Determination of 3-D Image Viewpoint Using Modified Nearest Feature Line Method in Its Eigenspace Domain

LINA*, BENYAMIN KUSUMOPUTRO++
*Faculty of Information Technology
Tarumanagara University
Jl. Letjen S.Parman 1, Jakarta 11440
INDONESIA
++Faculty of Computer Science
University of Indonesia,
Depok Campus 16424
INDONESIA

Abstract: Since viewpoint is one of the important information in face recognition system, a viewpoint estimation system is developed to determine the pose position of the 3-D human face images. The system is developed using projection of an unknown viewpoint of a spatial image into its eigenspace representation and calculated its distances to all of available lines connection between every two known viewpoints. We then developed FullyK-LT and SubsetK-LT methods, in order to transform the images from its spatial domain into eigenspace representation. In this paper, we propose a new method that modified the Nearest Feature Line method by adding more feature lines that can be acquired from the projections of every feature point to other feature lines. The developed system is then applied to determine the pose position of 3-D human face of Indonesian people in various positions and expressions. Using our method, it is proved that the increment number of feature lines improved the recognition rate significantly; due to it provide more feature information of an object. The experimental results shown that the highest recognition rate is 96.64% for SubsetK-LT using our Modified Nearest Feature Line method.

Key words: Viewpoint recognition system, Feature-line method, Karhunen-Loeve Transformation

1 Introduction

The growth of multimedia technology has triggered the needs of 3-D face recognition systems. Face identification is very important problem in law enforcement and forensics, authentication of access into building or automatic transaction machine, and searching for faces in video databases, intelligent user interfaces and others. This face recognition system can hopefully recognize 3-D human face images with various features, such as viewpoints changes, lighting, expression, etc [1]-[4].

Several methods, such as combined feature and template matching [2], template matching using Karhunen-Loeve transformation of large set of face images [3], have developed by many researchers, and good results can be obtained, especially for 2-D frontal or semi-frontal images. Since it is argued that 3-D recognition can also be accomplished by using linear combinations of as few as four or five 2-D viewpoint images [5,6], many researchers are trying to recognize more general view positions of images that cover the entire 3-D viewing sphere.

We have developed a 3-D face recognition system using a cylindrical structure of hidden layer neural network [7]. However, since the viewpoints of the given image is a priori unknown, we have assume that the face viewpoints of the given image is already known, which in turn, gave high recognition rate of the developed system [8, 9]. Since these viewpoints
are such important information for the 3D face recognition system, we develop pose position or viewpoints estimation system, to determine the pose position of the 2-D image correctly [10].

In this paper, we proposed a new method to estimate the viewpoint of 3-D objects that can also be used as a subsystem in our developed 3-D face recognition system. The developed viewpoint estimation system consists of a viewpoint features extraction subsystem that define the feature-space of the used system, and the viewpoint estimation subsystem that clustered the unknown viewpoint input face by comparing with viewpoints of faces which are stored in a gallery.

In the viewpoint features extraction subsystem, a feature-space is developed based on transforming every face in the spatial domain as a vector in the feature-space. As already written in [10], the authors have introduced two kinds of transformation methods, i.e., the Fully K-LT method and Subset K-LT method as the transformation procedures. Fully K-LT method is the Karhunen-Loeve transformation techneque that transforms all of the model images into only one eigenspace, while in Subset K-LT method; every group of images that lies in the same viewpoint will be transformed into every one-sub-eigenspace. By using this method, we can have many sub-eigenspace related with the number of viewpoints images.

In viewpoint recognition subsystem, a Nearest Feature Line method [11] will be used to cluster the unknown viewpoints of images by calculating its nearest distance with all of the feature points. The Nearest Feature Line (NFL) method assumes that at least we should have two feature points of each viewpoint class, and a line that can be draw between two points could represents some changes of faces due to lighting, expression, etc.

When there is an unknown face that would be determined its view position, it should be firstly transformed into its eigenspace representation by the K-L transformation. Clustering is then determined by calculating the nearest distance of the transformed-unknown point into all available feature lines in the eigenspace. We have proved that by using the conventional NFL techniques, the estimation rate of the system is not high enough. We then developed Subset K-LT method to increase the estimation rate of the system.

In order to further increase the viewpoint estimation rate of the system, we propose, in this paper, a new technique of increasing the number of feature lines without making any addition of the number of feature points. The increment of feature lines are accomplished by creating projection lines from each feature point to the other feature lines in the same class. The detail explanation about this technique including with its results will be discussed thoroughly in the next sessions.

2 Feature Extraction Subsystem

In this feature extraction subsystem, we use the Karhunen-Loeve transformation to develop a feature-space for higher discrimination power. Karhunen-Loeve transformation means to orthogonalize, which in turns increasing the separability of the extracted feature, and reducing its dimensionality [12]. This transformation will be used to project every image in the face gallery, into a feature point that has a smaller vector dimension.

As we have discussed earlier in this paper, the transformation process for image vectors is accomplished by the use the two kinds of Karhunen-Loeve transformation methods. In the Fully K-LT method, the unknown images that will be estimated its viewpoint should be projected into only one eigenspace. However, when using Subset K-LT method, we should firstly determine the matrix transformation of every viewpoint-class of the images. Suppose we have a class images with a viewpoints of 15°, then an eigenspace of 15° is constructed by using all of the images in this viewpoint-class. And this procedure is repeated until all sub-eigenspaces of all viewpoints of the images are completely constructed.

Using Subset K-LT method, therefore, the unknown images that will be estimated its viewpoint should be projected into all of the available eigenspaces. By using these two different methods, each image in spatial space is now represented as a feature point in only one eigenspace for Fully K-LT method or as a feature point in every one of the sub-eigenspaces for Subset K-LT method.
The standard analysis using its principal components of image data \( x(k) \in \mathbb{R}^n, k=0,1,...,m \) is done by determining a set of hidden linear representation parameters \( y_1(k), ..., y_m(k), \) called factors or features. These features could then be used for linear least mean-squared-error reconstruction of the original data. If there is enough correlation between the observation variables \( x_1(k), ..., x_d(k), \) then we can reconstruct the image with acceptable accuracy using a number of features \( m \) which is much smaller than the data dimension \( n. \) We often call \( n \) is the superficial dimensionality of the image data while \( m \) the intrinsic dimensionality of the image data.

The Karhunen-Loeve transformation method is firstly done by forming a base vector of image in \( d \) dimensions, i.e., \( z_n = [z_1, z_2, ..., z_d]. \) The next step is computing the average vector of \( M \) images through:

\[
\mu_z = \frac{1}{M} \sum_{n=1}^{M} z_n
\]

and determine the covariance matrix \( C_z : \)

\[
C_z = \frac{1}{M} \sum_{n=1}^{M} (z_n - \mu_z)(z_n - \mu_z)^T
\]

From this covariance matrix, we can derive a set of \( \lambda_z \) and \( e_z \) which are the eigen values and the eigenvectors. The eigenvectors \( e_z \) are orthonormal and the corresponding eigenvalues \( \lambda_z \) are nonnegative. Assuming that there are no repeated eigenvalues and that they are arranged in decreasing order, \( \lambda_1 > \lambda_2 > ... > \lambda_m, \) a matrix transformation is then constructed based on the importance of these eigen values. The \( e_z \) matrix transformation then be used to map a set of \( z_n \) image vectors to be a set of \( y_n \) feature vectors in eigen space, through :

\[
y_n = e_z^T (z_n - \mu_z)
\]

while the inverse reconstruction of \( z_n \) vectors can be done through:

\[
z_n = e_z^T y_n + \mu_z
\]

The importance of every eigenvalue that is related to its strength on giving an optimal matrix transformation for higher estimation rate, can be determined by computing the cumulative proportion of the eigen value using:

\[
a^k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{p-1} \lambda_j}
\]

The Karhunen-Loeve transformation matrix is then reformed by using this cumulative proportion, and we should recalculate the equations (3) and (4) to compute \( z_n' \) dan \( y_n'. \)

### 3 Viewpoint Recognition Subsystem

The viewpoint recognition subsystem used in this viewpoint estimation system is developed based on feature line distance calculation, for both FullyK-LT and SubsetK-LT. However, each method has different procedures on determining the viewpoint-class classification that will be discussed later. Each image in the spatial domain will be projected as a point in the eigenspace domain, and since we have only limited variations of images in the gallery, any variation could be approximated by point that lie in all of the possibilities of straight line between all of the points in that eigenspace domain. The straight line between those points of the same class is called feature-line (FL) of that viewpoint-class, and for each class we have:

\[
G_c = N_c (N_c - 1) / 2
\]

where \( G_c \) denotes number of feature lines and \( N_c \) the number of feature points from the viewpoint-class.

Suppose there is a new image with unknown-viewpoint that will be estimated. Assume that the unknown image is transformed into the eigenspace domain as \( x \) and its projection to the feature line of previous known points \( x_1 \) and \( x_2 \) is \( p. \) Projection-point \( p \) and its relation to \( x_1 \) and \( x_2 \) can be calculated through:

\[
p = x_1 + \mu (x_2 - x_1)
\]
with:

$$\mu = (x - x_1). (x_2 - x_1) / (x_2 - x_1). (x_2 - x_1)$$  \hspace{1cm} (8)

The minimum distance is then calculated for all of the viewpoint-class through:

$$d(x,p) = ||x - p||$$  \hspace{1cm} (9)

and is sorted in ascending order. The $$x$$ now is clustered into the viewpoint-class, which has the minimum distance.

In order to increase the viewpoint estimation rate of the system, we propose a new technique of increasing the number of feature lines without making any addition of the number of feature points. As already stated, the additional feature lines are done by creating projection lines from each feature point to the other feature lines in the same class, as can be illustrated in Fig.1.

![Fig.1 The process of forming the feature lines](image)

Assume we have three feature points $$x_1$$, $$x_2$$, and $$x_3$$, and in the conventional Nearest Feature Line (NFL) method, we can form only three feature lines to capture some variation from these two images. However, in our Modified Nearest Feature Line (M-NFL) technique, we add more feature lines by projecting each feature point to all available feature lines, i.e., point $$x_1$$ is projected to line $$x_3$$ to form the line $$x_1x_3$$, point $$x_2$$ is projected to line $$x_1x_3$$ to form the line $$x_2x_3$$, and point $$x_3$$ is projected to line $$x_1x_3$$ to form the line $$x_3x_1$$.

By the addition of feature line above, the total number of feature lines in this M-NFL method can be calculated through:

$$G = N_c (N_c - 1)^2 / 2$$  \hspace{1cm} (10)

as a substitution to the Eq.6.

By using this technique, every unknown image object in spatial domain will be firstly transformed to every eigenspace domain, and projected as point $$p$$ to all available feature lines through Eq.7, and calculated its nearest distance through Eq.9.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Train Images</th>
<th>Test Images</th>
<th>Training (degree)</th>
<th>Testing (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>52</td>
<td>0,60, 120,180</td>
<td>0,15,30,45, 60,75,90, 105,120, 135,150, 165, 180</td>
</tr>
<tr>
<td>%-age</td>
<td>23,53%</td>
<td>76,47%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>52</td>
<td>0,45,90, 135,180</td>
<td>0,15,30,45, 60,75,90, 105,120, 135,150, 165, 180</td>
</tr>
<tr>
<td>%-age</td>
<td>27,78%</td>
<td>72,22%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>52</td>
<td>0,30,60, 90,120, 150, 180</td>
<td>0,15,30,45, 60,75,90, 105,120, 135,150, 165, 180</td>
</tr>
<tr>
<td>%-age</td>
<td>35%</td>
<td>65%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**4 Experimental Result And Analysis**

The experimental procedure is conducted using a 3-D face database that consists of 4 Indonesian persons. The images are taken under four different expressions such as, neutral, smile, angry and laugh expressions. The 2-D images are given from 3-D human face image by gradually changing visual points, which is successively varied from -90° to +90° with an interval of 15°. Part of the 3-D human face database that is used in this experiment can be seen in Fig.2.

The overall data set of the experimental set-up is depicted in Table 1, including with its percentage of training/testing paradigm. The percentage of training/testing in Data#1 is 23.53%:76.47%, increased up to 27.78%:72.22% in Data#2, and up to 35%:65% in Data#3, respectively. It is hoped that after those experiments, we could have a clear picture of the minimum percentage in which the system still produces high estimation rate.
4.1 Viewpoint estimation by the use of FullyK-LT method

In the viewpoint estimation system using the FullyK-LT, the unknown viewpoint image $x$ is projected as $p$ into all available feature lines, and the minimum distance is then calculated between $x$ and its projected point $p$. Suppose that the minimum distance is determined to be in the line that connected two points of $x_1$ and $x_2$, and if those two points belong to the same viewpoint-class, then $x$ will be clustered into the same viewpoint-class with that of $x_1$ and $x_2$. However, if $x_1$ and $x_2$ are points that belong to different viewpoint-classes, then $x$ will be clustered into viewpoint-class that in between of those viewpoint-class in which $x_1$ and $x_2$ belongs.

Table 2. Recognition rate on determining the viewpoint using FullyK-LT method

<table>
<thead>
<tr>
<th>Data set</th>
<th>Recognition Rate (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
</tr>
<tr>
<td>1</td>
<td>26.92%</td>
</tr>
<tr>
<td>2</td>
<td>42.31%</td>
</tr>
<tr>
<td>3</td>
<td>28.85%</td>
</tr>
</tbody>
</table>

Results of experiments by using FullyK-LT method are presented in Table 2. As can be seen in this table, the highest estimation rate of the FullyK-LT is about 26.92% for Data#1, 42.79% for Data#2, and 30.29% for Data3, respectively. This results show that conventional NFL method could not determining the viewpoint of the unknown images with high estimation rate.

4.2 Viewpoint estimation by the use of Subset K-LT method

Let the unknown image in the spatial domain is transformed to be an unknown point of $x$ in every defined viewpoint-class in its eigenspace by the use of each SubsetK-LT matrix transformer. Then, project the transformed-point of $x$ into every available lines, and suppose the projection of the unknown point of $x$ become $p_r$, in each of the viewpoint-class, with $-90^\circ < r < +90^\circ$. The clustering mechanism is then determined by calculating the minimum distance of $xp_r$ which is the distance between point $x$ and $p_r$, using the Eq.9.

As we can see in the experimental data set, some of the unknown images have its viewpoint were in between of the known training set, and as the consequences, to have a higher estimation rate of the system, we can clustered the unknown point of $x$ into three categories of viewpoints, i.e., viewpoint-class $r$; or between viewpoint-class $(r - t)$ and viewpoint-class $r$; or between viewpoint-class $r$ and viewpoint-class...
(r + t); where t denotes the interval viewpoint of classes.

If the minimum distance \( \tilde{x}_r \) is nearly to 0 then \( x \) will be definitely clustered into viewpoint-class \( r \). However, if the distance \( \tilde{x}_r \) has a higher value, it means that the unknown viewpoint input image cannot be definitely clustered into viewpoint-class \( r \) or into viewpoint-class \( (r-t) \), or into viewpoint-class \( (r+t) \). If this is happened, another algorithm is developed to cluster the unknown viewpoint of an input image, which is illustrated in Fig. 3.

To precisely clustering the unknown viewpoint of an input image, we should calculated the minimum distance of the projected point to viewpoint-class of \( (r-t) \), \( \tilde{x}_{r-t} \), and compared with that of viewpoint-class of \( (r+t) \), \( \tilde{x}_{r+t} \). If \( \tilde{x}_{r-t} \) is higher than \( \tilde{x}_{r+t} \), then \( x \) will be clustered into viewpoint-class of \( (r-t) \) and viewpoint-class of \( r \). Otherwise, if \( \tilde{x}_{r-t} \) is lower than \( \tilde{x}_{r+t} \) then \( x \) will be clustered into viewpoint-class of \( r \) and viewpoint-class of \( (r+t) \).

Results of experiments by using SubsetK-LT method are presented in Table 3. As can be seen from Table 3, the estimation rate of the system is 48.08% for Data#1, which is increase up to 86.06% for Data#2, and the highest estimation rate is 90.38% for Data#3, respectively.

Those results show that even using lower training/testing paradigm such as in Data#2, the recognition rate of the system using SubsetK-LT is high enough to determine the viewpoint of the unknown image in the system. Comparing with that of the used FullyK-LT method, the estimation rate by using SubsetK-LT method is significantly high.

### 5 Additional of Feature Line for Higher Estimation Rate

As we have mentioned above, we propose a new technique by increasing the number of feature lines without making any addition of the number of feature points. As the FullyK-LT method is not appropriate in determining the viewpoint of the unknown images, in this experiments, we only used SubsetK-LT method as a platform on applying our M-NFL technique. Results of experiments by using SubsetK-LT method and the Modified Nearest Feature Line technique are presented in Table 4.

#### Table 4. Recognition rate on determining the viewpoint using SubsetK-LT method and Modified NFL technique

<table>
<thead>
<tr>
<th>Data set</th>
<th>Recognition Rate (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
</tr>
<tr>
<td>1</td>
<td>46.63%</td>
</tr>
<tr>
<td>2</td>
<td>85.09%</td>
</tr>
<tr>
<td>3</td>
<td>90.38%</td>
</tr>
</tbody>
</table>

As we can see in this table, the estimation rate is about 48.56% for Data#1, and could be increased up to 92.79% for Data#2 with 99% of cumulative percentage of the used eigenvalues. While the highest estimation rate is achieved when using Data#3, i.e., 96.64% with 99% of cumulative percentage of eigenvalues. This result has shown that our Modified NFL technique could increase the estimation rate of the developed system.
Table 5 shows more rigorously about the comparison of the NFL and our M-NFL technique. Second column shows the correct classification percentage using NFL that is classified correct also by our M-NFL method, while the third column shows the percentage of incorrect classification by NFL method but classified as correct classification by M-NFL method. The fourth column shows the inverse condition of the third column, in which the correct classification NFL method is classified as incorrect by M-NFL method.

It is shown that in every used data-set with its different cumulative percentage, the percentage of being classified as incorrect by NFL but classified as correct by M-NFL method (Column 3rd) is always higher than the inverse condition (Column 4th). Table 5 also shows that the percentage of the correct classification by both NFL and M-NFL methods is increasing along with higher percentage of the training/testing paradigm. Higher estimation rate along with higher training/testing paradigm is also happened in the Column 3, while being stable in the Column 4.

![Table 5](image)

This shows that increasing the percentage of training/testing paradigm is not affected much in the disadvantage of M-NFL, while it would increasing the advantage M-NFL method, compare with that of using NFL method. This comparison shows, that the nearest calculation through our method has higher percentage of correctness, which means that the addition of feature lines can improve the estimation rate in the viewpoint estimation system.

### 6 Conclusions

We had developed a viewpoint estimation system based on a viewpoint features selection subsystem and a viewpoint estimation subsystem. Fully K-LT Subset K-LT methods are utilized, and based on its results of experiments, the Subset K-LT method shows higher estimation rate compare with that of Fully K-LT method. To increase the estimation rate higher, we then developed a Modified NFL technique by increasing the number of feature lines without increasing the number of points in the eigenspace domain. Result of experiment shows that the used of M-NFL technique with Subset K-LT method performed on 99% of cumulative percentage of using eigenvalues has highest estimation rate of about 97%, compare with only about 30% when using NFL technique with Fully K-LT method using the same cumulative percentage value. It is shown that our model is robust enough to determine the viewpoint of unknown images. This system is now being developed as a subsystem in our 3D face recognition system.

### ACKNOWLEDGEMENT

The authors would like to express their gratitude to Ministry of Research and Technology of Indonesia for their support funding of this research through RUT X under contract number 14.06/SK/RUT/2003

### References:

3. M. Kirby and L. Sirovich, Application of the Karhunen-Loeve procedure for the...


