AI-Powered Management of Identity Photos for Institutional Staff Directories

Daniel Canedo[a], José Vieira[b], António Gonçalves and António J. R. Neves[c]
IEETA, DETI, LASI, University of Aveiro, 3810-193 Aveiro, Portugal

Keywords: Computer Vision, Face Verification, Deep Learning, Identity Photos, Photo Management.

Abstract: The recent developments in Deep Learning and Computer Vision algorithms allow the automation of several tasks which up until that point required the allocation of considerable human resources. One task that is getting behind the recent developments is the management of identity photos for institutional staff directories because it deals with sensitive information, namely the association of a photo to a person. The main objective of this work is to give a contribution to the automation of this process. This paper proposes several image processing algorithms to validate the submission of a new personal photo to the system, such as face detection, face recognition, face cropping, image quality assessment, head pose estimation, gaze estimation, blink detection, and sunglasses detection. These algorithms allow the verification of the submitted photo according to some predefined criteria. Generally, these criteria revolve around verifying if the face on the photo is of the person that is updating their photo, forcing the face to be centered on the image, verifying if the photo has visually good quality, among others. A use-case is presented based on the integration of the developed algorithms as a web-service to be used by the image directory system of the University of Aveiro. The proposed service is called every time a collaborator tries to update their personal photo and the result of the analysis determines if the photo is valid and the personal profile is updated. The system is already in production and the results that are being obtained are very satisfactory, according to the feedback of the users. Regarding the individual algorithms, the experimental results obtained range from 92% to 100% of accuracy, depending on the image processing algorithm being tested.

1 INTRODUCTION

In this paper, we propose image processing algorithms to be integrated into a photo management system, presenting a use-case implemented at University of Aveiro. With the help of the proposed solutions, the collaborators can update their identity photo autonomously, with the guaranty of a good image quality, following the institutional guidelines and validating the user identity.

Generally, photo management is done manually with considerable costs associated with human resources, especially in large institutions such as universities. Automating this process to some extent would lower those costs. However, one needs to be wary of automating certain tasks, such as face recognition, since it deals with sensitive information that if forged could undermine the purpose of such systems. Since Deep Learning algorithms are not perfect nor transparent in their decisions (Xu et al., 2019), the photo management system must deal with failures by providing feedback to users and Human Resources.

The proposed photo management system takes a submitted photo as an input, processes it through several image processing algorithms to extract relevant information, matches this data against the validity criteria imposed by the institution in which the system is operating and provides feedback regarding each single criterion as the output. Based on the provided feedback, the user can correct the submission, or contact the Human Resources reporting the problem. The whole pipeline of the photo management system, including its image processing algorithms, are detailed in Section 3. A use-case for this system at University of Aveiro is presented in Section 4. The proposed photo management system allows the collaborators of University of Aveiro to update their personal photo autonomously under certain guidelines by uploading a photo on its website. The image processing block of this system will verify and transform the submitted photo to follow those guidelines.

[a] https://orcid.org/0000-0002-5184-3265
[b] https://orcid.org/0000-0002-4356-4522
[c] https://orcid.org/0000-0001-5433-6667

Canedo, D., Vieira, J., Gonçalves, A. and Neves, A.
AI-Powered Management of Identity Photos for Institutional Staff Directories.
DOI: 10.5220/0011649000003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 804-811
ISBN: 978-989-758-634-7; ISSN: 2184-4321

This document is structured as follows: Section 2 presents the related work; Section 3 presents the methodology; Section 4 presents the use-case; Section 5 presents the results; Section 6 presents the conclusion.

2 RELATED WORK

Most of the institutions nowadays have a digital staff directory, public or private. There are clear advantages to keep this information updated. However, depending on the dimension of the institution, this task can be complex. Creating procedures to help each person to update their personal information are of extreme importance. Taking into consideration the focus of this work, there are cases where the user or collaborator have not updated their picture for a long time, which leads to outdated personal biometric information. This can be a problem for a photo management system during the face verification.

In the use-case of University of Aveiro that is presented later in this paper, some identity photos of students and collaborators come from their identity card and were taken in their adolescence. From adolescence to adulthood, there are significant facial changes. The aging problem in face recognition is not new (Singh and Prasad, 2018), however verifying the identity of someone using the photo of the identity card is a particular problem that the proposed photo management system may face.

(Albiero et al., 2020) proposed AIM-CHIYA (ArcFace Identity Matching on CHIYA) to tackle this problem. Firstly, they collected a dataset called Chilean Young Adult (CHIYA), that contains a pair of images for each person. Each pair of images is one image of the person's national identity card issued at an earlier age, and one current image acquired with a contemporary mobile device. They also detected and aligned the faces using RetinaFace (Deng et al., 2020). For the training strategy, they chose a few-shot learning with triple loss approach. Since they are only interested in matching between photos and identity card images, all the triplets are selected as photos to identity cards. That is, if the anchor is a photo face, the positive can only be an identity card face, constraining the negative to also be an identity card face. Then they performed transfer learning using a model that was trained in a larger in the wild dataset. This approach can be relevant for our work since depending on where the photo management system operates, it may have to deal with identity card faces as the reference faces for the face verification. This would have to be done at least one time per user, since after successfully updating their photo, the old identity card one is no longer the reference for future updates.

Another work found in the literature that reinforces this approach is DocFace+ (Shi and Jain, 2019). The authors of this work proposed a pair of sibling networks for learning domain-specific features of identity card faces and photo faces with shared high-level parameters. Afterwards, they trained their model with a larger in the wild dataset and then fine-tuned it using an identity card dataset. To overcome underfitting, the authors proposed an optimization method called dynamic weight imprinting to update the classifier weights.

A similar photo management system found in the literature is MediAssist (Cooray et al., 2006). MediAssist is a web-based personal photo management system that groups all the photos into meaningful events based on time and location information which are automatically extracted when uploading the photos to the system. The authors found that such events are useful both in the search and indexing operations. This is a semi-automatic system where the users can annotate who appears in the photos. The system receives the uploaded photo and performs face detection and identification based on body-patch similarity matching. Then, during the annotation, this photo management system can generate an automatic name suggestion for someone appearing in a given photo.

Another photo management system found in the literature is Face Album (Xu et al., 2017). This system is basically a mobile application that organizes photos by person identity. This system is also semi-automatic since users can correct misidentified faces. The authors use a light convolutional neural network (CNN) for face recognition and proposed an algorithm consisting of two pools: the certain pool which consists of clusters of identified faces, and the uncertain pool which consists of clusters of faces that are yet to be identified. If some faces form a convincing cluster within the uncertain pool, this cluster is moved to the certain pool as a new identity. Besides this automatic management, the user can interact with the system to correct misidentified faces and to identify faces in the uncertain pool.

While the purpose of the photo management systems found in the literature is to index and organize photos, mainly around identity, our system's purpose is to allow for an automatic picture update. This goal not only requires face recognition, but also a set of image processing algorithms to validate and transform the submitted pictures into an adequate format.

3 METHODOLOGY

This section addresses the pipeline of the proposed photo management system. Basically, this system is a web-service that implements several image processing algorithms. They should be computationally optimized to keep the processing time under control for the system to be responsive. This is intended for time constrained applications that may have numerous users using the system at the same time.

The main goal of this system is to receive a picture, analyze it, make the necessary adjustments to fit within certain guidelines, and return the adjusted picture with the analysis feedback. Therefore, the following image processing algorithms were implemented: color verification, face detection, face recognition, face alignment, cropping, head pose estimation, sunglasses detection, blink detection, gaze estimation, brightness estimation, and image quality assessment. The following subsections briefly explain each algorithm.

3.1 Color Verification

Institutions generally have colored images for the identity photos of their collaborators (a full RGB image). Therefore, the photo management system must be able to identify what is the color space of the submitted photo. A straightforward way to do this is to extract this information from the image metadata. However, this metadata can be edited before the submission, that is why the proposed system should have a way to complement the metadata information. The proposed algorithm splits and compares the channels of the image. If the channels are equal, it means the image is grayscale, if the channels are different, it means the image is colored. Thus, if the image metadata indicates that the image is RGB and the channels of the image are different, this algorithm validates the image as having the adequate color space.

3.2 Face Detection

The face detector on the proposed system uses Dlib library (King, 2009). This face detector uses the Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid, and sliding window detection scheme. The main reasons behind this choice are the lower processing time when compared to heavier Deep Learning options and the fact that in addition to the bounding box, it also returns a set of 68 facial landmarks. Some image processing algorithms implemented in this photo management system take advantage of these landmarks. To avoid false positives or multiple faces being detected in a single picture, only the largest face is returned, which generally is the face of the user.

3.3 Face Recognition

This algorithm is one of the most important in the proposed system since it deals with sensitive information, namely personal identity. Therefore, the Dlib ResNet model was deployed as the basis of the face recognition algorithm, which is reported to attain 99.38% accuracy on the standard LFW face recognition benchmark (Huang et al., 2008). This model maps a face to a 128-dimensional vector space, therefore two vector spaces will be needed for this algorithm: the candidate face and the reference face. These two vector spaces are compared with each other by calculating their Euclidean distance. The closer the distance, the more likely it is that they are the same person.

3.4 Face Alignment and Cropping

Generally, institutions want a uniform and adequate directory of pictures of their collaborators. To reach that, face alignment is performed in every submitted picture, using the facial landmarks returned by Dlib. This algorithm considers two facial landmarks on the eyes and if they form an angle of zero in the horizontal axis, the face is aligned. If the angle is not zero, a rotation transformation is applied to align those two facial landmarks with the horizontal axis, until the angle formed by them is zero. After the face is aligned, the cropping is done.

We also consider as requisite that photos should have a similar cropping as the identity card photos. Thus, to achieve that, the cropping algorithm of the photo management system takes the left, right, top, and bottom extremities of the facial landmarks. With these four points, the center point is found. From this center point, the algorithm expands a bounding box around it with a certain length, emulating the style of identity card photos. Finally, the resulting cropped image is resized to a fixed resolution to keep uniformity between all collaborators' pictures.

3.5 Head Pose Estimation

Another issue taken into consideration is the rotation of the face, considering that the photo should be frontal towards the camera. Therefore, the photo management system is equipped with an algorithm to estimate the head pose of the face present in the submitted photo. This algorithm makes use of the facial landmarks outputted by the Dlib face detector. By calculating the rotation and translation of key facial landmarks, it is possible to transform those points in world coordinates to 3D points in camera coordinates and project them onto the image plane. This results in the Euler angles which represent the orientation of the facial landmarks. For a face to be frontal, the pitch and yaw need to be close to zero degrees, the roll is disregarded because the face is previously aligned using this angle.

3.6 Sunglasses Detection

For an identity photo in a professional profile, generally sunglasses are not adequate. For this reason, the photo management system implements two algorithms to detect sunglasses: a low-level algorithm and a Deep Learning algorithm. The low-level algorithm uses the Dlib facial landmarks to get regions of interest right below the eyes and a region on the nose. The idea is to make a color comparison between the regions below the eyes and the nose. If there are sunglasses, the regions of interest below the eyes represent a part of the lenses instead of skin, resulting in a color disparity between those regions and the region of the nose. This algorithm triggers a difference for sunglasses, but not for glasses since the lenses are transparent. This is obviously desirable since there are people who need to use glasses in their daily life. Figure 1 illustrates this algorithm.

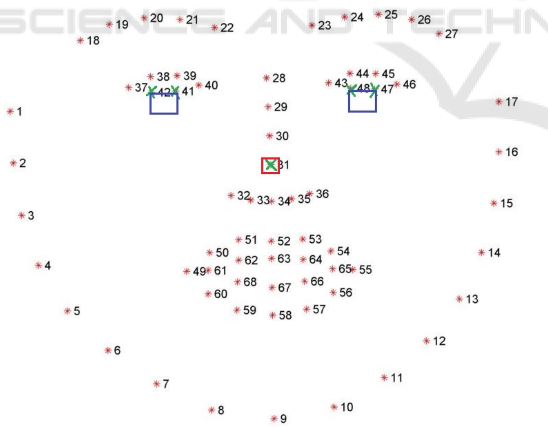

Figure 1: Low-level algorithm for sunglasses detection. A color comparison between the blue regions and the red region is made.

The photo management system implements another algorithm for this task based on Deep Learning. A hybrid dataset was created by combining a Kaggle dataset for glasses detection[1] and a Kaggle dataset for sunglasses detection[2]. This was done to improve the quality and robustness of the dataset and to tackle class imbalance, since sunglasses photos were quite outnumbered. Then, a model was trained using a transfer learning approach on a VGG16 network that was trained on the ImageNet dataset (Simonyan and Zisserman, 2014). This model is the heavier and more accurate solution of the photo management system for the sunglasses detection problem.

3.7 Blink Detection and Gaze Estimation

Since having collaborators with their eyes closed on their identity photos is generally undesirable, the photo management system implements an algorithm that estimates the eye aspect ratio (EAR) (Soukupova and Cech, 2016). This algorithm uses the facial landmarks of Dlib, namely the eyes landmarks. Figure 2 illustrates which landmarks are used by Equation 1 to estimate the EAR.

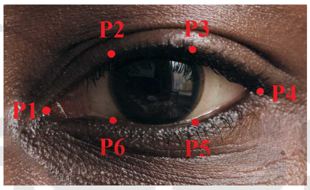

Figure 2: Eyes landmarks used by the blink detection algorithm.

$$EAR = \frac{||P2 - P6|| + ||P3 - P5||}{2||P1 - P4||} \tag{1}$$

The numerator of Equation 1 basically calculates the vertical distance of the eye, while the denominator calculates the horizontal distance. The EAR is mostly constant when the eyes are opened, but it is approximately zero when the eyes are closed. Therefore, with this algorithm, it is possible to verify if the eyes are opened as desired.

Not only it is desired that there are no pictures with eyes closed, but it is also desired that the gaze is directed to the camera. For this reason, the photo management system has a gaze estimation algorithm following the ideas presented in (Canedo et al., 2018). This algorithm also uses the Dlib facial landmarks, and Figure 3 shows which ones are used.

As it is possible to observe in Figure 3, an average between points P1 and P2 is made to obtain the top left corner and an average between points P4 and P5 is made to obtain the right bottom corner. With these

[1]https://www.kaggle.com/code/jorgebuenoperez/computer-vision-application-of-cnn

[2]https://www.kaggle.com/datasets/amol07/sunglasses-no-sunglasses

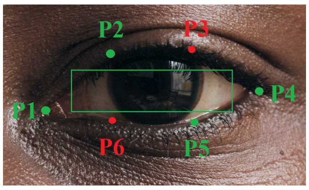

Figure 3: Eyes landmarks used by gaze estimation algorithm in green and the resulting region of interest.

corners, a region of interest is extracted. These averages are made such that the region of interest only catches the eye and not unnecessary features like eyelashes or skin which could deteriorate the algorithm performance. After obtaining the region of interest, a series of low-level image processing operations are conducted to detect the pupils. Firstly, this region is converted to grayscale and a bilateral filter is performed to smooth the image. Then a morphological operation is conducted to erode the image to remove unnecessary features and noise. Finally, it is applied an inverted binary threshold followed by Otsu thresholding (Xu et al., 2011). With this, the pupil is properly segmented. Then, the contours of the pupil are found, and the centroids of the eyes are calculated using the moments of the contours. By averaging the distance of both centroids to the eye's extremities, it is possible to estimate where the user is looking at.

3.8 Brightness Estimation and Image Quality Assessment

It is desirable that the submitted pictures are not overexposed nor underexposed. Therefore, the photo management system has an algorithm to estimate the brightness of the picture. This is simply done by converting the image to the HSV (Hue, Saturation, Value) color space, splitting the channels, and averaging the Value channel, since this channel corresponds to the brightness. The resulting average is then compared to a certain threshold to indicate if the image has adequate brightness.

It is also desirable that the submitted pictures have good quality. Therefore, the photo management system is equipped a model provided by the OpenCV library (OpenCV, 2022) called BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal et al., 2012). This model uses scene statistics of locally normalized luminance coefficients to quantify possible losses in the genuineness of the image due to distortions, which leads to the measure of quality. The features used derive from the distribution of locally normalized luminance and products of locally normalized luminance under a spatial natural scene statistic model. This algorithm not only has a very low computational complexity, but it also is highly competitive within the image quality assessment field, which makes it an ideal choice for an already feature-heavy photo management system.

4 IMPLEMENTATION OF A REAL USE-CASE

In order to validate the proposed photo management system, a prototype called *FotoFaces* was implemented at University of Aveiro, which serves as a reliable use-case. Figure 4 illustrates the process followed by this prototype.

As it can be observed in Figure 4, the user can submit a photo in the University of Aveiro's website and this photo goes through the *FotoFaces Analysis* block, which basically runs all the image processing algorithms described in Section 3 except for the face verification. If the picture meets all the requirements, then it goes through the face verification algorithm, and if it is successfully verified, the photo is updated. Otherwise, the human resources department *SGRH* will re-analyze the photo manually. This is still a closed prototype that is going through testing with the staff before opening it to the students. Figure 5 shows the website of the photo management system implemented in University of Aveiro.

To test this prototype, a popular and challenging dataset was chosen: LFW (Huang et al., 2008). This dataset contains challenging images that were taken in the wild, meaning they were not taken in a controlled environment. A new dataset[3] with 50 pairs of images was created using images from LFW. Each pair corresponds to a reference picture and a candidate picture. The reference picture represents a picture that is already in the system and that is used to verify the identity of a submitted photo. The candidate picture represents a submitted photo. LFW is already challenging by nature, but the pictures chosen for this dataset consider specific challenges that the image processing algorithms need to tackle. Therefore, it was chosen images with glasses, sunglasses, different ethnicities and genders, different lighting conditions, different head poses and gaze directions, different age and looks between the reference and candidate pictures, and so on. Finally, the candidate pictures were manually annotated with 1 or 0 for eight different categories sequentially (*1 = True, 0 = False*): frontal face, colored image, same person, no sunglasses, opened eyes, frontal gaze, adequate brightness, adequate image quality.

For instance, if a picture meets all these require-

[3] Available in tinyurl.com/y4vmyx27

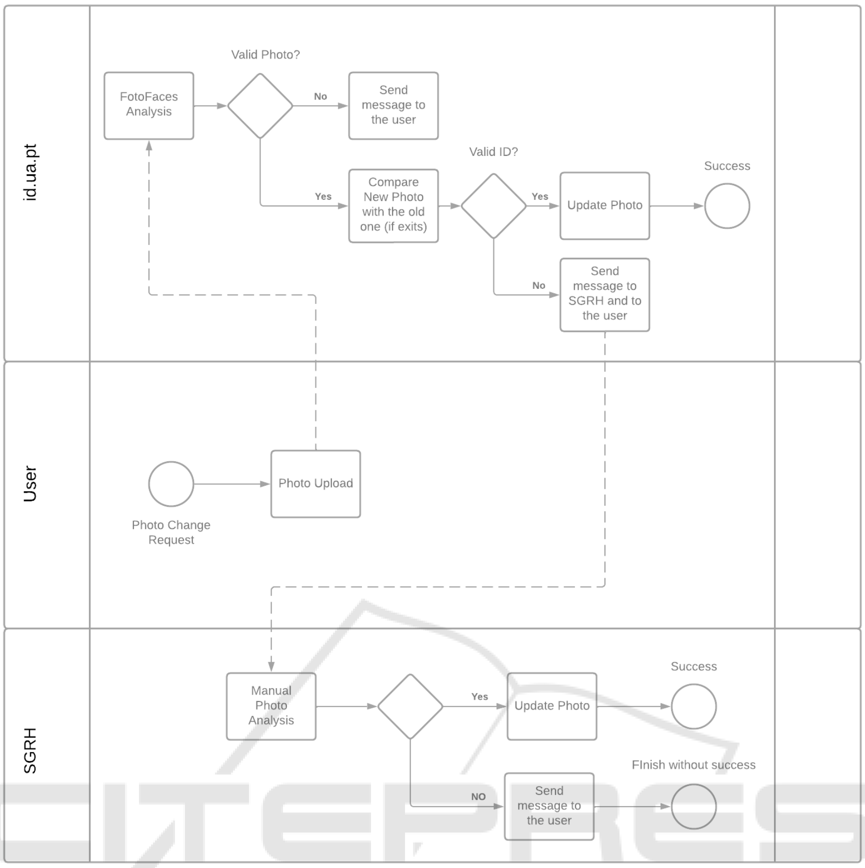

Figure 4: Photo management system prototype of University of Aveiro.

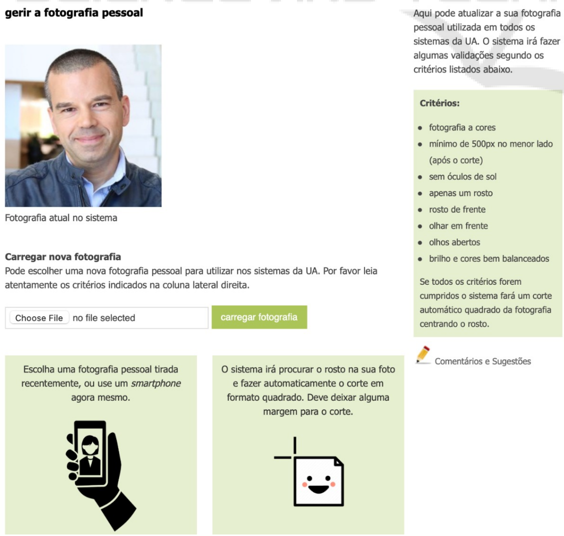

Figure 5: Webpage available to the staff of University of Aveiro for personal photo update.

ments the label is 11111111, if a picture does not meet the third and fourth requirements the label is 11001111, and so on. Figure 6 illustrates a pair of images considered for the dataset.

Figure 6: A pair from the dataset. Reference picture on the left, candidate picture on the right.

Looking at the candidate picture in Figure 6, it meets all the requirements except for the *no sunglasses* category. Since he is wearing sunglasses, the *opened eyes* and *frontal gaze* categories were also annotated as 0 (*False*) because the eyes are obstructed by the dark lenses. Therefore, the final label is 11100011. Table 1 shows what the dataset is consisted of.

Table 1: Dataset based on the LFW dataset, with 50 pairs of images.

Algorithm	True	False
Frontal	30	20
RGB	50	0
Verification	44	6
No Sunglasses	46	4
Opened Eyes	41	9
Gaze	16	34
Brightness	47	3
Quality	49	1

5 RESULTS

Two experiments were conducted. In the first one, each pair of images of the dataset was consecutively fed to the prototype shown in Figure 4 and the outputs of each image processing algorithm were directly compared with the labels for the pair in question. The outputs of each image processing algorithm do not influence each other. Table 2 shows the results of this experiment.

Table 2: First experiment: algorithms operate independently.

Algorithm	Accuracy
Frontal	1.00
RGB	1.00
Verification	0.96
No Sunglasses	0.96
Opened Eyes	0.88
Gaze	0.80
Brightness	0.92
Quality	1.00

As it can be observed in Table 2, the accuracy of each algorithm composing the photo management system is high. The poorest result is the gaze direction, with 80% accuracy.

For the second experiment, an improved approach was taken. It is clear that the image processing algorithms of the photo management system synergize and are dependent on each other. Most algorithms use the facial landmarks of Dlib (Figure 1) and are based on a geometric approach. For this reason, the ideal condition for most algorithms to function properly is when the face is frontal to the camera. Not only that, but it is possible to go further: whenever sunglasses are present, it is not desirable to run the algorithms that check if the eyes are opened or if the gaze is frontal to the camera, since the eyes are obstructed by the dark lenses. Additionally, if the eyes are closed, it is not desirable to run the algorithm that checks the gaze direction since it will not detect the pupils. Figure 7 shows how these changes impacted the previously independent pipeline and Table 3 shows the results.

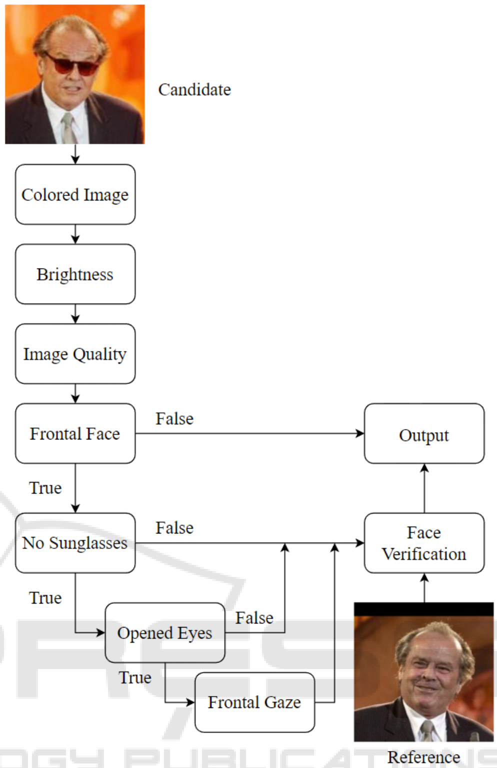

Figure 7: Pipeline for the second experiment.

Table 3: Second experiment: key algorithms operate dependently.

Algorithm	Accuracy
Frontal	1.00
RGB	1.00
Verification	1.00
No Sunglasses	0.97
Opened Eyes	0.96
Gaze	0.96
Brightness	0.92
Quality	1.00

As it can be observed in Table 3, the results were significantly improved from Table 2. This was possible by understanding the dependency and synergy between some image processing algorithms of the photo management system. For instance, constraining the sunglasses' detection algorithm and face verification algorithm only to frontal faces boosted the accuracy from 96% to 97% and 100%, respectively. Constraining the blink detection algorithm that checks if the eyes are opened to faces that are not wearing sunglasses boosted the accuracy from 88% to 96%. Finally, constraining the gaze estimation algorithm to faces that are not wearing sunglasses and have their eyes opened boosted the accuracy from 80% to 96%.

These experiments validate the potential of this prototype deployed at University of Aveiro. It is worth mentioning that the sunglasses detection algorithm used in this prototype was the low-level one and not the Deep Learning one (Subsection 3.6). This choice was done to keep the system responsive when in high demand.

6 CONCLUSION

A photo management system was presented in this document. This system analyzes submitted photos with image processing algorithms, enabling users to update their identity photo automatically. A use-case was tested, and it is currently in use at University of Aveiro. In this use-case, a prototype was implemented to allow the staff to update their identity photo. To validate this prototype, a dataset was built and annotated based on images from the LFW dataset. Two experiments were conducted to test the several image processing algorithms of the prototype. It was concluded that it is possible to take advantage of the dependency of certain image processing algorithms to boost their accuracy. The results obtained were quite satisfactory, considering the challenging nature of the LFW images. The results ranged from 92% to 100% accuracy, depending on the image processing algorithm being tested. The most crucial algorithm of such systems, which is face verification, attained 100% accuracy.

As for future work, this prototype can be improved in several aspects. The face verification algorithm needs to employ strategies to deal with aging, as discussed in Section 2, since there are institutions that have not updated the identity photos of their collaborators for a long time. This prototype also needs to be equipped with an emotion recognition algorithm, since generally identity photos require a neutral expression. Furthermore, it also needs an algorithm to detect hats and other accessories that are not usually welcome in a professional setting.

ACKNOWLEDGEMENTS

This work was supported in part by Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro.

REFERENCES

Albiero, V., Srinivas, N., Villalobos, E., Perez-Facuse, J., Rosenthal, R., Mery, D., Ricanek, K., and Bowyer, K. W. (2020). Identity document to selfie face matching across adolescence. In *2020 IEEE Int. Joint Conf. on Biometrics (IJCB)*, pages 1–9. IEEE.

Canedo, D., Trifan, A., and Neves, A. J. (2018). Monitoring students' attention in a classroom through computer vision. In *Int. Conf. on Practical Applications of Agents and Multi-Agent Systems*, pages 371–378. Springer.

Cooray, S., O'Connor, N. E., Gurrin, C., Jones, G. J., O'Hare, N., and Smeaton, A. F. (2006). Identifying person re-occurrences for personal photo management applications.

Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conf. on computer vision and pattern recognition*, pages 5203–5212.

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758.

Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708.

OpenCV (2022). Opencv. https://opencv.org/. Accessed: 2022-08-26.

Shi, Y. and Jain, A. K. (2019). Docface+: Id document to selfie matching. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(1):56–67.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, S. and Prasad, S. (2018). Techniques and challenges of face recognition: A critical review. *Procedia computer science*, 143:536–543.

Soukupova, T. and Cech, J. (2016). Eye blink detection using facial landmarks. In *21st computer vision winter workshop, Rimske Toplice, Slovenia*.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., and Zhu, J. (2019). Explainable ai: A brief survey on history, research areas, approaches and challenges. In *CCF Int. Conf. on natural language processing and Chinese computing*, pages 563–574. Springer.

Xu, X., Xu, S., Jin, L., and Song, E. (2011). Characteristic analysis of otsu threshold and its applications. *Pattern recognition letters*, 32(7):956–961.

Xu, Y., Peng, F., Yuan, Y., and Wang, Y. (2017). Face album: towards automatic photo management based on person identity on mobile phones. In *2017 IEEE Int. Conf. on acoustics, speech and signal processing (ICASSP)*, pages 3031–3035. IEEE.