Smile, You’re on Identity Camera

Ye Ning, Terence Sim
School of Computing, National University of Singapore
{yening,tsim}@comp.nus.edu.sg

Abstract

Inspired by recent advances in psychological studies on motion-based face perception, we examine in this paper, from the viewpoint of pattern recognition, the identity information behind a smile. A smile video database is collected, from which we compute dense optical flow fields and generate features by summing up the flow fields over time during the neutral-to-smile period. We investigate the relationship between smiles and identity by studying the class separability of the features. Our experiment results indicate a strong identity-specific characteristic of smile dynamics. Moreover, we compare the discriminating power of the features generated from different face regions as well as from different periods of motion.

1. Introduction

Human faces contain rich information about personal identities. Babies are known to use facial features, among others, to recognize individuals. However, in scientific studies, this face perception/recognition process has never been fully understood. Questions like “what kind of facial information is used” and “how it is used in identity recognition” are still unanswered. Research in face perception/recognition has primarily focused on analyzing static facial features, like shape, color, etc. However, recent studies in psychology and computer vision suggest that facial motion may also contribute to the recognition of personal identities to a considerable extent. This paper contributes to this research by examining, from the viewpoint of pattern recognition, the identity information behind smiles.

There are three psychological hypotheses [11] concerning the role of facial motion in human face perception/recognition. The one most related to our work is known as supplemental information hypothesis. It states that facial motion provides identity-specific dynamics signature which benefits the perception of identity. Supportive evidence have been reported by several psychological research groups independently. Lander et al. [6] asked their experiment participants to recognize famous faces from static images, multi-static images and moving images of facial expressions separately. The highest recognition rate was achieved with moving images. Similar observations were reported by Thornton and Kourtzi [12]. Pilz et al. [10] familiarized their participants with a set of frontal-view static face images showing a neutral expression, as well as with a set of frontal-view videos showing some facial expressions, and then asked them to recognize non-frontal neutral-expression static face images. Participants took less time for recognition when they were familiarized with videos. Lander et al. [7] asked their participants to recognize faces from real smile videos and synthetic smile videos. A higher recognition rate was reported using real smiles.

In the computer vision community, researchers try to extract discriminative features from facial motion and use them for computer-controlled identity recognition. Zhang et al. [14] estimated the elasticity of the mas- seter muscle from a pair of side-view face range images (neutral and mouth-open faces) to characterize a face. A verification rate of 67.4% was reported at a false alarm rate of 5%. Tulyakov et al. [13] extracted the displacement of a set of feature points from a pair of face images (neutral and smile) to characterize a face. An equal error rate of slightly less than 40% was reported. Liu et al. [8] used HMM (Hidden Markov Models) to model the global face motion from videos of specific tasks, e.g. walking. Subsequently, they used the HMMs for identification, achieving an error rate of 4%. Chen et al. [1] used the visual cues observed in speeches to recognize a person. Facial dynamics was extracted by computing a dense motion flow field from each video clip. A recognition rate of about 86% was achieved.

Compared with the aforementioned work, our study focuses on the core scientific question: *how discriminating is a smile?* We believe we are the first to investigate the inherent discriminating power in a smile. We show that smiles are highly discriminating, and moreover, that the lower face is up to 3 times more discriminating than the upper face in terms of the Bayes’ error
2. Data Collection and Feature Extraction

While most existing facial expression video databases provide one to three smile videos per subject, our research requires many more smile videos per subject to carry out reliable statistical analysis on class separability. Thus, the first step of our research is to build our own smile video database.

2.1. Smile Video Database

Our video database consists of 30 to 40 smile video clips from each one of the 10 objects, or 341 clips in total. Every clip records the facial motion of one subject performing a smile. The expression begins with a neutral face, moves to a smile, and then back again to the neutral expression. Videos were recorded at 15fps under the resolution of 768 by 1024 pixels. The subjects were asked to perform their own smiles as naturally as they could. Before recording, a template smile video was shown to the subject to remind him/her of the proper intensity of the smile (in order to avoid too small or too big smiles). Also, an LCD display was placed before the subject to let the subject see himself/herself during recordings, because we found that the subjects smiled more naturally when they were able to see themselves. We asked the subjects to take a rest after every 4 or 5 times of recordings. The whole recording was conducted in two sessions over two days to avoid fatigue.

2.2. Face Alignment and Feature Extraction

For face alignment, the positions of eyes were manually marked on the first frame of each video clip and then tracked through the entire clip. Videos were converted into image sequences. These images were then aligned by an in-plane 2D linear transformation with respect to the positions of eyes. After alignment, face regions were cropped and resized to 81 by 91 pixels. Since our subjects smiled at a consistent speed each time, we did not temporally align the video frames.

We compute dense optical flow fields between consecutive frames in the cropped face image sequence using the Lucas-Kanade algorithm [9]. Then, we estimate the normalized smile intensity in each frame as,

\[
I_{\text{smile}}^i = \frac{\sum_{x \in \text{pixels}} \| \sum_{t=1}^k f^t(x) \|^2}{\sum_{x \in \text{pixels}} \| \sum_{t=1}^k f^t(x) \|^2}
\]  

where \( I_{\text{smile}}^i \) denotes the normalized smile intensity in the \( i \)-th frame; \( x \) is a pixel; \( f^t(x) \) denotes the optical flow of pixel \( x \) in the \( t \)-th frame; and the \( K \)-th frame contains the smile apex (the peak of the curve in Figure 1). We then sum up all optical flow fields during the neutral-to-smile period,

\[
F(x) = \sum_{t=1}^K f^t(x)
\]

Thus, \( F(x) \) summarizes all pixel-wise observations of the facial motion from neutral to the smile apex. Figure 1 shows a plot of Eq.1 over the duration of a smile. For a \( w \times h \) pixel image, we stack all \( wh \) vectors \( F(x) \) into a long \( 2wh \times 1 \) column vector \( u \). Then we perform PCA (Principal Components Analysis [3]) to get,

\[
v = P_d u
\]

where the matrix \( P_d \) consists of the first \( d \) principal components. Please note that all the information we exploit for generating \( v \) comes from the smile dynamics along, that is, the optical flows. No texture or shape information is used.

3. Class Separability Analysis

We investigate the relationship between smile and identity by studying the class separability of the generated features. Since the features are generated purely from smile dynamics, a high class separability of the features will suggest a strong identity-specific characteristic of the smile dynamics.
We adopt the approach proposed by Cover and Hart [2]. They have proved that, when the number of samples $N$ approaches infinity, the following inequality holds,

$$ R^* \leq R \leq R^* \left( 2 - \frac{1}{\alpha R^*} \right) \tag{5} $$

or,

$$ \alpha - \sqrt{\alpha^2 - \alpha R} \leq R^* \leq R \tag{6} $$

where,

$$ \alpha = \frac{M - 1}{M} \tag{7} $$

Here $R^*$ denotes the Bayes’ error rate; $R$ denotes the 1NN error rate; $M$ denotes the number of classes (in our case, $M = 10$). And $R$ is computed as,

$$ R = \frac{|\{ v | \theta(v) \neq \theta(v_{nn}) \}|}{N} \tag{8} $$

where $v$ is the feature computed from Eq.4; $\theta(\cdot)$ denotes the labeling function; $v_{nn}$ denotes the nearest neighbor of $v$; and $N$ denotes the number of samples in total. In other words, $R$ is the fraction of those sample points whose class labels are different from those of their nearest neighbors.

According to [2], the bounds given by Eq.6 are tight. Although in real-world applications, it is impossible to get infinite number of samples (in our case, $N = 341$), it is a reasonable practice to indirectly measure the Bayes’ error rate using the 1NN error rate.

Figure 2(b) shows the band of the Bayes’ error rate $R^*$ estimated by Eq.6 at $M = 10$. The horizontal axis denotes the number of principal components $d$ used in dimension reduction (Eq.4). The upper and lower bounds of $R^*$ drop to 0.0029 and 0.0015 respectively at $d = 6$. After $d > 6$, both curves are largely flat, with minor ripples. Such a low error rate suggests clear separation between the underlying probability distributions of the 10 classes, which suggests a high class separability of the extracted features. This means that smile dynamics can be used to identify people.

### 3.2. Upper Face vs. Lower Face

We also examine the features generated from upper-face regions and lower-face regions separately to investigate which part of the face is more discriminating. Figure 3 shows the experiment results. We can see that the upper bound of the lower-face error rate (the blue curve with triangles) is always equal to or lower than the lower bound of the upper-face error rate (the dashed red curve with dots). The lower-face error rate can be less than the upper-face error rate by as much as 3 times at $d = 10$. This observation implies that the lower face is more discriminating than the upper face. This stands in contrast to static face recognition, where it has been shown that the upper face is more discriminating [5, 4].

---

Figure 2. Class separability studies

Figure 2(a) visualizes the features after being projected onto the first two principal components. Although the first two principal components preserve only 35.36% of the total energy, the projected features from the 10 classes (i.e. 10 subjects) form visually well-separated clusters (except for Class 3 and Class 5). Quantitative analysis of the class separability is carried out by estimating the Bayes’ error rate using the 1NN error rate (single nearest neighbor error rate).

### 3.1. The Bayes’ Error Rate

The Bayes’ error rate is the theoretical minimum rate any classifier can achieve. Therefore, the ideal way of measuring class separability is by calculating the Bayes’ error rate based on the underlying probability distributions of those classes. However, directly calculating the Bayes’ error rate is difficult in practice, because the calculation requires the probability density functions which are generally unknown in most applications. Various methods have been proposed to estimate the Bayes’ error rate from a set of observations.

---

We adopt the approach proposed by Cover and Hart [2]. They have proved that, when the number of samples $N$ approaches infinity, the following inequality holds,

$$ R^* \leq R \leq R^* \left( 2 - \frac{1}{\alpha R^*} \right) \tag{5} $$

or,

$$ \alpha - \sqrt{\alpha^2 - \alpha R} \leq R^* \leq R \tag{6} $$

where,

$$ \alpha = \frac{M - 1}{M} \tag{7} $$

Here $R^*$ denotes the Bayes’ error rate; $R$ denotes the 1NN error rate; $M$ denotes the number of classes (in our case, $M = 10$). And $R$ is computed as,

$$ R = \frac{|\{ v | \theta(v) \neq \theta(v_{nn}) \}|}{N} \tag{8} $$

where $v$ is the feature computed from Eq.4; $\theta(\cdot)$ denotes the labeling function; $v_{nn}$ denotes the nearest neighbor of $v$; and $N$ denotes the number of samples in total. In other words, $R$ is the fraction of those sample points whose class labels are different from those of their nearest neighbors.

According to [2], the bounds given by Eq.6 are tight. Although in real-world applications, it is impossible to get infinite number of samples (in our case, $N = 341$), it is a reasonable practice to indirectly measure the Bayes’ error rate using the 1NN error rate.

Figure 2(b) shows the band of the Bayes’ error rate $R^*$ estimated by Eq.6 at $M = 10$. The horizontal axis denotes the number of principal components $d$ used in dimension reduction (Eq.4). The upper and lower bounds of $R^*$ drop to 0.0029 and 0.0015 respectively at $d = 6$. After $d > 6$, both curves are largely flat, with minor ripples. Such a low error rate suggests clear separation between the underlying probability distributions of the 10 classes, which suggests a high class separability of the extracted features. This means that smile dynamics can be used to identify people.

---

**Table 2. Class separability studies**

<table>
<thead>
<tr>
<th>Class</th>
<th>2-D visualization of features</th>
<th>Upper bound of $R^*$</th>
<th>Lower bound of $R^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 2</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 3</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 4</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 5</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 6</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 7</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 8</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 9</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
<tr>
<td>Class 10</td>
<td><img src="image" alt="2-D visualization of features" /></td>
<td><img src="image" alt="Upper bound of $R^*$" /></td>
<td><img src="image" alt="Lower bound of $R^*$" /></td>
</tr>
</tbody>
</table>

---

**Figure 2. Class separability studies**

Figure 2(a) visualizes the features after being projected onto the first two principal components. Although the first two principal components preserve only 35.36% of the total energy, the projected features from the 10 classes (i.e. 10 subjects) form visually well-separated clusters (except for Class 3 and Class 5). Quantitative analysis of the class separability is carried out by estimating the Bayes’ error rate using the 1NN error rate (single nearest neighbor error rate).

**3.1. The Bayes’ Error Rate**

The Bayes’ error rate is the theoretical minimum rate any classifier can achieve. Therefore, the ideal way of measuring class separability is by calculating the Bayes’ error rate based on the underlying probability distributions of those classes. However, directly calculating the Bayes’ error rate is difficult in practice, because the calculation requires the probability density functions which are generally unknown in most applications. Various methods have been proposed to estimate the Bayes’ error rate from a set of observations.
3.3. Neutral-to-Smile vs. Smile-to-Neutral

Finally, we examine the features generated from the neutral-to-smile period (the red curve in Figure 1) and from the smile-to-neutral period (the blue curve in Figure 1) separately to investigate which period of motion is more discriminating. Figure 4 shows the experiment results. We can see that the two upper bounds (the two blue curves) overlap each other almost everywhere, so do the two lower bounds (the two red dashed curves). This observation implies that neutral-to-smile motion and smile-to-neutral motion provide almost the same amount of information about identity.

4. Conclusions

In this paper, we examined, from the viewpoint of pattern recognition, the identity information behind smiles. We generated features purely from smile dynamics, i.e. optical flows during smiles. We investigated the relationship between smiles and identity by measuring the class separability of those features. Our experiment result suggests that smiles are highly discriminating. We also compared the discriminating power of the features generated from different face regions (upper face vs. lower face) as well as from different periods of motion (neutral-to-smile vs. smile-to-neutral).

In future, we hope to construct a larger facial expression video database and study the identity information behind other facial expressions.

5. Acknowledgements

We would like to thank all the volunteers who were willing to contribute their lovely smiles to this research. We also want to thank Miss Li Jianran for prove-reading the article. Research was supported by the NUS Tallyface project (R252-000-261-422).

References