Pose Unconstrained Face Recognition based on SIFT and Alignment Error

Yongbin Gao
Department of Computer Science and Engineering
Chonbuk National University
Jeonju, Korea
gaoyongbin.sam@gmail.com

Hyo Jong Lee
Department of Computer Science and Engineering,
Center for Advanced Image and Information Technology
Chonbuk National University
Jeonju, Korea
hlee@chonbuk.ac.kr

***Abstract*—Pose variation is one of the key challenges for practical face recognition problem. Face recognition under well-controlled settings, like frontal face and good illumination, achieved high performance. But it fails when they are directly adopted to face recognition with large pose change. In this paper, we propose a novel framework using the combination of SIFT and alignment error (SIFT-AE) to perform pose invariant face recognition. SIFT (Scale Invariant Feature Transformation) is an effective local descriptor for face recognition under small pose change, which is scale and rotation invariant as well. However, the performance declined in case of large pose variance. To compensate this declination, Lucas-Kanade method is used to align the probe image and the gallery image, and the alignment error is deducted from the number of matching for SIFT algorithm. This alignment error provides additional information even in case of large pose change, while it is not distinctive alone. Therefore, the combination of SIFT and alignment error gains the performance for face recognition with large pose variance. Experiment results show our algorithm achieves impressive improvement compared with either SIFT or online alignment.**

Keywords—face recognition; SIFT; alignment error; Lucas-Kanade.

I. INTRODUCTION

Face recognition is used in a variety of applications. Such as online image search and tagging for personal photos, and human identification from surveillance for public security [1]. Face recognition can achieve high performance under controlled settings, such as limited pose like frontal image and good illumination. However, face recognition algorithms under these settings cannot be directly used for uncontrolled condition. Face recognition for uncontrolled pose and illumination still a challenging problem.

In this paper, we propose a robust face recognition algorithm that allows pose change. The main problem caused by viewpoint change is the larger distance for different viewpoints than different subjects. However, the distance between different subjects is the key feature used for face recognition. Therefore, the key issue for face recognition under viewpoint change turns to be the elimination of the distance between different viewpoints.

Normalization method is widely used to reduce the distance between different viewpoints. It can be performed by 2D or 3D way. For 2D model, Markov Random Fields (MRF) is used to find correspondences between two images [2, 3]. MRF seeks the 2D displacement of regions in two images by minimizing the energy, which is the summation of the distance of correspondences and smoothness of neighboring nodes. Another normalization method is proposed by Lucas and Kanade [4]. Lucas-Kanade method can be conducted as an online or offline way [5]. Online Lucas-Kanade performs the alignment for each probe image, and utilizes the alignment parameters to identify the subject. Offline Lucas-Kanade learns the Lucas-Kanade parameters from several sets of images, each set contains images from the same pose. The learned parameters are used for normalization of different viewpoints. Normalization can be performed in a way that either constructs the frontal face from the probe face or directly matches between these two faces. The former method utilizes the constructed face to identify subject, and the latter one uses matching scores instead. As for 3D model, Blanz et al. proposes a 3D morphable method to get face shape and texture coefficients by fitting 3D model to 2D face [6], and the similarity of these coefficients between two images is used for face recognition. It is reported that normalization method is an effective but time consuming face recognition algorithm, two minutes is reported to normalize one face [2]. Keypoint based face recognition is an alternative way [7]. Marsico et al. proposes a FACE framework that detects some keypoints using STASM algorithm [8], and construct half face using the middle line keypoints of the face, the left half is reflected from the constructed half face. FACE is an effective algorithm if the keypoints detection is of high accuracy. Otherwise, the performance declines significantly.

New classifiers or feature extraction methods are also proposed for pose invariant face recognition. For the new classifiers, Wolf et al. proposes one shot similarity (OSS) and two shot similarity (TSS) methods to calculate the similarity between two images by building two models with a third-party dataset without probe or gallery image in it [9]. These two models can be LDA or SVM and the average score of these models is considered as similarity. Cross-pose face recognition provides similar solution by introducing a third-party dataset [10]. Faces are linearly represented by the sub-dataset of the

same pose with it. Similarities of these linear coefficients from different images are used for face identification. New feature extraction method is also proposed to improve performance for pose changed face. Among which, tied factor analysis is an effective method to estimate the linear transformation and noise parameters in identity space [11].

Beside these algorithms, local descriptors are powerful in case of variance exists between gallery image and probe image due to its insensitive to scale, location or even affine transformation. Among these local descriptors, SIFT (Scale Invariant Feature Transformation) [12], Harris-Affine [13], Hessian-Affine [14] and Affine SIFT [15] are widely used for their invariance to scale or affine transformation. However, human face is not a planar, which contains significant 3D depth information. Affine transformation fails when directly using for human face.

In this paper, we propose to use SIFT and alignment error (SIFT-AE) for face recognition. SIFT is an effective local descriptor when pose change is below 15 degree. While the performance deteriorates for large pose change. In order to make up this deterioration, we align images from different pose using Lucas-Kanade method [5], and combine the alignment error with the number of matching for SIFT to represent the similarity of two images. This combination gains the reliability for large pose change.

The rest of this paper is organized as follows. Section II reviews the SIFT algorithm. We describe the proposed SIFT-AE algorithm in Section III. This algorithm includes image to image alignment and SIFT-AE framework. Section IV applies the above algorithm to FERET database, and presents the experiment results. Finally, we conclude this paper with future work in Section V.

II. SCALE INVARIANT FEATURE TRANSFORMATION

Scale Invariant Feature Transformation (SIFT) is an effective local descriptor, which is scale, rotation invariant. SIFT transforms image into scale invariant space and searches for the extrema as keypoints. After that, a descriptor is assigned to each keypoint using histogram of gradient. The main step of SIFT algorithm are as following [12]:

A. Scale space extrema detection

Image is transformed into different scales using Gaussian function. Extrema are localized by seeking maxima and minima over all scales after performing the Difference-of-Gaussian scheme, which is invariant to scale and orientation.

Difference-of-Gaussian is an approximation to the Laplacian of Gaussian. It is calculated by the difference of two neighboring scales as:

$$D(x,y,\sigma)=\left(G(x,y,k\sigma)-G(x,y,\sigma)\right)*I(x,y) \qquad (1)$$

where $\mathrm{D}(\mathrm{x},\mathrm{y},\sigma)$ is the Difference-of-Gaussian function with scale σ. G is the Gaussian function and k is the scale factor of nearby scales. I is the input image. Extrema are localized by comparing a pixel to its neighbor pixels at the same scale and adjacent two scales as shown in Fig. 1.

B. Keypoint Localization

Extrema are only keypoint candidates, it should be further refined by rejecting low contrast extrema and poor extrema that localized along an edge. This can be done by examining the nearby data for refined location and scale. Also, ratio of principal curvatures is used to determine whether the extrema localized along an edge. The ratio of principal curvatures is calculated indirectly by the trace and determinant of a 2*2 Hessian matrix.

$$\mathrm{H}=\begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad (2)$$

where D_{xx}, D_{yy}, D_{xy} are the derivatives estimated by differences of neighboring pixels.

C. Orientation Assignment

Orientation is assigned to each keypoint based on local image statistic feature. The orientation is the key feature to achieve rotation invariance for further descriptor, which is estimated by orientation histogram formed from the gradient orientations within a region around the keypoint. To gain the stability of the matching, multiple orientations are usually assigned to the same location and scale. Multiple orientations are generated by finding peaks within 80% of the highest peak orientation in the gradient orientation histograms.

D. Local Image Descriptor

Local image descriptor is assigned to each keypoint, which is distinctive and invariant to illumination or 3D viewpoint to some extent. Local keypoint descriptor is calculated around each keypoint by histogram of gradients after the coordinates of them are rotated to the keypoint orientation. The descriptor is transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

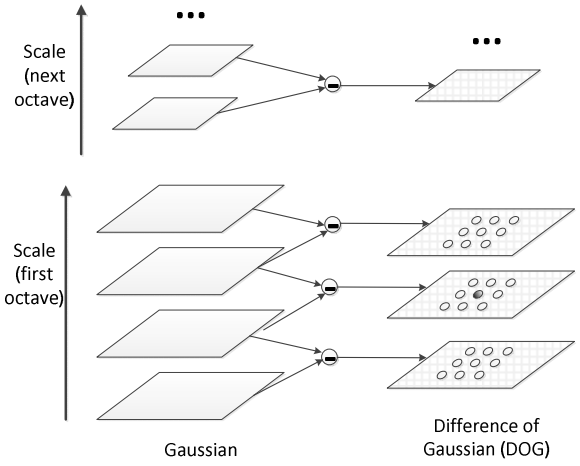

Fig. 1. Illustration of Difference of Gaussian (DOG), octaves from bottom to top are generated by down-sampling. The initial image is convolved with Gaussian filter using diffent scales for each octave. Difference of Gaussian images are generated from these Gaussian filtered image. Extrema are localized by finding the maxima and minima comparing with neighboring pixels in the current scale and adjacent scales as shown on right.

A keypoint descriptor is created based on the gradient and orientation in a region around the centre keypoint. The region is weighted by a Gaussian window. The region is divided into 4*4 subregions, and histogram of orientation with number of 8 bins is accumulated for each subregion. The length of each orientation in the histogram corresponds to the sum of the gradient magnitudes near that direction.

Image matching is performed after the location, scale, orientation and descriptor are generated for each keypoint. There are several methods reported for image matching and recognition of SIFT algorithm, such as BBF [16], Hough transform [17]. Nearest neighbour is the original and effective matching method for SIFT features. SIFT features are first pre-extracted from gallery images and stored in a database. When matching with a probe image, each SIFT feature from the probe image is compared with all gallery features in database. Nearest neighbour and second nearest neighbour are searched based on the Euclidean distance. The ratio of these two distances is compared with a threshold. Ratio that is smaller than the threshold is considered as a matching face.

The SIFT is an effective method under small viewpoint change since it is scale and rotation invariant, but it is not affine invariant. Affine SIFT is the extension of SIFT algorithm. Affine SIFT transforms an image into a series of simulated images by the change of longitude ϕ and latitude θ [15]. These simulated images are sampled to achieve a balance between accuracy and sparsity. However, Affine SIFT generates 61 images when the number of tilts set to 7. This increases the computation time too much, which is also unnecessary for face recognition. Moreover, human face contains 3D depth, while affine transformation is effective for planar object, simple affine transformation for a holistic face is not enough to represent the pose variant of face. In this paper, we propose to use SIFT-AE algorithm to combine the SIFT algorithm with alignment error from Lucas-Kanade method. The alignment error can complement the performance decline under large viewpoint change that SIFT cannot handle.

III. SIFT-AE

A. Image to Image Alignment

Image alignment is to find correspondences between two images, Lucas-Kanade algorithm is an effective image alignment method [18]. firstly we equally divide image into several subregions, for pixels in the same subregions, we assume they share the same warp parameters, Let the warp function be $x' = W(x, P)$, where $P = [p_1, p_2, \ldots p_m]^T$, For affine warp , m=6, and

$$W(x,P) = \begin{pmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{3}$$

Fig. 2 shows two images captured at two different poses, where I and T represent the probe image and the gallery image respectively. We divide image T into non-overlap subregions with same size. For each subregion r in T, we try to find a warp that aligns these two images. I_r is the corresponding subregion to T_r after warp transformation. The main objective for alignment is to minimize the error between the T_r and the warped subregions I_r as:

$$E_r = \sum_x \left(I_r \left(W(x,P) \right) - T_r(x) \right)^2 \tag{4}$$

The solution for Equation 4 is to iterate calculating a ΔP and update P till P converge. Lucas-kanade gives a solution for calculating ΔP by:

$$\Delta P = H_{img}^{-1} \sum_x \left(\nabla I_r \frac{\partial W}{\partial P} \right)^T \left(T_r(x) - I_r(W(x,p)) \right) \tag{5}$$

where $\nabla I_r = (\frac{\partial I_r}{\partial x}, \frac{\partial I_r}{\partial y})$ is the gradient of I_r . $\frac{\partial W}{\partial P}$ is the Jacobian of the warp (shown in Eq. 3). H_{img} is the pseudo Hessian matrix, which is given by:

$$H_{img} = \sum_x \left(\nabla I_r \frac{\partial W}{\partial P} \right)^T \left(\nabla I_r \frac{\partial W}{\partial P} \right) \tag{6}$$

We can now update the warp parameters $P \leftarrow P + \Delta P$ and iterate till the parameters P converge. This procedure is applied independently for every patch/subregion.

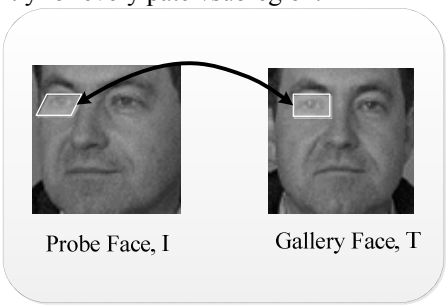

Fig. 1 Image to image alignment, image is divided into several subregions T_r , a warp between two subregions T_r and I_r is calculated by minimizing the alignment error.

B. SIFT-AE framework

Image to image alignment can be used online or offline. There are two kinds of online recognition methods. The first one is to calculate a match score for two images based on the warp parameters or alignment errors. Another one is to normalize images by transforming the profile face to its frontal face. Another scheme is off-line alignment. Warps parameters are trained from several sets of images, each set of images are from the same pose.

In our framework, online alignment is used to align gallery face and probe face, and the alignment error is combined with the matching number of SIFT algorithm to enable large viewpoint change. Let M_i be the number of matching between a probe face and the gallery face i for SIFT algorithm, and E_i is the alignment error after online alignment, which can be calculated as:

$$E = (\sum_{r} E_r) / N = \left(\sum_{r} \sum_{x \in r} \left(I_r \left(W(x, P_r) \right) - T_r(x) \right)^2 \right) / N \quad (7)$$

where P_r is the final alignment parameter for patch/subregion r, N is the total number of patches in a image.

Finally, the similarity between a probe face and the gallery face i is calculated as:

$$S_i = M_i + \lambda E_i \quad (\lambda < 0) \quad (8)$$

where λ is used to balance the importance between matching number and alignment error. $\lambda < 0$ means a big alignment error results in low similarity between two images. In our experiment, $\lambda = -1$ can achieve a good performance.

The reason we use the combination of SIFT and alignment error is the affine variance of SIFT as well as the significant 3D depth contained in human face. The performance of SIFT algorithm declines when large pose change exists. Meanwhile, alignment error provides distinctive information even for large pose change. While the alignment error alone is not discriminating enough for face recognition, the combination of both features can provide complement information for face recognition under large pose change.

IV. RESULTS

In our experiments, we used FERET [19] grey database to evaluate our algorithm. This database contains 200 subjects, each subject contains 9 images captured from different poses. For each subject, we use frontal image as gallery, and other 8 pose images as probe images, the pose angle of which are $-60^\circ, -40^\circ, -25^\circ, -15^\circ, 15^\circ, 25^\circ, 40^\circ$ and 60° degrees, respectively. Fig. 3 shows these face images in FERET database.

Fig. 3 Face images in FERET database with varying pose from 0^0, 60^0, 40^0, 25^0, 15°, -15°, -25°, -40°, -60°, respectively.

The proposed SIFT-AE algorithm is compared with SIFT [12], Alignment Error (AE) and Affine SIFT (ASIFT) [15]. The parameters used in our experiment for SIFT algorithm are: image is resized to a resolution of 800*800, and the ratio for nearest neighbour is set to 0.8. Image is divided into non-overlapping subregions for alignment. The size of which is 30*30 in our experiments. For Affine SIFT, number of tilt is set to 3 for the gallery image. This means ASIFT generates 10 viewpoints for a frontal image. Table I shows the comparison results of face recognition with ASIFT, SIFT and AE in FERET database. From the table, we know that SIFT can get similar results with SIFT-AE when a pose degree is between -15 to 15 degree, but SIFT-AE achieves better result than SIFT or AE under large pose different. Also, SIFT-AE performs better than ASIFT algorithm. Fig. 4 shows the face recognition rate in curve for the FERET database. From the table and figure, we can see AE is not distinctive alone, while combining it with SIFT, we can achieve 10% gain in recognition accuracy under pose degree larger than 40°.

TABLE I. EXPERIMENT RESULTS OF FACE RECOGNIZE IN FERET DATABASE (%)

Degree (°)	ASIFT [15]	SIFT [12]	AE	**SIFT-AE**
-40	56	48	18.5	61
-25	92.5	92	46.5	93.5
-15	99.5	100	83.5	99.5
15	99	99.5	86	99.5
25	93	96.5	44.5	97
40	58	53	13.5	64.5
Average	83	82	49	86

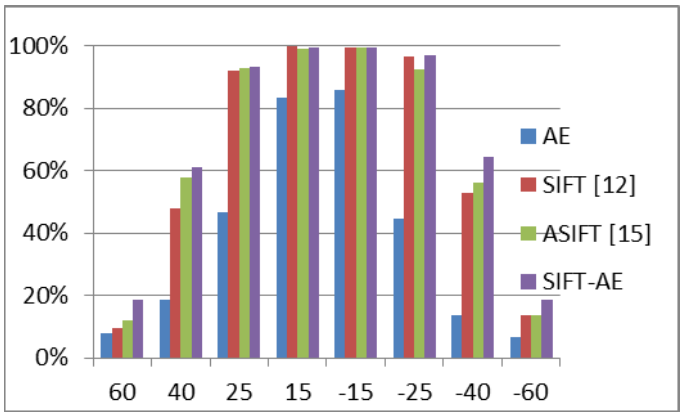

Fig. 4 Recognition rate of face with degrees varying from 0^0, 60^0, 40^0, 25^0, 15°, -15°, -25°, -40°, -60° in FERET database. Algorithms are SIFT, AE, ASIFT and proposed SIFT-AE.

V. CONCLUSION

In this paper, SIFT-AE is proposed for pose unconstrained face recognition. SIFT algorithm is scale and rotation invariant, which is powerful for small viewpoint changes in face recognition, but the performance of which declined when large viewpoint change exists. To complement this declination,

Lucas-Kanade is used to align the probe image and the gallery image, and the alignment error is deducted from the number of matching for SIFT algorithm. The combinations of SIFT and alignment error gains the performance in case of large pose variance. FERET database is used to test SIFT-AE, and experiment results show SIFT is an effective local descriptor when a pose degree is between -15 to 15 degree. Meanwhile, AE is not discriminating alone. However, as the combination of SIFT and AE, SIFT-AE achieves recognition rate above 10% better than SIFT or AE under large pose different. Also, our proposed SIFT-AE is better than Affine SIFT algorithm.

ACKNOWLEDGMENT

This work (Grants No. C0112553) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2013. This work was also supported by the Brain Korea 21 PLUS project, National Research Foundation of Korea.

REFERENCES

[1] G. Hua, M. H. Yang, E. L. Miller, Y. Ma, M. Turk, D.J. Kriegman and T. S. Huang, "Introduction to the Special Section on Real- World Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1921–1924, Oct. 2011.

[2] H. T. Ho, R.Chellappa, "Pose-Invariant Face Recognition Using Markov Random Fields," *IEEE Trans. Image Processing*, vol.22, no.4, pp.1573-1584, Apr. 2013.

[3] S. R. Arashloo and J. Kittler, "Energy Normalization for Pose-Invariant Face Recognition Based on MRF Model Image Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no.6, pp. 1274-1280, June. 2011.

[4] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision,* vol. 56, no.3, pp.221 – 255, Mar. 2004.

[5] A. B. Ashraf, S. Lucey and T. Chen, "Learning Patch Correspondences for Improved Viewpoints Invariant Face Recognition," *in Proc. IEEE conf. Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[6] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1-12, sep. 2003.

[7] M. D. Marsico, M. N. D. Riccio and H. Wechsler, "Robust Face Recognition for Uncontrolled Pose and Illumination Changes," *IEEE Trans. Systems, Man, and Cybernetics: Systems*, vol. 43, no. 1, pp. 149-162, Jan. 2013.

[8] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," *in Proc. Eur. Conf. Comput. Vis*., 2008, pp. 504–513.

[9] L. Wolf, T. hassner, and Y. Taigman, " Effective Unconstrained Face recognition by Combining Multiple Descriptors and Learned Background Statistics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1978-1990, Oct. 2011.

[10] A. Li, S. Shan, and W. Gao, "Coupled Bias–Variance Tradeoff for Cross-Pose Face Recognition," *IEEE Trans. Image Processing*, vol. 21, no. 1, pp. 305-315, Jan. 2012.

[11] S. J. D. Prince, J. H. Elder, j. Warrell, and Fatima, "Tied Factor Analysis for Face Recognition across Large Pose Differences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 970-982, June. 2008.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov. 2004.

[13] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *Int'l J. Computer Vision*, vol. 1, no. 60, pp. 3-86, 2004.

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L.V. Gool, "A Comparison of Affine Region Detectors," *Int'l J. Computer Vision*, vol. 65, no. 1-2, Nov. 2005.

[15] J. M. Morel and G. Yu, "ASIFT, A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp.438-469, 2009.

[16] J. Beis, and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in highdimensional spaces," *In Proc. Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 1000-1006.

[17] Hough, P.V.C. 1962. *Method and means for recognizing complex patterns.* U.S. Patent 3069654.

[18] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *In Proc. International Joint Conference on Artificial Intelligence,* 1981, vol. 2, pp. 674–679.

[19] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.