Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

Face Detection with a portable architecture in Real Time*

Francesco De Feo$^{1[0000-1111-2222-3333]}$ and Pasquale De Luca$^{1[0000-0001-7031-920X]}$

University of Salerno, Department of Computer Science, Fisciano, I-84084
{f.defeo9,p.deluca16}@studenti.unisa.it

Abstract. Nowadays, security is a top priority. In fact, biometrics uses cutting-edge technologies to identify terrorists and criminals. But the practice of distinguishing humans based on intrinsic physical or behavior traits goes back thousands of years.
With the widespread use of computers in the late 20th century, new possibilities for digital biometrics emerged and new technologies were generously used.
Among these, we remember high resolution security video cameras and drones. So, the aim of the present project is to study and explain the features of these technologies, especially the ones of the the Phantom 4 Pro+ aircraft and analyze its operating methods in order to identify human faces during live streaming of videos. For this purpose, it will be used Paul Viola and Michael Jones' face detection algorithm, which includes Haar features and cascade classifiers to identify faces, eyes and ears of an individual.

Keywords: Face detection · Drone · Real Time

1 Introduction

Nowadays, biometrics authentication is used in computer science to protect the access at several systems. It is way to identify and recognize people based on distinctive and measurable physiological characteristics that don't change over an individual's lifetime; they are related to the shape of the body, such as fingerprint, palm print, palm veins, hand geometry, retina, face or iris recognition. In the last few years there has been great progress in the automatic and instant verification of an individual's physical characteristics, especially by using new and original tools, such as drones and high-definition video cameras. To confirm what we said, it is enough to think that just three years ago, Alan Butler, an attorney at the Electronic Privacy Information Center (EPIC) was not sure about the development drones could have in surveillance systems.
In this context, we have chosen to analyze the potentiality of aircrafts, especially the ones of the Phantom 4 Pro+: the aim of the present project has become, so,

* Supported by organization x.

© 2019 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

to study and test its functionalities. More precisely, our goal is the execution of a precise face detection during a live streaming recorded by the drone: to achieve that, we would build a complete system architecture, from the recording of the video to its investigation.

1.1 State of the art

One of the oldest and most basic examples of how humans take advantage of biometrics is face recognition. Since the beginning of civilization, they have used faces to identify known (familiar) and unknown (unfamiliar) individuals. This task became more complicated as populations increased and as more convenient methods of travel introduced many new individuals into small communities and, so, other body characteristics were used as a formal means of recognition.

It can be said that the earliest form of biometrics appeared thousands of centuries ago. In ancient caves, walls were adorned with paintings created by prehistoric men who left numerous hand prints as a signature; also Babylonian business transactions were recorded in clay tablets that included fingerprints. Even in China, early merchants used them to settle business transactions, while parents used both fingerprints and footprints to discriminate a child from one another.

In early Egyptian history, traders were identified by their physical descriptors to differentiate between trusted traders of known reputation and those new to the market; also slaves were recognized by height and length of their arms in order to receive a correct salary payment.

But it was in the mid-1800s, that a more formal way to recognize and identify people became necessary, because of the rapid growth of cities due to the industrial revolution. Merchants and authorities faced with increasingly numerous populations, so they could no longer rely just on their local knowledge of individuals. This situation produced the need of a formal system to record criminals with their measured physical traits. So, in 1870, Alphonse Bertillon, a French anthropologist, developed a system of measuring various body dimensions, used for the first time in the Parisian prisons, that became known as *Bertillonage*. His method consisted on precise anthropometric measurements (such as height and width of arms and fingers, distance between elbows and end of the fingers, length of trunks and feet) calculated on prisoners' body because each skeleton is different from another one and the human backbone doesn't change beyond the twentieth year of age.

Then, in 1960 in USA the first real recognition system was developed and, in 1999, NSA (National Security Agency) installed many biometrics emplacements to discern among fingerprints, voices and faces; in favor of this, there was the idea that external support could be lost, stolen or left at home, while access keys can be forgotten, copied or revealed to someone else.

From this moment, lots of things changed and lots of advances were done: technology improved, and biometrics began to be used for many services and in various activities including banking, politics or research of "wanted" individu-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

als. Nowadays, biometrics is often used in daily life as a fingerprint recognition feature that allows users to unlock their devices.

For a full comprehension, we can divide biometrics features into two groups (see Fig. 1):

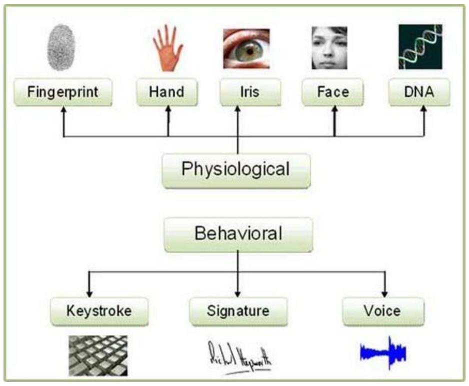

Fig. 1. Most famous and widespread biometrics systems

- *physiological*: these are related to human measurements, such as fingerprints, faces, palms, retina, iris, DNA, ear's shape, height or weight;
- *behavioural*: these are related to one's behaviour, such as handwriting, walk, keystroke, signature or voice.

Generally, the first ones are more reliable than the others, because individual's body shapes go through just a few changes during his whole life, while behavioural features are affected by psychological conditions, so they need to be changed frequently.

1.2 Drones

As said previously, among the most famous and widespread biometrics technologies we can find face recognition, that has recently gained success as a valid application for the analysis and the comprehension of images.

In the past, facial recognition software has relied on a 2D image to compare or identify another 2D image from the database, but a newly-emerging trend in facial recognition software uses a 3D model, which claims to provide more accuracy. They can also be used in real time to easily and quickly recognize who is the person located in front of the sensor (camera or webcam). Today, software for facial recognition deals with artificial intelligence algorithms, which make identification process automatic and almost immediate; this allows recognition or verification of one's identity just thanks to one or more images.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

Face detection can be done using both fixed framing cameras, which can record just what appears in front of them, so that large areas cannot be totally under their control, and auto-framing cameras, which make it is possible for the frame to move in relation to users' movements.

[SCRIVI SPECIFICHE TECNICHE, fotocamera, raggio di comunicazione etc] For our tests DJI Phantom 4 PRO+ model has been used. This model takes care of every issue a flyer could come across. With an interesting camera and a long flight time, this drone is a treat for all professional photographers, hobbyists, and flyers. It comes with a system that avoids obstacles and it has a 4K camera and lens with an amazing sharpness, so videos and photos have the best quality possible. Besides, it has an advanced 3 axis gimbal that makes sure that there are no unnecessary vibrations and in-flight movements that might actually compromise the camera's capabilities.

1.3 Related Studies

The realization of a biometric system is especially influenced by benefits that derive from implementation and overall cost, which include sensors and related infrastructures. We can define a biometric system as a technological system that uses information about a person (or other biological organism) to identify that person. A complete and good one should be easy to use for any user, well-received by population and sufficiently resistant in case of fraudulent attacks: in fact, a biometric system could be attacked in order to compromise information security (data integrity and availability). It involves running data through algorithms for a particular result, usually related to a positive identification of a user or other individual. Specifically, users enter an account, user name, or inserts a token such as a smart card, but instead of entering a password, a simple touch with a finger or a glance at a camera is enough to try to authenticate them. Then, the system had to decide whether a person is really who he declares to be: the authentication system verify the identity by searching in a database for a match based solely on the biometric.

This process of recognition could be static, if it happens using a single image, or dynamic, if face acquisition takes place through a series of frames: in the first case images are normally obtained observing one's posture, illumination and background, producing high quality photos. In the second case, instead, there are more frames extracted by the video but their quality is lower due to irregular backgrounds and postures.

Many researches have explained and examined in depth how algorithms of face detection and recognition work. Among these, one that turned out to be useful is the "Rapid Object Detection using a Boosted Cascade of Simple Features", written by Paul Viola and Michael Jones, the creators of the famous algorithm of the same name. In it, they presented Haar feature-based cascade classifiers, that is an effective object detection method. More in detail, it is a machine learning based approach where a cascade function is trained from a lot of positive and negative images.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

Part of their trick was to ignore the much more difficult problem of face recognition and concentrate only on detection, as we also did. They also focused only on faces viewed from the front, ignoring any seen from an angle. Given these bounds, they realized that the bridge of the nose usually formed a vertical line that was brighter than the eye sockets nearby. They also noticed that eyes were often in shadow and so formed a darker horizontal band. So, Viola and Jones built an algorithm that looks first for vertical bright bands in an image that might be noses; then looks for horizontal dark bands that might be eyes and, at last, it looks for other general patterns associated with faces.

The two same authors wrote another important paper,“Robust Real-Time Face Detection”, which describes a face detection framework that is capable of processing images extremely rapidly while achieving high detection rate. A basic implementation of it it’s included in OpenCV.

2 Methodology of the Study

2.1 Viola-Jones object detection framework

The ViolaJones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones [1, 2]. It can be trained to detect a variety of object classes, but it is generally used to solve the problem of face detection, as we also did during the development of this project.

The basic principle of this algorithm is to scan a sub-window capable of detecting faces across a given input image; although human can do this easily, a computer needs precise instructions and constraints. In order to obtain the expected outcomes, ViolaJones requires full view frontal upright faces pointing towards the camera. It might seems that these constraints could reduce the algorithm’s utility somehow, but it is due to the fact that the detection step is often followed by a recognition phase, so, actually, these limits on pose are quite acceptable.

First of all, we should specify that all human faces share some similar properties. For example, the eye region is darker than the upper-cheeks or the nose bridge region is brighter than the eyes. So, when you adopt this algorithm, you should look at the position and size of the eyes, the mouth and the bridge of nose and at the oriented gradients of the pixels intensities. These regularities may be matched using **Haar Features** (see Fig. 2), which involve the sums of image pixels within rectangular areas. It is a machine learning based approach created by Paul Viola and Michael Jones where a cascade function is trained from a lot of positive (face) and negative (non-face) images and it is used to detect objects in other images, by training on hundreds of thousands of face and non-face images to learn how to classify a new one correctly. So, we basically we call it *classifier*.

Algorithm is built by following steps:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

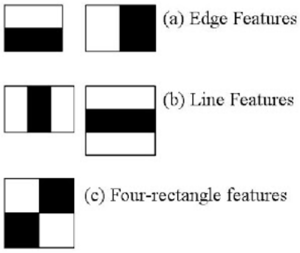

Fig. 2. Facial landmarks points

1. **Haar Feature Selection** - Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier, because we need to extract features from it. Each feature is a single value obtained by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle.
2. **Creation of an Integral Image** - All possible sizes and locations of each picture are used to calculate plenty of features. For each feature calculation, we need to find the sum of pixels under white and black rectangles. To solve this, they introduced the integral images: it simplifies calculation of sum of pixels, how large may be the number of them, making it an operation involving just four pixels.
3. **Adaboost Training** - Among all these calculated features, most of them are irrelevant [3]. A fast way to select the best ones out of 160000 (which are contained in a simple 24x24 window) or more is achieved by Adaboost [4, 5]; it selects nearly 6000 images and continues to work on them. In order to do it, we apply each and every feature on all the training images. For each feature, it finds the best threshold which will classify the faces to positive and negative; obviously, there would be errors or misclassifications. So, We just select the features with minimum error rate, which means they are the features that best classifies the face and non-face images. Final classifier is a weighted sum of these weak classifiers. It is called weak because it alone can't classify the image, but together with others forms a strong classifier. According to Viola and Jones, even 200 features provide detection with 95% accuracy.
4. **Cascading Classifiers** - In a photo, most of the image region is non-face region. So it is a better idea to have a simple method to check if a window is not a face region. If it is not, the algorithm discard it and doesn't process it again. Instead, it focuses on regions where there can be a face. This way, we can find more time to check a possible face region. For this reason, they introduced the concept of Cascade of Classifiers. Instead of applying all the 6000 features on a window, it groups features into different stages of classifiers and applies one-by-one (normally first few stages will contain very less number of features). If a window fails the first stage, it is discarded

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

and the remaining features on it aren't considered. Instead, if it passes, the algorithm applies the second stage of features and continue the process. The window which passes all stages is a face region.

3 Experiments and Results

3.1 Analysis of the problem

At the beginning, our goal was to detect faces and eyes in real time using a live video streaming produced by a Phantom 4 Pro+. In fact, with the spread of new technologies, such as drones, there have been a lot of innovations and new services have been offered in many fields like law enforcement, photography or, obviously, security.

Hence, the required creation of a support in the identification of human features that could happen in **real time**, taking advantage of high quality devices and high resolution cameras.

So, to achieve the aim of our project, we tried many softwares, applications and libraries: some of these gave us problems or seemed to slow down the visualization of the streaming, so we looked for the best free compromises we could find searching on the internet.

What we produced at the end of these studies was a real *system architecture* (see Fig. 3).

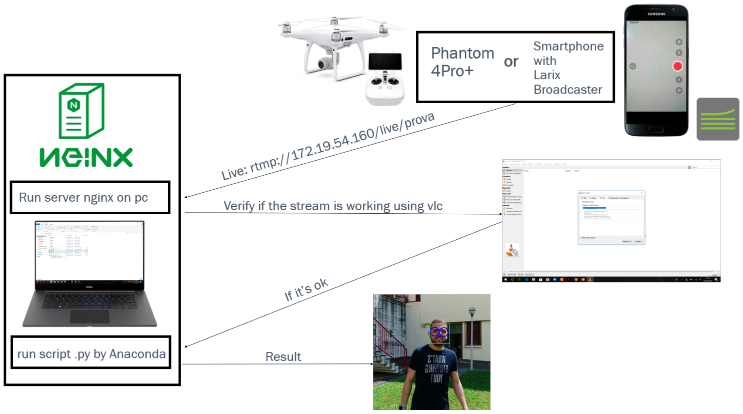
Phantom 4Pro+
or
Smartphone with Larix Broadcaster
NGINX
Live: rtmp://172.19.54.160/live/prova
Run server nginx on pc
Verify if the stream is working using vlc
If it's ok
run script .py by Anaconda
Result

Fig. 3. System architecture

3.2 Problems faced

Firstly, we deal to find a way to get the video of the live streaming filmed by the Phantom in real time, because face detection has to happen at the same time as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

the drone is recording; so, it is possible simultaneously see the video with eyes or faces identification and the one without it.

A system that could help us for this purpose was a server-based tool offered by Phantom Suite. Hence, we allows the transmition by using **RTMP** (Real-Time Messaging Protocol), a TCP-based protocol which maintains persistent connections and allows low-latency communication.

According to our specific of architecture, we chose **nginx** to stream the video, which uses an asynchronous event-driven approach to handling requests. Its modular event-driven architecture can provide more predictable performance under high loads. The default configuration file is named `nginx.conf`.

It was exactly this file that gave us many problems during this step. In fact, we had to change some parts of it and add others to allow a stream transmission using RTMP protocol.

3.3 Solutions

The only solution that could be useful to realize the processing of the live streaming was Python: so, we started using it and some of its libraries in order to write code about Viola Jones' algorithm.

OpenCV library appeared as the best solution for our goal. To perform a good implementation of our algorithm by using set of programming routines in OpenCV library. In fact, it was designed for computational efficiency[6] and with a strong focus on real-time applications.

Analysis of the script The script we wrote to detect faces involved the OpenCV library and the Haar Feature-based Cascade Classifiers.

OpenCV comes with a trainer as well as detector. If you want to train your own classifier for any object you can use OpenCV to create one. Dealing with detection, it already contains many pre-trained classifiers for face, eyes, smiles, etc. Those XML files are stored in the opencvdatahaarcascades folder.

Moreover, you have often to capture live stream with camera. OpenCV provides a very simple interface to this. Basically, you have to capture a video from the camera, convert it into gray-scale video and display it. To realize it, you had to follow a few steps and run some methods.

Firstly, we loaded the classifier, among all, we chose `haarcascade_frontalface_default.xml` to detect faces, `haarcascade_eye.xml` to detect eyes and other two to detect ears, that are `haarcascade_mcs_rightear.xml` and `haarcascade_mcs_leftear.xml`.

Then, **VideoCapture** object to capture a video has been used, this one is used as video buffer to send the classifier, which is based on `multiScale` detection by using an ad-hoc modelling of setup file.

Just changing the Classifier, we used this function to detect the other part of the face we were interested in. To perform last operations, in loop-for, several polygonal shapes have been drawn, by using related OpenCV routines.

To perform a good solution of the overall work the environment need to make a connection among drone network and server by using RTMP protocol.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

To confirm the efficiency of our architecture we show several test using differents video resolutions.

Risolution	**MT**	**Time**
360p	5	2 s
360p	10	2.5s
360p	15	3
720p	5	3s
720p	10	4s
720p	15	4.5s
720p	20	5s
1080p	5	5s
1080p	10	6s
1080p	15	7s
1080p	25	8s
4k	30	10s

Previous table shows 720x480 resolution as the best, even if produces at least 7 seconds of delay, related to the server nginx and to the algorithm used for the face detection.

3.4 Conclusions

In this thesis, we discussed the problem of face detection during a live video streaming recorded by Phantom 4 Pro+.

It interested us because of the great demand for this kind of identification in security services, agencies, investigative services and so on.

In fact, the human face plays an important role in our social interaction, conveying people's identity. Using the human face as a key to security, biometric face recognition technology has received significant attention in the past several years due to its potential for a wide variety of applications in both law enforcement and non-law enforcement.

As compared with other biometrics systems using fingerprint or palmprint and iris, face recognition has distinct advantages because of its non-contact process. Face images can be captured from a distance without touching the person being identified, and the identification does not require interacting with the person. In addition, face recognition serves the crime deterrent purpose because face images that have been recorded and archived can later help identify a person.

This, in addition to the progress of technologies, led to the use of new devices, such as drones, even for face detection.

So, the aim of the present project became to realize a new architecture that could help us to distinguish faces and its features, especially eyes and ears, while filming a live streaming by the above-mentioned aircraft (you can see first picture of this paragraph Fig.5 - to have an idea of how it really works).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019

doi:10.20944/preprints201909.0100.v1

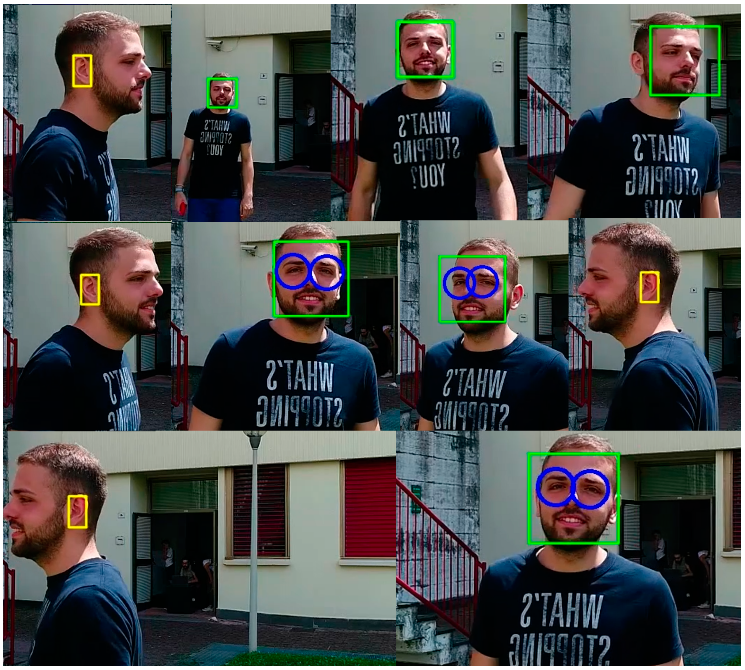

Fig. 4. Image selected from the live video showing different types of detection

As you can see in these images, while the drone is filming some people (members of our group), the script is running and getting results (displayed on the computer).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

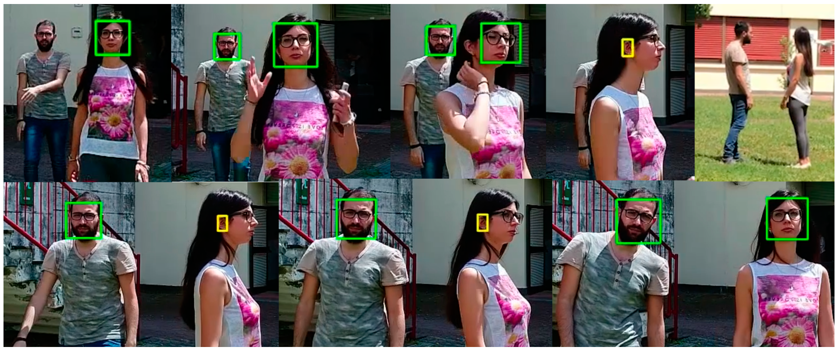

Fig. 5. Image selected from the live video showing different types of detection

Besides, we had also estimate some significant values: the greatest distance required to detect face is about 8 meters, while the greatest one required to detect eyes and ears is about 1 meter (see Fig.6). Moreover, the resolution chosen for the streaming of the drone is 1280x720, which is also the smallest it supports, while the one chosen for smartphones (using Larix Broadcaster) is 720x480.

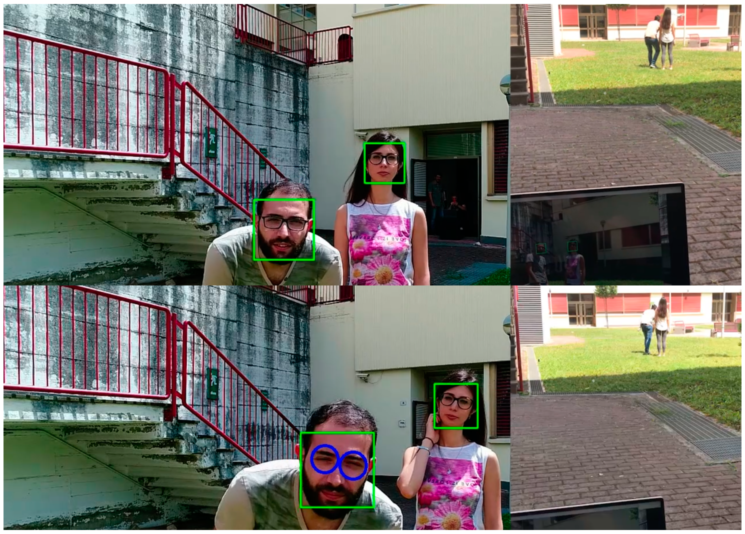

Fig. 6. Comparison among drone and computer visualization

Our experimental results highlight a small delay due to high number of operations according to algorithm. The delay is about 5-7 seconds of delay are due to the processing of the images and on calculating capacity. So, we have a total delay of about 7-8 seconds; however, we can still consider it as a real-time sys-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

tem, considering that we had worked using our average computers, but, if you have the opportunity to test it with a more powerful pc, you will sure notice the difference.

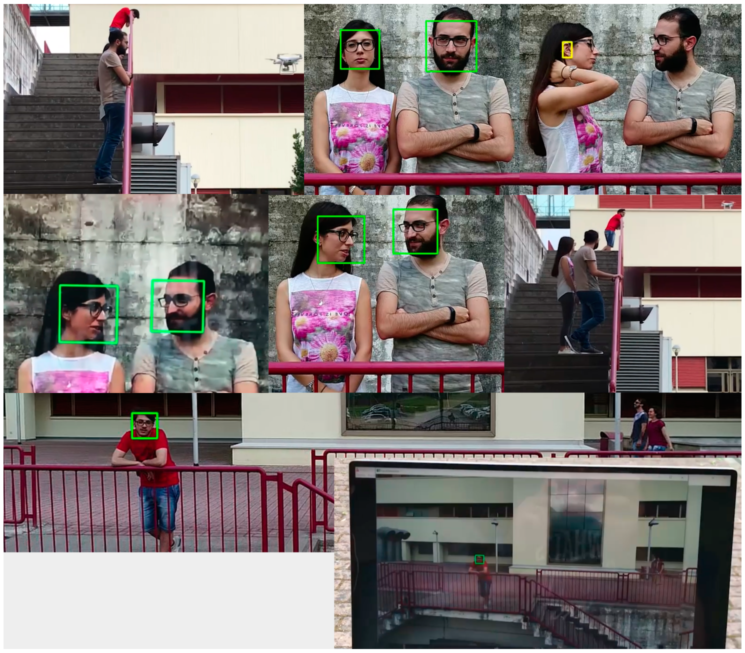

Fig. 7. Comparison between drone and computer visualization. More in details, the first picture of the second row is extracted from the video streaming on computer, while the next one represents the same instant processed by the script

4 Future work

In view of future works and researches in the same area of interest, there are several updates that could be done.

First, it could be helpful to add a grid on irises in order to identify them better and faster; keeping on thinking about eyes, it could be interesting to improve the precision in their recognition even if a person is far from the camera. For this reason, circles around eyes should modify their size adapting to the face on which they are drawn; in fact, now it is a fixed parameter.

Furthermore, it would be useful to implement methods to provide face recognition in addition to the detection. For this reason, feature extraction should be improved and mainly tested, because it is crucial for any recognition algorithm and system. A remarkable change would been noticed in the recognition rate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 September 2019 doi:10.20944/preprints201909.0100.v1

using different feature extraction. Specifically, in cropping an image before it is run through a recognition system, there would be still much work to be done in this area. It would be interesting to explore new techniques of preprocessing that would lead to the optimal recognition rates. In this case, face and ear biometrics has been used, but perhaps there are other metrics that can be combined that will lead to a more robust system. It would be interesting to see what kind of results could be achieved in a system like this with different metrics.

References

1. Viola, P., Jones Michael J.: Robust Real-Time Face Detection. In: International Journal of Computer Vision **57**(2), (2004).
2. Viola, P., Jones Michael J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Accepted Conference on Computer Vision and Pattern Recognition, (2001).
3. Freund Y., Schapire R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Journal of computer and system sciences **55**, pp. 119–139, (1997).
4. Papageorgiou C., Oren M., Poggio T.: A general framework for object detection. In: International Conference on Computer Vision, (1998).
5. Freund Y., Schapire R. E., Bartlett P., Lee W. S.: Boosting the margin: a new explanation for the effectiveness of voting methods. Ann. Stat., **26**(5), pp. 1651–1686, (1998).
6. P. De Luca, A. Galletti, G. Giunta, L. Marcellino, M. Raei - Performance analysis of a multicore implementation for solving a two-dimensional inverse anomalous diffusion problem. - AIP Proceeding 2019 NUMTA Conference.