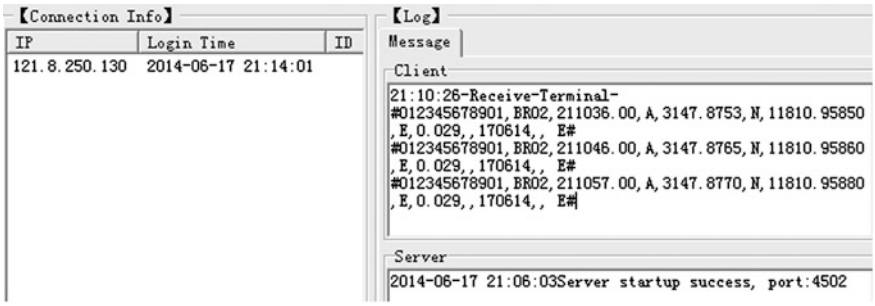


Fig. 57.6 Main interface of socket communication software

Fig. 57.7 Test vehicle positioning information

57.5.2.3 Implementing the Main Interface of Communication Software

The main interface of socket communication software is shown in Fig. 57.6. The terminal connection information is on the left side, and at the same time displays the IP and connection time; the received terminal data information is shown on the upper right, such as 21:10:26—Receive—Terminal—# 012345678901, BR02, 211036.00 A, 3147.8753 N, 11810.95850 E, 0.029, and 170614, and E# said in time 21:10:26, after receiving the vehicles position data that uploaded by Terminal 012345678901; the related operation information of server is showed at low right window.

Positioning the vehicle information on the electronic map combined with GIS after collected the vehicle information is as shown in Fig. 57.7.

57.6 Conclusions

This system successfully used the vehicle equipment to receive GPS satellites signals, to obtain the vehicle's real-time longitude and latitude location data after data processing. In turn, the information and command for monitoring center is sent from monitoring client to gateway server, and then to the specified onboard equipment by the gateway server, implementing two-way communication of information and data between communication network and monitoring center. System combines the common WEBGIS platform (such as MapXtreme platform of MapInfo Company or use JavaScript API of Baidu map) to publish the vehicle position information, real-time monitoring of movement of track of vehicles, convenient traffic managers to monitor vehicles. This system is convenient for the vehicle managers to monitor the vehicles that have improved the management efficiency of the company vehicles and the transport costs of production.

Acknowledgments This paper was supported by Jiangmen City Science and Technology Project, No: 2013ZJ008 and Guangdong Province Science and technology project. No: 2012B040500063.

References

1. Cheng Y (2009) The design and development of vehicle monitoring management system based on GPS_GIS_GPRS. Master thesis, Xi'an University of Science and Technology (in Chinese)
2. Han B (2003) GPRS principle and network optimization. Mechanical Industry Press, Beijing (in Chinese)
3. Liu J (1995) Global positioning system (GPS) principle and application. Surveying and Mapping Press, Beijing (in Chinese)
4. Zhang H, Cue WC (2009) The vehicle management system based on WebGIS of MapXtreme. Microcomput Inf 25:130–131 (in Chinese)
5. Liu H (2003) GPS vehicle supervision and control system based on GSM/GPRS remote communication system. Comput Meas Control 11(11):846–847 (in Chinese)
6. Information on http://www.51aspx.com

Chapter 58
An Underwater Image Classification Algorithm Based on PCA and D-S Evidence Theory

Yihua Shi and Daqi Zhu

Abstract In this paper, an integrated underwater image classification method is proposed by combing D-S (Dempster-Shafer) evidence theory and Principal Component Analysis (PCA) method. First, underwater image texture features are extracted and compressed by using PCA method. Then, the correlation coefficient between image feature and classification pattern is calculated, and the fusion belief function of every image classification pattern can be obtained through D-S evidence theory. Finally, by comparing the results between single PCA feature components and fusion data recognition, the superiority of the proposed integrated method in image recognition is demonstrated.

Keywords D-S evidence theory · PCA · Texture feature extraction · Belief function · Image classification

58.1 Introduction

Because of the complexity of the underwater environment, image processing in the natural underwater environment is a very complex task [1]. At present, the main challenge is the undetected environment and irregular refraction of light by current.

In order to reach a more precise underwater image classification, many methods have been proposed. SVM-based architecture has been applied successfully in underwater object detection problems [2]. This method is simple for calculation, but

Y. Shi · D. Zhu (✉)
Laboratory of Underwater Vehicles and Intelligent Systems, Shanghai Maritime University, Haigang Avenue 1550, Pudong New District, Shanghai 201306, China
e-mail: zdq367@aliyun.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_58

it relies heavily on color. To overcome the limitation of the underwater medium, a confidence-based approach to enhancing underwater acoustic image formation is proposed [3], but it requires relative image sequences.

The texture of an image is usually considered as gray scale or color distribution of some kind of regularity. Most research in texture classification focuses on the feature extraction part [4]. In the past, the common statistical method is gray matrix. At present, the method of gray level co-occurrence matrix (GLCM) integrating gradient is more advanced [5]. Although it is relatively simple and convenient for image recognition by the directional characteristics or properties of the individual scales of the texture image, it has lower recognition accuracy, expensive computation, and greater uncertainty.

Inspired by the advantage of Kernel Principal Component Analysis (KPCA) [6] feature detection and classification, Padmavathi et al. propose the KPCA-SIFT [7] feature detection method for underwater images by combining KPCA and Scale Invariant Feature Transform (SIFT) [8] together. This method is reliable and accurate. However, the computation is still expensive, and this approach is unable to classify when meeting with complex surrounding environment.

In this paper, in order to improve the accuracy of classification and reduce the running time of the algorithm, the D-S evidence theory is applied to the image recognition to obtain the characteristic parameters of common texture feature statistics, and a PCA algorithm is used to reduce the dimensionality of texture statistics data [9], and an integrated underwater image classification method is proposed by combing D-S evidence theory and PCA method. The proposed underwater image classification method can not only reduce computation complexity, but also improve the accuracy of image classification.

This paper is organized as follows: In the second section, the proposed integrated underwater image classification algorithm including texture features extracting compressing is described. Experiment results are presented in the third section. Finally, conclusions are drawn in Sect. 58.4.

58.2 The Proposed Integrated Underwater Image Classification Algorithm

Texture feature contains important information about the arrangement of object surface structure. Therefore, the texture feature has been widely used in image analysis and image classification. Our method of image classification is based on the texture feature. Figure 58.1 is the system chart of our method where F_1 and F_2 are principal components from the image. $m_t(u_0), m_i(u_1), \ldots, m_i(u_i)$ is the assignment to the value of the reliability function for each image. $m(u_0), m(u_1), \ldots, m(u_i)$ is the assignment to the value of the reliability function after D-S data fusion.

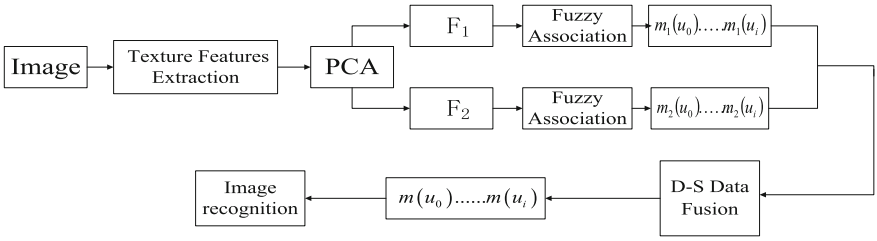


Fig. 58.1 PCA-DS image recognition system

58.2.1 Texture Analysis

Gray matrix describes the distribution of image texture of gray, while the gradient contains the image texture structure. The two characteristics are combined to form a gray level co-occurrence matrix (GLCM) [10] which can represent more comprehensive texture features.

Assume that a figure is $f(x, y), x = 1, 2, \ldots, M, y = 1, 2, \ldots, N$, and its gray level is L. The gray matrix of image extracted from the original image by the gradient operator is $g(x, y)$. Then, stretch the gradient image, and get a new gray-scale image $G(x, y)$. Its gray level is L_g, which is given as follows:

$$G(x,y) = \frac{g(x,y) - g_{\min}}{g_{\max} - g_{\min}}(L_g - 1), x = 1, 2, \ldots, M, y = 1, 2, \ldots, N \tag{58.1}$$

where $g_{\max} = \max \quad g(x, y)$, $g_{\min} = \min \quad g(x, y)$.

Then GLCM can be defined as $\{ H_{ij}, i = 0, 1, \ldots, L-1; j = 0, 1, \ldots, L_g - 1\}$. Here, H_{ij} is defined as the number of elements in set $\{(x, y)|f(x, y) = i, G(x, y) = j\}$. After $[H_{ij}]_{L\times L_g}$ is normalized, $[P_{ij}]_{L\times L_g}$ is obtained where $P_{ij} = H_{ij}/\sum_{i=0}^{L-1}\sum_{j=0}^{L_g-1} H_{ij}$. First, the image is derived, on the basis of which a GLCM is produced. After that, extract characteristic parameters of image texture attributes. The formulas of eight statics are given as follows [11]:

Large Gradient Strengths and Angular Second Moment:

$$f_1 = \sum_{i=0}^{L-1}\sum_{j=0}^{L_g-1}(j+1)^2 P_{ij}, \quad f_2 = \sum_{i=0}^{L-1}\sum_{j=0}^{L_g-1} P_{ij}^2 \tag{58.2}$$

Mean Gradient and Gray-Scale Mean Square Error:

$$f_3 = \sum_{j=0}^{L_g-1} j\left(\sum_{i=0}^{L-1} P_{ij}\right), \quad f_4 = \left\{\sum_{i=0}^{L-1}[i - \sum_{i=0}^{L-1} i(\sum_{j=0}^{L_g-1} P_{ij})]^2(\sum_{j=0}^{L_g-1} P_{ij})\right\}^{\frac{1}{2}} \tag{58.3}$$

Gradient MSE and Gray Entropy:

$$f_5 = \left\{ \sum_{j=0}^{L_g-1} [j - \sum_{j=0}^{L_g-1} j(\sum_{i=0}^{L-1} P_{ij})]^2 (\sum_{i=0}^{L-1} P_{ij}) \right\}^{\frac{1}{2}}, \quad f_6 = -\sum_{i=0}^{L-1} \left(\sum_{j=0}^{L_g-1} P_{ij} \right) \log_2 \left(\sum_{j=0}^{L_g-1} P_{ij} \right) \quad (58.4)$$

Gradient Entropy and Mixing Entropy:

$$f_7 = -\sum_{j=0}^{L_g-1} \left(\sum_{i=0}^{L-1} P_{ij} \right) \log_2 \left(\sum_{i=0}^{L_g-1} P_{ij} \right), \quad f_8 = -\sum_{i=0}^{L-1} \sum_{j=0}^{L_g-1} P_{ij} \log_2 P_{ij} \quad (58.5)$$

Although it is relatively simple and more convenient for image recognition by eight statics characteristics of the texture image above, the computation of this directional characteristics recognition method must be expensive. In this paper, a PCA algorithm is used to reduce the dimensionality of texture statistics data and reduce the computation.

58.2.2 Basic Concept of PCA

PCA is a powerful tool for data reduction and classification purposes [12], and it can express the data in such a way as to highlight their similarities and differences.

Assume that the original data $X = (X_1, X_2, \ldots, X_m)^T$ is a random vector which dimension is m. In this paper, X_i is an eight dimension characteristic parameter of image texture defined on the above. First, calculate the covariance matrix $A = E[XX^T]$. Then, calculate the eigenvectors and eigenvalues of the covariance matrix. The eigenvalues of the covariance matrix of X are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$, and the corresponding eigenvectors set to $P = [p_1, p_2, \ldots, p_m]$. Finally, we simply take the transpose of P and multiply it on the left of the original dataset $T = P^T X$. The principal component data $T_1, T_2, \ldots, T_n, n \ll m$, which can represent the main information of X are obtained, according to the significance level which can be got from eigenvalues. Usually, it is commanded that both reduce the data to one or two dimension and keep 90 % principal component.

58.2.3 D-S Evidence Fusion Theory

After compressing the dimensionality of texture statistics data, in order to improve the accuracy of image classification, D-S evidence theory is applied to the multi-parameter data fusion in this paper. Each feature parameter (sensor) forms an evidence set in multi-sensor information fusion. It can constitute the belief function assignment of object mode which represents the credible degree assuming of each

target mode. The multi-parameter data fusion is to combine several evidence groups to form a new comprehensive evidence group which means that unite each feature parameter (sensor) fusion belief assignment function, so as to provide target mode decision comprehensive accurate information.

58.2.3.1 Belief Function and Assignment

Θ is assumed as a finite set of mutually exclusive alternatives which is not empty, called frame of discernment. The frame of discernment Θ can be defined as:

$$\sum_{u} m(u) = 1, u \in \Theta, \quad m(\phi) = 0 \tag{58.6}$$

where $m(u)$ represents every possible hypothesis. The belief function is defined as follows.

$$m_j(u_i) = C_j(A_i)/\{\sum_{i=1}^{N_c} C_j(A_i) + N(1 - R_j)(1 - W_j\alpha_j\beta_j)\} \tag{58.7}$$

$$C_j(A_i) = G_j(A_i) = \exp(-\frac{(x_i - c)^2}{2\sigma^2}) \tag{58.8}$$

$$R_j = (W_j\alpha_j\beta_j)/(\sum_{k=1}^{N} W_k\alpha_k\beta_k), k = 1, 2\ldots, N \tag{58.9}$$

$$a_j = \max\{c_j(u_i)\}, i = 1, 2\ldots N_c, \ \beta_j = \lambda_j/\sum_{i=1}^{m} \lambda_i \tag{58.10}$$

where N is the dimension number of data. R_j is the reliability coefficient of data in j class. α_j is the maximum correlation coefficient of data in j class. β_j is the proportion of class j in data. In this paper, we use Gaussian subordinate function $G_j(X_i)$ to define correlation coefficient $C_j(A_i)$. The uncertainty belief function is given as follows.

$$m_j(\theta) = 1 - \sum_{i=1}^{N_C} m_j(u_i) \tag{58.11}$$

58.2.3.2 Combination Rule of D-S Evidence Theory

According to D-S evidence theory, supposed Θ contains two evidences, the corresponding belief function assignment are m_1 and m_2, whose focal elements are

$A_1, A_2 \ldots A_k$ and $B_1, B_2 \ldots B_k$ respectively. The combined belief function assignment can be defined as

$$\begin{cases} m(A) = \sum\limits_{A=A_i \cap B_j} [m_1(A_i) m_2(B_j)]/(1-C), A \neq \Phi \\ m(A) = 0, A = \Phi \end{cases} \tag{58.12}$$

where $C = \sum\limits_{\Phi=A_i \cap B_j} m_1(A_i) m_2(B_j)$ which reflects the degree of evidences' conflict. Φ represents empty set.

58.3 Experiments and Results

In order to test the proposed algorithm effectiveness, five underwater pictures are showed in Fig. 58.2 for the experiment. Figure (a) shows an ROV. Figure (b) is sea weeds and (c) is sea sediment. Figure (d) represents sea water. Figure (e) is the image to be classified, which is the ROV that we have already known.

58.3.1 Image Texture Statistics Extraction

When doing image texture statistics extraction, the gray matrix and GLCM of five images are first calculated. Then, calculate the eight image texture statistics through GLCM. The result is shown in Table 58.1.

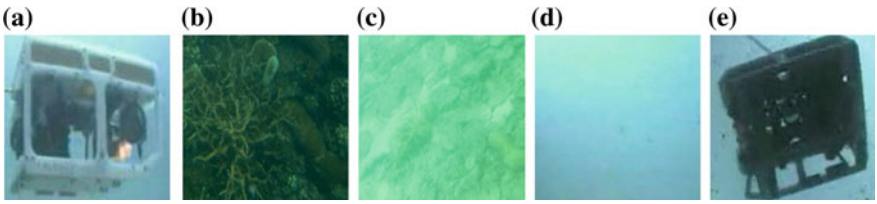


Fig. 58.2 Five underwater images. **a** is case 1. **b** is case 2. **c** is case 3. **d** is case 4. **e** is case 5

Table 58.1 Image texture statistics of five images

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
a	0.0067	5.0095	33.1000	11.4105	−30.5937	1.1028	0.1921	2.2823
b	0.0134	6.7994	15.1506	12.8277	8.7332	0.7573	0.2285	1.9834
c	0.0180	2.4119	13.4895	8.3037	−15.0444	0.7362	0.1187	1.8498
d	0.0156	0.0449	19.1229	1.1792	−0.2712	0.8413	0.0047	1.8450
e	0.0063	4.8710	43.3620	9.6363	−49.9046	1.1727	0.1490	2.3160

Table 58.2 Two principal components characteristics data of four classification images and one test image

	a	b	c	d	e (test image)
F_1	−2.0413	−5.5963	−2.8702	1.9415	0.5023
F_2	−16.8419	−11.0349	−6.9746	−14.6997	−17.1799

58.3.2 Dimension Reduction by PCA

This step processes the data from Table 58.1 through PCA data compression, which is compressed to two dimensions. Also, the same algorithm is applied to the test image. The compressed data retain 96.85 % information from f_1 to f_8. The result is shown in Table 58.2, and it is two principal components characteristics data of four classification images and the test image.

58.3.3 Combination of D-S Evidence

Through Table 58.2, with Eq. (58.8), the correlation coefficient of ROV can be calculated. Other correlation coefficients can also be calculated according to Eq. (58.8), and the results are shown in Table 58.3.

After we get the correlation coefficient, we can use them to calculate belief function. $N = 2$ is the number of main feature parameters. With Eqs. (58.9)–(58.11), the reliability coefficient β, the maximum correlation coefficient α and the relevant characteristics of value assignment R can be calculated.

The belief function of test image belonging to ROV is calculated by Eq. (58.7), and the belief function of test image belonging to other class and the degree of evidences' conflict can also be computed.

Finally, according to Eq. (58.12), we can combine two belief function of principal component into one. The belief function of other classes can also be calculated, which is shown in Table 58.4.

Table 58.3 Correlation coefficient of four classes

	$C(A_1)$	$C(A_2)$	$C(A_3)$	$C(A_4)$
F_1	0.5832	0.0451	0.3876	0.6079
F_2	0.5230	0.0777	0.0599	0.1899

Table 58.4 Single and fusion belief function

	$m(A_1)$	$m(A_2)$	$m(A_3)$	$m(A_4)$	$m(\theta)$	Result
F_1	0.3264	0.0252	0.2169	0.3402	0.0913	Uncertainty
F_2	0.3742	0.0556	0.0429	0.1359	0.3941	Uncertainty
D-S fusion	0.4552	0.0262	0.1577	0.3078	0.0531	ROV

In order to identify underwater image mode, the classification strategies of target mode are showed as follows:

(1) The determination of the target type should have the maximum value of belief function. And it must be less than a certain threshold, which is 0.4500.
(2) The difference between the target type and other types should be greater than a threshold, which is 0.1000.
(3) The uncertain belief function must be less than a certain threshold, which can be 0.0600.

Table 58.4 shows that if classifying only according to the feature statistical measures based on the classification strategies of target mode above, type of the test image cannot be determined by principal component F_1, and the principal component F_2 cannot classify the tested image. If only based on the maximum value of belief function in Table 58.4, the test image will be class fourth according to the belief function of F_1, which is sea water. But, in fact, the image to be tested is ROV. Moreover, the uncertainty belief function of F_2 is very high.

However, after D-S information fusion, the belief function of class 1 rises, while the value of other classes is down. Furthermore, through D-S information fusion, the belief functions can be classified and we can distinguish that the object to be classified is ROV, and the uncertainty belief function is relative low, which is 0.0531. This also meets the needs in practical situations that the target ROV can be distinguished from the complex underwater environment.

58.4 Conclusion

In this paper, an integrated underwater image classification method is proposed by combing D-S evidence theory and PCA method. The proposed underwater image classification method can not only reduce computation complexity, but also improve the accuracy of image classification. Although only four kinds of image are used to fusion recognition in this paper, more kinds of the identifying images can be classified if their belief functions can be obtained relatively accurately. Therefore, this method of fusion recognition can be applied to the actual image classification entirely.

Acknowledgments This project is supported by the Creative Activity Plan for Science and Technology Commission of Shanghai (13510721400, 14JC1402800) and the Innovation Program of Shanghai Municipal Education Commission (13ZZ123).

References

1. Rife JH (2003) Automated robotic tracking of gelatinous animals in the deep ocean. Stanford University, Stanford (Citeseer)
2. Mitra V, Wang C, Banerjee S (2006) Lidar detection of underwater objects using a neuro-SVM-based architecture. IEEE Trans Neural Netw 17(3):717–731

3. Murino V, Trucco A (1999) A confidence-based approach to enhancing underwater acoustic image formation. IEEE Trans Image Process 8(2):270–285
4. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
5. Liu L, Fieguth PW (2012) Texture classification from random features. IEEE Trans Pattern Anal Mach Intell 34(3):574–586
6. Kim KI, Park SH, Kim HJ (2001) Kernel principal component analysis for texture classification. Signal Process Lett IEEE 8(2):39–41
7. Padmavathi G, Muthukumar M, Kumar Thakur S (2010) Kernel principal component analysis feature detection for underwater images. Adv Mater Res 129:953–958
8. Wu Q, An J, Lin B (2012) A texture segmentation algorithm based on PCA and global minimization active contour model for aerial insulator images. IEEE J Sel Top Appl Earth Observ Remote Sens 5(5):1509–1518
9. Zhaohua W, Xiangdong H (2009) Principle of phase measurement and its application based on all-phase spectral analysis. J Data Acquis Process 24(6):777–782
10. Beliakov G, James S, Troiano L (2008) Texture recognition by using GLCM and various aggregation functions. In: IEEE international conference on fuzzy systems (IEEE world congress on computational intelligence), IEEE, pp 1472–1476
11. Blake A, Rother C, Brown M et al (2004) Interactive image segmentation using an adaptive GMMRF model. Computer vision-ECCV 2004. Springer Berlin, pp 428–441
12. Mustafa M, Taib MN, Murat ZH et al (2010) GLCM texture feature reduction for EEG spectrogram image using PCA. In: 2010 IEEE student conference on research and development (SCOReD), IEEE, pp 426–429

Chapter 59
Study on Prediction of the Total Power of Agricultural Machinery Based on Fuzzy BP Network

Hongfu Ai

Abstract It made full use of agriculture-related data, self-learning, and prediction ability of BP neural network which is in the field of artificial intelligence is applied, and combined with fuzzy set theory, in-depth study of the related algorithm, using C++ language programming, established the agriculture machinery corresponding to the total dynamic forecasting model, digging the potential of total power of agriculture machinery in this paper. The author hopes that a reasonable forecast of the development trend of the total power of agricultural machinery in China can be realized and reasonable targets for the development of agricultural mechanization in China.

Keywords The total power of agricultural mechanization · BP neural network · Fuzzy set · Prediction model

59.1 Introduction

Agriculture is an important pillar of economic development in China; Agricultural mechanization is the key link to improve the efficiency of agricultural production. The total power of agricultural machinery is an important index to reflect objectively and to measure the level of the development of Agricultural Mechanization. The total power of agricultural machinery refers to the summation of all kinds of agricultural machinery power, forestry, animal husbandry, and fishery production and transportation. It will have a positive effect on the development of agriculture in China which to grasp the development trend of the total power of agricultural machinery in general and work out the reasonable development plan of agricultural machinery power. That can make productivity grow rapidly. Achievements have been made from the research

H. Ai (✉)
The Information of Teaching and Management Center, Jilin Agricultural University, No 2888 Xin Cheng Street, Changchun, China
e-mail: aixin1115@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_59

of agricultural machinery total power. For example, in Gou Guohua's paper, the gray theory was applied to predict the total power of agricultural machinery [1, 2]. In Yu and Cheng's paper, the ARIMA model is used to predict the total power of machinery [3]. According to the characters of the total power of agricultural machinery and the factors of change [4, 5], to put forward the idea of combining fuzzy set negative feedback and establish the corresponding model of artificial neural network. Forecast the development trend of China's total power of agricultural machinery. Provide scientific reference basis for decision-making departments to formulate development planning of Agricultural Mechanization.

59.2 Fuzzy Set Theory

The concept of the fuzzy set is presented by L.A.Z a Den, a USA cybernetics expert. Fuzzy set took two absolute concept in ordinary set theory of flexibility, which is u either to A or does not belong to A, by the membership function instead of the absolute "belongs to" or "does not belong to the" relationship. Its membership arbitrary real values between 0 and 1

$$0 \leq \mu A(\mu) \leq 1 \tag{59.1}$$

$\mu A(\mu)$: A (fuzzy subset) membership function. It represents a collection of any element in the u membership degree of fuzzy subset of A. We can select according to the requirements of different λ (confidence level) to determine the relationship of subordination which can be expressed as:

$$0 \leq \lambda \leq 1 \tag{59.2}$$

When $A(\mu) \leq 1$ is $u \in A$, else $u \notin$ A. λ Is from 1 to 0, A is gradually expending. Therefore Fuzzy subset A is a set of boundaries with vacillation. It will become larger when λ becomes smaller. The membership function in the practical application of the selection is the key.

59.3 BP Artificial Neural Networks

Back Propagation Neural Networks is a multilayer feed forward network which is proposed by a team of scientists led by Rumelhart in 1986. The network is one kind according to the error back-propagation algorithm for training [6]. As suitable for nonlinear characteristic object analysis and prediction tools, it is with strong self-study ability. The network model is one of the most widely used, and has been widely used in various fields of Science and Engineering. BP artificial neural

network is often a three-layer structure, the input layer, the middle layer (hidden layer), and the output layer [7]. The algorithm is with error back propagation training algorithm. The main idea is that the learning process is divided into two stages: the first stage is the information forward propagation process; the second stage is the error back propagation process [8]. If the output information (The actual output value of network) of the output layer is different from the learning information (Expected value) in the sample output, then each layer recursive calculation of actual output value error between expected output values and network connection weights and neuron threshold will be corrected according to the error.

59.4 The Experimental Data and the Preprocessing

59.4.1 Data Source

The experimental data for this study is from China statistical Yearbooks (2013)—The total power of agricultural machinery of the national statistical data. The original data as shown in Table 59.1, the total power of agricultural machinery unit is million kilowatts.

59.4.2 Normalization Processing

In order to improve the model learning speed and prediction accuracy, the statistical data and the influence factors are the pretreatment. For the original statistical data, in order to the convenience processing, the most simple normalization method is used. The formula is as follows:

Table 59.1 Original data

Year	Total power	Year	Total power	Year	Total power	Year	Total power
1978	11749.90	1987	24836.00	1996	38546.90	2005	68397.85
1979	13379.50	1988	26575.00	1997	42015.60	2006	72522.12
1980	14745.75	1989	28067.00	1998	45207.71	2007	76589.56
1981	15680.10	1990	28707.70	1999	48996.12	2008	82190.41
1982	16614.21	1991	29388.60	2000	52573.61	2009	87496.10
1983	18021.90	1992	30308.40	2001	55172.10	2010	92780.48
1984	19497.22	1993	31816.60	2002	57929.85	2011	97734.66
1985	20912.55	1994	33802.50	2003	60386.54	2012	102558.96
1986	22950.00	1995	36118.05	2004	64027.91		

Table 59.2 Normalized data

Year	Normalized data	Year	Normalized data	Year	Normalized data	Year	Normalized data
1978	0.100000	1987	0.229197	1996	0.364562	2005	0.659275
1979	0.116089	1988	0.246365	1997	0.398808	2006	0.699993
1980	0.129577	1989	0.261096	1998	0.430323	2007	0.740150
1981	0.138802	1990	0.267421	1999	0.467725	2008	0.795446
1982	0.148024	1991	0.274144	2000	0.503045	2009	0.847828
1983	0.161922	1992	0.283225	2001	0.528699	2010	0.872650
1984	0.176488	1993	0.298115	2002	0.555926	2011	0.882958
1985	0.190461	1994	0.317721	2003	0.580181	2012	0.900000
1986	0.210577	1995	0.340582	2004	0.616131		

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \times 0.8 + 0.1 \tag{59.3}$$

In the formula y is normalized data, y' is real data, $y_{\min}$ is the minimum value in the set of data, and $y_{\max}$ is the maximum value in the set of data. The data processed are falling in between [0.1, 0.9]. After normalizing the data is shown in Table 59.2.

59.4.3 Fuzzy Processing

On the training sample of fuzzy input, according to the characteristics of data, membership function is selected as the formula:

$$\mu(x) = 1 - e^{-6x^2} (0 \le x \le 1) \tag{59.4}$$

Fuzzy data is shown in Table 59.3.

Table 59.3 Fuzzy data

Year	Fuzzy data	Year	Fuzzy data	Year	Fuzzy data	Year	Fuzzy data
1978	0.0487705	1987	0.2309932	1996	0.4854830	2005	0.8861890
1979	0.0651628	1988	0.2617551	1997	0.5485261	2006	0.9137021
1980	0.0805242	1989	0.2888379	1998	0.6038207	2007	0.935372
1981	0.0918358	1990	0.3006266	1999	0.6650699	2008	0.9577298
1982	0.1037681	1991	0.3132422	2000	0.7178372	2009	0.9725136
1983	0.1228648	1992	0.3304041	2001	0.7528142	2010	0.9825776
1984	0.1442181	1993	0.3587669	2002	0.7867455	2011	0.9858680
1985	0.1658793	1994	0.3963338	2003	0.8141931	2012	0.9992357
1986	0.1988542	1995	0.4400912	2004	0.8501451		

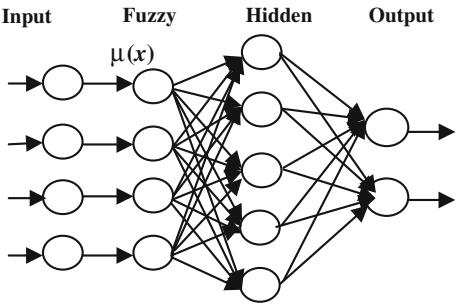


Fig. 59.1 Fuzzy topological structure of BP neural network

59.5 Create a Fuzzy BP Neural Network Prediction Model

59.5.1 The Establishment of the Network Topological Structure

According to the agricultural machinery total power of data complexity, combined with the characteristics of this prediction model, the fuzzy BP network model is designed to four layers, input layer, fuzzy layer, hidden layer, and output layer. And the hidden layer is only set to 1 layer. According to the actual application, after a great deal of experimental verification, the set of rules of BP network topology which is based on the fuzzy are following. Input variables of the network is four dimensions, total power of machinery is composed of four consecutive years of data. Fuzzy number layer nodes for the same four dimensions. The input layer data fuzzy processing. The output layer is two dimensional after the input variables by composition of 4 years data of 2 years forecast value.

The number of the hidden layer node is set according to the following rules: the number of nodes in the hidden layer is constantly adjusting, but there is general scope. The number of nodes in hidden layer is $L \prec n-1$ or $L = \log_2^{n+1}$ rounding, or $L \prec \mathrm{Sqrt}(m-n) + a\log_2^{n+1}$ (n is a upper layer of nodes, m is the next layer of nodes, a is an arbitrary constant). The design of hidden layer is with five neurons. The network topology is shown in Fig. 59.1

59.5.2 Total Power of Agricultural Machinery in the Fuzzy BP Neural Network Forecasting Process

The specific processes are described as below:

(1) The original data acquisition and the data are collected.
(2) The data was normalized and the difference of data is eliminated.
(3) It is fuzzy processing normalized data.

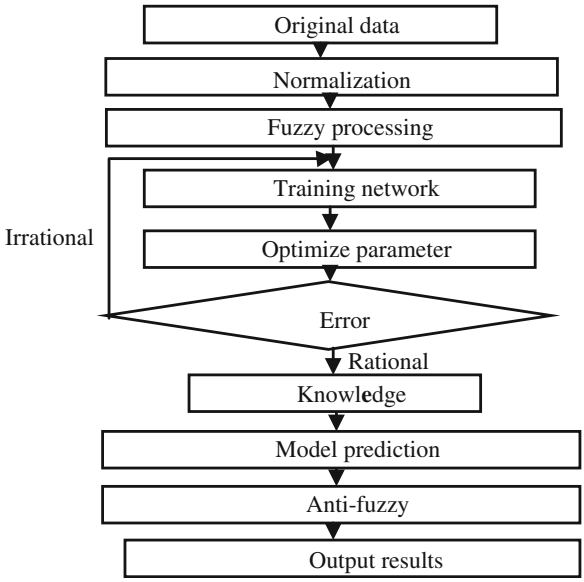


Fig. 59.2 The prediction procedure of fuzzy BP neural network

(4) The processed data is divided into two parts, training samples and testing samples, in which the training sample data of 30 groups, The test samples of five groups.
(5) Training samples by fuzzy BP neural network and making the corresponding optimization of parameters according to the effect of training to create a prediction knowledge base.
(6) To predict the total power of the National Agricultural.
(7) Anti-normalization and anti-fuzzy process of prediction results
(8) Output the predicted results and have a reasonable explanation.

Fuzzy prediction flowchart model is shown in Fig. 59.2

59.6 Prediction and Analysis

59.6.1 Parameter Setting

According to the actual situation of the research problem, combined with the basic principle of BP neural network, in the fuzzy BP neural network prediction model, the neural network artificial neurons are selected as follows: the input layer neurons using transparent activation function:

$$\text{transparence}(x) = x \tag{59.5}$$

The middle layer and the output layer neurons using sigmoid activation function.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{59.6}$$

The experimental parameters setting: after a lot of experiments, in order to achieve better convergence speed and better data, the model of the learning rate is set to 0.005, error is set to 0.0001.

59.6.2 The Predicted Results

The first 30 group of the 35 group statistical data obtained data is as the model of learning and training, through fuzzy BP neural network's training to form the knowledge base. The five group of data at last is testing samples. To validate the data samples. The test results in Table 59.4.

By fitting test, the model fit the data accords with the actual value, the error is small. Therefore, the model that is applied to predict the total power of agricultural machinery has certain feasibility. According to the fuzzy BP network knowledge base, reasonable application of the model, the prediction of the total power of agricultural machinery in China for the next 4 years is calculated. The predicted results are shown in Table 59.5.

Predicted by the model, we can see the next 4 years, the total power of agricultural machinery in China has shown a steady upward trend. The prediction results can be provided to relevant decision-making departments which make reasonable formulation of agricultural mechanization development planning. To further improve the environment for the development of China's agricultural mechanization, mobilize the enthusiasm of farmers in agricultural production; promote the steady development of agriculture in our country.

Table 59.4 Model test results

Year	Total power	Fitting data	Error rate
2008	82190.41	81265.42	1.13
2009	87496.10	86532.00	1.10
2010	92780.48	93820.46	1.12
2011	97734.66	99135.26	1.43
2012	102558.96	103564.91	0.98

Table 59.5 Forecast of China's total agricultural machinery power value (2013–2016)

Year	Total power (million kW)
2013	108648.36
2014	112486.52
2015	118201.00
2016	125823.47

59.7 Conclusions

With the advent of the era of big data, the effective use of the mass data of Agricultural Mechanization in China, it provides reasonable forecasting analysis and science and accurate service for Agricultural Mechanization development of China. Using the 1978–2012 Chinese total power of agricultural machinery data and the negative feedback artificial neural network algorithm based on Fuzzy in this research, the author established a reasonable prediction model. Reasonable analysis and forecast based on historical statistics, the corresponding knowledge base formed. Through the simulation experiment, results indicated that the model in the prediction of the total power of agricultural machinery has certain scientific. The establishment of the model prediction and analysis of total power of agricultural machinery in China provides a new idea and has certain application value.

References

1. Gou GH (2013) Study on forecasting total power of agricultural machinery based on grey theory and BP neural network. J Chin Agric Mech 34(5):35–37, 63 (In Chinese)
2. Xie J, Zhang Y (2008) Application of unbiased grey model in total power requirement of agricultural machinery forecasting. J Anhui Agric Sci 36(20):8397–8398 (In Chinese)
3. Yu G, Cheng J (2009) Predict and analysis of agricultural machinery total power in china-based on ARIMA model. Agric Mech Res 2:51–54 (In Chinese)
4. Fang H, Zhang S, Ding W (2007) Study on forecasting the yield in Maize Regional Test based on BP neural network. J Anhui Agric Sci 35:10969–10970 (In Chinese)
5. He Z-D, He R (2010) Analysis on the relationship between gross power of agricultural machinery and key influencing factors based on time series analysis. Chin Agric Mech 1:20–24 (In Chinese)
6. Yan PF, Zhang CS (2006) Artificial neural networks and evolutionary computation. Tsinghua University press, Beijing (In Chinese)
7. Li X (2013) Air quality forecasting based on GAB and fuzzy BP neural network. J Huazhong Univ Sci Tech (Nat Sci Ed) 10:18–22 (In Chinese)
8. Jiang Z, Mao B, Meng XX et al (2010) An air quality forecast model based on the BP neural network of the samples self-organization clustering. In: 2010 6th international conference on natural computation, Yantai, IEEE, pp 1523–1527

Chapter 60
Target Threat Assessment Based on Improved RBF Neural Network

Kehu Xu, Depeng Kong and Jinyu Chen

Abstract In order to solve the problem of threat assessment of informational armored unit, an improved neural network is designed, which based on the feature of fast and only best approximation. This RBF neural network can adjust the weight of assessment index according to different patterns of operations, and can effectively assess the operations target of informational armored unit. Example analysis shows that the improved method can satisfy the requirements of target threat assessment of different operation patterns and has higher practical value.

Keywords Threat assessment · Operation patterns · RBF neural network · Assessment index

60.1 Introduction

Threat assessment is an important part of the command and control system [1], which can assist commanders in the actual battlefield environment to make accurate and reasonable judgments and decisions to the enemy. Currently, the assessment method used most is a combination of qualitative and quantitative, such as AHP, Fuzzy sets, Bayesian network method [2–4], and rarely use radial basis function (RBF) neural network modeling and solving. This paper proposes an assessment method based on the improved RBF neural network, which can fully consider the impact on assessment index weights of different operation patterns [5]. This method is more consistent with the actual operational situation, reduces the effect of battlefield uncertainties, with high accuracy assessment.

K. Xu · D. Kong (✉) · J. Chen
Department of Control Engineering, Academy of Armored Force Engineering, PLA No 21, Dujiakan, Fengtai, Beijing, China
e-mail: 1358013459@qq.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_60

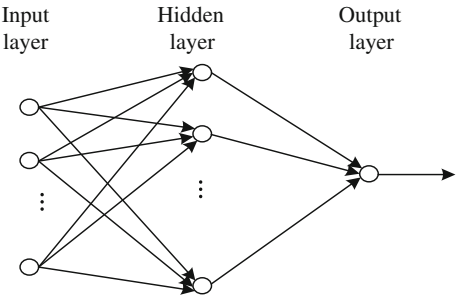


Fig. 60.1 The structure of traditional RBF neural network

60.2 The Basic Idea of Improved RBF Neural Network

The traditional RBF neural network is a kind of three feed-forward static network with strong input and output mapping function, the structure is shown in Fig. 60.1. The input layer is directly constituted by the source node, the function is only receiving the input signal and passing it to the hidden layer; the hidden layer nodes function constituted by a radial basis function, typically choose a Gaussian function, when the center of the Gaussian function (the center of the vector, the center width) determined and the mapping from the input layer to hidden layer is also determined; output layer to achieve a linear combination of the nonlinear basis functions of the hidden layer nodes output.

RBF neural network has four sets of parameters: center vector, center width, the output weights, and output adjustment coefficient, only choosing these parameters accurate and reasonable can play a nonlinear approximation capability of RBF neural network, RBF neural network has been proved that has the only best approximation properties and can approximate any continuous function with arbitrary precision [6]. This paper proposes an improved RBF neural network with four-layer structure shown in Fig. 60.2. The index adjusting layer is added before input layer in order to adjust evaluation index according to different operation patterns, which can overcome the shortcomings of traditional RBF neural network that the same assessment results occurs in different operation patterns.

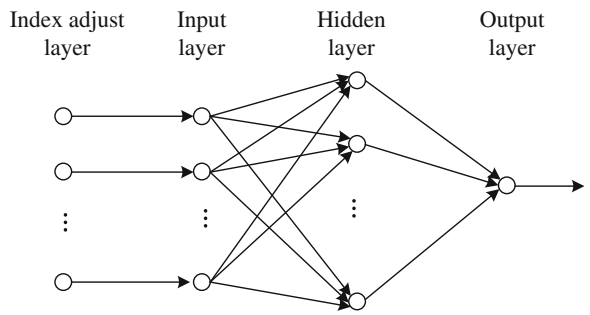


Fig. 60.2 The structure of improved RBF neural network

Table 60.1 Determination of evaluation index

Evaluation index	Target type	Enemy distance	Motor characteristics	Damage probability	Condition of intervisibility	Condition of terrain
Target threat degree	I^{TYPE}	I^{DIS}	I^{MOV}	I^{DES}	I^{VIEW}	I^{EARTH}

60.3 Assessment Index and Sample

Armored unit combat under conditions of information has many uncertainties, tense battlefield situation, fast-paced firepower, and different operation patterns presents different characteristics, which needs to assess the threat of the target to eliminate the target with large threat to our armored unit as soon as possible and to save ourselves as much as possible. Battlefield commanders need to grasp the situation accurately, select reasonable target assessment criteria, and be able to adjust evaluation index weights timely according to the changes of combat style, in order to make the battlefield situation more conducive toward our direction and achieve victory in battle. Select the evaluation index for the actual battlefield tactical situation: target type (I^{TYPE}), enemy distance (I^{DIS}), motor characteristics (I^{MOV}), damage probability (I^{DES}), condition of intervisibility (I^{VIEW}), condition of terrain (I^{EARTH}), shown in Table 60.1.

Sample set of target threat assessment index can be obtained from the above and the sample data can get from training simulation system, $\mathbf{T} = \{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_N\}$ $\mathbf{T}_i = [T_{i1} \quad T_{i2} \quad T_{i3} \quad T_{i4} \quad T_{i5} \quad T_{i6}] = [I_i^{\mathrm{TYPE}} \quad I_i^{\mathrm{DIS}} \quad I_i^{\mathrm{MOV}} \quad I_i^{\mathrm{DES}} \quad I_i^{\mathrm{VIEW}} \quad I_i^{\mathrm{EARTH}}]$.

T is a target threat sample set and N training samples are included, T_i as the i training sample, $I_i^{\mathrm{TYPE}} \quad I_i^{\mathrm{DIS}} \quad I_i^{\mathrm{MOV}} \quad I_i^{\mathrm{DES}} \quad I_i^{\mathrm{VIEW}} \quad I_i^{\mathrm{EARTH}}$ are the input part of training samples, y is the threat of the training sample i as the output part. Get each target threat index (the input part of the training samples) by index analysis method. Determine the comprehensive target threat degree (output part of the training sample) using AHP method, then revise the sample which individual error is larger through expert analysis and the final target threat sample set is got.

60.4 The Training of Improved RBF Neural Network

Due to the principle both improved RBF neural network and the traditional method to threat assess are basically the same, following are the training processes and methods used for improving RBF neural network based on target threat assessment. The model of target threat assessment is established based on the index in Table 60.1, and shown in Fig. 60.3.

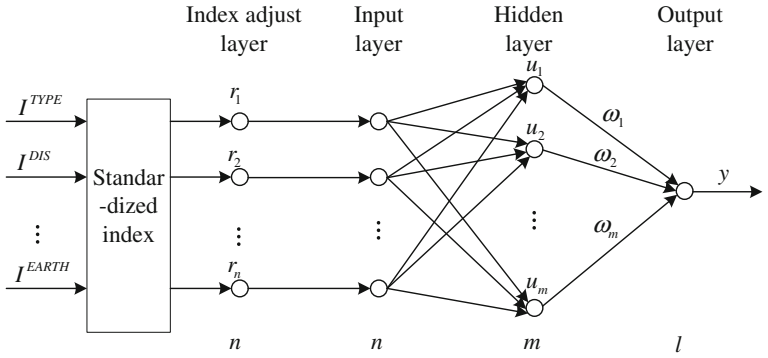


Fig. 60.3 The model of target threat assessment based on improved RBF neural network

60.4.1 Identify Index Adjustment Layer According to Operation Patterns

Information armored combat units can be broadly divided into three categories: offensive combat, defensive battle, and suffered battle. Under different operation patterns, usually adopt different strategies to attack objectives. Accordingly, the target threat assessment is different. As in offensive combat, and the target in the situation of defensive, then the target type I^{TYPE} of threat assessment index accounts for a weak weight; but in the case of defensive combat missions, target type I^{TYPE} need to be focused on considering target threat assessment, so it accounts for a larger weight. Therefore, the level of target threat and operation patterns link closely. And the adjustment layer reflects the impact of different operation patterns to assess each index of battlefield target threat. After normalized index, the threat index entering through the adjusting layer for adjusting, the adjustment vector is $R = [r_1\, r_2\, r_3\, r_4\, r_5\, r_6]$, and $r_i (i = 1, 2, \ldots, 6)$ is the weight of adjusted index. As an example of target type I^{TYPE}, the corresponding index adjusted weight is r_1, the value is as follows:

$$r_1 = \begin{cases} [0.8, 0.95] & \text{offensive combat} \\ [0.95, 1.05] & \text{defensive battle} \\ [1.05, 1.2] & \text{suffered battle} \end{cases}$$

Among them, r_1 is an interval value in the different operation patterns and can be converted in the different operation periods of the same operation pattern. In order to meet the requirements of continuity of the battlefield, r_1 need to adjust in the corresponding value interval. Due to the paper limitation, the impact of operation patterns to other indexes are not repeated.

60.4.2 Determination of Mathematical Model of RBF Neural Network

Since the each index of the detected target has the different physical meaning, the dimension and the range of the index is different and the gap is larger. To eliminate the effect, each index data of the target need to be standardized.

Suppose there are N training samples, which have n data in input section, the standardization process of the input section samples for the training as follows:

$$x_{ij} = \frac{T_{ij} - \overline{T}_j}{\sigma'_j}, \overline{T}_j = \frac{\sum_{i=1}^{N} T_{ij}}{N}, \sigma'_j = \sqrt{\frac{\sum_{i=1}^{N} \left(T_{ij} - \overline{T}_j\right)^2}{N-1}}, i = 1, 2, \ldots, N,$$
$$j = 1, 2, \ldots, N \tag{60.1}$$

T_{ij} is the input part of the training samples T_i before standardization, X_{ij} is the input part of the training samples T_i after standardization, the input part of number i training sample after standardization as $X_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$.

The output part of training samples in the sample generation stage have been limited in scope for being suitable to parameter optimization of RBF neural network, output part of the training samples after standardization is the same as the original data, i.e.,

$$\bar{y}_i = T_i^{\mathrm{TAR}}$$

The input layer can standardize and adjust training samples and then transfer to the hidden layer; the hidden layer nodes composite by radial function, usually chooses Gauss function:

$$u_k = \exp\left(-\frac{(x_i r_i - \mathbf{C}_k)^T (x_i r_i - \mathbf{C}_k)}{2\sigma_k^2}\right), k = 1, 2, \ldots, m \tag{60.2}$$

$C_k = [\, c_{k1} \quad c_{k2}, \quad \ldots, \quad c_{kn} \,]$ as the k hidden layer nodes center vector, σ_k as the k hidden layer nodes center width. RBF neural network hidden layer dimension can be determined based on the empirical formula $m = \frac{3}{2}n$, also can be determined by experiment for the more effective dimension.

The output layer is the linear combination of hidden layer node output, because threat value is the only one output of this problem, that is:

$$y = \sum_{k=1}^{m} w_k u_k \tag{60.3}$$

w_k is the weight coefficient of the k hidden layer nodes to output layer.

60.4.3 Learning Algorithm of RBF Neural Network

Regular RBF network structure is characterized by a network with P input nodes, N hidden node, 1 output node; the number of network hidden nodes is equal to the input sample, activation function of hidden nodes in the form of Gaussian function, and set all input samples as the center of radial basis functions, each RBF takes a unified extension constant [7]. Compared with the traditional RBF neural network, regularization RBF neural network has the following advantages: (1) Regularization network is a universal approximator, as long as there is enough hidden nodes, it can approximate any arbitrary precision multivariate function. (2) Regularization network obtains the best solution, the so-called "best" reflect in meeting both the approximation error and approximation smoothness of the curve. (3) Regularization network has less parameters for identification in case of the input with large dimensions and small amount of sample, reducing the model identification difficult.

According to the features and advantages of regularization RBF neural network, this network structure is chosen since the number of hidden layer nodes are equal to the input sample and set all input samples as the center of the radial basis function, the training process only needs to consider the expansion constant σ_k and the output node weights W_k.

Extended constant of the radial basis functions can be determined according to the spread of the data center, and set it as:

$$\sigma = \frac{d_{\max}}{\sqrt{2N}} \tag{60.4}$$

$d_{\max}$ is the maximum distance between the sample; N is the number of samples.

Adjust the output layer weights with a minimum mean square algorithm (LMS); the input vector of LMS algorithm is the output vector of hidden nodes. Weight adjustment formula:

$$\Delta w_k = \eta(d_k - w_k u_k)u_k, k = 1, 2, \ldots, N \tag{60.5}$$

Weights can be initialized to any value.

60.5 Simulation Example

In order to verify the reasonableness of the improved method, set the target at 60 different battlefield conditions; using the method in Sect. 60.2 to determine the target threat sample set, the select N = 50 samples as training samples, and the remaining 10 samples as test samples, results are shown in Table 60.2:

For comparison, using the same samples for training and testing with the traditional RBF neural network, the final results shown in Table 60.3:

Table 60.2 Results with improved RBF neural network

Number of target	Evaluation index						Threat value	Threat ranking
	I^{TYPE}	I^{DIS}	I^{MOV}	I^{DES}	I^{VIEW}	I^{EARTH}		
1	0.60	0.85	0.75	0.78	0.75	0.75	0.643	6
2	0.60	0.73	1.00	0.82	0.81	0.75	0.694	5
3	0.40	0.93	0.50	0.95	0.87	1.00	0.968	1
4	0.20	0.79	0.25	0.78	0.23	0.25	0.486	9
5	0.40	0.81	1.00	0.46	0.39	0.50	0.585	7
6	0.20	0.56	0.25	0.23	0.41	0.75	0.884	2
7	0.60	0.54	0.75	0.65	0.43	0.75	0.371	10
8	0.60	0.71	0.75	0.73	0.54	0.50	0.547	8
9	0.80	0.36	1.00	0.68	0.87	1.00	0.829	3
10	0.60	0.52	0.75	0.49	0.79	0.75	0.778	4

Table 60.3 Results with traditional RBF neural network

Number of target	Evaluation index						Threat value	Threat ranking
	I^{TYPE}	I^{DIS}	I^{MOV}	I^{DES}	I^{VIEW}	I^{EARTH}		
1	0.60	0.85	0.75	0.78	0.75	0.75	0.668	6
2	0.60	0.73	1.00	0.82	0.81	0.75	0.672	5
3	0.40	0.93	0.50	0.95	0.87	1.00	0.948	1
4	0.20	0.79	0.25	0.78	0.23	0.25	0.375	10
5	0.40	0.81	1.00	0.46	0.39	0.50	0.549	7
6	0.20	0.56	0.25	0.23	0.41	0.75	0.824	3
7	0.60	0.54	0.75	0.65	0.43	0.75	0.489	9
8	0.60	0.71	0.75	0.73	0.54	0.50	0.524	8
9	0.80	0.36	1.00	0.68	0.87	1.00	0.875	2
10	0.60	0.52	0.75	0.49	0.79	0.75	0.794	4

From the comparison of above data, following points are gotten:

(1) Assessment results from improved RBF neural network and the traditional are generally the same, improved methods in line with the actual tactics conditions of the battlefield.

(2) The sorting of objective 9 and 6, objective 4 and 7 changes with the appearance of a cross, comparing with assessment index finds that target type and mobility has a smaller impact on the assessment results in offensive operations, which in line with offensive combat characteristics.

(3) The changes of sorting cross only in the adjacent target sort, such as the target of sorting 2 and 3, sorting 9 and 10, indicating that improved RBF neural network is reasonable and effective.

The above points, basically illustrates the improved RBF able to fully consider the impact of combat on the threat assessment index, the model established in line with the battlefield tactical situations.

60.6 Conclusions

Information armored unit threat assessment model presented in this paper has a strong practicality, the improved RBF neural network designed in this paper can meet the requirements of solution and can provide support data for firepower optimal allocation of information armored units. To verify the effectiveness of the improved method, 10 samples are used for testing, whose results are shown in Tables 60.2 and 60.3. It is easy to find that the target threat sorting results with the improved method of output targets in line with the features of different operation patterns, which verify the effectiveness of the improved method. However, there are still two problems in the improved method: (1) Parameter of index adjusting layer are static and cannot meet the requirements for dynamically changing on battlefield. (2) Assessment index selection is fixed and should be chosen dynamically based on the actual battlefield conditions. How to improve the RBF neural network to solve the above two issues is the focus of future research work.

References

1. Xu KH, Zhang ZY, Huang DS (2013) Research on combat factor quantification of tank unit. J Acad Armed Forces Eng 27(1):48–53 (in Chinese)
2. Jia B, Sun J, Feng ZC (2011) An improved evaluating model of air strike target threat level. Command Control Simul 33(4):25–28 (in Chinese)
3. Zhang L, Tong YT, Xu YH (2009) Research on air target threatening order model of warship formation. Ship Electron Eng 29(6):136–138 (in Chinese)
4. Shui W, Ge Y, Han Y (2009) Algorithm of firepower threat level assessment based on Bayesian network. J Syst Simul 21(15):4625–4627 (in Chinese)
5. Du T, Zhu YG, Hang XL (2009) The evaluation module of tank firepower application automatic system for threat of object. Fire Control Command Control 34(4):138–141 (in Chinese)
6. Zhang GZ (2007) Intelligent control system and application. China Electric Power Press, Beijing (in Chinese)
7. Han LQ (2007) The artificial neural network theory, design and application. Chemical Industry Press, Beijing (in Chinese)

Chapter 61
Design and Implementation of an Embedded Vision Measurement Intelligent Sensor

Yuan Li, Yongbing Wang and Qinglin Wang

Abstract This paper develops an embedded vision measurement intelligent sensor (EVMIS) using ARM microprocessor, Linux system, a camera with a line-structured light projector. Images are captured by CMOS sensor, and are processed in the embedded system. With calibration, the developed vision sensor can realize visual measurement, and the profile information of the objects in 3D scene is extracted as the output of the sensor. The EVMIS is featured with fieldbus and Ethernet interface, which are suitable for industrial field. The experimental results verify the design and implementation of the proposed sensor.

Keywords Visual measurement · Intelligent sensor · Embedded system · Calibration

61.1 Introduction

The visual measuring system has developed from analog signal stage to digital manners. Distributed computing and embedded processing are the two tends of the visual measurement recently [1]. The early visual system acquires image by analog camera, and the computer converts analog signal to digital signal by image acquisition devices and then processes the digital images. However, this method generally cannot meet the requirements of the high-speed and high-definition image. With the development of computer technology, digital vision system has been widely used. At the present stage, digital visual system generally acquires image by digital camera which has the capability to interface directly to the PC (by IEEE1394, CameraLink, LVDS, USB, etc.). Because of the high bandwidth of

Y. Li (✉) · Y. Wang · Q. Wang
School of Automation, Beijing Institute of Technology,
No. 5 Zhongguancun South Street, Haidian District, Beijing
100081, People's Republic of China
e-mail: liyuan@bit.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_61

digital communication and the strong processing ability of computer, the digital visual system can meet most requirements of the image acquisition and processing. Smart cameras can capture and process images without PCs [2]. Smart cameras are generally implemented base on DSP, FPGA, and other high-performance processor. Smart cameras represent the development trend of advanced visual system; however, the smart camera has following limitations: First of all, the smart camera is expensive and not suitable for multinode application as a general sensor. Second, an engineer needs to grasp complex programming skill to configure image acquisitions and processing. Finally, the standard interface of a smart camera is not practical for industrial field where the field bus, serial port, and IO output are widely used. In industrial application, the ideal mode of the visual system is that it can acquire and process image independently as a sensor; can be configured by figure interface; calculate the 3D coordinates of the scene and transmit visual measuring data to other equipment by field bus. In order to achieve this visual system, in this paper, we design an EVMIS with the function of measurement and control.

Though the structured light measuring systems can meet the requirements in precision and speed, measurements are controlled based on either computer or the combination of computer and single-chip computer, resulted in the deficiency in the integration [3]. The EVMIS proposed in this paper is a kind of visual sensor, which can provide not only the functions of smart cameras [4–6], but also fieldbus interface and protocol. It can acquire, process, and transmit image information. The EVMIS is designed integrally in order to reduce the complexity and improve the reliability. This embedded visual sensor has the advantage of low cost, low power consumption, easy to use with other automation system, etc. [7].

Structured light is one method for acquiring 3D information of the scene [8]. It is widely used in the industrial environment. In this paper, the EVMIS consists of line-structured light projector, S3C6410 microprocessor, and CMOS camera. When the intrinsic and extrinsic parameters of the camera and the laser plane are calibrated, the 3D coordinates of feature points in the laser stripe image can be calculated according to triangle measurement principle. The original images and the measuring data can be output through the Ethernet, CAN bus, and IO, so the EVMIS has the function of intelligent sensor. The experimental results show that the system function is realized, practical, and suitable for measurement and visual inspection in the industrial field.

61.2 System Design

The EVMIS consists of three parts, they are the embedded hardware platform, embedded Linux system, and embedded application program. System block diagram shown in Fig. 61.1.

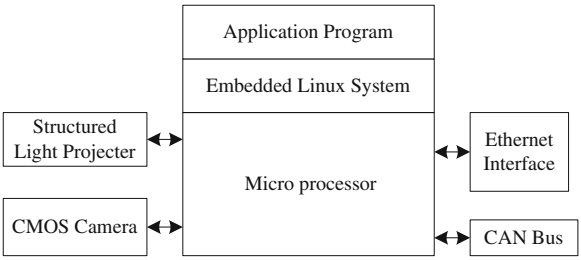


Fig. 61.1 System block diagram

61.2.1 Hardware

Embedded hardware platform uses S3C6410 as the processor. It adopts 64/32-bit intrinsic bus architecture and includes hardware peripherals such as a camera interface, system manager, and i2c-bus interface. The structured light projector and CMOS image sensor are used to capture stripe image. The communication interface includes Ethernet and CAN bus interface.

Images and 3D coordinates data can be transmitted through the network; 3D coordinates can be passed from one sensor to others through CAN bus. The status also can be output from IO port. The EVMIS has low system cost with a strong reciprocity and scalability.

61.2.2 Software

Based on the hardware platform, the embedded Linux operating system is customized. Embedded application software can be divided into the following functional modules: Image acquisition and processing module, CAN bus, and Ethernet module.

In user space, image acquisition calls “Video For Linux Two” which is the unified interface to access audio and video drivers. Image processing calls the library which is defined by custom. CAN bus communication module is extended by the SPI interface. Can Socket is called which provides a socket interface for Linux-based network layer. This design uses a connection-oriented socket to realize the Ethernet communication. The data format and structure of the Ethernet communication are defined in Table 61.1.

Table 61.1 Data format

	Head	Length	Code	Data	Check
Downlink data	*2B*[a]	*4B*	*0xA1*:read	*0000 0001b*[b]: IO state	*2B*
			0xB0:set	*0000 0010b*: Image data	
				0000 0100b: 3D coordinates	
				0000 1000b: Sequence	
				0000 0010b: Image data	
Uplink data	*2B*	*4B*	*0xA1*:read	IO state	*2B*
			0xB0:set	Image data	
				3D coordinates	
				Sequence	

[a] *2B* means two bytes
[b] *0000 0001b* is in the form of binary

61.3 System Implementations

61.3.1 Model and Calibration

The visual measuring model is pinhole model, and three kinds of coordinate systems exist in this model; the relation between them is shown in Fig. 61.2.

(u, v) is the pixel coordinate in the image coordinate system. (x_c, y_c, z_c) is the point in the camera coordinate system. The point (x_w, y_w, z_w) is in the world coordinate system.

In this paper, Faugeras's camera calibration method [9] is used to calibrate the camera. The relationship between the 3D coordinate of the scenes and the pixel point of the image is as follows.

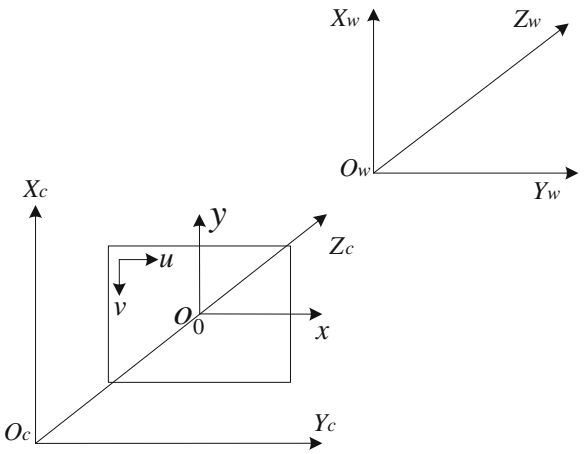


Fig. 61.2 Three kinds of coordinate system diagram

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 & 0 \\ 0 & k_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{P} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \boldsymbol{M}_{\boldsymbol{in}}^{'} \boldsymbol{M}_{\boldsymbol{w}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{61.1}$$

In the formula, k_x and k_y are amplification factors on the x-axis and y-axis separately. u_0, v_0 are image coordinate. $\boldsymbol{R}, \boldsymbol{P}$ are translation matrix and rotation matrix. $\boldsymbol{M}_{\boldsymbol{in}}^{'}$ is completely determined by k_x, k_y, u_0, v_0 and is called intrinsic parameter matrix. $\boldsymbol{M}_{\boldsymbol{w}}$ represents extrinsic parameter matrix.

In order to obtain the intrinsic and extrinsic parameter matrices, the stereo target is chosen as in Fig. 61.3. Set up world coordinate system on the center of the upper surface. Then the coordinates of the eight vertices are known in the coordinate system.

In the camera coordinate system, the equations of the two upper surfaces are as following.

$$\begin{aligned} a_x x + a_y y + a_z z - a_x p_x - a_y p_y - a_z p_z = 0 \\ a_x x + a_y y + a_z z - a_x p_x - a_y p_y - a_z p_z + d = 0 \end{aligned} \tag{61.2}$$

where $[a_x \quad a_y \quad a_z]^T$ is the direction vector of the Z_w axis in the camera coordinate system $O_c X_c Y_c Z_c$. In the camera coordinate system $O_c X_c Y_c Z_c$, $[p_x \quad p_y \quad p_z]^T$ is the position of the origin of the world coordinate system $O_w X_w Y_w Z_w$. d is the distance of the two top surfaces.

The point p_j in the laser stripe and the point $(x_{cj}, y_{cj}, 1)$ are on the same straight line. p_j is also in the laser plane. We can obtain the 3D coordinates of scenes by the following Eq. 61.3. *a, b, c* are the parameters of the laser plane equation.

$$\begin{cases} x = \frac{-x_{cj}}{ax_{cj}+by_{cj}+c} \\ y = \frac{-y_{cj}}{ax_{cj}+by_{cj}+c} \\ z = \frac{-1}{ax_{cj}+by_{cj}+c} \end{cases} \tag{61.3}$$

When the laser is projected on two surfaces of the trapezoidal target, the points in the laser stripe are both in the laser plane and the surface of the target. Therefore,

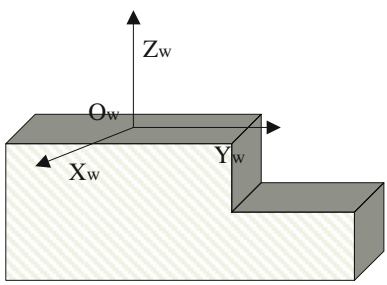


Fig. 61.3 The stereo target of the structured light calibration

these points meet their equations. We can obtain the laser plane equation based on the least square method by Eqs. 61.2 and 61.3.

61.3.2 Image-Processing Library

For structured light system, the feature extraction of the stripe is important. In this paper, image-processing library is designed as the standard library which can be called by the main program. By configurable software, users can program and configure to generate a sequence for image processing. This is discussed in another paper. When EVMIS receives the sequence by the network, it interprets and executes the sequence by calling the library.

61.3.3 Measurement Realization

The application model of the EVMIS is shown in Fig. 61.4.

As the condition of the intrinsic and extrinsic parameter matrices and the laser plane equation are obtained. The image coordinates of the feature points can be converted into the 3D coordinates by Eqs. 61.1 and 61.3. According to the data format defined in Table 61.1, the measuring results can be transmitted through the network. And the CAN bus interface is the communication interface among the sensors.

61.4 Experiment

The EVMIS prototype system is shown in Fig. 61.5.

The intrinsic and extrinsic parameters of the camera are calibrated by Faugeras linear model. The calibration target is shown in Fig. 61.3. And we need another four image coordinates from the laser stripe in both of the upper and lower surfaces. The camera parameter matrix is as follows.

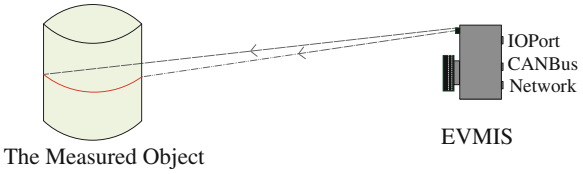


Fig. 61.4 The model of the EVMIS

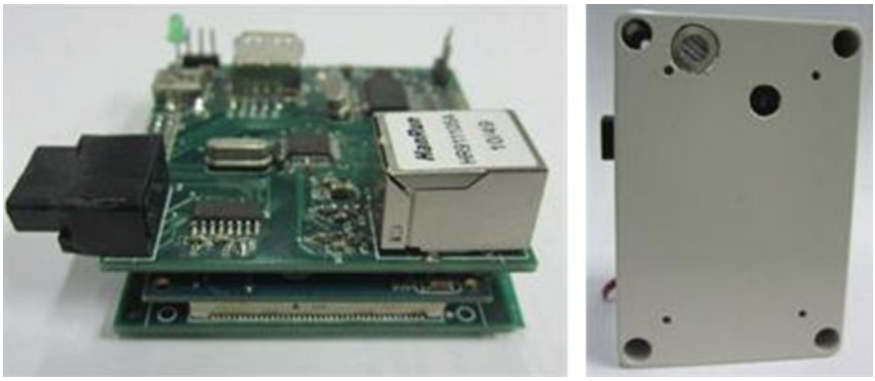


Fig. 61.5 Embedded vision measurement sensor prototype

$$\boldsymbol{M}_{\mathbf{in}} = \begin{bmatrix} 1298.2 & 0 & 339.2 \\ 0 & 1249.6 & 277.0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{M}_{w} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & -72.9 \\ 0.0 & 1.0 & 0.0 & 44.8 \\ 0.0 & 0.0 & -1.0 & 659.4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The equation of the laser plane is as $24x + 3y + 17z - 10000 = 0$.

The other 3D coordinates of the light stripe can be calculated by Eq. 61.3. And parts of results are shown in Table 61.2.

We adopt the EVMIS as the TCP server and PC as the TCP client. After the server and client connect successfully, we verify the network communication on PC by the test software which is programmed by QT. The software is shown in Fig. 61.6.

Table 61.2 Measurement data

Image coordinate (u_j, v_j)	World coordinate (x_{wj}, y_{wj}, z_{wj}) (*unit:* mm)
(211 92)	(−70.1 −105.2 710.4)
(207 134)	(−72.2 −81.2 709.3)
(198 197)	(−77.2 −45.5 710.3)
(193 240)	(−79.9 −21.0 710.0)
(249 267)	(−45.8 −5.3 659.4)
(243 312)	(−48.9 18.5 659.8)
(236 366)	(−52.4 47.0 660.0)
(230 420)	(−55.5 75.4 659.5)
(225 459)	(−58.0 96.0 659.6)

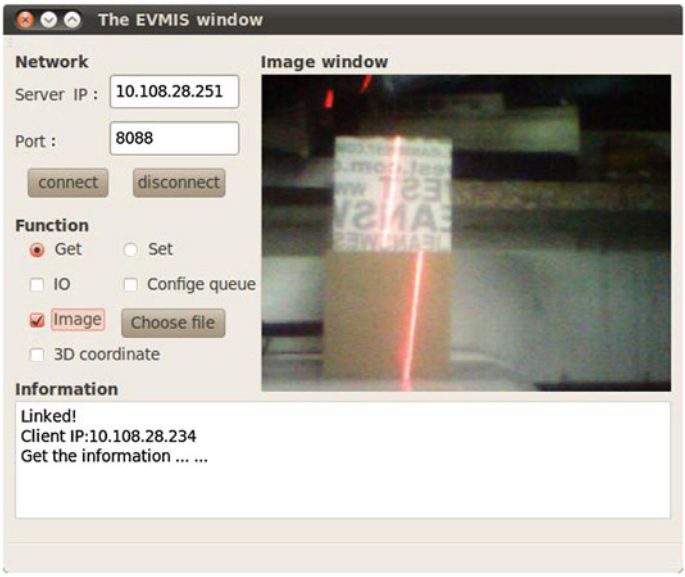


Fig. 61.6 The visual window of the test software

61.5 Conclusion

By the experimental verification, the EVMIS can acquire and process the image, obtain the visual measuring information, and be suitable for general industrial application. At present, although the EVMIS has many shortcomings such as the simple function, low processing ability, and low real-time performance. Considering this EVMIS is for industrial application, it can meet the general requirement of measuring and monitoring. And it is particularly suitable for a wide range of application.

Acknowledgments This research is supported by National Natural Science Foundation of China (no. 61375100 and no. 61472037).

References

1. Rinner B, Jovanovic M, Quaritsch M (2007) Embedded distributed middleware on smart cameras. In: International conference on acoustics, speech, and signal processing, pp 1381–1384
2. Wolf W, Ozer B, Lv T (2002) Smart cameras as embedded systems. Computer 35:48–53
3. Pages J, Salvi J, Garcia R et al (2003) Overview of coded light projection techniques for automatic 3D profiling. In: IEEE international conference on robotics and automation, pp 34–44
4. Caarls W, Jonker PP, Corporaal H (2002) SmartCam: devices for embedded intelligent cameras. In: Proceedings of the 3D progress workshop on embedded system, pp 14–17

5. Rana V, Matteucci M, Caltabiano D, Sannino R, Bonarini A (2008) Low cost smartcams design. ESTIMedia pp 27–32
6. Aggarwal A (2008) Embedded vision system (EVS). In: International conference on mechtronic and embedded systems and applications, pp 618–621
7. Bramberger M, Doblander A, Maier A, Rinner B (2006) Distributed embedded smart cameras for surveillance applications. Computer 39:68–75
8. Valkenburg RJ, McIvor AM (1997) Accurate 3D measurement using a structure light system. Image Vis Comput 16:68–80
9. Xu D, Tan M, Li Y (2008) Visual measurement and control for robots. Defence Industry Press, Beijing, pp 45–52 (in Chinese)

Chapter 62
Unsteady Aerodynamics Modeling Using SVM and Artificial Neural Network

Yichao Jiang, Qingjie Zhao and Jihong Zhu

Abstract Recently, more and more attention has been drawn by the aircraft's maneuvering problem. This problem is very significant for improving performance of the nonlinear and unsteady modeling methods used for aircrafts at high angles of attack. In this paper, support vector machine (SVM) and artificial neural network are introduced into unsteady aerodynamics modeling. The experimental results show that the generality and precision have been significantly improved using these two methods, which verifies that machine learning methods can be applied to unsteady aerodynamic modeling.

Keywords Unsteady aerodynamics · Machine learning method · System modeling

62.1 Introduction

Due to the development of aviation technology, lots of attention has been paid to the maneuvering problems of aircrafts at high angles of attack. Thanks to the application of aerodynamic thrust vector control and advanced aerodynamic control, the aircraft expands its flight envelope. At the same time when the radar systems and flight control systems are updated continuously, it is necessary to improve the maneuvering performance accordingly.

Y. Jiang · Q. Zhao (✉)
Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, People's Republic of China
e-mail: zhaoqj@bit.edu.cn

Y. Jiang
e-mail: yi_chao_jiang@163.com

J. Zhu (✉)
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, People's Republic of China
e-mail: jhzhu@tsinghua.edu.cn

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_62

The so-called maneuver at high angles of attack means that the aircraft completes the maneuver with great change that over the stall angle in a very short time. In this process, the airflow around the aircraft may appear flow separation, vortex breakdown, or other complex flow phenomenon. As a result, the corresponding aerodynamic force and moment show a high degree of nonlinear and unsteady hysteresis characteristics. Furthermore, the mobility and maneuverability of aircrafts may decrease. So it is import to investigate the aerodynamic characteristics of flow phenomenon [1, 2].

Numerous traditional modeling methods applied to unsteady aerodynamics model can be found in literatures, such as polynomial model [3], Fourier model [4, 5], state-space model [6], and fuzzy logic model [7]. There is no doubt that every method has its own advantages and disadvantages [8, 9]. In this paper, two machine learning methods are used to realize regression. Their effectiveness have been proved by the experiments compared with the traditional polynomial model.

62.2 Model Principle

For machine learning methods, the main process contains three steps: Extracting data, training the model, and testing the model. Assuming that the data is $\{(x_i, y_i)\}, i = 1, 2, 3, \ldots m$, $x_i \in R^n$, where we have defined x_i, n, y_i and m as the input of system, dimension of x_i, the actual output of system and the number of data, respectively.

62.2.1 SVM Regression Principle

SVM is based on VC (Vapnik-Chervonenkis) theory in statistical learning theory and structural risk minimization principle.

Assuming the regression model $f(x)$ is linear, like Eq. (62.1), where $\langle w, x_i \rangle$ representing inner product operation:

$$f(x) = \langle w, x \rangle + b, w \in R^n, b \in R \tag{62.1}$$

It is believed that the regression model approximately the real model when the condition $|f(x_i) - y_i| \leq \varepsilon, i = 1, 2, 3, \cdots m$ satisfied. The principle of SVM regression could be interpreted by Fig. 62.1, where the solid line represents $f(x)$, and the dotted lines are the upper and lower bounds. It could be found out that the circular points between two dotted lines meet the precision requirement.

The larger distance between the dotted line and solid line, the higher probability for results meet the requirements would get. So the problem becomes how to find the max value of the distance.

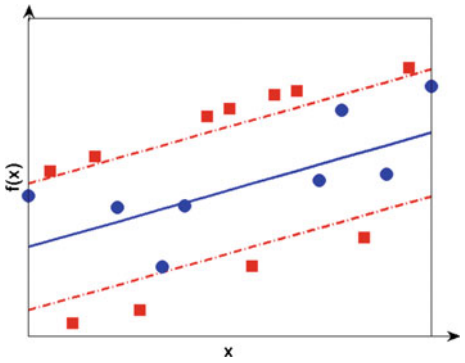


Fig. 62.1 Regression principle of SVM

$$d = \frac{\varepsilon}{\sqrt{1 + \|\omega\|^2}} \tag{62.2}$$

When the ε is fixed, this problem needs to find the minimal value of ω [10, 11].

$$Q = \min\left\{\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} L(f(x_i), y_i)\right\}, \quad \text{s.t.} \ |\langle \omega, x_i \rangle + b - y_i| \le \varepsilon \tag{62.3}$$

The C represents the factor of punishment, indicating the tolerance of error data. The L represents loss functions, and generally insensitive area function is chosen.

In order to solve the problem, slack variables ξ_i, ξ_i^* are introduced,

$$\begin{aligned} Q &= \min\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right\}, \\ \text{s.t.} \ \xi_i \ge 0, \xi_i^* \ge 0, \quad & y_i - f(x_i) \le \varepsilon + \xi_i, \quad f(x_i) - y_i \le \varepsilon + \xi_i^* \end{aligned} \tag{62.4}$$

Based on KKT (Karush-Kuhn-Tucker) condition, combining lagrangian method, the following equation could get:

$$\begin{aligned} f(x) &= \sum_{i=1}^{\text{nSV}} (\alpha_i - \alpha_i^*)\langle x_i, x \rangle + b \\ & s.t. \, x_i \in R^n, b \in R \end{aligned} \tag{62.5}$$

In this equation nSV is the number of support vectors. And α_i, α_i^* represent the Lagrangian multipliers. For nonlinear problems, we could use kernel instead of inner product in linear space, and then we could solve it in the same way mentioned above in high-dimension space.

Different kernels may affect the performance of SVM. There are some kernels used frequently, including linear kernel Eq. (62.6), polynomial kernel Eq. (62.7), and RBF kernel Eq. (62.8):

$$k\langle x_i, x_j\rangle = x_i x_j' \tag{62.6}$$

$$k\langle x_i, x_j\rangle = (x_i x_j' + 1)^d \tag{62.7}$$

$$k\langle x_i, x_j\rangle = \mathrm{e}^{\frac{-\left\|x_j - x_j\right\|^2}{2\sigma^2}} \tag{62.8}$$

62.2.2 Artificial Neural Network

The neural network simulates human's brain to make decision, and it can be treated as a black system consists of many neurons [12–14]. The basic neuron is shown in Fig. 62.2, where x is the input and it may be multidimensional. ω is the weight proportion of input data, threshold is bias, and f is transfer function.

There are many available transfer functions, such as Sigmoid function Eq. (62.9), hyperbolic tangent function Eq. (62.10), and Gaussian function Eq. (62.11).

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{62.9}$$

$$f(x) = \frac{e^x - e^x}{e^x + e^x} \tag{62.10}$$

$$f(x) = \exp\left[\frac{1}{2\sigma_i^2}\sum_j^m (x_j - \omega_{ji})^2\right] \tag{62.11}$$

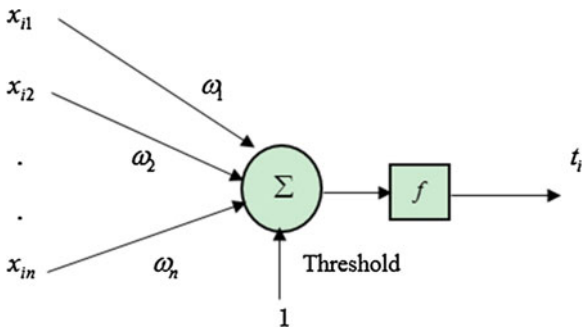


Fig. 62.2 The basic component of neuron

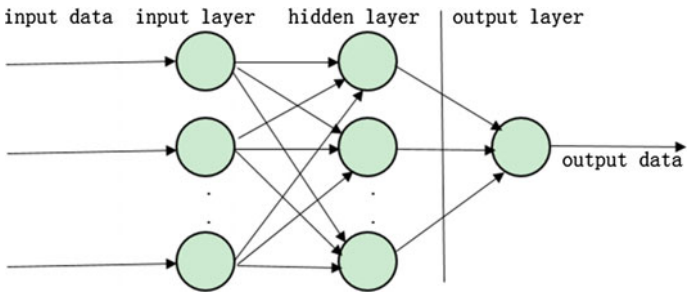


Fig. 62.3 A typical forward neural network with three layers

A common forward neural network contains three layers: Input layer, hidden layer, and output layer. For any layer, different number of neurons and different transfer functions could be set separately. The detailed message is shown in Fig. 62.3.

The input x is handled by input layer, hidden layer, and output layer, and then output $f(x)$.

62.2.3 Model Design and Regression Result Evaluation

Based on the knowledge of aerodynamics, the aircraft's rapid mobility at high angles of attack may lead to strong unsteady hysteresis on the aerodynamic force and moment. Under this condition, the force and moment not only depends on the current aircraft flight parameters, but also on the change rate of relevant flight parameters.

The nonlinear polynomial model shows that, for oscillations maneuver at large angles of attack in the longitudinal direction, the aircraft's aerodynamic coefficient and the aerodynamic moment are related to the angles of attack of the aircraft α, change rate of angles of attack $\dot{\alpha}$, and reduced frequencies k.

In oscillations maneuver experiments, the movement of the aircraft satisfy the following equations

$$\alpha = \alpha_m + \alpha_0 \cos(kt) \tag{62.12}$$

$$\dot{\alpha} = -k\alpha_0 \sin(kt) \tag{62.13}$$

$$\ddot{\alpha} = -k^2\alpha_0 \cos(kt) \tag{62.14}$$

where α_m represents mean angle, α_0 is the amplitude.

Assuming our dataset is $\{(x_i, y_i)\}, i = 1, 2, 3, \cdots m$, where x_i is input, y_i is output, and m represents the number of data. Based on these data, a system model can be built First, it is necessary to extract the features of dataset. For oscillations

maneuver at large angles of attack in the longitudinal direction, the output is defined as the aerodynamic force or moment. The angle of attack of the aircraft α, change rate of angle of attack $\dot{\alpha}$, and reduced frequencies k are chosen as the input. So $x_i = [k, \alpha, \dot{\alpha}]$, and y_i represents aerodynamic force and moment.

Three measures are used to evaluate the regression result, including MSE (Mean-Square-Error) Eq. (62.15), MRE (Mean-Relative-Error) Eq. (62.16), and R^2 Eq. (62.17)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{62.15}$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{62.16}$$

$$R^2 = 1 - \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \Bigg/ \left(\sqrt{\sum_{i=1}^{n} y_i^2} + \sqrt{\sum_{i=1}^{n} \hat{y}_i^2}\right) \tag{62.17}$$

In these three equations, n is the number of data, y_i is the origin actual data, and $\hat{y}_i$ is the regression result. The detailed evaluate standard is that if MSE and MRE approximating 0, and R^2 approximating 1, better regression result can be obtained.

62.3 Experiments

62.3.1 Model Training

The train data is obtained from reference [5], where the reduced frequency, mean angle, and amplitude are chosen as [0, 0.0320, 0.0782, 0.1116], [32.5°, 22.5°], [30°, 20°]. The lift coefficient data CL is chosen to train these three models and the result are shown in Fig. 62.4 and Table 62.1.

62.3.2 Model Testing

The test dataset is also obtained from reference [5], where the reduced frequency, mean angle, and amplitude are chosen as [0.0558, 0.0892], 27.5°, 25°, respectively. These three methods use this dataset to test their accuracy. The experiment result are shown in Fig. 62.5 and Table 62.2.

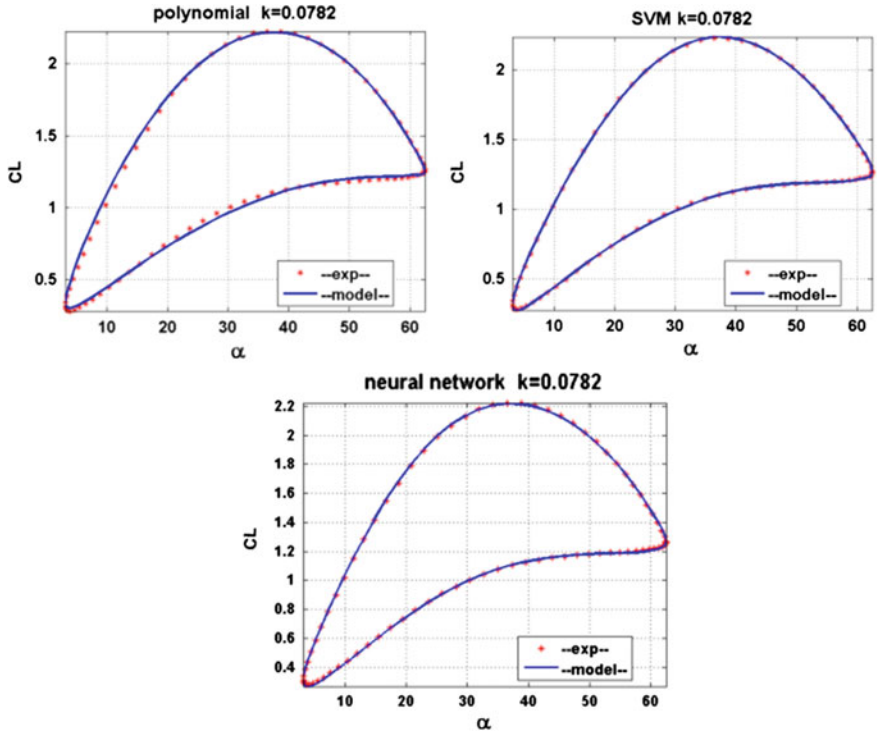


Fig. 62.4 The regression results of different model on the train dataset when k = 0.0782

Table 62.1 Different model's regression results on the train dataset

Model	MSE	MRE	R2
Polynomial	5.34e-5	0.0248	0.9910
SVM	4.56e-5	0.0064	0.9974
Neural network	5.01e-5	0.0083	0.9959

It is clear that in Fig. 62.5 and Table 62.2, the MSE and MRE values of SVM model and forward neural network model are much smaller than the polynomial model, and their R^2's values are higher, which means that SVM and forward neural network perform much better on the test dataset.

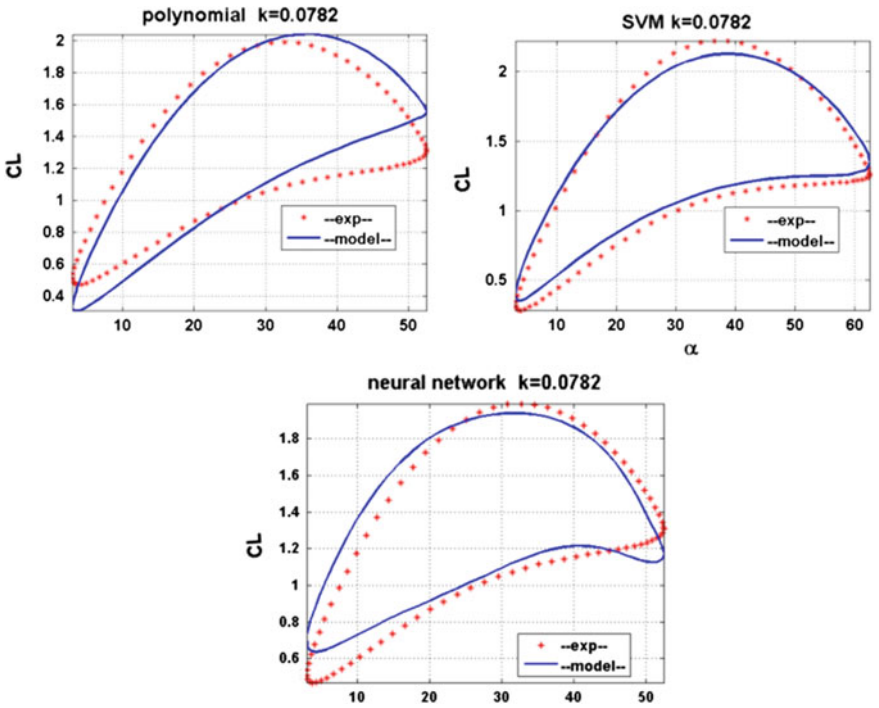


Fig. 62.5 The regression results of different model on the test dataset when $k = 0.0782$

Table 62.2 Different model's regression results on the test dataset

Model	MSE	MRE	R2
Polynomial	0.0255	0.1445	0.9383
SVM	0.0061	0.0869	0.9700
Neural network	0.0094	0.0787	0.9590

62.4 Conclusion

The experiments show that SVM and forward neural network methods do much better on train dataset and test dataset than the polynomial method. And they can be applied to unsteady aerodynamic modeling at high angles of attack. The detailed processes of model training and testing have been discussed in the section of model design. Subsequently, some further research on two major issues must be done. The first one is to modeling the other aerodynamic coefficients in the longitudinal and lateral motion. Due to the aforementioned experiments data obtained from the wind tunnel, the next issue is to find out how to apply these methods to the actual flight data.

References

1. Greenwell DI (2004) A review of unsteady aerodynamic modelling for flight dynamics of manoeuvrable aircraft. AIAA Pap 5276–5301
2. Wang Q, Cai JSH (1994) Advances in aerodynamic modeling of airplane at high angles of attack. Exp Meas Fluid Mech 8(3):7–16 (in Chinese)
3. Lin GF, Lan CE, Brandon JM (1997) A generalized dynamic aerodynamic coefficient model for flight dynamics applications. AIAA Pap 97–3643
4. Chin S, Lan CE (1992) Fourier functional analysis for unsteady aerodynamic modeling. AIAA J 30(9):2259–2266
5. Hu CC, Lan CE, Brandon JM (1993) Unsteady aerodynamic models for maneuvering aircraft. AIAA Pap 93–3626
6. Goman M, Khrabrov A (1994) State-space representation of aerodynamic characteristics of an aircraft at high angles of attack. J Aircr 31(5):1109–1115
7. Wang Z, Lan CE, Brandon JM (1998) Fuzzy logic modeling of nonlinear unsteady aerodynamics. AIAA Pap 4351
8. Zhang HY (2009) Comparative of evaluation of unsteady aerodynamic mathematic models of aircraft. Aerodynamic Research and Development Center, pp 11–12 (in Chinese)
9. Shi ZW, Wu GX (1999) The comparison of several unsteady aerodynamics models at high angle of attack. Acta Aerodyn Sin 17(04):452–460 (in Chinese)
10. Wang DC, Fang TJ, Tang Y et al (2003) Review of support vector machines regression theory and control. Pattern Recog Artif Intell 02:192–197 (in Chinese)
11. Wang B (2006) Identification and control of nonlinear systems with hysteresis characteristics based on support vector machine. Zhejiang University, pp 8–10 (in Chinese)
12. Shi ZW, Wang ZH, Li JC (2012) The research of RBFNN in modeling of nonlinear unsteady aerodynamics. Acta Aerodyn Sin 30(1):108–112 (in Chinese)
13. Gong Z, Shen HL (2008) Structure self-adapting ANN method in modeling of unsteady aerodynamics. Flight Dyn 25(4):13–16 (in Chinese)
14. Sinha M, Kuttieri RA, Ghosh AK et al (2013) High angle of attack parameter estimation of cascaded fins using neural network. J Aircr 50(1):272–291

Chapter 63
Reconstruction of Temperature Field from Sound Travel-Time Measurements: With or Without Considering the Bending of Sound Wave Paths

Hua Yan, Shuang Li and Xiaoning Wang

Abstract Commonly used acoustic reconstruction algorithms for temperature fields assume that the sound wave paths between transmitter and receiver pairs are straight. When a sound wave passes through a nonuniform temperature field, the sound wave path will bend due to refraction. A comparison study on reconstruction accuracy with and without considering the bending of sound wave paths is conducted in this paper. First, obtain a reconstruction temperature field without considering the bending of sound wave paths. Then, find the sound ray paths in the reconstructed field by developing triangular method and acquire a reconstruction field with considering the bending of sound wave paths. Various single hot temperature fields are reconstructed from sound travel-time simulation data. Suggestions are given based on the analysis of reconstruction accuracy and feature of temperature fields.

Keywords Acoustic tomography · Temperature field · Reconstruction algorithm · Sound refraction

63.1 Introduction

Temperature field reconstruction based on acoustic tomography [1–4] has many advantages such as nondestructive, noncontact sensing and quick in response. It uses the dependence of sound speed in materials on temperature along the sound propagation path. Acoustic pyrometers are its representative application in industry [4], whereas monitoring the temperature distribution of stored grain by acoustic tomography [5, 6] is a new application research being explored. In stored grain, sound is transmitted principally through the gas in the narrow passageways between

H. Yan (✉) · S. Li · X. Wang
School of Information Science and Engineering, Shenyang University of Technology, Shenyang, China
e-mail: yanhua_01@163.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_63

the grain kernels [7]. Therefore, the temperature distribution of stored grain could be monitored by acoustic tomography.

When a sound wave passes through nonuniform temperature fields, the sound wave path will bend due to refraction. However, commonly used acoustic reconstruction algorithms for temperature fields ignore such effect. In this paper, a reconstruction algorithm taking into account the bending of sound wave paths is proposed. A comparison study on reconstruction accuracy with and without considering the bending of sound wave paths is conducted. And suggestions are given based on the analysis of reconstruction accuracy and feature of temperature fields.

63.2 Theory of Acoustic Temperature Field Reconstruction

Temperature measurement by acoustic method is based on the principle that the sound velocity in a medium is a function of the medium temperature. The sound velocity c in a gaseous medium at an absolute temperature T is given by [1, 6]

$$c = z\sqrt{T} \tag{63.1}$$

where z is a constant decided by gas composition. To survey the temperature distribution in a space by acoustic tomography, several acoustic transceivers have to be installed on its periphery. Using a proper reconstruction algorithm, the sound velocity distribution in the space can be obtained from the travel-time measurements of sound signals between the transmitter and receiver pairs traveling through the space. Then the temperature distribution can be calculated by using Eq. (63.1).

63.3 Reconstruction Algorithm

The reconstruction algorithm used in this paper can be stated as follows.

Assuming the distribution of temperature is $T(x, y)$, the distribution of reciprocal of sound velocity is $f(x, y)$, we have

$$T(x, y) = 1/[z \cdot f(x, y)]^2 \tag{63.2}$$

$$t_k = \int_{l_k} f(x, y)\mathrm{d}l \quad (k = 1, 2, \ldots N) \tag{63.3}$$

where t_k is the sound travel-time of the kth sound path, N is the number of the sound travel paths.

Divide the measurement space into M grids (pixels) and express $f(x, y)$ as follows:

$$f(x,y) = \sum_{m=1}^{M} \varepsilon_m \frac{1}{\sqrt{(x-x_m)^2 + (y-y_m)^2} + G} \tag{63.4}$$

where (x_m, y_m) is the center coordinates of the mth grid, G is the shape parameter of basic functions. In this paper, G = 60, N = 24, M = 100.

Substituting Eq. (63.4) into (63.3) and defining

$$\begin{aligned} t_k &= \sum_{m=1}^{M} \varepsilon_m \int_{l_k} \frac{1}{\sqrt{(x-x_m)^2 + (y-y_m)^2} + G} dl, \\ a_{km} &= \int_{l_k} \frac{1}{\sqrt{(x-x_m)^2 + (y-y_m)^2} + G} dl \\ \boldsymbol{t} &= (t_1, \ldots, t_N)^T, \boldsymbol{A} = (a_{km})_{k=1,\ldots,N,m=1,\ldots,M}, \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M) \end{aligned} \tag{63.5}$$

we have

$$\mathbf{t} = \mathbf{A}\varepsilon \tag{63.6}$$

The regularization solution of equation Eq. (63.6) can be expressed as

$$\boldsymbol{\varepsilon} = \sum_{i=1}^{r} \left(\frac{\sigma_i^2}{\sigma_i^2 + \mu} \right) \frac{\boldsymbol{u}_i^T \boldsymbol{t}}{\sigma_i} \mathbf{v}_i \tag{63.7}$$

where μ is regularization parameter, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ are the singular values of model matrix $\boldsymbol{A}$, r is the number of nonzero singular values, $\boldsymbol{u}_\mathrm{i}$ and $\boldsymbol{v}_\mathrm{i}$ are the left and right singular value vectors of $\boldsymbol{A}$. Because of the ill-posedness of the inverse problem, the condition number of A is very large, that means very small singular values exist. A proper small positive regularization parameter can restrain errors of the solution effectively. μ controls the weight of measured data and experience in solution. If u is too small, the error of measured data cannot be well restrained, while if μ is too large, the solution will lose much detailed information. Usually, the regularization parameter is chosen by experience. In this paper μ = 0 since the travel-time simulation data used are noise free. When model matrix $\boldsymbol{A}$ is determined, the parameter vector $\boldsymbol{\varepsilon}$ can be calculated by using measured or simulated sound travel-time vector $\boldsymbol{t}$ and Eq. (63.7). Substituting $\boldsymbol{\varepsilon}$ into Eq. (63.4), the reciprocal distribution of sound velocity can be calculated, and the temperature distribution in free space can be obtained from Eq. (63.2).

The acoustic reconstruction method of 2D temperature fields considering the bending of sound wave paths proposed in this paper can be divided into three steps as follows:

Step 1. Reconstruct a temperature field from the sound travel-time vector $\boldsymbol{t}$ under the assumption that the sound wave paths are straight. The model matrix $\boldsymbol{A}$ used in this step can be pre-calculated on the basis of the locations of the transceivers and the coordinates of the grid centers.
Step 2. Trace the ray paths between the transmitter and receivers pairs in the reconstruction field by using the triangular forward deployment method.
Step 3. Recalculate the model matrix $\boldsymbol{A}$ based on the ray paths obtained in step 2. Then reconstruct a temperature field taking into account the bending of sound rays from the new model matrix $\boldsymbol{A}$.

A numerical method called acoustic ray tracing or ray tracing is one of the procedures for carrying out an analysis on the effect of refraction in wave propagation in a medium with a continuously varying index of refraction or acoustic velocity. In conventional acoustic ray tracing, the solution of the differential equation, called the acoustic ray equation, is evaluated numerically. Yamamoto et al. proposed a simple acoustic ray tracing method, the triangular forward development method, in a discrete acoustic field for application to such inverse problems as ultrasonic CT. In Yamamoto's method, the acoustic ray is formed geometrically by a linear approximation of the acoustic field in a triangular region which is deployed in front of the wave. Therefore, the algorithm becomes extremely simple and can be evaluated at a high speed [8].

63.4 Reconstruction with and Without Considering the Bending of Sound Wave Paths

In this paper, the space to be measured is assumed to be a square with 10 m long. Eight acoustic transceivers are amounted on the periphery. In order to make a comparison study on the reconstruction accuracy with and without considering the bending of sound wave paths, 48 one-peak temperature fields are reconstructed from simulated travel-times. The temperature distribution can be expressed as follows:

$$T(x,y) = a \cdot e^{-\frac{(x-p)^2+(y-p)^2}{f}} \cdot e^{-\frac{(x-p)^2+(y-p)^2}{10}} + 285 \tag{63.8}$$

Fixing $f = 200$, $p = 0$ and letting $a = 10, 30, 50, 70, 90, 110, 130, 150$, we get the 1–8th temperature fields. Fixing $f = 200$, $p = 1$ and changing a as above, we get 9–16th fields. Fixing $f = 200$, $p = 2$ and changing a as above, we get 17–24th fields. Then, fixing $f = 10$ and changing p and a as above, we get 24–48th fields. Parameter f controls the steepness of the temperature distribution. The smaller parameter f, the steeper the temperature change. Parameter a controls the difference between the maximum and minimum temperature in the field. The smaller parameter a, the smaller the difference. Parameter p controls the distance between the space center and the peak. The smaller parameter p, the closer the peak to the center.

In order to evaluate the accuracy of the reconstruction field, the root-mean-squared error E_{rms} is used which is defined as follows:

$$E_{\text{rms}} = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}[T(j) - \hat{T}(j)]^2}}{T_{\text{mean}}} \times 100\% \tag{63.9}$$

where $\hat{T}(j)$ and $T(j)$ are the reconstructed and the true temperature of mth pixel, respectively; T_{mean} is the true mean temperature, M is the number of pixel.

Above 48 temperature fields are reconstructed with and without considering the bending of sound wave paths. The travel-times are computer simulation values calculated using temperature field models, the locations of the transceivers and the relationship between the sound velocity and the temperature in gas. The root-mean-squared errors of these 48 temperature fields with and without considering the bending of the sound wave paths are given in Fig. 63.1. Four fields, that is the 3rd, 13th, 32nd, and 37th fields are selected as examples to demonstrate the detail of the comparison study.

The mathematical description of 3rd temperature field is given in Eq. (63.10). It has a peak located at (0, 0). The difference between the maximum and minimum temperatures in the field is about 43 K.

$$T(x, y) = 50 \cdot e^{-\frac{x^2+y^2}{200}} \cdot e^{-\frac{x^2+y^2}{10}} + 285 \tag{63.10}$$

The mathematical description of 13th temperature field is given in Eq. (63.11). The peak is located at (1, 1). The difference between the maximum and minimum temperatures in the field increases to about 82 K either.

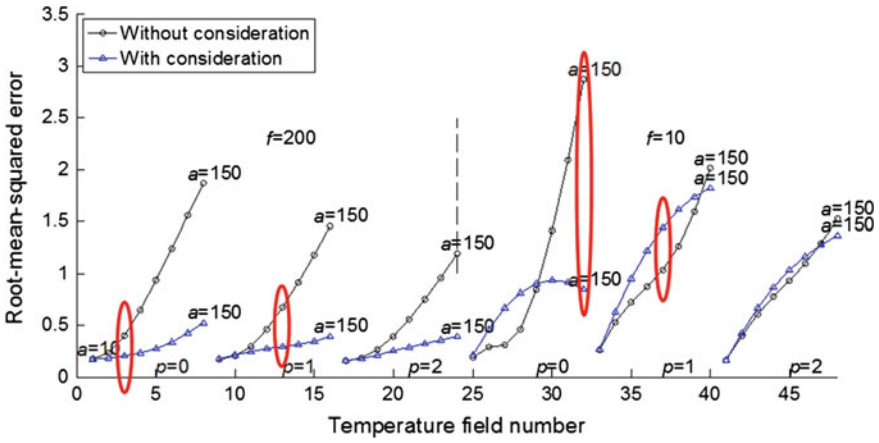


Fig. 63.1 The root-mean-squared errors of the 48 temperature fields with and without considering the bending of sound wave paths

$$T(x,y) = 90 \cdot e^{-\frac{(x-1)^2+(y-1)^2}{200}} \cdot e^{-\frac{(x-1)^2+(y-1)^2}{10}} + 285 \tag{63.11}$$

The mathematical description of 32nd temperature field is given in Eq. (63.12). It has a steep peak located at (0, 0). The difference between the maximum and minimum temperatures in the field is about 122 K.

$$T(x,y) = 150 \cdot e^{-\frac{x^2+y^2}{10}} \cdot e^{-\frac{x^2+y^2}{10}} + 285 \tag{63.12}$$

The mathematical description of 37th temperature field is given in Eq. (63.13). It has a steep peak located at (1, 1). The difference between the maximum and minimum temperatures in the field is about 81 K.

$$T(x,y) = 90 \cdot e^{-\frac{(x-1)^2+(y-1)^2}{200}} \cdot e^{-\frac{(x-1)^2+(y-1)^2}{10}} + 285 \tag{63.13}$$

Their model fields, reconstruction fields with and without considering the bending of sound wave paths, and sound paths are given in Fig. 63.2. The sound paths are traced in the reconstructed field without considering the effect of sound refraction. Following can be seen from Figs. 63.1 and 63.2.

If the peak is not abrupt, such as the 1–24th temperature fields in this paper, the reconstruction accuracy can be effectively increased by considering the bending of sound wave paths. Usually, the bigger the temperature difference of the field, the more the improvement. In addition, the influence of peak position on reconstruction accuracy is not evident.

If the peak is abrupt and located at or near the center of the measured space, but the temperature difference of the field is not big, such as the 25–29th fields in this paper, the reconstruction accuracy with consideration is near or slightly lower than that without consideration.

If the peak is abrupt and located at or near the center of the measured space, and the temperature difference of the field is big, such as the 30–32nd temperature fields in this paper, the reconstruction accuracy can be obviously increased by considering the bending of sound wave paths. The bigger the temperature difference of the field, the more the improvement.

If the peak is abrupt and farther away from the center of the measured space, such as the 33–48th fields in this paper, the reconstruction accuracy with consideration is near or slightly lower than that without consideration.

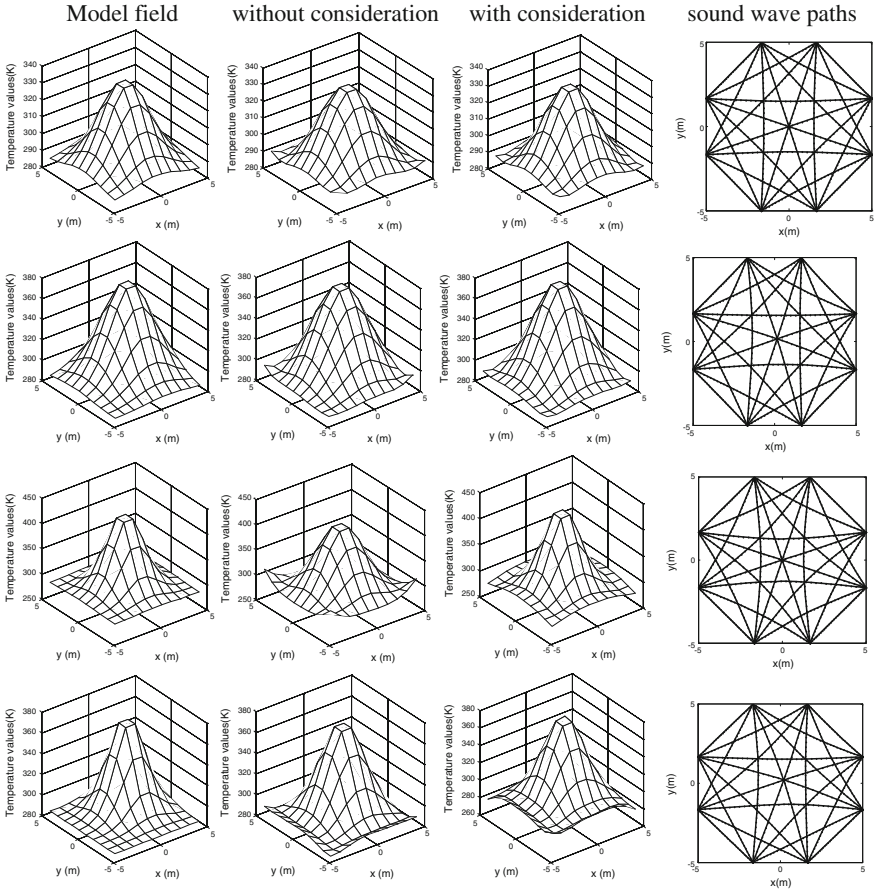


Fig. 63.2 The model fields, reconstruction fields with and without considering the bending of sound wave paths, and the sound wave paths. From top to bottom are the 3rd, 13th, 32nd, and 37th temperature fields

63.5 Conclusion

Commonly used acoustic reconstruction algorithms for temperature fields assume that the sound wave paths are straight. When a sound wave passes through a nonuniform temperature field, the sound wave path will bend due to refraction.

Following suggestions are given based on the analysis of reconstruction accuracy and feature of temperature fields.

(1) If the peak is not abrupt, the bending of sound wave paths should be considered.

(2) If the peak is abrupt and located at or near the center of the measured space, and the temperature difference of the field is big, the bending of sound wave paths should be considered.
(3) If the peak is abrupt and located at or near the center of the measured space, but the temperature difference of the field is not big, the bending of sound wave paths should not be considered.
(4) If the peak is abrupt and farther away from the center of the measured space, the bending of sound wave paths should not be considered.

Acknowledgments The authors thank the National Natural Science Foundation of China (No. 60772054, 61372154), the Program for Liaoning Excellent Talents in University (No. LR2013005) for supporting this research.

References

1. Bramanti M, Salerno EA, Tonazzini A et al (1996) An acoustic pyrometer system for tomographic thermal imaging in power plant boilers. IEEE Trans Instrum Meas 45:159–167
2. Holstein P, Raabe R, Müller R et al (2004) Acoustic tomography on the basis of travel-time measurement. Meas Sci Technol 15:1240–1248
3. Barth M, Armin R (2011) Acoustic tomographic imaging of temperature and flow fields in air. Meas Sci Technol 22:1–13
4. Wei F, Chen Y, Pan HC et al (2010) Experimental study on underwater acoustic imaging of 2-D temperature distribution around hot springs on floor of Lake Qiezishan China. Exp Thermal Fluid Sci 34:1334–1345
5. Yan H, Chen GN, Zhou YG et al (2012) Experimental study of sound travel-time estimation method in stored grain. J Comput 7:947–953
6. Yan H, Chen GN, Zhou YG et al (2012) Primary study of temperature distribution measurement in stored grain based on acoustic tomography. Exp Thermal Fluid Sci 42:55–63
7. Hickling R, Wei W, Hagstrum DW (1997) Studies of sound transmission in various types of stored grain for acoustic detection of insects. Appl Acoust 50:263–278
8. Hiroshi Y, Makoto T, Mitsuhiro U (1991) Acoustic ray tracing in discrete wave velocity by deployed triangle. Electron Commun Jpn 74(2):59–69

Erratum to: A Prediction Method for Wind Speed Based on the Correlation Analysis of Measured Data of Adjacent Wind Turbine

Yinsong Wang and Ziqing Su

Erratum to:
Chapter 9 in: Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336, DOI 10.1007/978-3-662-46469-4_9

The chapter author Yinsong Wang, affiliation was incorrect. The correct affiliation is given below:

North China Electric Power University, Baoding, China.

The online version of the original chapter can be found under DOI 10.1007/978-3-662-46469-4_9

Y. Wang (✉) · Z. Su
North China Electric Power University, Baoding, China
e-mail: 920028274@qq.com

Z. Deng and H. Li (eds.), *Proceedings of the 2015 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 336,
DOI 10.1007/978-3-662-46469-4_64