MIPRO 2010, May 24-28, 2010, Opatija, Croatia

Adaptation of SIFT Features for Face Recognition under Varying Illumination

J Križaj, V. Štruc, N Pavešić
Faculty of Electrical Engineering, University of Ljubljana
Complete Address: Tržaška 25, SI-1000 Ljubljana, Slovenia
Phone: (+386) 40-640 864 E-mail: janez.krizaj@fe.uni-lj.si

Abstract - *Scale Invariant Feature Transform* (or SIFT) is an algorithm used to detect and describe local features in images invariant to image scale, translation and rotation. All SIFT-based face recognition techniques found in literature so far rely heavily on the keypoint detector. The purpose of this detector is to locate interest points in the given image that are later used to compute the SIFT descriptors. While these descriptors are known to be among others (partially) invariant to illumination changes, the keypoint detector is not. To overcome the presented shortcoming of SIFT-based methods, a novel face recognition technique is proposed in this paper. The SIFT descriptors are computed at fixed points in the locations of the nodes on a regular grid, overlapping face image. By doing so, the need for keypoint detection on the test images is eliminated and greater robustness to illumination variations is achieved in comparison with related approaches from the literature. Experimental results, obtained on the Extended Yale face database B, demonstrate that better results are achieved with proposed technique in comparison with the remaining techniques assessed in our experiments, especially under severe illumination conditions.

I. INTRODUCTION

Scale Invariant Feature Transform (SIFT) proposed by Lowe in [1] represents one of the recent additions to the group of feature-based face recognition techniques. The SIFT features have many properties that make them suitable for matching different images of an object or a scene, such as invariance to image scaling and rotation, (partial) occlusion and to a certain extent also to changes in illumination and 3D camera view point. The SIFT technique works by first detecting a number of interest points (called keypoints) in the given image and then computing local image descriptors at the locations of these keypoints. When performing classification, each keypoint descriptor from the given image is matched independently against all descriptors extracted from the training images, and based on the matching outcome, the image is assigned to one of the classes from train set.

Even though the SIFT technique represent one of the state-of-the-art approaches to object detection/recognition and was successfully applied in general object detection and recognition tasks, it has some deficiencies when applied to the problem of face recognition. In comparison with general objects, faces have fewer structures with high contrast or high-edge response. Since keypoints corresponding to interest points representing distinctive facial features can be removed by the process of unreliable keypoints removal (discussed in further detail in Section II), where low-contrast keypoints and keypoints along edges are removed, it is important to adjust adequate thresholds in order to accept more keypoints. Another thing to be concerned is false matched keypoints. The majority of SIFT-based approaches employed for face recognition use different partitioning schemes to determine number of subregions of the facial image and then compare the SIFT descriptors only between corresponding subregions. Due to local matching, wrong matches between spatially inconsistent SIFT descriptors are partially eliminated. However, variable illumination still has significant influence on the detection of keypoints, since the keypoint detector intrinsic to the SIFT technique is not invariant to illumination.

To overcome the presented shortcomings of the original SIFT technique (when applied for face recognition), in this paper a novel SIFT-based approach for face recognition is proposed, where the SIFT descriptors are computed at fixed predefined image locations. With this procedure, the need for (keypoint-removal) threshold optimization and image partitioning is eliminated, while suggested approach gains greater robustness to the illumination variance than comparable SIFT adaptations found in the literature.

The proposed method, called Grid-SIFT (GSIFT), was compared to several other approaches, such as Principal Component Analysis – PCA [2], Linear Discriminant Analysis – LDA [3] and several other related SIFT-based approaches. Experiments obtained on the Extended Yale B face database show, that, under severe illumination conditions, better results can be achieved with the proposed approach than with other face recognition methods mentioned above.

The rest of the paper is structured as follows: In Section II brief reviews of the original SIFT technique is presented. In Section III existing SIFT-based face recognition techniques are described and their major shortcomings are highlighted. In Section IV our novel approach is introduced, while its feasibility for robust face recognition in difficult illumination conditions is assessed in Section V. The paper concludes with some final comments in Section VI.

II. THE SCALE-INVARIANT FEATURE TRANSFORM

In this section basics of the SIFT algorithm are summarized. As described in **[1]** SIFT consists of four computational stages: (*A*) scale-space extrema detection, (*B*) removal of unreliable keypoints, (*C*) orientation assignment, (*D*) keypoint descriptor calculation and (*E*) descriptor matching:

A. Scale-space extrema detection. In first stage, interest points called keypoints, are identified in the scale-space by looking for image locations that represent maxima or minima of the difference-of-Gaussian function. The scale space of an image is defined as a function $L(x,y,\sigma)$, that is produced from the convolution of a variable-scale Gaussian $G(x,y,\sigma)$, with the input image $I(x, y)$:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (1)$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

where σ denotes the standard deviation of the Gaussian function $G(x,y,\sigma)$. The difference-of-Gaussian function $D(x,y,\sigma)$ can be computed from the difference of two Gaussians nearby scales separated by a multiplicative factor k:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (3)$$

Local maxima and minima of $D(x,y,\sigma)$ are computed based on the comparison of the sample point and its eight neighbors in the current image as well as the nine neighbors in the scale above and below. If the pixel represents a local maximum or minimum, it is selected as a candidate keypoint.

B. Removal of unreliable points. The final keypoints are selected based on measures of their stability. During this stage low contrast points (sensitive to noise) and poorly localized points along edges (unstable) are discarded. Two criteria are used for the detection of unreliable keypoints. The first criterion evaluates the value of $|D(x,y,\sigma)|$ at each candidate keypoint. If the value is below some threshold, which means that the structure has low contrast, the keypoint is removed. The second criterion evaluates the ratio of principal curvatures of each candidate keypoint to search for poorly defined peaks in the Difference-of-Gaussian function. For keypoints with high edge responses, the principal curvature across the edge will be much larger than the principal curvature along it. Hence, to remove unstable edge keypoints based on the second criterion, the ratio of principal curvatures of each candidate keypoint is checked. If the ratio is below some threshold, the keypoint is kept, otherwise it is removed.

C. Orientation assignment. An orientation is assigned to each keypoint by building a histogram of gradient orientations $\theta(x, y)$ weighted by the gradient magnitudes $m(x,y)$ from the keypoint's neighborhood:

$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2} \quad (4)$$

$$\theta(x,y) = \tanh\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \quad (5)$$

where L is a Gaussian smoothed image with a closest scale to that of a keypoint. By assigning a consistent orientation to each keypoint, the keypoint descriptor can be represented relative to this orientation and, therefore, invariance to image rotation is achieved.

D. Keypoint descriptor calculation. The keypoint descriptor is created by first computing the gradient magnitude and orientation at each image point of the 16×16 keypoint neighborhood. This neighborhood is weighted by a Gaussian window and then accumulated into orientation histograms summarizing the contents over subregions of the neighborhood of size 4×4, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. Each histogram contains 8 bins, therefore each keypoint descriptor features 4×4×8 = 128 elements. The coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation to achieve orientation invariance and the descriptor is normalized to enhance invariance to changes in illumination.

E. Matching. When using the SIFT algorithm for object recognition, each keypoint descriptor extracted from the query (or test) image is matched independently to the database of descriptors extracted from all training images. The best match for each descriptor is found by identifying its nearest neighbor (closest descriptor) in the database of keypoint descriptors from the training images. Generally, many features from a test image do not have any correct match in the training database, because they were either not detected in the training image or they arose from background clutter. To discard keypoints whose descriptors do not have a good enough match in the training database, a subsequent threshold is used, based on which matches that are too ambiguous are rejected. If the distance ratio between the closest neighbor and the second-closest neighbor, (i.e., the closest neighbor that is known to come from a different object than the first) is below some threshold, than the match is kept, otherwise the match is rejected and the keypoint is removed. The object in the database with the largest number of matching points is considered the matched object, and is used for the classification of the object in the test image.

III. SIFT-BASED FACE RECOGNITION – LITERATURE REVIEW

One of the first attempts to use the SIFT algorithm for face recognition is proposed in [4]. The algorithm used here, differs from original SIFT algorithm in the implementation of the matching stage. Each SIFT descriptor in the test image is matched with descriptors of each training image. Matching is done using a distance based criterion. A descriptor from the test image is considered match with another descriptor from the training image, if the distance between the 2 descriptors is less than a specific fraction of the distance to the next nearest descriptor. The problem with this method is that it is very time consuming. Matching between two images has a computational complexity $O(n^2)$, where n is the average number of SIFT descriptors in each image.

In [5] two strategies for locally constrained SIFT descriptor matching are proposed. The first method matches only SIFT descriptors extracted from image windows corresponding to the mouth and the two eyes, while the second relies on grid-based matching. Local matching within a grid or a cluster, limits the SIFT features to match features from nearby areas only and also reduces the computational complexity to $O(n^2/s)$, where s is the number of grids or clusters. Fig. 1 illustrates that some keypoints are considered matched, even though they do not represent the same characteristic of the face. We would expect the distance between such keypoints to be high, since they correspond to different regions of the face, this is clearly not the case. Therefore better results are achieved, if certain subsets of SIFT keypoints are used for matching and only (spatially) corresponding subsets of SIFT descriptors are matched as in [5] and later in [6], [7], [8] and [9].

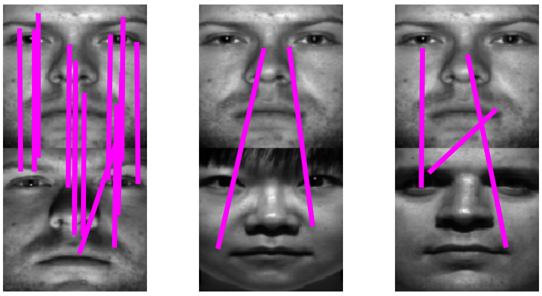

Fig. 1. Match results for one of the test images (top pictures) with a set of training faces (bottom) using the basic SIFT algorithm. The greatest number of matches are found, when two face images of the same person are compared. Some of the keypoints are matched even they do not correspond to the same person and/or to the same region of the face.

In [10] SIFT features are extracted from the frontal and half left and right profiles. An augmented set of SIFT features is then formed from the fusion of features from the frontal and side profiles of an individual, after removing feature redundancy. SIFT feature sets from the database and query images are matched using the Euclidean distance and Point pattern matching techniques.

In [11] Keypoints-Preserving-SIFT (KPSIFT) is proposed, which keeps all the initial keypoints for SIFT descriptor calculation. This procedure greatly differs from the basic SIFT approach, where unreliable keypoints are removed as explained in section 2. However, this removal can eliminate some keypoints and discard potentially useful discriminative information for face recognition. Intrinsic properties of the face images have to be considered, when setting the threshold values governing the process of keypoint removal. As shown in this approach, recognition rates improve when adjusting thresholds on low-contrast and edge keypoints in order to accept more keypoints.

IV. THE GRID-SIFT ALGORITHM

The techniques presented in previous section try to compensate the imperfections of the keypoint detector by means of imposing local matching constraints, by relaying on sub-windows of the images, by using clustering techniques or by deploying graph-matching techniques. In the remainder a simple procedure is presented, which eliminates the need for the keypoint detector and its shortcomings such as sensitivity to illumination and also prevents matching keypoints belonging to different regions of face. We based our method on presumption that each face was preliminary localized, so that each image consists only of a properly registered face region of a certain person.

A. Fixing the Keypoints

To achieve greater robustness on illumination variances, we avoid using the keypoint detector. Since faces are preliminary localized, for each face image we determine a regular grid overlapping face region. Locations of grid nodes serve as *fixed* keypoints for computation of SIFT descriptors. In these regions k = 100 descriptors were computed for each face image. Additionally LDA method was applied to descriptor data to maximize ratio between inter-class and intra-class variations. The advantages gained by this approach are illustrated in Fig. 2. The two images on the top depicts keypoint locations gained by original keypoint detector under different illumination conditions. The number of detected keypoints in the top right image is smaller than in the left image, moreover many of the keypoints are detected in different locations than in the left image, consequently reduction of keypoint matches is expected. In case SIFT descriptors are computed at predefined fixed locations, as presented in bottom images of Fig. 2, greater robustness to variable illumination can be achieved.

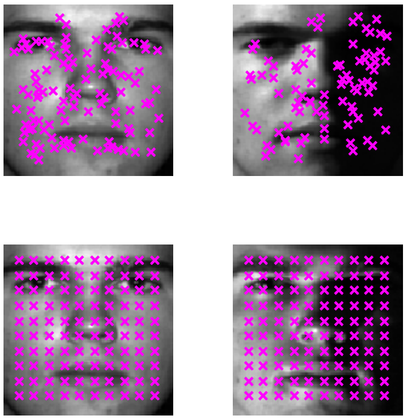

Fig. 2. Keypoints detected under different illumination conditions by original SIFT detector (top), and by the proposed method (bottom).

B. Descriptor Matching

Since all images have the same number of descriptors, the sum of the Euclidean distances between equally located descriptors of the two images to be compared is used as the matching criterion. Let us denote the sets of SIFT descriptors from the training images as $S_j = \{S_{i,j}(x_i,y_i);\ i = 1,2, ..., k\}$, where $j = 1, 2, ..., n$ denotes the training image index, n stands for the total number of training images, i represents the descriptor index, k denotes the number of fixed keypoint locations, and (x_i, y_i) is the location of the i-th SIFT descriptor. Let us further assume that the n training images correspond to N different classes (i.e. persons) with corresponding class labels $\omega_1, \omega_2, ..., \omega_N$. The matching procedure can then be written as

$$\delta_{SL_2}\left(S_g, S_t\right) = \min_j \delta_{SL_2}\left(S_j, S_t\right) \rightarrow S_t \in \omega_g \qquad (6)$$

where S_t stands for the set of SIFT descriptor extracted from the test image at the k predefined image locations, and the matching function is defined as

$$\delta_{SL_2}\left(S_p, S_r\right) = \sum_i \delta_{L_2}\left(S_{i,p}, S_{i,r}\right) \tag{7}$$

By the expression above, a given test image is assigned to the class ω_g, if the sum of the Euclidean distances between spatially corresponding descriptors of the test image and one of the training images from the g-th class is the smallest among the computed distances to all descriptor sets of n training images.

V. EXPERIMENTS AND RESULTS

The experiments were performed on the Extended Yale B (EYB) face database [12], comprised of 38 subjects, with approximately 64 frontal-view images of each subject, taken under different illuminations conditions. For the experiments, the images were sorted into five subsets (the numbers in the brackets next to the subset label in Table I represent the number of images in each subset). First subset (S1 in the remainder) contains images captured in relatively good illumination conditions, while for the image subsets labeled S2 to S5, the lighting conditions are going from good to worse. S1 is used as the training set, while images in the other subsets are used as test images.

All algorithms were implemented in Matlab relying partially on existing code available from [13] and [14]. The performance of the proposed approach was compared with some other face recognition techniques such as PCA, LDA, original SIFT algorithm and method from [6] called SIFT_CLUSTER, which relies on clustering of the SIFT keypoints. The performance of the listed methods is presented in Table I in the form of rank one recognition rates for each subset. As can be seen from the last row of Table I, with our method, denoted as GSIFT, better results are achieved in comparison with the recognition performance of the remaining techniques assessed in our experiments.

TABLE I
RANK ONE RECOGNITION RATES (IN %) OBTAINED ON THE EYB DATABASE IN THE COMPARATIVE ASSESSMENT.

Method	S2(456)	S3(455)	S4(488)	S5(752)
PCA	93.6	55.0	16.7	22.0
LDA	100	99.8	56.3	51.0
SIFT	100	45.7	25.7	11.2
SIFT_CLUSTER	100	100	66.8	64.9
GSIFT	**100**	**100**	**82.6**	**83.1**

VI. CONCLUSION AND FUTURE WORK

In this paper, a new approach called GSIFT is proposed, to solve the problems of SIFT technique when applied to face recognition. The results obtained by testing method on EYB database show that the performance of the proposed method is significantly better in the presence of severe illumination changes, when compared to common techniques, such as PCA or LDA and different SIFT-based recognition techniques from the literature. With the intention to be able to deal with possible variations in pose, we plan to enhance the GSIFT technique with a pose detector.

REFERENCES

[1]. °Lowe, D. G. *"Distinctive image features from scale-invariant keypoints".* International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.

[2]. M. °Turk, A. S. °Pentland. *"Face recognition using eigenfaces".* Proc. IEEE Conference on Computer Vision and Pattern Recognition. pp. 586-591, 1991.

[3]. K. °Etemad, R. °Chellappa. *"Discriminant Analysis for Recognition of Human Face Images".* Journal of the Optical Society of America, Vol. 14, No. 8, pp. 1724-1733, 1997.

[4]. °Aly, M. *"Face Recognition using SIFT Features".* http://www.vision.caltech.edu/malaa/research.php.

[5]. M. °Bicego, A. °Lagorio, E. °Grosso, M. °Tistarelli. *"On the use of SIFT features for face authentication".* Conference on Computer Vision and Pattern RecognitionWorkshop, pp. 35-35, 2006.

[6]. L. °Jun, Y. °Ma, E. °Takikawa, S. °Lao, M. °Kawade, B. L. °Lu. *"Person-Specific SIFT Features for Face Recognition".* IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 593-596, 2007.

[7]. D.R. °Kisku, A. °Rattani, E. °Grosso, M. °Tistarelli. *"Face Identification by SIFT-based Complete Graph Topology".* IEEE Workshop on Automatic Identification Advanced Technologies, pp. 63-68, 2007.

[8]. A. °Majumdar, R.K. °Ward. *"Discriminative SIFT features for face recognition".* Canadian Conference on Electrical and Computer Engineering, pp. 27-30, 2009.

[9]. D. R. °Kisku, M. °Tistarelli, J. K. °Sing, P. °Gupta. *"Face recognition by fusion of local and global matching scores using DS theory: An evaluation with uni-classifier and multi-classifier paradigm".* IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 60-65, 2009.

[10]. A. °Rattani, D. R. °Kisku, A. °Lagorio, M. °Tistarelli. *"Facial Template Synthesis based on SIFT Features".* IEEE Workshop on Automatic Identification Advanced Technologies, pp. 69-73, 2007.

[11]. Cong Geng, Xudong Jiang. *"SIFT features for face recognition".* 2nd IEEE International Conference on Computer Science and Information Technology, pp.598-602, 2009.

[12]. *The extended Yale Face Database B.* http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.

[13]. D.G. °Lowe, Software for SIFT. *http://people.cs.ubc.ca/~lowe/keypoints/.*

[14]. Vedaldi A., MATLAB/C implementation of the SIFT detector and descriptor. *http://www.vlfeat.org/~vedaldi/code/sift.html.*

[15]. L. °Wiskott, J. M. °Fellous, N. °Kruger, C. °Malsburg. *"Face recognition by elastic bunch graph matching".* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 775–779, 1997.

[16]. A. °Nefian, M. °Hayes. *An embedded HMM-based approach for face detection and recognition.* Proc. IEEE CASSP, pp. 3553-3556, 1999.