Preventing Impersonation in Face Authentication Using DCT Signs

Takaharu Kouda\textsuperscript{a,*}

\textsuperscript{a} Department of Electrical Engineering, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, JAPAN

*Corresponding Author: kouda@fukuoka-u.ac.jp

Abstract

The purpose of this study is to propose a method for detecting the changes in expression in a facial image from changes in DCT signs and determining whether the subject is a real person or a photograph. Facial detection was performed on two frames taken continuously with the camera, and DCT signs were used to determine whether or not changes in expression occurred in the detected facial images. First, we check whether there are any large changes such as changes in direction or tilting between the two frames of facial images. Next, we check for the presence of changes in expression using only the center portion of the facial image. If a change in expression occurs, the image taken with the camera is determined to be a real person. As a result, we were able to detect blinking or changes in expression from changes in DCT signs. We conclude that it is possible to determine whether a photographed subject is a photograph or a real person using the proposed method, and that this could be used to prevent impersonation in facial detection systems.

Keywords: Impersonation, Face Authentication, DCT signs.

1. Introduction

In recent years, smartphones and tablet computers have come into widespread use. Smartphones in particular have seen impressive spread and performance enhancement, and a great deal of information is stored and carried within the phones. Smartphones are also often used as terminals for web applications. For this reason, the risk of information leakage when a phone is lost has become a serious problem. Face authentication has been proposed as a method for solving this problem. Because user authentication using a facial image can be used easily and without physical contact, it is considered an effective method for the improvement of security. For this reason, many recent smartphones include a locking function using facial authentication. However, it is possible that other users could use a facial photograph to impersonate the owner and unlock the terminal. This is because the phone does not determine whether the subject in the image captured with its camera is a photograph or a real person. To prevent such type of impersonation, detecting blinking or changes in expression in the facial image and using them to determine whether the subject captured with the camera is a photograph or a real person is vital to security.

Existing methods employ blink detection as a method for preventing facial authentication impersonation. Eye image pattern matching and detecting opening and closing of eyelids using facial detection have been used as methods for detecting blinking\textsuperscript{(1,2)}. To detect blinking, accurate detection of eye location is essential. However, if the subject is closing his/her eyes or wearing glasses, the location of the eye cannot be accurately detected. Performing both facial detection using the generally used Haar-like features and eye detection requires a significant amount of processing time. To improve user convenience, it is important to perform processing in real-time.

The purpose of this study is to propose a method for detecting the changes in expression in a facial image from changes in DCT signs and determining whether the subject is a real person or a photograph.

Facial detection was performed on two frames taken continuously with the camera, and DCT signs were used to determine whether or not changes in expression occurred in the detected facial images. First, we check whether there are any large changes such as changes in direction or tilting between the two frames of facial images. Next, we check for the presence of changes in expression using only the
center portion of the facial image. If a change in expression occurs, the image taken with the camera is determined to be a real person.

As a result, we were able to detect blinking or changes in expression from changes in DCT signs. We conclude that it is possible to determine whether a photographed subject is a photograph or a real person using the proposed method, and that this could be used to prevent impersonation in facial detection systems.

2. Proposed Method

2.1 Feature of DCT Signs

The DCT are commonly used in the field of image compression and pattern recognition. DCT and inverse DCT are defined as follows:

\[
F(u,v) = \frac{2C(u)2C(v)}{W} \sum_{m=0}^{W-1} \sum_{n=0}^{H-1} f(m,n) \times \cos \left(\frac{(2m+1)u\pi}{2W}\right) \cos \left(\frac{(2n+1)v\pi}{2H}\right) 
\]

IDCT

\[
f(m,n) = \sum_{u=0}^{W-1} \sum_{v=0}^{H-1} C(u)C(v)F(u,v) \times \cos \left(\frac{(2m+1)u\pi}{2W}\right) \cos \left(\frac{(2n+1)v\pi}{2H}\right)
\]

where,

\[
C(w) = \begin{cases} 
1 & (w = 0) \\
\sqrt{2} & (w \neq 0) 
\end{cases}
\]

Inverse transformation in the case of DCT signs only is referred to as Sign-only Synthesis (SOS). The DCT SOS is defined as

\[
f(m,n) = \sum_{u=0}^{W-1} \sum_{v=0}^{H-1} C(u)C(v)e^{i\theta} \times \cos \left(\frac{(2m+1)u\pi}{2W}\right) \cos \left(\frac{(2n+1)v\pi}{2H}\right)
\]

where,

\[
\theta = \begin{cases} 
0 & F(u,v) \geq 0 \\
\pi & F(u,v) < 0 
\end{cases}
\]

Figure 1 shows the sample of SOS processing. Looking at Figure 1, we can see that SOS detects the edge of the image. From this, we can say that the DCT coefficient signs contain image edge information. Therefore, the DCT sign is used for such as face recognition.

2.2 Facial Detection

Here, we perform face detection using two continuous frame images taken with the camera. Haar-like features are used for facial detection. The detected facial region is scaled and trimmed to 64×64 pixels. Figure 2 shows an example of the trimmed facial image. As Figure 2 shows, the facial image includes background in addition to the face.

2.3 Detecting Changes in the Facial Image

In this study, because we only perform facial detection and do not detect eyes, we do not normalize the facial image based on the position of the eyes. For this reason, it is necessary to differentiate changes in the detected facial image due to changes in the face’s direction or tilt from changes in expression. Here, we check for large changes
between the two frames of the detected facial image. First, we reduce the trimmed facial image to 32×32 pixels. This is done to reduce the effects of changes in expression on large changes in the facial image such as changes in facial direction. Next, we perform DCT on the reduced facial image. Next, we extract 128 large amplitude DCT coefficient signs and their positions from the current frame of the facial image. Using the extracted DCT sign positions, we extract 128 DCT signs from the previous frame of the facial image. Finally, we compare the DCT signs. If more than 125 of the 128 signs match, we determine that there are no changes in facial direction or tilt between the two frames. If less than 125 DCT signs match, we determine that there was a change in the direction or tilt of the face between the two frames and perform facial detection again. The number of DCT signs to extract and their threshold were determined experimentally.

2.4 Detecting Changes in Expression

The facial image shown in Figure 2 includes information unrelated to changes in expression, such as face outline, bangs, and the background. In order to remove these portions, we extract a 32×40 pixel region from the center of the image. An example of the extracted region is shown in Figure 3. First, we perform DCT on the extracted image. Next, we extract 56 large amplitude DCT coefficient signs and their positions from the current frame of the image. However, the positions of the extracted DCT signs are limited to the positions of the white pixels in the mask pattern shown in Figure 4. This mask pattern was created in advance using the facial images of multiple people. The white pixels in the mask pattern are DCT sign positions which change easily with changes in expression, and which do not change easily with image blur. As shown in Figure 5, changes in DCT signs due to changes in expression often appear in the high frequency component. However, high frequency component DCT signs change easily with image blur (Fig. 6). When a picture moves in front of the camera, the camera is sometimes unable to focus and takes a blurry photograph. In order to ensure that such blur is not detected as a change in expression, the mask pattern shown in Figure 4 is introduced. Next, using the extracted DCT sign positions, we extract DCT signs from the previous frame of the facial image. Finally, we compare the DCT signs. If more than 8 of the 56 signs have changed, we determine that the subject of the photograph has blinked or changed his or her expression.
3. Experiment

Table 1 shows results for an experiment with five participants. Participants blinked and changed expressions in front of the camera for one minute. Table 1 shows that over 92.9% of blinks and changes in expressions were detected. However, we found that changes in expressions were more difficult to detect in participants wearing glasses (Person ID is B and D). When wearing glasses, frame information appears in large amplitude DCT coefficients. It is thought that for this reason, there was less information on changes in expression in the extracted DCT signs. Even when a photograph was moved in front of the camera, the detected facial image was almost never determined to be a live person. This confirms that it is possible to determine whether the subject of a photograph taken with a camera is a real person or a photograph with the proposed method. In addition, the processing times required when experimenting on a low-spec PC are shown in Table 2. Table 2 demonstrates that the proposed method executes at a high speed since only facial detection is performed.

4. Conclusions

The purpose of this study is to propose a method for detecting the changes in expression in a facial image from changes in DCT signs and determining whether the subject is a real person or a photograph.

Facial detection was performed on two frames taken continuously with the camera, and DCT signs were used to determine whether or not changes in expression occurred in the detected facial images. First, we check whether there are any large changes such as changes in direction or tilting between the two frames of facial images. Next, we check for the presence of changes in expression using only the center portion of the facial image. If a change in expression occurs, the image taken with the camera is determined to be a real person.

As a result, we were able to detect blinking or changes in expression from changes in DCT signs.

In conclusion, the information obtained in this experiment is given below.

(1) It was possible to detect the presence of changes in expression through detecting changes in DCT signs in the facial image.

(2) By dividing processing into two steps (detection of facial image changes and detection of expression changes), it was possible to determine whether the

![Fig. 5. The DCT sign which is easy to change by an expression change.](image5)

![Fig. 6. The DCT sign which is easy to change by the blur of the image.](image6)

<table>
<thead>
<tr>
<th>Person ID</th>
<th>The number of expression change</th>
<th>The number of detections</th>
<th>Detection rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>35</td>
<td>34</td>
<td>97.1</td>
</tr>
<tr>
<td>B</td>
<td>28</td>
<td>25</td>
<td>89.3</td>
</tr>
<tr>
<td>C</td>
<td>31</td>
<td>30</td>
<td>96.8</td>
</tr>
<tr>
<td>D</td>
<td>33</td>
<td>29</td>
<td>87.9</td>
</tr>
<tr>
<td>E</td>
<td>31</td>
<td>29</td>
<td>93.5</td>
</tr>
</tbody>
</table>

Table 2. Experimental Result of Low-spec PC.

<table>
<thead>
<tr>
<th>CPU</th>
<th>Intel Atom Z550 2.0GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>2GB</td>
</tr>
<tr>
<td>Face detection [fps]</td>
<td>15</td>
</tr>
<tr>
<td>Detection of expression change [fps]</td>
<td>13</td>
</tr>
</tbody>
</table>
subject was a photograph or a real person using only facial detection without performing eye detection.

(3) Short processing times allowed for the detection of changes in expression in real-time, even on a low-spec PC. This suggests that this system could be implemented on portable terminals such as smartphones.

If this system is used on portable terminals such as smartphones, environmental lighting on the face may change based on where the user is located. Therefore, future research will explore the issue of dealing with changes in lighting.

References