Published in: [Pattern Recognition Letters] journal.
This version of the article has been accepted for publication, after peer review and is subject to Elsevier terms of use.
The Version of Record is available online at: http://dx.doi.org/10.1016/j.patrec.2021.04.006

Remote 3D face reconstruction by means of autonomous unmanned aerial vehicles

Andrea F. Abate, Luigi De Maio*, Riccardo Distasi, Fabio Narducci

Department of Computer Science, University of Salerno, Salerno, Italy

A B S T R A C T

The 3D face model of an individual includes a rich amount of information useful in many application scenarios. Currently, the 3D reconstruction from images is obtained through multiple acquisitions at different distances and angles in a controlled environment. “In the wild’ conditions, or inadequate acquisition systems, can make 3D reconstruction a strongly error-prone problem. This paper illustrates a system that creates 3D face models from images taken by unmanned aerial vehicles in a completely automatic way. The proposed method is adaptive, dynamic, and contactless. No human intervention is required for image acquisition. Experimental results are described—in particular, the comparison between the ideal 3D reconstructions, obtained in controlled and cooperative conditions, and the reconstructions obtained by the drone during flight at different resolutions. The results show good overlap of the models and a low co-registration error. Given the free mobility of the UAVs, the system is suitable, among other applications, for biometrics and open-air access control over large areas.

1. Introduction

The 3d model of a face provides a rich amount of information useful for biometric recognition. Having a 3d model available allows to check the acquisition of faces in non-frontal poses with greater precision. Biometrics measure physical traits or behavioral characteristics of an individual. Among several different purposes, they can help to detect, identify and recognize subjects in surveillance systems. Examples of biometric traits are face, ear, iris, gait, voice, typing speed patterns and, more generally, everything that can be measured [1]. During the biometric acquisition process, the interaction between a subject and the system should be as non-invasive as possible. In the presence of large numbers of people, it is crucial to automatize such a process. Contactless and unconstrained interfaces are preferred whenever possible [2].

Recently, Unmanned Aerial Vehicles (UAVs) have appeared as an opportunity for biometric purposes. UAVs can replace systems based on camera grids, making it possible to monitor large areas [3], sparing the use of a large number of sensors. Aerial vehicles are equipped with cameras and possibly other sensors, some of which are integrated into mobile devices such as smartphones.

As a result of technological progress, these sensors have changed how UAVs interact with the environment and with people. In particular, on-board cameras allow more advanced mobile framing. For example, the classic Pan, Tilt and Zoom mode (PTZ), where only the camera has moving parts, can be extended to Throttle, Pitch, Roll, Yaw. The degrees of freedom of the aerial vehicle, which can freely move in the working environment, allow a greater set of options for the shot and hence for the acquisition most suitable to biometric goals.

The flight of these devices is usually managed by a pilot through a RC (Remote Controller). In some cases, however, the pilot might not be able to promptly react in critical situations; hence the need for automating processes such as object tracking, take-off, landing, return to home, etc. The autonomous flight of UAVs is also desirable in order to reach specific targets or avoid obstacles [4].

This paper proposes the implementation of an adaptive biometric acquisition system and provides an example of automatic acquisition for the 3D face reconstruction. The solution is achieved through an Android application using the drone DJI Phantom 4 Pro+ [5] and DJI Mobile SDK [6] libraries. The drone adapts to the position of the subjects face in the frame, flying autonomously to take photos from an adequate distance and perspective view. The operator performs simple start and stop operations of the system.

The pilot-drone interface to the system is simple, since it only entails starting and stopping the application. Besides, the only assumption is the physical presence of a human subject in the scene. Full freedom of movement is granted by the contactless mode: the subjects can take control of the drone by getting its attention. The drone then detects the user’s face and starts hovering around to

take photos. The captured images are processed to produce the 3D model.

The technique proposed in this paper can be exploited as an acquisition method for the creation of biometric face datasets with the added value of 3D information. It can also be part of a gate access system for wide areas with low presence density, or with higher density if fleet management is integrated. Another possible use of this technique could be surveillance in areas that are dangerous for environmental or human factors, or where fixed cameras are likely to be vandalized or stolen.

The paper is organized as follows. Section 2 contains a survey of autonomous control systems for UAV and their application to biometrics. Section 3 illustrates the system architecture. Section 4 discusses the experimental results achieved. Finally, Section 5 draws some conclusions.

2. Related work

2.1. Background

Physical and behavioral biometrics can be combined to build multiple biometric systems able to prevent access to restricted areas or to protect private data in a more secure way. A typical example of behavioral biometrics is the pattern by which a person looks at a face, as studied in Cantoni et al. [7]. The fixation points over time allow discriminating a subject from another.

Biometric identification and recognition are performed through the matching of biometric templates. Face biometrics solutions are often based on 2D models representing the structure of the face. Alternatively, 3D models can be used for more details [8]. 3D face models have depth information absent from the typical 2D models discussed in face recognition literature. Therefore, they add accuracy to identification and recognition, but increase the computational workload [9]. 2D and 3D models can be used in combined approaches such as so-called multimodal algorithms [10]. An example method of face recognition fusing 2D and 3D models is shown in Lin et al. [11].

The acquisition of a 3D face model usually takes place in controlled systems designed for this specific purpose, mostly through the use of a 3D scanner. A good example is the system used for the construction of the FLORENCE 3D FACE dataset [12]. It is also possible to obtain a 3D model through the use of monocular cameras. This requires taking simultaneous shots from various specific angles and then processing them with reconstruction algorithms. Such an approach has been used for the construction of the dataset COMPACT [13]. Another example of acquisition is the use of face images captured from random angles such as single frames extracted from videos [14]. Recent research has shown that, even with a single frame, it is possible to obtain accurate models by the alignment of preloaded models, as shown in Jiang et al. [15] or by the use of neural networks [16].

The model can be built in a later postprocessing phase or in real time, depending on the complexity of the task and the computational power available.

2.2. Drones in biometric system

UAVs, commonly known as *drones*, are used for various purposes such as surveillance, photography, hobbies, interactive social context, rescue, and more [17]. They are becoming more and more ubiquitous. Sometimes, the human factor decreases the probability of success of a mission. For this reason, there are appropriate interfaces to maximize automation and minimize the possible impact of human error. One such interface is [18]. In the present work, however, the interface is reduced to a minimum and the control is carried out by the software.

The conventional user interfaces for human-drone interaction are remote controls, phones and simple gestures where the users can interact with a drone in a natural manner [19]. In some cases, it is possible to use biometrics itself as an interface to human drone interaction, in a non conventional way. An example would be aiming and launching flying robots on user-defined trajectories. In [20], the authors present a user interface using face data. The method is easy to learn and requires no user instrumentation. The drone is sent along a desired 3D trajectory beyond line-of-sight. As a different take on this problem, an example of combined interface gestures and face pose estimation can be found in Nagi et al. [21]. Combining the use of behavioral biometrics and drone flight is also possible. In one study, the drone follows trajectories related to the emotion of the pilot [22]. Another human-robot interaction application that uses face information for control can be found in Yao et al. [23].

Some recent biometric datasets of face images have used drones to obtain images from above. Representative examples include the DroneFace dataset [24] and the multimodal dataset MUBIDUS-I. In the latter case, images were acquired in both controlled and non-controlled conditions [25]

3. Method and tools

A prior task to recognition is detection. For detection, we have used the Google Mobile Vision API [26], now a part of the Machine Learning Kit. The Mobile Vision API library was chosen over the classic OpenCV [27] because of better portability and scalability. The Google API and the DJI SDK can be used on any DJI remote flight control system, without need for special computing power. Had OpenCV been used for the detection stage, there would have been more limitations.

The detection algorithm used in the API is based on a machine learning approach for object localization that guarantees support over time. To make use of the Phantom 4 Pro, we employed the DJI SDK, mainly the Mobile SDK, to create a customized mobile app unlocking the full potential of the DJI aerial platform.

The Android app was built in the Android Studio IDE with Android DJI SDK full compatibility [28].

The drone is initially in stand-by on the ground. In the first step, the application signals the drone to raise the camera and wait for face detection within the scene. Once a face is detected, the UAV takes off and aligns the camera with it both horizontally and vertically, then the preset sequence starts. At this stage, the flight follows a geometric figure pattern, such as a rhomboid, around the subject while taking a shot of the scene. Other flight paths, such as circles, are possible. What is important is that the subject is always framed at various angles. For simplicity, only rhomboids were used.

Finally, the 3DF Zephyr software form 3DF Flow [29] processes the video acquired by the drone and produces an accurate 3D reconstruction of the subject under consideration. A repository containing the source code of the proposed solution is available online at https://github.com/ldema/Remote3D.

3.1. Biometrics

The objective of the present work is reconstructing the subject's face through a 3D representation of the biometric trait. This puts our proposal within the area of biometric collection systems—more specifically, in the class of subsystems that address the issues of biometric trait representation and manipulation. The problem we tackle is that of achieving a robust and reliable 3D representation of a face acquired from a monocular system (the drone). The higher the quality of the reconstruction, the better the performance expected by a face recognizer based on such reconstruction.

The term 'Biometrics' refers to all measurable human physical and behavioral characteristics. Some of these features are unique to each individual, and the metric used can be specific. This information labels and describes individuals. It can be used for security purposes in authentication or identification; in some cases, it can even detect pathology. Such data are therefore strictly personal, and their use is regulated by specific laws about their treatment and storage. The ethical and safety problems springing from the collection and processing of biometric data are still largely uncharted, and a substantial amount of work is being done on the matter.

The acquisition of single or multiple traits can be carried out in controlled conditions or in so-called 'wild' environments, in a multimodal way, at different times, with multiple sensors, with or without the consent of the subject. The characteristic of an ideal biometric trait are: (i) the Universality: every person should possess the trait; (ii) the Uniqueness: the trait should differ between individuals to ensure distinctiveness; (iii) the Permanence: the feature should be reasonably unchangeable over time, in order to remain meaningful; (iv) the Measurability/collectibility: the acquisition or measurement of the trait should be easy; (v) the Performance: the technology used for the traits should provide adequate accuracy, speed, and robustness; (vi) the Acceptability: the relevant population should accept capture and storage of that biometric trait; (vii) the Circumvention resistance: it should be hard to imitate a biometric trait with an artifact or substitute. The proposed system meets all those properties. It is commonly agreed that the human face satisfies the aforementioned properties. In fact, the face is considered a hard biometric trait [8].

3.2. Unmanned aerial system

A system consisting of a UAV, a ground-based controller, and a communication system between the two is called a UAS (Unmanned Aircraft System). These aerial vehicles are classified differently by defense agencies or civilians, with continuously evolving categories. They can be classified by their size, the ranges they can travel, their endurance in the air, and in some case their cost, as explained in Watts et al. [30]. The evolution of commercial models in terms of performance and sensors has reached levels that can be compared to those of professional models. The model chosen for this study is such a case.

Originally, these vehicles were used for missions considered "dull, dirty or dangerous" for humans, usually military [31]. However, their use quickly expanded in all human activities: daily life, leisure, and scientific research. Many benefits came in the security sector, where UAVs help overcome architectural barriers and can be equipped with modern technology. The different types of possible on-board sensors allow deployment in several critical situations.

UAVs may be flown either by a human pilot under remote control (RPAS - Remotely Piloted Aircraft System), by onboard computers autonomously, or through a remote application—as in our case. Drones often have the ability to be programmed to create intelligent flight systems like the DJI drone, one of the most widely used commercial systems.

3.3. Hardware

Acquisition devices The hardware used is not particularly high-end. The most exotic item is a DJI Phantom 4 Pro+ drone with a 30 min flight time, 7 km control distance, 72 km/h flight speed, 4 K 60 fps video resolution, 30 M sensors distance, Obstacle sensing 5 Directions, 6 Camera navigation system, main camera specification (FOV) 84° 8.8 mm/24 mm (35 mm format equivalent) f/2.8–f/11 auto focus at 1 m $-\infty$. In addition a smartphone Samsung S8 with 12 MP, f/1.7, 26 mm (wide), 1/2.55″, 1.4 µm, dual pixel PDAF, OIS, back camera.

Computation system The reconstruction was computed on an Asus laptop with Intel® Core™ i7 6700HQ Processor, Intel® HM170 Chipset, NVIDIA® GeForce® GTX 1070 with 8 GB VRAM, 16 GB DDR4 2133 MHz SDRAM.

3.4. Software

Google mobile vision API The Mobile Vision API is a framework that allows the identification of specific objects, like faces, barcodes, and text, present in photos or videos. The framework is part of the ML Kit (Machine learning for mobile developers Kit), an easy-to-use and powerful package that brings Googles expertise in machine learning to mobile developers. It includes pretrained detectors that work in real-time. They locate and describe the object in any orientation within images or video frames. The Face detector used in this study returns the coordinates of the areas containing human faces. Furthermore, it is able to extract the positions of facial landmarks such as contours of facial features: eyes, ears, cheeks, nose, and mouth. The response time of this detector allows real-time processing of the video stream from the drone.

DJI mobile SDK software The DJI Mobile SDK is a software development kit designed to give developers software access to the capabilities of DJIs aircraft. The SDK provides high level calls for low level functionality such as flight stabilization, battery management, and signal transmission and communication, thus simplifying application development. The SDK includes a library/framework that can be imported into an Android or iOS app. Many product features and capabilities are accessible to developers through the SDK: developers can automate the flight, control the camera and gimbal, receive real time video and sensor data, download saved media from the product, and monitor the state of the other components. Many of these capabilities have been used in this work. Some issues that need to be addressed include the following. The aircraft can have a mass of up to 2.80 kg and move at speeds of up to 20 m/s. While the ability to programmatically change position is quite powerful, attention has to be paid to the things and people around it (kinetic energy). DJIs aircraft moves in space shared by people, structures and possibly other aircraft. DJI provides a geofencing system to prevent aircraft from entering critical space. However, developers and users should still pay constant attention to aircraft movements to avoid accidental and risky impacts (share space). Wireless connectivity can be unpredictable in challenging wireless environments. It can take hundreds of milliseconds for a command to be transferred, assuming it ever does. There can be unexpected consequences in the physical world (highly asynchronous process).

Fig. 1 illustrates how the DJI Mobile SDK fits into a mobile application, and how it is connected to a DJI aircraft. Fig. 2 shows the flight simulator during an application test. A mobile application is built with the DJI Mobile SDK based on the relevant platform (iOS or Android), and it is run on a Mobile Device such as an Apple iPhone, iPad, Nexus phone, Nexus tablet, or others.

3DF ZEPHYR software By means of the 3DF Zephyr software, it is possible to reconstruct 3D models of objects from multiple photos. The software does not require specific hardware requirements, but it is able to exploit GPU computing power if one is present. It is built on top of proprietary reconstruction technology, basically photogrammetry techniques. If the framed object is stationary and the photos are in focus, the resulting model has high accuracy. The software extracts the best focused frames and discards the ones that are not similar enough. Details of this step are better discussed in Section 4.2. For our experiments, the default parameters settings were used, with focus set at high level. During the processing step, the software selects the best frames for re-

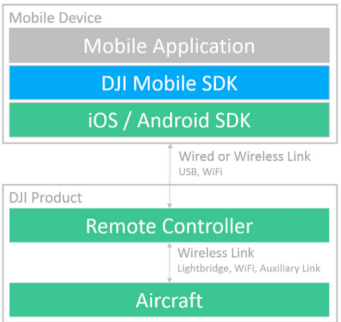


Fig. 1. For a handheld camera product, the remote controller is replaced by a handheld controller. There is no additional aircraft or wireless link.

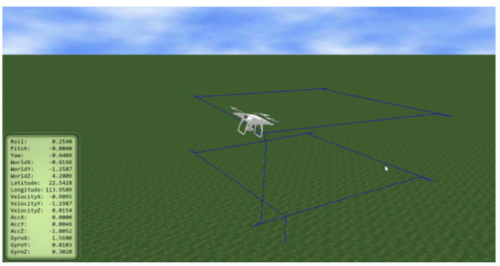


Fig. 2. A screen shot of the flight simulator.

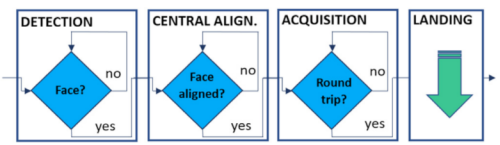


Fig. 3. Summary of the first system component: drone flight for data collection.

Fig. 4. (Left) The drone is in loop until a face is detected in the scene. (Right) The drone has detected a face and is preparing to take off.

construction among the ones extracted. 3DF Zephyr can create and export meshes and point clouds in the most common 3D formats to generate video animations, digital terrain models, sections, and contour lines. It can infer angles, areas and volumes. 3DF Zephyr is a computationally demanding application, especially with high resolution images, but multiple CPU cores and Nvidia CUDA technology, if available, can speed up the computation.

Fig. 5. Sample of half body 3D reconstructions from Drone at different resolutions (FHD on the left and C4K in the middle) and Mobile (FHD resolution).

4. Experiment design

The system is mainly composed by three software modules. The first module automates the acquisition process of the subject, recording video frames once a human face is detected in the video stream. The drone autonomously flies and corrects its route to provide different perspective acquisitions of the subject. The second module implements the 3D reconstruction process. It discards the useless frames: those with excessive blur, with suboptimal angular views, and other defects). Some samples of half body models are shown in Fig. 5. The third module of the pipeline performs the co-registration of the models so obtained. The 3D model of a subject reconstructed by the drone during flight is co-registered with the model of the same subject acquired in ideal and cooperative conditions. As expected, the accuracy of the model reconstructed by the acquisitions from the drone is as much higher as the co-registration error is low.

The following sections present the interface of the system, the testbed and the software product used to manage the point clouds. The experimental results in Section 4.3 discuss the level of performance achieved and make a few observations inferred from the behavior of the system.

4.1. Human interaction and interface

When acquiring raw data, the drone goes through a succession of several phases and states in a specified order, much like a finite automaton. A summary of the phases and phase transitions is illustrated in Fig. 3.

The first phase is called `DETECTION`. The Phantom 4 is at ground level, in the initial state called `LANDED`. When the user presses the start button on the RC, the automation starts. First of all, a command is sent to the drone to rotate the gimbal upwards by 29.5°, which is the maximum possible rotation for the camera. Next, the application starts streaming from the vehicle, and the streamed data is processed by Google Mobile Vision API's real time face detection. When a face is detected in the scene, the UAV takes off and rotates the gimbal to 0°, which is the default camera position. This new state is called `TAKE-OFF`. Fig. 4 shows an example of real execution of this state.

In the next phase, called `CENTRAL ALIGNMENT`, the drone, after taking off, tries to align the face in the center of the frame both vertically and horizontally. After alignment, the gimbal starts the video acquisition and the drone stops, passing to the state `STALLED FLIGHT` to perform Phase 3 (`ACQUISITION`).

The `ACQUISITION` phase consists in actually acquiring the face data. This is carried out by flying around the subject. Ideally, the height stays at human eye-level. Height stabilization is obtained by appropriately modulating the parameters Yaw and Pitch in rapid succession. For safety reasons, it is possible to immediately stop the drone from continuing the trajectory at any time by pressing a dedicated button.

The final phase, `LANDING`, happens when the trajectory has completed and all the data have been acquired. The drone goes back to ground level and returns the camera to the original 29.5° orientation.

Table 1
Quantitative data of the experimental session.

Data information	Drone	Drone	Drone	Drone	Mobile
Video resolution	FHD	2.7 K	4 K	C4 K	FHD
Video duration (sec)	[90,120]	[90,120]	[90,120]	[90,120]	[25,35]
Video-frames used	[80,130]	[80,120]	[75,110]	[80,110]	[50,75]
Processing time (min)	[90,120]	[90,120]	[120,180]	[120,180]	[50,75]
Point cloud size	[1k,2k]	[2k,3k]	[3k,4k]	[3k,3.5k]	[4.5k,6k]

4.2. 3D reconstruction

After the automated drone acquisition has captured a video recording of the subject, 3D reconstruction by the 3DF Zephyr software can start. First of all, the video is imported into the software. Then, keyframes are extracted from the video. 3DF Zephyr can import not only images, but also videos. When doing so, a video is divided into frames that are subsequently treated as individual still images. In this phase, it is possible to set the frame rate for extraction (the frame-per-second, FPS), the automatic analysis of the blurring artifacts in each frame and the thresholds for rejecting the outlier frames that are not similar enough to the others. If necessary, it is also possible to intervene manually and cut some parts of the video that are not needed for extraction - e.g., those not including the subject.

The next phase, called `STRUCTURE FROM MOTION`, processes all the images initially loaded into the software, normalizes their orientation, and produces a first rendering of the photographed subject defined as a cloud of scattered points.

Fig. 5 shows an example of results from the last two phases after imposing a texture on the three-dimensional model.

4.3. Results and discussion

Starting from the system above described, 3D models of 20 participants have been constructed in different resolutions and at a distance ranging from 2.5 to 3 m. The set of subjects used for the experiments is made up of males, aged between 40 and 50, with a height ranging from 1.7 to 1.8 m, of Caucasian ethnicity, without occlusions on the face area except for one subject wearing a mustache. The environment is outdoor and consists of an open space without obstacles between the drone and the subject to be acquired. The area behind the drone is also free of obstacles for at least 2 m. The shots were taken with minimal wind and in light conditions either sunny, but without shadows in the shooting area, or cloudy. Four different resolutions have been considered: FHD, 2.7 K, 4 K, and C4 K. The 3D models contain a wealth of information. In this work, we focus on the 3D reconstruction of the face and the alignment of the models obtained.

A collection of quantitative data from the experimental session is summarized in Table 1. 'Processing time' refers to the 640×480 crop of each videoframes, corresponding to the bounding boxes containing the half body of the subjects. The main contribution of these quantitative results is about the effect of video resolution on the computing time and on the extraction of point cloud, which in turn should impact on the accuracy of the reconstructed model. It can be observed that even if the processing time is reasonably limited and the variations on higher resolutions can be controlled, the number of points extracted from the videos at 4K resolution is higher than in the other conditions. This suggests that C4 K resolution can be more affected by noise, while resolutions lower than 4K do not capture sufficient information and do not achieve a quality that can produce accurate 3D models. The experimental results shown in the following further confirm this observation.

To evaluate the accuracy of the 3D reconstruction from the drone, the crops of the 3D face have been aligned with the ideal model acquired in controlled laboratory conditions through a mobile device. The ICP (Iterative Closest Point) method has been used to align the models [32], and the RMSE (Root-Mean-Square Error) as a metric of co-registration error to estimate how the 3D models fit each other, according to the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{Drone}_i - \mathrm{Mobile}_i)^2} \tag{1}$$

where N is the total number of points in the 3D cloud saved by the co-registration process, while Drone_i and Mobile_i are the corresponding points in the two clouds found through the ICP algorithm—in other words, they represent the closest points found during the co-registration. Table 2 shows the mean RMSE achieved on all participants acquired and reconstructed in the experimental session. The table also reports the mean variance of the alignment of the point clouds. As expected from the preliminary considerations on the size of the point clouds, the video resolution at 4K achieves the most promising results. The mean error is the lowest among all resolutions considered. The impact of this result can be also visually inspected in Fig. 6.

Taking a subject as a reference from the dataset, the crop of the face has been extracted from the half body 3D model. Visual results of such a crop are shown in Fig. 7b, where the region of interest has been segmented from all 3D models at different video resolution and compared with the model from mobile acquisition. The overlap among the 3D models can be also better appreciated in Fig. 8. Different colors show how points from model clouds are

Table 2
Results of the alignment of the 3D models in terms of RMSE and variance of the point clouds.

	Drone video resolution			
	FHD	2.7 K	4 K	C4 K
RMSE	0.0356	0.0270	0.0264	0.0275
variance	0.0007	0.0005	0.0004	0.0005

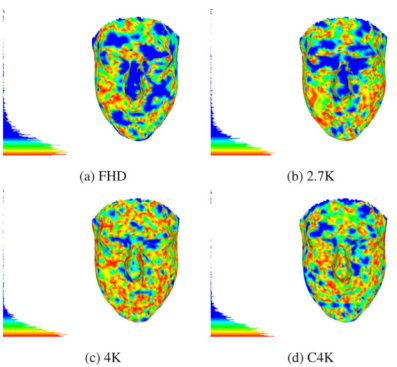
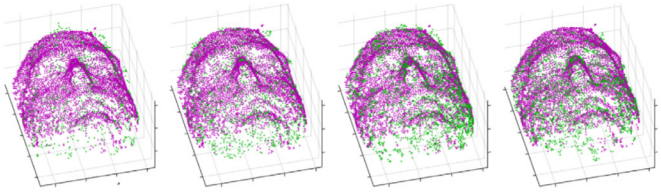


Fig. 6. Hausdorff distance map of Mobile vs. Drone at different video resolutions.

(a) (left to right) Mobile FHD, Drone FHD, 2.7K, 4K, C4K.

(b) (left to right) Mobile FHD vs Drone FHD, 2.7K, 4K, C4K.

Fig. 7. Top: Crops of the 3D faces extracted from the 3D models acquired in controlled conditions by mobile device and by drone at different resolutions. Bottom: Visual results of the overlap between the 3D model from mobile acquisition and those acquired by the drone at different resolutions. When a model prevails on the other, it means the overlap is not satisfactory. The more interleaved the textures, the more accurate the overlapping.



(a) FHD (b) 2.7K (c) 4K (d) C4K

Fig. 8. Point cloud registration of mobile vs. drone at different resolutions

co-registered. Similarly to what happens in Fig. 7b, where the overlapping error is low if the two textures are interleaved, also in Fig. 8 the point clouds are well co-registered when they both contribute to colorize the plot. It can be observed that the overlap is the highest at 4 K drone resolution. The same evidence comes out from the comparative results in Table 2, where the 4 K resolution achieves the lowest mean RMSE error and mean variance.

Fig. 6 shows the distance maps obtained by comparing the ideal 3D model from the mobile acquisitions and the one generated by the acquisitions of the drone during flight. In particular, the figure shows how the map varies among the comparison obtained at different video resolutions. The distance metric considered is the Hausdorff distance and the distributions from nearest to farthest pixels is plotted on the left of each map. By analyzing the histograms, it can be observed that the best performing reconstruction is obtained at 4 K resolution.

When the two models overlap, a high concentration of points accumulates at bottom of the histogram (red/yellow color) while the blue points are the farthest. At FHD and 2.7 K, wide surfaces of non ideal overlapping can be noticed. Resolutions of 4 K and C4 K are more or less equivalent, but on careful inspection, the map corresponding to comparison with 4K resolution is more compacted on lowest values, thus resulting in slightly better overlap than at C4 K.

5. Conclusions

3D representations of human faces can add significant information to improve the performance of biometric recognition approaches. In this paper, an automatic method for 3D reconstruction of faces through a single dynamic source is proposed.

The 3D model is computed by using a commercial UAV that autonomously flies along a dynamic trajectory to capture pictures of the face. The images are processed to create a faithful 3D reconstruction, suitable for biometric purposes as well as for any application fields where these models can be used. Since the obstacle avoidance sensors do not allow approaching the subject any closer than 1.5 m, obtaining a geometric representation accurate enough to provide adequate 3D reconstruction quality is a non-trivial task. The proximity sensors can be disabled by the DJI SDK, but it is not recommended for user safety, since the whole procedure is automated. Keeping this limitation active, we showed that an accurate 3D reconstruction can be achieved by drones during flight.

The 3D models of the faces obtained in the wild scenario have been compared with controlled acquisitions in laboratory environment, collected through a common mobile device limiting the camera noise. The co-registration procedure implemented through the Iterative Closest Point algorithm has shown that the models obtained by the drone at different resolutions are in large part comparable with a model obtained in the absence of environmental noise. As expected, the most promising results have been achieved by higher resolutions of the camera embedded in the aerial vehicle. However, from an experimental analysis carried out on 20 participants, the 4K video resolution has achieved a higher accuracy compared to C4 K, with a saving in computing time.

The next steps in this direction will be to solve some Remote Controller performance issues and move most of the 3D model computation directly into it.

Another useful development would be to further specialize the drone for biometric applications. Other physical biometric features could be addressed, e.g., iris or ear. Also, the system could be used to capture soft/behavioral traits, in applications such as gait analysis.

The achieved results show that it is possible to automate capture and storage of face biometric information with the use of UAVs. They also suggest that the proposed approach can make it feasible to monitor wide open-access areas without human intervention, at least during the detection phase.

Technological evolution in terms of computing power and battery consumption will hopefully allow to move all the computational part onboard in real time in the near future. The new sensors available on board with future UAVs will make it easier and faster to acquire 3D models of the face.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20, doi:10.1109/TCSVT.2003.818349.

[2] R. Donida Labati, A. Genovese, V. Piuri, F. Scotti, Toward unconstrained fingerprint recognition: a fully touchless 3-D system based on two views on the move, IEEE Trans. Syst. Man. Cybern. 46 (2) (2016) 202–219, doi:10.1109/TSMC.2015.2423252.

[3] K. Blint, Uavs with biometric facial recognition capabilities in the combat against terrorism, in: 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), 2018, pp. 000185–000190, doi:10.1109/SISY.2018.8524800.

[4] P. Chen, C. Lee, UAVnet: an efficient obstacel detection model for UAVwith autonomous flight, in: 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), 2018, pp. 217–220, doi:10.1109/ICoIAS.2018.8494201.

[5] DJI, Dji phantom 4 pro+, 2020, https://www.dji.com/it/phantom-4-pro.

[6] DJI, Dji mobile sdk, 2018, https://developer.dji.com/.

[7] V. Cantoni, M. Porta, L. De Maio, R. Distasi, M. Nappi, Towards a novel technique for identification based on eye tracking, in: 2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2012, pp. 1–4, doi:10.1109/BIOMS.2012.6345780.

[8] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: a survey, Pattern Recognit. Lett. 28 (14) (2007) 1885–1906, doi:10.1016/j.patrec.2006.12.018. Image: Information and Control

[9] D. Giorgi, M. Attene, G. Patane, S. Marini, C. Pizzi, S. Biasotti, M. Spagnuolo, B. Falcidieno, M. Corvi, L. Usai, L. Roncarolo, G. Garibotto, A critical assessment of 2D and 3D face recognition algorithms, in: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 79–84.

[10] A. Rama, F. Tarrs, J. Rurainsky, P. Eisert, 2D-3D mixed face recognition schemes, Recent Adv. Face Recognit. (2008), doi:10.5772/6398.

[11] S. Lin, M. Bai, F. Liu, L. Shen, Y. Zhou, Orthogonalization-guided feature fusion network for multimodal 2D + 3D facial expression recognition, IEEE Trans. Multimed.a PP (2020), doi:10.1109/TMM.2020.3001497. 1–1

[12] A.D. Bagdanov, A. Del Bimbo, I. Masi, The florence 2D/3D hybrid face dataset, in: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, in: J-HGBU 11, ACM, New York, NY, USA, 2011, pp. 79–80, doi:10.1145/2072572.2072597.

[13] M. WÅodarczyk, D. Kacperski, W. Sankowski, K. Grabowski, Compact: biometric dataset of face images acquired in uncontrolled indoor environment, Comput. Sci. 20 (1) (2018), doi:10.7494/csci.2019.20.1.3020. https://journals.agh.edu.pl/csci/article/view/3020

[14] M. Zollhfer, J. Thies, P. Garrido, D. Bradley, T. Beeler, P. Prez, M. Stamminger, M. Niener, C. Theobalt, State of the art on monocular 3D face reconstruction, tracking, and applications, Comput. Graph. Forum 37 (2018) 523–550, doi:10.1111/cgf.13382.

[15] L. Jiang, J. Zhang, B. Deng, H. Li, L. Liu, 3d face reconstruction with geometry details from a single image, IEEE Trans. Image Process. PP (2017), doi:10.1109/TIP.2018.2845697.

[16] X. Han, H. Laga, M. Bennamoun, Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era, IEEE Trans. Pattern Anal. Mach. Intell. (2019).

[17] R. Tariq, M. Rahim, N. Aslam, N. Bawany, U. Faseeha, DronAID : a smart human detection drone for rescue, in: 2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT IoT (HONET-ICT), 2018, pp. 33–37, doi:10.1109/HONET.2018.8551326.

[18] S. Luongo, M. Di Gregorio, G. Vitiello, A. Vozella, Human machine interface issues for drone fleet management, in: T. Ahram, W. Karwowski, R. Taiar (Eds.), Human Systems Engineering and Design, Springer International Publishing, Cham, 2019, pp. 791–796.

[19] J.R. Cauchard, A. Tamkin, C.Y. Wang, L. Vink, M. Park, T. Fang, J.A. Landay, Drone.io: a gestural and visual interface for human-drone interaction, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2019, pp. 153–162, doi:10.1109/HRI.2019.8673011.

[20] J. Bruce, J. Perron, R. Vaughan, Readyaimfly! hands-free face-based HRI for 3Dtrajectory control of UAVs, in: 2017 14th Conference on Computer and Robot Vision (CRV), 2017, pp. 307–313, doi:10.1109/CRV.2017.39.

[21] J. Nagi, A. Giusti, G.A.D. Caro, L.M. Gambardella, Human control of UAVs using face pose estimates and hand gestures, in: 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2014, pp. 1–2.

[22] J.R. Cauchard, K.Y. Zhai, M. Spadafora, J.A. Landay, Emotion encoding in human-drone interaction, in: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2016, pp. 263–270, doi:10.1109/HRI.2016.7451761.

[23] N. Yao, E. Anaya, Q. Tao, S. Cho, H. Zheng, F. Zhang, Monocular vision-based human following on miniature robotic blimp, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 3244–3249, doi:10.1109/ICRA.2017.7989369.

[24] H.-J. Hsu, K.-T. Chen, Droneface: An open dataset for drone research, in: Proceedings of the 8th ACM on Multimedia Systems Conference, in: MMSys17, Association for Computing Machinery, New York, NY, USA, 2017, p. 187-192, doi:10.1145/3083187.3083214.

[25] L. De Maio, R. Distasi, M. Nappi, MUBIDUS I - multibiometric and multipurpose dataset, in: Proc. SITIS 2019 - The 15th International Conference on Signal Image Technology & Internet based Systems, 2019, pp. 748–753, doi:10.1109/SITIS.2019.00124.

[26] Google, Mobile vision, 2020, https://developers.google.com/vision.

[27] G. Bradski, The OpenCV Library, 2000,

[28] Google, Android studio developers, 2020, https://developer.android.com/studio.

[29] 3Dflow, 3DF zephyr sofware, 2020, https://www.3dflow.net/it//.

[30] A. Watts, V. Ambrosia, E. Hinkley, Unmanned aircraft systems in remote sensing and scientific research: classification and considerations of use, RS 4 (2012) 1671–1692, doi:10.3390/rs4061671.

[31] L. Takayama, W. Ju, C. Nass, Beyond dirty, dangerous and dull: what everyday people think robots should do, in: 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2008, pp. 25–32.

[32] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.