Novel Matching Methods for Automatic Face Recognition Using SIFT

Ladislav Lenc and Pavel Král

Department of Computer Science and Engineering,
University of West Bohemia, Plzeň, Czech Republic
{llenc,pkral}@kiv.zcu.cz

Abstract. The object of interest of this paper is Automatic Face Recognition (AFR). The usual methods need a labeled corpus and the number of training examples plays a crucial role for the recognition accuracy. Unfortunately, the corpus creation is very expensive and time consuming task. Therefore, the motivation of this work is to propose and implement new AFR approaches that could solve this issue and perform well also with few training examples. Our approaches extend the successful method based on the Scale Invariant Feature Transform (SIFT) proposed by Aly. We propose and evaluate two methods: the Lenc-Kral matching and the SIFT based Kepenekci approach [7]. Our approaches are evaluated on two face data-sets: the ORL database and the Czech News Agency (ČTK) corpus. We experimentally show that the proposed approaches significantly outperform the baseline Aly method on both corpora.

Keywords: Automatic Face Recognition, Czech News Agency, Scale Invariant Feature Transform.

1 Introduction

Automatic Face Recognition (AFR) consists of automatic identification of a person from a digital image or from a video frame by a computer. This field became intensively studied in the last two decades. Concerning other biometrics methods, AFR seems to be one of the most important ones.

The spectrum of applications utilizing AFR is really broad: access control to restricted areas, surveillance of persons, various programs for sharing and labeling of photographs, social networks and many others.

The most of the current AFR approaches perform well when high quality images available (well aligned, unified pose, etc). Unfortunately, the performance of such methods degrades significantly, when this assumption is violated. This issue is often handled by using more training examples/person (also our case). For creation of the correct face models a training corpus with enough training examples is necessary. Unfortunately, the corpus creation is very expensive and time consuming task. Our motivation is thus to propose and implement new AFR approaches that have high recognition accuracy also with few training examples

L. Iliadis et al. (Eds.): AIAI 2012, IFIP AICT 381, pp. 254–263, 2012.
© IFIP International Federation for Information Processing 2012

(close to one). The main contribution of this work consists in proposing new matching methods and their integration to the SIFT algorithm.

The outcomes of this work shall be used by the Czech News Agency (ČTK) as follows. ČTK disposes a large database of photographs. A certain number of photos is manually annotated (i.e. the photo identity is known). However, another photos are unlabeled; the identities are thus unknown. The main task of our application consists in the automatic labeling of the unlabeled photos. Note that only few labeled images of every person are available.

The rest of the paper is organized as follows. The following section presents a short review of automatic face recognition approaches. Section 3 describes the SIFT algorithm and the proposed matching approaches. Section 4 evaluates the approaches on two corpora. In the last section, we discuss the results and we propose some future research directions.

2 Related Work

Thanks to the intensive research in the past years, many successful AFR methods were developed. The first attempts were based upon simple measures between important facial features [3]. The main drawback of such methods is the need of manual face labeling. Later, several approaches reducing the facial vector dimensionality were developed. One of such methods is the successful Eigenfaces approach [13,14] which is based on the Principal Component Analysis (PCA). Another method belonging to this group are Fisherfaces [2]. This approach uses Linear Discriminant Analysis (LDA) and Independent Component Analysis [12].

In the last ten years, a lot of attention was given to the feature based methods. The core of such methods is creating of a feature representation of the face. The facial image is inspected and the points of interest are detected. Then the features are created in the detected points. Some of those methods utilize Gabor wavelets to extract the features (e.g. Elastic Bunch Graph Matching (EBGM) [15] and the Kepenekci method [4]).

Recently, the Scale Invariant Feature Transform (SIFT) [9] proposed by David Lowe has been also used to create the facial features leading to high recognition accuracy. It has the ability to detect and describe local features in images. The features are invariant to image scaling, translation and rotation. The algorithm is also partly invariant to changes in illumination. The SIFT algorithm was originally developed for object recognition. The features of the reference and test images are compared using the Euclidean distance of their feature vectors.

2.1 SIFT for Face Recognition

One of the first applications of this algorithm for the AFR is proposed in [1] by Aly. It takes the original SIFT algorithm and creates the set of descriptors as described in Section 3. Each image is represented by the set of descriptors corresponding to the features.

First, the feature vectors are extracted from all gallery images. The test face is then matched against the faces stored in the gallery. The face, that has the largest number of matching features is identified as the closest one. The feature is considered to be matched if the difference between similarities of two most similar gallery features is higher than a specified threshold. In this work, the ORL and Yale databases are used for testing. It is reported that the recognition rate is 96.3% and 91.7% respectively. The results are compared with Eigenfaces [13,14] and Fisherfaces [2].

In [5], another approach using SIFT is presented. This method is called Fixed-key-point-SIFT (FSIFT). Contrary to the previous method, the SIFT keys are fixed in predefined locations determined in the training step as follows.

In the training step, the key-point candidates are localized in the same manner as in the original SIFT. A clustering algorithm is then applied to this key-point candidate set. The number of clusters is set to 100. The centroids of the clusters are used as the fixed key-point locations. The number of features thus remains constant. The distance between faces can be computed as the sum of Euclidean distances between the corresponding features. The reported recognition rate for the Extended Yale Database is comparable to the previously described approaches.

The proposed approaches, which use the features resulting from the SIFT algorithm, are described in more detail in the sequel.

3 Method Description

The first step is the determination of extrema in the image filtered by the Difference of Gaussian (DoG) filter. The input image is gradually down-sampled and the filtering is performed in several scales. Figure 1 demonstrates the process of creation of the DoG filters at the different scales [10].

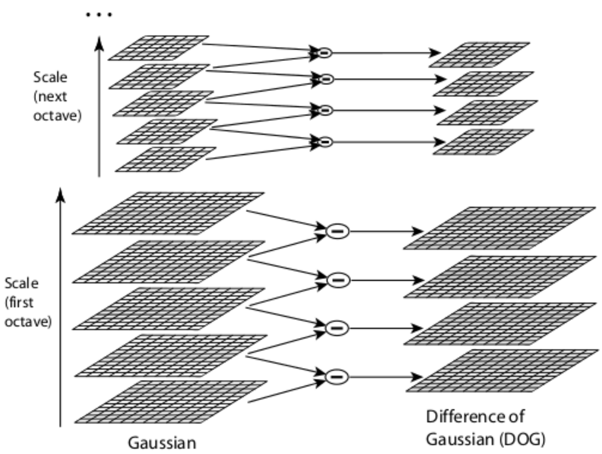

Fig. 1. Difference of Gaussian filters at the different scales [10]

In the next step, the detected key-points are further examined to choose the "best" candidates. Only points with high enough contrast are used and also points near edges are discarded. Then, orientation is assigned to each key-point. The resulting set of points is then used for creation of feature vectors (descriptors). Each descriptor contains a vector of the length 128 and also the coordinates of the point.

Figure 2 shows how two images of the same object (a face) with varying scale and orientation are matched.

Fig. 2. Matched key-points in two different views of the same object (face)

3.1 SIFT Features Extraction

The SIFT algorithm has basically four steps: extrema detection, removal of key-points with low contrast, orientation assignment and descriptor calculation [5].

To determine the key-point locations, an image pyramid with re-sampling between each level is created. It ensures the scale invariance. Each pixel is compared with its neighbours. Neighbours in its level as well as in the two neighbouring (lower and higher) levels are examined. If the pixel is maximum or minimum of all the neighbouring pixels, it is considered to be a potential key-point.

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

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

The final step is the creation of the descriptors. 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.

Figure 3 shows the SIFT features detected in the example images from the ČTK face corpus.

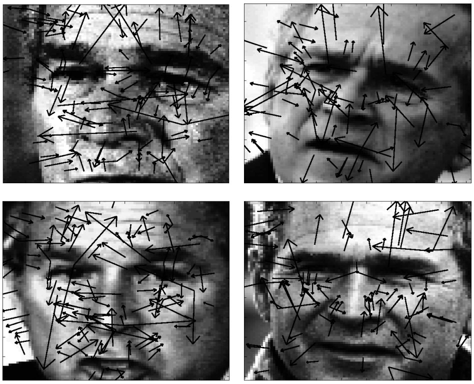

Fig. 3. Examples of detected SIFT features with orientation

The SIFT algorithm is described in detail in [9,10,5]. An implementation example can be found in [11].

3.2 Aly Matching

The first approach computes the number of the gallery image feature vectors that are matched against the test face feature vectors. For each test feature vector the similarities to all of the gallery feature vectors are computed. The cosine similarity of two feature vectors f_1 and f_2 is computed as follows:

$$S(f_1, f_2) = \frac{f_1 \cdot f_2}{\|f_1\| \|f_2\|} \tag{1}$$

The two most similar gallery feature vectors are determined. If the difference between these two similarities is higher than a prespecified threshold the feature vector is considered to be matched. For each gallery face, the number of matched feature vectors is computed. The recognized face is the one with highest number of matched feature vectors.

3.3 Lenc-Kral Matching

The first proposed approach computes a sum of similarities between pairs of image feature vectors. For each feature vector of the test face the most similar feature vector of the gallery face is identified. The sum of the highest similarities is computed and is used as a measure of similarity between two faces.

Speaking in more mathematical terms, let T be a test image represented by m feature vectors $t_1, t_2, .., t_m$. Let G be a gallery of images composed of N images $G_1, G_2, .., G_N$. Let every gallery image G_i be represented by n_i feature vectors $g_1, g_2, .., g_{n_i}$. Similarity of two feature vectors $S(t, g)$ is computed by the cosine similarity (see Equation 1). For each feature vector t_i of the recognized face T we determine the most similar vector $g^j_{max_i}$ of one gallery image G_j:

$$g^j_{max_i} = \arg\max_{G_j}(S(t, g)) \tag{2}$$

The sum of those similarities is computed as follows:

$$D(T, G_j) = \sum_{i=1..m} g_{max_i} \tag{3}$$

where m is the number of test image feature vectors. The recognized face is then determined by the following equation:

$$\hat{G}_i = \arg\max_{G_i}(D(T, G_j)) \tag{4}$$

3.4 Kepenekci Matching

This approach has been initially used by Kepenekci in [4] with Gabor wavelets. Author shows that this approach exhibits high recognition accuracy. Therefore, we decided to adapt this approach and integrate it with the SIFT.

Kepenekci combines two methods of matching and uses a weighted sum of the two values as a result. The cosine similarity is employed for vector comparison.

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

$$\sqrt{(x_t - x_g)^2 + (y_t - y_g)^2} < distanceThreshold \tag{5}$$

where x and y are 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 an average of similarities between each pair of corresponding vectors as:

$$OS_{T,G} = mean\left\{S(t, g), t \in T, g \in G\right\} \tag{6}$$

Then, the face with the most similar vector to each of the test face vectors is determined. The C_i value informs how many times the 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. Weighted sum of these two similarities is used for similarity measure:

$$FS_{T,G} = \alpha OS_{T,G} + \beta \frac{C_G}{N_G} \tag{7}$$

The face is recognized as follows:

$$F\hat{S}_{T,G} = \arg\max_{G}(FS_{T,G}) \tag{8}$$

4 Experimental Setup

4.1 Corpora

ORL Database. The ORL database was created at the AT & T Laboratories[1]. The pictures of 40 individuals were taken between April 1992 and April 1994. For each person 10 pictures are available. Every picture contains just one face. They may vary due to three following factors: 1) time of acquisition; 2) head size and pose; 3) lighting conditions. The images have black homogeneous background. The size of pictures is 92×112 pixels. A more detailed description of this database can be found in [8].

Czech News Agency (ČTK) Database. This corpus is composed of the images of individuals in uncontrolled environment that were randomly selected from the large ČTK database. All images were taken during a long time period (20 years or more). The detection of faces was made automatically utilizing the OpenCV library. They were automatically resized to the size 92×92 pixel and transformed to grayscale. The resulting corpus contains images of 63 individuals, 8 images for each person. Note that orientation, lighting conditions and background of images differ significantly. Performing accurate face recognition using this dataset is thus very difficult.

Figure 4 shows one example from this corpus. This corpus is available for free for the research purpose upon request to the authors.

Fig. 4. Examples of one face from the ČTK face corpus

4.2 Experiments

All experiments were performed on two datasets: the ORL dataset and the ČTK corpus mentioned before. Previously, we tested the ČTK database using the Eigenfaces approach. It exhibited very low accuracy. Therefore, we use the SIFT based methods for experiments. We used the successful Aly method (see Section 3.2) as a baseline. In all cases the cross-validation was used to ensure more reliable results.

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

We made a series of experiments for each dataset. The size of the training set is gradually increased from 1 image/person to $N-1$ images/person (N is the total number of images/person). We used 9 different set-ups for the ORL dataset and 7 set-ups for the ČTK dataset. To allow a straightforward comparison of these methods, we evaluated each set-up with three previously described matching schemes.

Table 1. Recognition rate of the different matching schemes for the ORL dataset according to the different training set size

Matching scheme	Aly	Lenc-Kral	Kepenekci
Training Set	**Recognition rate (%)**		
1 of 10	61.25	78.75	80.56
2 of 10	78.72	88.24	90.15
3 of 10	85.36	92.46	94.24
4 of 10	88.83	95.67	97.25
5 of 10	92.42	96.75	97.92
6 of 10	95.27	97.86	97.86
7 of 10	96.88	98.65	98.65
8 of 10	98.36	98.86	99.17
9 of 10	99.00	99.00	99.25

Table 1 shows the recognition rates of the different test set-ups for the ORL dataset. This table shows that the scores of the proposed Lenc-Kral approach are significantly higher than the original Aly method especially where not enough training examples available. The second proposed approach (SIFT based Kepenekci method) have slightly better recognition accuracy than both other approaches.

Table 2. Recognition rate of the different matching schemes for the ČTK corpus according to the different training set size

Matching scheme	Aly	Lenc-Kral	Kepenekci
Training Set	**Recognition rate (%)**		
1 of 8	9.78	12.95	19.73
2 of 8	14.18	19.11	27.78
3 of 8	16.90	24.29	31.75
4 of 8	20.40	28.89	37.10
5 of 8	22.93	31.92	41.18
6 of 8	24.12	34.27	43.85
7 of 8	25.79	36.71	46.63

Table 2 shows the recognition accuracy of the experiments on the ČTK corpus. The recognition accuracy is significantly lower than such in the case of the ORL database probably due to the different orientation of the images (see Figure 4).

This table also shows that both proposed methods significantly outperform the baseline Aly approach for all training examples in all cases.

5 Conclusions and Perspectives

In this paper, we presented two new AFR methods: namely Lenc-Kral matching and SIFT based Kepenekci approach. Both methods are based on the SIFT features. The experiments show that both proposed methods outperform the baseline Aly approach. The recognition accuracy on the ORL corpus is significantly higher particularly when the training set is small. In the case that only one training example per person is used, the Lenc-Kral and the Kepenekci matching increase the recognition rate respectively by 17% and by 19%, over the baseline. The results on the ČTK dataset show the difficulties of the face recognition in the real conditions. However, the recognition accuracy is in all cases significantly higher than in the Aly method.

The first perspective consists in combining this method with another successful method in order to further improve the recognition accuracy. Particularly, the adapted Kepenekci approach [7] based on the Gabor wavelets could be a suitable choice due to its high recognition accuracy. Another perspective consists in the use of confidence measures in the post-processing step [6]. The confidence measure technique will be used to detect and remove incorrectly recognized examples from the result set.

Acknowledgements. This work has been partly supported by ČTK and by the UWB grant SGS-2010-028 Advanced Computer and Information Systems. We also would like to thank ČTK for providing the photographic data.

References

1. Aly, M.: Face recognition using sift features (2006)
2. Belhumeur, P.N., Hespanha, J.A.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (1997)
3. Bledsoe, W.W.: Man-machine facial recognition. Tech. rep., Panoramic Research Inc., Palo Alto, CA (1966)
4. Kepenekci, B.: Face Recognition Using Gabor Wavelet Transform. Ph.D. thesis, The Middle East Technical University (2001)
5. Krizaj, J., Struc, V., Pavesic, N.: Adaptation of sift features for robust face recognition (2010)
6. 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)
7. Lenc, L., Král, P.: Gabor wavelets for automatic face recognition. In: 38th International Conference on Current Trends in Theory and Practice of Computer Science, Špindlerův Mlýn, Czech Republic (January 2012)
8. Li, S., Jain, A.: Handbook of face recognition. Springer (2005)

9. Lowe, D.G.: Object recognition from local scale-invariant features. In: International Conference on Computer Vision (1999)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 2 (2004)
11. Lowe, D.: Software for sift (2004), http://people.cs.ubc.ca/~lowe/keypoints/
12. Shakhnarovich, G., Moghaddam, B.: Face recognition in subspaces. In: Handbook of Face Recognition, pp. 141–168. Springer (2004)
13. Sirovich, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America 4 (1987)
14. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on in Computer Vision and Pattern Recognition (1991)
15. Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. Intelligent Biometric Techniques in Fingerprint and Face Recognition (1999)