Algorithms for Acceleration of Image Processing at Automatic Registration of Meeting Participants

Alexander Ronzhin[1], Irina Vatamaniuk[2], Andrey Ronzhin[1,3], and Miloš Železný[4]

[1] SPIIRAS, 39, 14th line, St. Petersburg, 199178, Russia
[2] SUAI, 67, Bolshaya Morskaia, St. Petersburg, 199000, Russia
[3] ITMO University, 49 Kronverkskiy av., St. Petersburg, 197101, Russia
[4] University of West Bohemia, Pilsen, Czech Republic
{ronzhinal,vatamaniuk,ronzhin}@iias.spb.su
zelezny@kky.zcu.cz
www.spiiras.nw.ru/speech

Abstract. The aim of the research is to develop the algorithms for acceleration of image processing at automatic registration of meeting participant based on implementation of blurriness estimation and recognition of participants faces procedures. The data captured by the video registration system in the intelligent meeting room are used for calculation variety of person face size in captured image as well as for estimation of face recognition methods. The results shows that LBP method has highest recognition accuracy (79,5%) as well as PCA method has the lowest false alarm rate (1,3%). The implementation of the blur estimation procedure allowed the registration system to exclude 22% photos with insufficient quality, as a result the speed of the whole system were significantly increased.

Keywords: face recognition, intelligent meeting room, rapid image processing, image blur estimation, local binary patterns.

1 Introduction

Currently, there is a wide range of intelligent spaces prototypes (e.g. intelligent meeting rooms [1,2,3]). For developing services of intelligent room the basic information about the room, meeting participants (number of people, their identities, location, etc.) and their interaction should be acquired. For analyzing of acquired data the methods of audio and video signal processing are used: location and tracking [4,5,6], speech recognition [7], estimation of head orientation and face recognition [8], etc. Implementation of such methods provides valuable data that can be used for annotation of meetings, as well as to provide necessary context data to build real-time support services in intelligent room [9]. Such services require identification of meetings participants without any constraints on head-pose, illumination, and other uncontrolled condition [10]. In addition, the face resolution varies and generally is low according to the distance between camera and participant. Such problem may be decided by usage of video data from multiple views, provided by several cameras.

This paper is organized as follows. Section 2 discusses the methods of biometric identification based on face recognition. Section 3 describes the specifics of the

A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, pp. 89–96, 2014.
© Springer International Publishing Switzerland 2014

developed method of automatic registration of meeting participants. Section 4 presents the experiments, conditions and results.

2 Biometric Identification Methods Based on Face Recognition

The most common biometric identification technologies are fingerprint and retina, voice, signature, face, palm and finger geometry recognition [11]. From commercial point of view the popularity of biometric identification technology is distributed as follows [12]: fingerprint recognition takes the first place, occupying 54% of the market. Second place is taken by hybrid technology (14.4%), where several types of biometric data are simultaneously used for user recognition. Face recognition technology (11.5%) takes third place, further recognition technologies for palm geometry (10%), retinal (7.3%), votes (4.1%), and signature (2.4%), typing rhythm (0.3%) are arranged. First contactless biometric systems were based on the text-dependent methods of speaker determining, superior systems identify the speaker's voice on any phrase with duration sufficient for decision making. In multimodal biometric systems, the analysis of the speaker position, his/her head location, changing the geometry of the face, its brightness and other parameters [12] are estimated.

Face recognition technology is a well compromise between security and convenience for a user, as well as between the security and confidentiality of personal data. In most cases, at face recognition two main steps are carrying out: 1) detecting the position of the user's face in the image with a simple or complex background [13]; 2) analysis of the facial features to identify the user. Detection of the presence and position of a user's face is carried out by analyzing the pixels belong to the foreground (face area) and background of the image [14]. On images, where the background is clean, i.e. uniform and solid, not difficult to detect the area of the face, but when it is composed of several layers, which are attended by other objects, this problem becomes quite complex. Typically, methods based on identifying the key points of the face, such as eyes, nose, lips, or analyzing the color space of the image, as well as methods of using other features of the face are used for detection of a face region. After segmentation of a face region, is necessary to perform normalization of parameters such as size, orientation, brightness and other characteristics. Image normalization is important to identify the key points of the face, relative to which correction of parameters will be performed. Only after the normalization procedure the procedure for calculating features and generation of personality biometric template, which is stored in the database, be performed.

Today, the most common methods of face recognition are the principal component analysis (PCA) [15], the Fisher linear discriminate analysis (LDA) [16] and the local binary patterns (LBP) [17]. The application of the considered methods can improve the efficiency of face recognition in the case of a small number of training samples and in the processing of digital images of large size under a small sample, as well as reduce the dimension of the facial features and improve the speed of image processing.

2.1 Principal Component Analysis

The purpose of face recognition system is the division of the input signals (image files) into several classes (users) [18]. The application of such systems is relevant for a wide range of tasks: images and movies processing, human-computer interaction, identification of criminals, etc. Input signals may contain a lot of noise due to various conditions such as lighting, users pose, their emotions, different hairstyles. However, the input signals are not completely random and even more common features partly present in each incoming signal. Among the common features in input data the following objects can be observed: eyes, mouth, nose, and the relative distances between these objects. Such common features in the fields of research on face recognition are called eigenfaces [15] or principal components. The subspace of eigen features in the image is calculated by the following formula: $\hat{\Phi} = \sum_{i=1}^{M} w_i u_i, (w_i = u_i^\top \Phi)$, where Φ is an input image, $w_i u_i$ are the feature vectors; M is the total amount of images in training database.

After that, the resulting matrix $\hat{\Phi}$ is converted into the eigenvectors of the covariance matrix C corresponding to the original face images:

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^\top.$$

Further the calculation of the Mahalanobis distance is performed by formula:

$$e^2 = \sum_{i=1}^{M} \frac{1}{\lambda_i} (w_i - w_i^k)^2,$$

where k is a number of used eigenfaces, λ is a scalar of eigenvalues.

After the calculation of e^2 value, its comparison with the pre-selected threshold is performed for the belongingness definition of the analyzed face to users faces, which are added to the training database.

2.2 Fisher Linear Discriminate Analysis

The main idea of LDA is to find such a linear transformation to separate features clusters after transformation [19], which is achieved due to the scattered matrix analysis [20]. This method selects M-class scatter for matrix S_b and S_w between- and within-classes as follows:

$$S_b = \sum_{i=1}^{M} Pr(C_i)(\mu_i - \mu)(\mu_i - \mu)^\top, S_w = \sum_{i=1}^{M} Pr(C_i) \sum i,$$

where $Pr(C_i)$ is a priori probability of class C_i, which takes the value $\frac{1}{M}$ with the assumption of equal priori probabilities; μ - the overall average vector; $\sum i$ average scatter of sample vectors of different classes C_i around their representation in the form of the mean vector μ_i.

Distribution of class features can be calculated using the ratio of the scatter matrices S_b and S_w determinants: $J(A) = \arg\max_A \frac{A S_b A^\top}{A S_w A^\top}$, where A is a matrix with size $m \times n$, where $m \leq n$.

For optimization of the previous formula the approach described in the paper [20] is used, as a result the formula becomes: $S_b A = \lambda S_w A$, where λ is the largest generalized eigenvalues of S_w.

Solution of previous equation is to calculate the inverse matrix S_w, solution of the eigenvalue problem for the matrix $S_w^{(-1)} S_b$ described in [20]. However, this method is numerically unstable since performs a direct appeal to the high dimension matrix of probabilities. In practice, the most commonly used algorithm LDA is based on finding matrix A, which can simultaneously diagonalized matrix S_b and S_w: $AS_wA^\top = I, AS_bA^\top = \Lambda$, where Λ is a diagonal matrix with elements sorted in descending order.

2.3 Local Binary Patterns

One of the first researches dedicated to texture description based on LBP is [21]. The LBP operator describes a pixel 3x3-neighborhood in the binary form. When neighbor's value is greater than the center pixel's one, its labeled "1". Otherwise, its labeled "0". This gives an 8-digit binary number. The LBP is considered uniform pattern, if it contains two or less bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. The LBP operator can be extended to use circular neighborhoods of any radius and number of pixels by bilinear interpolation of pixel values.

A histogram of the labeled image $f_l(x, y)$ contains information about the distribution of the local micro patterns, such as edges, spots and flat areas, over the whole image. It can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x, y) = i\}, i = 0, 1, \dots, n-1,$$

in which n is a number of different labels produced by the LBP operator, and IA is 1 if A is true and 0 if A is false. For efficient face representation, three different levels of locality are used: a pixel-level (labels for the histogram), a regional level (the labels sum over a small region), and a global description of the face (the regional histograms). For this purpose the image is divided into regions $R_0, R_1, \dots, R_{(m-1)}$ and the spatially enhanced histogram is defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x, y) = i\} I\{(x, y) \in R_j\}, i = 0, 1, \dots, n-1, j = 0, 1, \dots, m-1.$$

Obviously, some of the regions contain more information, useful for face recognition than others. Taking this into account, each region is assigned a weight depending on the importance of the information it contains. The described methods of face recognition used in conjunction with blur estimation methods in the system of preparation and support of meetings carried out in the smart room speed up the registration process.

3 Automatic Participant Registration System

The developed automatic participant registration system has two algorithms for face capture with different quality. At the first one the rapid procedure of face recognition is used. It based on capture one photo, which includes the view of all participants with low resolution and following face recognition. At this stage the image patches with participant faces has resolution around of 30x30 pixels. The faces unrecognized during the first level of processing further are separately captured by pan-tilt-zoom camera with high resolution at the second algorithm of registration system work. At that the captured face region has resolution higher than 200x200 pixels.

There are two algorithms for image capturing and processing. In the first algorithm a high-resolution image (1280x1024 pixels), which is a group photo of sitting participants in chairs, is processed for finding their faces by face detection procedure [22]:

$$D_{v\_h}^{roi} = \begin{cases} 1, \sum_{n=1}^{N} a_n h_n(I) \geq \frac{1}{2} \sum_{n=1}^{N} a_n, \\ 0, \text{otherwice} \end{cases}$$

where h_n is a set of features $n = 1, 2, \ldots, N$, which are used in the Haar cascades; a_n - features weight coefficient, I - input image.

To reduce image processing time for each chair area (k) with the possible appearance of a participant's face was determined. Due to the fact that chairs have a static position, this procedure is performed once for the selected configuration of the chairs, as follows:

```
{The beginning of the first algorithm}
for(k=0;k < Number_of_Chairs; k++)
    FaceRegion[k] = FaceDetection (area[k]);
    if(FaceRegion[k])
        FaceRecognition(FaceRegion);
        SaveImage(FaceRegion);
    end if
end for
{The beginning of the second algorithm}
for(i=0;i < Number_of_Unregistered_Participants; i++)
    FaceRegion = FaceDetection (InputImage);
    if(FaceRegion)
        blurriness estimation (InputImage(FaceRegion));
{The blurriness estimation procedure executes blurred photos from future processing}
        if(FaceRegion not blurred)
            FaceRecognition(FaceRegion);
            SaveImage(InputImage);
        end if
    end if
end for
```

Each founded region is processed by face recognition function, which identifies all participants with low charge of computational resource and time, because the average size of such region is around of 30x30 pixels. The second algorithm is aimed to identify unregistered participants, which faces haven't been recognized by previous algorithm. The first step of this algorithm is capturing close-up photo of participant with resolution 704x576 pixels. For blur estimation the NIQE (No-Reference Image Quality Assessment) [23] method was implemented. Such blurriness happens when participant moves in photographing moment or may be particularly closed by other participant or if camera, from which algorithm receives image, still haven't focused. If the participant hasn't been identified, then the procedure for registration of new participant is started, where his/her photo is used for focusing attention on the audiovisual speech synthesis system.

Introduction of the blur estimation procedure as preliminary stage of photo processing allows the registration system to exclude 22% photos with high resolution but insufficient quality from face recognition stage, as a result the speed of the whole system were significantly increased. Implementation of the blur estimation procedure on the first level of processing of the photos with resolution around 30x30 pixels did not give positive results, because such low resolution is insufficient to make decision about image blurriness.

4 Experiments

For the experimental evaluation of a method of automatic registration of participants during the events in the intelligent meeting room the accumulation of participants photos was produced only at the first algorithm of the system. As a result, the number of accumulated photos was more than 55,920 for 36 participants. The training database contains 20 photos for each participant. At the preparatory stage of experiments have been decided to determine the threshold for the three face recognition methods: 1) PCA; 2) LDA; 3) LBP. During this experiment threshold was calculated for each participant added to the recognition model, a maximum value of a correct recognition hypothesis for the LBP ranged from 60 to 100, for the PCA from 1656 to 3576, for the LDA from 281 to 858. As a consequence, for the further experiments were selected the minimum threshold value - 60, 1656 and 281, for these methods, respectively. Table 1 presents the average values of face recognition accuracy, as well as first (False Alarm (FA) rate) and second (Miss rate (MR)) type errors for each selected method.

Table 1. Average values of face recognition accuracy, FA and MR

Method	FA, %	MR, %	Accuracy, %
LBPH	12	8,5	79,5
PCA	1,3	23,5	75,2
LDA	19,2	7,8	73

The high value of false positives and miss rate errors was due to the fact that the photos were stored in the course of actual of events, without a prepared scenario and focusing participants on a single object. Hereupon at the time of photographing participants can move freely, according to their face in the photos could be blurred or partially hidden.

For estimation of influence of participants face size change on recognition rate was decided to divide them into several groups. Figure 1a shows distribution of participants by variety difference of their face sizes in ranges from 0 to 10, from 10 to 20, and so on. Most of participants have difference between minimum and maximum face size in range from 30 to 40 pixels. Figure 1b shows distribution of recognition rate for three methods for the groups of participants. From figure 1b it is obvious that with increasing of the

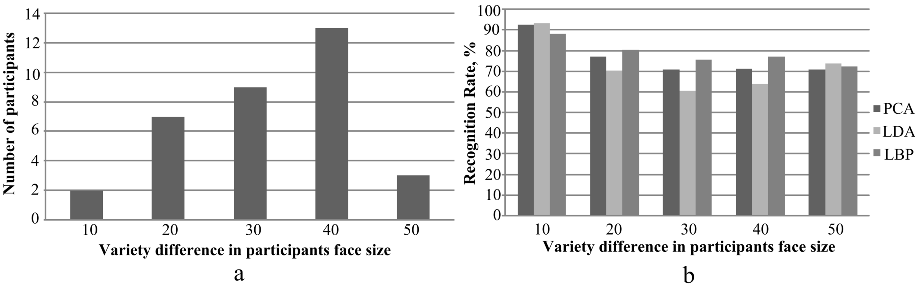

Fig. 1. Distribution of participants and recognition rate by variety difference of their face sizes

participant's face size variety difference the recognition accuracy gradually decreases, this is due to the fact that, at the normalization of images to a uniform size distortion in the certain facial features like eyes, nose, mouth may occur.

Considering the experimental conditions (different distances from the camera to a participant, lighting, movement of participants while taking pictures), influencing on quality and quantity of extracted facial features from the image, which are directly influenced on the accuracy of recognition and occurrence of false positives, we can conclude that the best results are shown by method LBP 79,5%.

5 Conclusion

Development of intelligent meeting rooms, as well as realization meetings support services based on natural and unobstructive method of interaction between users and intelligent space is a relevant subject of research. Application of biometric identification technology based on face recognition methods provides automation of registration processes of meeting participants.

During the research an algorithm for acceleration of image processing at automatic registration of meetings participants personalities based on blurriness estimation and recognition of participants faces. Experimental estimation of the face recognition methods was provided on database with more than 55 thousands photos for 36 participants. During experiments three face recognition methods LBPH, PCA and LDA were compared. The results shows that LBP method has highest recognition accuracy (79,5%) as well as PCA method has the lowest false alarm rate (1,3%).

Acknowledgments. This research is supported by the Scholarship of the President of Russian Federation (Project no. СП-1805.2013.5), Government of the Russian Federation, Grant 074-U01, and European Regional Development Fund (ERDF), project New Technologies for Information Society (NTIS), European Centre of Excellence, ED1.1.00/02.0090.

References

1. Yusupov, R.M., Ronzhin, A.L., Prischepa, M., Ronzhin, A.L.: Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall. ARC 72(7), 1389–1397 (2011)

2. Aldrich, F.: Smart Homes: Past, Present and Future. In: Harper, R. (ed.) Inside the Smart Home, pp. 17–39. Springer, London (2003)
3. Nakashima, H., Aghajan, H.K., Augusto, J.C.: Handbook of Ambient Intelligence and Smart Environments. Springer, Heidelberg (2010)
4. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
5. Ronzhin, A., Budkov, V.: Speaker Turn Detection Based on Multimodal Situation Analysis. In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS (LNAI), vol. 8113, pp. 302–309. Springer, Heidelberg (2013)
6. Zhang, C., Yin, P., et al.: Boosting-Based Multimodal Speaker Detection for Distributed Meeting Videos. IEEE Transactions on Multimedia 10(8), 1541–1552 (2008)
7. Karpov, A., Markov, K., Kipyatkova, I., Vazhenina, D., Ronzhin, A.: Large vocabulary Russian speech recognition using syntactico-statistical language modeling. Speech Communication 56, 213–228 (2014)
8. Ekenel, H.K., Fischer, M., Jin, Q., Stiefelhagen, R.: Multi-modal Person Identification in a Smart Environment. In: Proc. of CVPR 2007, pp. 1–8 (2007)
9. Stiefelhagen, R., Bernardin, K., Ekenel, H.K., McDonough, J., Nickel, K., Voit, M., Wlfel, M.: Audio-Visual Perception of a Lecturer in a Smart Seminar Room. Signal Processing 86(12), 3518–3533 (2006)
10. Ekenel, H.K., Fischer, M., Stiefelhagen, R.: Face Recognition in Smart Rooms. In: Popescu-Belis, A., Renals, S., Bourlard, H. (eds.) MLMI 2007. LNCS, vol. 4892, pp. 120–131. Springer, Heidelberg (2008)
11. Schneiderman, H., Kanade, T.: Object detection using the statistic of parts. International Journal of Computer Vision 56(3), 151–177 (2004)
12. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: A survey. Pattern Recognition Letters 28(14), 1885–1906 (2007)
13. Gorodnichy, M.D.: Video-Based Framework for Face Recognition in Video. In: Proc. CRV 2005, pp. 330–338 (2005)
14. Castrillón-Santana, M., Déniz-Suárez, O., Guerra-Artal, C., Hernández-Tejera, M.: Real-time Detection of Faces in Video Streams. In: Proc. of CRV 2005, pp. 298–305 (2005)
15. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proc. of IEEE CVPR, pp. 586–591 (1991)
16. Belhumeur, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. TPAMI 19(7), 711–720 (1997)
17. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
18. Georgescu, D.: A Real-Time Face Recognition System Using Eigenfaces. Journal of Mobile, Embedded and Distributed Systems 3(4), 193–204 (2011)
19. Yang, J., Yu, Y., Kunz, W.: An Efficient LDA Algorithm for Face Recognition. In: The Proc. Sixth Int. Conf. Control, Automation, Robotics and Vision (2000)
20. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press, New York (1990)
21. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. TPAMI 24, 971–987 (2002)
22. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: Proc. of IEEE ICCV, pp. 734–741 (2003)
23. Mittal, A., Soundarajan, R., Bovik, A.C.: Making a 'Completely Blind' Image Quality Analyzer. IEEE Signal Processing Letters 20(3), 209–212 (2013)