An Improved Face Clustering Method Using Weighted Graph for Matched SIFT Keypoints in Face Region

Ji-Soo KEUM and Hyon-Soo LEE, Members

1. Introduction

Recently, various methods have been proposed using facial information (e.g., face recognition, authentication, clustering, and retrieval). In particular, methods for finding a person in video and photos have received much attention [1], [2]. Using such a method in video and photo collections, we can easily track individuals and automatically construct a photo album of a specific person.

In order to determine whether a queried person is the subject of interest, an approach based on eigenfaces was widely applied in face recognition [3]. The eigenface-based approach basically reduces the dimensionality of the feature space for the face region and then apply a pattern classification algorithm to verify individuals. Before we apply these methods to facial information-based identification problems, we should normalize the size of the face region to match the size of the images in the database. Also, the proposed method shows acceptable performance in a constrained environment but remains sensitive to changes of illumination, pose, and facial expression.

The scale invariant feature transform (SIFT) feature was proposed for object and scene recognition [4]. This method is used to extract distinctive features from images. SIFT is invariant to the size and orientation of the keypoint and robust to changes in lighting conditions. For these reasons, SIFT is widely used in general object recognition, near-duplicate detection, and object tracking [5], [6].

Based on the distinctive ability of SIFT, it is also employed in face recognition and clustering. However, face regions are non-rigid and smooth compared to general objects. Therefore, some studies have investigated the ability of SIFT in face recognition [7], [8]. A person-specific SIFT feature was proposed with local and global similarity from the sub-region in [7]. In addition, a new approach was proposed to remove unreliable SIFT keypoints [8]. Instead of extracting a more distinctive feature based on the naive SIFT feature, some research has focused on the matching scheme using constraints on the face region. In another work [2], the visual language model was constructed to determine whether two faces are the same individual for face clustering. A graph-based approach was suggested to find the same people in news photographs and video collections [1]. Another study proposed the geometric constraint and unique-match constraint to overcome the problem of the matching method in [4].

Through the review of face clustering methods using SIFT keypoint matching, we determine that a new discriminative parameter is necessary, and that the combination with existing matching schemes and constraints is required to improve performance. Therefore, we focus on the extraction of a new parameter and the combination with recently well applied constraints. The proposed parameter is calculated using the analysis of orientation of matched keypoints with constraints. By combining the proposed parameter with generally used average similarity, we can improve the ability to distinguish individuals. The improved performance for several datasets is shown in the experimental results.

2. Description of the Proposed Method

Figure 1 shows the block diagram of the proposed method based on the weighted graph using average similarity and orientation matching ratio.

2.1 SIFT Keypoints Matching with Constraints

Before matching the keypoints between two images, we first find the face region using the method proposed in [9]. The SIFT keypoints that have been detected in the face region are used for face clustering.

First, a minimum similarity of keypoints between two face regions is specified to exclude keypoints that have low similarity. The SIFT keypoints are represented as $P^A = \{p_1^A, p_2^A, \ldots, p_m^A\}$, $P^B = \{p_1^B, p_2^B, \ldots, p_n^B\}$ for the detected face region in image $A$ and image $B$, where $m$ and $n$ are the number of SIFT keypoints, respectively. The similarity matrix $S$
The keypoint matching phase is that we consider more candidates for keypoints of interest and automatically eliminate keypoints that have low similarity. Therefore, we identify more reliable keypoints before applying the geometric constraint. For example, if our method finds only one keypoint that has the lowest Euclidean distance, it sometimes fails to find the true keypoint that is located at a similar position. Therefore, we allow identification of keypoints with multiple matches, because the geometric constraint effectively eliminates false matches.

In SIFT keypoints matching with constraints phase, we newly proposed SIFT keypoints matching scheme and employed two constraints to verify correct matches.

2.2 Orientation Analysis for Matched Keypoints

The SIFT algorithm returns a 128-dimensional descriptor, orientation, and magnitude for a keypoint of interest [4]. Most face clustering methods use only the descriptor in keypoint matching [1], [2], [7], [8]. And, a research showed that the possibility for usability of orientation coherence in face recognition [10].

In this research, we analyze the orientation information in detail and newly define a parameter called orientation matching ratio for the matched keypoints. In order to use additional information for an extracted keypoint, we analyze the orientation information returned from the SIFT algorithm for the matched keypoints in the keypoint matching phase. The detected face region is surrounded by a reminded face region (e.g., forehead, cheek, and lower jaw). Therefore, the orientation of an extracted keypoint is similar, though some changes of facial expression exist.

Figure 2 shows the verification scheme of matched keypoints with the orientation condition. We verify matched keypoints using the orientation and allowing angle. The keypoint $p^A_1$ is matched with the keypoint $p^B_1$ in Fig. 2, where the orientation difference is within the allowable angle range (e.g., $\pm 10^\circ$). Also, the keypoint $p^A_2$ is matched with the keypoint $p^B_2$ by the previous similarity processing and two constraints. However, we discard this matching in the calculation of the matching ratio because the orientation difference

![Verification of matched keypoints with orientation condition](image-url)
is greater than the allowed angle range. By applying this constraint, we identify more correct matches combined with similarity and other constraints.

We define the orientation matching ratio $O_{mr}$ using the orientation condition as follows:

$$O_{mr} = \sqrt{\frac{\text{NoS}_{AB}}{\text{min}(\text{NoMA}, \text{NoMB})}}$$  \hspace{1cm} (3)

where $\text{NoMA}$ and $\text{NoMB}$ are the number of matched keypoints in the similarity comparison phase for matches $P_A \rightarrow P_B$ and $P_B \rightarrow P_A$, respectively. $\text{NoS}_{AB}$ is the number of keypoints that satisfies the geometric and OOS constraints together with the proposed orientation condition, and $O_{mr}$ has a value between 0 and 1. It is 1 when the matched keypoints perfectly satisfy all conditions; otherwise it closes to 0.

Figure 3 shows an example of SIFT keypoint matching for each step. Figure 3 (a) shows initial matching using the similarity, and Figs. 3 (b) and 3 (c) show the results of filtering for geometric constraint and OOS conditions, respectively. The keypoints in Fig. 3 (d) are verified with the proposed orientation condition, where the arrow indicates the keypoint orientation.

2.3 Weighted Graph Construction

The face clustering was performed using the weighted graph $G(V, E)$, where the vertices $V$ are faces, the edge $E$ corresponds to weight by the average similarity $S_{avg}$, and $O_{mr}$ was previously defined for matched keypoints between image $A$ and image $B$. The weight of edge $E$ between face $A$ and face $B$ is

$$E = w \cdot S_{avg} + (1 - w) \cdot O_{mr}$$  \hspace{1cm} (4)

where $S_{avg}$ is calculated as

$$S_{avg} = \frac{\sum_{r^i \in R^A} S_{r^i,m^i}}{|R^A|}$$  \hspace{1cm} (5)

$w = [0, 1]$ is a weighting parameter for the construction of the weighted graph, which is determined experimentally. $r^A_i$ is a keypoint for the reminded keypoints in the $R^A$ subset of $P^A$ that satisfies minimum similarity and geometric and OOS conditions. $m^B_i$ is a keypoint that matches the keypoint $r^A_i$ in face $B$, and $|R^A|$ is the cardinality.

The weighted graph is converted into binary form using a given threshold. Previous research [1] found the densest component after graph cutting in order to find a highly connected sub-graph. Identification of the densest component allows the most similar person to be chosen because the component result is closest to the data from the most similar individual and distant from the others.

In the face clustering phase, we constructed the weighted graph using newly proposed similarity value and orientation matching ratio. And, we employed the finding of the densest component in our weighted graph instead of applying other graph theoretical clustering analysis.

3. Experimental Results

3.1 Datasets

We tested the proposed method using three different data sets: the ORL faces, Caltech Faces, and FERET database [11]–[13]. We performed experiments consisting of 25 rounds by changing individuals, images, the threshold for geometric constraint, and the weighting parameter $w$ for the graph edge. We randomly selected five images of each individual from the ORL and Caltech datasets. We also randomly selected individuals from the FERET datasets and used four $(fa, fb, hr, hl)$ images to evaluate performance for different facial directions. Table 1 shows the conditions of the experimental datasets, such as selected individuals, images, the average size of the detected face region, and the average number of SIFT keypoints.

3.2 Performance Comparison

We compared the performance of our method with that of a previous work [1]. In the face clustering, we consider parameters such as the geometric distance and weighting parameter of the graph. The geometric distance was changed from 0.1 to 1.0 with a 0.1 interval. We consider only the average similarity when $w$ is 1.0; when $w$ is 0.0, the orientation matching ratio is considered. The threshold for binary graph generation that outputs the best performance in F-Ratio was automatically selected. We used precision (PRC), recall (RCL), and F-Ratio for performance evaluation.

Table 2 shows a performance comparison of the previous graph-based method [1] and the proposed method. All parameters are adjusted to output best performance, and

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Individuals</th>
<th>Images</th>
<th>Face size</th>
<th>Keypoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL [11]</td>
<td>5, 10, 15, 20</td>
<td>5</td>
<td>65 × 65</td>
<td>42</td>
</tr>
<tr>
<td>Caltech [12]</td>
<td>5, 10, 15, 20</td>
<td>5</td>
<td>235 × 235</td>
<td>188</td>
</tr>
<tr>
<td>FERET [13]</td>
<td>5, 10, 15, 20</td>
<td>4</td>
<td>118 × 118</td>
<td>48</td>
</tr>
</tbody>
</table>

### Figure 3

Example of SIFT keypoints matching. (a) initial matching of keypoints using similarity, (b) after applying geometric constraint, (c) matched keypoints after applying OOS condition, (d) result of orientation verification for reminded keypoints.
Table 2  The result of performance comparison of face clustering.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individuals</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>PRC</td>
<td>0.635</td>
<td>0.487</td>
<td>0.416</td>
<td>0.344</td>
</tr>
<tr>
<td>RCL</td>
<td>0.706</td>
<td>0.622</td>
<td>0.544</td>
<td>0.532</td>
</tr>
<tr>
<td>F-Ratio</td>
<td>0.660</td>
<td>0.539</td>
<td>0.463</td>
<td>0.414</td>
</tr>
<tr>
<td>PRC</td>
<td>0.851</td>
<td>0.812</td>
<td>0.762</td>
<td>0.704</td>
</tr>
<tr>
<td>RCL</td>
<td>0.710</td>
<td>0.634</td>
<td>0.595</td>
<td>0.574</td>
</tr>
<tr>
<td>F-Ratio</td>
<td>0.768</td>
<td>0.707</td>
<td>0.662</td>
<td>0.626</td>
</tr>
</tbody>
</table>

Fig. 4  Comparison of F-Ratio depending on the geometric distance.

Fig. 5  Comparison of F-Ratio depending on the weighting parameter $w$.

the values are averaged for 25 rounds. In the experimental result, the proposed method outperforms the previous method. In particular, the proposed method shows improved performance on precision, which indicates that the proposed method reliably reduces false alarms. Through the performance comparison, we can summarize as follows. When many keypoints exist in the face region, the proposed method can improve precision and recall. However, when there are few keypoints and faces of the dataset have different directions (such as in the FERET database), the two methods show relatively lower performance compared with other datasets (such as Caltech Faces).

Figure 4 shows the performance comparison depending on the geometric distance for two methods when the number of individuals is 5. The best performance is the same as that shown in Table 2. The proposed method shows high performance in all ranges of geometric distance. Figure 5 shows the result of comparison depending on the weighting parameter $w$ for the proposed method. When $w$ is between 0.8 and 0.9, the proposed method shows the best performance. The proposed method uses matched keypoints that have the highest similarity instead of finding a keypoint that has the lowest Euclidean distance. Therefore, the performance of two methods is not identical when the weighting parameter was 1.0. Through the change of performance depending on $w$, we can see that the selection of $w$ is not sensitive, and the proposed similarity and other conditions show reliable performance combined with the proposed orientation matching ratio.

4. Conclusion

In this paper, we have proposed an improved face clustering method using a weighted graph that consists of average similarity and orientation matching score. It is an unsupervised face clustering method using weighted graph construction. The experimental results show that the proposed method can be efficiently applied in face clustering, and confirms that the method improves overall performance for different datasets.

Acknowledgment

This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (Ministry of Education, Science and Technology) [NRF-2011-355-H00012].

The authors would like to thank the anonymous reviewers for their helpful suggestions.

References


