MULTI-FORESEE

COST Action CA16101

MULTi-modal imaging of FOREnsic SciEnce Evidence
Tools for Forensic Science

Final Report of 2017

Chair of the Action:	Dr. Simona Francese (UK)
Vice Chair of the Action:	Prof. Massimo Tistarelli (IT)
Working Group 2 Leader:	Dr. Alesandro Trivilini (CH)

Report of WG2’s activities of 2017

The following document contains all information regarding the WG2’s activities , meeting and workshops, and results of the year 2017.

Basing on the main objectives of the Action, the WG2 started discuss and working to identify the main topics to consider for the Round Robin Study.

During the first meeting we had in Porto on June 2017, WG2 discussed about all considerable identification of the imaging analysis topic and techniques:

A. **<u>Facial recognition / facial detection:</u>**
   From end-user point of view the discussion identified several issues, such as: low quality of the images; camera position (too high, resulting in bad face images); CCTV video too many codes, not standard format; increasing number of cameras, resulting in an increasing amount of data to be analyzed and stored.

   From Academia point of view the discussion identified several issues: many research on the subject but not applied on real cases; automatic facial recognition: deep learning techniques problems (few data available; not real database for testing/ improving the techniques; difficult having access to police database due to privacy/ security issued; usability in court: experts need to interpret the results/ evidence; the results of manual facial recognition are more used in court (Poland/ Denmark)).

B. **<u>Fingerprints</u>**
   From end-user point of view the discussion identified the following issue: technique is used and works fine and quire fast (20 minutes to get a list of 15 possible candidates, then manual control) 67% of the case is correct (they are developing/implementing a new algorithm that will result correct in 80% of the cases) documentation and the input of the data in the database).

   From Academia point of view the discussion identified the following issue: the finger prints technique works in 99% of the cases. This is another gap between research and real world.

C. **<u>Iris recognition</u>**
   From end-user point of view the discussion identified the following issue: not used in UK. Again problem with the quality of the images.

   From Academia point of view the discussion identified the following issue: difficulty of the analysis.

D. **<u>Soft biometrics</u>** (gait analysis / general proportion /height/color of the hair / clothes)
From end-user point of view the discussion identified the following issue: used, but very weak. Gait analysis used in Denmark with the support of photogrammetric technique.

E. **<u>Handwriting</u>**

In general it could be useful in case of digital signature by considering: Multimodalities crime scene reconstruction; 3D documentation of the crime scene using laser scanner and photogrammetric (UK and Denmark); 3D models of the scene allow freeing of the scene. Used for drawing bullet trajectory; Location of the actual victim in the scene; 3D of crime scene is way to visualize all the info in one view.

Other information that could be added: blood pattern analysis. It has been proposed the idea of creating database of crime scene that could be use in the case of pedo-pornography: in such way the perpetrator and the room may be linked to more victims.

F. <u>Video enhancement</u>

The WG discussion agreed that this is at the base of many analysis, such as facial recognition.

In this matter, the WG2 decided to focus and planning the Round Robin Study on **Facial recognition**, and organized the preparation of the framework of real data necessary for performing the RRS. Dr. Dariusz Zubs (Poland) offered to provide data for the study ny the end of July 2017. He used its own data/video or he can extract data from a public database (it has been suggested a possible database: one million celebrities database).

In the same way, Dr. Giuseppe Amato (Italy) may find some other data.

After this, the RRS was performed by: Dr. Giuseppe Amato (Italy); Dr. Anastasios Tefas (Greece). All details about the two Face recognition analysis are available as annexes in this document.

The WG2, after the second meeting held in Krakovia on Oktober 2017 (see participants on table below), decided to keep attention on Face recognition analysis and starting to consider the problem of overlapping of fingerprints.

List of participants:

NAME	COUNTRY	ROLE	EMAIL
Alessandro Trivillini	Switzerland	Academia	Ok
Dariusz Zuba	Poland	End user - Institute of Forensic Research-	Ok
Aleksandra Karczmarek	Poland	End user	Dariusz’ colleagues
Jerzy Brzowski	Poland	End user	
Wojciech Czubak	Poland	End user	
Nikolous Passalis	Greece	Academia	passalis@csd.auth.gr
Florin Alexa	Romania	Academia	ok
Andres Udal	Estonia	Academia	Andres_udal@ttu.ee
Ivana OGNJANOVIC	Montenegro	Academia	ivana.ognjanovic.edu@gmail.com
Ramo SENDELJ	Montenegro	Academia	ramo.sendelj@gmail.com
Stefan Rodiger	Germany	Academia	
Joseph Vella	Malta	Academia	Joseph.g.vella@um.edu.mt

Gholamreza ANBARJAFARI	Estonia	Academia	Shb@ut.ee
Claudio Vairo	Italy	Academia	Ok
Kamal NASROLLAHI	Denmark	Academia	
Chiara Villa	Denmark	Academia	Ok

Table 1: List pf participants of 2 meeting in Krakovia

Lugano, Thursday, March 29, 2018

Dr. Alessandro Trivilini
WG2 Leader

Annex 1

WG2 RRS Report on Face Detection and Recognition

Nikolaos Passalis and Anastasios Tefas
Greece

Introduction

The main goal of this study was to examine the performance of existing face detection and recognition tools on real Closed-Circuit Television (CCTV) footage. More specifically, six short video sequences depicting two different persons were used to evaluate the performance of state-of-the-art face detection and recognition algorithms using an existing library, the Python "*face recognition*" library [1]. The experimental results suggest that face detection and recognition using such low quality and low resolution video is an especially challenging task that can be only performed accurately when a clear view of the person of interest can be acquired from the footage. Finally, we demonstrate that detecting the whole body of a person, instead of its face, can be performed more reliable. This highlights the potential of using person detection as a preprocessing step before face detection/recognition.

Methodology and Evaluation Protocol

The main aim of this study was to evaluate the performance of existing and readily available tools for face detection and recognition. To this end, an open source Python library, the "face recognition" library [1], was used. This library provides an easy to use interface to another lower level library, the well-known "*dlib*" library [2], that is capable of efficiently performing various machine learning tasks using state-of-the-art techniques.

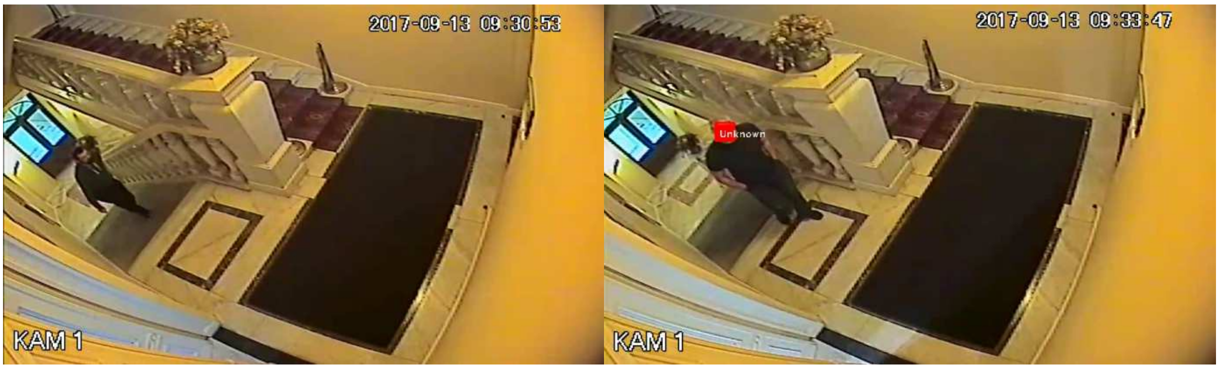
A state-of-the-art deep face detector provided by the dlib library was used for detecting the faces [2]. This detector is capable of achieving an impressive 99.38% detection accuracy on the Labeled Faces in the Wild benchmark [3]. For recognizing the identity of a person, a face encoding using the 68-point landmarking model of dlib was extracted and matched to the database of the known persons. The matching threshold was set to 0.5, i.e., a person was identified whenever the euclidean distance between the face encoding vectors extracted from a known person (stored in our database) and a detected person was smaller than 0.5. The 68-point landmarking model identifies various salient landmark points on the face, such as the location of eyes, mouth, nose, etc, that can be then used to recognize the identity of different people.

For the conducted experiments, the following evaluation protocol was used. First, one face image was chosen from each person using the supplied CCTV footage. This corresponds to the realistic scenario where a suspect has been identified by the end users in one frame and we are interested in recognizing him in other existing CCTV footage. The selected face images are shown below (Fig. 1a and 1b):

Then, the faces were detected and the face encodings were extracted from these images (database images) and used to identifying the persons in the rest of the supplied CCTV footage.

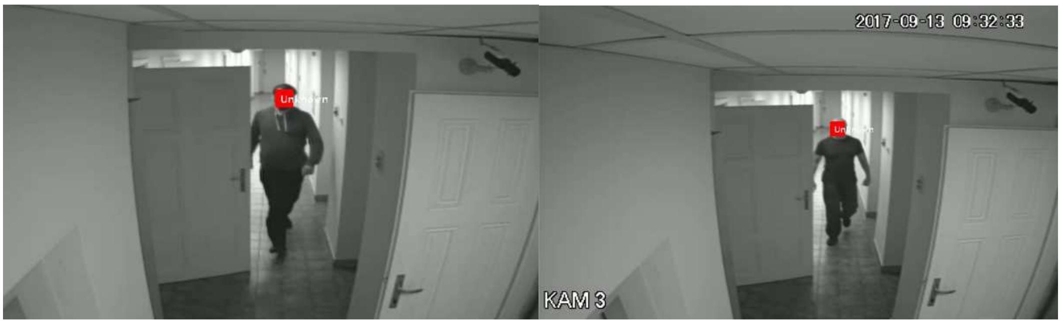
Evaluation Results

The first CCTV video was especially challenging since the size of the depicted faces was very small. This is demonstrated in some example frames of the first video that are shown below:

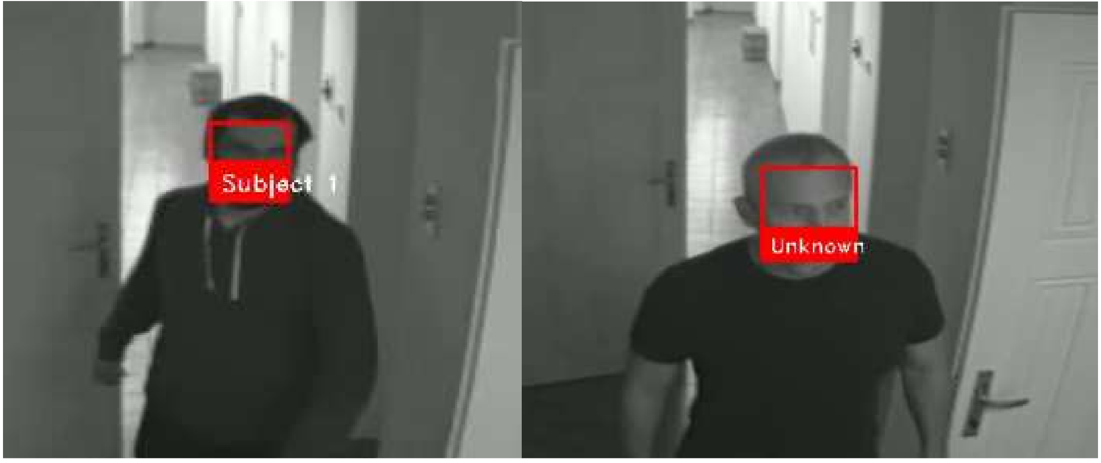


Detecting small objects (e.g., objects that are often smaller than 16 x 16 pixels) is a difficult problem that even state-of-the-art detectors are unable to tackle efficiently. Note that even when we were actually able to detect a face (right frame), we were unable to recognize the identity of the corresponding person, since it was not possible to accurately identify facial landmarks in such small face images.

In the second video, the face detections were more stable, since the depicted faces were larger. Some detection results are shown in the frames below:

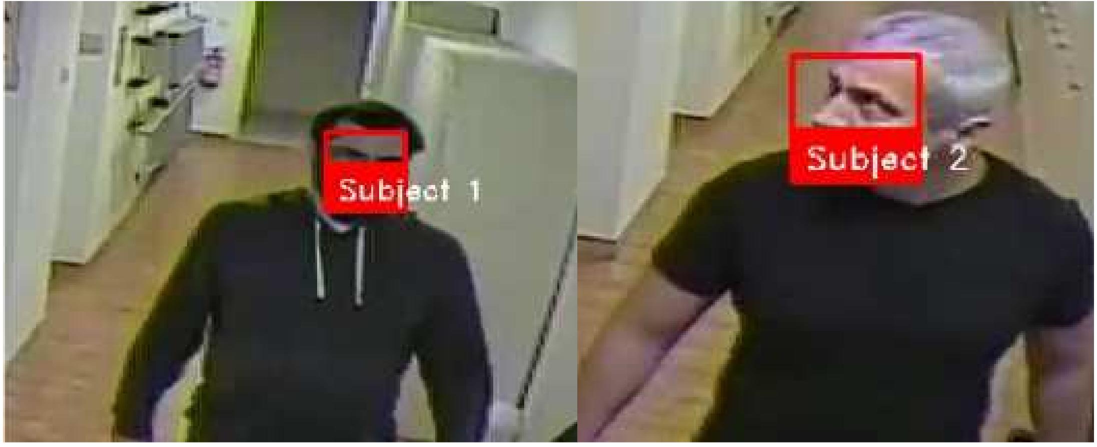


We were also able to identify the first person (left frame) when a larger face image was acquired:



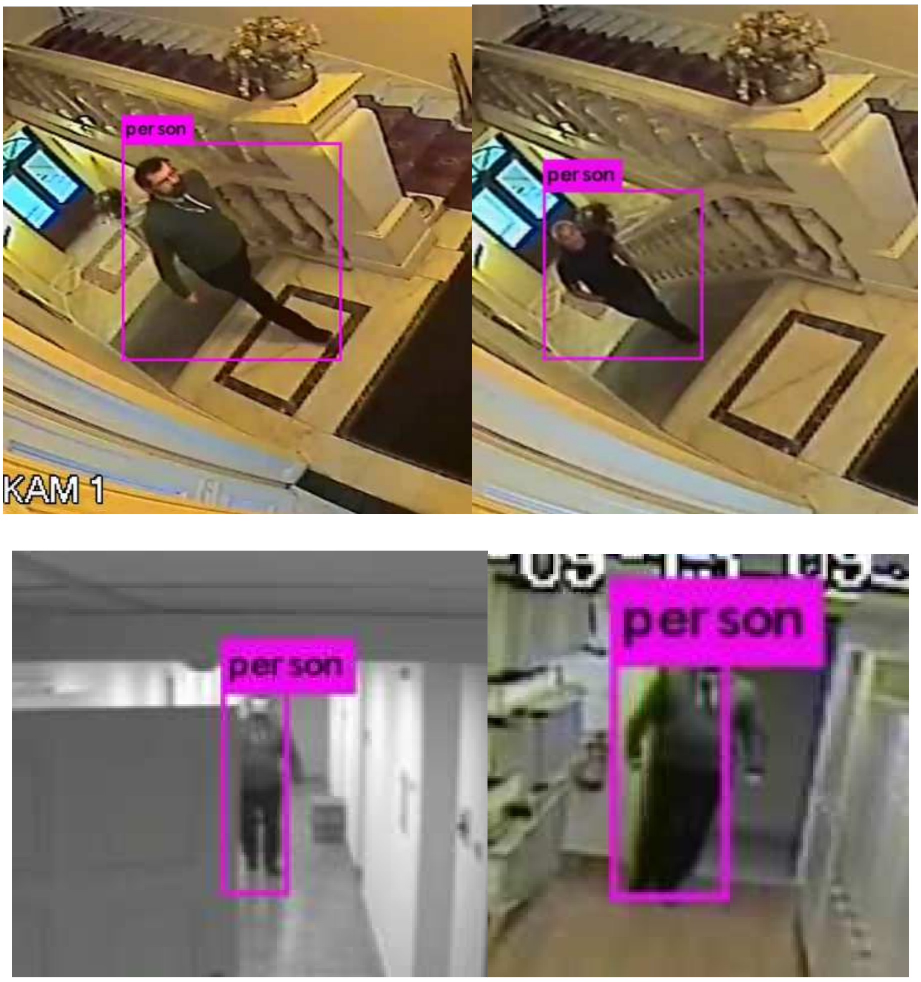
However, this was not possible for the second subject (right frame). It worths mentioning that the face image used to match the first subject (Fig. 1a) was extracted from this video, therefore there was smaller distribution shift compared to the second subject.

Finally, the results for the third video are shown below:



For this video, we were able to detect and accurately recognize both persons, even though the face image used for the first person (Fig. 1a) was substantially different from the face depicted in the frame above (left frame). For the second person (right frame), the face was matched only when the facial pose matched the original one (Fig. 2b) demonstrating the need for techniques that are able to align the detected face image and reconstruct a neutral face (centered and without any facial expression) before performing the actual recognition task. A technique that can be used to achieve this is deep autoencoders [4].

Since detecting faces in CCTV footage is especially difficult, we used an additional evaluation setup, where the whole body had to be detected. After detecting the whole body of a person, it is easier to identify where the face is (we expect that the face will be located in the upper half of the detected bounding box of a person). For the conducted experiments a state-of-the-art object detector was used, the YOLO detector [5]. The experimental results are shown below:



The persons were detected reliably even at smaller scale (since they are always significantly larger than a single face). This highlights the potential of using person detection as a preprocessing step before performing the task of face detection to acquire more accurate detection results.

Conclusions

In this report we evaluated the performance of existing tools for face detection and recognition using real CCTV footage. It was demonstrated that face detection and recognition from CCTV footage is especially hard due to the low quality of the videos and the small size of the depicted faces. Furthermore, the pose and the facial expression seem to significantly affect the recognition precision, suggesting that techniques that are able to reconstruct a neural and centered version of the face can improve the recognition accuracy. Finally, it was demonstrated that person detectors, that are able to reliably detect persons from CCTV footage, can be used as a preprocessing step before face detection to acquire more accurate results.

References

[1] Face recognition library, online resource available at https://github.com/ageitgey/face_recognition.

[2] King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*(Jul), 1755-1758.

[3] Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Vol. 1, No. 2, p. 3). Technical Report 07-49, University of Massachusetts, Amherst.

[4] Nousi, P., & Tefas, A. (2017, August). Discriminatively Trained Autoencoders for Fast and Accurate Face Recognition. In *International Conference on Engineering Applications of Neural Networks* (pp. 205-215).

Annex 2

WG2 RRS Report on Face Detection and Recognition

A Facial Landmarks Features

Giuseppe Amato and Claudio Vairo
Italy

Facial landmarks are key points along the shape of the detected face, that can be used as face features to improve face recognition, to align facial images, to distinguish males and females, to estimate the head pose, and so on.

Key points from landmarks are rarely used as representation of face verification tasks, typically *facial nodal points* are used instead. As nodal points, we can either use directly some of the facial landmarks or we can compute some new points starting from the facial landmarks. For example, the eyes, the nose, and the mouth are very representative parts of a person's face, so points relative to these parts of the face can be relevant to represent that face. In particular, for example, for the eyes, we can use the centroid of the eye instead of using the facial landmarks that constitute the contour of the eye.

In order to perform the face detection and to extract the facial landmarks from an image, we used the dlib library[2]. The dlib facial landmark detector is an implementation of the approach presented by Kazemi et al. in [15]. It returns an array of 68 points in form of (x,y) coordinates that map to facial structures of the face, as shown in Figure 1.

[2] http://dlib.net/

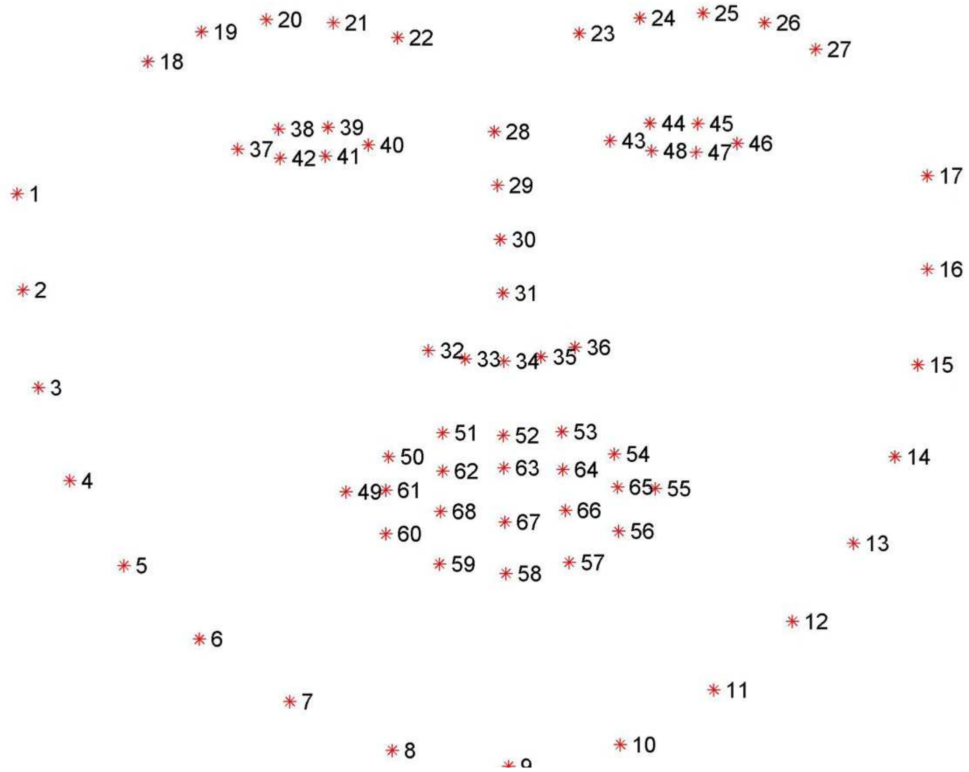


Figure 1. 68 facial landmarks.

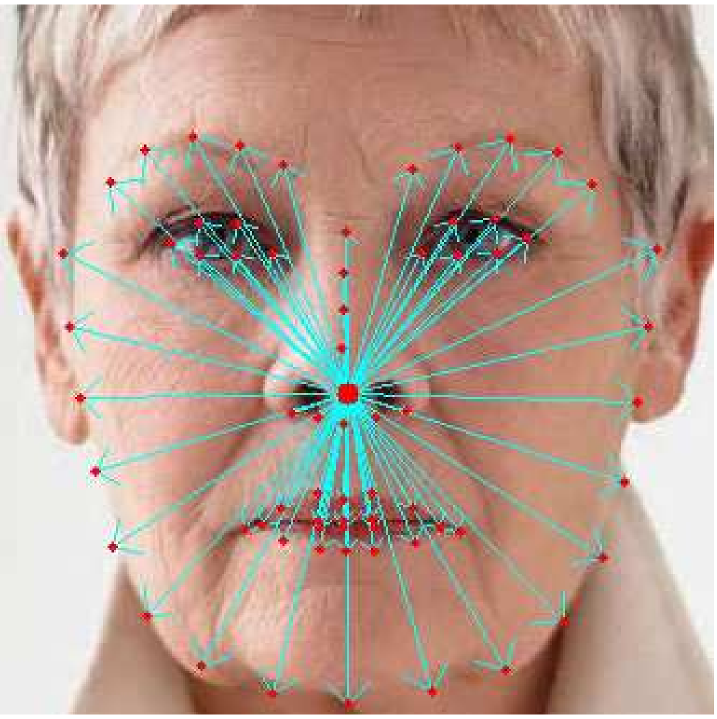


Figure 3. Distances from the centroid of the face to all 68 facial landmarks, used to build the 68-points features.

The distances between nodal points and facial landmarks can be used to build a feature of the face that can be compared with other faces features. In particular, we computed three features based on the distances between nodal points and facial landmarks: the *5-points* feature, the *68-points* feature and the *Pairs* feature. All the distances used to compute these features are normalized to the size of the bounding box of the face. In particular, each distance is divided for the diagonal of the bounding box.

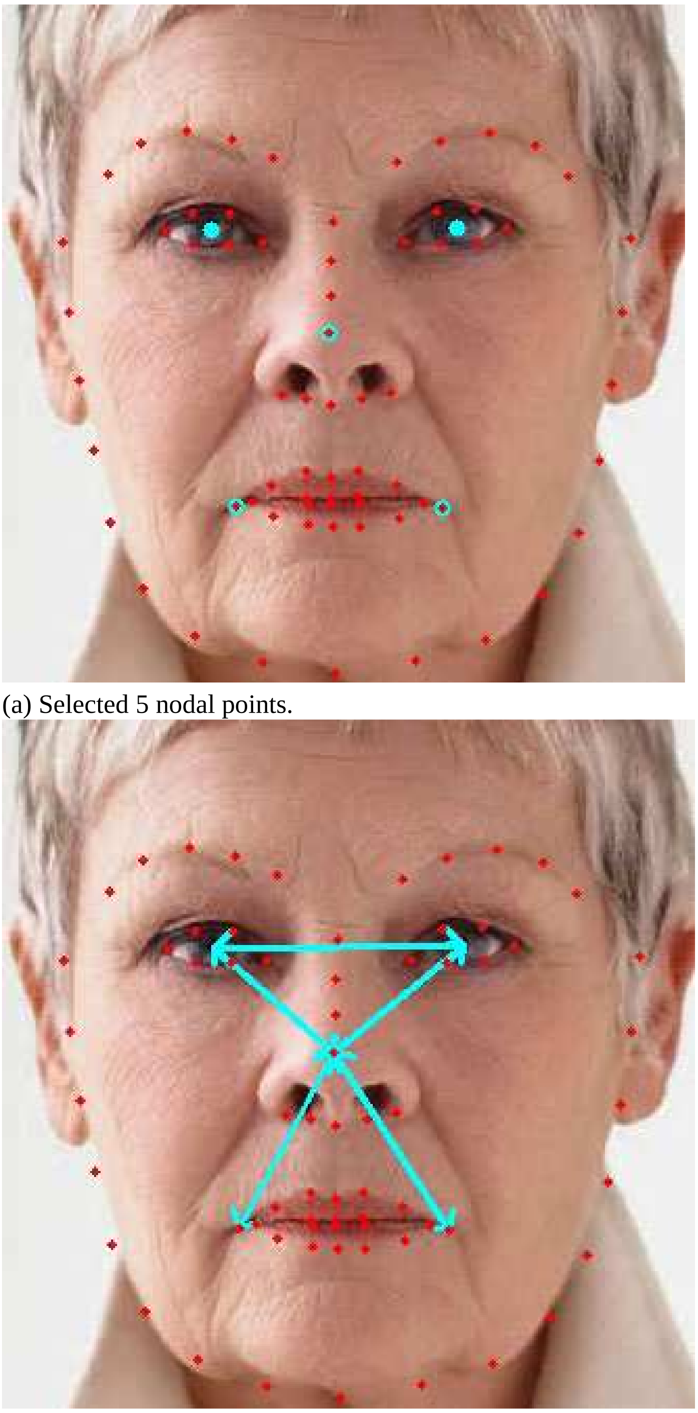
1) 5-points feature: In order to build the 5-points feature, we used five specific nodal points: the centroids of the two eyes, the center of the nose, and the sides of the mouth. The centroids of the two eyes are computed from the six facial landmarks for each eye returned by the dlib library. For the nodal points of the nose and of the mouth, instead, we used directly some of the facial landmarks, respectively the landmark #31 for the nose and the landmarks #49 and #55 for the sides of the mouth (see Figure 2(a)). We used these nodal points to compute the following 5 distances (see Figure 2(b)):

- left eye centroid - right eye centroid
- left eye centroid – nose
- right eye centroid - nose
- nose - left mouth
- nose - right mouth

This produces a 5-dimensional float vector that we used as 5-point feature of the face.

2) 68-points feature: For the 68-points feature, we computed the centroid of all the 68 facial landmarks returned by the dlib library and we computed the distance between this point and all the 68 facial landmarks (see Figure 3). This produces a 68-dimensional float vector that we used as 68-feature of the face.

3) Pairs feature: The pairs feature is obtained by computing the distance of all unique pairs of points taken from the 68 facial landmarks computed on the input face, as suggested in [9]. This produces a vector of 2,278 float distances that we used as Pairs feature of the face.



(a) Selected 5 nodal points.

(b) 5 nodal points distances.

Figure 2. Nodal points and distances used to build the 5-points features.

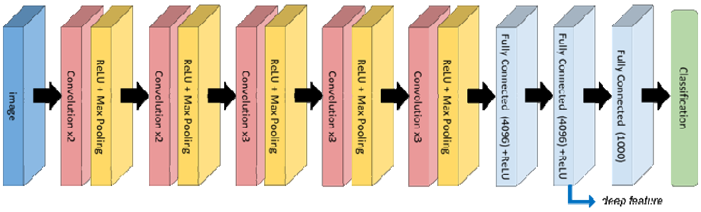
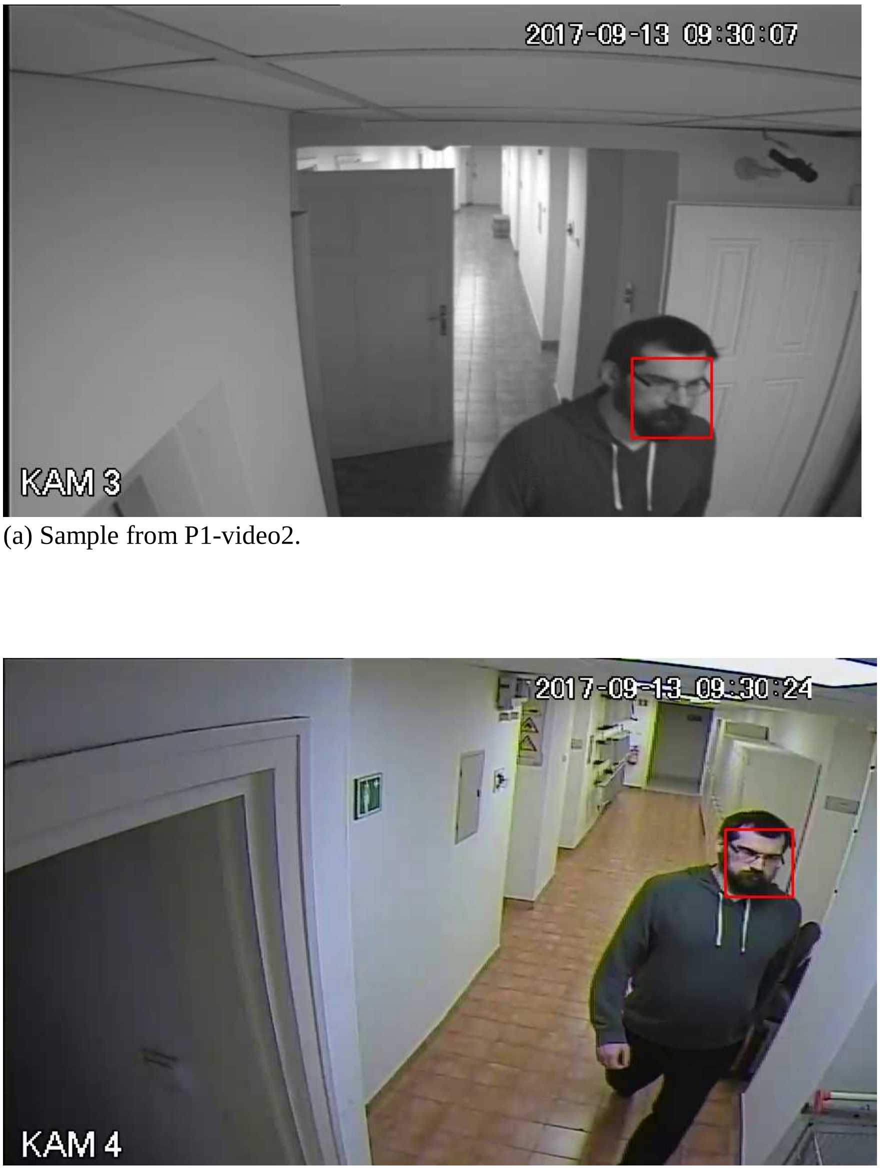


Figure 4. Structure of the VGG-Face CNN used to extract the deep features.



(a) Sample from P1-video2.

(b) Sample from P1-video3.

Figure 5. Samples of videos for Person1.

B. Deep Features

Deep Learning [16] is a branch of machine learning that uses lots of labeled data to teach computers how to perform perceptive tasks like vision or hearing, with a near-human level of accuracy. In particular, in computer vision tasks, Convolutional Neural Networks (CNNs) are exploited to learn features from labeled data. A CNN learns a hierarchy of features, starting from low level (pixels), to high level (classes). The learned feature is therefore optimized for the task and there is no need to handcraft it. Deep Learning approaches give very good results in executing tasks like image classification, object detection and recognition, scene understanding, natural language processing, traffic sign recognition, cancer cell detection and so on.

However, CNNs are not good only for classification purposes. In fact, as said before, each convolutional layer of a CNN learns a feature of the input image. In particular, the output of one of the bottom layers before the output of the network, is, in fact, a high-level representation of the input image, that can be used as a feature for that image. We call *deep feature* this representation of the image. This feature can be compared to other deep features computed on other faces, and close deep features vectors mean that the input faces are semantically similar. Therefore, if their distance is below a given threshold, we can conclude that the two faces belong to the same person.

For this work, we used the VGG-Face network [3] that is a CNN composed of 16 layers, 13 of which are convolutional. We took the output of the fully connected layer 7 (FC7) as deep feature, that is a vector of 4,096 floats (see Figure 4).

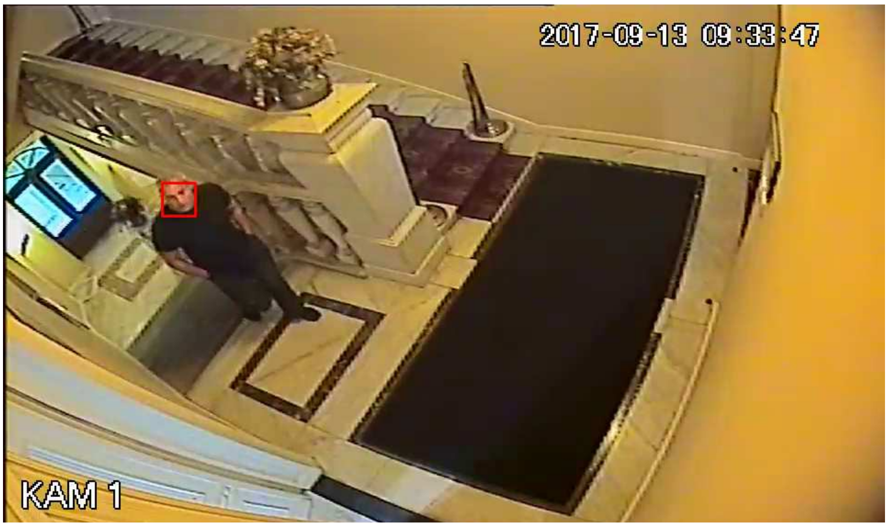
III. EXPERIMENTAL EVALUATION

In this section, we describe the experiments performed to compare the accuracy of the different features described in Sections II-A and II-B in performing the face verification task. We first describe the test set used in our experiments, that is constituted by six videos acquired by surveillance cameras deployed in some of the corridors of the Instytut Ekspertyz Sdowych in Krakow and by the famous face dataset LFW, that we used as confusion set. We then present an analysis of the distances computed over the facial landmarks and, finally, we report some accuracy results obtained by our experiments on the considered features.

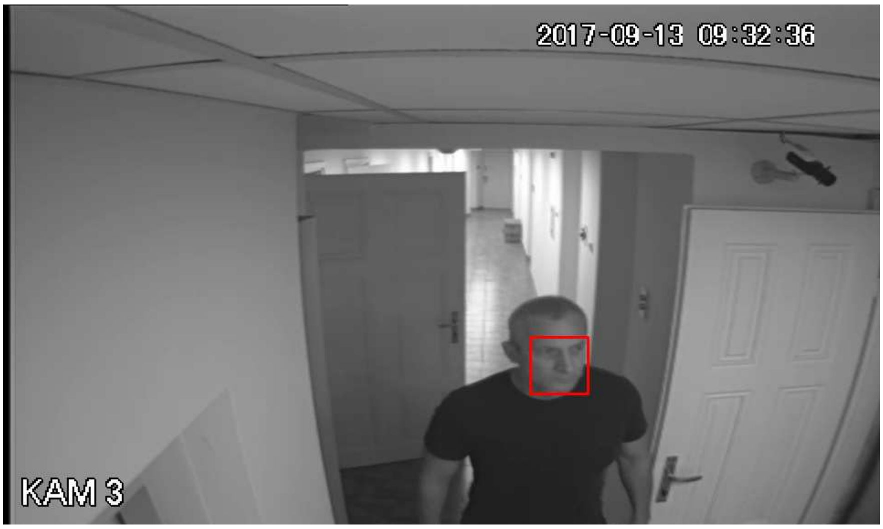
A. Test set

We used six videos as test set, provided by the EU Frame- work Programme Horizon 2020 COST Association COST Action CA16101[3]. These videos are taken from three differ- ent surveillance cameras deployed in the Instytut Ekspertyz Sdowych in Krakow and they capture two different persons (we call them ”Person1” and ”Person2”). Each of them is recorded in all the environments where the cameras are installed. So we have three videos for Person1 and three videos for Person2. For each video, we analyzed each frame independently. In particular, for each frame, we executed the face detection phase, and for the frames where a face has been detected, we executed the facial landmarks detection algorithm. We then computed the 5-points, 68-points and Pairs features, by exploiting the 68 detected landmarks.

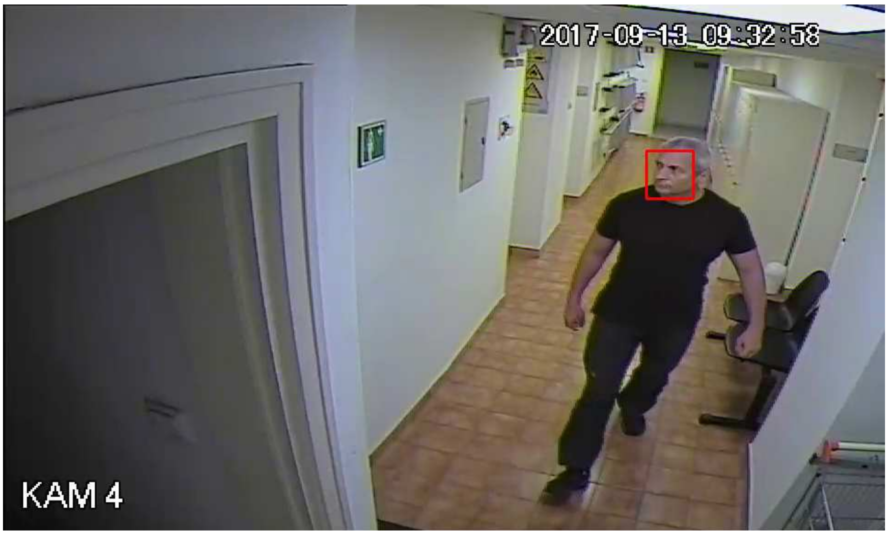
[3] http://www.cost.eu/COST Actions/ca/CA16101



(a) Sample from P2-video1.



(b) Sample from P2-video2.



(c) Sample from P2-video3.

Figure 6. Samples of videos for Person2.

The videos used in our experiments are very challenging because the resolution is low (768x576), and the person is in the foreground of the scene. We have obtained 59 total frames containing faces in all the six videos, that are composed as follows:

Person1 (P1):

- video1: 0 faces detected (the face was never recorded clearly in the video);
- video2: 5 faces detected;
- video3: 19 faces detected;

Person2 (P2):

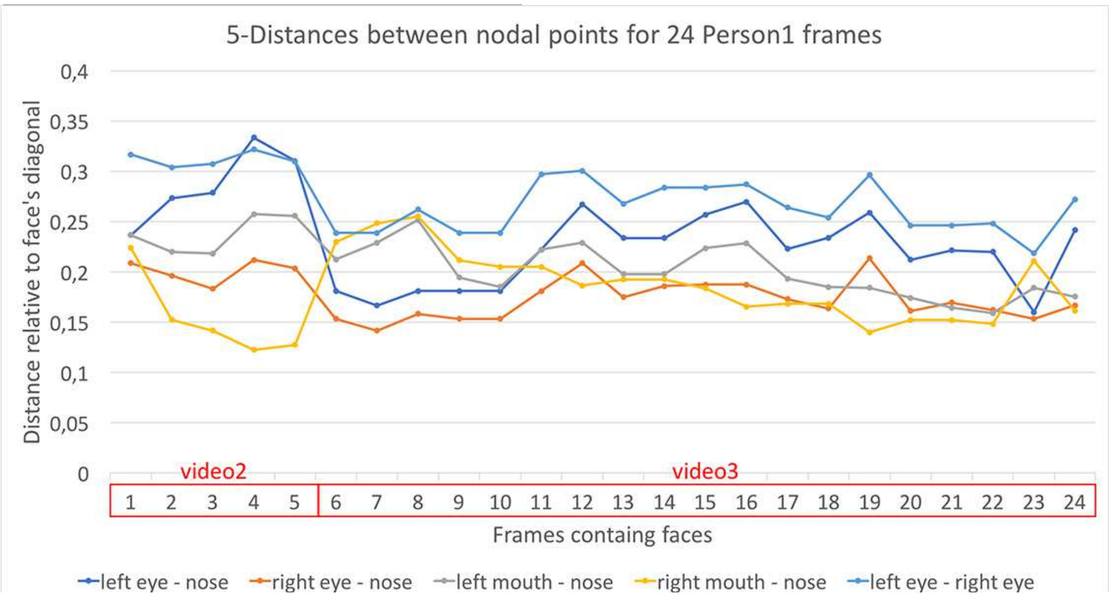
- video1: 5 faces detected;
- video2: 16 faces detected;
- video3: 14 faces detected;

Figure 5 and 6 show some samples of the Person1 videos and Person2 videos, respectively.

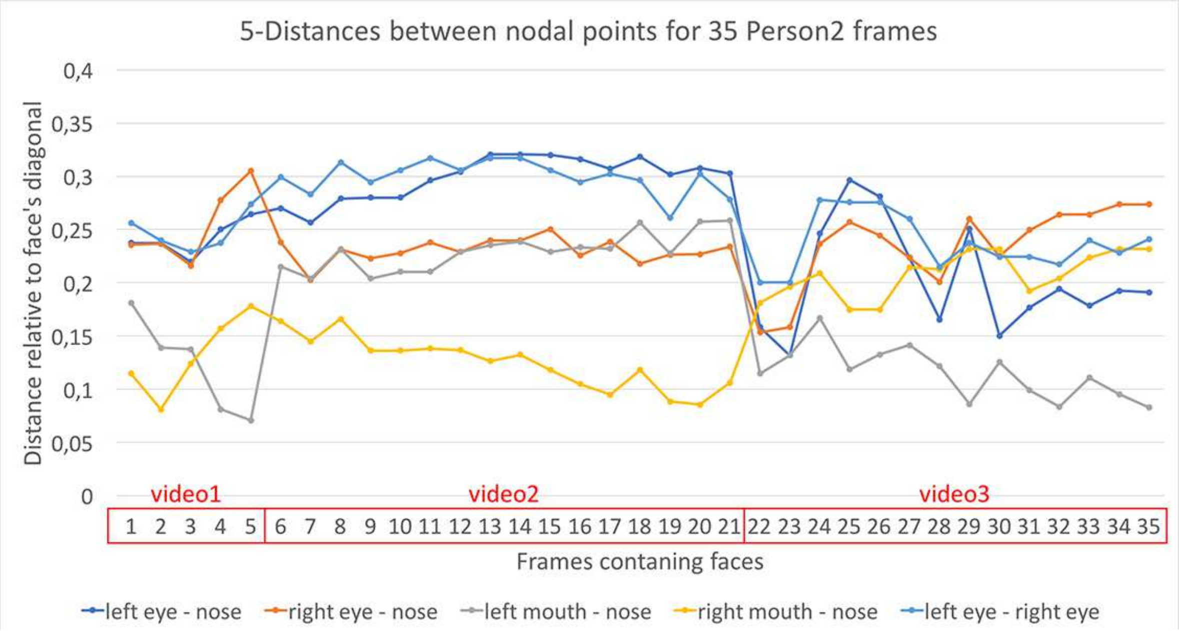
B. Facial landmarks distances measurements

In order to understand if there is a way to better exploit the distance between facial landmark points, we have performed an analysis and computed some measurements on the distances between 5 nodal points and on the distances between the 68 facial landmarks and the centroid, in different frames collected by the sampling videos that we used as test set.

Figures 7 and 8 show, respectively, the trend of the components of the 5-points and 68-points features in different frames of the videos, for both persons. Please, recall that Person1 face has been detected in just two videos, while Person2 face has been detected in all three videos. It is possible to notice that, for frames of the same video, the lines of the distances are quite regular, while they have a great difference when moving to another video. This shows that, while a person is seen by the same camera, with the same angle of view, it is possible to use the distance of facial landmarks to recognize a person by its face with good accuracy.



(a) 5-points features for Person1 videos.



(b) 5-points features for Person2 videos.

Figure 7. Distances between the 5 nodal points in different frames of Person1 (a) and Person2 (b) videos.

We also computed the average and the variance of the distances between nodal points and facial landmarks reported in Figure 9. In particular, Figure 9(a) reports the average and the variance of the distances between the 5 nodal points and Figure 9(b) reports the average and variance of the distances between the centroid of the face and the 68 facial landmarks. In both cases, the average and the variance are computed on the distance of the same pair of points in all the different frames of Person1 and Person2 videos. The figure shows that the variance is very small in almost every pair of points, and also that the average value of the two persons is quite different in four of five pairs of the nodal points (Figure 9(a)) and in lots of 68 facial landmarks (Figure 9(b)). This means that, by analyzing consecutive frames of a video, when this is feasible, it is possible to increase the possibility to recognize a certain person by using the distance of the facial landmarks.

C. Classification Accuracy

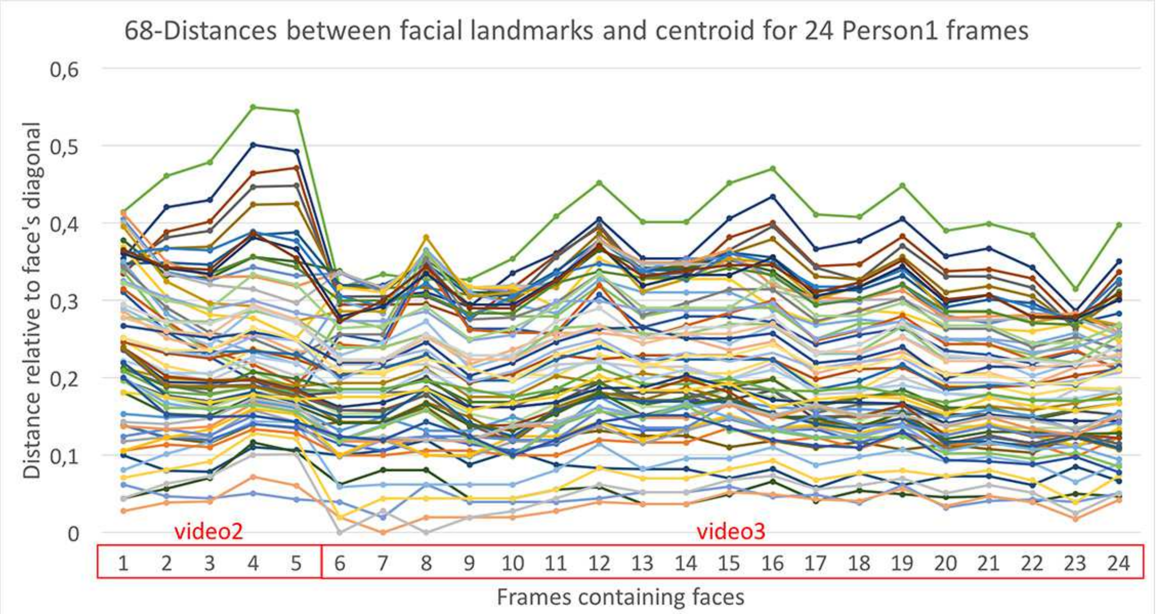
We performed some experiments to compare the accuracy in performing the face verification task by using the four different features described above. To this purpose, the faces extracted from the videos were merged with LFW, that has been used as distractor.

LFW is a very famous face dataset, which contains around 13 thousand faces and 5,750 different identities. All images in LFW are 250x250 pixels and the face is aligned to be in the center of the image. However, there is a lot of background in the images, sometimes capturing also other people faces. This could lead to multiple face detection. Therefore, we cropped each image in the LFW dataset to the size of 150x150 pixels, by keeping the same center, in order to cut the background and avoid multiple face detection. Also in this case, we performed the face detection and we computed the facial landmark points by using the dlib library (Figure 10 show some examples of LFW faces with facial landmarks highlighted). We merged the LFW dataset with the 59 faces that we detected in the test videos and we created a unified dataset. We then extracted the four different features (5-points, 68-points, Pairs and deep features), from all the faces in the new dataset.

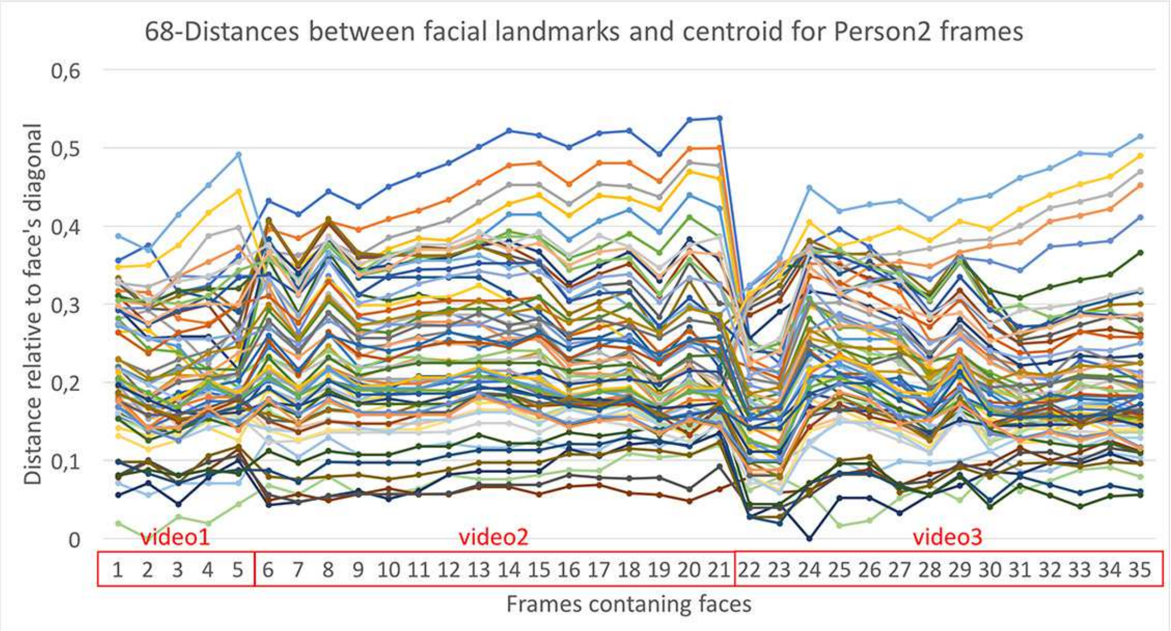
We used each of the faces detected in the test set videos as a query for a NN search in the unified dataset. We used the Euclidean distance as dissimilarity measure between features and sorted the entire dataset according to this distance with the given query, from the nearest to the farthest. We discarded the first result of each query since it is the query itself.

Figure 11 reports some query examples with the Top5 results, for all the features analyzed. For each feature, we report the best and the worst result, in which the biggest number of, respectively, correct and wrong matches in the first five results is obtained. The best result of 5-points feature only got three correct matches in the Top5 results, while all the other features got all correct matches in the Top5 results. The worst result is the same for all the facial landmarks features, that is no correct match in the Top5 results. On the other hand, the deep feature worst result only has one wrong match, that is ranked in the last of the Top5 results.

The different size of the faces detected in the videos is due to the different size of the bounding box of the face computed by the face detector library. This is caused by the different position of the person in the scene with respect to the camera; a bigger face means that the person is closer to the camera.

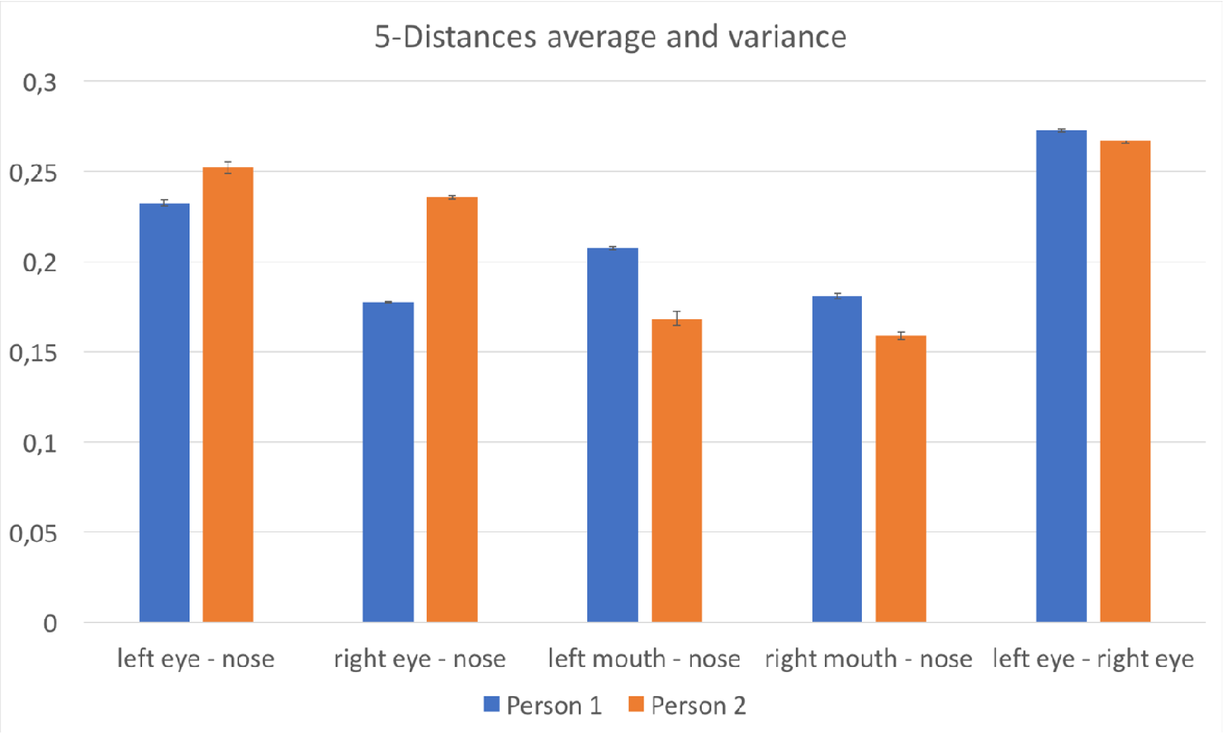


(a) 68-point features for Person1 videos.



(b) 68-points features for Person2 videos.

Figure 8. Distances between the 68 facial landmarks and the face centroid in different frames of Person1 (a) and Person2 (b) videos.



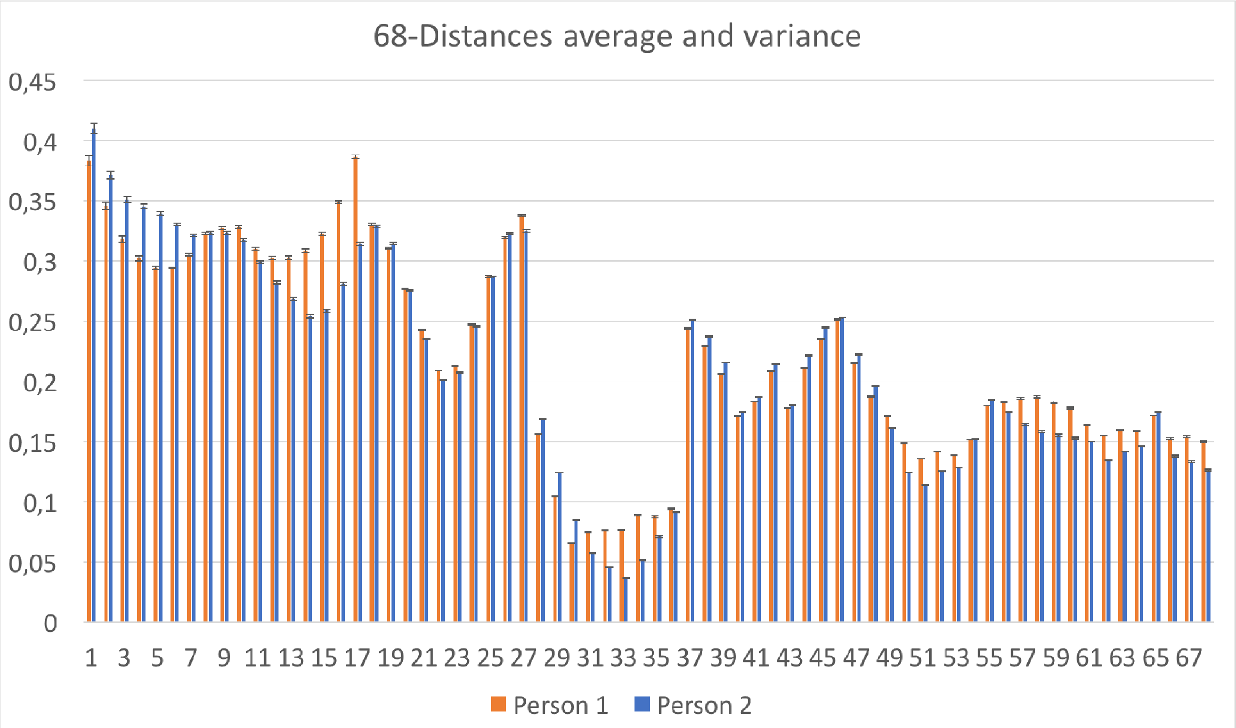
(a) 5- points feature average and variance for Person1 videos.

Feature	*mAP*
◦ 5-points feature	0.03
◦ 68-points feature	0.06
◦ Pairs feature	0.07
◦ Deep feature	0.81

TABLE I. MEAN AVERAGE PRECISION COMPUTED FOR ALL THE FOUR DIFFERENT FEATURES.

We compared all the four different features by computing the mean Average Precision (mAP) on the results of the queries, so we measured how well the results are ordered according to the query. In particular, for each query, we sum the number of correct results, weighted for their position in the result set, and we divide this value for all the correct elements in the dataset. We then average the precision of all queries, thus obtaining the mean Average Precision for each feature.

The results are reported in Table I. They show that the 68- points feature is two times better than the 5-points feature, and the Pairs feature slightly improves the 68-points feature result. However, the deep feature is more than one order of magnitude better than all the features based on the facial landmarks.



(b) 68-points feature average and variance for Person2 videos.

Figure 9. Average and variance of the 5 and 68 distances for Person1 (a) and Person2 (b).

Feature	*Top1*	*Top5*
◦ 5-points feature	24%	47%
◦ 68-points feature	51%	76%
◦ Pairs feature	64%	78%
◦ Deep feature	97%	98%

TABLE II. TOP1 AND TOP5 ACCURACY COMPUTED FOR ALL THE FOUR DIFFERENT FEATURES.

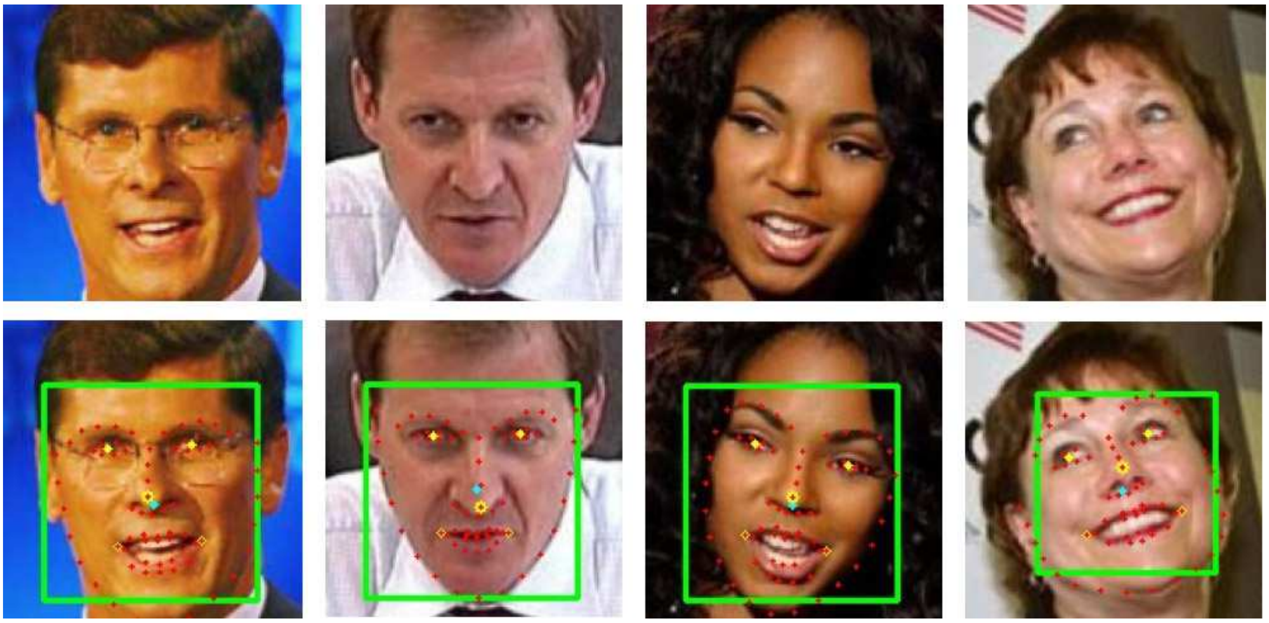


Figure 10. Some examples from LFW dataset and the corresponding detected faces with facial landmarks.

We also computed the Top1 and Top5 accuracy for all the features considered. The Top1 accuracy counts the percentage of queries in which the first person of the result set is the same person of the corresponding query. The Top5 accuracy considers the first five persons of the result set to check if the correct one is present. Table II shows that 5-points feature works very bad in this scenario with small and low-resolution faces with a Top1 accuracy of only 24% and a Top5 accuracy of 47%. The 68-points feature and the Pairs features, improve the Top1 accuracy of more than twice with respect to the 5- points feature, and up to 78% in case of the Top5 accuracy. Also in this case, however, the deep feature works much better obtaining a 97% Top1 accuracy and a 98% Top5 accuracy.

The facial landmarks have indeed the property of being an accepted proof in trials, and it can be used to classify people in some conditions and with a certain accuracy; however, the deep feature shows much better performances, especially in challenging scenarios with low-resolution faces.

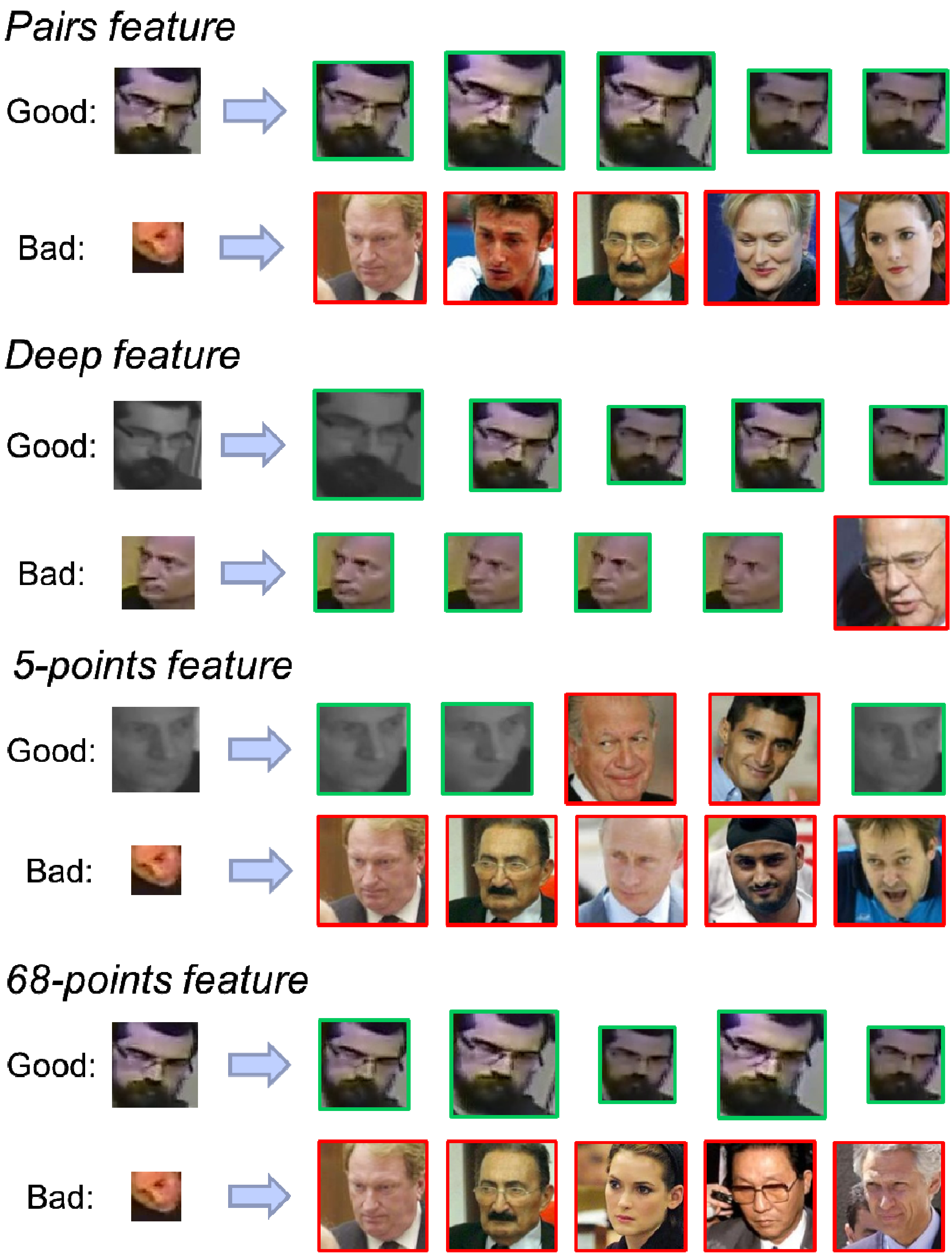


Figure 11. Query examples for the all the kinds of features, with Top5 results. For each feature, the best and the worst results are reported. The different size of the faces from the videos is due to the different size of the face detected in different frames, where the person is closer/farther to the camera.

- T. Ahonen, A. Hadid, and M. Pietika ̈inen, “Face recognition with local binary patterns,” Computer vision-eccv 2004, 2004, pp. 469–481.
- [2] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 12, 2006, pp. 2037–2041.
- [3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in British Machine Vision Conference, 2015.
- [4] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “La- beled faces in the wild: A database for studying face recognition in unconstrained environments,” Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.
- [5] P. Verlinde, G. Chollet, and M. Acheroy, “Multi-modal identity verifi- cation using expert fusion,” Information Fusion, vol. 1, no. 1, 2000, pp. 17–33.
- [6] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face rep- resentation by joint identification-verification,” in Advances in neural information processing systems, 2014, pp. 1988–1996.
- [7] C. Sanderson, M. T. Harandi, Y. Wong, and B. C. Lovell, “Combined learning of salient local descriptors and distance metrics for image set face verification,” in Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. IEEE, 2012, pp. 294–299.
- [8] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, “Attribute and simile classifiers for face verification,” in Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009, pp. 365–372.

[9] A. G. Rassadin, A. S. Gruzdev, and A. V. Savchenko, “Group-level emotion recognition using transfer learning from face identification,” arXiv preprint arXiv:1709.01688, 2017.

[10] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, “Targeting ulti- mate accuracy: Face recognition via deep embedding,” arXiv preprint arXiv:1506.07310, 2015.

[11] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.

[12] J. Park, K. Lee, and K. Kang, “Arrhythmia detection from heartbeat using k-nearest neighbor classifier,” in Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on. IEEE, 2013, pp. 15–22.

[13] D. Wang, C. Otto, and A. K. Jain, “Face search at scale,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, 2017, pp. 1122–1136.

[14] G. Amato, F. Carrara, F. Falchi, and C. Gennaro, “Efficient indexing of regional maximum activations of convolutions using full-text search engines,” in Proceedings of the 2017 ACM on International

Conference on Multimedia Retrieval. ACM, 2017, pp. 420–423.

[15] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.

[16] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, 5 2015, pp. 436–444.