



Data Collection and Image Processing Tool for Face Recognition

Francimar Rodrigues Maciel, Sergio Cleger Tamayo, Aasim Khurshid^(✉),
and Pauliana Caetano Caetano Martins

SIDIA Instituto de Ciência e Tecnologia, Manaus, Brazil
{francimar.maciel,sergio.tamayo,aasim.khurshid,
pauliana.caetano}@sidia.com
<https://www.sidia.org.br/en/home/>

Abstract. Many biometric systems are being used to identify transactions and increase security levels. These systems analyze the different registers that can recognize a person, for example, fingerprint, face, voice, and iris. Face recognition systems are widely studied for security, surveillance applications, transaction, and general services. The accuracy of these systems depends mainly on two closely related factors, quality data and machine learning techniques used. In this paper, we present a data collection and image analysis tool for face recognition with evolved parameters (ergonomic and visual) setting. The proposed tool is capable of collecting face data with various poses while making the user interaction intuitive and comfortable. The details of the different stages of study, along with discussions, is presented based on results extracted from 79 users.

Keywords: Data collection · Face recognition · Robust face data collection

1 Introduction

Many biometric systems have been proposed in the literature to allow access to secure systems [1, 2]. These biometric systems include fingerprint recognition, face recognition, voice recognition, iris recognition, and palm recognition [1, 3]. Face recognition is one of the most studied biometric systems [4]. In the security and surveillance applications, a high recognition rate is mandatory.

Recently, machine learning algorithms have produced high accuracy in face recognition systems [5]. Machine learning algorithms have two building blocks, which are: data and the algorithm. However, machine learning based face recognition methods requires a large number of labeled samples which are expensive and time consuming to collect. The performance of these methods often improves with the amount and quality of the available data.

There are two possibilities to obtain a large amount of data, i.e., Collection of face data from users, and data augmentation from limited available data.

Masi et al. [6] discussed the need of collecting huge numbers of face images for effective face recognition, and proposed an augmentation method to enrich the existing dataset by introducing face appearance variations for pose, shape, and expression. The methods of “one-to-many augmentation” can mitigate the challenges of data collection, and they can be used to increment the datasets [7]. They are categorized into four classes: data augmentation, 3D model, CNN model, and GAN model.

- **Data augmentation:** Data augmentation methods consist of photometric and geometric transformations. Transforms include a range of operations from the field of image manipulation, such as shifts, flips and zooms [7];
- **3D model:** To enrich the diversity of training data, different generic 3D models are used for rendering to augment faces;
- **CNN model:** Rather than reconstructing 3D models from a 2D image and projecting it back into 2D images of different poses, CNN models can generate 2D images directly;
- **GAN model:** Generative Adversarial Network (GAN) is also used for image augmentation, which combines prior knowledge of the face data distribution (pose and identity perception loss). Wang et al. [8] compared traditional transformation methods with GANs to the problem of data augmentation in image classification.

The specific data augmentation techniques used for a training dataset must be chosen carefully considering the context of the training dataset and knowledge of the problem domain. Besides, it can be useful to experiment with data augmentation methods in isolation and test to see if they result in a measurable improvement to model performance, perhaps with a small prototype dataset, model, and training. These techniques are robust but can be computationally intensive.

On the other hand, collecting real-world face database is expensive and time-consuming. However, real-world collected data provide better contextual meaning and allows the classifier to learn efficiently. For this reason, a data collection tool can help improve classification accuracy, especially for small and secure systems. In this paper, we propose a data collection and image processing tool that can be used to collect data for facial recognition. Our proposed tool is evaluated and updated from the feedback of users in three stages. Importantly, this tool allows for capturing 96 facial pose variations while making user-interaction with the system pleasant. Rest of the paper is organized as follows: Sect. 2 describes the stages considered in the data collection process and the main visual and ergonomic parameters. Next, the evolution of the developed tool is presented in detail in Sect. 3, followed by the user evaluation in Sect. 4. The conclusions and future research lines are placed in Sect. 5.

2 Data Collection Process

The user image capturing process may take a few seconds from image acquisition to processing and then subsequent use by recognition system. Some aspects may

affect depending on the task that will be performed, for example, distance with which it was taken to photo and capture angle. Different studies recognize a necessary number of photos (approximately 30) but do not define the values of visual and geometric parameters present in the interaction with users [4,6,7]. Based on this analysis, a data collection tool that is composed of several visual and ergonomic parameters with minimal interaction is proposed that should be evaluated with users for parameters adjusting in any recognition system.

The following parameters are considered in each iteration:

- **Participants height:** This aspect was considered to determine the height of camera from ground for final data collection;
- **Camera height:** The current height of the camera from ground level;
- **Camera angle:** The current angle of camera with respect to horizontal axis;
- **Capture stages:** The stages of user’s image capture process;
- **Yaw angle:** The maximal horizontal rotation of the head by users;
- **Pitch angle:** The maximal vertical rotation of the head by users;
- **Number of images:** The number of face images captured per user;
- **Discarded faces:** The average number of discarded faces per user;
- **Average capture Time:** The average capture time per user;
- **Worst capture Time:** The worst capture time per user.

Mennesson et al. [9] showed that the degree of head yaw rotation is very important for the task of face detection (e.g $\pm 15^\circ$). The authors further commented about how the number of detected faces decreases to zero with a Gaussian decay when user pose is far from the frontal face. Evidently that the maximum of detected frontal faces is obtained with a yaw angle near zero degree (a frontal face).

Visual and ergonomic concepts were studied to facilitate data collection. During the process, three major challenges need to be addressed:

1. To guide the user naturally considering comfort while moving his head;
2. Considering that most of the users are not familiar with face recognition technology, an efficient visual language is necessary to give instructions, when something is not going well;
3. Identify external factors that can influence the quality of experience while using Face Recognition.

The first step was searching parameters that could be used to mediate the human-technology communication. In this regard, three aspects of user interface were observed [10,11]:

1. Physical aspects (operating with a device as a physical object);
2. Handling aspects (the logical structure of interaction with the interface);
3. Subject-object-directed aspects (the mapping of objects “in the computers” with the objects in the real world).

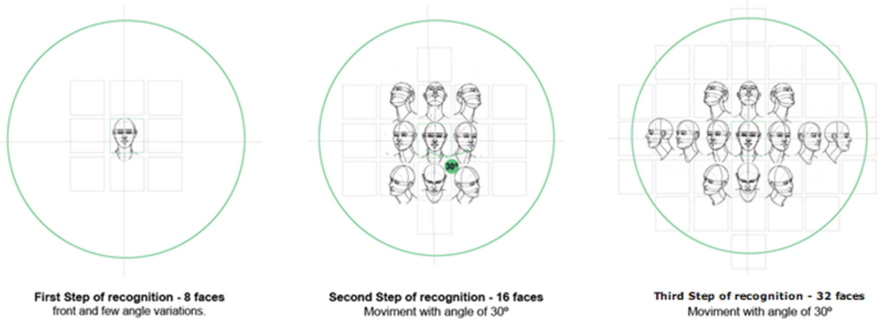


Fig. 1. Second stage - rectangular matrix.

3 Image Processing Tool

In the first stage, users start from a frontal position and perform yaw and pitch movements of the face responding to the text indications received from the device.

In the second stage, the registration process was divided into three steps for a total of 56 faces as shown in Fig. 1. The principal problem in this stage was that the interface didn't provide comfort and freedom for users. The users commented that the process was slow and the matrix interface was artificial and they needed mechanic movements.

In the third stage, the register was divided into two steps for a total of 96 faces. As you can see in Fig. 2 the initial steps of the previous stage were merged. In order to improve the human-computer interaction, after testing with the user, we detected improvement points that were implemented in this phase, for example:

- Facilitate the movements of the head at the capture time;
- To reduce time and effort to capture faces;
- Increase the amount of captured faces.

The strategy tested at this stage showed one critical result. A significant number of users failed to turn their face in the 30° for yaw and pitch angles. This leads to a result that the during the test phase, these angles are not expected to go beyond this limit.

The fourth stage was divided into two steps. The registered number of faces is set to 96, but changes are made in user interaction. In the first step, the register was divided into four quadrants where eight images were taken with 5° of head variation in each quadrant. In the second step, we divide into eight pieces where four photos were captured per quadrant, with a maximum head variation of 15°.

This last strategy provided a more intuitive and comfortable interaction, because by reducing the head angle movement, and the capture time, we promoted more natural movement for the user.

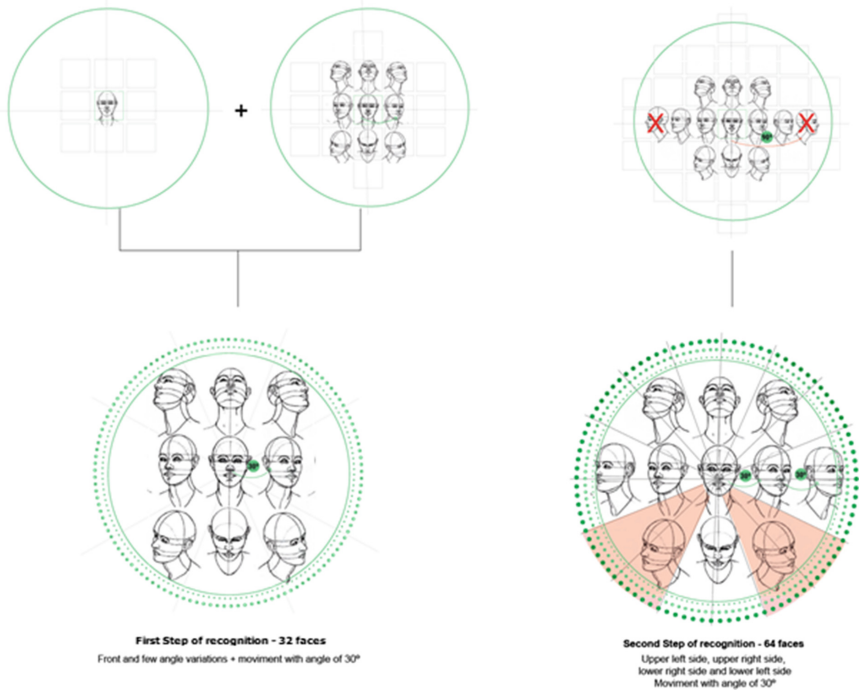


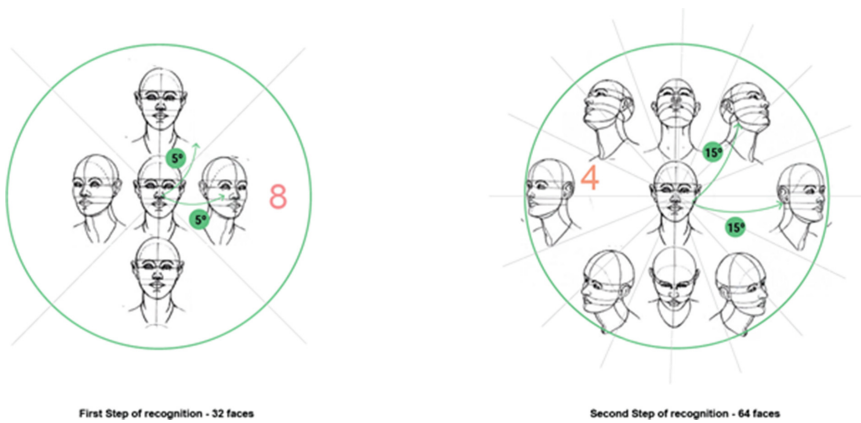
Fig. 2. Third stage - pie chart.

4 Users Evaluation

The experiment took place in an environment with controlled lighting conditions, where participants were tested individually with an average time of 3 min per user. A total of 79 users served as participants for this experiment. Their ages ranged from 21 to 45 years.

Table 1 shows experimental evaluation and the parameter settings. Based on the parameters from the state-of-the-art in face recognition [7], the initial parameters for the first Proof of Concept (PoC) were established. To cover all possible face poses and shapes, n numbers of images are captured ($n = 96$, in current experiments).

About ergonomic parameters, the feedback in all stages allowed to adjust these parameters, such as camera height and camera angle. The camera height is changed to 131 cm in the second stage and 140 cm in the last stage, in response to user discomfort in the experiments. The camera angle was only increased in the last stage for usability reasons. Another interesting element was to capture the rotations (yaw, pitch, and roll) in an angle greater than 15°, a situation that made the user lose control and attention. This difficulty was removed by making improvements in the design (Fig. 3).

**Fig. 3.** Fourth stage.**Table 1.** Parameter settings.

	First stage	Second stage	Third stage	Fourth stage
Participants	25	40	7	7
Shorter (cm)	161	150	155	160
Taller (cm)	180	187	179	170
Camera height	140	131	131	140
Camera angle	60	60	60	70
Capture stages	3	1	2	2
Yaw angle	≤ 30	≤ 30	≤ 15	≤ 15
Pitch angle	–	≤ 20	≤ 15	≤ 15
Face by User	30–35	56	96	96
Discarded faces (%)	–	10	–	–
Average capture Time	150 s	180 s	180 s	120 s
Worst capture Time	2 min	5 min	3 min	3 min

Among the functionality parameters, two steps were achieved in the final stage to complete the register, being insufficient and very ambitious to achieve it in one stage, and considered as excessive more than two stages. The different parameters adjustments allowed reducing the average registration time to two minutes, with the variance of one minute, highlighting as an acceptable time of user interaction.

The proposed changes meet the requirements from development and design teams. In the future, we intend to improve the communication when the user is not doing the correct head movements.

5 Conclusions

In this method, we propose a data collection and image processing tool for face recognition applications. We first analyze the process of the facial data collection and explains the data collection phases. The most important geometric and visual parameters are discussed and analyzed. In the final stage of the data collection tool, those parameters are selected that conform the system requirements and also allows a comfortable user interaction with the system. The parameters and their adjustments, although considered in other studies, show their importance in specific people and contexts. In conclusion, this system allows to collect facial data with important poses covered in the most user friendly manner and that in addition, high quality collected data can be obtained for subsequent tasks of face recognition in different scenarios.

Acknowledgment. The authors would like to thank the support provided by SIDIA Instituto de Ciência e Tecnologia and teams. The authors also like to thank the users, who helped create the system by providing useful feedback during the process.

References

1. Blasco, J., Chen, T.M., Tapiador, J., Peris-Lopez, P.: A survey of wearable biometric recognition systems. *ACM Comput. Surv.* **49**(3), 43:1–43:35 (2016)
2. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. *IBM Syst. J.* **40**(3), 614–634 (2001)
3. Delac, K., Grgic, M.: A survey of biometric recognition methods. In: 46th International Symposium on Electronics, Zadar, Croatia, pp. 184–193 (2004)
4. Özdil, A., Özbilen, M.M.: A survey on comparison of face recognition algorithms. In: 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), pp. 1–3, October 2014
5. Prihasto, B., et al.: A survey of deep face recognition in the wild. In: 2016 International Conference on Orange Technologies (ICOT), pp. 76–79, December 2016
6. Masi, I., Tran, A.T., Leksut, J.T., Hassner, T., Medioni, G.G.: Do we really need to collect millions of faces for effective face recognition? *CoRR abs/1603.07057* (2016)
7. Wang, M., Deng, W.: Deep face recognition: a survey. *CoRR abs/1804.06655* (2018)
8. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. *CoRR abs/1712.04621* (2017)
9. Mennesson, J., Dahmane, A., Danisman, T., Bilasco, I.M.: Head yaw estimation using frontal face detector, pp. 517–524, January 2016
10. Bødker, S.: A human activity approach to user interfaces. *Hum.-Comput. Interact.* **4**(3), 171–195 (1989)
11. Kaptelinin, V., Nardi, B.A.: *Acting with Technology: Activity Theory and Interaction Design*. MIT Press, Cambridge (2009)