Improving Face Recognition Systems Security Using Local Binary Patterns

Abdolhossein Fathi\(^\text{a}\),\(\ast\), Fardin Abdali-Mohammadi\(^\text{a}\)

\(\text{a}\) Department of Computer and Information Technology, Faculty of Engineering, Razi University, Kermanshah, Iran.

\(\ast\) Corresponding author.

Email addresses: a.fathi@razi.ac.ir (A. Fathi), fardin.abdali@razi.ac.ir (F. Abdali-Mohammadi)

ISSN: 2322-4460 © 2015 JComSec. All rights reserved.

1 Introduction

Traditional authentication systems widely use knowledge-based and token-based mechanisms. The knowledge-based systems employ username and password for authentication, and the token-based systems use tokens, namely, things usually carried by a person which can identify him, such as keys, passport, ID card, etc. The main drawbacks in the traditional systems are: using simple passwords (which are simply guessed, disclosed or stolen by attackers), counterfeiting, losing ID documents, and the wear and tear of these documents. The new authentication systems are designed and manufactured based on biometric features. These systems use biological and behavioral information of persons such as face [1], fingerprint [2], voice [3], palm print [4], iris [5], ear [6] and gait [7] for identification and authentication purposes.

Face is a very prominent biometric feature that can be used to identify persons uniquely. Facial Images contain such important information as a persons identity, gender, expression, age, ethnicity, etc. Thus, the idea of using face for identification purposes was initiated and later developed in the form of face-based biometric systems. During the past two decades, special focus has been placed on peoples face as a biometric feature for recognition and authentication purposes [8–12].

Although biometrical features for each user are unique, they suffer from spoofing or copy attacks. Since user’s face is visible, its voice is recordable and
fingertips are left everywhere, the spoofing attack is a fatal threat of the biometrical authentication systems.

Face recognition systems also suffer from different spoofing attacks like photograph-based, video-based and 3D model-based attacks [13]. The photograph-based attack is the most common way to spoof face recognition systems, because it is easy to obtain an image of the face of a valid user. The constructing of 3D model of valid user’s face is another way for spoofing face recognition systems. This way is more difficult than photograph-based methods. Video of valid user can also be obtained easily using high quality cameras. Video based spoofing is more difficult to detect, because videos have more physiological information like eye blink, head movements and facial expressions.

There are some efforts on anti-spoofing for face recognition systems [13–21]. Most of these methods only concentrated on the face liveness detection for photograph-based spoofing attacks. Eye blink detection is the most conventional method used for face liveness detection [13–15]. Eye blink is a biological activity of eyes which some methods tried to use it as liveness measure. These methods detect and track the persons eyes and if they have a blink, they assume the input subject is live user. However, simple eye blink detection method may fail when attackers use video-based spoofing.

Head and lips motion analysis are other techniques used for detecting face liveness [16–18]. Frischholz et al. [16] tried to estimate head position and extract head motion as a measure for face liveness when user moves and nods his/her head. Kollreider et al. [17] also tried to use the optical flow of face region as a measure of face liveness. Rua et al. [18] tried to detect the face liveness by measuring the degree of synchrony between the lips movement and the voice which is extracted from a video sequence.

Image quality degradation detection is also explored for photograph-based spoofing detection [19, 20]. Li et al. [19] used the deference between the Fourier coefficients of real images and imposter images to detect photograph-based spoofing from live faces. Maatta et al. [20] used the deference between the shape and texture of real images and imposter images to detect liveness. Thermogram of input image, which was obtained by infrared camera, was also used by Socolinsky et al. [21] for liveness measurement in photograph-based spoofing detection.

However, simple eye blink detection, head or lips motion extraction methods may fail when attackers use video-based spoofing, because videos have more physiological information like eye blink, head movement and facial expressions.

In this paper, we improve face recognition system security by employing efficient scene texture analyzing method to overcome the video-based spoofing attacks. To this end, the scene of input and reference images are divided into same non-overlapped blocks and the texture pattern of each block is extracted by local binary pattern (LBP) operator. The similarity of LBP pattern of corresponding block in the input image and reference background image is used as a measure for detecting video-based spoofing attacks. To increase the reliability of proposed method and also to reduce the effect of noise on LBP operator, the texture pattern of Y, Cb and Cr color channels are extracted independently. The majority of similarity of corresponding channels in each block of the input image and reference image is used as the similarity measure of that block. By obtaining the similarity of all blocks, we can easily discriminate video-based attacks from live users.

This paper is organized as follows: the proposed scene texture analyzing method for video-based spoofing detection is described in Section 2. Experimental results are reported in Section 3. Conclusion will be given in Section 4.

2 The proposed scene analyzing method for video-base spoofing detection

In the video-based spoofing, attackers prepare a clip of valid user and play it back on the authentication camera using Laptop, Tablet or mobile devices. These equipment should be placed close to the camera and this causes to cover almost all the background of the system, in addition to the space covered by user body. Therefore, the backgrounds of the spoofing videos are different from the original (reference) background of the system based on their scene contexts. Also when spoofing videos are captured in the authentication environment, their camera angles have been different.

To overcome the video-based spoofing attacks, the proposed system should analyze the scene of the input frames and compare them with the reference scene. Since the lighting conditions change the pixel values, direct comparison of pixels may increase the false rejection and may not have suitable performance. Therefore, we perform the comparison between structural textures of scene’s blocks. To this end, local binary patterns (LBP) operator is used [22]. LBP describes the texture information in term of local structure primitives using the value of P equally spaced neighborhood points on the circle with radius R. Each local texture pattern is represented with binary codes \(BC_{P,R}\) which is obtained by comparing the neighbor-
hood points \( \{b_0, b_1, \ldots, b_{p-1}\} \) with the value of central pixel \( b_c \):

\[
BC_{P,R}(i) = \begin{cases} 
1 & b_i \geq b_c \\
0 & b_i < b_c 
\end{cases} \quad i = 0 \cdots P - 1 
\]  

(1)

In the classical LBP (proposed by Ojala et al. as LBP\textsuperscript{prv2} [22]) all patterns categorized into two classes: uniform and non-uniform patterns; and only uniform patterns are selected as local texture features:

\[
LBP_{P,R} = \left\{ \sum_{i=0}^{P-1} BC_{P,R}(i) \right\} \quad if \ U(BC_{P,R}) \leq 2 \\
\quad P + 1 \quad otherwise 
\]  

(2)

The uniform patterns contain at most two bitwise transitions from 0 to 1 or vice versa in the obtained binary code when it is considered as circular structure:

\[
U(BC_{P,R}) = \left| BC_{P,R}(P - 1) - BC_{P,R}(0) \right| + \sum_{i=0}^{P-2} \left| BC_{P,R}(i) - BC_{P,R}(i + 1) \right| 
\]  

(3)

Since in the LBP\textsuperscript{prv2} non-uniform texture patterns are discarded, we employed a general extension of LBP operator (LBP\textsuperscript{GRI}) , proposed by Fathi and Naghsh-Nilchi [23], that attends to use all uniform and non-uniform patterns using a proper rotation-invariant scheme:

\[
LBP_{P,R}^{GRI} \left( \begin{array}{c} 
\sum_{i=0}^{P-1} BC_{P,R}(i) \\
(P - 1)(\lambda - 1) + 1 + \\
\sum_{i=0}^{P-1} BC_{P,R}(i) \end{array} \right) \quad if \ \lambda(BC_{P,R}) = 0 \\
(P - 1)(\lambda - 1) + 1 + \\
\sum_{i=0}^{P-1} BC_{P,R}(i) \quad if \ \lambda(BC_{P,R}) > 0 
\]  

(4)

In the LBP\textsuperscript{GRI} operator all local texture patterns are clustered based on their roughness value (\( \lambda \)). The value of roughness is equal to half of uniformity measure (U). This operator tried to describe texture patterns efficiently and has more compatibility with distribution of patterns in real images (see [23]). Figure 1 shows an example of extracting the LBP values.

To compare the scene of the input frames with the reference image, first we transform the input image from RGB color space to YCbCr space, and then, we divide the whole image into non-overlapped blocks with 10×10 pixels. Then the texture patterns of Y, Cb and Cr channels of each block are extracted by applying LBP\textsuperscript{GRI} at the center point of block. This LBP used 16 points as shown in Figure 2. The value of each point that not falls into a pixel will be computed by average of surrounded pixels.

To increase the speed of the proposed method, the LBP operator at one scale is applied on each block. Therefore by comparing the obtained LBP texture patterns of Y, Cb and Cr channel of all blocks in the input and reference images, the scenes similarity of them is determined and the video based spoofing can be detected. Two blocks is similar if at least the LBP patterns of two channels are equal. The scene similarity is obtained by calculating the similarity (or dissimilarity) of all blocks as below:

\[
S(LBP_R, LBP_I) = 1 - \frac{1}{T} \sum_m \sum_n \sum_{C \in Y,Cb,Cr} \left( \sum_{(m,n)_{m,n} \neq (m,n)} |LBP_{P,R}^{GRI}(m,n) - LBP_{P,R}^{GRI}(m,n)| \right) < 1 
\]  

(5)

where T is the total number of blocks and operator returns one if the LBP patterns of color channel C (Y, Cb and Cr) of the \((m,n)_{m,n}\) block in the input image (LBP\textsubscript{I}) and reference image (LBP\textsubscript{R}) is not identical; otherwise it returns zero. If the LBP patterns of two color channels are different, dissimilarity is occurred and 1 is added to the number of existing dissimilar blocks. To accept the input video as live video, the scene similarity measure S should be greater than

<table>
<thead>
<tr>
<th>Start Point</th>
<th>154</th>
<th>136</th>
<th>105</th>
<th>114</th>
</tr>
</thead>
<tbody>
<tr>
<td>154</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>154</td>
</tr>
<tr>
<td>95</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>99</td>
</tr>
</tbody>
</table>

Figure 1. The details of obtaining the LBP value for LBP\textsuperscript{prv2} and LBP\textsuperscript{GRI} operators.

\[BC_{3,1} = 11001100\]

\[U(BC_{3,1}) = 4\]

\[LBP_{prv2} = 9\]

\[LBP_{GRI} = 9\]
threshold $T_\beta$ (we experimentally set $T_\beta$ to 0.3, see Section 3).

3 Experimental Results

The performance of the proposed method against the photograph- and video-based spoofing attacks was evaluated on two datasets: real videos, and fake clips. The first dataset consists of 90 clips from 30 valid users that the system captured at three different times for each person. The second dataset was used for evaluating the anti-spoofing capability of the proposed method against video and photograph attacks. For each valid user two fake clips, one indoor and one outdoor, were captured in the place that its scene differ from the system environment. For each valid user also one photograph attack was simulated by moving the photo of it in the front of camera. Each video clip is captured with 30 fps and its frame size is 320×240 pixels. Some samples of these datasets along with similarity values and the obtained LBP of different color channel of RGB, HSV and YCbCr color spaces are shown in Figure 3.

To evaluate the proposed method, the detection performance is measured using true positive ($TP$), false positive ($FP$), true negative ($TN$) and false negative ($FN$) metrics. $TP$ is defined as the number of real videos correctly detected; $FP$ is defined as the number of fake clips detected as real video; $TN$ is defined as the number of fake videos correctly detected; and $FN$ is defined as the number of real videos not detected by the system. By using these metrics we can obtain more meaningful performance measures like sensitivity and specificity values as below:

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{Specificity} = \frac{TN}{FP + TN} \tag{7}
\]

We evaluated the ability of the proposed method by applying it to the all clips in two datasets. First we evaluated the effect of different color spaces and the value of scene similarity threshold ($T_\beta$) on the performance of the proposed method. To this end, the proposed method was applied on the three channels of RGB, HSV and YCbCr color spaces. In each color space the performance of the proposed method in detecting real and fake clips was calculated when different values for scene similarity threshold ($T_\beta$) have been used. The obtained results are shown in Figure 4.

From the obtained results, all three color spaces have similar sensitivity values in all range of threshold ($T_\beta$). For similarity threshold $T_\beta$ greater than 0.4, the specificity value of all color spaces are same. But specificity value of YCbCr color space for similarity threshold $T_\beta$ less than 0.3 is better than other color spaces. Therefore this color space was selected in final scene analyzing method.

The proposed scene texture analysis can resist against different spoofing attacks when the similarity threshold $T_\beta$ was set to 0.5 or greater. But in this range of $T_\beta$ the sensitivity of the proposed method for real videos is low, and not only all spoofing clips but also greater than 30% of real valid users will be rejected as spoofing cases. To increase the sensitivity of the proposed method, we set threshold $T_\beta$ to 0.3. In this case both the sensitivity and specificity of the proposed method will be greater than 98%.

Although the spoofing attacks using the system background in the fake clips is theoretically possible, however, it sensitizes to viewpoint, scale, and illumination. It is too hard to align the region of interest in a spoofing clip to corresponding region in the reference image. Also from the viewpoint of the fixed camera, the scale and coordinate of objects in the reference image are different from the scale and coordinate of objects in the spoofing clips. However, if the backgrounds of the spoofing clip and reference image are very simple and similar, such as white wall background, it would be hard to separate them with background texture analysis. In this case we should change the background scene, use additional steps like facial motion analysis or employ multi camera scene analysis.

We also evaluated the robustness of the proposed method against Gaussian white noise that may affect the input images. In these experiments input clips were contaminated by Gaussian white noise at different standard deviations: $\sigma = 5, 10, 15, 20$ and 30. These noises only added to one channel of YCbCr color space. We also applied LBP on blocks of gray scale version of images, which was obtained from color image, and compared them to detect spoofing attacks. In these experiments $T_\beta$ was set to 0.15. The obtained results are summarized in Table 1.

From the obtained results, the proposed method for calculating the similarity measure using the majority of three channels of YCbCr color space shows high performance in all noisy conditions and can resist against noise. On the contrary, by extracting the LBP patterns and calculating similarity measure only from gray scale images, the sensitivity of the proposed method was decreased dramatically, especially for high intensity noises. But the specificity values remain high.

For further evaluation, we compared the proposed method with some of existing methods. In this experiment the results of other methods are reported from
LBP\_R = 6  
LBP\_G = 0  
LBP\_B = 6  
LBP\_H = 6  
LBP\_S = 0  
LBP\_V = 6  
LBP\_Y = 6  
LBP\_Cb = 0  
LBP\_Cr = 6

(a)

S = 0.4857  
LBP\_R = 6  
LBP\_G = 8  
LBP\_B = 3  
LBP\_H = 1  
LBP\_S = 15  
LBP\_V = 6  
LBP\_Y = 6  
LBP\_Cb = 0  
LBP\_Cr = 6

(b)

S = 0.0091  
LBP\_R = 35  
LBP\_G = 38  
LBP\_B = 11  
LBP\_H = 10  
LBP\_S = 24  
LBP\_V = 11  
LBP\_Y = 37  
LBP\_Cb = 11  
LBP\_Cr = 6

(c)

S = 0.0156  
LBP\_R = 26  
LBP\_G = 14  
LBP\_B = 23  
LBP\_H = 1  
LBP\_S = 10  
LBP\_V = 26  
LBP\_Y = 26  
LBP\_Cb = 18  
LBP\_Cr = 21

(d)

Figure 3. Some samples of reference image (a), real video of valid user (b) and fake clips (c, d) along with similarity values (S) and the obtained LBP of different color channel of RGB, HSV and YCbCr color spaces.

Figure 4. The obtained results of the proposed method when different color spaces and different values for scene similarity threshold (T\_β) have been used. a) The obtained sensitivity values. b) The obtained specificity values.

Table 1. The obtained results (%) in the presence of Gaussian noise.

<table>
<thead>
<tr>
<th>Color</th>
<th>Noise</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
<th>Noise Free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YCbCr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>98.9</td>
<td>98.9</td>
<td>98.9</td>
<td>98.9</td>
<td>98.9</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>100</td>
<td>98.9</td>
<td>97.8</td>
<td>98.9</td>
<td>97.8</td>
<td>98.9</td>
</tr>
<tr>
<td>Gray</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>93.3</td>
<td>88.9</td>
<td>82.2</td>
<td>72.2</td>
<td>47.8</td>
<td>96.7</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>93.3</td>
<td>93.3</td>
<td>91.1</td>
<td>92.2</td>
<td>90.0</td>
<td>95.6</td>
</tr>
</tbody>
</table>

their paper on their datasets, because neither their datasets nor implementation codes available. The comparison is summarized in Table 2. Based on this table, the proposed method has high performance on han-
Table 2. The comparison of the proposed method with some existing methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>User Collaboration</th>
<th>Irresistible</th>
<th>Photo Imposter</th>
<th>Video Imposter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liting et al. [13]</td>
<td>Medium</td>
<td>Video</td>
<td>98.7 %</td>
<td>-</td>
</tr>
<tr>
<td>Pan et al. [14]</td>
<td>Medium</td>
<td>-</td>
<td>98.0 %</td>
<td>95.6 %</td>
</tr>
<tr>
<td>Szwoch et al. [15]</td>
<td>Medium</td>
<td>Video</td>
<td>84.6 %</td>
<td>-</td>
</tr>
<tr>
<td>Kollreider et al. [17]</td>
<td>Low</td>
<td>Video</td>
<td>99.5 %</td>
<td>0</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>-</td>
<td>-</td>
<td>98.9 %</td>
<td>98.9 %</td>
</tr>
</tbody>
</table>

4 Conclusion

This paper presents a new method to empower the face recognition systems against video-based spoofing by employing efficient scene texture analysis. The scene of input and reference images are divided to 10×10 non-overlapped blocks and the texture pattern of three channels of YCbCr color space in each block are extracted by local binary pattern (LBP) operator. The majority of similarity of three channels is used as the similarity measure of each block. The similarity of all blocks in the input image and reference image is used as a measure for detecting video-based spoofing attacks. The proposed algorithm shows the sensitivity and specificity values greater than 98%, when it was evaluated against different video-based spoofing attacks in real environments especially in presence of noise. In the future work we can concentrate on the problem of simple backgrounds and photograph-based attack when the background of imposter image was cut. In this case we should employ multi camera scene analysis or additional steps like eye blink count detection or lip motion analysis.

Acknowledgements

This paper is a part of Research Plan entitled as Anti-Spoofing method for face recognition systems with Code 1987 at Engineering Faculty of Razi University.

References


Abdolhossein Fathi was born in Delfan, Iran, in 1978. He received his B.Sc. in Computer Engineering from Iran university of Science and Technology in 2001 and his M.Sc. in Computer Architecture from Sharif University of Technology in 2003. He also received his Ph.D. Degree from the University of Isfahan in 2012. He is an assistant professor at Razi University, Kermanshah, Iran. His research interests include Signal and Image Processing, Biometric, Data Compression, Computer Vision, Pattern Recognition and Medical Image Analysis.

Fardin Abdali-Mohammadi was born in KhorramAbad, Iran, in 1979. He received his B.Sc. in Computer Engineering from Sharif University of Technology in 2001 and his M.Sc. in Software Engineering from Amirkabir University of Technology in 2003. He also received his Ph.D. Degree from University of Isfahan in 2012. He is an assistant professor at Razi University, Kermanshah, Iran. His research interests include Biomedical Signal and Image Processing, Bioinformatics, Medical Informatics, Machine Learning, Scientific Computation.