Confidence Measure for Experimental Automatic Face Recognition System

Pavel Král[1,2] and Ladislav Lenc[1,2]

[1] Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
[2] New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
{pkral,llenc}@kiv.zcu.cz
http://nlp.kiv.zcu.cz

Abstract. This paper deals with automatic face recognition in order to propose and implement an experimental face recognition system. It will be used to automatically annotate photographs taken in completely uncontrolled environment. Recognition accuracy of such a system can be improved by identification of incorrectly classified samples in the post-processing step. However, this step is usually missing in current systems. In this work, we would like to solve this issue by proposing and integrating a confidence measure module to identify incorrectly classified examples. We propose a novel confidence measure approach which combines four partial measures by a multi-layer perceptron. Two individual measures are based on the *posterior* probability and two other ones use the *predictor* features. The experimental results show that the proposed system is very efficient, because almost all erroneous examples are successfully detected.

Keywords: Face Recognition, Czech News Agency, Confidence Measure, Multi-layer Perceptron, Scale Invariant Feature Transform (SIFT)

1 Introduction

Automatic face recognition consists in the use of a computer for identification of a person from a digital photograph. This area has been focused on by many researchers and many algorithms have been proposed during the past two decades. Nowadays, face recognition can be seen as one of the most progressive biometric authentication methods and represents a key task in several commercial or law enforcement applications as for example surveillance of wanted persons, access control to restricted areas, automatic annotation of the photos used in the recently very popular photo sharing applications or in the social networks, etc.

The majority of the current applications achieves high recognition accuracy only in the particular conditions (sufficiently aligned faces, similar face pose and lighting conditions, etc.). Unfortunately, their recognition results are significantly worse when these constraints are not fulfilled. Many approaches to resolve this issue have been proposed, however only few of them perform well in a fully uncontrolled environment.

In our previous work, we proposed the SIFT based Kepenekci face recognition method [19] and showed that it significantly outperforms other approaches particularly on uncontrolled face images. However, its recognition accuracy is still not perfect. Therefore, we proposed in [17] two Confidence Measure (CM) approaches in order to detect and handle incorrectly recognized examples. These approaches are based on the *posterior* class probability. We experimentally showed that these approaches are very promising in our task. However, the further improvement of the results is beneficial.

The main goal of this paper consists in proposing a novel confidence measure approach which will be integrated into our experimental automatic face recognition system. This method combines two previously proposed measures with two novel ones in a supervised way using a multi-layer perceptron classifier. The novel measures are based on the *predictor* features which characterize our face model.

The proposed system will be used by the Czech News Agency (ČTK)[3] to annotate people in photographs during insertion into the ČTK database[4]. Its main strength is to successfully process photos of a great number of different persons taken in a totally uncontrolled environment. The system (with the source code) is publicly available for research purposes for free.

The following section gives a brief overview of important face recognition and confidence measure approaches. Some existing face recognition systems are also mentioned at the end of this section. Section 3 describes our face recognition and confidence measure methods. Section 4 details the architecture of the proposed system. Section 5 evaluates and compares the performance of our system on the ČTK corpus. In the last section we discuss the achieved results and give some further research directions.

2 Related Work

This section is composed of three parts. The successful face recognition approaches are described in the first part, while the second part is focused on the confidence measure task itself. This section further summarizes some existing face recognition systems.

2.1 Face Recognition

One of the first successful approaches is Eigenfaces [34]. This approach is based on the Principal Component Analysis (PCA). Unfortunately, it is sensitive on

[3] http://www.ctk.eu
[4] http://multimedia.ctk.cz/en/foto/

variations in lighting conditions, pose and scale. However, the PCA based approaches are still popular, as shown in [24]. Another method, the Fisherfaces [6], is derived from Fisher's Linear Discriminant (FLD). According to the authors, this approach should be less sensitive to changing lighting conditions than Eigenfaces.

Independent Component Analysis (ICA) can be also successfully used in the automatic face recognition field [3].Contrary to Eigenfaces, ICA uses higher order statistics. It thus provides more powerful data representation. The authors showed that ICA performs slightly better than PCA method on the FERET corpus.

Another efficient face recognition approach is the Elastic Bunch Graph Matching (EBGM) [7]. This approach uses features constructed by the Gabor wavelet transform. Several other successful approaches based on Gabor wavelets have been introduced [30]. Some approaches [29] combine the pre-processing with Gabor wavelets with well-established methods such as Eigenfaces, Fisherfaces, etc. Kepenekci proposes in [15] an algorithm that addresses the main issue of Elastic Bunch Graph Matching, manual labelling of the landmarks. The proposed method outperforms the classical EBGM.

Other successful approaches [1, 33] use so-called, Local Binary Patterns (LBP) for facial feature extraction. A modification of the original LBP approach called Dynamic Threshold Local Binary Pattern (DTLBP) is proposed in [21]. It takes into consideration the mean value of the neighbouring pixels and also the maximum contrast between the neighbouring points. It is stated there that this variation is less sensitive to the noise than the original LBP method. Another extension of the original method is Local Ternary Patterns (LTP) proposed in [32]. It uses three states to capture the differences between the center pixel and the neighbouring ones. Similarly to the DTLBP the LTP is less sensitive to the noise. The so called Local Derivative Patterns (LDP) are proposed in [37]. The difference from the original LBP is that it uses features of higher order. It thus can represent more information than the original LBP.

Speeded-Up Robust Features (SURF) [4] is another recent method used for automatic face recognition. This method is invariant to face rotation. To ensure rotation invariance, one orientation is assigned to each key-point. The computation is based on the circular neighbourhood of the key-points.

Recently, the Scale Invariant Feature Transform (SIFT) is successfully used for face recognition [2]. The main advantage of this approach is the ability to detect and describe local features in images. The features are invariant to image scaling, translation and rotation. Moreover, they are also partly invariant to changes in illumination. Therefore, this approach is beneficial for face recognition in real conditions where the images differ significantly. Another approach based on the SIFT, called Fixed-key-point-SIFT (FSIFT), is presented in [16].

For further information about the face recognition, please refer to the survey [5].

2.2 Confidence Measure

Confidence measure is used as a post-processing of the recognition to determine whether the result is correct or not. The incorrectly recognized samples should be removed from the recognition set or another processing (e.g. manual correction) can be further realized.

This technique is mainly used in the automatic speech processing field [27, 36] and is mostly based on the *posterior* class probability. However, two other groups of approaches exist [14]. The first one uses a classifier in order to decide whether the classification is correct or not. This classifier uses a set of the so-called *predictor* features which should have a maximal discriminability between the correct and incorrect classes. The second group uses a likelihood ratio between the *null* (a correct recognition) and the *alternative* (an incorrect recognition) hypotheses.

The confidence measure can be successfully used in other research areas as shown in [28] for genome maps construction, in [12] for stereo vision or in [23] for handwriting sentence recognition.

Another approach related to the confidence measure is proposed by Proedrou et al. in the pattern recognition task [26]. The authors use a classifier based on the nearest neighbours algorithm. Their confidence measure is based on the algorithmic theory of randomness and on transductive learning.

Unfortunately, only few works about the confidence measure in the face recognition domain exist. Li and Wechsler propose a face recognition system which integrates a confidence measure [20] in order to reject unknown individuals or to detect incorrectly recognized faces. Their confidence measure is, as in the previous case, based on the theory of randomness. The proposed approaches are validated on the FERET database.

Eickeler et al. propose and evaluate in [11] five other CMs also in the face recognition task. They use a pseudo 2-D Hidden Markov Model classifier with features created by the Discrete Cosine Transform (DCT). Three proposed confidence measures are based on the *posterior* probabilities and two others on ranking of results. Authors experimentally show that the *posterior* class probability gives better results for the recognition error detection task.

Note that the most of the proposed approaches are unsupervised. However, the supervised [31] and semi-supervised [10] methods have been also proposed.

2.3 Face Recognition Systems

As already shown above, numerous papers presented in the face recognition domain concentrate only on the recognition task itself. Unfortunately, to the best of our knowledge, relatively few works about whole face recognition systems exist.

One example of such a system is proposed in [22]. The system compensates the face position and also solves partial occlusion and different facial expressions. Only one training example per person is used.

Another face recognition system is described in [35]. The training images are well aligned (acquired in controlled conditions) whereas the recognized images are real-world photos. The system is based on the Sparse Representation and Classification (SRC) [13] algorithm. It achieves very good results on the FERET database.

Campadelli et al. present in [9] another face recognition system. The authors localize the face in the images and then compute the facial features. Their face recognition algorithm is based on the EBGM, but the fiducial points are detected completely automatically. The system is evaluated on the FERET corpus. The authors show that their system has recognition scores comparable to the elastic bunch graph matching.

For additional information about the face recognition and the face recognition systems, please refer to the survey [38]. The authors also mention some commercial face recognition systems. Unfortunately, neither the system architecture nor the approaches used are usually reported.

3 Confidence Measure for Face Recognition System

3.1 Face Recognition

For the face recognition task, we use our previously proposed SIFT based Kepenekci method [19] which uses the efficient SIFT algorithm for parametrization and adapted Kepenekci matching [18] for recognition. This method was chosen, because as proven previously, it significantly outperforms other approaches particularly on lower quality real data.

SIFT parametrization This algorithm creates an image pyramid with resampling between each level to determine potential key-point positions. Each pixel is compared with its neighbours. Neighbours in its level as well as in the two neighbouring levels are analysed. If the pixel is maximum or minimum of all neighbouring pixels, it is considered to be a potential key-point.

For the resulting set of key-points their stability is determined. The locations with low contrast and unstable locations along edges are deleted.

The orientation of each key-point is computed next. The computation is based on gradient orientations in the neighbourhood of the pixel. The values are weighted by the magnitudes of the gradient.

The last step consists in the descriptor creation. The computation involves the 16×16 neighbourhood of the pixel. Gradient magnitudes and orientations are computed in each point of the neighbourhood. Their values are weighted by a Gaussian. For each sub-region of size 4×4 (16 regions), the orientation histograms are created. Finally, a vector containing 128 (16×8) values is created.

Adapted Kepenekci matching This approach combines two methods of matching and uses the weighted sum of the two results.

Let T be a test image and G a gallery image. For each feature vector t of face T we determine a set of relevant vectors g of face G. Vector g is relevant iff:

$$\sqrt{(x_t - x_g)^2 + (y_t - y_g)^2} < distanceThreshold \tag{1}$$

where x and y are the coordinates of the feature vector points.

If no relevant vector to vector t is identified, vector t is excluded from the comparison procedure. The overall similarity of two faces OS is computed as the average of similarities between each pair of corresponding vectors as:

$$OS_{T,G} = mean\left\{S(t,g), t \in T, g \in G\right\} \tag{2}$$

Then, the face with the most similar vector to each of the test face vectors is determined. The C_i value denotes how many times gallery face G_i was the closest one to some of the vectors of test face T. The similarity is computed as C_i/N_i where N_i is the total number of feature vectors in G_i. The weighted sum of these two similarities is used for similarity measure:

$$FS_{T,G} = \alpha OS_{T,G} + \beta \frac{C_G}{N_G} \tag{3}$$

The face is recognized by the following equation:

$$\hat{FS}_{T,G} = \arg\max_G (FS_{T,G}) \tag{4}$$

The cosine similarity is used for vector comparison.

3.2 Confidence Measure

Posterior class probability approaches Let $P(F|C)$ be the output of the classifier, where C is the recognized face class and F represents the face features. The values $P(F|C)$ are normalized to compute the *posterior* class probabilities as follows:

$$P(C|F) = \frac{P(F|C).P(C)}{\sum_{I \in \mathcal{FIM}} P(F|I).P(I)} \tag{5}$$

$\mathcal{FIM}$ represents the set of all individuals and $P(C)$ denotes the *prior* probability of the individual's (face) class C.

We propose two different approaches. In the first approach, called ***absolute confidence value***, only faces $\hat{C}$ complying with

$$\hat{C} = \arg\max_C (P(C|F)) \tag{6}$$

$$P(\hat{C}|F) > T \tag{7}$$

are considered as being recognized correctly.

The second approach, called ***relative confidence value***, computes the difference between the *best* score and the *second best* one by the following equation:

$$P\Delta = P(\hat{C}|F) - \max_{C \neq \hat{C}} (P(C|F)) \tag{8}$$

Only the faces with $P\Delta > T$ are accepted. This approach aims to identify the "dominant" faces among all the other candidates. T is the acceptance threshold and its optimal value is adjusted experimentally.

Note that these two measures working separately were already presented in [17]. However, their description is important in the context of the whole composed approach.

Predictor feature approaches As already stated, this type of approaches uses the features with a maximal discriminability between the correct and incorrect classes to classify the recognition results. Two measures are proposed next.

The first one is based on the number of vectors in the model with the highest output value during the recognition task (i.e. the recognized face model). The number of vectors is given by the results of the SIFT algorithm. A face model with a high number of vectors is more general and it can be more likely identified as a good one. Conversely, a few vector face model is more accurate. Therefore, when this model is chosen as a good one (the highest output value) we assume that it is very probable that the recognition is correct.

Let V be the number of vectors in the face model and let T be the acceptance threshold. Only the faces where $V < T$ are accepted. The optimal value of the threshold T will be set experimentally. This measure is hereafter called the ***vector number*** approach.

The second measure uses a standard deviation of the similarities among images in the recognized face model. Let the recognized model M be composed of the images $I_1, I_2, ..., I_N$. The S measure is defined as follows:

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (FS_{I_i, M \setminus I_i} - \mu)^2} \tag{9}$$

where $FS_{I_i, M \setminus I_i}$ is the similarity (see Equation 3) of the image I_i and a model $M \setminus I_i$ created from the remaining images from model M and μ is computed by the following equation:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} FS_{I_i, M \setminus I_i} \tag{10}$$

Similarly as in the case of the *vector number* measure we suppose that higher standard deviation characterizes a more general face model and vice versa. Therefore, only the recognition results where $S < T$ are accepted. The optimal value of the acceptance threshold T will be set experimentally. This measure is hereafter called the ***standard deviation*** approach.

Composed supervised approach Let R_k be the score obtained by a partial unsupervised measure k described above and let variable H determines whether the face image is classified correctly or not. A Multi-layer Perceptron (MLP) which models the *posterior* probability $P(H|R_1, .., R_N)$ is used to combine all partial measures in a supervised way. Note that the variable N represents the number of measures to combine

In order to identify the best performing topology, several combinations and MLP configurations are built and evaluated. The MLP topologies will be described in detail in the experimental section.

4 System Description

The proposed system has (as shown in Figure 1) a modular architecture. It is composed of five modules (see the rectangles) connected by dependencies (see the oriented edges). The input image and the recognition results are represented by parallelograms. The storage of the face representation is shown by the *Face Gallery* sign.

The first module $M1$ deals with face extraction. This module converts a color image into its grey-scale representation, then it performs face detection. The detected face is further extracted from the image in the next step. This module also detects the eyes in the detected face region and transforms and resizes the face.

The second module $M2$ is used to create the face representation. It detects the SIFT key-points and creates a set of SIFT descriptors for a representation of the face image.

The next module $M3$ is used to create a face model M. It uses the SIFT representations of the face images (output of module $M2$) and saves them to the face gallery.

The fourth module $M4$ deals with face recognition. A recognized face is compared to the face models stored in the *Face Gallery* and the most similar model is chosen as the recognized one.

The last *confidence measure* module $M5$ is dedicated to identifying whether the recognition result is correct or not. This unique step is particularly important, because when the user knows that the recognition is probably not correct, he can manually correct the recognition result.

Modules $M1$ and $M2$ are used in both face representation (or modelling) and face recognition tasks. However, module $M3$ is used only for face representation and modules $M4$ and $M5$ are used only for recognition. The last remark is that every module should be used separately in order to create another face processing system.

5 Experiments

5.1 Czech News Agency Corpus

This corpus is composed of images of individuals in an uncontrolled environment that were randomly selected from the large ČTK database. All images were taken over a long time period (20 years or more). The corpus contains grey-scale images of 638 individuals of size 128×128 pixels. It contains about 10 images for

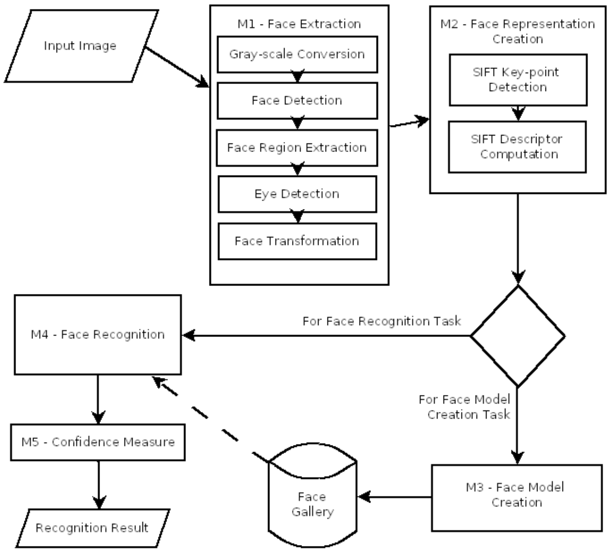


Fig. 1. System Architecture

each person. The orientation, lighting conditions and image backgrounds differ significantly.

Figure 2 shows four examples of one face from this corpus. This corpus is available for free for research purposes at `http://home.zcu.cz/~pkral/sw/` or upon request to the authors.

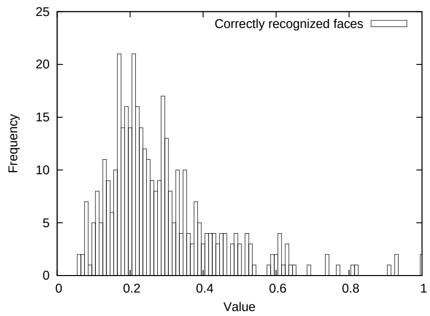
Fig. 2. Examples of one face from the ČTK face corpus

5.2 Recognition Results with Confidence Measure

Three experiments are described next. The first experiment analyses the discriminability of the proposed partial measures by histograms. This experiment is realized in order to show the suitability of the proposed measures. The second experiment reports the results of the measures also used separately. In the last experiment, we show the classification results of the whole composed approach.

Discriminability of the proposed measures In the first experiment, we would like to analyse the discriminability of the proposed partial measures. We created two histograms for every measure in order to analyse the distribution of the correctly and incorrectly classified faces. The reported output densities of the measures are based on the 638 values (the number of individuals in the corpus). Note that all output values are normalized to the interval [0..1].

Figure 3 shows the output densities of the correctly and incorrectly classified faces when the *absolute confidence value* measure is used. These histograms show that the majority of the correctly recognized face examples has higher output values than the incorrectly recognized ones. This fact confirms our assumption that the first measure is suitable for our task and should be useful to be integrated to the whole composed method.



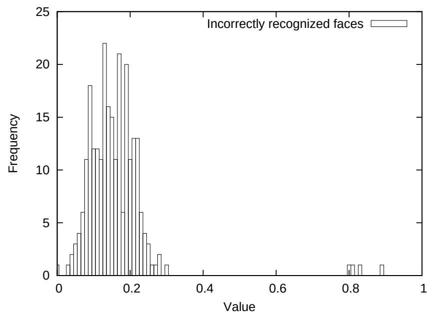


Fig. 3. Histograms of the correctly (left) and incorrectly (right) classified faces using the *absolute confidence value* measure

Figure 4 plots the output densities when the *relative confidence value* measure is used. These histograms show clearly that the discriminability of this measure is better than the previous one. Almost all correctly recognized face examples have higher output values than the incorrectly recognized samples. Therefore this measure should be suitable for our task and we decided to combine it with the other ones by an MLP. Moreover, we assume that this measure used separately outperforms the previously proposed one.

Figure 5 depicts the output densities when the *vector number* measure is used. These histograms show that this measure is less discriminant than the two ones presented previously. However, the correctly recognized examples have slightly inferior output values than the incorrectly ones. This fact confirms our assumption (see Sec. 3.2) that the confidence of a few vector model is high. We assume that this measure will bring poor results if used separately. However, it can add some further information when it will be combined with the other approaches. Therefore, we decided to integrate it into the whole composed approach.

The output densities of the last *standard deviation* measure are reported in Figure 6. The discriminability of these two histograms are limited and it is diffi-

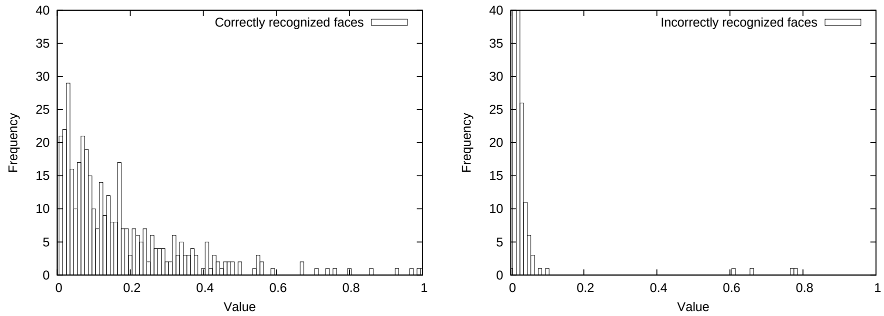


Fig. 4. Histograms of the correctly (left) and incorrectly (right) classified faces using the *relative confidence value* measure

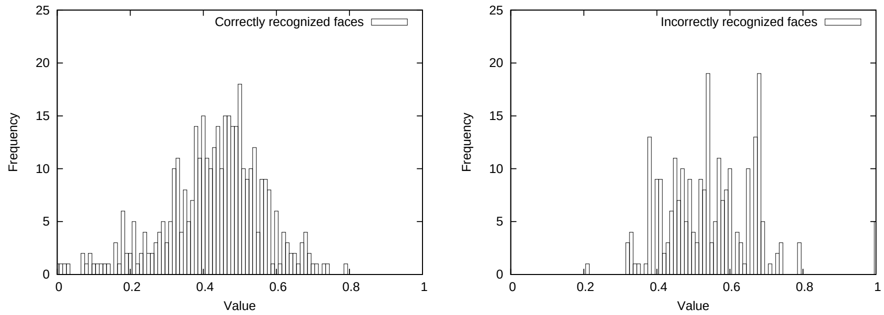


Fig. 5. Histograms of the correctly (left) and incorrectly (right) classified faces using the *vector number* measure

cult to propose some conclusions about this measure. However, we decided to use this measure in the further experiments and verify its usefulness experimentally.

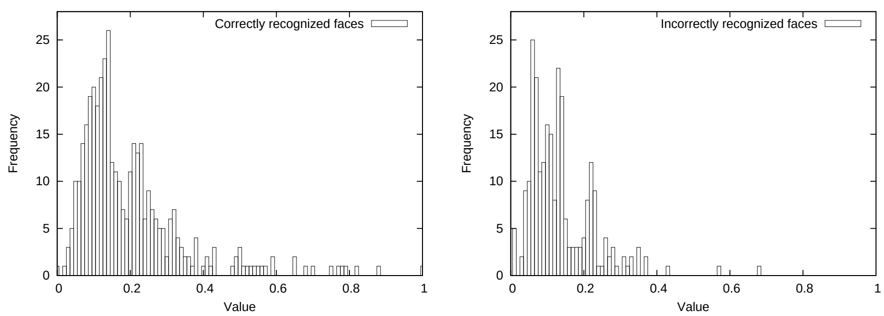


Fig. 6. Histograms of the correctly (left) and incorrectly (right) classified faces using the *standard deviation* measure

To summarize:

- *relative confidence value (rel)* measure is the best proposed one;
- *absolute confidence value (abs)* method has also very good separation abilities;
- *vector number (vect)* measure can bring some complementary information for our task;
- contribution of the *standard deviation (sd)* measure is questionable and must be confirmed experimentally.

Accuracy of the separate measures In the second experiment we would like to show the performance of the above described measures used separately without any combination. As in many other articles in the confidence measure field, we will use the Receiver Operating Characteristic (ROC) curve [8] for evaluation of this experiment. This curve clearly shows the relationship between the true positive and false positive rates for the different *acceptance* threshold.

Figure 7 shows the results of the proposed *absolute confidence value*, *relative confidence value*, *vector number* and *standard deviation* measures used separately. This experiment shows that the *relative confidence value* method significantly outperforms the all other approaches.

We can further deduce that our assumption in the fourth proposed measure was not correct. Based on this experiment we can consider that the dependence between the value of the standard deviation and the correctly recognized faces is reversed. We modify the definition of such measure as follows: only the faces where $S > T$ are accepted.

After this modification we can conclude that all proposed measures are suitable for our task in order to identify incorrectly recognized faces. Note that the corrected version of the ROC curve of the fourth *standard deviation* measure is reported with the *modified standard deviation* caption.

We will further compare the results of the separate measures with the whole composed approach. Therefore, we created Table 1 to show the scores of the separate measures with optimal threshold configurations. The F-measure (F-mes) [25] is used as an evaluation metric, the Precision (Prec) and Recall (Rec) are also reported in this table. Precision and recall have the similar importance for our application, therefore the optimal threshold $\hat{T}$ value has been defined for the "best" compromise between these two values:

$$\hat{T} = \arg\min_{T} \mid 1 - \frac{Prec}{Rec} \mid \qquad (11)$$

Accuracy of the whole composed approach In the last experiment, we will evaluate the results of the whole composed confidence measure method. First, we will show the impact of the use of an MLP classifier with the separate measures. Then, we compare and evaluate all possible combinations of the proposed measures in order to show the complementarities among them.

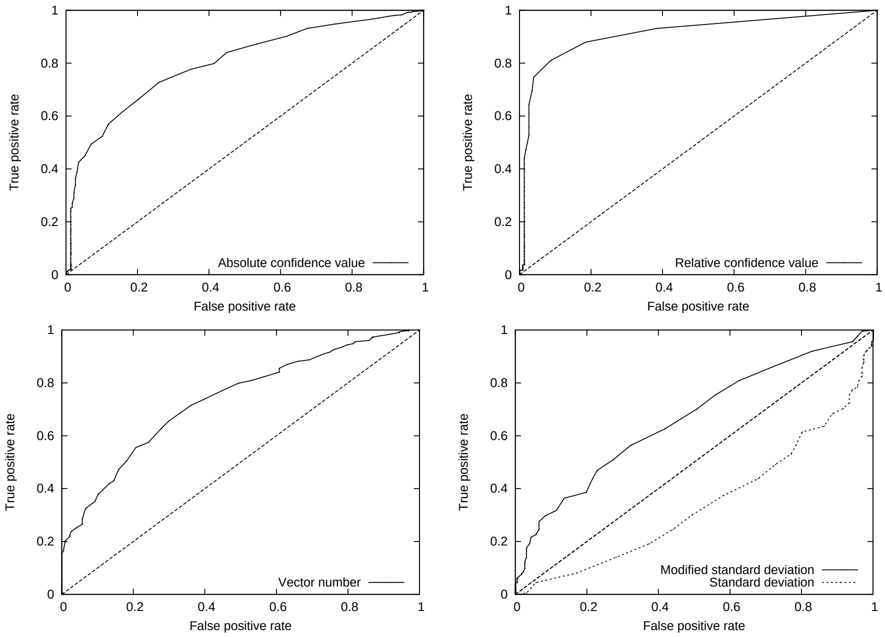


Fig. 7. ROC curves of the four proposed measures used separately. The corrected *standard deviation* measure is reported with the *modified standard deviation* label.

Table 1. Performance of the measures used separately [%]

Confidence Measure	**Prec**	**Rec**	**F-mes**
absolute confidence value	65.7	60.6	63.0
relative confidence value	69.6	60.8	64.9
vector number	62.2	63.5	62.8
standard deviation	58.9	60.3	59.6

Several MLP configurations are tested. The best MLP topology uses three layers. The number of the input neurons corresponds to the number of measures to combine, 10 neurons are in the hidden layer and two outputs are used to identify the *correctly* and *incorrectly* recognized faces. This MLP topology was set empirically on a small development corpus which contains 120 examples (i.e. 120 confidence values).

The results of this experiment are reported in Table 2. These results show that the separate measures used with an MLP have better F-measure values (except *sd* approach) than used in the unsupervised way. A successive addition of the measures improves progressively the F-measure value. When all measures are combined, the resulting F-measure is close to 100%. This figure also shows that all measures bring complementary relevant information and are thus useful to be integrated to the whole composed approach (i. e. the whole combined approach gives the best recognition score).

Table 2. Performance of all combinations of the measures by an MLP classifier [%]

Confidence Measure	Prec	Rec	F-mes
1. Separate measures			
abs. confidence value (abs)	92.5	64.8	76.2
rel. confidence value (rel)	96.2	80.4	87.6
vector number (vect)	55.4	84.9	67.0
standard deviation (sd)	54.0	65.3	59.1
2. Combinations of two measures			
abs, rel	97.2	83.5	89.8
abs, sd	70.4	55.8	62.2
abs, vect	95.8	75.8	84.6
rel, sd	95.8	84.3	89.7
rel, vect	97.7	85.6	91.2
sd, vect	67.6	90.6	77.4
3. Combinations of three measures			
abs, rel, sd	96.7	90.0	93.2
abs, rel, vect	97.2	93.7	95.4
abs, sd, vect	93.4	90.5	91.9
rel, sd, vect	94.8	94.8	94.8
4. Combination of all measures (the whole approach)			
abs, rel, sd, vect	100	99.5	99.8

6 Conclusions and Perspectives

We proposed and evaluated a novel confidence measure approach and integrated it in the experimental automatic face recognition system as a new module. The proposed approach combines two measures based on the *posterior* probability and two ones based on the *predictor* features in a supervised way with an MLP. We experimentally showed that the proposed approach is very efficient, because it detects almost all erroneous examples. We further showed that it is possible to use all four proposed measures separately. However, every measure brings complementary information and it is thus beneficial to combine all measures in the composed approach.

To summarize, the main scientific contribution of this paper consists in:

1. proposing two novel measures based on the *predictor* features;
2. proposing a combined supervised confidence measure approach which combines the measures from two groups of methods; two ones based on the *posterior* class probability and the other two ones on the *predictor* features;

3. evaluation of the proposed method in the face recognition task on the real ČTK data;
4. integration of the proposed confidence measure approach into the experimental automatic face recognition system.

The first perspective consists in proposing of semi-supervised confidence measures. In this approach, the CM model will be progressively adapted according to the recognized data. We will further integrate other more suitable features into our model. Another perspective consists in the use of our confidence measure approach in the task of automatic creation of the face corpora.

Acknowledgements

This work has been partly supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems and by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090. We also would like to thank Czech New Agency (ČTK) for support and for providing the photographic data.

References

1. Ahonen, T., Hadid, A., Pietikinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J. (eds.) Computer Vision - ECCV 2004, Lecture Notes in Computer Science, vol. 3021, pp. 469–481. Springer Berlin Heidelberg (2004), http://dx.doi.org/10.1007/978-3-540-24670-1_36
2. Aly, M.: Face recognition using sift features (2006)
3. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Transactions on Neural Networks pp. 1450–1464 (2002)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. 110(3), 346–359 (Jun 2008), http://dx.doi.org/10.1016/j.cviu.2007.09.014
5. Beham, M.P., Roomi, S.M.M.: A review of face recognition methods. International Journal of Pattern Recognition and Artificial Intelligence 27(4) (2013)
6. Belhumeur, P.N., Hespanha, J.a.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence (1997)
7. Bolme, D.S.: Elastic Bunch Graph Matching. Ph.D. thesis, Colorado State University (2003)
8. Brown, C.D., Davis, H.T.: Receiver operating characteristics curves and related decision measures: A tutorial. Chemometrics and Intelligent Laboratory Systems 80(1), 24–38 (2006)
9. Campadelli, P., Lanzarotti, R.: A face recognition system based on local feature characterization. In: Advanced Studies in Biometrics, pp. 147–152. Springer (2005)
10. Deng, J., Schuller, B.: Confidence measures in speech emotion recognition based on semi-supervised learning. In: INTERSPEECH (2012)

11. Eickeler, S., Jabs, M., Rigoll, G.: Comparison of confidence measures for face recognition. In: FG. pp. 257–263. IEEE Computer Society (2000), http://dblp.uni-trier.de/db/conf/fgr/fg2000.html#EickelerJR00
12. Hu, X., Mordohai, P.: A quantitative evaluation of confidence measures for stereo vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(11), 2121–2133 (2012)
13. Huang, K., Aviyente, S.: Sparse representation for signal classification. Advances in neural information processing systems 19, 609 (2007)
14. Jiang, H.: Confidence measures for speech recognition: A survey. Speech Communication 45(4), 455–470 (2005)
15. Kepenekci, B.: Face Recognition Using Gabor Wavelet Transform. Ph.D. thesis, The Middle East Technical University (2001)
16. Krizaj, J., Struc, V., Pavesic, N.: Adaptation of sift features for robust face recognition (2010)
17. Lenc, L., Král, P.: Confidence measure for automatic face recognition. In: International Conference on Knowledge Discovery and Information Retrieval. Paris, France (October 26-29 2011)
18. Lenc, L., Král, P.: Novel matching methods for automatic face recognition using sift. In: 8th AIAI (Artificial Intelligence Applications and Innovations) Confence. Halkidiki, Greece (September 27-30 2012)
19. Lenc, L., Král, P.: Face recognition under real-world conditions. In: International Conference on Agents and Artificial Intelligence. Barcelona, Spain (February 14-18 2013)
20. Li, F., Wechsler, H.: Open world face recognition with credibility and confidence measures. In: Audio-and Video-Based Biometric Person Authentication. pp. 462–469. Springer (2003)
21. Li, W., Fu, P., Zhou, L.: Face recognition method based on dynamic threshold local binary pattern. In: Proceedings of the 4th International Conference on Internet Multimedia Computing and Service. pp. 20–24. ACM (2012)
22. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(6), 748–763 (2002)
23. Marukatat, S., Artières, T., Gallinari, P., Dorizzi, B.: Rejection measures for handwriting sentence recognition. In: Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on. pp. 24–29. IEEE (2002)
24. Poon, B., Amin, M.A., Yan, H.: Performance evaluation and comparison of pca based human face recognition methods for distorted images. International Journal of Machine Learning and Cybernetics 2(4), 245–259 (2011)
25. Powers, D.: Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. Journal of Machine Learning Technologies 2(1), 37–63 (2011)
26. Proedrou, K., Nouretdinov, I., Vovk, V., Gammerman, A.: Transductive confidence machines for pattern recognition. In: ECML'02. pp. 381–390 (2002)
27. Senay, G., Linares, G., Lecouteux, B.: A segment-level confidence measure for spoken document retrieval. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. pp. 5548–5551. IEEE (2011)
28. Servin, B., de Givry, S., Faraut, T.: Statistical confidence measures for genome maps: application to the validation of genome assemblies. Bioinformatics 26(24), 3035–3042 (2010)
29. Shen, L.: Recognizing Faces - An Approach Based on Gabor Wavelets. Ph.D. thesis, University of Nottingham (2005)

30. Shen, L., Bai, L.: A review on gabor wavelets for face recognition. Pattern Analysis & Applications (2006)
31. Sukkar, R.A.: Rejection for connected digit recognition based on gpd segmental discrimination. In: Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on. vol. 1, pp. I–393. IEEE (1994)
32. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. Image Processing, IEEE Transactions on 19(6), 1635–1650 (2010)
33. Timo, A., Hadid, A., Pietikinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 2037–2041 (2006)
34. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on In Computer Vision and Pattern Recognition. Computer Vision and Pattern Recognition (1991)
35. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.: Toward a practical face recognition system: Robust alignment and illumination by sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(2), 372–386 (2012)
36. Wessel, F., Schluter, R., Macherey, K., Ney, H.: Confidence measures for large vocabulary continuous speech recognition. Speech and Audio Processing, IEEE Transactions on 9(3), 288–298 (2001)
37. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. Image Processing, IEEE Transactions on 19(2), 533–544 (2010)
38. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. Acm Computing Surveys (CSUR) 35(4), 399–458 (2003)