A Supervised Face Recognition in Still Images using Interest Points

Guilherme F. Plichoski[1], Guilherme Metzger[2], Chidambaram Chidambaram[2]

[1]Graduate Program in Applied Computing – State University of Santa Catarina
Joinville – SC – Brazil

[2]Department of Production Engineering and Systems – State University of Santa Catarina
Joinville – SC – Brazil

guilherme.plichoski@edu.udesc.br, gui_metzger@hotmail.com, chidambaram@udesc.br

Abstract. *In recent decades, face recognition (FR) has been studied due to technological advances and increased computational power of equipments. This happens also by the emergence of concern with security issues, and the possibility of its application in various domains. In this context, this study was developed in order to present an approach for recognition of individuals through facial images. For this, we used interest points detectors called SIFT (Scale Invariant Feature Transform) and SURF (Speeded up Robust Features), which are invariant to certain complicating factors found in the recognition process, such as lighting changes, scale and rotation. Using the face images of 138 individuals, the results obtained from the experiments show that the approach is suitable for face recognition.*

1. Introduction

During the last decades, the FR becomes one of important research field due to its application for security and other related areas. Consequently, much effort have been put on this area which resulted in many new approaches to increase the robustness of the FR systems. FR received attention from different areas, for example, image processing, computer vision, artificial intelligence and evolutionary computing [Zhao et al. 2003]. Among biometric recognition systems, face biometrics plays an important role in research activities and security applications. Development of this work is motivated by the fact that FR is still an active research area in computer vision applications. Due to the development of new technologies, nowadays, a huge amount of digital images are available acquired under different imaging conditions. These conditions generally add noise, blur, pose changes, occlusion, scale and illumination variation to images. Consequently, recognizing faces from these images becomes a challenging task.

In this context, facial recognition (FR) presents advantages in relation to other methods based on biometry, since most of technologies requires some voluntary action from the users. In voice recognition, iris and fingerprints, for example, the user needs to approach to the capture tool to collect information, but face images may be captured in public places without the knowledge of users. Automatic FR systems became necessary to overcome security problems. Due to this issue, FR has been studied in the past decades by research community, trying to overcome some problems in this complex process. In real life FR applications, the images that are captured by cameras may have variation in scale,

illumination and pose, face occlusions and facial expressions. These issues are difficult to compensate in FR process. To do that, different techniques are applied from several fields, trying to innovate the process and reach for better performance and accuracy.

Basically, we need three steps to achieve FR process: preprocessing and normalization, features extraction and finally matching. In preprocessing step, images are modified through filters according to the type of images and to the goal that should be achieved. Feature extraction methods are classified in three groups: structural, holistic and hybrid methods [Chidambaram et al. 2012]. Structural methods uses geometric measurement such as points and edges, meanwhile, holistic methods work with all face region such as Principal Component Analysis (PCA). Hybrid methods are generally developed using both methods. Structural and hybrid methods should have more discriminative information because it uses parts of the image to create the set of features. To compensate the image variations that appear in different image scenarios, many robust methods for image recognition have been proposed. Among these methods, SIFT [Lowe 1999] and SURF [Bay et al. 2008] have been used in different applications of object recognition and FR in recent decades.

Interest points can be an effective way to detect face images in complex scenarios. To detect interest points, local image features which are invariant to illumination, scale, translation and rotation are identified. Instead of single pixels, which are not representative, interest points gather data from a neighborhood of pixels, providing information that is more relevant and describing the local image features like shape, color and texture [Chidambaram 2013]. Interest points have been used in a vast amount of computer vision applications such as FR [Chidambaram et al. 2012, Ameen et al. 2017, Piotto and Lopes 2016, Fernandez and Vicente 2008, Lei et al. 2009, Križaj et al. 2010] and other image recognition approaches [Mehrotra et al. 2013]. Compared to low-level features like color and edges, interest points are considered more stable and reliable [Chidambaram et al. 2012]. SIFT is considered invariant to changes in illumination, scale and orientation [Lowe 2004]. The SURF algorithm also invariant to scale and rotation, however, it is considered as faster than other feature matching algorithms [Bay et al. 2008].

Based on this context, the main objective of this work is to evaluate the discriminative power of features extracted from both SURF and SIFT as a supervised approach in FR using images obtained under different conditions such as illumination variation, scale and facial expressions [Chidambaram et al. 2012]. In addition to the application of SIFT and SURF, the contribution of this work will also include the identification of image conditions under which the methods achieves a good rate of recognition. Based on the results, other complementary methods of feature extraction can be added to the present approach to improve the results which can effectively be useful for real applications.

This paper is organized as follows: in Section 2, we explain about the fundamentals of interest point detectors SURF and SIFT; in section 3, the methodology of the present work is detailed illustrating the recognition process. In Section 4, experiments and results obtained using different types of face images are described which includes the brief description of image database. Final conclusions and future work directions are drawn in Section 5.

2. Interest Point Detectors

In this section, the general description about SURF and SIFT is provided.

2.1. SIFT (Scale Invariant Feature Transform)

SIFT was first presented by [Lowe 1999]. This approach transforms an image in a significant set of local features, which are invariant to translation, scale and rotation, besides that it is partially illumination invariant.

Interest points are detected using a cascade filtering process that uses efficient algorithms to identify candidate locations [Lowe 1999]. Four main stages are performed to feature extraction, scale-space detection, interest points detection, assigning guidance and interest points description. This approach generates a large number of features that densely cover the image in the full range scale and locations.

In order to detect local maximums and minimums based on the DoG function, each point is compared with its eight closest neighbors points in the same scale and with its nine closest neighbors in inferior and superior scale. In this stage, some points are rejected considering low-computational cost, low contrast and sensitive to noises [Lowe 2004]. Due to a consistent guidance based on local image properties assigned to each interest point, its descriptor can be represented related to this orientation and therefore becomes invariant to image rotation. This is obtainded using gradient magnitude and orientation [LOWE, 2004]. These steps assign to each interest point location, scale and orientation in the image, assuring invariance to its elements. In addition to this, to calculate the highly representative descriptor to image local region, which should be most invariant to other aspects such as illumination and 3D viewpoints, the approach proposed by Edelman, Intrator and Poggio [Chidambaram 2013] is used in SIFT. To obtain the descriptor invariant to illumination variation, the vector obtained from the previous steps is normalized to a unitary vector.

2.2. SURF (Speeded up Robust Features)

This is scale and rotation invariant interest points detector, SURF, was first presented at work by [Bay et al. 2008]. This method performs better compared to SIFT with a low computational cost [Mehrotra et al. 2013]. SURF is a robust scale and rotation invariant interest point detector and descriptor that focuses the spatial distribution of gradient information within the interest point neighborhood [Pan et al. 2013]. This approach detects interest points using a fast Hessian detector based on approximation of the Hessian matrix for a given image point and computes Haar wavelet responses around an interest point for its orientation assignment and features description [Pan et al. 2013]. There are mainly two steps to perform SURF, orientation assignment, which finds a rotation-invariant orientation based on information from a circular region around the interest point and feature description using Haar wavelet responses.

To determine the orientation of feature points, Haar wavelet responses are calculated for a set of pixels within a circular neighborhood around a detected point. To extract the descriptor a square region centered around a key-point is constructed, then this region is divided into 4x4 square sub-regions, within Haar wavelets are calculated for 5x5 distributed sample points. Therefore, each sub-region generates four values to the descriptor vector leading to an overall vector of length 64 [Lei et al. 2009].

SURF uses the same approach as SIFT but with some variations. SURF uses the sign of the Laplacian to have a sharp distinction between background and foreground features. It uses only 64 dimensions and SIFT uses a 128 dimensional vector [Mehrotra et al. 2013]. This reduces time consumed by SURF compared to SIFT. Experiments in the same sample image performed by authors [Mehrotra et al. 2013] show that SIFT found 207 keypoints in 1216 milliseconds while SURF found 268 keypoints in 89 milliseconds. SURF uses a Hessian matrix and describes a distribution from a window around the interest points as descriptors [Mehrotra et al. 2013]. A sample image with interest points obtained from experiments performed by authors [Lei et al. 2009] is shown in Figure 1.

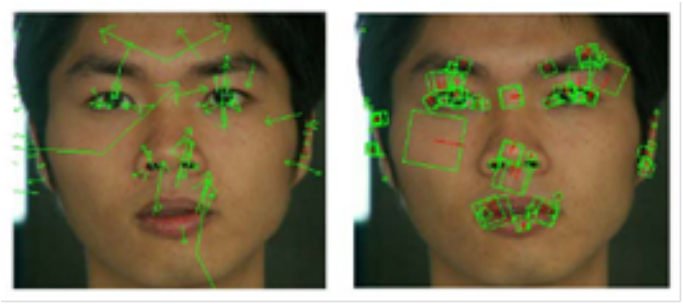

Figure 1. SIFT key-points (left) and SURF key-points (right).

3. Face Recognition

The main objective of this work to compare the performance of SURF and SIFT for FR using different set of face images obtained under different conditions. This section also addresses the other important aspects of the study such as the definition of thresholds and similarity measure.

The first main step is to find all the interest points in the query image and the base image. Then, each of the interest points descriptors from the base image are compared to each of the descriptors from the query image, in order to find the two nearest neighbors for each interest point. One nearest neighbor is defined as the interest point with the smallest distance to another point. In fact, the thresholds define this control. In this work, using a small set of face images, the thresholds are defined for each method. A representation of this process is shown in Figure 2.

During the querying process, the descriptor vector of the query face image will be compared with the descriptor vector of all face images in the database (1 to n). The matching procedure is based on Euclidean distance measure. The classification process in this approach is done using the K-NN (k-nearest neighbor) algorithm in which k=1. In other words, the retrieved descriptor vector that presents a minimum distance among all comparisons will be considered as a correct match. However, the selection of correct match depends on the repeatability rate as defined by Equation 3. The repeatability rate determines the maximum similarity between two face images. To select the best match, the maximum repeatability rate among all comparisons should be considered. Furthermore, the Euclidean distance measure of descriptor vector of query image and database images must respect the threshold distances that are defined by Equations 1 and 2. To

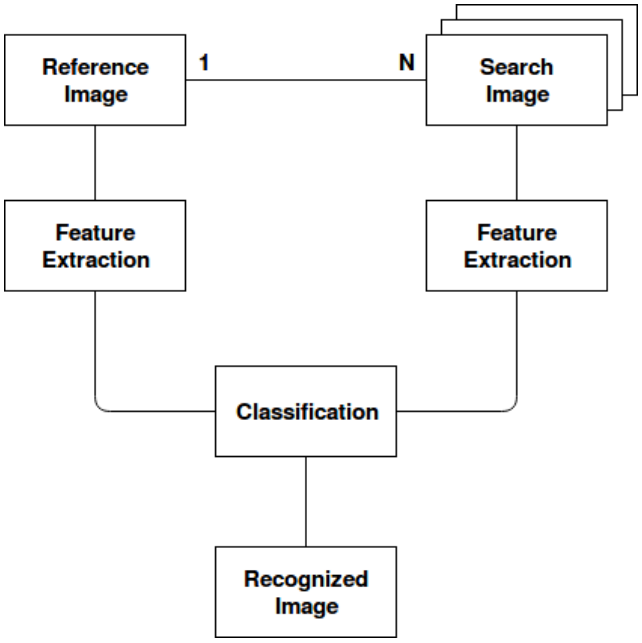

Figure 2. Recognition process.

produce high recognition rate, two distance thresholds, coordinate distance (TCDE) and descriptor distance (TDDE), must be defined through preliminary experiment analysis.

3.1. Interest Points Evaluation Criteria using Thresholds

The main task in FR is to find out the similar features between two face images. Generally, features should have some specific properties that can be used in matching images, for example, robustness and distinctiveness. Robustness refers to the invariant features to illumination, scale and pose variations and distinctiveness indicates the uniqueness of features. Large number of features can be extracted from face images using different algorithms. The main fact is that such features should be highly distinctive and provide a basis for the recognition task. Interest points can also be treated in the same way [Trujillo and Olague 2008].

Repeatability is defined by the image geometry. Measurements of repeatability will quantify the number of repeated points detected under varying conditions such as image rotation, scale change, variation of illumination, presence of noise and viewpoint change. The percentage of detected points that are repeated in both images is defined as the repeatability rate [Chidambaram 2013].

In summary, the percentage of points repeated in the two images being compared is defined as the repeatability rate. A point is considered repeated if it lies in the same coordinates on both images. Due to the several variations or transformations present in real-world conditions, a point is generally not detected exactly at the same position, but, in some neighborhood. Thus, an acceptable error needs to be established when measuring the distance between the coordinates of two images. Hence, the set of repeated interest points on images I_j and I_k, denoted by R_{ipj}, k, is defined as:

$$Rip_{j,k} = \{x_i | \sqrt{(x_i^j - x_i^k)^2} < T_{CDE}\} \quad (1)$$

where x_i^n = (x_i^n, y_i^n) denotes the *i-th* coordinate (x, y) in the image n and T_{CDE} represents the acceptable distance error between the coordinates of interest points on different images (Coordinate Distance Error - CDE).

If the point is classified as repeated, then an acceptable distance error for the associated descriptors also needs to be defined:

$$RIP_{j,k} = \{x_i \in Rip_{j,k} | \sqrt{\sum_{i=1}^{n} (d_i^j - d_i^k)^2} < T_{DDE}\} \quad (2)$$

where d_i^n denotes the *i-th* position of the descriptor vector related to the interest point x_i of image n, and T_{DDE} represents the admissible distance error between two descriptor vectors (Descriptor Distance Error - DDE).

The repeatability rate, R, of interest points extracted from two images, Im_j and Im_k, is defined by the following equation:

$$R = \frac{RIP_{j,k}}{min(IP_j, IP_k)} \quad (3)$$

where $RIP_{j,k}$ denotes the repeated interest points obtained by Equation 1, and IP_j and IP_k represent the total of number of interest points detected on images Im_j and Im_k, respectively. The image with minimum number of interest points is considered since the number of detected points may be different for the two images.

4. Experiments and Results

4.1. Image Database

To evaluate the potential of SIFT and SURF, we have conducted the experiments using the database [das Chagas Prodossimo et al. 2012]. These face images change in pose, scale and illumination. In this work, we have used seven categories of images from the database of 138 individuals and the total number of image is 966. The selected images are organized in two main categories: Controlled illumination (CIL) and Uncontrolled Illumination (UCIL). For CIL, a specific controlled lighting system was provided meanwhile for UCIL, just the lighting condition provided by the fluorescent lamps of the room was utilized. In the experiments of this section, a subset of seven classes, namely, frontal viewpoint with neutral face (under CIL, UCIL and lateral-light illumination conditions) and frontal viewpoint with changes on the facial expression and scale (both under CIL and UCIL illumination conditions). The images obtained under lateral illumination conditions (UCIL_LATFRONTAL) differs from the other categories because light source is positioned at individuals to the left side to simulate a partial illumination condition on one side of the face. These classes and the corresponding labels are summarized in Table 1. Since the CIL-Frontal images are used as the base images, they were captured following the standardization rules provided by [Nist. 2007]. The CIL_EXPR and UCIL_EXPR images were taken with the same procedure as frontal images but with individual smiling

Table 1. Face Images Categories.

Label	Illumination	Type
CIL_FRONTAL	Controlled	Neutral
CIL_EXPR	Controlled	Facial Expression
CIL_SCALE	Controlled	Scale
UCIL_LATFRONTAL	Uncontrolled Lateral	Neutral
UCIL_FRONTAL	Uncontrolled	Neutral
UCIL_EXPR	Uncontrolled	Facial Expression
UCIL_SCALE	Uncontrolled	Scale

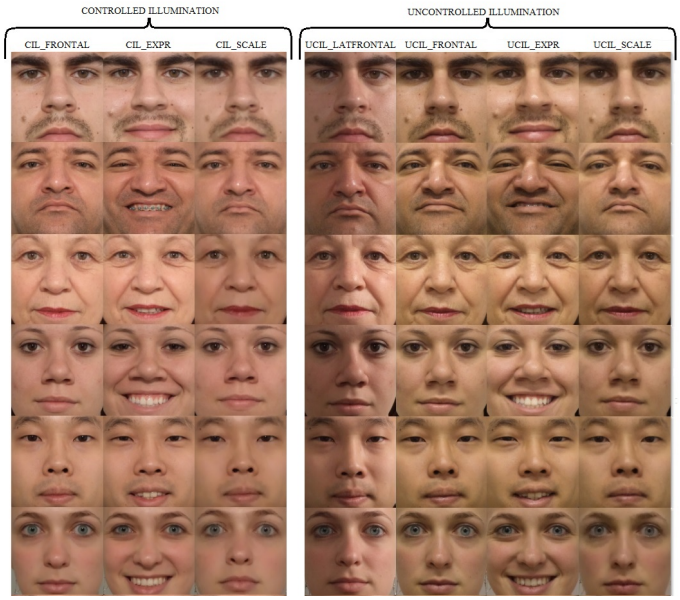

Figure 3. Database Sample.

providing the face expression. The CIL Images were obtained using a light source centered in person's face 2.5 meters away [Chidambaram 2013]. Samples images are shown in Figure 3.

After the image acquisition, normalization of face images is a critical issue in FR systems. Many FR methods require normalized face, for example, holistic features approaches. Facial features that are usually normalized include size, orientation and illumination. In a previous work [das Chagas Prodossimo et al. 2012], by detecting eyes, the database of face images was normalized for size and orientation, except for illumination. Similar to the FR works found in the literature, the single face images were cropped into the size of 550 x 550 pixels. In all experiments, the faces from the first category, CIL-Frontal, are maintained as the base images.

4.2. Threshold Definition

To find the best values to each threshold, a set of experiments was conducted using a small set of images, varying possible threshold values. The category CIL_FRONTAL was used as base images. Our search images were chosen randomly formed by 20% of the 828 images from the other categories.

To define T_{CDE} value for SIFT method, we performed experiments varying values between 10 to 50, with an interval of 10 and to define T_{DDE} we used values between 100 and 500 with 100 interval. The results are illustrated in Figure 4. After performing the experiments, the values 10 and 300 was chosen to T_{CDE} and T_{DDE} respectively because it showed best recognition rate than others combination of thresholds indicated in the 4 marked by an arrow.

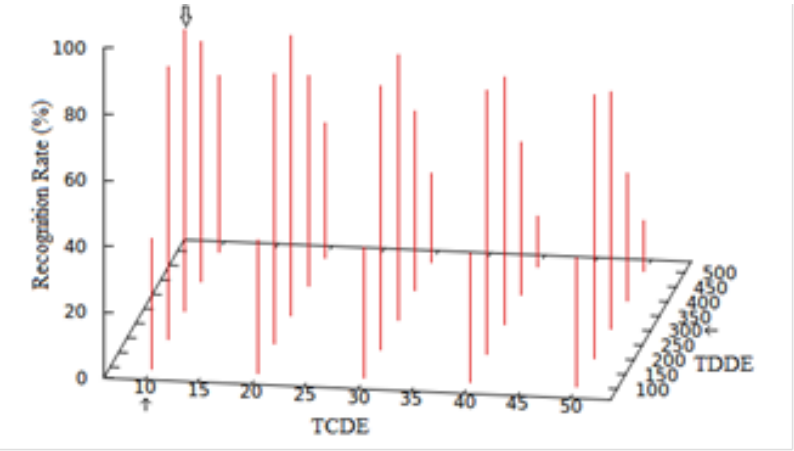

Figure 4. Threshold Analysis for SIFT.

In order to define T_{CDE} and T_{DDE} values for SURF method, the experiments were done varying the values between 10 to 70, with an interval of 10 and 0.2 to 0.7 with 0.1 interval respectively. The recognition rates for each combination are shown in Table 2 (the best values are in bold). The best results were obtained for T_{CDE} equal to 20 for almost all T_{DDE}. Therefore, the final experiments were performed with T_{CDE} value equal to 20 and T_{DDE} value equal to 0.4, which yielded the highest recognition rate.

Table 2. Thresholds Analysis for SURF (%).

	T_{DDE}					
T_{CDE}	0.2	0.3	**0.4**	0.5	0.6	0.7
10	84.06	86.23	84.78	86.96	84.78	84.06
20	85.51	88.41	**89.86**	86.23	87.68	84.78
30	86.23	88.41	87.68	85.51	84.78	84.78
40	86.23	83.33	86.96	83.33	82.61	76.81
50	85.51	84.78	86.96	84.06	78.26	75.36
60	84.78	84.78	83.33	83.33	77.54	76.09
70	84.06	84.06	83.33	81.88	76.81	73.91

4.3. Experimental Results

The recognition experiments were divided by the database categories, except CIL_FRONTAL which is defined as base image. Additionally one more set was defined using six category of images (ALL_IMAGES). All other categories are treated as query images. Table 3 presents the results obtained for each category.

Table 3. Recognition Rates (%).

Images	SIFT	SURF
CIL_EXPR	92.03	89.13
CIL_SCALE	92.75	95.65
UCIL_LATFRONTAL	78.26	69.57
UCIL_FRONTAL	87.68	83.33
UCIL_EXPR	65.22	32.61
UCIL_SCALE	75.36	62.32
ALL_IMAGES	81.88	72.1

According to obtained results, as shown in Table 3, Both methods performs well with the categories CIL_EXPR and CIL_SCALE achieving overall recognition rates above 90%. At the same time, both SIFT and SURF produce rates above 80% with the images obtained from uncontrolled condition (UCIL_FRONTAL). The results demonstrates clearly the discriminative power of features extracted from SIFT in comparison with SURF. Furthermore, both methods are invariant up to certain level regarding illumination variation as shown by the category UCIL_FRONTAL. Both methods suffer with uncontrolled illumination and facial expressions in which the recognition rate reached at the lowest value among all categories. Finally, the experiment done with all images (ALL_IMAGES) reached at 81.88% for SIFT and 75.1% for SURF. It highlights the robustness of the methods.

5. Conclusion

Facial recognition under uncontrolled conditions has been a subject of interest to research community mainly in the past two decades. Many works aiming to find robust solutions and assertive paths to FR problem were proposed, motivated mainly by security concerns. Thus, the present paper is proposed to study the interest points detectors as face features, focusing on SIFT and SURF methods. As a result, we obtained an average rate of 81.88% for SIFT method and 72.1% for SURF method, showing that interest points usage is viable to FR systems. The categories of face images with controlled illumination presented recognition rates higher than the others, for example, with uncontrolled illumination. From the experiments, we can strongly understand the discriminative power of both methods in FR process. In addition to this, we can note that recognition process is a complex task and it is influenced strongly by the image variations and lighting conditions. Hence, one possible way to achieve high rates is by combining other methods together with interest points detectors. For future works, based on the above discussions, we plan to attempt complementary techniques with SURF and SIFT methods, in the same database, with the goal to achieve an even higher recognition rates. Also, another direction for future research will be to apply the same methods in another database with different conditions. Finally, we intend to perform a comparison between recognition rates reached in this work using other methods applied to the same database.

References

Ameen, M. M., Eleyan, A., and Eleyan, G. (2017). Wavelet transform based face recognition using surf descriptors.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3):346–359.

Chidambaram, C. (2013). *A contribution for single and multiple faces recognition using feature-based approaches*. PhD thesis.

Chidambaram, C., Marçal, M. S., Dorini, L. B., Neto, H. V., and Lopes, H. S. (2012). An improved ABC algorithm approach using surf for face identification. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 143–150. Springer.

das Chagas Prodossimo, F., Chidambaram, C., Hastreiter, R. L. F., and Lopes, H. S. (2012). Proposta de uma metodologia para a construção de um banco de imagens faciais normalizadas. In *VIII Workshop de Visão Computacional*, page S.1:sn.

Fernandez, C. and Vicente, M. A. (2008). Face recognition using multiple interest point detectors and SIFT descriptors. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–7. IEEE.

Križaj, J., Štruc, V., and Pavešic, N. (2010). Adaptation of SIFT features for face recognition under varying illumination. In *MIPRO, 2010 Proceedings of the 33rd International Convention*, pages 691–694. IEEE.

Lei, Y., Jiang, X., Shi, Z., Chen, D., and Li, Q. (2009). Face recognition method based on SURF feature. In *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on*, pages 1–4. IEEE.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Mehrotra, H., Sa, P. K., and Majhi, B. (2013). Fast segmentation and adaptive SURF descriptor for iris recognition. *Mathematical and Computer Modelling*, 58(1):132–146.

Nist. (2007). Face recognition format for data interchange - best practices. *USA*.

Pan, H., Zhu, Y., and Xia, L. (2013). Efficient and accurate face detection using heterogeneous feature descriptors and feature selection. *Computer Vision and Image Understanding*, 117(1):12–28.

Piotto, J. G. S. and Lopes, F. M. (2016). Combining surf descriptor and complex networks for face recognition. In *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on*, pages 275–279. IEEE.

Trujillo, L. and Olague, G. (2008). Automated design of image operators that detect interest points. *Evolutionary Computation*, 16(4):483–507.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458.