Regular Article

Poliface: A Multi-pose Synchronous Imaging System

Viet-Bac Nguyen[1], Van-Lieu Vu[2], Duc-Toan Truong[3], Viet-Anh Dao[1], Dang-Ha Nguyen[1], Thom Tran Thi[1], Xiem HoangVan[4], Hoang-Anh Nguyen The[1]

[1] Vietnam - Korea Institute of Science and Technology Ministry of Science and Technology, Hanoi, Vietnam
[2] Department of General Administration and Market Development, Nacentech Technology and Business Incubator Center, National Center for Technological Progress, Hanoi, Vietnam
[3] Laser and Photonics Technology Department, National center for laser technology, National Center for Technological Progress, Hanoi, Vietnam
[4] VNU-University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

Correspondence: Xiem HoangVan, xiemhoang@vnu.edu.vn; Hoang-Anh Nguyen The, anhnth@most.gov.vn
Communication: received 31 January 2024, revised 20 July 2024, accepted 25 September 2024
Online publication: 22 October 2024, Digital Object Identifier: 10.21553/rev-jec.361

Abstract– **To enhance the accuracy of face recognition technology in real-world scenarios, it is necessary to train deep learning models on datasets that contain a large number of labeled human face images under multiple poses, lighting, and accessory variations. In this paper, we introduce a novel acquisition system named the Poliface. This system can capture multiple high-resolution images simultaneously around the human head. We designed this system with a well-built aluminum structure, control electronic circuits, and high-performing in-house software. The results demonstrate the precise operation and exceptional stability of this system. Using this Poliface system, we have collected over 6 million photos, which can be used to train and evaluate facial recognition models, and exploited for three-dimensional (3D) virtual face reconstruction.**

Keywords– **Multi-angle face photography, synchronous capturing, photogrammetry, and 3D human face modeling.**

1 Introduction

With the outstanding progress of deep learning technology, facial recognition systems have achieved an impressive 99% accuracy [1–3] on certain well-known databases [4, 5]. Nevertheless, in unconstrained scenarios such as security camera systems (CCTV cameras) or low-resolution camera devices, the accuracy of state-of-the-art face recognition models decreases. To address these issues, researchers introduced face image generation models like TP-GAN [6] and CAPG-GAN [7] designed to produce frontal faces from profile images. These models show benefits in enhancing face recognition accuracy, however, they are heavily reliant on the quality of the training and evaluation datasets.

Facial databases available on the Internet are classified into structured and unstructured types. The former databases comprise facial images that have unevenly distributed face pose angles. Such databases, including LFW [4], IJB-A [8], and VGGface2 [2], are original from online sources. Notably, accurately labeling the pose angles for these databases presents a significant challenge due to their diverse and long-tailed distribution of pose variations. The latter category contains databases such as CMU PIE [9], CAS-PEAL-R1 [10], and CMU Multi-PIE [11], which are great for providing accurately labeled datasets. However, these datasets predominantly focus on the face rotation angle along the vertical axis Oz (yaw angle) and often neglect the tilt angles along the Oy axis (pitch angle). In real-world scenarios captured by CCTV camera systems or mobile devices, objects in frames often have significant yaw and pitch poses simultaneously. This variation decreases face recognition accuracy, especially when dealing with quickly moving subjects or low-light conditions. Hence, high-quality face databases with multi-yaw and multi-pitch variations in different light modes play an important role in terms of enhancing face recognition technology.

Numerous research groups worldwide have undertaken the development of facial image datasets, such as K-FACE [12], V-FACE [13], and M2FPA [14], which adhere to the specified criteria. The M2FPA dataset established in 2019 comprises approximately 520,800 photos collected from 300 volunteers. The data acquisition system employs 62 cameras symmetrically distributed across six image layers, which form a spherical cap. This camera system captures images under seven distinct lighting conditions, combined with four facial statuses (normal, wearing glasses, smiling, surprised).

The K-FACE dataset collected by Kim *et al.* [12] at the Institute of Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology (KIST) contains over 1 million frontal facial images captured at 27 different angles of 1000 Korean individuals. This dataset includes three expressions (neutral, happy, annoyed) in combination with four accessories (mask, glasses, sunglasses, hat) under 35 different lighting conditions. Building on the concept of the K-FACE dataset, the V-FACE dataset encompasses data from nearly 300

Vietnamese individuals, covers seven facial expressions (neutral, happy, sad, afraid, angry, surprised, annoyed), and comprises approximately 3 million photos. Hence, the V-FACE dataset surpasses K-FACE in both qualitative and quantitative aspects.

Nevertheless, the design of the data collection systems in [12–14] predominantly focuses on the frontal face images and ignores the information area behind the head. This ignored information behind the head can be used for generating 3D face models, which aim to support identity recognition technologies. Recent studies demonstrate that 3D facial reconstruction achieved through photogrammetry or a deep neural network increases face recognition accuracy in various scenarios, such as the subject wearing a mask or glasses [15].

The existing methods that reconstruct face models can be generally categorized into two classes. The first one, called Photogrammetry, gathers physical information about objects and the environment through recording, measuring, and interpreting photographic images. Photogrammetry methods require numerous images of target objects from different perspectives and substantial graphic computing resources. In contrast, the other one employs deep learning networks for 3D model reconstruction, which has the advantages of speed and performance. However, the resulting 3D model in this latter method heavily relies on fundamental human face models, such as the Basel Face Model [16], Faces Learned with an Articulated Model and Expressions [17], or Universal Head Model [18].

To enhance and refine the data for these foundational models, we propose a comprehensive facial image data collection system, which obtains facial information and details from the back of the head. Using this system, over 6 million facial photos have been collected under 11 lighting conditions and with 7 different facial expressions [13] at 33 camera positions. This extensive dataset not only contributes to base face models but also advances the associated technologies discussed above and has the following features:

- Large-scale
- Accurate and diverse poses
- High-resolution 2976×1984 (pixel×pixel)
- Inclusion of accessories
- Diversity in terms of gender and age

In summary, the main contributions of this paper are as follows:

- Introduction of a novel multi-angle synchronous image data acquisition system.
- Introduction of a high-quality facial database containing approximately 6 million images of 570 subjects with variations in yaw, pitch, and illumination.

The structure of this paper is as follows. After this introduction, Section 2 describes the structure and main components of the Poliface system, including both the software and hardware development. Section 3 details the image data warehouse, while Section 4 discusses the experiments and modeling results. Finally, Section 5 gives some conclusions and future works.

2 The Poliface System

Figure 1 provides an overview of the structure of the Poliface system, which comprises two distinct parts that collectively cover 180 degrees of the human face and the entire back of the head. Moreover, we develop software that controls different components, such as light control, photo triggers, and cameras, and has a function to label facial data.

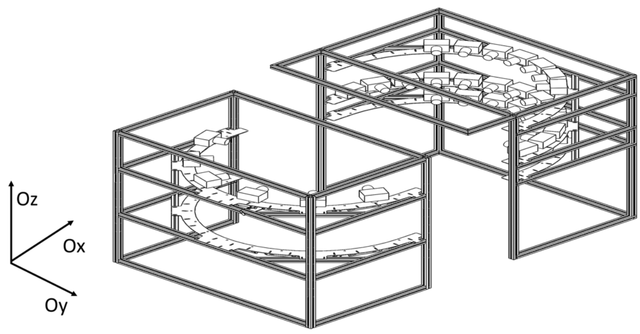

Figure 1. 3D design of the Poliface system.

2.1 Hardware Design

The hardware system includes the following main blocks:

- Skeleton of acquisition system.
- Capture image block: Four camera groups.
- Lighting block: 20 light modules.
- Power block: Three DC 7.5 V sources supplying power to the entire camera; three power sources for the lighting system.
- Control block: Three computers + display screen; control software is installed in each computer, two circuits to control light intensity, and two photo trigger circuits.

The main structure of the Poliface system is built from shaped aluminum bars, which have cross-sectional dimensions of 4×4 $(cm \times cm)$. The front and rear structures have respective width, depth, and height dimensions of $164 \times 150 \times 175$ $(cm \times cm \times cm)$ and $164 \times 100 \times 160$ $(cm \times cm \times cm)$. The aluminum bars were shaped, drilled, and connected with specialized bolts to form a strong structure. To enable mobility, eight load-bearing wheels are mounted at the bottom of the Poliface system. The structure is well built with five hemisphere bars designed for mounting specialized Digital Single Lens Reflex cameras, such as the Canon 1500D and 2000D.

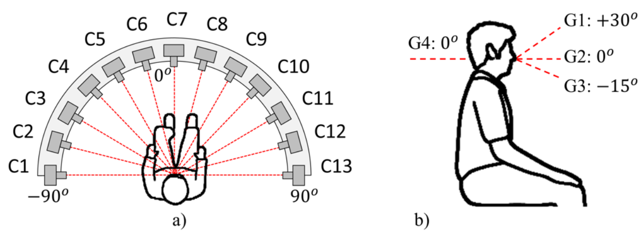

Figure 2. (a) Positions and yaw angles of the cameras in group 2; (b) Pitch angles of camera groups 1, 2, 3, and 4.

The cameras are connected to hemisphere bars using head ball devices. These head balls enable the cameras to be securely positioned on a stand and provide flexibility to adjust the camera angles. For instance, the pitch angle value could be from -45^o to 90^o, and the yaw angle could be from -90^o to 90^o. All 33 cameras are separated into four groups. The front system consists of the first three groups, G1 (C14 – C20), G2 (C1 – C13), and G3 (C21 – C27), which consist of 7, 13, and 7 cameras, respectively. These groups are arranged from top to bottom. In the rear system, six cameras (G4) are labeled from C28 to C33 and installed on a hemisphere bar. Each pair of adjacent cameras of the first three groups forms an angle of 15^o. This design ensures that all cameras direct and focus on the center of the hemisphere bars, as shown in Figure 2a. The camera's pitch angles of groups 1, 2, and 3 are set to 30^o, 0^o, and -15^o, respectively (Figure 2b). Additionally, the hemisphere bar of group 2 has a webcam to display the real-time image of the subject. We change the positions of camera group 4 into 4 cases, as shown in Table I, to optimize the yaw and pitch camera angles when collecting data behind the participants's heads. In the first two cases, six cameras in group four are installed in two hemisphere bars, which have 0^o and -15^o pitch angles. The yaw angle between the adjacent cameras in the first two cases is 45^o and 15^o, respectively. In the last two cases, all six cameras in group four are located in the same bar at 0^o on the vertical axis. However, the yaw angle of adjacent cameras changes from 30^o to 15^o. Moreover, to investigate light impact on the quality of images and 3D facial models of participants, we acquire images systematically and synchronously in different light intensities, such as 1000 lux (L1) and 200 lux (L3).

Table I
Different Camera Positions of Group 4

Case	C28	C29	C30	C31	C32	C33
	((Yaw angle, pitch angle))					
1	(-45,-15)	(0,-15)	(45,-15)	(-45,0)	(0,0)	(45,0)
2	(-15,-15)	(0,-15)	(15,-15)	(-15,0)	(0,0)	(15,0)
3	(-75,0)	(-45,0)	(-15,0)	(15,0)	(45,0)	(75,0)
4	(-45,0)	(-30,0)	(-15,0)	(15,0)	(30,0)	(45,0)

The lighting block comprises 20 sub-modules. Each module includes the LED circuit, protective cover, and light diffuser cover. These light modules are assembled to a steel plate using a U-shaped structure. The U-shaped frame provides the flexibility to adjust the lighting angle from -45^o to 45^o. The power supply diagram for both the lighting system and camera system is shown in Figure 3, which provides a clear illustration of the electrical connections and components involved.

Figure 3 describes the power supply arrangement for the lighting and camera systems. The power source is derived directly from the 220VAC/50Hz power grid system, facilitated by the AC-DC rectifier circuit and the subsequent output voltage stabilization circuit. This

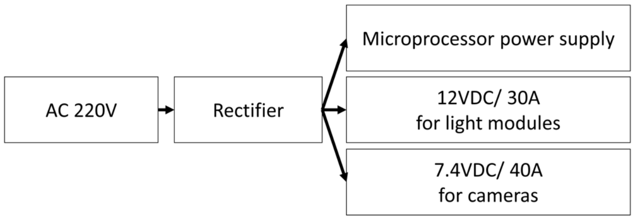

Figure 3. The diagram of the power supply for the camera and lighting system.

DC power source block is categorized into three distinct voltage groups as outlined below:

- Group 1: A specific mention of the +5V supply is made to the programmed microprocessor, responsible for overseeing lighting modes and issuing commands for the camera to capture photos as needed.
- Group 2: The +12V power is allocated for the operation of the LED lighting system.
- Group 3: The +7.4V power source is dedicated to powering the cameras, ensuring their proper operation.

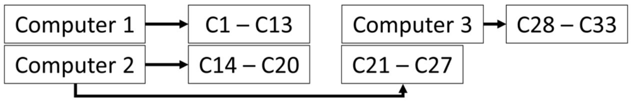

Figure 4. Camera groups are controlled by computers.

These power components are designed and manufactured with fuses and output voltage feedback circuits. These safety features help protect the lights, cameras, and microprocessors in case a short circuit happens or overload during system operation. The Poliface system has computers 1, 2, and 3, which connect to camera group 2, camera groups 1 and 3, and camera group 4, respectively (Figure 4). All cameras connect to the computers using USB 3.0 Type A to Mini USB Type B cables.

The light intensity control circuit, depicted in Figure 5, receives port signals from the computer and enables it to alter the voltage according to the setting parameters for each light module. This flexibility allows the fine adjustment of the light intensity for each light module. The central processor serves as the essential part of the Poliface system, receives control data from the computer software via the RS232 protocol, decodes

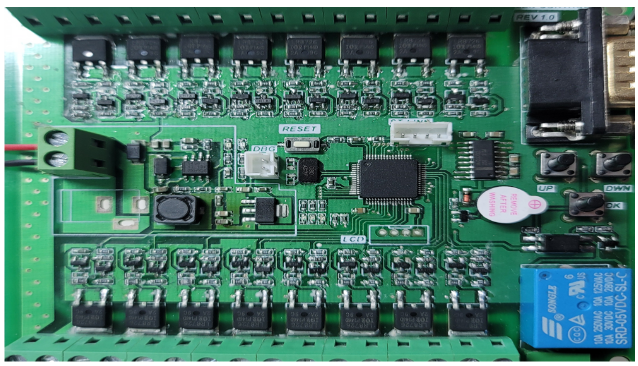

Figure 5. The electronic circuit controls the lighting system.

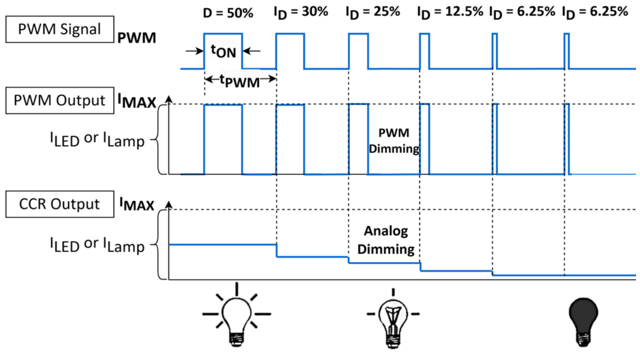

Figure 6. Dependence on the light intensity of the lamp occurs in the pulse width modulation method.

these signals, and subsequently generates control signals for the power stage. The power stage regulates the current supplied to the lights and adjusts the light intensity to the desired levels by manipulating the width of the control pulse width modulation (PWM). This method is illustrated in the diagram below (Figure 6). The formula for calculating duty in the PWM method is as follows

$$D = \frac{t_{on}}{t_{pwm}}, \tag{1}$$

where:

t_{on}: is the active pulse time.

t_{pwm}: is the total PWM pulse cycle time.

The electronic circuit triggers the image capture process and receives signals from the computer software through the USB to COM connection, as illustrated in Figure 7.

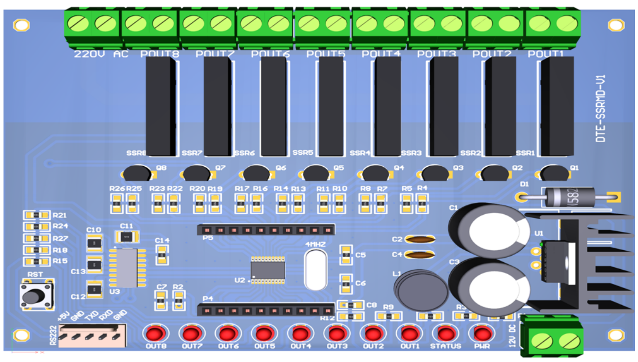

Figure 7. The 3D design of the electronic circuit enables simultaneous capturing of images.

The photo capture command from the software is transmitted to the central processing circuit using the RS232 connection method. The digital signal acquired from the computer software undergoes analysis within the central processing circuit and then generates a control signal at voltage levels supplied to the signal divider circuit. The post-processed signal is directly sent to all cameras to initiate the capture process, as depicted in Figure 8. The processing steps are in the following sequence:

- Step 1: The software receives control signals and transmits capture control commands to the central processing circuit via the RS232 protocol.
- Step 2: The central processing circuit receives the command, verifies the correct capture signal, and subsequently triggers the capture signal through the signal divider circuit.
- Step 3: The signal splitter circuit divides the signal into photography signals, activating the simultaneous shooting mode for all cameras within the system.

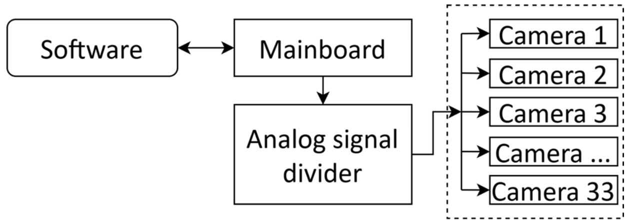

Figure 8. Diagram depicting the operation of the simultaneous imaging trigger circuit.

The number of photos per session exceeds 11,000 (approximately 20 GB), so optimizing the photo capture and storage process becomes crucial. The research team conducted tests on two high-speed image approaches, as illustrated in the diagram presented in Figure 9.

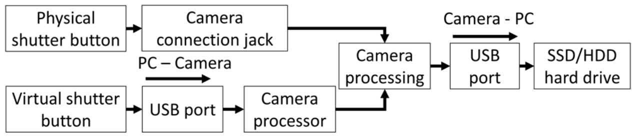

Figure 9. Two processes of high-speed image capture and storage.

The duration for capturing and saving data from the camera to computer memory depends on various factors, such as the number of photos and the method used to send the shooting signal to the camera. The image data acquired from the Poliface system is concurrent. To minimize data storage time, the research team tested the following three approaches: Activating photo capture using command lines (M1), capturing photos using trigger functions integrated with an electronic circuit shown in Figure 7 (M2), and triggering photography using software provided by the KIST Institute combined with a physical button (M3).

Figure 10 shows the time required for capturing and saving 1, 2, and 6 images simultaneously using the mentioned solutions. When saving a single image, all

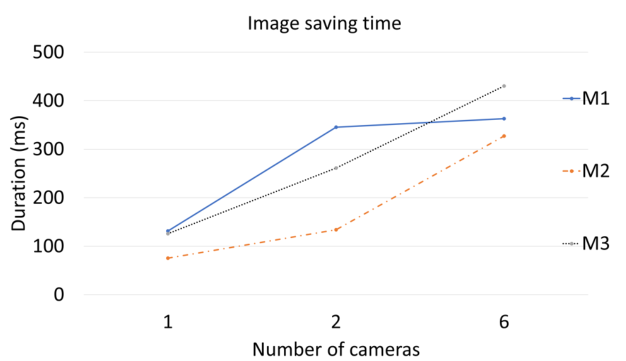

Figure 10. The time requirement for capturing and saving images by using different methods.

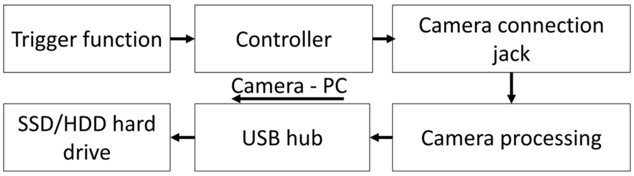

Figure 11. The optimal solution for image capturing with minimal data storage time.

three solutions yield similar results. However, for saving 2 and 6 images simultaneously, option M2 demonstrates considerable time savings compared to options M1 and M3. Based on these findings, the research team integrates command lines and electronic control circuits, as illustrated in Figure 11. This approach is the most effective in reducing the overall time required for capturing and storing images.

2.2 Software Development

2.2.1 Software interface: The software interface of our image data acquisition system is in Figure 12. Within this interface, the primary blocks comprise:

- Camera management block
- Camera parameter control block
- Image label management block
- Light control block
- Image display block

Figure 13 illustrates the main blocks and their specific functions. The camera management block lists the cameras, each identified by its installation location. The camera parameter control block displays settings for lens aperture, shutter speed, light sensitivity, and photo file format. The software interface facilitates the reading and display of specific configurations for each connected camera. Modifications to camera parameters within this module are globally applied to all connected cameras that share those parameters with the computer. The light control module comprises a table enumerating parameters that regulate the system's light intensity. This module allows for direct manipulation of light intensity settings and saves the current parameters into a new data file. The image label management block manages information related to the object number, shooting session number, emotion code, lighting condition code, and data storage location, ensuring that corresponding information for each captured object is accurately saved. The image display block features a screen that displays images directly obtained from the camera specified in the camera management block. Moreover, it includes a screen showing the most recent image captured from the front of the subject and another screen displaying photos taken during the previous session.

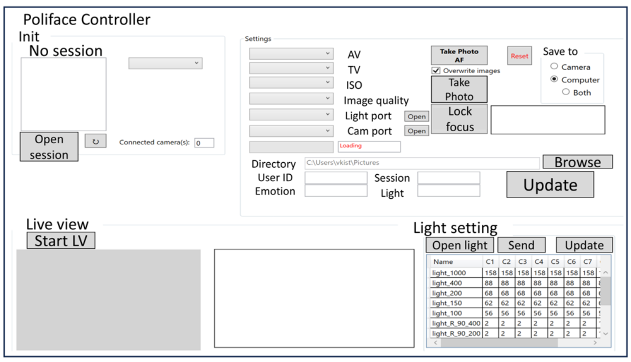

Figure 12. The software interface controls the operation of the Poliface system.

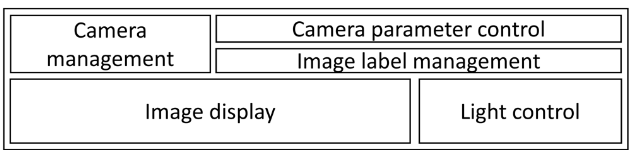

Figure 13. Diagram of main blocks of the software that controls the operation of the Poliface system.

2.2.2 Algorithm diagram: Figure 14 shows the algorithm of the operational process of the software. When the software starts, the program scans all available camera devices linked to the current computer and presents them within the camera management block. For each camera specified in the camera management block, the software retrieves parameters associated with lens aperture, shutter speed, light sensitivity, and photo quality, displaying them in the respective block. The software retrieves light control parameters from the data file saved in the software installation folder. The software automatically prompts the user to declare the settings if the data file is missing. Users are responsible for updating shooting session information, including photo subject number, session code, expression code, and lighting condition code. After verifying all pertinent information, the user can activate the shooting function of all connected cameras simultaneously. Two options are available for this activation: Using the shutter trigger button or employing a central processing circuit combined with a computer program. Utilizing handheld buttons presents disadvantages in terms of device stability. With a substantial number of shots taken over an extended period, it diminishes the trigger's lifespan, necessitating the replacement of buttons every 2-3 months. However, employing a microcontroller circuit in conjunction with software to trigger simultaneous image capture offers a solution to this problem.

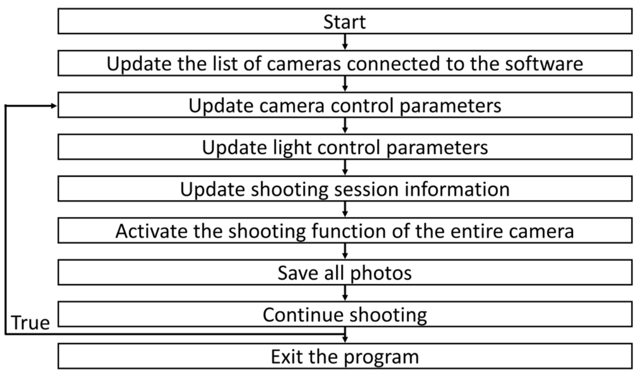

Figure 14. The algorithm diagram of the Poliface system software.

3 Image Data Warehouse

The research team analyzes data from 566 volunteers, which contains approximately 5,883,570 photos. Efforts were made to achieve an equitable gender distribution, with 53.3% male and 46.7% female participants (Figure 15a). The average ages of male and female volunteers are 22.29 and 23.48 years, respectively, and the overall average age is 22.85 years. Figures 15b, 15c, and 15d describe the age distribution.

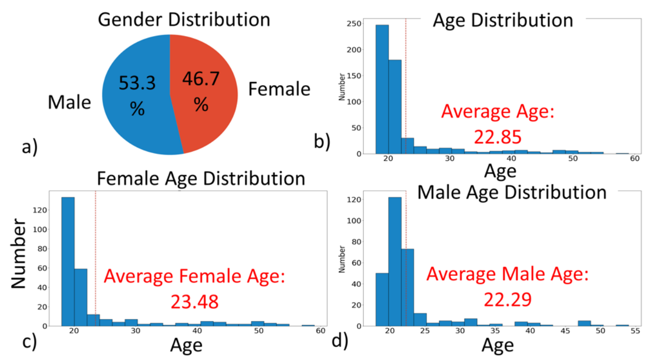

Figure 15. Male/female ratio in the collected data set (a); Average age distribution (b); Average age distribution of females (c); Average age distribution of males (d).

4 Experiment and Evaluation

4.1 Settings

We operate the Poliface system in a dark room, and the subject illuminated by a controlled lighting system during the photography. We cover this system with black fabric to eliminate undesirable features and ambient light. After carefully setting the camera's focus and light modes, we start concurrently taking photos of all cameras. A data sample from all 33 cameras is in Figure 16.

Figure 16. Synchronized photo data sample from 33 cameras.

4.2 Evaluation

To evaluate the performance of the Poliface system, we conducted a random statistical analysis of data from five different experiments. In each experiment, we took photos continuously for approximately an hour to test the stability of the data storage block, camera response, lighting block, and trigger block. The record data in Table II shows that the average time per image capturing of groups 1-2-3 and 4 are 9.8 and 8.4 seconds, respectively, which change slightly. In addition, other blocks work without problems, which means the Poliface system can work long-term with high stability.

Table II
Statistics of Data Obtained from the Poliface System

No.	Camera group	Image's number	Capacity (GB)	Average time per shutter click (seconds)
1	1-2-3	10395	15.5	11
	4	3150	3.6	7
2	1-2-3	10395	15.6	10
	4	3320	3.7	9
3	1-2-3	10395	15.8	9
	4	3242	3.7	8
4	1-2-3	10395	14.7	11
	4	3380	3.8	8
5	1-2-3	10395	15.9	8
	4	3250	3.7	10

4.3 3D Modeling Results

To optimize the camera positions of group 4 for 3D human face model reconstruction, we collect data in four cases mentioned in Table I. Then, we create back heads and face models using camera group 4 and all 33 cameras. As shown in Figure 17, back head models (Figure 17b) reconstructed by lower light intensity images have less information compared with these others (Figure 17a). Moreover, when reducing the yaw angle of adjacent cameras in group 4 from 45^o and 30^o to 15^o, the back head models become more realistic with more details and texture. In particular, the results in cases 2 and 4 show that the more data from cameras at different pitch angles, the better 3D modeling. Therefore, we use 33 images collected in case 2 to reconstruct full 3D facial models in different light modes (Figure 17c and Figure 17d). Based on results shown in Figure 17, the optimized camera position of group 4 is Case 2 in Table I.

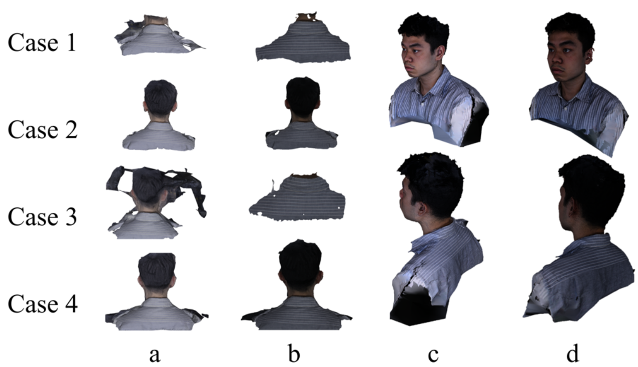

Figure 17. Back head and 3-D human face models reconstruction using 6 and 33 images in different light intensities and camera positions shown in Table I; 1000 lux for Figures a and c; 200 lux for Figures b and d.

4.4 2D and 3D Facial Synthesis

Benefiting from 3D base models reconstructed using the Poliface images, we can remove masks or hair that covers faces in frontal facial images. The results of removing obstructed face masks and hair are in Figure 18b and Figure 18d. Moreover, we can capture frontal faces (Figure 18f) or 2D face view at large yaw

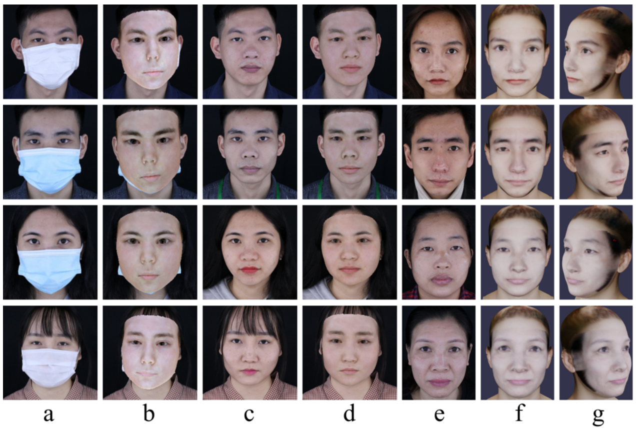

Figure 18. The 2D and 3D facial results are synthesized using a single frontal input image. Columns a and c/e are frontal faces with and without masks, respectively; columns b and d are 2D faces synthesized of the input face of columns a and c, respectively. Columns f and g are 3D face models under frontal and profile views, respectively.

angles from 3D facial models as shown in Figure 18g. These images help to identify people in a database of face recognition systems.

5 Conclusion

In this paper, we assert the design of a 360^o synchronous multi-angle imaging system. Testing data analysis has yielded positive results, which meet the predefined objectives and prove the stability of the Poliface system. The optimization of camera positions is determined to have the best 3D face models. In the future, ongoing research and improvements are necessary to reduce data transmission speed from the camera to local storage. Integrating additional image processing directly into the software, such as face detection, expression recognition, and facial tilt angle calculation (pose estimation), is also considered to improve the performance of this Poliface system. Our facial database, which includes more than 6 million images, is secured and stored in our local computers and online servers so researchers can request access by emailing corresponding authors.

6 Acknowledgment

This research is funded by the Ministry of Science and Technology of Vietnam (Grant No. 02.M02.2022).

References

[1] O. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in *Proceedings of the British Machine Vision Conference 2015 (BMVC 2015)*. British Machine Vision Association, 2015, pp. 1–12.

[2] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A Dataset for Recognising Faces across Pose and Age,” in *Proceedings of the 2018 13th IEEE international conference on automatic face and gesture recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[3] X. Wu, R. He, Z. Sun, and T. Tan, “A light CNN for deep face representation with noisy labels,” *IEEE transactions on information forensics and security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[4] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,” in *Proceedings of the Workshop on faces in’Real-Life’Images: detection, alignment, and recognition*, 2008.

[5] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, “The megaface benchmark: 1 million faces for recognition at scale,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4873–4882.

[6] R. Huang, S. Zhang, T. Li, and R. He, “Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision*, 2017, pp. 2439–2448.

[7] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun, “Pose-Guided Photorealistic Face Rotation,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8398–8406.

[8] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, “Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A,” in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1931–1939.

[9] T. Sim, S. Baker, and M. Bsat, “The CMU Pose, Illumination, and Expression (PIE) database,” in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. IEEE, 2002, pp. 53–58.

[10] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, “The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 1, pp. 149–161, 2008.

[11] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[12] Y. Choi, H. Park, G. P. Nam, H. Kim, H. Choi, J. Cho, and I. Kim, “K-FACE: A large-scale KIST face database in consideration with unconstrained environments,” *ArXiv*, vol. abs/2103.02211, 2021.

[13] H. Duong, T. Tran, V.-B. Nguyen, V. A. Dao *et al.*, “V-FACE: A Large-Scale Vietnamese Face Image Database in Unconstrained Environments,” *International Journal of iRobotics*, vol. 5, no. 2, pp. ˜36–41, 2022.

[14] P. Li, X. Wu, Y. Hu, R. He, and Z. Sun, “M2FPA: A Multi-Yaw Multi-Pitch High-Quality Dataset and Benchmark for Facial Pose Analysis,” in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10 043–10 051.

[15] D.-H. Nguyen, K.-A. Han Tien, T.-C. Ma, and H.-A. Nguyen The, “A New 3D Face Model for Vietnamese Based on Basel Face Model,” in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*. Springer, 2022, pp. 408–420.

[16] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, “A 3D Face Model for Pose and Illumination Invariant Face Recognition,” in *Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. Ieee, 2009, pp. 296–301.

[17] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, “Learning a model of facial shape and expression from 4D scans,” *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 194–1, 2017.

[18] S. Ploumpis, E. Ververas, E. O’Sullivan, S. Moschoglou, H. Wang, N. Pears, W. A. Smith, B. Gecer, and S. Zafeiriou, “Towards a complete 3D morphable model of the human head,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4142–4160, 2020.

Viet-Bac Nguyen received his B.Sc. and M.Sc. degrees in Aerospace Engineering from Hanoi University of Science and Technology in 2019 and 2020, respectively. His research interests include machine learning, deep learning, and data preparation for training AI models. Currently, he is a member of the Information Technology Division at the Vietnam-Korea Institute of Science and Technology in Vietnam.

Van-Lieu Vu graduated with a Master's in 2002 from the Faculty of Electrical and Electronics Engineering at Ho Chi Minh City University of Technical Education, and holds a Master of Science in Automation from HUST. Currently working at NACENTECH under the Ministry of Science and Technology. He has more than 10 years of experience in the field of electronics, automation, laser, and optics research. Main research directions: Design and manufacture industrial laser equipment and medical laser equipment; Designing factory automation systems. Key member of many scientific research projects at ministerial and national levels such as: Research and integrate Nd:YAG pulsed laser equipment to weld metal materials; Research and manufacture integrated 3D engraving equipment system in glass using 532 nm wavelength laser; Research and deploy solid laser equipment with capacity of 400W, wavelength of 1064nm to serve special mechanical processing technology.

Duc-Toan Truong received the B.Eng. degree in Automation Engineering from Phuong Dong University, in 2011, and the M.S. degree in Computer Science from the VNU University of Engineering and Technology, in 2023. From 2012 to 2015, he was a Researcher at the Center for Technology Development and Transfer, National Center for Technological Progress, where he worked on system design and PCB design. Since 2015, he has been with the National Center for Laser Technology, National Center for Technological Progress, focusing on embedded programming and computer programming. His main areas of expertise include system design, PCB design, embedded programming, and software development.

Viet-Anh Dao received a B.S. in Information Technology at VNU University of Engineering and Technology in 2021. My domain researches include computer vision, computer graphics, cloud computing, and deep learning. I participated in many research projects on computer vision and cloud computing at the Vietnam Korea Institute of Science and Technology. I currently work as an AI deployment engineer in the AI research lab of Samsung SDS Vietnam.

Dang-Ha Nguyen received his BSc degree in Computer Science from VNU University of Engineering and Technology in 2022. His research interests include AI, Computer Vision, and Computer Graphics. He is currently a researcher in the IT Department at the Vietnam-Korea Institute of Science and Technology.

Thom Tran Thi received her Engineering degree in Electronics and Telecommunication in 2020 and her M.S degree in Electronics Engineering in 2023, both from Hanoi University of Science and Technology (HUST). Currently, she is working at the Faculty of Communication Engineering, School of Electrical and Electronics Engineering, HUST, Vietnam. Her current research interests are in the application of machine learning techniques to 5G problems, image processing, and computer vision.

Xiem HoangVan is the Head and founding member of the Department of Robotics Engineering, Vietnam National Uni versity– University of Engineering and Tech nology (VNU-UET). He was the former Di rector of the Center for Quality Assurance in VNU– UET (2021– 2022 term). He re ceived a Ph.D. degree from Lisbon Univer sity, Portugal, in 2015, a M.Sc. degree from Sungkyunkwan University, South Korea, in 2011, and BE degree from Hanoi University of Science and Technology, 2009, all in Electrical and Computer Engineering. He has published nearly 100 papers on image/video processing and robotics vision. He is an editor of the Frontiers in Signal Processing Journal and VNU-Journal of Science and reviewed for a number of top IEEE, Elsevier and Springer Journals. He has received a number of prestigious awards including 2023 IEEE RIVF, 2022 REV-ECIT, 2018 IWAIT and 2015 PCS best paper awards, his work on distributed video coding received the 2021 REV-AWARD, 2020 Innova Patent Silver Award (Croatia), and the Fraunhofer Portu gal Challenge Award 2015. He is one of ten young research scientists (one of two in ICT) was awarded the Golden Globe in Science and Technology 2019 and in 2021– 2022 term, he was one of the youngest associate professors appointed by the State council for professorship in Vietnam.

Hoang-Anh Nguyen The was born in Thanh Hoa, Vietnam, in 1986. He reveived B.Sc. from National University, Hanoi, Vietnam in Electrical Engineering in 2008 and M.S. in Electrical Engineering and M. E. in Modeling and Simulation from Old Dominion University, Norfolk, Virginia, USA in 2010 and 2012, respectively. He got his Ph.D degree in Computer Science from University of Engineering and Technology, Vietnam National University, Hanoi in 2020. Currently, he is a researcher at Vietnam – Korea Institute of Science and Technology (VKIST), Ministry of Science and Technology (MOST), Vietnam. Since December 2020, he has led VKIST – Nacentech Joint Lab of VKIST focusing on AI technologies that include Neuroscience and BCI. His research includes Neurotechnology, Deep Learning and Image processing/Computer graphics.