Face Templates Creation for Surveillance Face Recognition System

Tobias Malach[1,2] and Jiri Prinosil[3]
[1]Department of Radio Electronics, Brno University of Technology, Brno, Czech Republic
[2]EBIS, spol. s r.o., Brno, Czech Republic
[3]Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

Keywords: Face Templates, Template Database Creation, Face Recognition System Application, Real-world Conditons.

Abstract: This paper addresses the problem of face templates creation for facial recognition system. The application of a face recognition system in real-world conditions requires compact and representative face templates in order to maintain low error rate and low classification time. Contemporary face template creation methods are not suitable for face recognition systems with large number of users as they produce many templates per person. These templates are often redundant and their high number requires long classification time. The paper presents four approaches to face templates creation that produce one to three face templates per person. The influence of different face template creation approaches was assessed on PubFig and IFaViD database. The achieved results show that appropriate face template creation methods have a significant influence on face recognition system performance.

1 INTRODUCTION

The machine recognition of human faces is being gradually implemented into various real-world applications. In the past, face recognition algorithms had severe limitations that prevented their usage in practical applications. An intensive research made feasible the partial elimination of these limitations, Zhao & Chellappa (2006). Thus face recognition systems have achieved acceptable results in appropriately designed applications. Nevertheless unsolved issues in near real-time applications with large number of users still remain e.g. long classification time and extensive template databases.

The face recognition research has been recently focused on face description by efficient features and a development of complex classifiers. There are other aspects that influence the face recognition system's performance, e.g. the creation of face templates. Face templates directly influence the classifier's performance, the system's sensitivity on a face pose and illumination, etc.

This paper presents novel approaches to face templates creation and describes their impact on recognition performance. The proposed approaches aim to reduce the number of face templates per person, which accelerates the classification process and reduces redundancy among one individual's templates. The face template creation research is a part of the project IVECS (Intelligent Video modules for Entrance Control Systems). IVECS deals with a development of face recognition system which is intended to be used in surveillance CCTV (Closed Circuit Television) systems.

In an ongoing text, the developed face recognition system is described. Subsequently new approaches to face template creation are proposed. The influence of different face template creation methods is evaluated on images from the PubFig Database and the IFaViD database described in Bambuch, Malach & Malach (2012). The IFaViD database was assembled from video sequences captured by the CCTV surveillance system. Test results are then credible considering the use in surveillance systems. The results indicate that the influence of face template creation has a significant impact on a face recognition system's performance.

1.1 Related Work

There are very few available publications describing the implementation of face recognition systems into real-world application. The work of Stallkamp, Ekenel, & Stiefelhagen (2007) presents overall face

Malach T. and Prinosil J..
Face Templates Creation for Surveillance Face Recognition System.
DOI: 10.5220/0004906307240729
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 724-729
ISBN: 978-989-758-018-5

recognition system implementation. Stallkamp et al. used a k-means algorithm to create face templates for K-NN (K-Nearest Neighbour) classifier. Using k-means algorithm, 10 face templates per person were created. However Stallkamp et al. didn't examine the influence of face templates on recognition performance.

2 FACE RECOGNITION SYSTEM DESCRIPTION

The developed face recognition system is to be part of a surveillance camera system. There are specific operating conditions due to the application in such a system. Firstly frontal face images are only provided in special cases. Generally face images are non frontal. Secondly an illumination intensity and sources varies. Last but not least aspect is that significant percentage of images is of low quality. Consequences of system's operating conditions introduce difficulties to the face recognition process.

The system's design should take into account user's requirements e.g.: near real time recognition, low error rate etc.

To fulfil requirements on the system and to cope with difficulties introduced by real operating conditions, we have utilized approaches that have already been proved to be reliable. The tested face recognition system is described in an ongoing text.

2.1 Face Acquisition with Viola-Jones Detector

The Viola-Jones face detector introduced in Viola & Jones (2001) has been used for face detection. This algorithm makes fast and reliable face detection feasible. After face detection, which roughly determines face position, face alignment is performed. We have proposed a framework, which aligns and crops the face image according to a position of found facial features such as eyes.

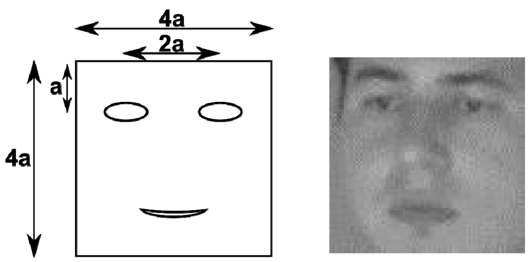

Figure 1: Example of face an alignment with relative distances among image margins and eye centers.

2.2 Face Description with LBPH Features

Once the face is detected and aligned, the features describing the face are extracted. We have applied LBPH (Local Binary Patterns Histograms) features presented by Ahonen, Hadid & Pietikäinen (2006). LBPH features appear to have both high discriminative power and low computational complexity. We have utilized following setup of LBPH features: radius of LBP 2, neighbors number 7, and grid 8 x 8.

A significant aspect for the practical implementation of a face recognition system is the length of a feature vector. The LBPH feature vector has a length of 8192 in integer precision. When considering a face template database consisting of hundreds of templates, a database access time and classification time have to be taken into account. From this point of view, LBPH features seem to be a compromise between discriminative power, computational complexity and a database access time.

2.3 Classification Scheme

A NN (Nearest Neighbors) based classifier has been utilized. The NN-based classifier seems to be suitable for classifying LBPH feature vectors because of their high dimensionality.

The NN classifier compares a feature vector (representing an unknown individual) to all face templates. The classifier utilized in our work uses a threshold which defines maximum dissimilarity between an unknown feature vector and a face template. The feature vector is assigned to a face template whose mutual dissimilarity is minimal and is smaller than the threshold. If the distance among face templates and the feature vector exceeds set threshold, the feature vector is expected to represent an individual without a face template – an impostor.

The NN-based classifier exploits chi square distance χ^2 to express mutual relations among feature vector and templates. The chi square metric is defined as follows:

$$\chi^2(T,S) = \sum_i \frac{(T_i - S_i)^2}{T_i + S_i}, \quad (1)$$

where T and S are feature vectors and i is the dimensionality of T and S respectively. We have adopted the chi square metric defined above in our approach.

3 FACE TEMPLATES CREATION METHODS

The face recognition scheme described in the previous section has been used to conduct all tests and remains unchanged. The investigated and modified part of the recognition system was the face template database. Several different face template databases have been created using the following methods.

3.1 The Most Similar Face Method

Many of current approaches create face templates by projecting a training database into the feature space, Zhao & Chellappa (2006). Each training face image creates one face template. This approach results in a large face template database which is not suitable for the developed system due to the large number of database entries, time-demanding classification. Therefore we propose the most similar face method that selects one face template per person out of all available face templates of one individual.

The template having the smallest accumulated distance (the chi square distance according to formula (1)) to all other templates of one individual is chosen as the ultimate template. The suggested approach is considered as a baseline method for this work.

3.2 Centroid Templates

The centroid template method is based on the idea that the feature vectors of one individual may be represented by one point in a multi-dimensional feature space – the cluster's centroid. This approach expects that all features have the same significance. The result face template is computed as a mean vector in all dimensions. The face template $T_{i,j}$ for an individual i in the dimension j is determined as follows:

$$T_{i,j} = \frac{1}{n_i} \sum_{n} F_{i,j,n} \, , \tag{2}$$

where $F_{i,j,n}$ is a value of the n^{th} feature vector of the i^{th} individual in the j^{th} dimension, and n_i is the total number of feature vectors for the i^{th} individual. This approach produces face templates which should be closest to all feature vectors in one cluster.

The Centroid templates method appears to be a simple and straightforward approach to face template creation. This method produces face templates which does not correspond to any real face; face templates are synthetic. This feature may prove impractical in cases when training face images were captured in diverted poses. This may result in a complex cluster which cannot be correctly represented by its centroid. In such a case the representation of a cluster as its centroid may be misleading and may result in a higher error rate. The medoid templates method is proposed to cope with this feature.

3.3 Medoid Templates

The Medoid templates method finds the most similar feature vector with respect to all the feature vectors of one individual. This approach was originally proposed in the work of Prinosil (2013). The medoid templates method generalizes Prinosil's formal approach and adds an outliers removal feature. The medoid templates method uses the following framework.

Distances among all feature vectors extracted from face images of one individual are calculated. The outliers removal is maintained by thresholding the distances D. the distances D are computed according to formula (1); distances exceeding the defined threshold T are excluded. The remaining distances $D(H_x,H_y)$ are scaled using the following formula:

$$D'(H_x, H_y) = \sqrt{\frac{T - D(H_x, H_y)}{T}} \, . \tag{3}$$

The scaled distances $D'(H_x,H_y)$ are added for each feature vector. The resulting face template corresponds to the feature vector with the highest total of scaled distances $D'(H_x,H_y)$.

Face templates creation using medoids should provide a better representation of complex clusters.

3.4 Multiple Face Templates (MFTM)

A cluster analysis of feature vectors extracted from face images captured by surveillance camera system indicates that clusters of one person's features cannot be sufficiently represented by one point – one face template. This is caused by the cluster's complexity which is a consequence of real-world conditions and an imperfect face alignment. To cover complex clusters, we have utilized a method producing multiple face templates for one individual.

The multiple face templates are obtained using the hierarchical k-means decomposition algorithm which was presented by Muja & Lowe (2009). This algorithm clusters feature vectors using the k-means algorithm into B regions; B is called the branching factor. Every region is clustered repeatedly until the number of feature vectors is larger than B. The application of such a recursive algorithm on the feature vectors of one individual results in the hierarchical tree representing distance relations among feature vectors. Once the hierarchical tree is constructed, the desired number of face templates has to be extracted. For this purpose, a cut in the hierarchical tree which produces the desired number of clusters/face templates is taken. The cluster's representatives are chosen as face templates.

4 TESTING METHODOLOGY

The influence of the proposed template creation methods on face recognition system performance was evaluated. Tests were conducted on images from two datasets. The first one is the PubFig database described in Kumar, Berg, Belhumeur & Nayar (2009) and the second is the IFaViD (IVECS Face Video Database). The PubFig database has been used because of the significant variability in face appearance. The IFaViD database has been used because its images correspond to the intended use of the developed face recognition system. Therefore IFaViD test results are supposed to be closer to the system's performance in real-world application.

4.1 The PubFig Database

The PubFig (Public Figures Face Database) is a standardized database for the performance evaluation of face recognition systems. For testing purposes we have divided the database into two non-overlapping parts, the training set and the test set. The training set consists of images of 86 different people; each person has exactly 40 training images. The test set consists of the images of 194 people. The number of test images per person varies; the total number of test images is 4,552.

4.2 The IFaViD Database

The IFaViD is a database of video sequences captured by the CCTV surveillance system. The original IFaViD contains 5,357 video sequences of 275 individuals. For testing purposes of this work, single images were used instead of video sequences. The number of single images totalled 16,442.

The video sequences have been captured by different cameras with different placements. Cameras captured one of the defined scenarios (a time-ordered sequence of human actions during video sequence capture). Images of two scenarios were used, scenarios are defined as follows:

- Scenario A: a person walking through a door frame or a corridor.
- Scenario B: a person requesting a closed door or a gateway access via an identification device.

Detailed description of scenarios, video sequences capturing and IFaViD assembly was described in Malach, Bambuch & Malach (2012) or Bambuch, Malach & Malach (2012).

4.2.1 The IFaViD Training Set

The IFaViD contains unique training set for each scenario. The training set contains manually sorted face images. The number of training images of one individual ranges between 40 and 150 images. Training images were extracted from video sequences that had been captured by a CCTV system. The training set does not overlap with the test set.

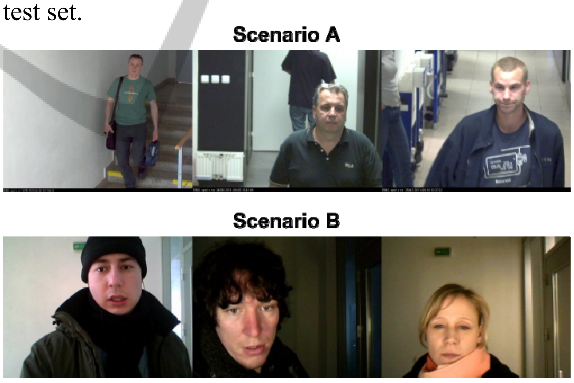

Figure 2: Example images from IFaViD database.

The IFaViD database was used for testing because IFaViD video sequences provide a trustworthy representation of the environment and operating conditions of a system's intended application. The results obtained are supposed to be more credible compared to results conducted on standard test databases.

In our research, we used face images only, not original video sequences. The single image approach was used in order to emphasize the impact of different methods on face template creation. Single face images were extracted from original IFaViD video sequences using the Viola-Jones face detector.

5 TEST RESULTS

The results of proposed methods for face templates creation are presented in the form of ROC (Receiver Operating Characteristics) curves in figure 3. ROC curves were obtained by a sweeping classifier's threshold. The threshold had a range of between 0 and 10,000 with steps of 100.

The medoid template method contains a threshold to remove outliers. This threshold was swept in all tests in the range from 2,500 to 6,000 with the step 500. The following graphs present the best-reached result of the medoid template method only. Threshold values are shown on graph labels (e.g. Medoid xxxx).

The results of the test conducted on the PubFig database indicate, that there is a significant performance difference when template databases containing one face template, were used. The centroid method outperforms both the medoid method and the baseline approach. Results also show that multiple template method may improve the system's performance.

The results in scenario A show that the medoid method performs relatively better than in the test conducted on PubFig. This could be caused by a high variability in the training images for scenario A. Multiple face template methods outperformed all single template methods. The three multiple template approach (MFTM 3) in particular has been shown to increase the CCR (Correct Classification Rate) by 10 percent compared to other methods. Considering the results in scenario A and the complex training database for scenario A, it can be stated that complex clusters are represented better by multiple-face templates.

Finally, the results in scenario B indicate that the multiple face template method has no meaning for training sets with low face pose variability. The centroid method produces the same results as the multiple face template method. The medoid method and the baseline approach seem to perform the same for databases with low variability.

A complex view of test results shows that different databases or scenarios require a different approach to the face template creation. Moreover, performance differences among tested methods can vary by up to 20% of the correct classification rate.

The overall system performance in all tests is relatively low compared to the state of the art approaches reported in recently published papers. This is caused by the influence of real-world conditions during the images captured by surveillance camera systems.

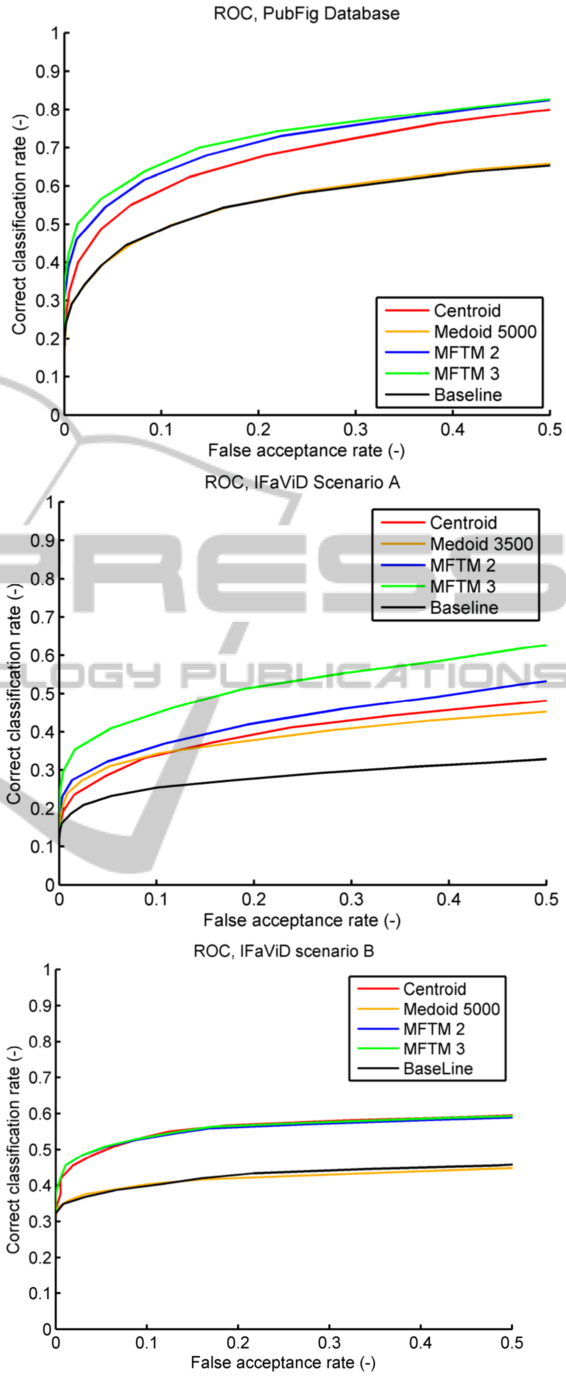

Figure 3: ROC curves describing face recognition system on different datasets. The Medoid stands for the medoid template method followed by the threshold value. The MFTM stands for the Multiple Face Template Method followed by the number of created face templates. The Baseline stands for the most similar face method.

6 CONCLUSIONS

This paper presents new approaches to face templates creation. Proposed methods were designed for real-world face recognition systems with large number of users in order to enhance face recognition system performance and to accelerate the classification process. The proposed methods produce one, two or three templates per person which reduces template database access time, classification time and redundancy among person's templates. This results in a more efficient recognition process. The impact of different face template creation methods has been examined on two databases, the PubFig database and the IFaViD database.

Firstly the tested face recognition system and its key algorithms have been described in detail.

Subsequently the examined methods for face template creation have been described. Face template examination and optimization is a novel step in face recognition research. Remarkable results have been obtained despite the fact that the proposed approaches to face template creation are only the first step in our research.

The next section has described the IFaViD and the PubFig testing and training databases and briefly summarizes their properties.

Finally the test results have been presented in the last section. The most remarkable finding is that different methods for face template creation have a significant influence on system recognition performance. The next finding is that there may be differences in recognition performance when different one-face template methods are used. The centroid approach outperforms the rest of one-face template methods. Moreover it does not require any extra setup, just like the medoid template method. The multiple template methods have outperformed all other methods and seem to offer a promising approach as they enable complex cluster representation.

ACKNOWLEDGEMENTS

Research described in the paper is financially supported by the Ministry of industry and trade of Czech Republic under grant IVECS, No. FR-TI3/170 and by Ministry of Education, Youth and Sports and European Structural Funds under grant SIX, No CZ.1.05/2.1.00/03.0072.

REFERENCES

Ahonen, A., Hadid, A. & Pietikäinen, M. (2006) *Face Description with Local Binary Patterns: Application to Face Recognition*. In Pattern Analysis and Machine Intelligence, *IEEE Transactions on*, vol.28, no.12, pp 2037-2041.

Bambuch, P., Malach, T., & Malach, J. (2012) Video database for face recognition. In *Technical Computing Bratislava 2012*. Bratislava, 7th November 2012. Bratislava: RT Systems, s.r.o & Systémy Priemyslovej Informatiky s.r.o. pp 12.

Kumar, N., Berg, A. C., Belhumeur, P. N. & Nayar, S. K. (2009) Attribute and Simile Classifiers for Face Verification. In 12th *International Conference on Computer Vision 2009*. Kyoto, 29th September – 2nd October 2009. pp. 365-372.

Malach, T., Bambuch, P. & Malach, J. (2012). Face Detection in video sequences. In *Radioelektronika 2012*. Brno, 17th -18th April 2012. Brno: Department of Radio Electronics, Brno University of Technology. pp 228-292.

Muja, M. & Lowe, G. D. (2009) Fast approximate nearest neighbors with automatic algorithm configuration. In *Conference on Computer Vision Theory and Applications*. Lisboa, 5th to 8th February 2009. Lisboa: INSTICC Press. pp. 331-340.

Prinosil, J. (2013) Local Descriptors Based Face Recognition Engine for Video Surveillance Systems. In *36th International conference on telecommunications and signal processing*. Rome, 2nd to 4th July 2013.

Stallkamp, J., Ekenel, H. K. & Stiefelhagen, R. (2007) Video-based Face Recognition on Real-World Data. In *IEEE 11th International Conference on Computer Vision.* Rio de Janeiro, 14th to 21st October 2007. pp: 1-8.

Viola, P. & Jones, M. (2001) Robust Real-Time Object Detection. *International Journal of Computer Vision.*

Zhao, W. & Chellappa R. (2006) *Face Processing: advanced modeling and methods*. Academic Press.