



Article

Design of a Multi-Vision System for a Three-Dimensional Mug Shot Model to Improve Forensic Facial Identification

Samuele Giuliani ¹, Francesco Tosti ¹, Pierpaolo Lopes ², Claudio Ciampini ³  and Carla Nardinocchi ^{1,*} 

¹ Dipartimento di Ingegneria Civile Edile Ambientale, Sapienza University of Rome, 00184 Roma, Italy; samuelegiuliani@gmail.com (S.G.); fra.tosti@outlook.it (F.T.)

² Comando Carabinieri Tutela Ambientale, 00100 Roma, Italy

³ Reparto Investigazioni Scientifiche Roma, 00191 Roma, Italy; claudio.ciampini@carabinieri.it

* Correspondence: carla.nardinocchi@uniroma1.it

Abstract: A traditional mug shot is a front and side view of a person from the shoulder up, taken by law enforcement. Forensic science is exploring the benefit of working with 3D data offered by new technologies, and there is an increasing need to work with 3D mug shots. Among the various available techniques, a multi-view photogrammetric approach achieves the highest accuracy in the shortest acquisition time. In this work, a multi-view photogrammetric system for facial reconstruction based on low-cost cameras is developed with the aims of verifying the performance of such cameras for the production of a 3D mug shot with submillimetre accuracy and assessing the improvement of facial matching using a 3D mug shot over traditional 2D mug shots. The tests were carried out in both a virtual and a real-world environment, using either a virtual or a 3D-printed 3D model. The outcome is a point cloud, which describes the face. The quantitative analysis of the errors was realized through the distances between the mesh of the acquired 3D model and the point cloud. A total of 80% of the points with a distance of less than ± 1 mm was obtained. Finally, the performance on facial recognition of the 3D mug shot is evaluated against the traditional 2D mug shot using the NeoFace Watching software (NeoFACE) with a score increment of up to 0.42 points, especially in scenarios where the suspect is not captured from a frontal view.



Citation: Giuliani, S.; Tosti, F.; Lopes, P.; Ciampini, C.; Nardinocchi, C.

Design of a Multi-Vision System for a Three-Dimensional Mug Shot Model to Improve Forensic Facial

Identification. *Appl. Sci.* **2024**, *14*, 9285.

[https://doi.org/10.3390/](https://doi.org/10.3390/app14209285)

[app14209285](https://doi.org/10.3390/app14209285)

Academic Editor: Mauro Lo Brutto

Received: 3 September 2024

Revised: 3 October 2024

Accepted: 7 October 2024

Published: 12 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: 3D model; mug shot; multi-view photogrammetric system; face matching

1. Introduction

Forensic Face Recognition (FFR) has become a critical tool in investigations due to the proliferation of surveillance cameras or mobile phones that continuously produce trace images. It consists of comparing images representing faces and evaluating the results in a forensic contest. The comparison involves two images, the trace image, which is typically taken under uncontrolled conditions, and the reference image, which is taken under controlled conditions. In the trace image, the identity of the depicted subject is contested or uncertain. In the reference image the identity of the subject is known; this one is the mug shot. A traditional mug shot is a front and side view of a person from the shoulder up, taken by law enforcement agencies. Its primary purpose is to provide a photographic record of arrested individuals for identification by victims, the public, and investigators. In a mug shot, the facial expression must remain neutral, the eyes must be open, and no hair or other objects should obscure the face, as required by police standards for passport photographs. In forensics images are used for different purposes. The ultimate aim of FFR is the evaluative purpose, that is to interpret the result of the comparison between the trace and mug shot of the suspect to present it as evidence in court [1]. As the FFR represents a mean to determine the strength of evidence used in a court of law, it must meet different legal requirement depending on the country. The specific guidelines regarding procedures, quality principles, and approaches, are those provided by

the European Network of Forensic Science Institute (ENFSI) and by the Facial Identification Scientific Working Group (FISWG) to guarantee the reliability of the analysis process. Both ENFSI [2] and FISWG [3] define all the workflow of FFR and also which method may be used for image comparison. Among them, the most recommended is the morphological analysis, whereby facial images are compared feature by feature methodically using a check list. Currently, 2D mug shots are used in FFR, but the increasing number of surveillance (CCTV) cameras is creating a large repository of trace images that need to be matched against millions of existing mug shots. The comparison with traditional front and side mug shots is complicated by the fact that CCTV cameras mainly capture images from above. This is one of the problems that the investigator encounters: the images of the anonymous person are captured by cameras with significantly different angles than those of the conventional mug shot (front and right profile). For this reason, facial comparison may not be possible (ENFSI) or may even produce false negatives. Figure 1 [2] shows the variation in the appearance of the same individual at different camera angles.

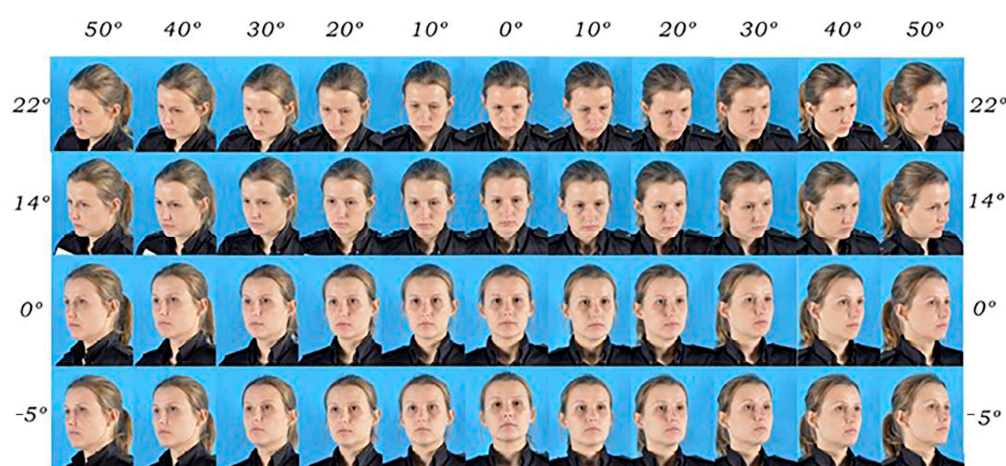


Figure 1. Variation in the appearance of the same individual at different camera angles (from [2]).

New research in the field of the forensic science is exploring the benefits of new technologies and the use of 3D tools in the analysis, interpretation, and presentation of forensic data is increasing in the criminal justice system [4]. As a result, the evolving needs of the forensic community are moving away from traditional portrait-based facial recognition methods [5] toward a 3D mug shot. Unlike 2D mug shots, which are fixed in perspective, 3D mug shots are flexible in perspective, improving comparison according to the investigators' needs. Nevertheless, the creation of a 3D mug shot is a more challenging process than that of a 2D one. This is due to the necessity for highly precise 3D reconstruction to prevent the distortion of the subject's facial features.

The FISWG have developed a detailed facial features checklist document to produce standardization in examinations. Every facial component is described by means of characteristics and descriptors such as size and shape. Moreover, it specifies that, with the term distance, the precise value of a dimension is not the intended determinant, but rather the relative size of this dimension with reference to a morphological analysis. Figure 2 [3] shows examples of alterations to the position among facial components and the effect those position have on the overall face composition. It is important to emphasize that the respect of proportions required by the FICWG guidelines is met if the reconstruction minimizes differences from the real face.

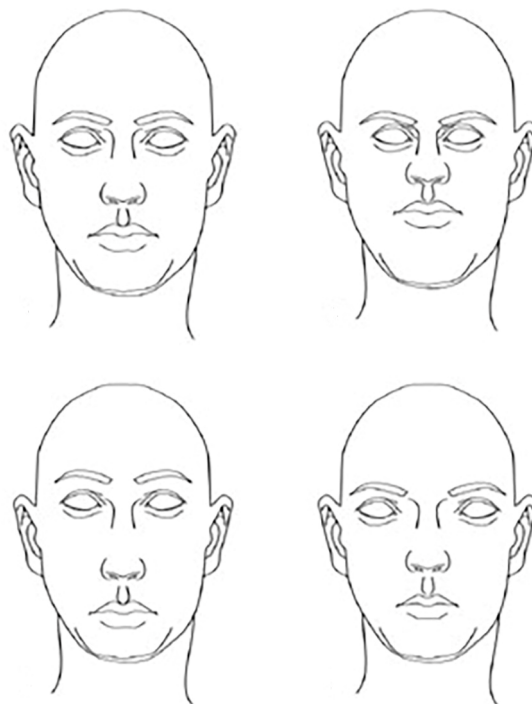


Figure 2. Examples of alteration to the positions among facial components and the effect those positions have on the overall face/head composition (from [3]).

According to this requirement, it can be assumed that a 3D face reconstruction (3DFR) that is closer to the subject will produce a 3D mug shot that is close to the real one. It is difficult to define the required accuracy, relative or absolute, because, to the best of our knowledge, we have not encountered any explicit references to establish an error threshold, but at this preliminary stage, a threshold of no more than 1 mm error seems reasonable. Techniques using laser or structured light have achieved sub-millimetre accuracy in the production of 3D models. However, the long acquisition time makes them less suitable for producing a 3D face model in a non-cooperative scenario, such as that of an arrested person. In [6], Schippers et al., 2024 compare a stereo photogrammetric system (3dDM) with two handheld structured light scanners, the Artec Eva and the ArtecEva Spider from Artec 3D. The acquisition time is 20 s for the first system and 60 s for the other one against 0.15 ms for the photogrammetric system. A short exposure time is achieved through camera settings like shutter speed. Typically, fast shutter speeds, such as 1/250th of a second (4 ms) or faster, will freeze and capture quick-moving action, resulting in a clear image of a subject that would otherwise be blurred.

In recent years, 3D facial reconstruction technology based on imagery has made rapid development. Systems for 3D face reconstruction (3DFR) based on imagery are currently being developed by several researchers [7–9]. More recent approaches are those presented in [10] where Zhang et al., 2024 summarize the most relevant algorithms for 3D facial reconstruction based on a single image or two images as the frontal and lateral images of a conventional mug shot. More specifically, methods based on a 3D Morphable Model and deep learning have attracted research interest. Zhang et al., 2024 also describe the available public dataset that allows realization of a quantitative analysis with respect to a ground truth and a comparison between methods. Reconstruction algorithms may be based on mug shots or sequences of video frames, and their aim is to improve the matching against an existing mug shot database. In [11] La Cava et al., 2022 provide a comprehensive and up-to-date review of the state of the art in 3DFR algorithms for forensic application. Algorithms suitable for forensic applications should satisfy constraints leading to the legal validity of the conclusion during a lawsuit or in the investigation phase. The 3DFR approaches were divided into two main groups based on whether they are evidence-

based, such as CCTV frames or based on a 2D mug shot or photographs—the so called model-approach. The approaches were evaluated against the essential requirements of a forensic system, including robustness to facial ageing and pose variation, robustness to occlusions, use of facial scars and marks, and adherence to biometrics characteristics. The model-based approach allows the strategy of introducing a gallery of various predefined poses in the 2D domain to enhance the representation capability. The author analyses the 3DFR obtained from the 2D mug shots (frontal and lateral view) or from one or multiple images. The conclusion of La Cava et al., 2022 is that a rigorous photogrammetric approach based on a large number of images is the recommended solution to ensure that the criteria are met in a way that is both effective and reliable, specifically in terms of high quality of biometric characteristics. Photogrammetry uses two or more cameras to extract 3D information about the target object, such as a face [12]. The output is a point cloud that can be used to generate a 3D model of the object. Texture information (RGB values) can be associated with each point to produce a photorealistic model. The use of a large number of synchronized cameras is essential to capture every part of the face redundantly due to the complexity of facial geometry. A large number of overlapping images is mandatory not only to guarantee the coverage of the whole face but also to optimize the image orientation and image alignment. Photogrammetric systems for facial or human body reconstruction are effectively applied in applications in the entertainment industry and in the medical field. The commercial company, Xangle Studio [13], provides 3D human body reconstruction for film visual effects and games. It uses about 100 full-frame cameras for head reconstruction and about 200 full-frame cameras for body reconstruction. 3D stereo photogrammetric imaging systems are increasingly used in clinical and research settings for facial surgery and rhinoplasty [14]. The 3dMD System [15] is one of the most widely used imaging systems currently on the market [16,17]. In this system, a random light pattern is projected onto the subject while precisely synchronized cameras capture images from different angles according to an optimal configuration. The accuracy reported by the manufacturer is 0.2 mm. In [6], Schipper et al., 2024 verify the system reliability of the 3dDM system by scanning the head of a mannequin and the faces of healthy volunteers several times, achieving a mean error of 0.23 mm on a few reference distances. The Botscan system by Botspot [18] uses 70 synchronized DSLR cameras to simultaneously capture images of a standing person for 3D reconstructions. In [19] (Michuenzi et al., 2018) compare measurements acquired on a 3D model realized with the Botscan system and Agisoft software against the measurements extracted from forensic photographs. The measurements obtained by photogrammetry were significantly more accurate than those obtained by standard forensic methods based on 2D mug shots, with mean differences of 1.5 mm compared to 3.6 mm. In [20], Leipner et al., 2019 present a 3D mug shot system specifically developed for forensic identification using a photogrammetric approach. The system comprises 26 digital single-lens reflex (DSLR) CANON EOS 80D cameras arranged in a semicircle with a radius of 1.46 m. The model is scaled with a single reference distance, and its validation is primarily focused on capturing the morphological facial features analyzing different focal distances. Full-frame cameras, known for their high performance and high-resolution characteristics, ensure accuracy, minimal noise, and distortion, especially when paired with top-quality lenses. The primary goal of this research is to evaluate the potential of using low-cost cameras to achieve submillimetre accuracy in reconstructing 3D facial models. In collaboration with the Scientific Investigations Department (RIS) in Rome, we conducted a pilot study, developing a multi-view system utilizing high-resolution Raspberry sensors [21]. These sensors offer a significant cost advantage compared to full-frame cameras. The system's performance was assessed by acquiring a 3D model and analyzing the point-wise reconstruction error.

2. Materials and Methods

Before delving into the methodology, it is crucial to understand the entire photogrammetric process involved in producing a 3D mug shot. An object, defined within an estab-

lished 3D world frame, is captured using two or more images. Each image is acquired from a unique camera position, defining a 3D camera frame with its origin at the camera’s projection centre and its z-axis aligned with the optical axis of the camera. For each camera frame, the three coordinates of the projection centre in the 3D world frame and the three rotations needed to align the 3D camera frame with the 3D world frame define the orientation parameter of the six cameras. The external parameters establish the relationships between the 3D world frame and the individual camera frames. The photogrammetric process can be divided in three stages. The first stage is the acquisition step (Figure 3a), where each image is captured from a different camera location according to a predetermined network design that defines the orientation parameters, T , of each camera. To ensure accurate facial reconstruction, the images should provide full coverage from ear to ear, with at least a double coverage for every facial feature. Additionally, cameras must be synchronized to capture all images simultaneously in a single shot minimizing motion blur caused by subject movements.

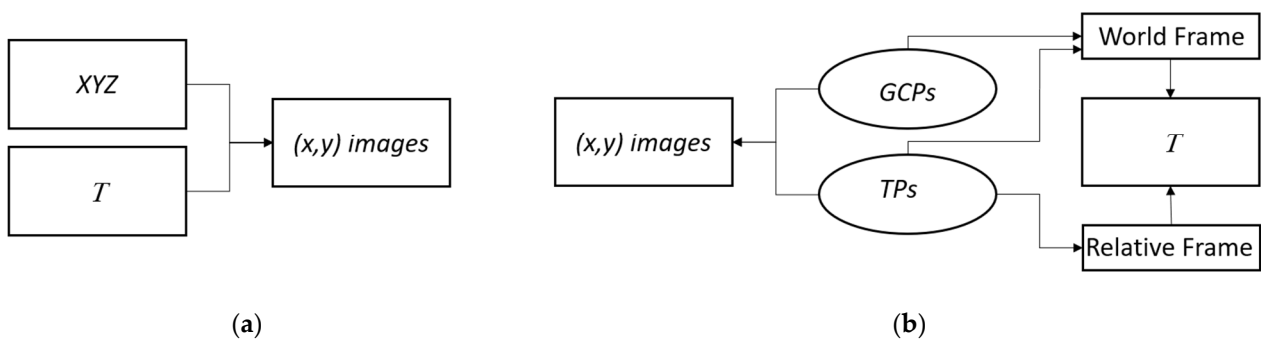


Figure 3. The photogrammetric stage: (a) acquisition stage; (b) orientation stage.

The second stage (Figure 3b) involves the orientation process, where the orientation parameters, T , are estimated using a set of points that allow the writing of projective relationships between the images. Key points for image alignment are primarily tie points (TPs), which connect images sharing common object features. Automatic methods for detecting corresponding interest points between images often generate a very large number of TPs. Ground control points (GCPs), in addition to tie points (TPs), are a limited set of points with precisely known coordinates in an external reference frame. For high-accuracy surveys, these points are often marked and can be located in images with sub-pixel precision. The knowledge of TPs is crucial for the second stage of the process and is derived entirely from the images. This step can be approached in two different ways, depending on whether we are working solely with TPs or with TPs and GCPs. In the first case, the object will be reconstructed in a relative frame, whereby the orientation parameters of all cameras are determined relative to each other (relative orientation). The resulting 3D model is out of scale and not oriented. In the second case, the knowledge of the GCPs provides a 3D model that is both scaled and oriented (external orientation). The third stage (Figure 4) involves object reconstruction based on the acquired images and known orientation parameters. This process relies on a dense matching algorithm and generates a point cloud representing the object either in a relative frame or in the world frame.

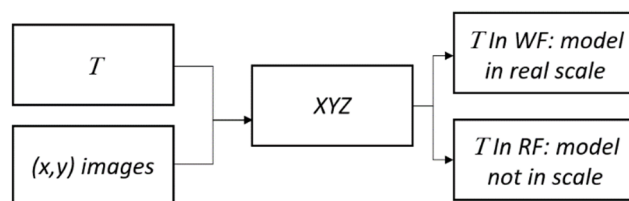


Figure 4. The photogrammetric stage: reconstruction stage.

As in any photogrammetric survey for automated surface reconstruction, image orientation is a fundamental aspect of the process. To improve the robustness of the multi-view system in determining orientation parameters, a higher coverage than double is recommended. Additionally, accurate face reconstruction must account for intrinsic facial movements. This necessitates very short image acquisition times, achievable only with multiple synchronized cameras.

Our study presents some preliminary results of a feasibility study for a low-cost photogrammetric system for 3D face reconstruction in forensic, aiming for submillimetre accuracy. The low-cost camera we tested is a Raspberry, which is particularly interesting because it supports C-mount lenses. The main challenge with low-cost cameras is their lower signal-to-noise ratio due to smaller sensors. Additionally, smaller sensors limit the field of view, which is another obstacle to overcome. The price ratio between a Raspberry and a DSLR is approximately 1/10. This is crucial in the forensic field where there is a great interest in producing a more affordable system that guarantees the same accuracy. As an example, in Italy, only the 'Arma dei Carabinieri' has a network of at least 500 systems distributed throughout the country for producing traditional mug shots. The system was tested in two phases: in a virtual environment and in a real-world setting. In both phases, we conducted a quantitative analysis of the reconstruction error using a 3D virtual model that is our ground truth. This reference model was captured directly within the virtual environment, allowing a point-by-point measurement of the reconstruction error between our reconstructed model and the ground truth.

It is important to note that in the real-world setting, we captured a 3D printed version of the reference model using the Raspberry Pi cameras. The virtual model was printed using a Stratasys J750 by Stratasys [22] 3D printer. At full scale, this printer has a declared accuracy of up to 0.2 mm for rigid material, which is negligible compared to our acceptable error range of less than 1 mm. Figure 5 shows the virtual model on the left, some details of its meshes in the middle, and the printed model on the right. Reconstruction errors were then calculated by measuring the point-by-point distance between our reconstructed model and the virtual ground truth. Therefore, the reconstruction error of the real-world test also introduces the intrinsic printer error. Reconstruction errors are visually represented as a colour map overlaid onto the 3D model, providing a clear visualization of error distribution across the facial surface.

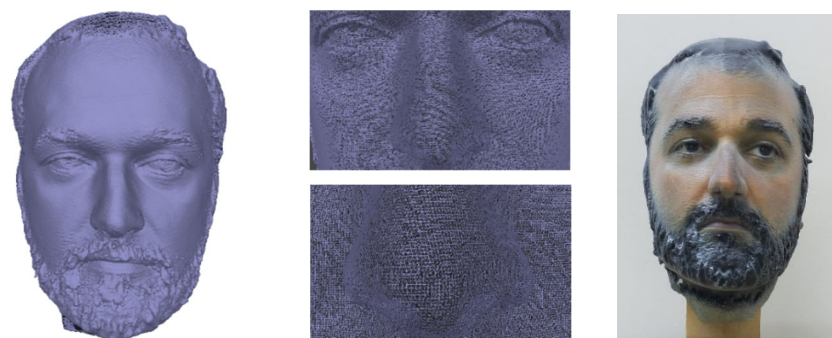


Figure 5. The virtual model on the (left), details of the virtual model in the (middle), and the printed model on the (right).

Virtual test aims to anticipate some practical aspects of the study by realizing an investigation into the accuracy and completeness of the body model reconstruction, as a function of camera network configuration and of camera orientation. The virtual environment was created in 3DStudio Max 2022 of Autodesk [23]. The first test was carried out using the true camera orientation parameters to produce the 3D face reconstruction. It served as a crucial benchmark for the following tests. Subsequently, we investigated the two photogrammetric orientation approaches and the camera configurations to determine which workflow most closely approximated the accuracy achieved with true orientation parameters.

In the real setting, we transferred the results obtained from the simulation network to evaluate the performance of such a sensor. We had only three Raspberry Pi cameras, which were moved along the acquisition position. Three cameras were less than ideal for our multi-view photogrammetric system. However, as we were capturing a static object (the 3D printed 3D model) camera synchronization was not essential. Synchronization becomes critical when capturing images of a real person, especially in non-collaborative context. Despite the lack of synchronization among all the frames, since we took all the image in sequence with the relocation of the cameras to cover all the angles, we successfully obtained the 3D models for three volunteers who maintained their positions throughout the acquisition process. The volunteer's collaboration allowed us to verify the performance of dense matching on real subjects, which is influenced by camera resolution. For one of them we produced a complete satisfactory 3D facial model that allow us to test the face matching software (NeoFace Watch ver5.1.3.15) routinely used by investigators to assess potential improvements in matching accuracy compared to traditional 2D mug shots. It is important to emphasize that this last test, which focuses on improving facial recognition, is still in its early stages and we will be presenting preliminary results.

When designing the camera configuration, we considered the following factors: a minimum triple coverage of the face; the angle between adjacent camera below 30° to improve image alignment and subsequent orientation; and a camera-to-subject distance of 1.5 m or less to meet the space constraints of the RIS investigator. The most basic camera configuration that meets these guidelines is a single row of eight cameras arranged in a semicircle with a radius of 1.5 m, covering a 180° arc. The angle between adjacent cameras is 25.74° , the distance between the projection centres of adjacent cameras is 40.39 cm and a field of view of 20° ensures triple coverage. Starting from this basic configuration, we gradually increased the number of cameras to achieve additional facial coverage. This involved adding cameras above and below the initial row, creating both side and lateral overlap. The final configuration consisted of 24 cameras, through the addition of two more rows of 8 cameras positioned half a metre above and below the first row, resulting in a total cluster of 24 cameras. The cameras in the additional rows were tilted 18° along the x-axis, downwards for the 8 cameras in the row above the first and upwards for the 8 cameras in the row below. Figure 6 shows the most basic camera configuration composed by 8 cameras and the complete 24-camera configuration.

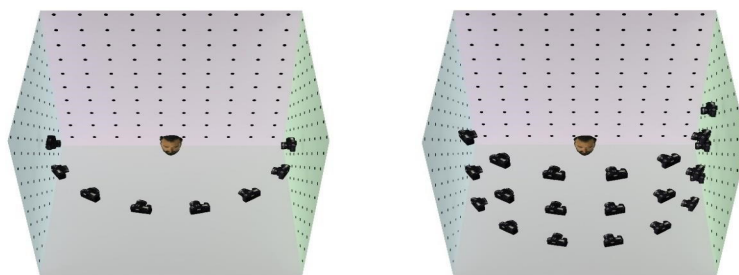


Figure 6. Camera Network Configuration: from 8 cameras to 24 cameras.

We used the software Agisoft Metashape ver.1.7 [24] to solve the image orientation and image matching step. At first Metashape detects correspondences across the photos, then it applies a greedy algorithm to find approximate camera location and refines them later using a bundle adjustment algorithm to obtain an accurate camera position, orientation, and distortion parameter. The image alignment is followed by the dense matching (reconstruction stage) whose outcome is the point cloud of the acquired object. All reconstruction errors were calculated using the Cloud-to-Mesh Distance plugin in the open-source software Cloud Compare [25].

3. Results

3.1. Test in a Virtual Environment

These tests were conducted within the virtual environment of 3DStudio Max. We modelled a 2.0 m \times 2.0 m box and positioned the 3D model (in *.obj format) at its centre. Circular targets of 1.5 cm in diameter were inserted along the wall, used as GCPs. The coordinates of the GCPs established the world frame, with the origin situated at the centre of the box, the xy axes aligned with the floor, and the z-axis pointing upwards. Within the software, we configured the virtual camera parameters (focal length, sensor dimension, and resolution) to match those of our Raspberry Pi camera. Images were acquired within 3DStudioMax. Figure 7 shows the 24 images of the virtual model obtained in 3DStudio Max.

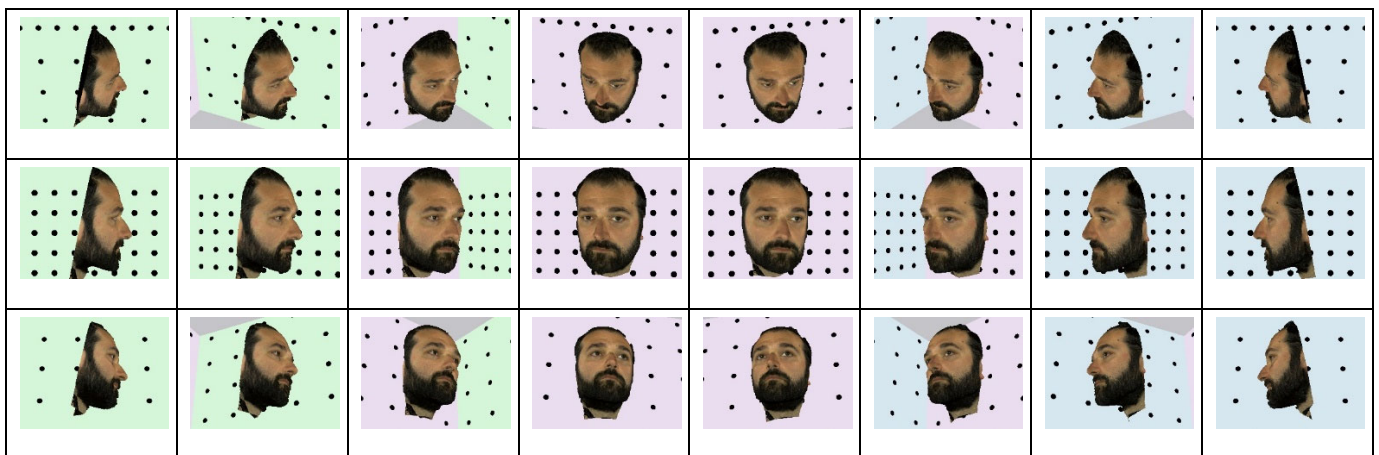


Figure 7. A total number of 24 frames acquired within 3DSMax 2022.

The first row of cameras corresponds to the upwards cameras, the second row corresponds to the frontal cameras, and the last row corresponds to the downward-facing cameras. In the first test, we imported in the Agisoft software the camera orientation parameters set in 3DStudioMax to acquire the images (true orientation parameters) via the open-source industry standard Alembic (*.abc) format. This format allows seamless communication between 3DStudioMacix and Agisoft. Since these parameters were directly imported, reconstruction errors in this test solely depend on the dense matching algorithm and the camera network. We compared the ground truth directly to the resulting point cloud, which was oriented and scaled. No significant difference was observed between using 8 cameras ($sqm = 0.037$ mm) and 24 cameras ($sqm = 0.036$ mm). In both cases, the mean error was zero as expected from a normal distribution of the error. Additionally, there was no significant difference in coverage observed between the two configurations. Figure 8 shows the colour map of the signed reconstruction error within ± 3 standard deviation (σ), superimposed on the 3D model. Reconstruction errors within ± 0.02 mm are represented in green, errors up to 0.1 mm in red, and errors down to -0.1 mm are indicated in blue.

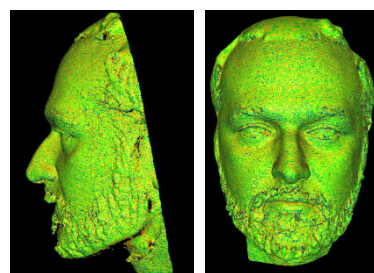


Figure 8. Reconstruction error superimposed on the 3D model. Error between ± 0.02 mm represented in green; error up 0.1 mm in red and error down to -0.1 mm in blue.

The reconstruction error ranges from -0.2 mm to 0.2 mm depending on noise measurement, and it represents the (ideal) limit we can obtain with our system; they are comparable to the intrinsic error of the 3D printed model, and this additionally confirms that we can neglect the intrinsic error of the 3D printer as it did not affect our result.

The next tests show the influence of the orientation parameter in the 3D reconstruction. The images acquired within 3DStudioMax (Figure 7) were aligned by Agisoft using a sparse cloud of TPs. We explored the two photogrammetric approaches: working in a relative frame using only TPs and in a real-world frame using both TPs and GCPs. The world-frame test (the frame defined in 3DStudio Max) allowed us to evaluate both the quality of orientation parameters and the reconstruction errors. An accurate camera orientation is crucial for achieving high-quality reconstructions. To improve the estimation of the camera orientation parameters, we experimented with increasing the number of GCPs. However, we found that this approach did not yield results comparable to those obtained using the relative frame and 24-camera configuration. This observation was supported by the results with true camera orientation parameters, which provide a baseline for understanding the error distribution in a system without the influence of orientation parameter errors, as well as providing reference values for camera orientation. When using eight cameras, we measured positional errors ranging from 1.0 to 1.5 centimetres for all cameras. This is significantly larger than the 1–2-millimetre errors observed using a triple row of cameras. Positional errors were measured as the distance between the true position of each camera's centre of projection and the estimated one. The orientation parameters, calculated as the difference between the three true axes angles and the estimated ones, were significantly less accurate with eight cameras. These cameras exhibited errors between 10^{-1} and 10^{-2} degrees for all three axes, compared to 10^{-3} to 10^{-4} degrees observed with a triple row of cameras. Figure 9 illustrates the errors distribution for the 8-camera configuration (left) and the 24-camera configuration (right). Errors within $\pm 3\sigma$ are represented. Green indicates errors within ± 0.02 mm, red indicates errors up to 0.5 mm, and blue indicates errors down to -0.5 mm.

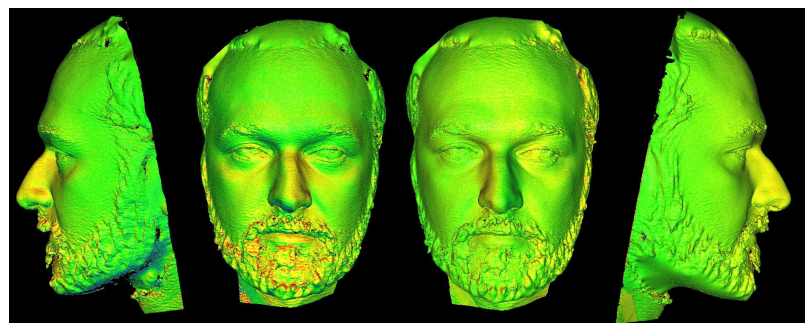


Figure 9. Reconstruction errors. Error up to 0.5 mm are represented in red and error down to -0.5 mm in blue.

The mean distance error was 0.06 mm for the 8-camera configuration and 0.03 mm for the 24-camera configuration. Figure 10 shows the histogram of the errors ranging from -0.5 mm to 0.5 mm. The yellow line represents the reconstruction error for both the 8-camera and the 24-camera configurations obtained using the true camera orientation parameters, serving as our reference for subsequent outcomes. The blue and the red lines display the results obtained by estimating the orientation parameters in the virtual environment using, respectively, 8 cameras and 24 cameras. The black line shows the distribution of the reconstruction error obtained with the full camera's configuration working in a relative frame. The 3D model was scaled within Agisoft using a known distance (scale bar) selected between two background GCPs and aligned by the Iterative Closest Point (ICP) plugin of Cloud Compare. Scaling accuracy was verified by measuring 10 check distances between different pairs of background GCPs, resulting in a mean error of 0.44 mm.

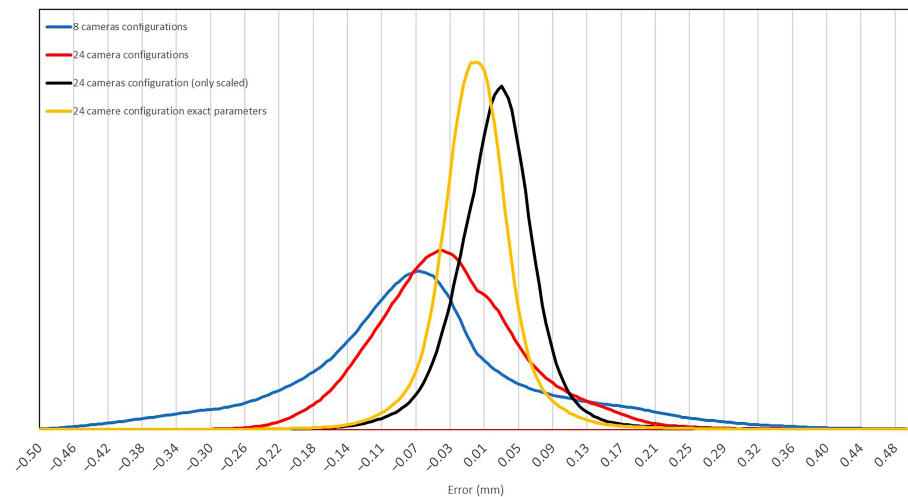


Figure 10. Error distribution in the range from -0.5 mm to 0.5 mm for the following processing: using the project orientation parameter (yellow); using estimated camera parameters and the 8-camera configuration (blue); using estimated camera parameters and 24-camera configuration (red); working in a relative frame and a 24-camera configuration (black).

We tested different setups to minimize the number of cameras and our results indicate that cameras with reduced fields of view, due to their sensor dimensions, require three rows of cameras to create a robust image bundle even in the relative frame.

3.2. Test in a Real Setting

Testing in a real setting was realized by working in a relative frame and using three row of cameras. Due to the practical considerations involving the repositioning of three connected cameras, we added three additional cameras. While theoretically 24 cameras would have been sufficient, the 27-camera network can be seen as a logical extension of our simulated configuration. Minor adjustments to the camera angles, from 25° to 21° (9 cameras in a row), had a negligible effect on the overall result of the virtual tests. We conducted these tests in a real environment using the Raspberry Pi High-Quality Camera (Raspberry HQ), a low-cost camera with a $1/2.3''$ sensor (Figure 11). The Raspberry HQ [17], based on the Sony IMX477R sensor, has a resolution of 12.3 Megapixel (4056×3040). The sensor has a width of 6.287 mm and a pixel dimension of $1.55 \mu\text{m}$. Thanks to the global shutter it can also operate with shorter exposure times down to $30 \mu\text{s}$ given enough light. Unlike previous models of Raspberry Camera models (Module V1, Module V2, Module 3, and Module 3 wide), the Raspberry HQ is compatible with C-Mount and CS-Mount lens. For our test, we selected the CHIOPT fixed lens model FA1610A with a 16 mm focal length. This setup provides a field of view of approximately 22° and a ground pixel size of 0.145 mm at a distance of 1.5 m. The camera's compact dimensions ($38 \text{ mm} \times 8 \text{ mm} \times 18.4 \text{ mm}$) facilitate its integration into an engineered system. The Raspberry HQ connects to the Raspberry Pi via a 15 pins MIPI CSI serial interface specifically designed for the camera connection.

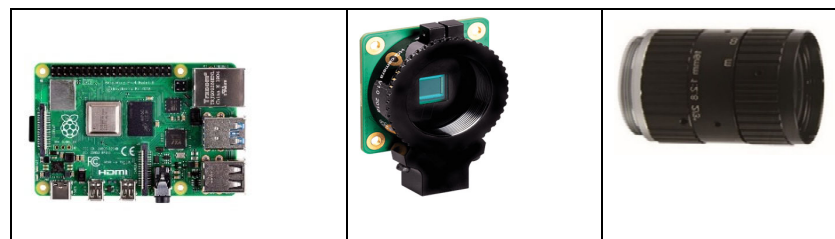


Figure 11. Raspberry Pi, the HQ Camera, and the CHIOPT lens.

The cameras are synchronized via GPIO ports, using a script that captures the images and saves them to external memory. This system can support multiple Raspberry Pi cameras. We created an environment that closely approximates the virtual one.

Figure 12 shows the three synchronized cameras in an environment with an area of $1.5 \text{ m} \times 2.6 \text{ m}$. Behind the object, we placed a surface with dots at a distance of 7 cm to extract the known distances used to rescale the point cloud. The cameras were arranged in a circular pattern at a distance of 1.5 m from the object. The three rows of acquisition were spaced 0.5 m apart, closely following the virtual configuration. To conduct the tests, the three synchronized cameras were moved to nine different positions: three positions for each camera row (two lateral and one frontal). The head of the object was positioned at the height of the middle row. Such small sensors are more sensitive to low light conditions. Therefore, we used a professional umbrella setup behind the cameras, pointed upward, using cold light. The series of tests included the reconstruction of a 3D-printed 3D model, which allowed for the quantitative analysis of the reconstruction errors, as well as the reconstruction of three real individuals who were able to maintain their position for the duration of the nine shots.

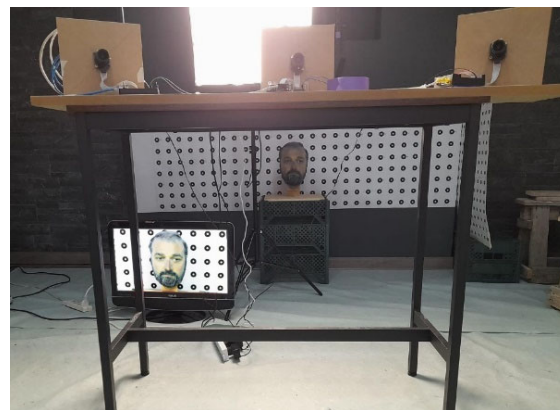


Figure 12. Three synchronized cameras. The up row's three cameras in frontal shooting position.

3.2.1. Tests with the 3D-Printed 3D Model

This work is in its initial stages, and it has been developed utilizing just three synchronized cameras. However, since the object was static in these tests, the lack of camera synchronization did not impact the outcome. Figure 13 displays the 27 images captured during this test. The signalized points in the background were only used to extract scale bars to scale the model.



Figure 13. The 27 images acquired on the 3D model.

Two distinct models were produced: the first scaled using a single distance scale, and the second scaled applying a scale factor estimated through ten known distances (Figure 14).

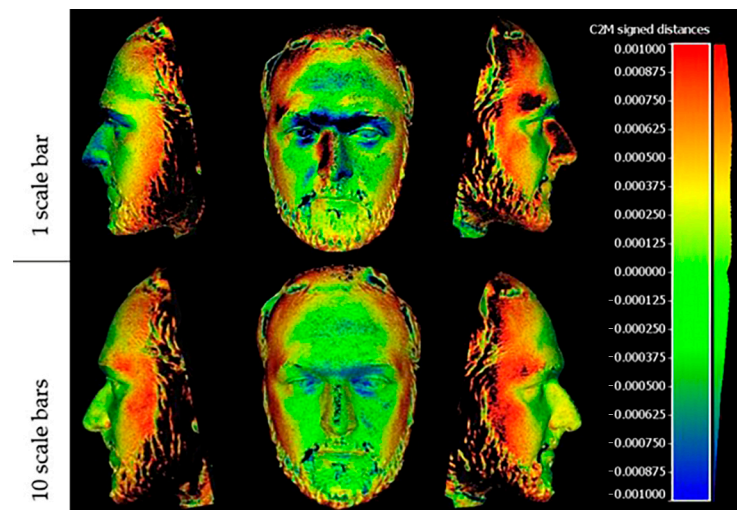


Figure 14. Reconstruction errors. Errors up to ± 1 mm are mapped on the model. The green colour corresponds to a few tenths of a millimetre.

Figure 15 shows the histogram of the reconstruction errors within the range from -5 mm to 5 mm. The blue line plot shows the error of the model scaled with a single distance of approximately 60 cm in length, whereas the orange line plot describes the error of the model scaled with the scale factor estimate using 10 known distances, each with a mean length of 60 cm. The standard deviation of the errors for the model scaled with a single distance is 1.4 mm, whereas it is 0.9 mm for the model scaled using 10 distances. Furthermore, 80% of the points in the model scaled with 10 distances have a distance from the reference model within ± 1 mm, compared to 65% in the model scaled with a single distance.

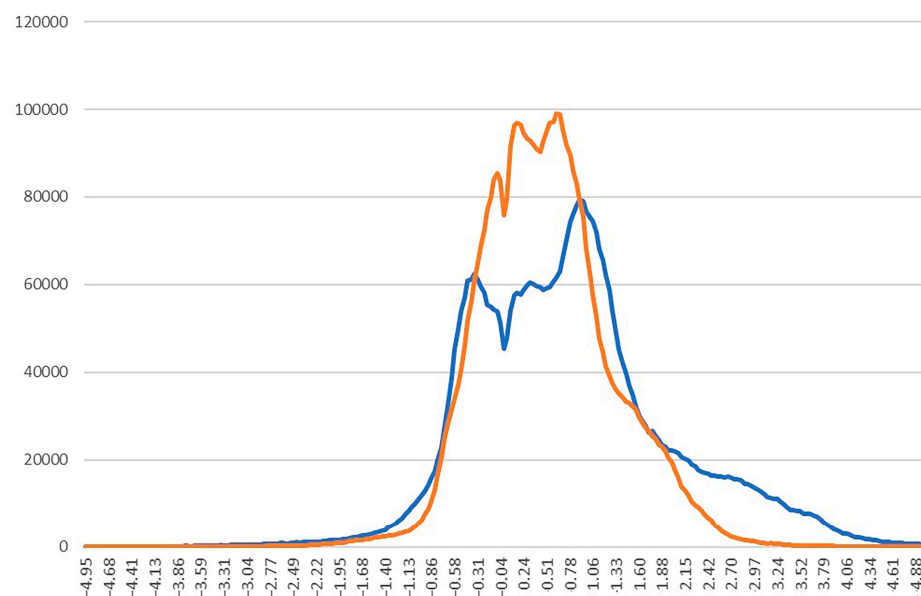


Figure 15. Error distribution in the range from -5 mm to 5 mm for the model scaled with a single distance (blue line) and with 10 distances (red line).

3.2.2. Three-Dimensional Mug Shot vs. 2D Mug Shot

We tested our system with seven volunteers. We are aware of the limitations of this kind of test due to the lack of synchronization of all the frames among all the frames, since we took all the image in sequence with the relocation of the 3 cameras to cover all the angles. Images were taken in groups of 3 , so that any movement of the subject would prevent

the face from becoming reconstructed. Despite that, three out of seven subjects were able to maintain their position for the entire acquisition step that lasted several minutes. This allowed us to obtain images that were suitable for producing a quasi-complete 3D face model. Results could only be analyzed through visual analysis. Figure 16 shows half of the original frontal images with the half-face of the 3D model. For one volunteer, the final subject in Figure 16, who positioned his head against the wall for stability, we were able to reconstruct a complete 3D model. This allowed us to evaluate the performance of facial recognition using the 3D mugshot compared to the conventional 2D mugshot within NeoFace Watching [26] software routinely used by RIS investigators.



Figure 16. Comparison of imagery of half face (on the left) and 3D model of half face (on the right).

This software performs a one-to-one comparison between a trace image and a database of 2D mug shot, producing, as a result, a score from 0 to 1. A value close to 0 represents ‘no match’, while close to 1 confirms a good matching. A score exactly equal to 1 indicates the same person. This is one of the techniques used by the investigators to confirm the identity of a suspect.

The subject depicted in Figure 17 shows the images of the volunteer used in this test. Figure 17a shows the original frontal image, Figure 17d shows the lateral image. Figure 17b shows the textured point cloud produced by Agisoft Metashape (black areas represent unreconstructed points), and Figure 17c shows a lateral view of the 3D model. The texture covers small data gaps but cannot cover a large area without introducing deformation. By comparing the image with the model, we can observe that small details, such as nevi and wrinkles expression, are represented.

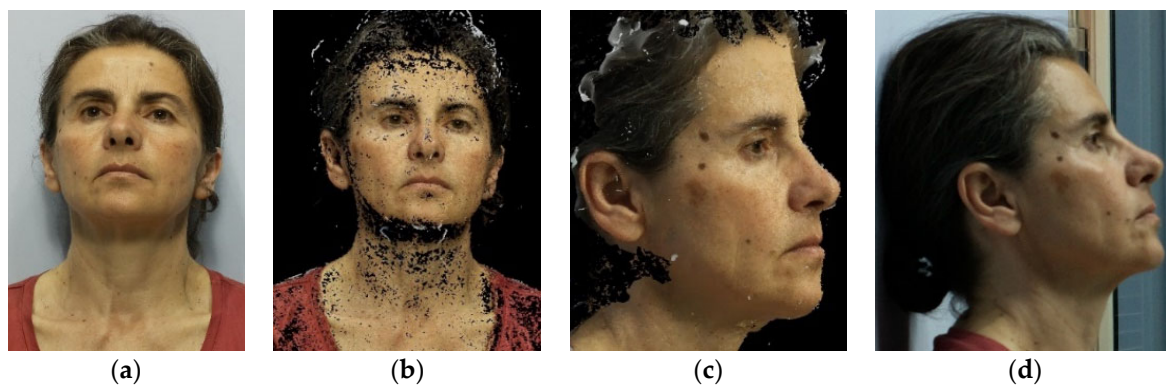


Figure 17. (a) Original frontal image; (b) point cloud, frontal view; (c) textured model, lateral view; (d) original lateral image.

From the 3D mug shot, we extracted fifteen poses in different orientations, as illustrated in Figure 18, and we included them in the mug shot database.



Figure 18. Frames extracted from the 3D model to improve face recognition.

Initially, we sought to find similarities between the frontal image of Figure 17a (used as trace image) and the fifteen frames extracted from the 3D model, as well as the frontal image itself. As expected, we obtained a matching score equal to 1 when comparing the frontal image (Figure 19(1)) with itself and a score higher than 0.97 for four frames extracted from the 3D model (Figure 19(2–5)), confirming the reliability of the produced model (Figure 19). The subfigure 1–5 are the output of the NeoFace Watch software.

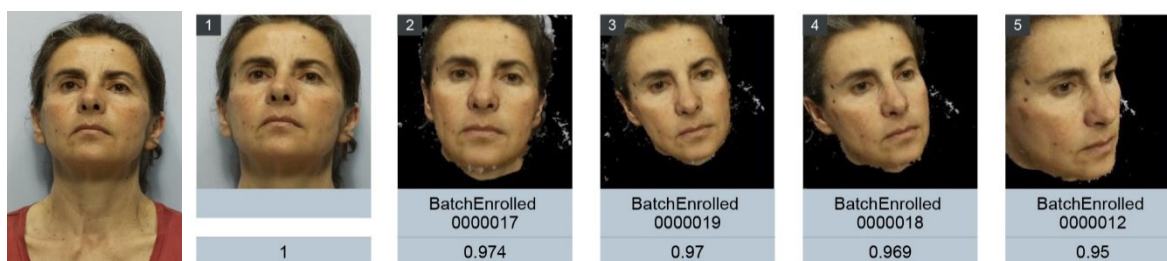


Figure 19. Matching result of the original frontal image against itself and several poses of the 3D model.

To further examine the performance of face matching of a 3D mug shot against the 2D mug shot we extracted 12 frames from videos acquired by two surveillance cameras (Figure 20). The subject is shown in different poses and at varying scales due to the distances from the CCTV cameras. The two video surveillance cameras have a resolution of 1920 × 1080. To select the area of the face, the first eight frames were cropped at a resolution of 686 × 571 pixels, (Figure 20(1–8)) while the last four at a resolution of 592 × 601 pixel (Figure 20(9–12)). Following the investigator’s methodology, for each video frame, we measured the matching score obtained with the conventional 2D mug shot and with the 15 images extracted from the 3D mug.

More specifically, we compared the matching score of each frame obtained with the 2D mug shot to the best score obtained within the fifteen poses of Figure 17. Figure 21 depicts these results. The blue columns show the scores obtained with the poses extracted from the 3D model while the orange columns show the scores obtained with the 2D mug shot. There is a consistent positive difference between the score obtained with the images extracted from the 3D model compared to the 2D images.



Figure 20. The 12 selected frames of the subject.

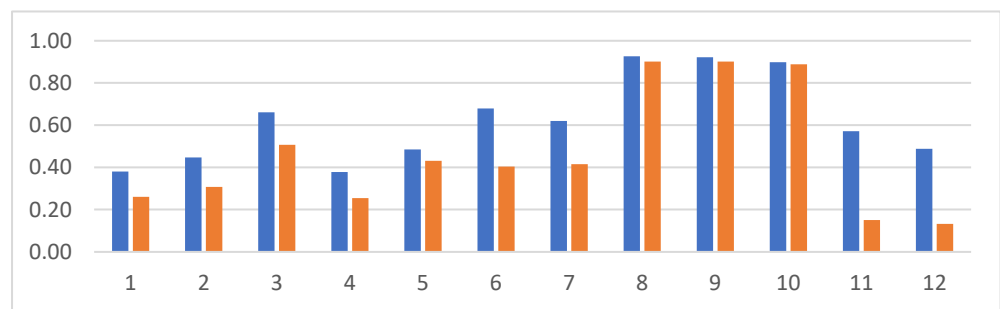


Figure 21. The matching score of each frame obtained with the 2D mug shot (in orange) was compared with the best score obtained within the fifteen poses (in blue).

The mean difference is 0.16 points, with a maximum value of 0.42 points obtained with frame 11, which represents a challenging pose for a conventional 2D mug shot. It can be observed that there is not a significant discrepancy in frontal poses, such as frames 8, 9, and 10, while there is a notable enhancement in the other poses. Furthermore, given that a score exceeding 0.5 points is considered valid for further investigation of the suspect, it can be seen that four or five additional matches were included in the investigations.

4. Discussion

Research in the forensic context is pushing for acceptance of the 3D data offered by the new technology, which must prove the strength of the evidence. Morphological features must be preserved, and accurate metric reconstruction is essential to maintain proportionality. These advances aim to increase the evidential value and acceptance of forensic facial

recognition technology and 3D mug shot within the legal system and, therefore, facial reconstruction for application in a forensic context requires more rigorous approaches. This study proposes a multi-view photogrammetric approach utilizing low-cost Raspberry Pi cameras to create high-precision 3D mugshots with submillimetre accuracy.

We first designed the network in a virtual environment to verify the camera configuration, then we compared the outcomes obtained working in a world frame through GCPs distributed in the environments and in a relative frame. The results of the tests realized using the external orientation parameters set in the virtual environment compared with those obtained using estimated orientation parameters, highlight the critical influence of the external orientation task. It is more prone to error than the alignment in a relative frame. Our results indicate that cameras with reduced fields of view, due to their sensor dimensions, require three rows of cameras to produce a robust image bundle oriented in a relative frame. Tests realized in a real setting emphasize the importance of accurately estimating the scale factor whenever it is necessary to extract metric information from the face. Specifically, we observed that 80% of the points had a distance of less than ± 1 mm when multiple scale bars were used to estimate the scale factor. More specifically, we used 10 scale bars extracted from the GCPs behind the model.

Furthermore, despite the limited number of synchronized cameras, we were able to obtain the 3D models from three volunteers, who were able to maintain their position for several minutes. More specifically, one of them was a complete model, which allowed us to evaluate the improvement in forensic recognition of the 3D mugshots over the 2D mugshots, showing an increase in matching score of up to 0.42 points, especially in scenarios where the subject is not captured from a frontal view.

In reality, the images of a subject to be identified are rarely taken from a frontal or profile view. Surveillance cameras mainly capture people from above and this factor can be decisive in investigations. Therefore, having a 3D model of a suspect's face would allow forensic experts in facial recognition to improve the alignment of the known face with anonymous images captured by CCTV. The study represents an initial feasibility analysis for the design of a multi-vision system for facial reconstruction. Further development, such as acquiring a complete synchronized multi-view system and creating a dedicated environment for system development can only enhance these preliminary results.

Author Contributions: Conceptualization, P.L., C.C. and C.N.; data curation, S.G. and F.T.; formal analysis, C.N.; methodology, P.L., C.C. and C.N.; software, S.G. and F.T.; validation, S.G. and F.T.; writing—original draft, C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due it was required only informed consent.

Informed Consent Statement: Written informed consent has been obtained from all subjects involved in the study to publish this paper.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Maëlig, J.; Champod, C. Automated face recognition in forensic science: Review and perspectives. *Forensic Sci. Int.* **2020**, *307*, 110124. [CrossRef]
2. ENFSI. Best Practice Manual for Facial Image Comparison. ENFSI-BMP-DI-01. Version 01—January 2018. Available online: <https://enfsi.eu/wp-content/uploads/2017/06/ENFSI-BPM-DI-01.pdf> (accessed on 5 October 2024).

3. FICWG. Facial Image Comparison Feature List for Morphological Analysis, version 2.0; 11 September 2018. Available online: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf (accessed on 5 October 2024).
4. Carew, R.M.; French, J.; Morgan, R.M. 3D forensic science: A new field integrating 3D imaging and 3D printing in crime reconstruction. *Forensic Sci. Int. Synerg.* **2021**, *3*, 100205. [[CrossRef](#)]
5. Jain, A.K.; Klare, B.; Park, U. Face Matching and Retrieval in Forensic Applications. *IEEE Multimed.* **2012**, *19*. [[CrossRef](#)]
6. Schipper, J.A.M.; Merema, B.J.; Hollander, M.H.J.; Spijkervet, F.K.L.; Dijkstra, P.U.; Jansma, J.; Schepers, R.H.; Kraeima, J. Reliability and validity of handheld structured light scanners and a static stereophotogrammetry system in facial three-dimensional surface imaging. *Sci. Rep.* **2024**, *14*, 8172. [[CrossRef](#)]
7. Dindaroğlu, F.; Kutlu, P.; Duran, G.S.; Görgülü, S.; Aslan, E. Accuracy and reliability of 3D stereo-photogrammetry: A comparison to direct anthropometry and 2D photogrammetry. *Angle Orthod.* **2016**, *86*, 487–494. [[CrossRef](#)]
8. van Dam, C.; Veldhuis, R.; Spreeuwiers, L. Towards 3d facial reconstruction from uncalibrated cctv footage. In Proceedings of the Information Theory in the Benelux and the 2nd Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux, Boekelo, The Netherlands, 24–25 May 2012; p. 228.
9. van Dam, C.; Veldhuis, R.; Spreeuwiers, L. Face reconstruction from image sequences for forensic face comparison. *IET Biom.* **2016**, *5*, 140–146. [[CrossRef](#)]
10. Zhang, B.; Li, Y.; Li, X.; Liu, D.; Li, Z.; Sun, X. A Review of Research on 3D Face Reconstruction Methods. In *Proceedings of the ICIIT '24: Proceedings of the 2024 9th International Conference on Intelligent Information Technology, New York, NY, USA, 23–25 February 2024*; pp. 70–76. [[CrossRef](#)]
11. La Cava, S.M.; Orrù, G.; Goldmann, T.; Drahansky, M.; Marcialis, G.L. 3D Face Reconstruction for Forensic Recognition—A Survey. In Proceedings of the 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022. [[CrossRef](#)]
12. Kraus, K. Volume 1: Fundamentals and Standard Processes. In *Photogrammetry*; Dummler: Nahe, Germany, 1993.
13. Xangle Studio. Available online: <https://xangle3d.com> (accessed on 24 July 2024).
14. Heike, C.L.; Upson, K.; Stuhau, E.; Weinberg, S.M. 3D digital stereophotogrammetry: A practical guide to facial image acquisition. *Head Face Med.* **2010**, *6*, 18. [[CrossRef](#)] [[PubMed](#)]
15. 3dMD. Available online: <https://3dmd.com> (accessed on 24 July 2024).
16. Lane, C.; Harrell, W., Jr. Completing the 3-dimensional picture. *Am. J. Orthod. Dentofac. Orthop.* **2008**, *133*, 612–620. [[CrossRef](#)] [[PubMed](#)]
17. Aynechi, N.; Larson, B.E.; Leon-Salazar, V.; Beiraghi, S. Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labelling before image acquisition. *Angle Orthod.* **2011**, *81*, 245–252. [[CrossRef](#)] [[PubMed](#)]
18. Botscan by Botspot. Available online: <https://botspot.de/botscan-neo/> (accessed on 24 July 2024).
19. Michienzi, R.; Meier, S.; Ebert, L.C.; Martinez, R.M.; Sieberth, T. Comparison of forensic photo-documentation to a photogrammetric solution using the multi-camera system “Botscan”. *Forensic Sci. Int.* **2018**, *288*, 46–52. [[CrossRef](#)] [[PubMed](#)]
20. Leipner, A.; Obertová, Z.; Wermuth, M.; Thali, M.; Ottiker, T.; Sieberth, T. 3D mug shot—3D head models from photogrammetry for forensic identification. *Forensic Sci. Int.* **2019**, *300*, 6–12. [[CrossRef](#)] [[PubMed](#)]
21. Raspberry. Available online: <https://www.raspberrypi.com/products/raspberry-pi-high-quality-camera/> (accessed on 20 July 2024).
22. Stratsys. Available online: https://pdf.aeroexpo.online/pdf/stratsys-gmbh/j750-spec-sheet/170653-5605-_3.html (accessed on 12 July 2024).
23. 3DStudioMax. Available online: <https://www.autodesk.com/products/3ds-max/features> (accessed on 29 July 2024).
24. Agisoft Metashape. Available online: www.agisoft.com (accessed on 29 July 2024).
25. Open Source Project CloudCompare. Available online: www.cloudcompare.org (accessed on 29 July 2024).
26. NEC. Available online: <https://www.nec.com/en/global/solutions/biometrics/face/neofacewatch.html> (accessed on 16 April 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.