IMPOSTOR DETECTION USING FACIAL STEREOSCOPIC IMAGES

Abdelaali Benaiss[†]*, Usman Saeed*, Jean-Luc Dugelay*, Mohamed Jedra*[†]

* EURECOM Sophia Antipolis
2229 Route de Cretes, Sophia Antipolis, France.
{Usman.Saeed, Jean-Luc.Dugelay}@eurecom.fr
website: www.eurecom.fr

[†] Laboratoire Conception et Systèmes
Faculté des Sciences, Avenue Ibn Batouta, Rabat, Maroc.
{benaiss, jedra}@fsr.ac.ma
website: www.fsr.ac.ma

ABSTRACT

Biometric systems are prone to impostor attacks, which in case of high security applications could be critical. These attacks range from simple attacks consisting of stealing someone's password or complex as cryptographic attacks on the transmission channel. In this study we have focused on combating replay attacks, in which an impostor uses a pre-recorded image of the client. The system captures stereoscopic images of the face and then decides if the images belong to a real face or a poster from depth calculated using triangulation.

1. INTRODUCTION

Biometric systems have become a common place in our society these days. They are usually based on something you know or something you have, e.g. password, ID card, fingerprints, voice, face etc. Each biometric has some advantages or disadvantages associated with it. Some are easily accepted by users, others are unique. Biometric systems based on facial images [1] have the advantage that they are easily accepted by users and are non intrusive but lack the robustness of some other biometrics such as fingerprints.

Whatever unique characteristics biometrics may have, all have one problem in common, they are susceptible to attacks [2]. Attacks come in a variety of form and sites [3, 4], from attacks on the capturing device, transmission channel, data storage or even the recognition system itself. The most common attacks take place on the capture device and are usually of three kinds. The first are called coercive attacks, which occur when a person physically forces a client to attack the system by placing him in front of the sensor, the second are impersonation attacks, where a person changes his appearance to match a client. The third are replay attacks where an impostor uses a pre-recorded image of a client to deceive the system.

In this paper we present an initial system based on stereoscopic images that tries to avert replay attacks in a face recognition scenario. Two images of the face are taken by two identical webcams, next face is detected in one of the images and matched with the second image, then depth is estimated using triangulation of stereo images and finally a decision is made whether the images belong to a client or impostor based on depth map.

2. PROPOSED METHOD

In this section we will first describe the experimental setup of the system and then proceed to an algorithmic description of the system that enables us to differentiate between an impostor and a client.

2.1 System Setup

Our system (c.f. Figure 1) consists of two webcams, each associated with a coordinate system, $R_r(O_r, X_r, Y_r, Z_r)$ for the right webcam and $R_l(O_l, X_l, Y_l, Z_l)$ for the left webcam. Both the webcams have a focal length of 3.85 mm, image resolution of 640 X 480 pixels at 30 f/s. The optical axis of both the cameras must be parallel and the distance between the two cameras (baseline) is fixed at 7 cm.

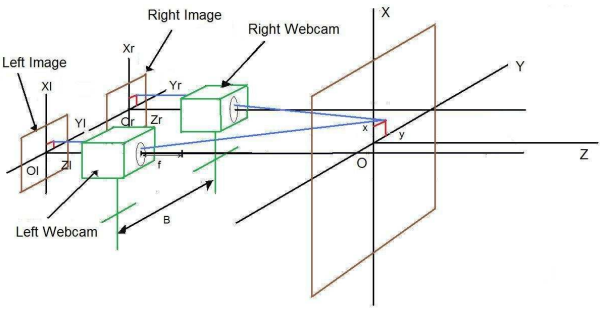

Figure 1 – System Setup.

2.2 Detection Algorithms.

The impostor detection algorithm can be best explained as a series of several image processing steps starting with acquisition, detection, registration and finishing with geometric calculations. Figure 2. gives a description of the steps involved.

2.2.1 Acquisition

Images are acquired from the left and right webcams simultaneously. The webcams must be aligned with parallel optical axis and the face roughly in the middle of the frame, occupying at least half of the frame.

2.2.2 Face Detection

The face detector module is based on cascade of boosted classifiers approach proposed by [5]. Instead of working with direct pixel values this classifier works with a representation called "Integral Image", created using Haar-like features. The advantage of which is that they can be computed at any scale or location in constant time. The learning algorithm is based on AdaBoost, which can efficiently select a small number of critical visual features from a larger set, thus increasing performance considerably.

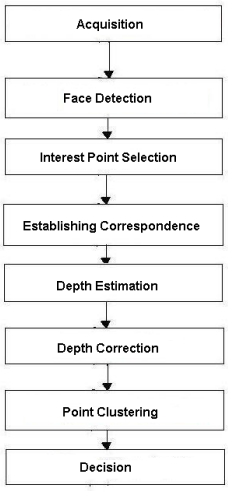

Figure 2– Steps of Detection Algorithm.

The classifier has been trained with facial feature data provided along the Intel OpenCV library [6]. Once the face is detected, as shown in figure 3. the background is removed.

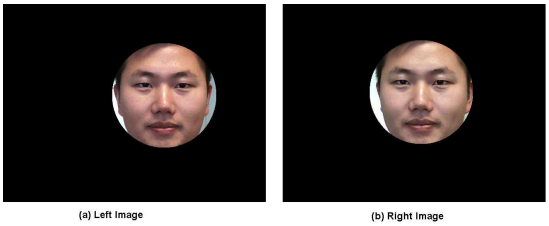

Figure 3– Face Detection

2.2.3 Interest Point Selection

The next step involves the selection of prominent points (refer figure 4.) within the region of the image where the face has been detected. We have applied the Harris corner and edge detector [7] to find such points. The Harris operator is based on the local auto-correlation function.

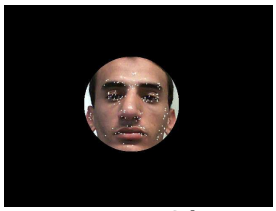

Figure 4– Point Selection

2.2.4 Establishing Correspondence

This step consists of finding points selected in the second image. The algorithm proposed by Bouguet [8] is based on the Lucas Kanade [9] optical flow with a pyramidal extension, which enable us to calculate the optical flow at several hierarchical intervals.

2.2.5 Depth Estimation

After establishing the correspondence between the left and the right image, the depth can be calculated by basic triangulation.

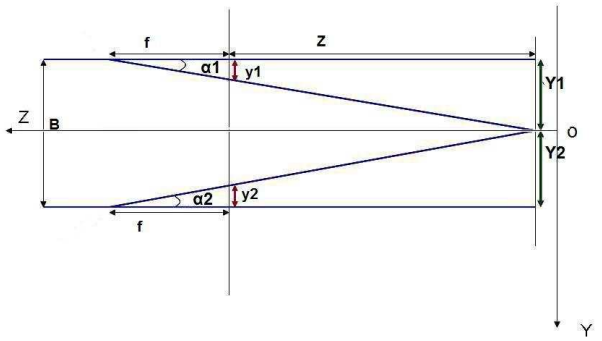

Figure 5– Triangulation

From figure 5. triangulation can be formulated as

$$\tan(\alpha_1) = \frac{y_1}{f} = \frac{Y_1}{Z+f} \cong \frac{Y_1}{Z}$$

$$\tan(\alpha_2) = \frac{y_2}{f} = \frac{Y_2}{Z+f} \cong \frac{Y_2}{Z}$$

$$y_2 - y_1 = \frac{f}{Z}(Y_2 - Y_1)$$

where,

$$y_2 - y_1 = d$$

$$Y_2 - Y_1 = B$$

then,

$$Z(d) = \frac{fB}{d}$$

As it is evident from the above equations, depth can be calculated using three values, first is the disparity *d*, which is the distance between corresponding point in the left and right image, the second is the focal length *f*, the third is the baseline *B* (the distance between the optical centers of the two cameras).

2.2.6 Depth Correction

Some errors were observed in the previous step of depth calculation and it was judged imperative that they must be handled in a generalized manner, thus the following clipping was applied to all the points based on μ , the average depth of all the points.

$$Z = \begin{cases} \mu \text{ if } \mu - \beta \le Z \le \mu + \beta \\ Z \text{ otherwise} \end{cases}$$

Where β is calculated empirically and will be explained in the experiment section.

2.2.7 Point Clustering

This step is specifically applied to reduce the number of calculations in the subsequent steps. It consists of clustering the points based on the depth value into *n* groups. The number of groups *n* was experimentally estimated and will be explained in the experiment section. Next the centeroid for each group was calculated.

2.2.8 Decision

The decision regarding the authenticity of the person in front of the camera is based on the fact that in case of a picture the depth map would resemble that of a plane surface. From the *n* centroids calculated above we select 3 points A, B, C and a fourth point as M, then we compute the vectors AB, AC and AM. Now if the determinant of these vectors is equal to zero we can establish that they indeed belong to the same plane. We repeat this process with all the centroids and then calculate the average. The algorithm is defined as:

Sum: variable containing the sum of determinants, initialized to zero
Counter: variable for number of determinants, initialized to zero.
B: a table of size 3 x n for centroids
D: a table of size 3 x 3 for vectors
for i ←1 to (n-3)
{
$D_1 = B_{i+1} - B_i$
$D_2 = B_{i+2} - B_i$

for j←1 to n
{
if j > i+3 or j < i
{
$D_3 = B_j - B_i$
sum = sum + det(D)
counter = counter+1
}
}
}

The average then is taken as

$$average = \frac{\sum_{i=1}^{C_n^4} \sum_{j \neq i}^{n-3} \sum_{j \neq i+1 \; j \neq i+2} |(B_{i+1} - B_i)(B_{i+2} - B_i)(B_j - B_i)|}{C_n^4}$$

The final decision is taken based on a threshold *T*.

$$Decision = \begin{cases} imposter & if \quad average < T \\ client & otherwise \end{cases}$$

3. EXPERIMENTS AND RESULTS

3.1. Database

Experiments were carried out on an internal database of 12 people, due to the fact that no standard database is available for testing replay attacks. The database consists of 2 images per person, left and right image; giving 24 images for the client database. The impostor database is constructed by using an image of the client's face and varying the angle of image in front of the system. The first is perpendicular; the second is inclined towards the left and the other towards the right. This gives a total of 3 x 12 x 2, i.e. 72 images for the impostor database.

(a)

(b)

Figure 6– Database (a) Clients (b) Impostors

3.2. Estimation of β

The parameter β described above is estimated from the database. It is calculated as,

$$\beta = \max_j \sqrt{\sum_{i=1}^{N} (Z_i - \mu_j)}$$

Where, μ is the average depth of all the points on an impostor image *j* and Z_i is the depth value of a point *i*.

3.3. Estimation of n

The value of *n*, which is the number of groups, is estimated by calculating the EER at different values of n. The results obtained by varying the value of n from 4 to 38 are depicted in the figure. 7

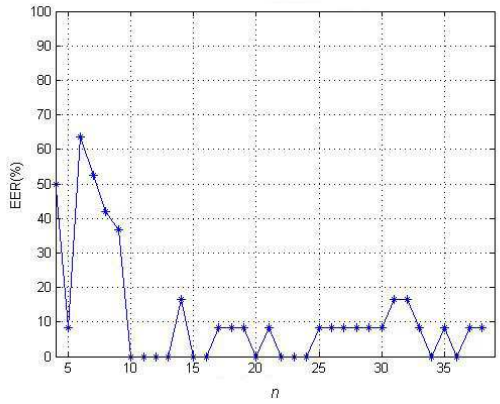

Figure 7– EER with varying *n*

3.4. Estimation of T

To estimate the value of T we compute T-FA / T-FR which are T dependent FAR and RRR, keeping n = 10 and varying T from 0 to 700.

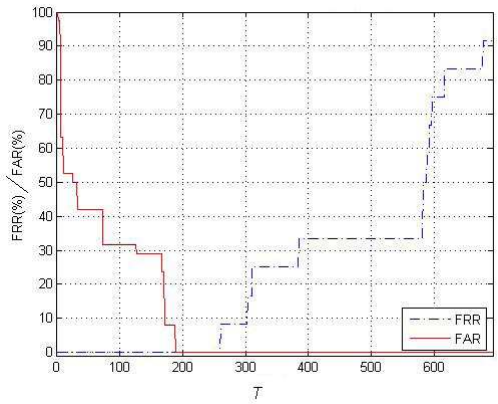

Figure 8– T-FA / T-FR for *n= 10.*

As it is evident the value is minimal between 189 and 260, so we select the threshold T as average of these values, i.e. 224.5.

3.5. Results

Once all the parameters have been fixed, we proceed to calculate the FA/FR for the entire database, which results in an EER almost equal to 0. Although this validates our assertion but we must highlight here the fact that the size of the database is quite limited.

4. CONCLUSIONS

In this paper we have presented an imposter detection system using stereoscopic images of the face. Two images of the face are taken and depth is estimated using triangulation, next this depth map is analyzed to verify whether the image belongs to a real client or is an imposter.

Significant amelioration can be made to face recognition systems by using our imposter detection as a pre-filtering method. Another related study could be to use images at different zoom levels obtained by a single camera instead of stereo images for imposter detection.

REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *In ACM Comp. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] I. Buhan, P. Hartel, "The State of the art in Abuse of biometrics", *Technical Report TR-CTIT-05-41 Centre for Telematics and Information Technology*, University of Twente, Enschede, ISSN 1381-3625, 2005.

[3] U. Uludag and Anil K. Jain, "Attacks on biometric systems: a case study in fingerprints", *In proc. SPIE*, vol. 5306, pp.62-633, 2004.

[4] J-F. Bonastre, D. Matrouf, C. Fredouille, "Effect of voice transformation on impostor acceptance", *In Proc. Of ICASSP*, 2006.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *In Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 511-518, 2001.

[6] http://www.intel.com/technology/computing/opencv/

[7] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *In Proceedings of 4th Alvey Vision Conference*, pp.147-151, 1988.

[8] J-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker : Description of the algorithm," *In Intel Research Laboratory, Technical Report*, 1999.

[9] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *In Proceedings of DARPA Image Understanding Workshop*, pp. 121–130, 1981.