Haar-like Rectangular Features for Biometric Recognition

Nasrollahi, Kamal; Moeslund, Thomas B.; Rashidi, Maryam

Published in:
International Conference on Biometrics

DOI (link to publication from Publisher):
10.1109/ICB.2013.6612952

Publication date:
2013

Document Version
Early version, also known as pre-print

Haar-like Rectangular Features for Biometric Recognition

Kamal Nasrollahi, Thomas B. Moeslund, and Maryam Rashidi
Visual Analysis of People Laboratory
Sofiendalsvej 11, 9200 Aalborg, Denmark
kn@create.aau.dk

Abstract

Developing a reliable, fast, and robust biometric recognition system is still a challenging task. This is because the inputs to these systems can be noisy, occluded, poorly illuminated, rotated, and of very low-resolutions. This paper proposes a probabilistic classifier using Haar-like features, which mostly have been used for detection, for biometric recognition. The proposed system has been tested for three different biometrics: ear, iris, and hand vein patterns and it is shown that it is robust against most of the mentioned degradations and it outperforms state-of-the-art systems. The proposed system has been tested using three public databases.

1. Introduction

Biometric recognition, the process of human identification based on their behavioural and biological characteristics, is of great importance in many real-world applications, like for example in access control, surveillance and forensic applications. Such characteristics include but not limited to face, ear, iris, hand geometry, hand vein patterns, fingerprints, fingers' veins patterns, etc. Three of these biometrics, ear (due to its stable shape during the life and its relative invariance against change in pose and facial expressions), hand vein patterns (due to its uniqueness and robustness against imposter attacks as it is located beneath the skin), and iris (due to its uniqueness in the texture) have been chosen in this paper to develop a biometrics identification based system.

The recognition methods in such systems can be generally divided into two groups: appearance based and feature based. The classifiers in the first group use the gray scale values of the input images directly, while in the second group they work on some features that are extracted from the input images. Examples of the first groups are Principal Component Analysis (PCA)-based methods, Independent Component Analysis (ICA) algorithms, Linear Discriminant Analysis (LDA), and Neural Networks (NN). These classifiers have been well applied to the chosen biometrics. For example for ear images in [2] a PCA method and in [4] a NN method have been used. For hand vein patterns a 2D PCA classifier in [6], and ICA based classifiers in [8] have been reported. For iris recognition a Probabilistic NN (PNN) in [11] and an LDA classifier in [12] have been used.

Several features based classifiers are also reported in the literature for the chosen biometrics. For example for ear images a Force Filed Transform based classifier in [7], a structural feature based method in [13], and a Gabor filter based method in [11] can be mentioned. Among the feature based methods for hand vein pattern in [16] the end points and the crossing points of the veins, in [17] the Scale Invariant Feature Transform (SIFT) features, in [5] Non-negative Matrix Factorization based features and in [3] Line Edge Mapping based features have been used. In [11] Gabor filters have been used for iris recognition.

The input images, from which the biometric data are extracted, in the appearance based algorithms usually need to be registered to a fixed frame. Therefore these methods are sensitive to registration errors. The performance of the feature based methods directly depends on the effectiveness and robustness of the employed features. Furthermore, the performance of both group of algorithms degrades when the input images are noisy, occluded by some obstacles, of low resolution, and not properly illuminated.

This paper proposes a feature based approach for biometric recognition which is shown to be robust against most degradations and poor imaging conditions. Furthermore, it is shown that it does not need a prior registration of the input images to a fixed frame. The employed features are obtained from integral images using Haar-like rectangular filters which have mostly been used for rapid and robust object detection using a boosted cascade of simple weak classifiers [15]. To the best of our knowledge, these features have not been used for recognition in the literature, except in [20] where they have been used for face recognition in the same way they have been used for detection using the same classifiers which has resulted in a poor recognition rate around

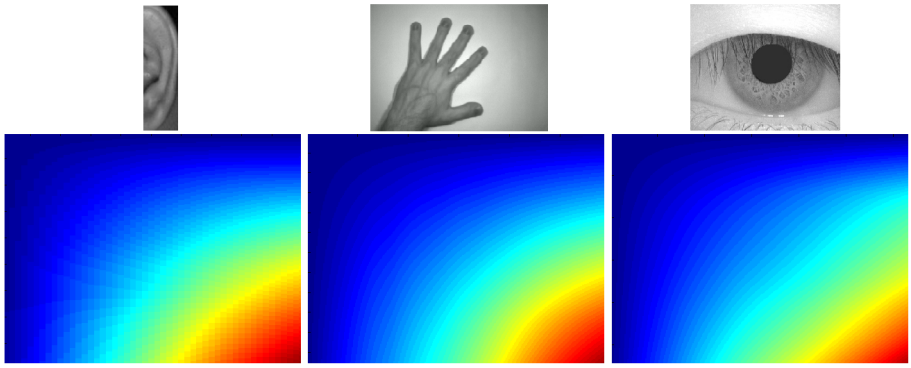

Figure 1. Three different biometric images and their corresponding integral images. Thought the visualizations of the integral images seem very similar, it is shown in this paper that they carry important information which can be used for recognition purposes.

70%. Though the detection performance of Haar-like features is very good, they result in poor recognition rates if they are used with the same original boosted cascade classifier. The poor recognition performance of this classifier might has been the reason for ignoring these features for recognition purposes. In this paper we show that a set of these features combined with a PNN classifier results in a robust and fast biometric based recognition system, for different biometrics.

In a currently under-review paper [1] (uploaded as an anonymous supplement to this paper) we have shown that a combination of these features (61 features) can be used for face recognition using a PNN classifier. Here we show that extending the number of the features to 115 using the same classifier can indeed be used for recognition purposes using other biometrics than face. This results in a generic biometrics based identification system. To document the generality of the system we have used ear, hand vein patterns and iris images. The experimental results on public databases show that the proposed method outperforms the state-of-the-art biometric systems.

The rest of this paper is organized as follows: in section 2 we revisit the integral images and the Haar-like rectangular features that are extracted from these images. Section 3 provides a brief description of the employed classifier. The experimental results using public databases are reported in section 4. Finally section 5 concludes the paper.

2. Integral Images and Haar-like Features

In integral images which were first introduced by Viola and Jones [15] each pixel simply represents a row and column cumulative information of all the pixels that are located on the left and above of the current pixel. In another words, given an input image as i, its counterpart integral image is defined as:

$$ii(x,y) = \sum_{k=1}^{x}\sum_{l=1}^{y} i(k,l) \tag{1}$$

Using a temporary variable which calculates the cumulative row (or the cumulative column) for a pixel, the integral value of that pixel can be calculated using a one pass algorithm over the image. For example if we assume that the temporary variable, t, calculates the cumulative row of every pixel as:

$$t(x,y) = t(x-1,y) + i(x,y) \tag{2}$$

Then the integral value of each pixel is obtained as:

$$ii(x,y) = ii(x,y-1) + t(x,y) \tag{3}$$

where the initial conditions are zero. The integral images of three biometrics obtained by Eq. 3 are shown in Figure 1.

The reason for defining the integral images is that they can very efficiently be used for producing the values of Haar-like rectangular features. These features in the simplest case consist of two rectangular regions: a black region and a white region (see Figure 2 (a)). The value of the feature is equal to the value of the white region minus the value of the black region. The integral image is used for calculating the values of these regions. Having an integral image, the value of such a region is equal to summation of integral pixels located on the vertexes of the main diagonal of the region minus the summation of the integral pixels located on the vertexes of the other diagonal of the region. For example for the zoomed feature shown in Figure 2 (a), the values of the white and black regions are equal to: $white = A+E-F-B$, and $black = B+D-E-C$. Then, the value of the feature is equal to: $white - black$. If the number of the regions of a filter is more than two, the final value of the feature simply equals to the summation of values of the white regions minus the summation of the values of the black regions.

The main difference in using these features for recognition and detection is shown in Figure 2(b). When they are used for detection, they are usually applied to a resolution pyramid of the integral images of the input image to find

the area of the interest in the image. To do so, the rectangular filters of the features scan every possible positions of the input image in the resolution pyramid and calculate the value of the feature in those positions. Then, the object of interest is found if the values of a set of these features are higher than some thresholds. However, when these rectangular features are used for recognition, like in this paper, it is assumed that the object of interest is already detected. Therefore, the rectangular filter is resized to the size of the input image, while the ratios of white and black regions are not changed, and the value of the feature is calculated.

It can be seen from Figure 3 that many features (115 features) have been used for the recognition purpose in this paper. Following [19] these features are generated by consecutive division of the entire area of the input image to 2, 3, ..., 20 regions.

Having calculated the values of the Haar-like rectangular features of Figure 3, they are fed to the classifier which is explained in the next section.

3. The Employed Classifier

PNN was first introduced in [14] and later on used in many pattern recognition applications. The PNN's success in classification problems is mainly due to its ability in learning complicated spaces without getting lost in local minimums.

PNN uses the Parzen window to estimate the Probability Distribution Functions (PDF)s of the classes that are involved in the recognition process. Such a window is of the following form:

$$f_j(s) = \frac{1}{\sigma_j n_j} \sum_{k=1}^{n_j} W(\frac{s - s_{kj}}{\sigma_j}) \tag{4}$$

where f_j is the PDF of the jth class, n_j is the number of the samples of this class, σ_j is a smoothing parameter, s_{kj} is the kth training sample of the jth class, s is the unknown sample, and W is a weighting function which in PNN is replaced by an exponential function to make the employed PDFs of the PNN to be of a Gaussian form as:

$$f_j(s) = \frac{1}{\sqrt{2\pi}\sigma_j n_j} \sum_{k=1}^{n_j} \exp(\frac{s - s_{kj}}{2\sigma_j^2}) \tag{5}$$

since the number of the involved features in this paper are more than one (115) s becomes a vector, therefore the distance between the training samples and the unknown sample is measured using the l_2-norm function resulting in the following PDF:

$$f_j(s) = \frac{1}{(2\pi)^{p/2}\sigma_j n_j} \sum_{k=1}^{n_j} \exp(-\frac{||s - s_{kj}||^2}{2\sigma_j^2}) \tag{6}$$

where p is the size of the feature vectors (here 115). PNN applies this PDF using a four-layered structure:

- Input layer which provides the input data to the next layer.
- Pattern layer which calculates the normalized distance of the training samples to the unknown sample (parameter of $\exp$ function in the above PDF).
- Summation layer which applies the summation and the normalization parts of the above PDF.
- Output layer which applies a winner-takes-all process to the outputs of the previous layer.

4. Experimental Results

In this section we first mention the details of the implemented system, then the employed public databases are described. Then, the proposed system is compared against state-of-the-arts using the public databases. Then, several tests are performed on the proposed system to show its robustness against different imaging conditions and degradations. Finally, the training performance of the proposed system and its speed are further analyzed.

4.1. Implementation

The number of the processing units in each of the layers of the employed PNN is as follows:

- Input layer: 115 units which is equal to the size of the used feature vector in this paper.
- Pattern layer: number of the samples of each class multiplied by the number of the classes.
- Summation layer: the same as the number of the classes.
- Output layer: one unit showing the recognized class's identifier.

The used PNN uses gradient descent algorithm for training. To do so, each database is partitioned into three parts: training, cross-validation and testing parts. The results reported in this paper are obtained when the sizes of these partitions are respectively 60%, 15%, and 25% of the size of the databases. Though it seems that the size of the training portion is much larger than the other two portions, it is shown in section 4.5 that as long as the number of the training samples per subjects are high enough reducing the size of the training portion does not reduce the recognition rate of the system too much. Furthermore, since the number of the training samples in some of the employed databases are very few, around 3, the databases are duplicated by a factor of three in which the added images are generated from the existing ones by by a random operation like adding random Gaussian noise, random uniform noise, random translations, or random rotations, all by small random parameters.

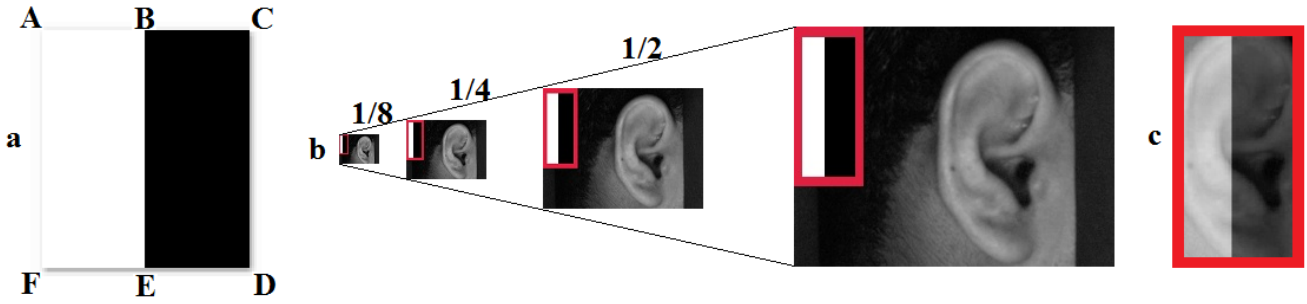

Figure 2. a) a typical Haar-like rectangular feature and its application for b) detection, and c) recognition purposes.

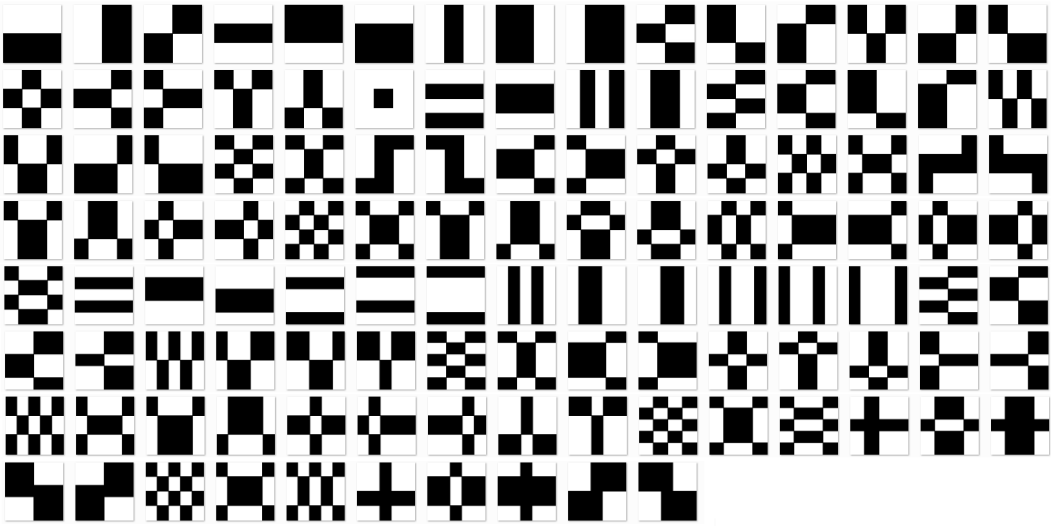

Figure 3. The employed Haar-like rectangular features, inspired by [19].

4.2. Databases

Three public databases have been used in this paper, one for each of the three different biometrics, to show the generality of the proposed system. These databases are as follows:

- **Ear Database** which contains two sub-databases:
  - **Ear Database 1 (ED1)** [11] which contains 465 images of 125 subjects taken in an indoor environment without any additional illumination. There are at least three samples per subject in this database (Figure 4(a)).
  - **Ear Database 2 (ED2)** [11] which contains 753 images of 221 subjects. Figure 4(b) shows some of the samples of this database. The images of these two databases are of size 50×180 pixels.
- **Hand Vein Database (HVD)** [18] which contains 1200 near infrared images of the vein patterns of the left hands of 100 subjects. This database contains four sub-databases as:
  - **HVD1**: The hand is in a normal condition.
  - **HVD2**: The user has just carried a bag.
  - **HVD3**: The user has just pressed a ball, and
  - **HVD4**: The user has just cooled his/her hand by ice.

  The images of these four databases are of size 300×240 pixels. Sample images of these databases are shown in Figure 4(c).
- **Iris Database (ID)** [10] this database contains 2240 images of 224 subjects each providing 10 grayscale iris images. The sizes of the images are 320×240 pixels (Figure 4(d)).

The images in the employed databases are not aligned to a strictly fixed frame and there are some rotations, scaling and translations among both the samples of the same subjects and from images of a subjects to another one.

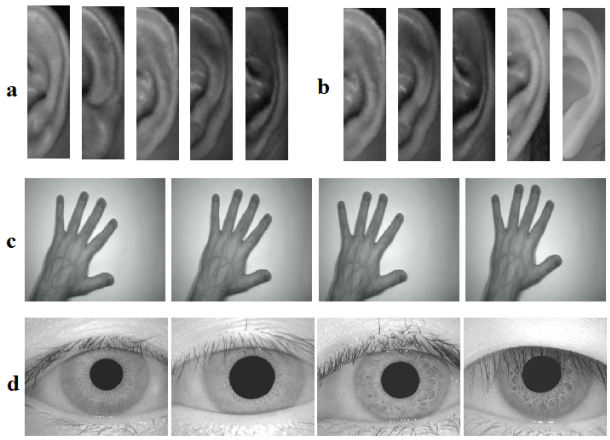

Figure 4. The employed databases: a) ED1, b) ED2, c) HVD1-HVD4, and d) ID.

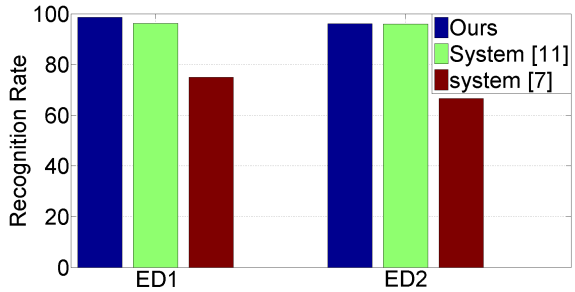

Figure 5. Recognition rates of the proposed system vs. state-of-the-art ear-based recognition systems.

4.3. Comparing to the state-of-the-arts

The results of the proposed system are compared here against the state-of-the-art systems using the above mentioned databases. The results of the proposed system for the ear are compared against the very recent publication of [11] in which the localized orientation information and local graylevel phase information are used in complex Gabor filters. Besides the best results reported in [11], the Force Field Transform (FFT) method of [7] has also been included in the comparisons. The results of the FFT method on the employed ear databases are taken from [11]. The comparison of the recognition rates is shown in Figure 5.

The results of the proposed system for hand vein patterns are compared against the systems that are reported in [18]. The first and second systems (System 1 and System 2 in Figure 6) are based on ICA Architecture 1 and Architecture 2, respectively [8], the third system (System 3 in Figure 6) Non-negative Matrix Factorization [5] and the last system (System 4 in Figure 6) Line Edge Mapping [3]. The results are shown in Figure 6.

The results of comparing the recognition rate of the proposed system for iris recognition against the state-of-the-art

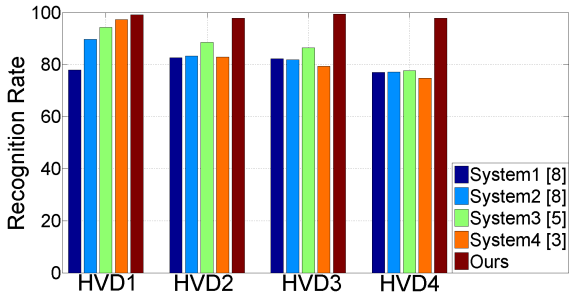

Figure 6. Recognition rates of the proposed system vs. state-of-the-art hand vein patterns-based recognition systems.

Method	S1	S2	S3	S4	Ours
Recognition Rate	86.9	88.8	89.9	91.8	93.9

Table 1. Recognition rates of the proposed system vs. state-of-the-art iris-based recognition systems, please see the text for the descriptions of S1-S4.

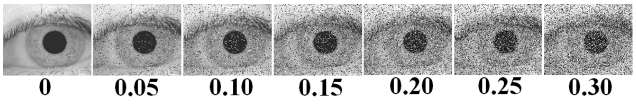

Figure 7. A typical image from ID database corrupted by different levels of a salt and pepper noise.

iris recognition systems are shown in Table 1. The involved systems are: S1 is a decision tree based classifier, S2 is an appearance based PNN classifier, S3 is a Support Vector Machine classifier, S4 is a fuzzy binary decision tree-based classifier, and PS is the proposed system. The results of S1-S4 on the employed iris database are taken from [9]

It should be mentioned that in contrary to most prior works on iris and hand vein recognition, the proposed method does not need precise extraction of iris texture and hand patterns. The images shown in Figure 4 are fed to the proposed method directly without any preprocessing.

Having shown that the proposed method outperforms the state-of-the-art systems in terms of recognition rates, its robustness is analyzed in the next subsection.

4.4. Robustness of the Proposed System

The robustness of the proposed system against three imaging degradation has been studied here: noise, occlusion, and downsampling. In the first case, different levels of salt and pepper noise has been added to the input images. The noise level, d, changing from 0.05 to 0.3 affect $d \times num(i)$, where $num(i)$ is the number of the pixels of the input image, i. A typical image affected by different levels of noise is shown in Figure 7.

For each level of corruption, all the images of the database are affected by the noise and a new database is

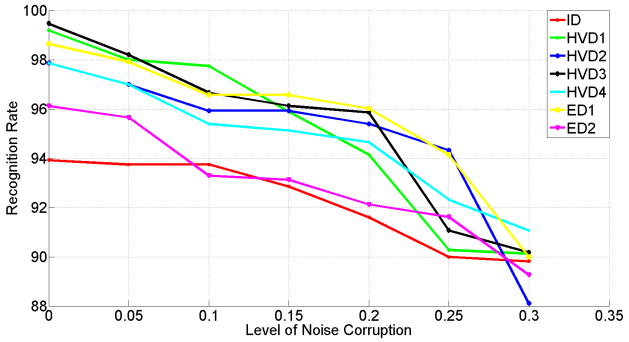

Figure 8. The recognition rate of the proposed system against different levels of salt and pepper noise corruption using the employed databases.

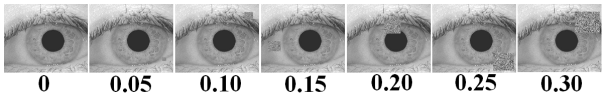

Figure 9. A typical image from ID database occluded by different levels of occlusions.

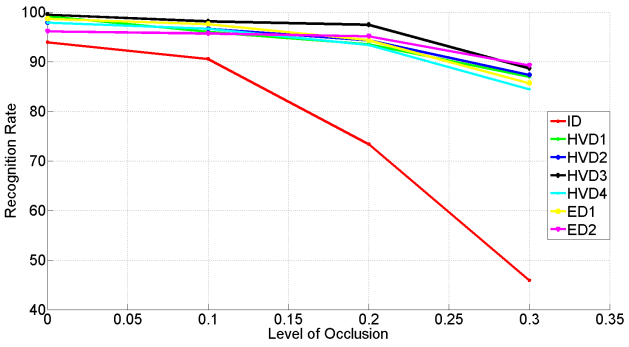

Figure 10. The recognition rate of the proposed system against different level of occlusion using the employed databases.

generated. Then the system gets learned and tested using this database by portions mentioned in section 4.1. The results are shown in Figure 8.

In the second case, different portions of the images (changing from 0.05 to 0.3 of the size of the images) of database are occluded by blocks of noise where the location and the exact shape of the block (horizontal or vertical rectangles) are determined randomly. Such different levels of occlusion are shown in Figure 9.

For each level of occlusion a new database is generated by which the proposed system is trained and tested. The results are shown in Figure 10.

In the third case, the images are down-sampled by the following factors: 1/2, 1/4, 1/8, and 1/16 (Figure 11).

The recognition rates of the proposed system for these down-sampled images are shown in Figure 12.

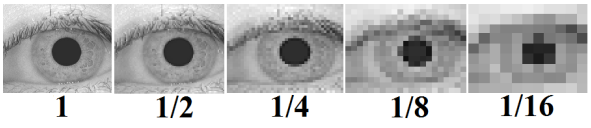

Figure 11. A typical image from ID database downsampled by different downsampling factors.

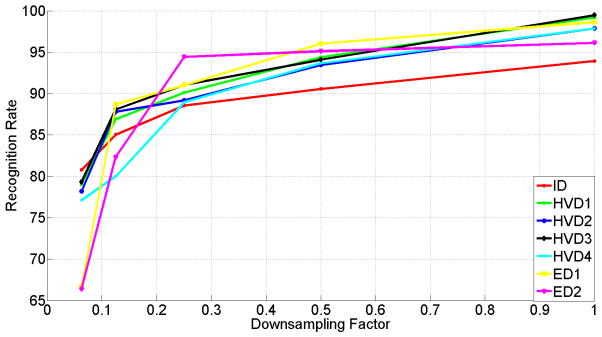

Figure 12. The recognition rate of the proposed system against different downsampling factors using the employed databases.

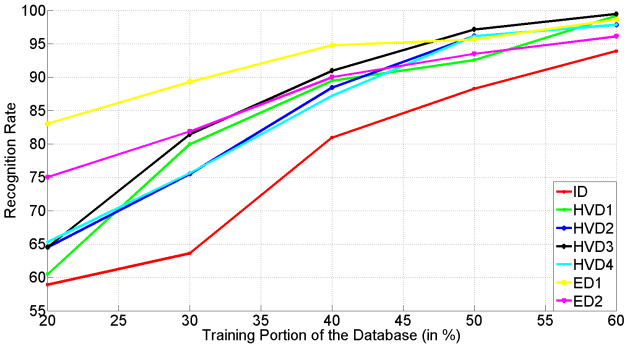

Figure 13. The recognition rate of the proposed system against different amounts of training data using the employed databases.

4.5. Further Analysis of the Proposed System

In this section the training of the proposed system and its speed are analyzed. As mentioned in section 4.1 the above reported results are obtained when the employed databases are partitioned into parts of sizes 60%, 15%, and 25% of the entire database for training, cross-validation, and testing, respectively. Though this might seem as a drawback for the proposed system that it uses around 60% of the databases for training, it is shown in Figure 13 that using smaller portions of the databases for training still produces high recognition rates. The only thing that needs to be respected is that the proposed system needs around 6 images per subject for training.

Table 2 shows the timing requirements of the proposed system for both feature extraction and classification on an Intel 3.4 GHz CPU with 16GB RAM. It can be seen that the

Time required for \ Databases	ID	HVD1	HVD2	HVD3	HVD4	ED1	ED2
Feature Extraction	0.281	0.113	0.116	0.101	0.119	0.106	0.107
Classification	0.014	0.012	0.012	0.012	0.012	0.013	0.015

Table 2. The average required time (in seconds) of testing the proposed system per image for the employed databases, when 25% of each database has been used for testing.

proposed system is working very fast.

5. Conclusion

This paper has proposed a biometrics-based identification system using Haar-like features which have mostly been used for detection. These features are used for training a probabilistic classifier which at the end results in a generic biometrics-based identification system. To document the generality and the efficiency of the proposed system, the experimental results on public databases of three very different biometrics are reported. It is shown that the proposed system outperforms state-of-the-art systems.

Studying the importance of the individual features and their relative weighting are among our future works for extending this work.

References

[1] Anonymous Currently Under-review Submission. Robust real-time face recognition using haar-like features obtained from integral images. 2013.

[2] K. Chang, K. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1160–1165, sept. 2003.

[3] B. Gokberk, A. Salah, and L. Akarun. Rank-based decision fusion for 3d shape-based face recognition. In *Signal Processing and Communications Applications Conference, 2005. Proceedings of the IEEE 13th*, pages 364–367, may 2005.

[4] L. Gutierrez, P. Melin, and M. Lopez. Modular neural network integrator for human recognition from ear images. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–5, july 2010.

[5] P. O. Hoyer and P. Dayan. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

[6] C.-B. Hsu, S.-S. Hao, and J.-C. Lee. Personal authentication through dorsal hand vein patterns. *Optical Engineering*, 50(8):087201–087201–10, 2011.

[7] D. J. Hurley, M. S. Nixon, and J. N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98(3):491–512, 2005.

[8] A. Hyvrinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[9] A. Kumar, M. Hanmandlu, A. Das, and H. Gupta. Biometric based personal authentication using fuzzy binary decision tree. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 396–401, april 2012.

[10] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, 43(3):1016–1026, 2010.

[11] A. Kumar and C. Wu. Automated human identication using ear imaging. *Pattern Recognition*, 45:956–968, 2012.

[12] C. Liu and M. Xie. Iris recognition based on dlda. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 489–492, 2006.

[13] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi. Shape and structural feature based ear recognition. In S. Li, J. Lai, T. Tan, G. Feng, and Y. Wang, editors, *Advances in Biometric Person Authentication*, volume 3338 of *Lecture Notes in Computer Science*, pages 663–670. Springer Berlin Heidelberg, 2005.

[14] D. Specht. Probabilistic neural networks. *Neural Networks*, 3:109–118, 1990.

[15] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1, 2001.

[16] K. Wang, Y. Zhang, Z. Yuan, and D. Zhuang. Hand vein recognition based on multi supplemental features of multi-classifier fusion decision. In *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on*, pages 1790–1795, june 2006.

[17] Y. Wang, Y. Fan, W. Liao, K. Li, L. Shark, and M. Varley. Hand vein recognition based on multiple keypoints sets. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 367–371, april 2012.

[18] A. Yuksel, L. Akarun, and B. Sankur. Biometric identification through hand vein patterns. In *Signal Processing and Communications Applications Conference (SIU), 2010 IEEE 18th*, pages 708–711, april 2010.

[19] C. Yuriy. Face detection c++ library with skin and motion analysis. In *Biometrics AIA 2007 TTS*, 2007.

[20] Z. Zhu, T. Morimoto, H. Adachi, O. Kiriyama, T. Koide, and H. J. Mattausch. Multi-view face detection and recognition using haar-like features. In *COE Workshop*, 2006.