Haar-like Features for Robust Real-Time Face Recognition

Nasrollahi, Kamal; Moeslund, Thomas B.

Published in:
IEEE International Conference on Image Processing

DOI (link to publication from Publisher):
10.1109/ICIP.2013.6738633

Publication date:
2013

Document Version
Early version, also known as pre-print

HAAR-LIKE FEATURES FOR ROBUST REAL-TIME FACE RECOGNITION

Kamal Nasrollahi and Thomas B. Moeslund

Laboratory of Visual Analysis of People (VAP)
Aalborg University
Sofiendalsvej 11, 9200 Aalborg, Denmark

ABSTRACT

Face recognition is still a very challenging task when the input face image is noisy, occluded by some obstacles, of very low-resolution, not facing the camera, and not properly illuminated. These problems make the feature extraction and consequently the face recognition system unstable. The proposed system in this paper introduces the novel idea of using Haar-like features, which have commonly been used for object detection, along with a probabilistic classifier for face recognition. The proposed system is simple, real-time, effective and robust against most of the mentioned problems. Experimental results on public databases show that the proposed system indeed outperforms the state-of-the-art face recognition systems.

***Index Terms*—** Haar-like rectangular features, face recognition, integral images.

1. INTRODUCTION

Robust and real-time face recognition is a very important step in many applications like access control, surveillance scenarios, gaming, human-computer interaction, etc. Many methods have been proposed in the literature for solving this problem, indulging but definitely not limited to: Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Support Vector Machines (SVM) [3] Independent Component Analysis (ICA) [4], Local Binary Pattern (LBP) [5], and more recently Sparse Representation (SR) based methods [6]-[7]. A recent survey on face recognition algorithms can be found in [8].

Though many of the available face recognition approaches result in real-time high recognition rates when the input image(s) follow some certain conditions, they mostly fail when the input is of low-resolution, rotated, noisy, occluded, contains facial expressions, is not in focus, and not properly illuminated [8]. These issues make the face recognition problem still a challenging problem especially when face images are of very low-resolution like 6×6. The common approach for dealing with such faces is to use super-resolution algorithms [9] and [10] for enhancing the quality and increasing the size of the input images. Beside imposing extra computations to the recognition system such super-resolution algorithms very often constrain their input images to have some specific conditions. For example they mostly need multiple probes of the same object where the probes should have relative sub-pixel displacements. Then, these very low-resolution images should be registered to a common framework using a proper registration algorithm. Finding the right registration algorithm is itself a challenging problem. These, all impose extra computations on the final system and make it more complicated and at the same time more sensitive to errors in any of these steps. When it comes to the rotated input faces, the face recognition problem gets more unstable and in most of the cases using a 3D model of the face [11] and [12] is required which again fail when faces are of very low-resolution. For dealing with occlusion and noise among other methods [8], SR methods [6]-[7] which generally try to represent a probe as a sparse linear combination of the galleries have gained much attention.

The proposed system in this paper deals with many of the mentioned challenges in a robust and real-time approach. It uses Haar-like features extracted from integral images to train a Probabilistic Neural Network (PNN) as the classifier. Integral images were introduced by Viola and Jones for rapid Haar-like features extraction for the purpose of object detection by AdabBoost classifier in [13] more than a decade ago. Since then, Haar-like rectangular features extracted from integral images and AdaBoost classifier have widely been used for detecting objects of different categories from human faces [13], to hand gesture [14], to vehicles [15], to name a few.

Though AdaBoost with Haar-like features is very efficient for object detection, it does not result in good performance for object recognition. For example it has been used for face recognition in [16] the same way it has been used for face detection, and it has resulted in a recognition rate of 75% on a local database. This is actually the only available work in which Haar-like features obtained from integral images have been used for face recognition. In all the other research papers involving Haar-like features, these features have been used for detection and later on the detected object has been recognized using a different set of features/methods. Possibly the weak recognition performance of the AdaBoost classifier has been the reason for neglecting the interesting Haar-like features for the recognition purposes. The proposed system in this paper revisits Haar-like features but with a different classifier for the purpose of face recognition. To do so, the input face image of the system is first converted to an integral image, then the rectangular Haar-like filters are applied to the integral image to extract the corresponding features, these features are then fed to a previously trained PNN classifier for the final recognition.

The rest of this paper is organized as follows: the next section revisits the integral images of Viola and Jones and the Haar-like features which have been extracted from these integral images for the purpose of recognition. Section 3 explains the employed classifier. Experimental results and in-depth evaluations of the proposed system using three different public databases are presented in Section 4 where the proposed system is compared against state-of-the-art face recognition systems in different testing scenarios. Finally Section 5 concludes the paper.

2. INTEGRAL IMAGES AND HAAR-LIKE FEATURES

Following [13] for a given input image as i its counterpart integral image, ii, is defined as:

$$ii(x,y) = \sum_{x' \leq x} \sum_{y' \leq y} i(x', y') \tag{1}$$

It means that the value of the integral image at a given location is equal to the summation of all the pixels located at the above and the left of that location. Using a temporary image variable as:

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

with the following initial conditions:

$$t(x,-1) = 0, ii_1(-1,y) = 0 \tag{3}$$

the integral image can be calculated in one pass over the image using:

$$ii(x,y) = ii(x-1,y) + t(x,y) \tag{4}$$

Having obtained the integral image of the input image, it is used to extract Haar-like features. The rectangular Haar-like features that have been used for face recognition in this paper are similar to those used for face detection in [13]. These features are shown in Fig. 1. Integral images provide the possibility of rapid calculation of rectangular Haar-like features. The value of each feature is simply obtained by subtracting the (summation of the) value(s) of the black area(s) from the (summation of the) value(s) of the white area(s). The value of the black and white areas for example for the zoomed filter in this figure are simply equal to $B + D - E - C$ and $A + E - F - B$, respectively.

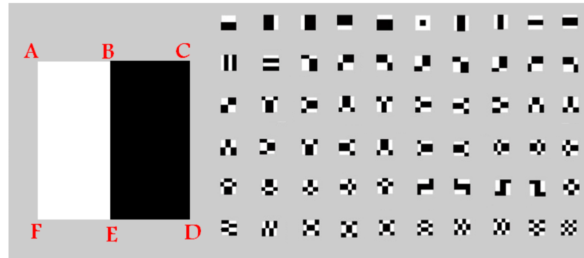

Fig. 1. The rectangular Haar-like features used in this paper for face recognition.

The main difference between the application of these features for detection like in [13]-[15] and for recognition as in here is that in the detection case these filters are scanning every pixel locations of the input image in many different scales while in the recognition case it is assumed that the face image is already detected, therefore the rectangular filters are resized to the same size of the face image.

3. THE CLASSIFIER

This section provides a brief introduction to the employed classifier, a PNN, which was first introduced in [17] and later on used in many pattern recognition problems, e.g. [18]-[19], to name a few. PNN is a fast implementation of Kernel Discriminant Analysis and can perform non-liner classification without getting lost in the local minimums. This classifier is obtained by replacing the sigmoid activation function of the feed forward neural networks by a Gaussian function as:

$$f_A(\boldsymbol{X}) = \frac{1}{m(2\pi)^{p/2}\sigma^p} \sum_{i=1}^{m} \exp[-\frac{(\boldsymbol{X}-\boldsymbol{X}_{A_i})^T(\boldsymbol{X}-\boldsymbol{X}_{A_i})}{2\sigma^2}] \tag{5}$$

where $\boldsymbol{X} = [X_1, X_2, ..., X_p]$ is the p-dimensional input feature vector (here the Haar-like rectangular features), i is the pattern number, m is the total number of training patterns, $\boldsymbol{X}_{A_i}$ is the ith training pattern from class θ_A, and σ is a smoothing parameter.

PNN composes of four layers: an input layer which provides the input data to the next layer, a pattern layer which calculates the normalized distance of the training samples to the unknown sample, a summation layer which applies a summation and a normalization operations to the data coming from the second layer, and an output layer which applies a winner-takes-all process to the outputs of the third layer. The number of the input units in the input layer of the developed PNN for the proposed system is equal to the number of the employed Haar-like rectangular features, 61 features, and the number of the output units in the output layer is the same as the number of the desired classes, the number of the units in the pattern layer is equal to the the number of the samples of each class multiplied by the number of the classes, and the number of the units in the summation layer is the same as the number of the classes.

4. EXPERIMENTAL RESULTS

The proposed system has been tested using three public databases: ORL ATT Face Database [20], UMIST [21], and Faces94 [22]. The number of the images in these databases are 400 (of 40 subjects), 564 (of 20 subjects), and 3060 (of 153 subjects), and the size of the images are 92×112, 92×112 , and 105×120 pixels respectively. These images contain variations in head-pose, expression, and illumination conditions. The first two databases have been used for comparing the proposed system against state-of-the-art systems. The third database has been added to the other two when the system has been further evaluated.

For doing every test the employed databases are divided into three parts for training, cross-validation, and testing. The best results for the proposed system are obtained when the sizes of these parts are 60%, 15%, and 25%, respectively, of the entire database. However, it is shown later in this section that using smaller portions of the databases for training does not reduce the recognition rate of the system too much as long as there are enough number of samples per each subject.

4.1. The Proposed System vs. State-of-the-art

To compare the proposed system against state-of-the-art systems some known and advanced algorithms like PCA [1], LDA [2], SVM [3], ICA [4], and LBP [5], and some very recent systems DLRDSR [7], FDDL [23], RPCA [24], and SR [6] are chosen. The results of the very recent works of RPCA, SR, FDDL, and DLRDSR are published in [7]. The proposed system has been compared against these algorithms in three different testing scenarios as explained below.

In the first test the images from ORL and UMIST databases without any modifications are fed to the proposed system and the nine other face recognition algorithms. Similar number of images have been used for both training and testing of all the systems. The resulted recognition rates are shown in Fig. 2.

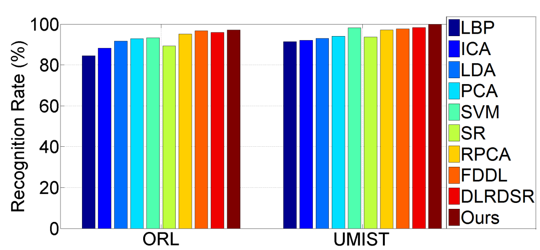

Fig. 2. The recognition rates of the proposed system vs. state-of-the-art algorithms using ORL and UMIST databases.

It can be seen from this figure that the proposed system clearly outperforms the other methods. Indeed the recognition rate of the proposed system for the UMIST databases is 100%.

In the second and third tests the robustness of the proposed system against noise in the input image and occlusion are studied, respectively. To corrupt the images for the second test an uniform noise has been used where the level of the corruption changes from zero to 25% of the entire pixels (Fig. 3 top). To generate occluded images, blocks of noise which cover zero to around 50% of the images are used. The location of the block of noise and the exact shape of the noise block are defined randomly (Fig. 3 bottom).

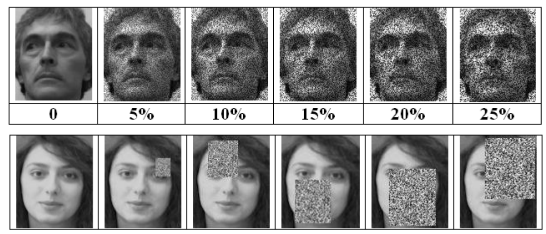

Fig. 3. A sample image from top) UMIST and bottom) ORL databases, corrupted by different levels of an uniform noise and occlusion, respectively.

The results of the second and third tests are shown in Fig. 4 and Fig. 5, respectively.

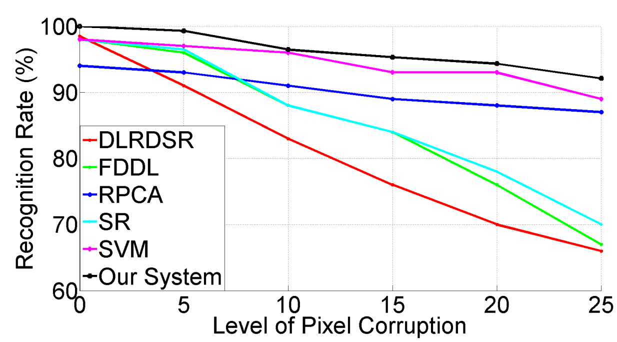

Fig. 4. The proposed system vs. other face recognition systems where the inputs taken from UMIST are corrupted by different level of uniform noise.

4.2. Further Evaluation of the Proposed System

Having shown the robustness of the proposed system against noise and occlusion in the input, in this subsection we first show that it is robust against down-sampling as well. Then, the effect of the size of the training database has been studied. Next, it is shown that the proposed system works in real-time. For all of these further evaluations the third database, Faces94, has also been added.

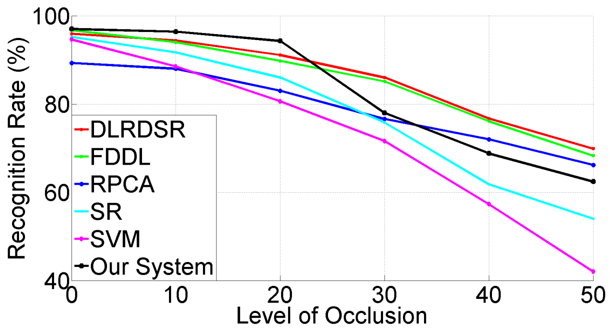

Fig. 5. The proposed system vs. similar systems using occluded input images from ORL database.

To show the robustness of the proposed system against down-sampling, its recognition rate has been monitored when the sizes of the input images decrease. To do so, all the images in all the three databases are down-sampled by factors of 2, 4, 8, and 16 to generate new databases. Then the new databases are used for training and testing the proposed system. The results are illustrated in Fig. 6 for all the three employed databases.

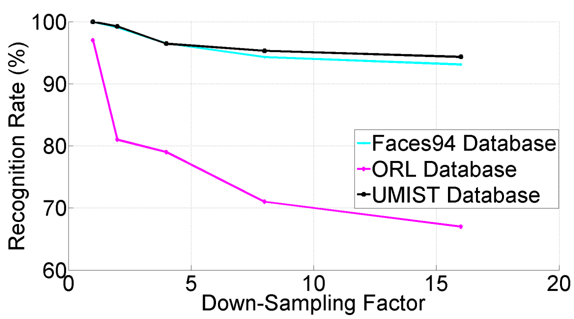

Fig. 6. The recognition rates of the proposed system vs. different down-sampling factors using three public databases: Faces94, ORL Face Database, and UMIST.

The smallest down-sampled images, almost of size 5×7 pixels, hardly resemble a face, but the recognition rates of the proposed system for such images are still very high at least for images coming from Faces94, and UMIST databases. Despite the fact that the sizes of the images in ORL database are very similar to those in Faces94, the recognition rates of the system for images coming from ORL are lower than those for Faces94. This is because the number of the samples per test subject in ORL is less than half of those for Faces94. Therefore, as long as the amount of the training samples per subject are high enough (clarified below) the system can recognize faces of very small sizes.

Now the results of analyzing the proposed system for finding a training database of a proper size and the required number of training samples per objects are given. To do so, the recognition rates of the proposed system are monitored when different portions of the available databases are used for training. The results are shown in Fig. 7. It can be seen from this figure that as long as the amount of the training samples are high enough (depending on the required recognition rate) the drop in the recognition rate of the system is not too much.

Database \ Time for	A	B	C	D
ORL	1.63	5.7	2.4	0.0243
UMIST	6.30	9.3	4.6	0.0358
Faces94	17.68	25.3	11.8	0.0179

Table 1. The speed of the system (in seconds). A, B, and C are the required time for feature extraction, training and cross-validation, testing, respectively and D is the average testing time per image.

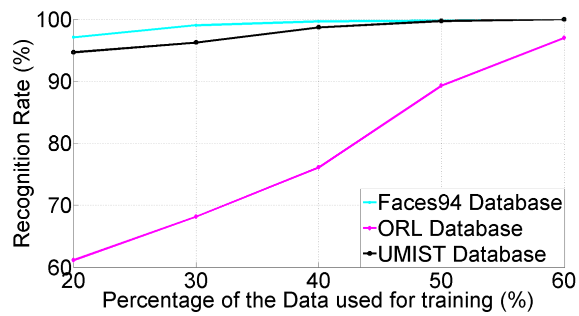

Fig. 7. The effect of the size of the training data on the recognition rate of the proposed system.

The required time for each part of the system on a 3.4GHz Intel CPU is shown in Table 1. The average time shown in the last column of the table, which is simply the average of the first three columns, shows that the system works in real-time.

5. CONCLUSION

This paper has presented a novel approach to face recognition using Haar-like features extracted from integral face images. The extracted features are then used for learning the complicated space of human faces in the Haar-like feature domain using a probabilistic classifier. The proposed system has been tested using public databases and the obtained results show that it indeed outperforms the state-of-the-art face recognition systems. Several experimental results have been considered for testing the system against different imaging degradation including noise corruption, occlusion, head-pose and down-sampling (to very small sizes). Sensitivity analysis of the involved features and more interestingly generalization of the system to objects of other types rather than human faces are among our future options for extending this system.

6. REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[3] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: global versus component-based approach," in *Proceedings of IEEE International Conference on Computer Vision*, 2001, vol. 2, pp. 688–694.

[4] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1466, 2002.

[5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns:application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[6] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2008.

[7] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2586–2593.

[8] R. Chellappa, J. Ni, and V.M. Patel, "Remote identification of faces: Problems, prospects, and progress," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1849–1859, 2012.

[9] P. Hennings-Yeomans, S. Baker, and B.V.K. Vijaya-Kumar, "Simultaneous super-resolution and feature extraction for recognition of low resolution faces," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[10] W.W.W. Zou and P.C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012.

[11] J. Yu, B. Bhanu, and N. Thakoor, "Face recognition in video with closed-loop super-resolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 39–45.

[12] D. Zhang, J. He, and M. Du, "Morphable model space based face super-resolution reconstruction and recognition," *Image and Vision Computing*, vol. 30, no. 2, pp. 100–108, 2012.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted classifier of simple features," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 511–518.

[14] Q. Chen, N.D. Georganas, and E.M. Petriu, "Hand gesture recognition using haar-like features and a stochastic context-free grammar," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 8, pp. 1562–1571, 2008.

[15] A. Haselhoff and A. Kummert, "A vehicle detection system based on haar and triangle features," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2009, pp. 261–266.

[16] 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 *Proceedings of COE Workshop*, 2006.

[17] D.F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, pp. 109–118, 1990.

[18] C.H. Chen and C.T. Chu, "Low complexity iris recognition based on wavelet probabilistic neural networks," in *Proceedings of IEEE International Joint Conference on Neural Networks*, 2005, vol. 3, pp. 1930–1935.

[19] Z. Sankaria and H. Adeli, "Probabilistic neural networks for diagnosis of alzheimer's disease using conventional and wavelet coherence," *Journal of Neuroscience Methods*, vol. 197, no. 1, pp. 165–170, 2011.

[20] ATT Laboratories Cambridge, "The ORL database of faces," *http://www.cl.cam.ac.uk/research/dtg/*.

[21] Image Engineering Laboratory, "The Sheffield (UMIST) face database," *http://www.sheffield.ac.uk/eee/research/iel/research/face*.

[22] L. Spacek, "Collection of facial images: Faces94," *http://cswww.essex.ac.uk/mv/allfaces/faces94.html*.

[23] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proceedings of IEEE International Conference Computer Vision*, 2011, pp. 543–550.

[24] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," *Neural Information Processing Systems*, vol. 14, no. 1, pp. 234–778, 2009.