

Comparative Analysis of the Variability of Facial Landmarks for Forensics Using CCTV Images

Ruben Vera-Rodriguez, Pedro Tome, Julian Fierrez, and Javier Ortega-Garcia

Biometric Recognition Group - ATVS, Escuela Politecnica Superior
Universidad Autonoma de Madrid, Avda. Francisco Tomas y Valiente, 11
Campus de Cantoblanco - 28049 Madrid, Spain
{ruben.vera,pedro.tome,julian.fierrez,javier.ortega}@uam.es

Abstract. This paper reports a study of the variability of facial landmarks in a forensic scenario using images acquired from CCTV images. This type of images presents a very low quality and a large range of variability factors such as differences in pose, expressions, occlusions, etc. Apart from this, the variability of facial landmarks is affected by the precision in which the landmarks are tagged. This process can be done manually or automatically depending on the application (e.g., forensics or automatic face recognition, respectively). This study is carried out comparing both manual and automatic procedures, and also 3 distances between the camera and the subjects. Results show that landmarks located in the outer part of the face (highest end of the head, ears and chin) present a higher level of variability compared to the landmarks located in the inner face (eye region, and nose). This study shows that the landmark variability increases with the distance between subject and camera, and also the results of the manual and automatic approaches are similar for the inner facial landmarks.

Keywords: Forensics, face recognition, video surveillance, data analysis.

1 Introduction

Automatic face recognition over forensic caseworks is still a challenge for the research community. Large amounts of research are being carried out trying to compensate variability sources (such as illumination, pose, facial expressions, occlusions, etc.) that affect significantly reducing the performance of the face recognition systems. In a forensic scenario, these variability factors are crucial, because forensic examiners have to frequently deal with face images extracted from CCTV cameras and other low quality sources, which make the task really difficult.

Many different techniques have been developed to automatically tag facial landmarks on a face [1–5]. These techniques achieve good results over good quality and frontal faces, but are still not that good for the cases of having high variability and low quality images. On the other hand, humans are subjective

and do not work as systematically as computers. For this reason, in practice forensic examiners make use of semiautomatic systems, which can help in the suspects identification tasks [6].

Among the tasks carried out by forensic examiners, they analyse the intra-variability of two face images, a set of gallery images (with known identity) and the probe image. In an anthropometric analysis they extract manually a set of facial landmarks, then compute some distances between them, which can be used as features in their analysis. Figure 1 shows a diagram of this procedure.

This paper focuses on analysing the variability of facial landmarks in a forensic scenario over a database of face images acquired from CCTV images. This landmarking variability is affected by two factors, on the one hand the accuracy of the process of landmark tagging, which can be done manually or automatically and can vary significantly due to the quality of the images, and on the other hand it is also affected by the intrinsic variation of the landmarks, due to changes in pose, expression or occlusions among others.

In this paper, we carry out the study using SCface database, which is comprised of CCTV images at three different distances (1, 2.6 and 4.20 meters) between the camera and the persons. We analyse both the effect of the distance between the subject and the camera in the landmark variability, and we also compare the variability of an automatic system compared to a manual landmark tagging imitating the work of a forensic examiner. Some of the findings of this study are that in general facial landmarks located in the outer part of the face (highest point on the head, ears and chin) have a high level of variability, due possibly to hair occlusions. Regarding the distances between the camera and the persons, the variability increases gradually with the distance. Surprisingly, very similar results are achieved for both manual and automatic approaches, although not all the landmark points were able to be tagged by the automatic system. The findings of this paper could be included in the work carried out by a forensic examiner within a anthropometrical facial analysis.

The remainder of the paper is organized as follows. Section 2 describes the data used in the experimental work of this paper, comprised of 130 persons and 3 distances between camera and subjects. Section 3 describes the task of landmark tagging and the image processing. Section 4 describes the experimental results achieved and finally Section 5 draws the final conclusions and future work.

2 SCface Database

This section describes the subset of the SCface database [7] used in our experiments. SCface is a database of static images of human faces with 4.160 images (in visible and infra-red spectrum) of 130 subjects.

The dataset used in this paper is divided into 6 different subsets: *i*) mugshot images, which are high resolution frontal images, and *ii*) five visible video surveillance cameras. Each of these subsets contains 130 images, one per subject. The images were acquired in an uncontrolled indoor environment with the persons walking towards several video surveillance cameras having different qualities.

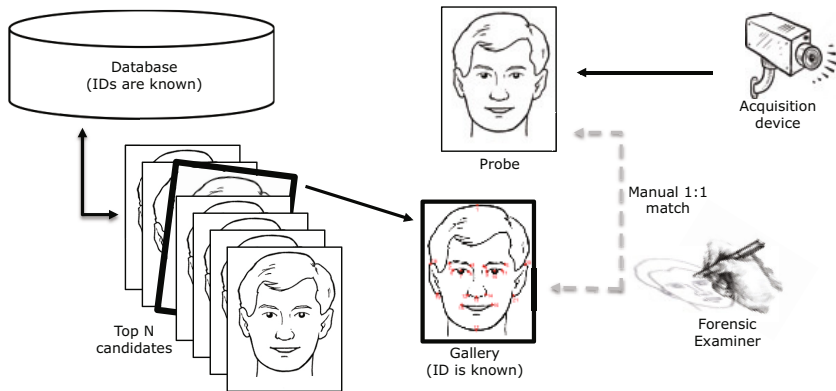


Fig. 1. General procedure followed by a forensic examiner to compare two face images

Additionally the images were acquired at three different distances: 1.00 meters (Close), 2.60m (Medium) and 4.20m (Far) respectively while the subject walked towards the cameras. Fig. 2 shows an example of a mugshot image, and the images acquired by one of the surveillance cameras. As can be seen there is a considerable scenario variation in terms of quality, pose and illumination. The effect of the pose is specially important due to the different angles between the person and the cameras.

This database is of particular interest from a forensic point of view because images were acquired using commercially available surveillance equipment, under realistic conditions. One of its drawbacks is that it is just comprised of one mugshot session, so it is not possible to study the landmark variability for the mugshot images, as several pictures per person are needed. For this study, we use the 5 available images per person and per distance to analyse the variability of the facial landmarks (1950 images in total, 3 distances \times 5 cameras \times 130 persons). Also, we carry out this study both in a manual way imitating the work that a forensic examiner would perform, and using an automatic system to detect the facial landmarks.

3 Facial Landmark Tagging and Image Processing

This section describes the process of facial landmark tagging and image processing in order to analyse the variability of facial landmarks.

The first step after database collection was to define a set of facial landmarks to include in this study. A set of 21 facial landmarks was defined following recommendations from the Spanish Guardia Civil [8], Netherland Forensic Institute [9] and ENFSI [10], including the irises (2 landmarks), inner and outer eye corners (4), eyebrow ends (4), mouth corners (2), nose corners (2), center of the nose (1), chin (1), upper and lower ears ends (4) and highest point on the head (1). Figure 3 shows the 21 facial landmarks considered in this study.

SCface database

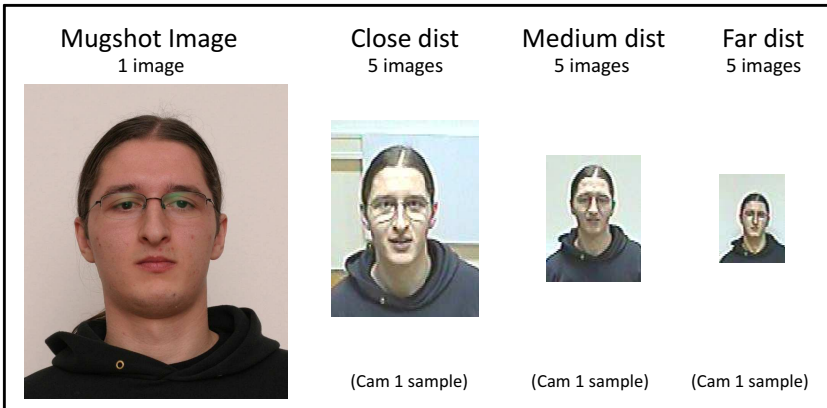


Fig. 2. Example of images of SCface database. High quality mugshot image, and 3 CCTV images acquired at three distances: close, medium and far, for one of the five CCTV cameras.

The process of facial landmark tagging was carried out both manually and automatically. The manual landmark tagging was carried out by the same person, imitating the work of a forensic examiner. In this case the set of 21 landmarks was tagged for the whole database. On the other hand, it is interesting to compare this experimental work with an automatic system. For this case, Luxand FaceSDK [5] was used, which is a high performance face recognition commercial software based on facial landmarks features. In this case, a set of 13 facial landmarks (in red in Fig. 3) were considered, as the automatic system was not able to locate most of the other 8 remaining landmarks.

A second stage of image processing was carried out in order to normalise the facial images to the same size and position. Thus, the midpoint between the eye corners (midpoint between points 6 and 8, and midpoint between 9 and 11) was computed and used instead of the irises positions to align the faces, because the position of the irises can vary if the person does not look at the camera directly. The positions of these two points were fixed having 75 horizontal pixel between them following the recommendation from the ISO standard [11]. Therefore, translation, rotation and scaling of the original images was carried out to normalize the database. This was done in the same way for images collected at different distances between the camera and person. Figure 4 shows an example of the three CCTV face images shown in Figure 2 but size normalised, and showing the positions of the 21 facial landmarks in red and the positions of the center of the eyes in green. As can be seen, this is a challenging scenario for both manual and automatic landmark tagging due to the low quality of the images to analyze.

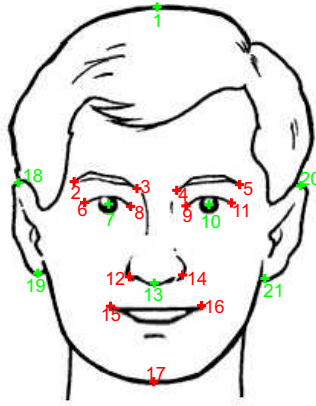


Fig. 3. Set of 21 facial landmarks defined. In red are the landmarks considered for automatic tagging. The manual process considers the whole set of 21 landmarks.

4 Experimental Results

This section describes the experimental work carried out to analyse the variability of facial landmarks considering a forensic scenario using images from CCTV cameras. Three different experiments were designed and are described here.

4.1 Person Specific Landmarking Variability

In this experiment a person specific landmarking variability (LV) was studied. Thus, the 5 available facial images per person and per distance were considered. The mean and standard deviation for each facial landmark were computed for the two (x,y) spatial dimensions ($\sigma_{x,i}$, $\sigma_{y,i}$, with $i=1, \dots, 13$ or 21 depending on the automatic or manual landmark tagging process), assuming following a gaussian distribution. Figure 5 shows two example face images superimposing for each facial landmark the result of tagging the 5 available images. An ellipse around each facial landmark is computed using as the radii $2\sigma_{x,i}$, $2\sigma_{y,i}$. Throughout this paper the variability of the different facial landmarks was computed as $\pm 2 \times \text{mean}(\sigma_{x,i}, \sigma_{y,i})$, covering this way a 95.44% of the hypothesized gaussian distribution. For example, in the image shown in Figure 5(a) the landmark for the highest point on the head shows a variability of ± 25.4 pixels considering a normalization of the face images with 75 horizontal pixels between the eye positions.

This procedure was followed for the 130 persons comprising the database, and it was found that the variability of the facial landmarks, specially for the outer ones varies significantly from person to person. The variability of these landmarks on the outer part of the face (highest point on the head, chin and ears) is very dependent on hair occlusions, more frequent in women than men for the population considered.

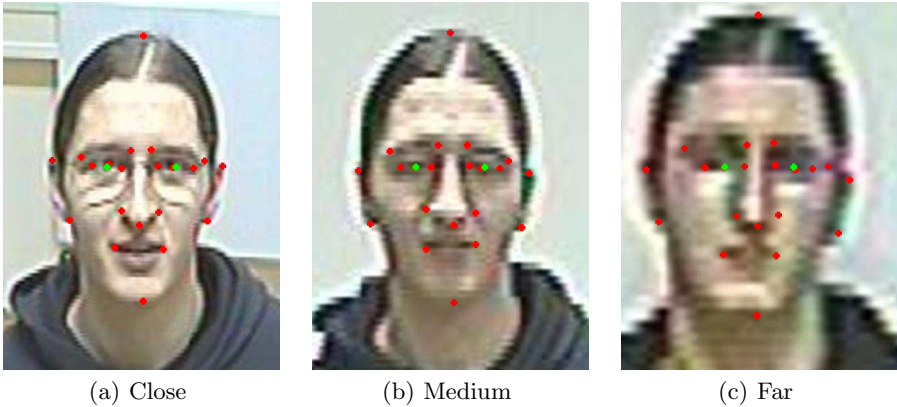


Fig. 4. Same example as in Fig. 2 (only for the CCTV images) but normalizing the faces with 75 pixels between the center of the eyes. Also, the 21 manual facial landmarks are shown (red), plus the center of the eyes (green).

4.2 Distance Specific Landmarking Variability

This section reports the experimental results achieved for the global landmarking variability (i.e., an average of the individual results) considering the effect of the 3 distances between the camera and the persons. In order to compute a global landmarking variability (LV), the mean of the different individual values of the variability of each facial landmark is computed, following the equation:

$$LV_i = \frac{1}{N} \sum_{j=1}^N (\sigma_{x,i,j} + \sigma_{y,i,j}) \quad (1)$$

where $i=1, \dots, 13$ or 21 (for automatic or manual tagging respectively) are the landmarks and $j=1, \dots, N$, being N the maxima number of persons in the database, 130 in this case. This procedure is followed for each of the three distances considered.

Figures 6(a,c,e) show the results achieved for the case of manual landmark tagging, and Figures 6(b,d,f) show the same but for the case of automatic landmark tagging. Here we focus on the analysis of the distance for the manual case, as next section compares the case of manual vs. automatic landmark tagging.

As can be seen, there is a clear increment of the landmark variability regarding the acquisition distance between the subject and the camera for all the facial landmarks considered (except for the ears). The outer facial landmarks (highest point on the head, ears and chin) present the highest variability, then the mouth and nose areas, and the parts with the least variability are the eyes and eyebrows. It is worth noting that the normalisation of the faces was done using the center of the eyes, so it is also natural that these parts present less variability than the rest. It is also worth noting that as the landmark tagging is carried out over the original size images, close images have a bigger size compared to face images, as

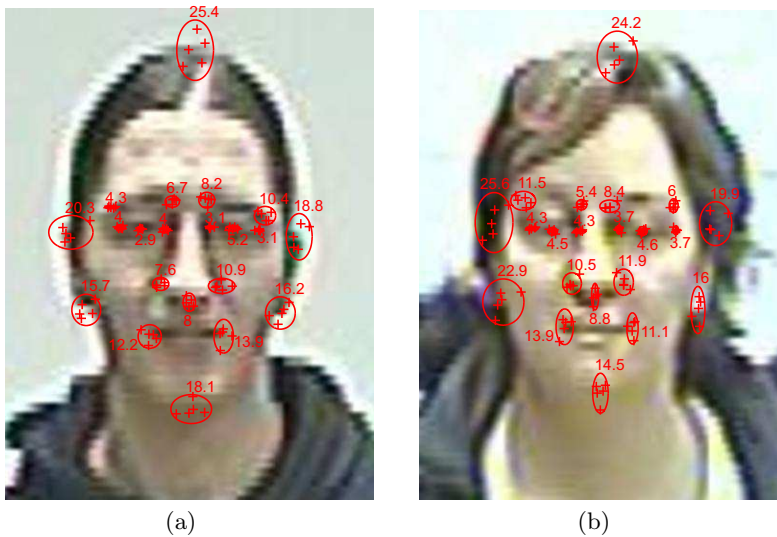


Fig. 5. Examples of the manual landmarking variability for two persons present in the SCface database for images taken at 2.60 meters distance between the person and the camera.

can be seen in Figure 2. Therefore, the process of landmark tagging can be done with more precision for the close images and therefore reducing the landmark variability.

4.3 Manual vs. Automatic Landmarking Variability

Figure 6 show the landmark variability for both manual (a, c, e) and automatic (b, d, f) procedures. The number of facial landmarks tagged is different in both cases, 21 for manual and 13 for automatic tagging, as described above. The results show that the landmark variability is very similar for the set of common landmarks. Specifically, for the close images, the landmarks located in the ocular region present lower variability for the automatic system compared to the manual case, while the landmarks located in the mouth region present a higher variability for the automatic system.

For the medium distance, both ocular and mouth region present in general a lower variability for the automatic system, but in the far distance where the quality of the images is very low the manual procedure achieves a lower landmark variability. It is also worth noting that the automatic system only considers 13 facial landmarks as it was not able to locate correctly most of the remaining 8 landmarks (mainly the outer ones), but in general it achieves better results than expected a priori.

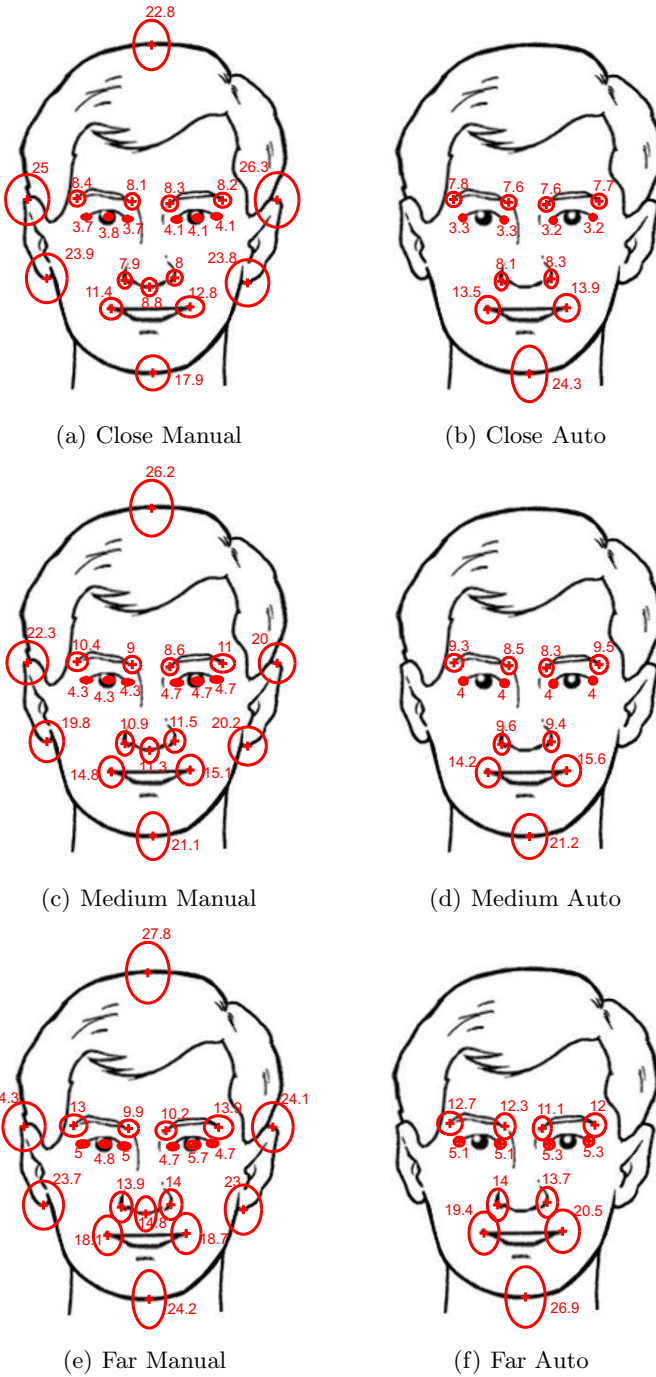


Fig. 6. Results of the landmarking variability for the three distances considered between the persons and the camera: far (3 meters), medium (2 meters) and close (1 meter).

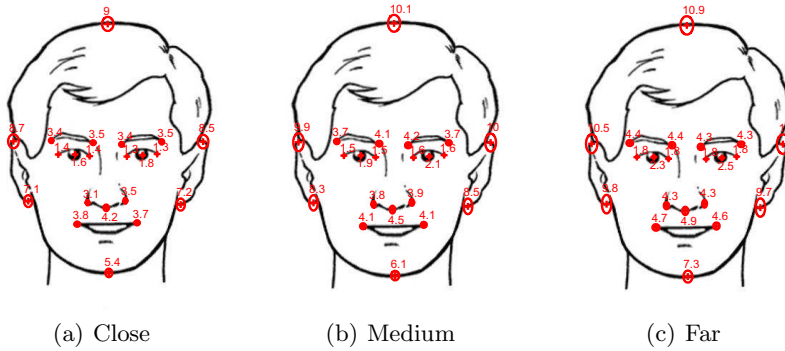


Fig. 7. Results of the landmarking variability for the three distances considered between the persons and the camera: close (1 meter), medium (2 meters) and far (3 meters), for a high quality database of mugshot images [12].

4.4 CCTV Images vs. Mugshot Images Landmarking Variability

For completeness, results achieved in this analysis can be compared with a previous similar work [12], using in this case a database of mugshot images acquired in a controlled scenario. In this case, the images were of a much higher quality as the images from SCface database considered here. Also, three distances between the subject and the camera were analysed (1, 2 and 3 meters). In this case, there were 8 images per distance and per person, for a total of 50 persons (1200 images in total).

Results achieved for this study using a manual landmark approach are shown in Figure 7. As can be seen, the landmark variability is much lower in all three distances compared to the results achieved in this paper. In this case, also landmarks from the inner parts of the face are more consistent than those from the outer parts. This significant difference of the variability is mainly due to the quality of the images considered, in this case of high quality and therefore achieving lower landmark variability. It is also worth mentioning that SCface database was acquired in uncontrolled conditions while the other database was acquired in a controlled scenario.

5 Conclusions

This paper reports a study of the variability of facial landmarks over a CCTV database with low quality images and large range of variability factors. Face images are taken with three distances between the persons and the 5 CCTV cameras (1, 2.60 and 4.20 meters). 21 facial landmarks were defined and the database was manually tagged imitating the procedure followed by a forensic examiner. Also, an automatic system was used to tag 13 out of the 21 landmarks defined. The main conclusions are that the landmarks located in the outer part of the face have a much higher variability compared to the landmarks placed

near the eyes. A reason for this is mainly that these are the areas which can have hair occlusions more frequently like the highest point on the head and the ears.

Regarding the distances between the camera and the person, the landmarking variability increases with the distance. Comparing the two manual and automatic tagging approaches, the results show that the landmark variability is very similar for the set of common landmarks, having in some cases lower variability for the automatic system.

A final comparison of these results with a previous study over a controlled database of mugshot images, shows that the CCTV images present a significantly higher landmark variability, which is mainly due to lower quality of the images making very difficult to tag the facial landmarks with high precision.

Acknowledgments. R. Vera is supported by a Juan de la Cierva Fellowship from the Spanish MINECO. This work has been partially supported by a contract with Spanish Guardia Civil and projects BBfor2 (FP7-ITN-238803), Bio-Shield (TEC2012-34881), Contexts (S2009/TIC-1485), TeraSense (CSD2008-00068) and “Catedra UAM-Telefonica”.

References

1. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3D Face Recognition. *Int. Journal of Computer Vision* 90(3), 331–349 (2010)
2. Arca, S., Campadelli, P., Lanzarotti, R.: An Efficient Method to Detect Facial Fiducial Points for Face Recognition. In: *Proc. of International Conference on Pattern Recognition (ICPR 2004)*, vol. 1, pp. 532–535 (2004)
3. Arca, S., Campadelli, P., Lanzarotti, R.: A Face Recognition System based on Automatically determined Facial Fiducial Points. *Pattern Recognition* 39(3), 432–443 (2006)
4. Beumer, G.M., Tao, Q., Bazen, A.M., Veldhuis, R.N.J.: A Landmark Paper in Face Recognition. In: *Proc. of Int. Conference on Automatic Face and Gesture Recognition (FG 2006)*, USA, pp. 73–78 (2006)
5. Luxand FaceSDK, <http://www.luxand.com/facesdk/>
6. Jain, A.K., Klare, B., Park, U.: Face Matching and Retrieval in Forensics Applications. *IEEE Multimedia* 19(1), 20 (2012)
7. Grgic, M., Delac, K., Grgic, S.: Sface - surveillance cameras face database. *Multimedia Tools Appl.* 51(3), 863–879 (2011)
8. Spanish Guardia Civil, <http://www.guardiacivil.es/>
9. Netherlands Forensic Institute (NFI), <http://www.forensicinstitute.nl>
10. European Network of Forensic Science Institutes, <http://enfsi.eu/>
11. ISO/IEC JTC 1/SC 37 N 504, Biometric Data Interchange Formats Part 5: Face Image (2004)
12. Vera-Rodriguez, R., Tome, P., Fierrez, J., Exposito, N., Vega, F.J.: Analysis of the Variability of Facial Landmarks in a Forensic Scenario. In: *Proc. of International Workshop on Biometrics and Forensics (IWBF 2013)*, pp. 1–4 (2013)