An improved Face Recognition Based on Scale Invariant Feature Transform (SIFT): Training for Integrating Multiple Images and Matching by Key Point’s Descriptor-Geometry

Sunghoon Kim, Youngjin Kim and Seunghyung Lee

Department of Computer Engineering, Kyung Hee University, Korea;
shoonkim@khu.ac.kr, jerryjin@khu.ac.kr, shlee7@khu.ac.kr

Abstract

Objectives: We proposed for reduction of the computational complexity and improvement of the recognition precision in the face recognition system using Scale Invariant Feature Transform (SIFT) local feature approaches.

Methods/Statistical Analysis: The first one is a novel training procedure for the integration of multiple training images. This training procedure performs to remove redundant local features and blend different local features from face image. The second one is a proposed matching scheme not only considering the similarity of key point’s descriptor but also the geometry property through one-to-one matching between query and references images.

Findings: This research finds the optimal settings of parameter for the proposed face recognition system based on SIFT. First, we have analyzed the change of recognition rate according to the resolution of face image in the proposed system. Then, to effect of the reduced number of key points per subject on the recognition rate and the resolution of face image were analyzed with the multiple templates per subject. As a result, we observed that the proposed template training procedure using Lowe’s key points detection method with 50×61 resolution of face images achieves higher recognition rate than the holistic approaches. The usage of Geng’s key point detection method in the proposed system obtains higher recognition rate than the usage of Lowe’s method. From the experimental result with ORL databases, the proposed face recognition system gives 99.5% of rate, which shows the higher performance than the previous ones. In addition, the proposed integration method of multiple training images reduces the number of key points by average 49.84% than the method of using multiple templates.

Improvements/Applications: The experimental result of the proposed system in two well-known face databases shows that the computational quantities are reduced effectively compared to other SIFT based methods, and it gives better performance on face recognition accuracy.

Keywords: Face Recognition, Image Matching, Scale Invariant Feature Transform (SIFT), Template Synthesis

1. Introduction

Person’s identification based on the visual cues of human face, is required to realize surveillance, authentication and content based retrieval in various situations such as security system, human-computer (or robot) communication and context-aware computing. Many solutions were suggested to solve the issues of facial recognition in the study of both pattern recognition and computer vision research. However, the facial recognition is a difficult problem because the feature of the face can be transformed easily under the influence of pose, facial expression and illumination. In addition, faces have the similar form, which includes the same local parts such as nose, lips and eyes etc. To enhance the ability of face recognition kit, it needs to apply efficient features. Which can give clear representations of similarity from subject’s diverse face images and distinctive classification properties among different subjects.

Generally, to extract significant features, the approaches based on linear transformation such as Principal Component Analysis (PCA)\(^1\)\(^2\) and Linear
Discriminant Analysis (LDA) are widely used in face recognition. These approaches firstly obtained the reduced dimensional features from face image with n×m pixels by using eigenfaces or fisher faces. Then the pattern classification techniques are applied to identify individuals. These approaches have been recently developed into various models that include Intrinsicfaces and feature fusion approach based on 2DPCA. In general, these approaches are requiring a pre-processing process to normalize the query face image into the same conditions to the reference images on training. So an acceptable performance is hard to be obtained due to the difficulty of detected the face range which are satisfy the conditions of training form the complex image. Moreover, the extracted features are highly susceptible to variations such as lighting, pose, and face expressions. Image matching system using the Scale Invariant Feature Transform (SIFT) descriptor was stated by Lowe for object recognition. In Lowes system, key points, which are unaffected to change of the scale, rotation and lighting condition, are firstly extracted independently from query and reference images. On next step, the matching algorithm performs to find the base match of each key point pair in query image and reference images. Finally, the similarities between the query and each reference image are evaluated by geometrical verification from pairs of matched key points. It has shown a good performance by using SIFT descriptor in many applications for object, scene recognition and object tracking, as well as the other machine vision applications. The SIFT descriptor is also used as feature representation for face recognition task in many prior works. The initial attempts of some studies try to solve the issue of local mismatching on pairs of matched key points while face is changeable and smooth compared to general objects and has a lot of similar regions. Facial recognition based on SIFT takes usage of an overlapping sub-image matching strategy. A personally exclusive feature matching strategy applies global similarity with local similarity from the grouping of key points by using K-means clustering scheme. In addition, the matching strategy using facial landmarks, which consists of a right, eye, a left eye and a region of mouth and nose was proposed. In these approaches, the performance of face recognition is deeply affected by the controlled condition, segmented, normalized or the detected result of landmarks. Another method proposed a 2-stage image matching technique based on Lowe’s matching algorithm with numerous amount of key points using a modified key point descriptor. This method finds a set of candidate reference images in a first matching, and then a second matching re-matches the query-image key points and the ones in each individual candidate image. Another issue is to grow the computational burden in recognition process because of a use of the multiple reference images per subject. The synthesis method, which focus on getting rid of redundant key points in the multiple training images. Although they synthesis method shows a decrease in the number of key points, it cannot increase the recognition rate because they do not consider the synthesis of key points in the variation of the face. The objective of this essay is improvement of identification capability and the reduction of computational quantities in the face recognition based on SIFT. We first propose an approach on the integration of the multiple training images, and it performs to remove redundant key points and blend different key points in the changing of the face pose and expression. In the proposed integration method, we use the mean shift clustering which is possible to group the key points into homogeneous and generate the integrated template. Secondly, we proposed a matching strategy for face identification based on Lowe's matching algorithm. This proposed matching strategy compares a query image with the integrated template one-on-one by using computation of key point's descriptor similarity and geometrical property. In the experimental results of face recognition tasks, the proposed method is found to be more accurate than other methods; also the computational quantities are reduced effectively.

2. Related Work

2.1 Scale Invariant Feature Transform

Lowe made proposal of the SIFT descriptor and matching strategy to implement object and scene recognition systems. SIFT descriptor has detection for points of local feature by detecting scale-space extreme, localizing key point, assigning orientation and key point description stages that can find stable points through the change of the scale, rotation and lighting. Firstly, action of detecting scale-space extrema is to identify potential points that are unaffected by scale and orientation by using a Difference-of-Gaussian (DoG) method. The scale space within image is defined as \( L(x, y, \sigma) \) which is produced from the
convolution of a variable-scale Gaussian $G(x, y, \sigma)$ with the input image $I(x, y)$

\[
(x, y) = G(x, y, \sigma) * I(x, y) \quad (1)
\]

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

where $\sigma$ means for the standard deviation of Gaussian function. DoG can be taken from the difference of two close scales separated through constant multiplicative factor $k$.

\[
D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3)
\]

\[
= L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)
\]

By using DoG images, a candidate key point is selected only if a sample point describes a local maximum or minimum by comparing with eight neighbors in the current scale in addition to the nine neighbors in scale above and below. And then, the key point localization is to select stable key points as candidates. Candidates which are sensitive to noise are removed by application of threshold on key point contrast. By using the ratio based on Hessian matrix, it eliminates candidates that have strong response along edge. Next, the orientation assignment is to assign a consistent orientation and a gradient magnitude to each key point. It calculates the gradient magnitude $m(x, y)$ from the samples within a region around the key point, and then the orientation is assigned by creating a histogram of gradient orientations $\theta(x, y)$ weighted by the gradient magnitudes

\[
m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (5)
\]

\[
\theta(x, y) = \tan^{-1}\left(\frac{L(x+1, y) - L(x-1, y)}{L(x,y+1) - L(x,y-1)}\right) \quad (6)
\]

As explanation, $L$ denotes Gaussian smoothed image. Finally, the key point descriptor is composed by using the gradient magnitude and orientations in the key point’s 16×16 neighborhood. The key point descriptor has 128 elements computed from orientation histograms for summary of the contents over 4×4 sub-regions with 8 orientation bins. In reference 14, a modified key point description that is able to capture the information of many smooth facial areas such as forehead, cheeks and chin is proposed. This description selects a candidate key point for comparison of a sample point with its eight neighbors in the its scale and the corresponding one neighbor in the scale above and below in DoG images. To prevent the removal of key points. Which are represented discernible areas such as mouth corners and wrinkles, it does not use the way to delete key point with low contrast and on the edges. Instead it eliminates insignificant clues and keeps distinctive ones by using the similarity related to others. Figure 1 shows the detected key points by the original method and the modified method respectively by using the size of 50×61, 60×74, 70×86 and 80×98

![Figure 1](image-url)

(a) Example of detected key points by the original method.
(b) Example of detected key points by the modified method.
images. We can see that the modified method in Figure 1 (b) obtains more key points than Low’s detection method on the smooth area of face. But the increasing amount of the detected key points according to the image size appears steeper than the Low’s approach. Therefore, it may cause the increasing computational complexity while applying to the recognition of various face sizes. In this paper, key points are extracted by using the explained two approaches mentioned above, and we compared the performance of the face recognition on each approach in the experiments.

2.2 Mean Shift
The proposal of Mean shift algorithm is made by\(^\text{15}\) and it is a non-parametric analysis technique for speculation of the density gradient. This can be used for the object tracking\(^\text{16,17}\) and clustering\(^\text{18}\) in computer vision and image processing. Mean shift clustering is performing to decompose the data set into homogeneous patterns in feature space by using the filter, which estimates the local density gradient of similar patterns. This speculation of gradient is gained by an iterative process, which finds the peaks in the local density. The members of the each cluster consist of all patterns that are shown upwards to the same peak. The maximum peaks in local density are gained as the mean-shift vector. The definition of mean-shift vector of \(x\) is shown.

\[
M(x) = \frac{\sum_{i=1}^{n} x_i g \left( \frac{x - x_i}{h} \right)}{\sum_{i=1}^{n} g \left( \frac{x - x_i}{h} \right)} - x
\]  

(7)

Where the \(n\) data point \(x_i\), \(x\) is the center of the kernel, while \(h\) is a bandwidth parameter, \(g(x)\) is \(d\)-vitiate kernel, and is defined as

\[
g(x) = 2\pi^{\frac{d}{2}} \exp \left( -\frac{1}{2} \|x\|^2 \right)
\]

(8)

Where \(x \in \mathbb{R}^d\) is a point located in \(d\)-dimensional feature space. For all time, the mean-shift vector point toward the direction of the maximum increase in the density. So the maximum peaks in local density are obtained by successive computation of the mean-shift vector and translation of the kernel \(g(x)\) by the mean-shift vector. At the end of the process, it is guaranteed to converge at a nearby point where the estimate has zero gradients. Denote by \(\{y_j\}_{j=1,2,\ldots}\) the sequence of successive locations of the kernel \(g(x)\), where, from (7),

\[
y_{j+1} = \frac{\sum_{i=1}^{n} x_i g \left( \frac{x - x_i}{h} \right)}{\sum_{i=1}^{n} g \left( \frac{x - x_i}{h} \right)} \quad j = 1,2,\ldots
\]

(9)

Is the weighted mean at \(y_j\) computed with kernel \(g(x)\) and \(y_j\) is center of the initial position of the kernel.

3. Proposed Method

The proposed face recognition system based on SIFT is made of training process and a recognition process. Figure 2 shows the overview of proposed system. In the training process, we firstly extract key points from the multiple training images of a subject by using SIFT. Next, a similarity matrix is creating by computing similarity between two images in the training set, and then we choose an image that is the best representation of the training set to be a template. The training set is restructured without the selected image. After that we select an image with the maximum score by calculating the similarity scores between the current template and each image in the training set for updating template. If the score of selected image is greater than \(T_{ms}\), the selected image is merged with the current template. Alternatively, the current template is recorded as the integrated template set and we calculate a similarity matrix for selection of a new template. The training process is repeated until the number of images in training set is zero and finally it produces one trained template in the integrated template set. The proposed recognition method is based on Lowes matching algorithm. However, one major difference is that a key point in query image is searched from individual integrated template instead of all the integrated templates. In addition, we use a computation of key point’s descriptor similarity and geometrical property to determine the identity of a query image. The recognition process begins with the key point descriptor-based matching and the geometrical property verification. Next, the matched pairs between a query
image and an integrated template are found by using key point’s descriptor and the transformed geometrical information which is applying the obtained affine transform parameters in the geometrical verification. Final step computes the similarity scores based on the accumulated sum of matched key point pairs.

3.1 Training Process

Under this Section, we discuss the details of the similarity matrix computation, template selection, similarity calculation and template updating in the proposed training process to generate combined templates from a training set of each subject. The training process starts with utilizing SIFT to represent the local feature for each subject which has a training set with N images. To select a template that best represents images in training set, a similarity matrix is computed as the similarity between two different images. A similarity matrix SM is defined as

\[ \text{SM}_{ij} = \frac{\text{SE}(I_i, I_j)}{m_i}, \quad i = 1, 2, ..., N, j = 1, 2, ..., N, i \neq j \]  

(10)

where similarity evaluation function \( \text{SE}(I_i, I_j) \) in section 3.2 is used to compute similarity of keypoints in the reference image \( I_i \) and keypoints in image \( I_j \). \( m_i \) is the number of keypoints in image \( I_j \). If \( i = j \) then \( \text{SM}_{ij} \) is zero. The represent- ability score \( rS_i \) is calculated as following:

\[ rS_i = \frac{1}{N-1} \sum_{j=1}^{N} \text{SM}_{ij} \]  

(11)

A reference image \( I_i \) with the maximum value of \( rS_i \) is chosen as the template of this subject, since it is more representative than the other images in training set. The training set of current subject is reset to \( N - 1 \) images which are the remained training images without selected image. The updating procedure, which synthesizes the training images to the template, is performed iteratively by the order of image similarity with the current template. For one of iteration, to select the most similar image in the training set with the template, the similarity score is calculated as the similarity of current template \( T_{E_c} \) on the image \( I_k \). The formula is \( rS_{E_c} = \text{SE}(T_{E_c}, I_k) \). Next, a candidate image for merging is selected as image \( I_k \) with the maximum value of \( rS_{E_c} \). Then, we applied a threshold of maximum score \( T_{ms} \) to consider whether the candidate image is suitable to merge the current template. If the similarity score of candidate image is greater than \( T_{ms} \) the reconfiguration of training set is performed without the candidate image then the updating template step is carried out. Alternatively, the current template is recorded in the reference templates and it fulfills a selection of new template in performed of current subject. A higher value \( T_{ms} \) may increment the number of templates per subject, so it can increase the computational difficulty greatly in the recognition process. In the case of low value

![Figure 2. Overview of proposed face recognition system (left: training process, right: recognition process).](diagram)
It may lead to the templates configuration with the incorrect geometrical position of key points. So we select $T_{gs} = 0.45$ carefully which have been tested as the best performance. The updating template stage is the task of assigning the key points which bundle up similar they in the current template and the selected image, into a key point of new template by using mean-shift in Section 2.2. To perform the mean-shift clustering by descriptor and geometrical information of key point, we use a modify way of the image segmentation method based on mean shift. Then, the results of clustering are depending on the two bandwidth parameters $h_g$ and $h_p$ of Mean-Shift where $h_g$ is the geometrical part and $h_p$ the part of descriptor. To determine the identity of a test face image among the integrated templates, the descriptors and position represented by the key points of query image are compared with those in reference templates. Through calculation of the similarity evaluation function, the subject of template with the maximum similarity to the query image is chosen to the identity of the query image.

3.2 Recognition process

To determine the identity of a test face image among the integrated templates, the descriptors and position represented by the key points of query image are compared with that of reference templates. Through calculation of the similarity evaluation function, the subject of template with the maximum similarity to the query image is chosen to the identity of the query image. The similarity evaluation function $SE(A, B)$ commences with the key point descriptor-based matching step, where $A$ is the key points of the template (or the reference image) and $B$ is the key points of query image. In the key point descriptor-based matching step, we adopt Lowe’s matching method to find a set of the matched key point in the descriptors feature space. Given the Euclidean distance $d_{ij}^p$ from the descriptor $P_i^B$ of one key point in $B$ to its first and second nearest $P_i^A, P_j^A$ in $A$, if the ratio between two distances is $d_{ij}^p < 0.8$, $P_i^B$ is paired with $P_i^A$. This step only takes the descriptor’s space into account and ignores the geometrical positions of key points. So, it always includes some mismatches in the result. Here we obtain a valid set of matched keypoint pairs by applying RANSAC (Random Sample Consensus) [6] which is to execute geometric consistency checking. The projection which is to transfer the position $(x_i^B, y_i^B)$ of keypoint in B to the geometric space $(x_i^A, y_i^A)$ of A, is fulfilled by the affine transform

$$
\begin{bmatrix}
  x_i^A \\
  y_i^A
\end{bmatrix} =
\begin{bmatrix}
  \alpha_1 & \alpha_2 & t_x \\
  \alpha_3 & \alpha_4 & t_y
\end{bmatrix}
\begin{bmatrix}
  x_i^B \\
  y_i^B
\end{bmatrix}
$$

where $\alpha_1, \alpha_2, \alpha_3$ and $\alpha_4$ are the coefficients about scale, rotation, compress, and stretch, while $t_x$ and $t_y$ are the translations along X- and Y-coordinates, respectively. The six affine transform parameters are calculated based on the set of matched key point pairs. We use the transformed position of key points in $B$ to calculate the similarity of a $B'$ keypoint $i$ to $A$ in the two space of descriptor and geometric. The geometrical distance $d_{ij}^g = \sum_{j=1}^m |T_{gs} - T_{gp}|$, where $m$ is the number of key points in $A$ ) between the a keypoint $i$ of B and every keypoints in $A$ obtain by calculating Euclidean distance of the transformed position of $B'$ keypoint $i$ with the position of every keypoints in $A$. A descriptor of $B'$ key point is calculating Euclidean distance with the descriptor of every key point in $A$ and their descriptor distances are recorded as $d_{ij}^p$. The similarity of a $B'$ keypoint $i$ to $A$ is defined as

$$
S_i = \begin{cases} 1 & D_i \neq 0 \\ 0 & \text{otherwise} \end{cases}
$$

$$
D_i = \left\{ d_{ij}^p | d_{ij}^p < T_g \text{ and } d_{ij}^p < T_p \right\}
$$

where $T_g$ and $T_p$ are the thresholds of geometric and descriptor, respectively. The similarity score $SE(A, B)$ and the similarity distance $SE_d(A, B)$ are calculated as the following.

$$
SE(A, B) = \sum_{i=1}^n S_i
$$

$$
SE_d(A, B) = \sum_{i=1}^n \min_{D_i \neq 0}(D_i)
$$

where $n$ is the number of key points in $B$. The identity of the query image is determined as that of the integrated template consisting of greatest similarity score. If there are multiple integrated templates having the same highest similarity score, the template with the minimum value of similarity distance among them is determined.

4. Experimental Results

In this section, we show how we try to find the optimal settings of parameter for the proposed face recognition
system based on SIFT. To test the effects of the proposed training procedure based on Mean-Shift, our method is comparing with the template selection and synthesis method proposed by. Next, the face recognition performance is analyzed according to the threshold values, image resolution and key point detection methods on the ORL database. The ORL database contains 400 images taken from 40 subjects and the size of 92×112 images with some lighting, pose and expression variation per subject. Each subject has 10 images with index \( i \) from 1 to 10.

4.1 Training process
The face recognition accuracy of the proposed method can be influenced by adjusting two bandwidth parameter \( h_g, h_p \) of Mean-Shift in training procedure and two threshold value \( T_g, T_p \) in recognition procedure. In order to determine the optimal values of parameters in training procedure, we first fix the value \( h_p \) and find the optimal values for \( h_g \). The parameter \( h_g \) is verified with the value from 1, 2, 3, 4, 5, and 6. Then we find the optimal value \( h_p \) by the determined \( h_g \). Next, we check the recognition rate according to the changes of the threshold value \( T_g \) and \( T_p \). All these experiments are taken with the ORL database, which is resized into 50×61 pixels. The key points are detected by using Lowe’s method. The training images consist of five images, which are selected randomly in each subject, and the recognition rate is calculated by testing of 200 remaining images. 10 runs of experiments are performed while each run has distinct training image set. Table 1 shows the performance for different \( h_g \) values with \( h_p = 0.4 \). The average recognition rate is taken by the values which are max recognition rate among performance according to the changes of the threshold value \( T_g \) and \( T_p \). It can be observed that the best recognition rate is achieved when \( h_g = 5 \), and the changes of \( h_g \) does not affect the number of key points per subject’s template. Next work has to determine the optimal parameter \( h_p \) value from 0.1, 0.2, 0.3, 0.4 and 0.5. Table 2 lists the effect of different value of \( h_p \) on the recognition performance. While increasing the value of \( h_p \), the number of keypoints per subject’s template starts to decrease by merging of key points in training images. So it is possible to decrease the computational complexity of the recognition process. Note that the average number of key points in 5 training images per subject is 147.87. However, after a certain value, higher \( h_p \) value leads to merge differing keypoints widely that might degrade the performance. From our experiment, \( h_p = 0.4 \) provides the most optimal recognition rate. So we selected that \( h_g = 5 \) and \( h_p = 0.4 \) as the optimal parameter values for Mean-Shift in the proposed training procedure. To obtain the optimal threshold values of \( T_g \) and \( T_p \), we set \( T_p \) in a range of \([0.01, 1]\) with quantized by 0.05, and \( T_g \) which equal to 1, 2, 3, 4, 5, and 6. Figure 3 shows recognition rate according to the changes of the threshold values and the result of experiment is obtained by using \( h_g = 5 \) and \( h_p = 0.4 \) of the optimal parameter values.

![Figure 3. Recognition rate by varying the threshold value T_g and T_p.](image-url)
We can see that the maximum average recognition rate keeps on increasing while the $T_g$ is increasing from 1 to 5 and recognition rate begins to decrease after $T_g = 5$. We observed that the maximum average recognition rate can be obtained in 0.2 - 0.3 range of $T_p$. So we select $T_g$ and $T_p$ as 5 and 0.25 respectively, which yields the best recognition rate.

### 4.2 Validation of Training Procedure

To test the effect of the proposed training procedure based on Mean-Shift, we repeat the same experiment by the different methods of key point detection and the changing number of training images. We change number of training images from 2 to 5 with 1 increment sequentially, and each increment is performed ten times by selecting images randomly in the ORL database in which images are resized into 50× 61 pixels. In this simulation, we use Lowe's and Geng's methods for detecting key point that is described in section 2.1. The proposed template training procedure provides recognition accuracy at 96.15% and 95.65% by using Lowe's and Geng's keypoint detection with 5 training images respectively. Table 3 provides result comparing recognition performance by Geng's the template selection, synthesis and matching. It can be observed that the proposed templates training procedure is more accurate than Geng's method while focus on the removing redundant information. The proposed method gives a lower standard deviation and is not much difference in recognition rate between the two detection methods for key point. So, it demonstrates that the variation of a composition of training images and a method of key point detection are less affect the proposed training procedure. Looking at the recognition rate and the number of key point per subject, depending on the detection method in the proposed procedure of table 3, the Geng's method than the Lowe's method improve the recognition rate a little; on the other hand it generate the template with a large number of key points. So the use of the Geng's method than the Lowe's method in the proposed training procedure amplifies the computational cost by the effect of growing number of key points procedure amplifies the computational cost by the effect of growing number of key points.

| $h_g$ | Recognition performance |
|-------|-------------------------|
|       | Average recognition     |
|       | rate (%)                |
|       | Average number of keypoints per subject's template |
| 1     | 91.25 117.02            |
| 2     | 93.45 110.44            |
| 3     | 94.20 109.12            |
| 4     | 95.10 108.56            |
| 5     | 96.15 108.25            |
| 6     | 95.90 108.04            |

| $h_p$ | Recognition performance |
|-------|-------------------------|
|       | Average recognition     |
|       | rate (%)                |
|       | Average number of key    |
|       | points per subject's template |
| 0.1   | 95.35 145.58            |
| 0.2   | 95.35 134.84            |
| 0.3   | 95.50 120.76            |
| 0.4   | 96.15 108.56            |
| 0.5   | 95.90 97.24             |

| The number of training images | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 |
|------------------------------|---|---|---|---|---|---|---|---|
| Average recognition rate     | 66.66 | 66.86 | 75.42 | 82.00 | 82.60 | 89.14 | 92.41 | 96.12 |
| Standard deviation            | 3.43  | 3.28 | 2.61 | 2.21 | 2.61 | 2.08 | 1.82 | 1.22 |
| Average number of keypoint per subject | 47.81 | 60.36 | 87.51 | 100.26 | 53.14 | 70.02 | 92.03 | 108.56 |

| The number of training images | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 |
|------------------------------|---|---|---|---|---|---|---|---|
| Average recognition rate     | 75.41 | 78.00 | 81.00 | 84.90 | 83.06 | 89.71 | 92.91 | 96.65 |
| Standard deviation            | 2.58  | 2.74 | 3.31 | 3.11 | 2.54 | 1.48 | 2.09 | 1.49 |
| Average number of keypoint per subject | 97.07 | 106.72 | 164.05 | 181.40 | 95.58 | 131.10 | 162.76 | 192.14 |
4.3 Recognition Performances

In the subsection, we need comparison of recognition rate of the proposed face recognition system and some other methods. Two databases, ORL and Yale are applied for testing the face recognition rate. The first 5 images of each individual from the ORL Database were chosen for training, the rest are used for testing. Table 4 summarizes the face recognition rate of some well-known and recent methods in ORL Database. The number in the bracket is the resolution associated with the used face image for each method. In Holistic approaches of Table 4, the results of PCA and Feature Fusion Approach (FFA) is referenced from reference 4 and 5 gives the result of Eigen faces, Fisher faces and Intrinsic faces. Among the SIFT approaches, the Geng_Syn uses the synthesized templates by Geng’s template training method with the key point detection. The Geng_MT uses the multiple templates per subject with Geng’s the key point detection method and the matching framework method and without the template training method. The results of SIFT, Geng_Syn and Geng_MT are taken from reference 14. The proposed LKD and proposed GKD are the result of proposed training procedure by using Lowe’s and Geng’s key point detection respectively. From this, it is observed that the recognition accuracy of the proposed GKD and Geng_MT is superior as compared to other approaches. However, the proposed system uses the integrated templates by removing the redundant key points from the training images, whereas the templates in Geng_MT are containing all key points of training images. So the proposed system requires less computation quantities than Geng_MT. The Yale Face Database consists of 165 grayscale images in 15 individuals. There are 11 images per subject, one per variations of illumination conditions, different face expressions and the face wearing glasses or not. The original size of the images in Yale Face database is 320 × 243 pixels with gray levels. For the experiments, each image was normalized in scale into various sizes. The 1, 2 and 3 face images of each subject were, respectively, selected as training images and the others were used as query images. Table 5 shows a compared recognition rate with PCA and FFA (which are taken from reference 4) by Yale database. It is noticeable that the performance of the proposed system is better than others.

Table 5. The recognition rate of methods on Yale database

| Methods       | Recognition rate | Image size |
|---------------|------------------|------------|
| PCA4          | 83.3%            | 50x61      |
| FFA4          | 73.3%            | 70x86      |
| Proposed LKD  | 90.0%            |            |
| Proposed GKD  | 90.83%           |            |

4.4 Resolution of Face Image

In this subsection, we show the change of recognition rate according to the resolution of face image in the proposed system. Then, the effect of the reduced number of key points per subject on the recognition rate and the resolution of face image are analyzed with the multiple templates per subject. In this experiment, we evaluate by using the ORL and Yale databases which are into resized different sizes. Two databases for training and testing conditions are the same as the subsection 4.3. Figure 4 shows the recognition rate for a different resolution of face images varying from 40 × 49 to 80 × 98. In case the LKD PTM which are used Lowe’s key points detection method achieved a best recognition accuracy of 96.5% and 90% for ORL and Yale databases respectively by using 50 × 61 resolution of face images. The GKD PTM by Geng’s key points detection method achieved a best recognition rate of 99.5% in ORL database with 60×74 resolution of face images; and a best recognition about Yale database obtained 90.83% in 70×86 resolution of face images. Figure 5 provide the results comparing the number

Table 4. The recognition rate of methods on ORL database

| Holistic approaches          | PCA4 | FFA4 | Eigen faces5 | Fisher faces5 | Intrinsic face5 |
|------------------------------|------|------|--------------|---------------|----------------|
| Recognition rate             | 90.5%| 92.0%| 87.5%        | 92.0%         | 95.5%          |
| Image size                   | -    | -    | 32x32        | 32x32         | 32x32          |

| SIFT approaches              | SIFT14 | Geng_Syn14 | Geng_MT14 | Proposed LKD | Proposed GKD |
|------------------------------|--------|------------|-----------|--------------|--------------|
| Recognition rate             | 90.0%  | 89.0%      | 99.5%     | 96.5%        | 99.5%        |
| Image size                   | 50x57  | 50x57      | 50x57     | 50x61        | 60x74        |
An improved Face Recognition Based on Scale Invariant Feature Transform (SIFT): Training for Integrating Multiple Images and Matching by Key Point’s Descriptor-Geometry

of keypoint in all templates by the proposed training procedure and the approach of multiple templates per subject with the varying the resolution of face images. In ORL database when 50×61 resolution of face images, we can see that the number of key points in templates by the proposed LKD PTM has decreased to 37.30% and 69.10% than the LKD MTM and GKD MTM respectively by approached multiple templates. In Yale database also, the proposed LKD PTM has a small number of key points in templates than the multiple templates methods. From results of this simulation shown in Table 4 and 5, we observed that the proposed template training procedure using Lowe’s key points detection method with 50×61 resolution of face images achieves higher recognition rate than the holistic approaches. The usage of Geng’s keypoints detection method in the proposed system obtains higher recognition rate than the usage of Lowe’s method. Instead it requires much more number of key points. In addition, the creation of integrated templates in accordance with the proposed template training method reduces the number of key points by average 49.84% and 21.64% than the method of using multiple templates, in ORL databases with 5 training face images and Yale database with 3 training face images, respectively.

5. Conclusion

SIFT which detects local features invariant to change of the scale, rotation and illuminating conditions in image, is an efficient feature extraction system for object and scene recognition. However, usage of SIFT for face recognition has many problems because face has the landmark of doubleness, non-rigid and smooth character compared to general objects. In this paper, we firstly proposed a new training procedure for the integration of templates to remove redundant features and blend different features in
the multiple training images. Moreover, in order to solve the problem of locational mismatching, we proposed the modified matching scheme. Which performs the one-to-one matching between query images and references templates by considering the similarity of key point’s descriptor and geometrical property. The experimental outcomes on two well-known face databases demonstrate the performance improvement by using the proposed method. The two proposals for face recognition based on SIFT are contributed to reduce the computational complexity through reduction of key points number per trained template and also improve the recognition accuracy. In future, we plan to apply the proposed training method to a SIFT based face clustering solution for facial expression recognition problem.

6. References

1. Kirby M, Sirovich L. Application of the karhunen-loeve procedure for the characterization of human faces. Pattern Analysis and Machine Intelligence, IEEE Transactions. 1990 Jan; 12(1):103-08.
2. Yang J, Zhang D, Yang JY. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions. 2004 Jan; 26(1):131-37.
3. Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherface: Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions. 1997 Jul; 19(7):711-20.
4. Xu Y, Zhang D, Yang J, Yang JY. An approach for directly extracting features from matrix data and its application in face recognition. Neuro Computing. 2008 Jun; 71(10-12):1857-65.
5. Zang F, Zang J. Face recognition using Elasticfaces. Pattern Recognition. 2012 Nov; 45(11):3866-76.
6. Lowe DJ. Distinctive image features from scale-invariant key points. International Journal of Computer Vision. 2004 Nov; 60(2):91-110.
7. Battiato S, Gallo G, Puglisi G, Scellato S. SIFT features tracking for video stabilization. In: Image Analysis and Processing. 14th International Conference on, Modena. 2007 Sep, p.825-30.
8. Chang L, Duarte MM, Sucar E, Morales EF. A Bayesian approach for object classification based on clusters of SIFT local features. Expert Systems with Applications. 2012 Feb; 39(2):1679-86.
9. Zhao WL, Ngo CW. Scale-rotation invariant pattern entropy for key point based near-duplicate detection. Image Processing, IEEE Transactions. 2009 Feb; 8(27):412-23.
10. Bicego M, Lagorio A, Grosso E, Tistarelli M. On the use of SIFT features for face authentication. In: Computer Vision and Pattern Recognition Workshop, IEEE International Conference. 2006 Jun, p. 35.
11. Cho M, Park H. A robust key points matching strategy for SIFT: An application to face recognition. Springer Berlin Heidelberg, 2009 Dec, p.716-23.
12. Luo J, Ma Y, Takikawa E, Lao S, Kawade M, Lu BL. Person-specific SIFT features for face recognition. In: Acoustics, Speech and Signal processing, IEEE International Conference, 2007 Apr, p.II-593-II-596.
13. Cruz C, Sucar LE, Morales EF. Real-time face recognition for human-robot interaction, In Automatic Face & Gesture Recognition. 8th IEEE International Conference on, Honolulu, HI. 2008 Sep, p. 1-6.
14. Geng C, Jiang X. Face recognition based on the multi-scale local image structures. Pattern Recognition. 2011 Oct-Nov; 44(10-11):2565-75.
15. Fukunage K, Hostetler L. The estimation of the gradient of a density function with applications in pattern recognition. Information Theory, IEEE Transactions. 1975 Jan; 21(1):32-40.
16. Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift, Computer Vision and Pattern Recognition, IEEE Conference on, Hilton, Head Island, SC. 2000,2, p.142-49.
17. Zhou H, Yuan Y, Chunmei S. Object tracking using SIFT features and mean shift. Computer Vision and Image Understanding. 2009 Mar; 133(3):345-52.
18. Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence. IEEE Transactions. 2002 May; 24(5):603-19.