Application of Three-Level Handprinted Documents Recognition in Medical Information Systems

Jerzy Sas[1] and Marek Kurzynski[2]

[1] Wroclaw University of Technology, Institute of Applied Informatics,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
jerzy.sas@pwr.wroc.pl

[2] Wroclaw University of Technology, Faculty of Electronics,
Chair of Systems and Computer Networks,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
marek.kurzynski@pwr.wroc.pl

Abstract. In this paper the application of novel three-level recognition concept to processing of some structured documents (forms) in medical information systems is presented. The recognition process is decomposed into three levels: character recognition, word recognition and form contents recognition. On the word and form contents level the probabilistic lexicons are available. The decision on the word level is performed using results of character classification based on a character image analysis and probabilistic lexicon treated as a special kind of soft classifier. The novel approach to combining these both classifiers is proposed, where fusion procedure interleaves soft outcomes of both classifiers so as to obtain the best recognition quality. Similar approach is applied on the semantic level with combining soft outcomes of word classifier and probabilistic form lexicon. Proposed algorithms were experimentally applied in medical information system and results of automatic classification of laboratory test order forms obtained on the real data are described.

1 Introduction

Automatic analysis of handwritten forms is useful in such applications where direct information insertion into the computer system is not possible or inconvenient. Such situation appears frequently in hospital medical information systems, where physicians or medical staff not always can enter the information directly at the system terminal. Form scanning is considered to be especially useful in laboratory support software, where paper forms are still frequently used as a medium for laboratory test orders representation. Hence, in many commercially available medical laboratory systems a scanning and recognition module is available.

Typical form being considered here has precisely defined structure. It consists of separated data fields, which in turn consist of character fields. In our approach we assume that the whole form contents describes an object from the finite set of items and the ultimate aim of form recognition is selecting of relatively small

J.L. Oliveira et al. (Eds.): ISBMDA 2005, LNBI 3745, pp. 1–12, 2005.

subset of objects. Therefore, instead of using the classic pattern recognition approach consisting in indicating a single class, we will apply “soft” recognizer ([3]) which fetches the vector of soft labels of classes, i.e. values of classifying function.

In order to improve the overall form recognition quality, compound recognition methods are applied. Two most widely used categories of compound methods consist in combining classifiers based on different recognition algorithms and different feature sets ([4]). Another approach divides the recognition process into levels in such a way, that the results of classification on lower level are used as features on the upper level ([2]). Two-level approach is typical in handwriting recognition, in which the separate characters are recognized on the lower level and next on the upper level the words are recognized, usually with the use of lexicons.

In this paper, the method which uses both classifier combination and multilevel recognition is described. Probabilistic properties of lexicon and character classifier are typically used to build Hidden Markov Model(HMM) of the language ([11]). We propose another approach to the word recognition, in which probabilistic lexicon is treated as a special kind of classifier based on a word length, and next result of its activity is combined with soft outcomes of character classifier based on recognition of character image. Soft outcomes of a word classifier can be used next as data for semantic level classifier, which - similarly as previously - combined with object lexicon - recognizes the object described by the whole form.

The contents of the work are as follows. Section 2 introduces necessary background. In section 3 the classification methods on successive levels of object recognition problem are presented and concept of fusion strategies of character-based and lexicon-based classifiers are discussed. The proposed algorithms were practically implemented in application for automatic processing of laboratory test order forms in hospital information system. The system architecture and some implementation details are described in section 4. Results of experiments on proposed method efficiency are presented in section 5

2 Preliminaries

Let us consider a paper form F designed to be filled by handwritten characters. The form consists of data fields. Each data field contains a sequence of characters of limited length coming from the alphabet $\mathcal{A} = \{c_1, c_2, ..., c_L\}$. We assume that the actual length of filled part of data field can be faultlessly determined. The set $\mathcal{A}$ can be different for each field. Typically we deal with fields that can contain only digits, letters or both of them. For each data field there exists a probabilistic lexicon $\mathcal{L}$. Lexicon contains words that can appear in the data field and their probabilities:

$$\mathcal{L} = \{(W_1, p_1), (W_2, p_2), ..., (W_N, p_N)\}, \tag{1}$$

where W_j is the word consisting of characters from $\mathcal{A}$, p_j is its probability and N is the number of words in the lexicon.

The completely filled form describes an object (e.g. a patient in medical applications) and the data items written in the data fields are its attributes. The form contents, after manual verification is entered to the database, which also contains the information about the objects appearance probability. An example can be a medical information system database, where the forms contain test orders for patients registered in the database. The patients suffering from chronic diseases are more frequently examined, so it is more probable that the form being recognized concerns such a patient. Thus, this data base can be treated as a kind of probabilistic lexicon containing objects recognized in the past and the information about probability of its appearance, viz.

$$\mathcal{L}_{\mathcal{B}} = \{(b_1, \pi_1), (b_2, \pi_2), ..., (b_M, \pi_M)\}. \tag{2}$$

Our aim is to recognize the object $b \in \mathcal{L}_{\mathcal{B}}$ on the base of scanned image of a form F and both lexicons (1), (2). The recognition process can be divided into three levels, naturally corresponding to the three-level form structure:

- character (alphabetical) level – where separate characters are recognized,
- word level – where the contents of data fields is recognized, based on the alphabetical level classification results, their probabilistic properties and probabilistic lexicon (1),
- semantic level – where the relations between fields of the form being processed and lexicon (2) are taken into account to further improve the recognition performance.

In the next section the classification methods used on the successive levels of recognition procedure are described in details.

3 Three-Level Form Recognition Method

3.1 Character Recognition on the Alphabetical Level

We assume that on character (alphabetical) level classifier Ψ_C is given which gets character image x as its input and assigns it to a class (character label) c from $\mathcal{A}$, i.e., $\Psi_C(x) = c$. Alternatively, we may define the classifier output to be a L-dimensional vector with supports for the characters from $\mathcal{A}$ ([4]), i.e.

$$\Psi_C(x) = [d_1(x), d_2(x), ..., d_L(x)]^T. \tag{3}$$

Without loss of generality we can restrict $d_i(x)$ within the interval $[0, 1]$ and additionally $\sum_i d_i(x) = 1$. Thus, $d_i(x)$ is the degree of support given by classifier Ψ_C to the hypothesis that image x represents character $c_i \in \mathcal{A}$. If a crisp decision is needed we can use the maximum membership rule for soft outputs (3), viz.

$$\Psi_C(x) = arg\,(\max_i \ d_i(x)). \tag{4}$$

We have applied MLP-based classifier on this level. The vector of support values in (3) is the normalized output of MLP obtained by clipping network output values to $[0, 1]$ range and by normalizing their sum to 1.0.

Independently of nature of classifier Ψ_C, support vector (3) is usually interpreted as an estimate of *posterior* probabilities of classes (characters) provided that observation x is given ([4], [9], [10]), i.e. in next considerations we adopt:

$$d_i(x) = P(c_i \mid x), \quad c_i \in \mathcal{A}. \tag{5}$$

3.2 Data Field Recognition on the Word Level

Let the length $\mid W \mid$ of currently recognized word $W \in \mathcal{L}$ be equal to n. This fact defines the probabilistic sublexicon $\mathcal{L}_n$

$$\mathcal{L}_n = \{(W_k, q_k)_{k=1}^{N_n} : W_k \in \mathcal{L}, \mid W_k \mid = n\}, \tag{6}$$

i.e. the subset of $\mathcal{L}$ with modified probabilities of words:

$$q_k = P(W_k / \mid W_k \mid = n) = \frac{p_k}{\sum_{j:|W_j|=n} p_j}. \tag{7}$$

The sublexicon (6) can be also considered as a soft classifier Ψ_L which maps feature space $\{\mid W_k \mid : W_k \in \mathcal{L}\}$ into the product $[0,1]^{N_n}$, i.e. for each word length n produces the vector of supports to words from $\mathcal{L}_n$, namely

$$\Psi_L(n) = [q_1, q_2, ..., q_{N_n}]^T. \tag{8}$$

Let suppose next, that classifier Ψ_C, applied n times on the character level, on the base of character images $X_n = (x_1, x_2, ..., x_n)$, has produced the sequence of character supports (3) for the whole recognized word, which can be organized into the following matrix of supports, or matrix of *posterior* probabilities (5):

$$D_n(X_n) = \begin{pmatrix} d_{11}(x_1) & d_{12}(x_1) & \dots & d_{1L}(x_1) \\ d_{21}(x_2) & d_{22}(x_2) & \dots & d_{2L}(x_2) \\ \vdots & \vdots & \dots & \vdots \\ d_{n1}(x_n) & d_{n2}(x_n) & \dots & d_{nL}(x_n) \end{pmatrix}. \tag{9}$$

Now our purpose is to built soft classifier Ψ_W (let us call it *Combined Word Algorithm* - CWA) for word recognition as a fusion of activity of both lexicon-based Ψ_L and character-based classifier Ψ_C:

$$\Psi_W(\Psi_C, \Psi_L) = \Psi_W(D_n, \mathcal{L}_n) = [s_1, s_2, ..., s_{N_n}]^T, \tag{10}$$

which will produce support vector for all words from sublexicon $\mathcal{L}_n$.

Let $\mathcal{N} = \{1, 2, ..., n\}$ be the set of numbers of character positions in a word $W \in \mathcal{L}_n$ and $\mathcal{I}$ denotes a subset of $\mathcal{N}$. In the proposed fusion method with "interleaving" first the algorithm Ψ_C applied for recognition of characters on positions $\mathcal{I}$ on the base of set of images $X^{\mathcal{I}} = \{x_k : k \in \mathcal{I}\}$, produces matrix of supports $D^{\mathcal{I}}$ and next - using these results of classification - the lexicon $\mathcal{L}_n$ (or algorithm Ψ_L) is applied for recognition of a whole word W.

The main problem of proposed method consists in an appriopriate division of $\mathcal{N}$ into sets $\mathcal{I}$ and $\bar{\mathcal{I}}$ (complement of $\mathcal{I}$). Intuitively, subset $\mathcal{I}$ should contain these positions for which character recognition algorithm gives the most reliable results. In other words division of $\mathcal{N}$ should lead to the best result of classification accuracy of a whole word. Thus, subset $\mathcal{I}$ can be determined as a solution of an appropriate optimization problem.

Let $W^{\mathcal{I}} = \{c_{i_k} : \ k \in \mathcal{I}, \ c_{i_k} \in \mathcal{A}\}$ be any set of characters on positions $\mathcal{I}$. Then we have following posterior probability:

$$P(W^{\mathcal{I}} \mid X^{\mathcal{I}}) = \prod_{k \in \mathcal{I}} d_{k\, i_k}(x_k). \tag{11}$$

The formula (11) gives conditional probability of hypothesis that on positions $\mathcal{I}$ of word to be recognized are characters $W^{\mathcal{I}}$ provided that set of character images $X^{\mathcal{I}}$ has been observed.

Applying for remaining part of the word sublexicon $\mathcal{L}_n$, we can calculate conditional probability of the whole word $W_j \in \mathcal{L}_n$, which constitutes the support (10) for word W_j of soft classifier Ψ_W:

$$s_j = P(W_j \mid X^{\mathcal{I}}) = P(W^{\mathcal{I}} \mid X^{\mathcal{I}})\, P(W_j \mid W^{\mathcal{I}}). \tag{12}$$

The first factor in (12) is given by (11) whereas the second one can be calculated as follows:

$$P(W_j \mid W^{\mathcal{I}}) = \frac{q_j}{\sum_{j : W_j contains W^{\mathcal{I}}} q_j}. \tag{13}$$

Since the support vector (12) of the rule Ψ_W strongly depends on the set $\mathcal{I}$ hence we can formulate the following optimization problem:

It is neccesary to find such a subset $\mathcal{I}^*$ of $\mathcal{N}$ and such a set of charcters $W^{\mathcal{I}^*}$ which maximize the maximum value of decision supports dependent on sets $\mathcal{I}$ and $W^{\mathcal{I}}$, namely

$$Q(\Psi_W^*) = \max_{\mathcal{I}, W^{\mathcal{I}}} \max_{j=1,2,\ldots,N_n} s_j(\mathcal{I}, W^{\mathcal{I}}). \tag{14}$$

The detailed description of suboptimal solution of the problem (14) which was applied in further experimental investigations can be find in [8].

3.3 Complete Form Recognition on the Semantic Level

For recognition of the whole form (object) on the semantic level we propose procedure called *Combined Semantic Algorithm* (CSA), which is fully analogous to the approach applied on the word level, i.e. relation between word classifier Ψ_W and probabilistic lexicon (2) is exactly the same as relation between the character recognizer Ψ_C and word lexicon (1). In other words, the form lexicon is treated as a special kind of classifier producing the vector of form supports (probabilities)

$$\pi = (\pi_1, \pi_2, ..., \pi_M), \tag{15}$$

which next are combined with soft outcomes (10) of word classifier Ψ_W.

Let suppose that form to be recognized contains K data fields. Recognition of kth data field containing word from sublexicon (6) which length is equal to n_k, has provided the vector of supports (10)

$$\Psi_W(\Psi_C, \Psi_L) = \Psi_W(D_{n_k}, \mathcal{L}_{n_k}) = (s_1^{(k)}, s_2^{(k)}, ..., s_{N_{n_k}}^{(k)})^T, \tag{16}$$

given by formula (12).

Now, repeating recognition method from section 3.2 and optimization of fusion procedure with "interleaving" for support vectors (16) for $k = 1, 2, ..., K$ instead of matrix (9) and probabilities (15) instead of (8) we get according to (11), (12) and (13) support vector which soft classifier on the semantic level Ψ_B gives to forms from the lexicon (2)

$$\Psi_B = (\sigma_1, \sigma_2, ..., \sigma_M). \tag{17}$$

As previously, that σ_i can be interpreted as an estimate of *posterior* probability of the object described by form $b_i \in \mathcal{L}_{\mathcal{B}}$ provided that observation of character images of all data fields $X_B = (X_{n_1}, X_{n_2}, ..., X_{n_K})$ are given and both lexicons (1) and (2) are available, viz.

$$\sigma_i = P(b_i \mid X_B, \mathcal{L}, \mathcal{L}_{\mathcal{B}}). \tag{18}$$

The crisp decision is possible by selection the object b^* from (2) for which support value (probability) σ^* is the greatest one.

In application like the one described here, probabilistic lexicons are derived from the contents of database, where previously recognized and verified forms are stored, It may happen that the object described by a form is not registered in the database yet. Forcing to always recognize one of registered objects would be unreasonable. In particular at the early stages of the recognition system operation, when the database contains few records it cannot be used as a reliable objects lexicon. In our approach, the database is periodically tested in order to estimate the probability P_{new}. P_{new} is the probability that the verified form being entered describes the object not registered in the database yet. Before the new record is entered to the database, it is tested if the object described by the record is already in the database. The new record is appropriately flagged depending on the test result. By analyzing the flags associated with certain number of recently entered records we can estimate P_{new}. If the probability σ^* is greater than P_{new} then b^* found by CSA is the final recognition. Otherwise CSA result is rejected and it is assumed that the object described by the form is not contained in the data base.

4 Application of Three-Level Form Recognition Concept in the Laboratory Orders Form Recognition Module

The concept described in previous sections has been applied to laboratory orders form recognition module in a hospital information system (HIS). In some cases specimens for laboratory test are taken in locations distant from central laboratory, e.g. in outpatient clinic. It is convenient to transfer information about ordered tests in paper form together with specimens. To improve operation of central laboratory, which processes hundreds of orders daily, the method of fast and reliable entering of test orders data into the information system controlling automatic laboratory devices is necessary.

For each specimen, individual order form is filled by the medical assistant in outpatient clinic where specimen is taken. The form for particular specimen is identified by its symbol represented by barcode label stick both to specimen container and to the corresponding form. In the central laboratory the forms are scanned, their contents is recognized and entered into HIS database. Next, each recognized form contents is manually verified by operator by comparing recognized data with the image of the form displayed on the screen. Finally, verified data are used to register test order record in HIS database, where it is later used to control bi-directional laboratory devices. The system architecture and data flow is presented on Fig.1. It consists of the following modules:

form design module – allows system administrator to design various form variants containing required subsets of tests being supported by the laboratory,

scanning module – controls the farm of scanners connected to the system and manages the form images repository,

lexicon extraction module – updates periodically probabilistic lexicons both for word and semantic level using actual contents of HIS database,

recognizer module – performs soft recognition of form images, fetches results of soft recognition of isolated data fields as well as the results of soft identification of patient, for which the form has been prepared,

manual form verification module – provides user interface for thorough verification of form recognition results. Support vectors that are results of soft recognition in recognizer module are used to build ordered list of alternative values for isolated fields and for the whole patient data identification section. They are used as combo boxes connected with corresponding fields in case where manual correction is needed.

The test order form is presented on Fig. 2. It consists of three sections: ordering institution/physician data, patient identification data and ordered tests list. Two-level recognition is applied to ordering institution/physician data because there are no clear relations between data fields in this section and hence third level cannot be defined. Full three-level concept is applied to patient identification data. The patient data contain: name, first name, sex, birth date, social security identifier. Probabilistic lexicons for all data fields in patient section are derived from HIS database contents using lexicon extraction module. In case of date field, probabilistic dictionary is applied only to 4-digit year section. The module updates lexicons periodically (every night) using current database contents.

The ordered laboratory tests are identified by marking check boxes in the lower section of the form. The count of check boxes is limited by the area of test

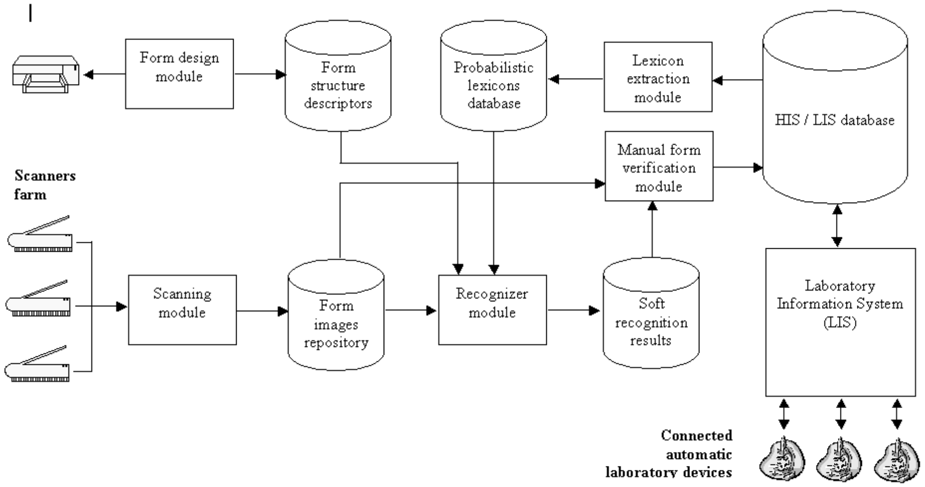
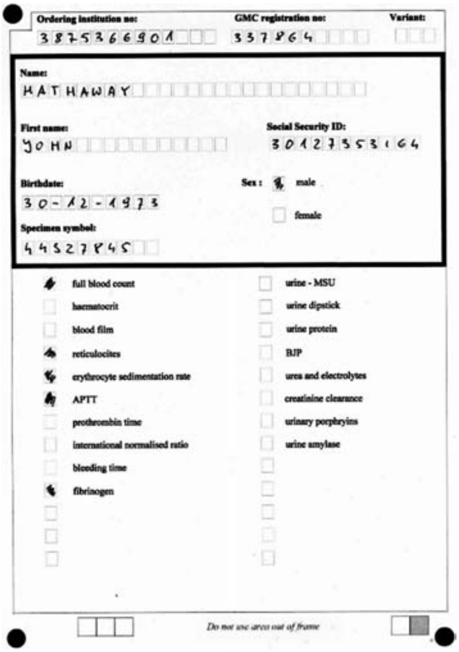

Form design module
Form structure descriptors
Probabilistic lexicons database
Lexicon extraction module
HIS / LIS database
Manual form verification module
Scanners farm
Scanning module
Form images repository
Recognizer module
Soft recognition results
Laboratory Information System (LIS)
Connected automatic laboratory devices

Fig. 1. Form recognition module architecture and data flow



Ordering institution no: 3875366901 GMC registration no: 337864 Variant:

Name: HATHAWAY

First name: JOHN

Social Security ID: 30127353164

Birthdate: 30-12-1973

Sex : ☒ male
□ female

Specimen symbol: 44527845

☒ full blood count
□ haematocrit
□ blood film
☒ reticulocites
☒ erythrocyte sedimentation rate
☒ APTT
□ prothrombin time
□ international normalised ratio
□ bleeding time
☒ fibrinogen
□
□
□

□ urine - MSU
□ urine dipstick
□ urine protein
□ BJP
□ urea and electrolytes
□ creatinine clearance
□ urinary porphryins
□ urine amylase
□
□
□
□
□

Do not use area out of frame

Fig. 2. Laboratory test order form image

selection section on the form and it is significantly lower that the total number of tests, carried out by the laboratory. To assure flexibility of the system, many variants of forms can be designed in the system. Particular form variant can be used to order a subset of tests, usually related each to other, e.g. test in hematology, urine tests, clinical biochemistry etc. The user can define any number of form variants containing tests subsets using form design module. In different form variants the checkbox in given position represents different laboratory tests. Assignment of tests to particular check box positions is stored in variant description data structure used during final form contents interpretation. The form variant is defined by numeric field printed on the form. Correct recognition of form variant is absolutely essential for system usability and even for patient safety. To assure maximal accuracy and human-readability of form variant identification, the numerical variant symbol is pre-printed using fixed font.

4.1 Character Classification and Features Extraction

MLP has been used as the character level soft classifier. Feature extraction and MLP architecture is based on methods described in [1]. Directional features set has been selected as the basis for character classification due to its superior efficiency reported in literature and ease of implementation. The directional features describe the distribution of stroke directions in 25 evenly spaced subareas of the character image. The set of eight direction planes is created. Direction plane is an array of resolution equal to image resolution. Each plane corresponds to one of eight direction sections as shown on Fig. 3. According to the concept described in [1] for each image pixel the Sobel operator on image brightness $g(i,j))$ is calculated giving image brightness gradient. The brightness gradient vector is decomposed into two components g_k and g_{k+1} parallel to lines surrounding its section k. The lengths of components are then added to cells of corresponding direction planes. Finally, each plane is filtered using Gaussian filter resulting in 5x5 grid of values. 200 elements feature vector is built of all of grid values calculated for all plane arrays. MLP with 200 inputs and number of outputs corresponding to count of classes was used as character classifier. Data fields in test order form can be divided into purely numerical or alphabetical.

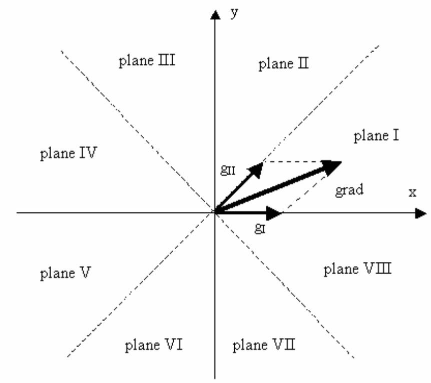


Fig. 3. Laboratory test order form image

We applied three independently trained classifiers for: printed numerical fields (ordering institution ID, form variant), handwritten numeric fields and handwritten alphabetical fields. The count of nodes in hidden layer was determined experimentally by maximizing character recognition accuracy. MLP for handwritten letters recognition contains 80 units in hidden layer while hidden layer of MLP for numerals recognition consists of 60 units.

5 Experiment Results

The aim of experiments was to assess the increase of form recognition accuracy resulting from application of described methods on word and semantic levels.

Character classifiers were trained using sets of character image database collected from almost 300 individuals. The training set for letters contained 9263 images. For numerals classifier training the set of 3130 images was used. For assessment of isolated characters classification accuracy, the characters extracted from real laboratory test order forms were used. The images extracted from 376 manually verified forms were used. Achieved accuracy in recognizing isolated handwritten characters level was: 90.7% for letters and 98.1% for digits. Probabilistic lexicons on word and semantic levels were derived from real data in HIS system database containing 53252 patient records.

In the system being described here, all automatically recognized forms are manually verified before they are entered into HIS database. To simplify the manual data correction it is expected that the system suggest alternate values for erroneous fields. The values are presented to the user as combo boxes filled with items ordered according to their support factors evaluated by soft recognition classifier on word level. It is expected that the correct value is located close to top of the list, so user can select it without typing but rather by simply clicking on the list. In the same way, the user can select the complete set of patient data from the list of suggested patients already registered in HIS database, ordered according to support values produced by recognition algorithm on the semantic level. In assessing the recognition algorithm it is therefore not essentially important if the actual class is the one with highest value of support factor, but rather if it is among k ($k = 1, 2, 5$) classes with highest support values. The recognizer performance was therefore evaluated in three ways, using as the criterion the number of cases where actual class is among $k = 1, 3, 5$ classes with highest support factors.

On the word level the approach described in this article has been compared to two simple approaches. The first one is based only on the results of soft recognition on the character level. Support factor for a word is calculated as a product of support factors for subsequent letters. Only 5 words with highest values calculated in this way are taken into account. The second simple approach calculates support values in the same way, but the set of allowed words is defined by the lexicon. Probabilistic properties of the lexicon however are not used. Experiments have been performed for three levels of lexicon completeness: $p = 0.75$, $p = 0.90$ and $p = 1.0$, where p is the probability that actual word belongs to

Table 1. Names recognition accuracy

Criterion	S	SL p=0.75	CWA p=0.75	SL p=0.90	CWA p=0.90	SL p=1.00	CWA p=1.00
1 of 1	88.6%	90.7%	94.1%	92.3%	95.2%	93.4%	96.3%
1 of 3	90.2%	93.1%	94.6%	94.4%	95.5%	95.7%	97.1%
1 of 5	94.1%	94.1%	95.7%	96.0%	97.3%	96.5%	98.1%

Table 2. Surnames recognition accuracy

Criterion	S	SL p=0.75	CWA p=0.75	SL p=0.90	CWA p=0.90	SL p=1.00	CWA p=1.00
1 of 1	84.3%	87.5%	89.4%	89.8%	93.1%	91.6%	94.4%
1 of 3	91.2%	93.1%	93.9%	94.2%	95.9%	95.3%	97.1%
1 of 5	95.5%	94.2%	95.2%	96.0%	96.8%	96.0%	97.6%

the lexicon. Results for names and surnames recognition are presented in tables below. S and SL denote here two described above simple reference algorithms. CWA denotes combined word algorithm described in section 3.2.

Similar experiment has been performed to assess the accuracy CSA algorithm on semantic level. Results are presented in table 3.

Table 3. Patient identification accuracy

Criterion	S	SL p=0.75	CSA p=0.75	SL p=0.90	CSA p=0.90	SL p=1.00	CSA p=1.00
1 of 1	67.3%	77.7%	80.6%	83.4%	88.3%	89.9%	92.8%
1 of 3	73.7%	81.4%	84.8%	85.1%	87.5%	91.6%	93.4%
1 of 5	78.2%	84.0%	85.9%	88.0%	89.4%	92.3%	93.6%

6 Conclusions

Experiments described in previous section have shown that application of proposed algorithms on both word and semantic levels significantly improves isolated data and patient recognition accuracy. In case of complete name and surname lexicons, average reduction of error rate on word level is 43% and 37% correspondingly. In case of patient identification on semantic level error is reduced by 23%. Obtained results, due to reduction of necessary corrections, contribute to making form verifier work more efficient, easier and less error prone.

Described here methods have been implemented in laboratory test order forms recognition subsystem cooperating with large hospital information system. Elimination of necessity of retyping of most data present on data forms reduced the average operator time needed for single form processing many times and in result reduced also laboratory operation costs.

Acknowledgement

This work was financed from the State Committee for Scientific Research (KBN) resources in 2005–2007 years as a research project No 3 T11E 005 28.

References

1. Liu C., Nakashima K., Sako H., Fujisawa H.: Handwritten Digit Recognition: Benchmarking of State-of-the-Art Techniques. Pattern Recognition, Vol. 36. (2003) 2271-2285
2. Lu Y., Gader P. Tan C.: Combination of Multiple Classifiers Using Probabilistic Dictionary and its Application to Postcode Generation. Pattern Recognition, Vol. 35. (2002) 2823-2832
3. Kuncheva L.: Combining Classifiers: Soft Computing Solutions. In: Pal S., Pal A. (eds.): Pattern Recognition: from Classical to Modern Approaches. World Scientific (2001) 427-451
4. Kuncheva L.I.: Using measures of similarity and inclusion for multiple classifier fusion by decision templates. Fuzzy Sets and Systems, Vol. 122. (2001) 401-407
5. Sas J., Kurzynski M.: Multilevel Recognition of Structured Handprinted Documents – Probabilistic Approach. In: Kurzynski M., Puchala E. (eds.): Computer Recognition Systems, Proc. IV Int. Conference, Springer Verlag (2005) 723-730
6. Sas J., Kurzynski M.: Application of Statistic Properties of Letter Succession in Polish Language to Handprint Recognition. In: Kurzynski M. (eds.): Computer Recognition Systems, Proc. IV Int. Conference, Springer Verlag (2005) 731-738
7. Sas J.: Handwritten Laboratory Test Order Form Recognition Module for Distributed Clinic. J. of Medical Informatics and Technologies, Vol. 8. (2004) 59-68
8. Kurzynski M., Sas J.: Combining Character Level Classifier and Probabilistic Lexicons in Handprinted Word Recognition – Comparative Analysis of Methods. In: Proc. XI Int. Conference on Computer Analysis and Image Processing, LNCS Springer Verlag (2005) (to appear)
9. Devroye L., Gyorfi P., Lugossi G.: A Probabilistic Theory of Pattern Recognition. Springer Verlag, New York (1996)
10. Duda R., Hart P., Stork D.: Pattern Classification. John Wiley and Sons (2001)
11. Vinciarelli A. et al.: Offline Recognition of Unconstrained Handwritten Text Using HMMs and Statistical Language Models. IEEE Trans. on PAMI, Vol. 26. (2004) 709-720