

Score Fusion of 3DLBP and ACDN+P for 3D Face Recognition

João Baptista Cardia

*Faculdade de Tecnologia
"Dep. Júlio Julinho Marcondes de Moura"
Garça, São Paulo, Brasil*

Aparecido Nilceu Marana

*Faculdade de Ciências
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Bauru, São Paulo, Brasil*

Abstract — Many situations of our everyday life require our identification. Biometrics-based methods, besides allowing such identification, can help to prevent frauds. Among several biometrics features, face is one of the most popular due to its intrinsic and important properties, such as universality, acceptability, low costs, and covert identification. On the other hand, the traditional automatic face recognition methods based on 2D features can not properly deal with some very frequent challenges, such as occlusion, illumination and pose variations. In this paper we propose a new method for face recognition based on the fusion of 3D low-level local features, ACDN+P and 3DLBP, using depth images captured by cheap Kinect V1 sensors. In order to improve the low quality of the point cloud provided by such devices, Symmetric Filling, Iterative Closest Point, and Savitzky-Golay Filter are used in the preprocessing stage of the proposed method. Experimental results obtained on EURECOM Kinect dataset showed that the proposed method can improve the face recognition rates.

Keywords-Biometrics; Face Recognition; 3D Local Features; Symmetric Filling ; Iterative Closest Point; Savitzky-Golay Filter; Kinect.

I. INTRODUCTION

In everyday life we have the constant need for assuring our identity (e.g. accessing Facebook or other social network). Normally the authentication is made using something the person possesses (e.g. keys, id cards) or something it knows (e.g. passwords, codes, personal data) [1]. However those more traditional methods have some downsides. The first one is that some impostor can get a hold of another persons id card and, having the original document in hands, passes itself as the genuine person, without never being caught [1].

The second one is that all those methods can only do positive authentication and, in several cases, it is needed to perform negative authentication. An example of this need is for registering on a governmental social program, it is needed to know if a person is not already signed up for the benefit, this way avoiding duplicity of records. In this case the duplicity can be harmful, since it is normally a sign of someone trying to fraud the social program.

One way to solve this is utilizing Biometrics based methods instead of knowledge or possession. Biometrics utilizes physical (face, fingerprint, or iris) or behavioral characteristics (gait, signature, or voice) for identifying subjects. The utilization of biometrics can help to reduce drastically fraudulent operations [1].

Of all the types of biometric characteristic the face is one of the most interesting. Human beings utilize faces naturally to recognize different individuals and in automatic identification systems, the use of facial traits presents several advantages

over other types of biometric characteristic. For instance, is possible to capture a face from a considerable distance, in a covert manner, this makes face one of the best choices when discretion is a requirement of the identification application.

In constrained applications automated facial recognition systems proposed so far can perform better than the human visual system [2], but the real challenge arise when dealing with uncontrolled environments, where the faces can present severe variations in illumination, pose, and facial expressions. One way to increase the automated facial recognition systems accuracy is to utilize different sensing devices such as 3D or infrared cameras [3]. However, the main issue with these devices is their high costs.

The proposed work tries to utilizes data from a Kinect 1.0 sensing device to perform facial recognition. This is done extending the work proposed in [4] creating a new local characteristic and doing score level fusion with the 3DLBP proposed in [5]. It is also accessed how utilizing a combination of Symmetric Filling, Savitzky-Golay filter, and Iterative Closest Point (ICP) can help to increase the performance of the proposed method.

This work is structured as follows: In the section II the original ACDN characteristics are discussed and explained, in section III the 3DLBP method is discussed and explained, in section V the pre-processing, the P characteristic and the score fusion of the proposed method is discussed, in section VII the experiments and the dataset are explained, and in section VIII the final conclusions of this work are discussed.

II. FACE RECOGNITION BASED ON LOCAL LOW-LEVEL FEATURES

The method proposed in [4] focus on putting together low-level features taking only in consideration the nose (rigid), eyes and forehead (non-rigid). This section explains the core of this method.

A. Feature Extraction

In [4] it is utilized empirical results to develop binary masks to crop the nose, mouth, eyes and forehead from the 3D scans. To reduce the size of the characteristics extracted, all cropped features are re-sampled at uniform intervals (2mm) and only the seeds are stored. Examples of the masks and the cropped features are given in the Figure 1.

After the pose correction, the range image is transformed into a point cloud. This is done by making the x and the y indices from a (x,y,z) matrix and z the depth data. Each region from the face is represented as multiple spatial triangles, which

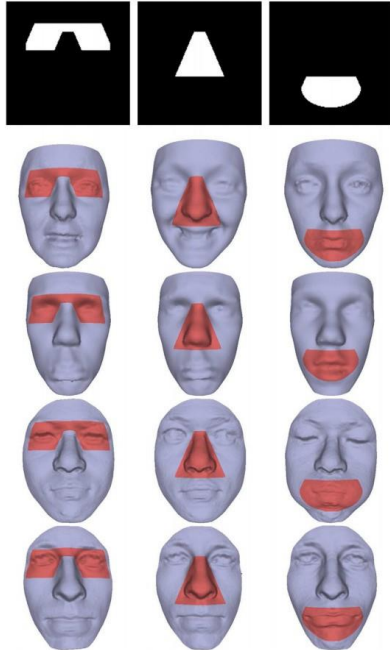


Fig. 1. Binary mask cropping the areas from the face [4].

are made from a vertex selected utilizing the nose tip and two others random points chosen from the corresponding local surface region. From each triangle four types of geometric features are proposed [4], showed in Figure 2:

- A: The angle of two lines generated by two random points in the nose tip;
- C: The radius of the circle circumscribed in the triangle formed by the nose tip point and two random points;
- D: The distance of a line between two random points;
- N: The angle between a line defined between two random vertices and the z.

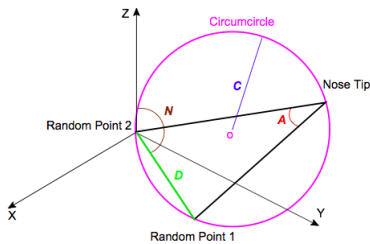


Fig. 2. An example of the four geometric features [4].

After the feature extraction they are normalized into (-1,+1) and grouped into histograms. The histogram generation is done by counting how many entries fall into each of the m bins.

The Figure 3 shows the comparison of four histograms with a dimension of 180. It is important to note that even for different expressions the histograms from the same subject remain stable, while for another person it exhibits a visible difference.

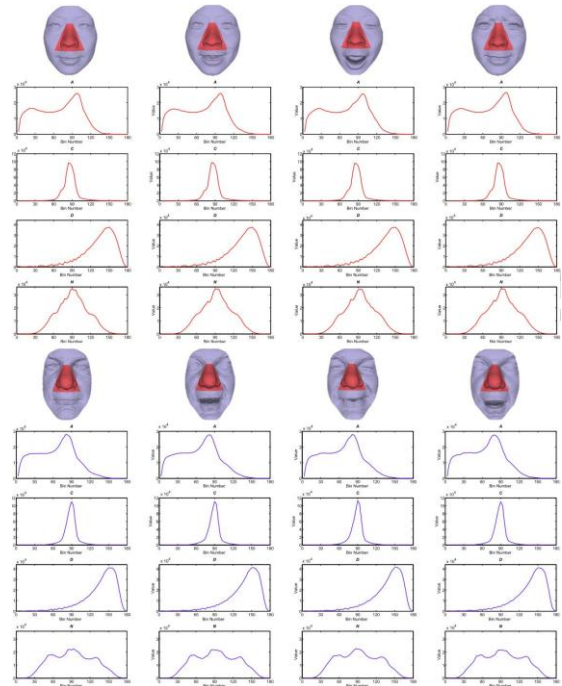


Fig. 3. Comparison of histograms generated from the same subject [4].

III.3D LOCAL BINARY PATTERN

The LBP operator only takes in consideration the signal of the comparison between a region and its kernel. The operator was originally proposed for texture description and cannot deal with the behavior of depth values. One can observe that if two central points on different samples have highest (or lowest) depth values than their neighbors, they will have the same operator value, even if they are from different subjects [5]. This would be common on points belonging to the nose tip of a face subject, for instance.

To deal with situations like the authors of [5] proposed the 3D Local Binary Patterns (3DLBP). This variation of the original operator considers not only the signal of the difference, but also the absolute depth difference. In [5] is stated that more than 93% of all depth differences (DD) with $R = 2$ are smaller than 7. Due to this property the absolute value of the DD is stored in three binary units ($i_2 i_3 i_4$). Therefore, it is possible to affirm:

$$|DD| = i_2 \cdot 2^2 + i_3 \cdot 2^1 + i_4 \cdot 2^0 \quad (1)$$

There is also i_1 , a binary unit defined by:

$$i_1 = \begin{cases} 1 & \text{if } DD \geq 0; \\ 0 & \text{if } DD < 0. \end{cases} \quad (2)$$

Those four binary units are divided into four layers and, for each of those layers, four decimal numbers are obtained: P_1, P_2, P_3, P_4 . The value of the P_1 has the same value as the original LBP. For matching, the histograms of the local

regions (P_1, P_2, P_3, P_4) are concatenated. The Figure 4 shows the process for the generation of the 3DLBP, given an image.

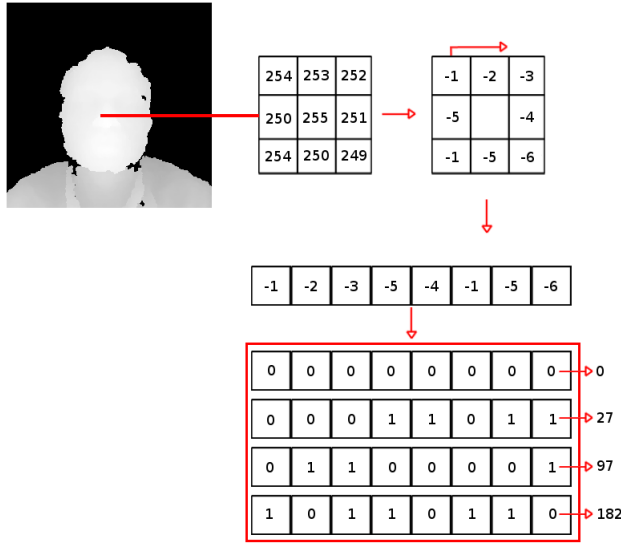


Fig. 4. The full process of the 3DLBP proposed by [5]. Each of the differences is encoded into the layers (layer 2, 3 and 4) and the signal into the layer 1.

IV. ITERATIVE CLOSEST POINT

The Iterative Closest Point (ICP) tries to find a rigid transformation that minimizes the least-square distance between two points. This can be done with a closed-form solution proposed in [6]:

$$E = \sum_{i=1}^n w_i |A_i - R(B_i - B_c) - T|^2 \quad (3)$$

The solution finds the translation factor T and the rotation R . T is the difference of the weighted centroids of the corresponding points while R has a more complex solution and is entirely described in [6].

A. Savitzky-Golay filter

The Savitzky-Golay filter is a lowpass filter based on least-squares polynomial approximation [7]. In [8] the authors show that fitting a polynomial to a set of inputs and then evaluating the polynomial resulting at a single point within the approximation interval is the same as a discrete convolution with a fixed impulse response [7].

Simply put, the Savitzky-Golay filter tries to find filter coefficients c_n that preserves the higher moments of the filtered data. It utilizes the idea of a window moving through the function data and approximate it by a polynomial of higher order, quadratic or quartic [9].

For each data point a polynomial is least-square fit to all $n_L + n_R + 1$ in the moving window, being n_L all the number points to the left of the data point and n_R and the number points to the right. After this g_i is set to the value of the polynomial at position i [9].

It is possible to define g_i as [9]:

$$g_i = \sum_{n=-n_L}^{n_R} c_n + f_i + n \quad (4)$$

being n the data point and $f_i \equiv f(t_i)$ e $t_i \equiv t_0 + i\Delta$, this is for some constant sample spacing Δ and $i = \dots -2, -1, 0, 1, 2, \dots$.

After the windows moves to another data point the process starts all over again.

V. PROPOSED METHOD

In this work the proposed method has three main steps, Pre-Processing, Feature Extraction and Matching.

A. Pre-Processing

For pre-processing first the face is segmented, then it goes through the Symmetric Filling process and, lastly, new depth maps are generated.

1) *Face segmentation*: For face segmentation all points that fall inside a circle with radius $R = 70mm$ centered at the nose tip are segmented. After this the face is translated, having its nose at the origin.

2) *Symmetric Filling*: After the face segmentation, a mirrored image is generated by replacing the X points with their opposite numbers. If the Euclidean distance from a XY coordinate in the mirrored image to its closest neighbor in the original image is less than δ , then the point is removed. Otherwise, it will be added to the main point cloud.

The main objective of this step is to deal with missing data. The human face is not completely symmetric, but the difference in the two sides of the face is lower than the one caused by different identities. In [10] the authors utilize $\delta = 2mm$ (this value was found empirically).

After the Symmetric Filling process it is possible that the faces are not perfectly aligned, for correcting this the Iterative Closest Point (ICP) algorithm is applied. To smoother even further a Savitzky-Golay filter is applied to the resulted Cloud Point.

B. Generation of Depth Maps from the Cloud Points

In our method, aiming to increase the robustness of the 3D face recognition from kinect data, we generate depth maps from the cloud points, outputted by the Kinect device.

In order to generate a depth map from the cloud points, a circular region with radius R is cropped centered at the nose tip. Then, the cropped image goes through the symmetric filling process. Finally, the resulting face image is fitted to a smooth surface using an approximation approach, which is done with an open source code¹ written in MATLAB. The result of this process is a 100 X 100 matrix, stored in an image file. Since the image file cannot contain decimal values, all the depth values are rounded.

¹<http://mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit>

VI.FEATURE EXTRACTION

For each face two different feature extraction methods are applied. First the ACNDP characteristics (ACDN is explained in the section II and the P is proposed in this work) are extracted. For each characteristic there is a histogram with 14 bins and, the final descriptor for this characteristic, is the concatenation of each characteristic. This process is made on the cloud point after the symmetric filling and ICP fitting.

Second the 3DLBP is applied to the new depth maps generated. Each face is divided in 64 micro-regions and, each of those regions, has a histogram with 14 bins. The final descriptor of the face is the concatenation of the histogram from each region.

1) *The P characteristic:* Based on the methodology in [4], another characteristic was determined, the P feature. The P feature is the angle of the intersection from the line defined by the random point and the nose-tip and the Z axis.

A. Matching

The current work focus in score-level fusion. The final classification score is given by:

$$FS = 3DLBP SC * w_1 + ACDNP SC * w_2 \quad (5)$$

being FS the final score, 3DLBP SC the score from the 3DLBP method, w_1 the weight for the 3DLBP score, ACDNP SC the score from ACDN+P characteristics, and w_2 the weight for the ACDN+P characteristics.

For each probe is calculated the probability of him being each one of the subjects in the gallery, this is done with SVM and, each method, has a different SVM trained for it. After this the scores (the probability of the subject in the probe being the subject n in the gallery) are fused with the previous formula, the class with the highest probability is the identity assigned to the face in the probe.

VII.EXPERIMENTAL RESULTS

In our work, the experiments were carried out utilizing the following subsets of the EURECOM database:

- Gallery: open mouth, smiling, and illumination variation faces;
- Probe: neutral face.

Figure 5 shows examples of depth maps of some Gallery and Probe faces used in our experiments.

First, in order to show the importance of the ICP and Savitzky-Golay filter in the enhancement of the quality of the point cloud generated by the Kinect sensor, we evaluated each characteristic of the ACDN+P separated in a process without the ICP and the Savitzky-Golay filter. The results of this experiment (referred here as experiment 1) are compared with the results of the ACDN+P in the full pre-processing proposed in this work.

Then, in the second experiment, we evaluated the 3DLBP method on two different data sets: (i) the original Kinect depth images; and (ii) the new depth maps generated as described in subsection V-B. This experiment also aimed to compare the results obtained in this work with the results presented in [11].



Fig. 5. Depth maps of some Gallery and Probe faces utilized in our experiments.

TABLE I
RESULTS FOR THE EXPERIMENT UTILIZING THE ACDN+P FEATURES.

Feature	Recognition %
A	84 missclassified [80.77%] on 104 subjects
C	101 missclassified [97.12%] on 104 subjects
D	98 missclassified [94.23%] on 104 subjects
N	82 missclassified [78.85%] on 104 subjects
P	87 missclassified [83.65%] on 104 subjects
ACDN+P	62 missclassified [59.62%] on 104 subjects

A. Experiment 1

In this experiment, the ACDN+P low level local features were extracted from the rigid regions of the face: the nose and eyes/forehead regions. Then, the histograms of such features were calculated and concatenated into the feature vectors. Finally, the feature vectors were used by the SVM classifier in order to classify the faces.

Table I presents the result for one part of the experiment. They were carried out with the same sub set of images from the EURECOM Kinect Faceset and separately for each feature and the fusion between them. In this experiment specifically the ICP and Savitzky-Golay filter are not applied.

Table II presents the results obtained with this experiment. One can observe that the new low level local P feature proposed in this work increased significantly the recognition rate: from 28.85% to 40.38%.

Table III presents the results obtained with ACDN+P features with and without the preprocessing stage. One can observe that the application of the Symmetric Filling, ICP, and Savitzky-Golay filter contributed significantly to improve the recognition results: from 40.38% to 68.26%.

B. Experiment 2

In this experiment, the recognition of the faces were carried out by using:

TABLE II
RECOGNITION RATES OBTAINED WITH THE TWO SETS OF LOW LEVEL LOCAL FEATURES: ACDN AND ACDN+P.

Feature	Recognition Rate
ACND	28.85%
ACDN+P	40.38%

TABLE III
RECOGNITION RATES OBTAINED WITH THE ACDN+P LOW LEVEL LOCAL FEATURES WITH AND WITHOUT THE PREPROCESSING STAGE.

Feature	Recognition Rate %
Without preprocessing	40.38%
With preprocessing	68.26%

- ACDN+P features;
- 3DLBP features;
- score fusion of ACDN+P and 3DLBP features.

Figure 6 shows the CMC (Cumulative Matching Characteristics) curves of the 3DLBP, HAOG, ACDN+P, and the fusions between HAOG (Histogram of Averaged Oriented Gradients) and 3DLBP, and ACDN+P and 3DLBP features. The results about HAOG are 3DLBP were obtained in [11].

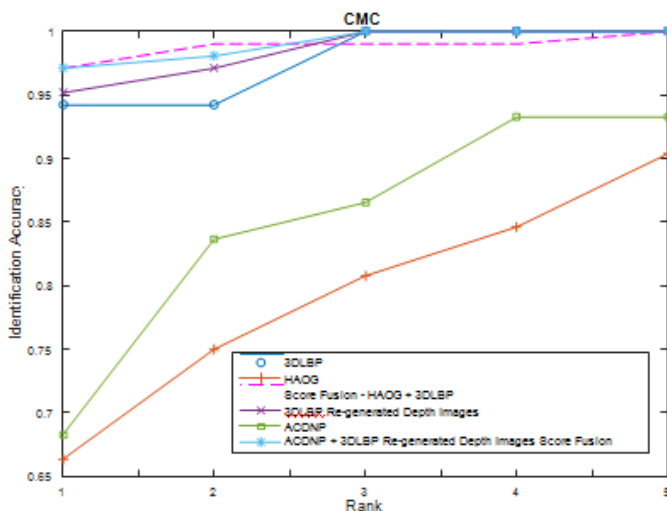


Fig. 6. CMC curves of the 3DLBP, HAOG, ACDN+P, and the fusions between HAOG and 3DLBP, and ACDN+P and 3DLBP features.

In this experiment, the optimum weight scores for the 3DLBP and ACDN+P fusion were 0.9 and 0.1, respectively. For the 3DLBP and HAOG fusion, according [11], the optimum weights were 0.8 and 0.2, respectively. One can observe that both fusions increase the overall recognition performance, and that the fusion proposed in this paper, ACDN+P and 3DLBP features, is slightly better, reaching 100% at rank 3.

One can also observe that the new generated depth map performs better than the original Kinect depth map. This is due the symmetric filling process that increases the quality of the samples.

Table IV shows the Identification Accuracy for the ACDN+P, 3DLBP, HAOG, and their fusion. One can observe that the benefit of the fusion and the new depth images are clear.

Figure 7 shows the ROC (Receiver Operating Characteristics) curves of the fusions of HAOG and 3DLBP, and ACDN+P and 3DLBP. It is possible to observe that, even though both curves are close each other, the fusion of ACDN+P and 3DLBP presented a slight advantage.

TABLE IV
COMPARISON OF THE RESULTS IN THE EXPERIMENT 2

Feature	Rank 1	Rank 2	Rank 3 %
ACDN+P + 3DLBP	97.11%	98.07%	100%
HAOG + 3DLBP	97.11%	99.03%	99.03%
ACDN+P	68.26%	83.65%	86.53%
3DLBP New depth map	95.19%	97.11%	100%
3DLBP Kinect depth map	94.23%	94.23%	100%
HAOG	66.34%	75.00%	80.76%

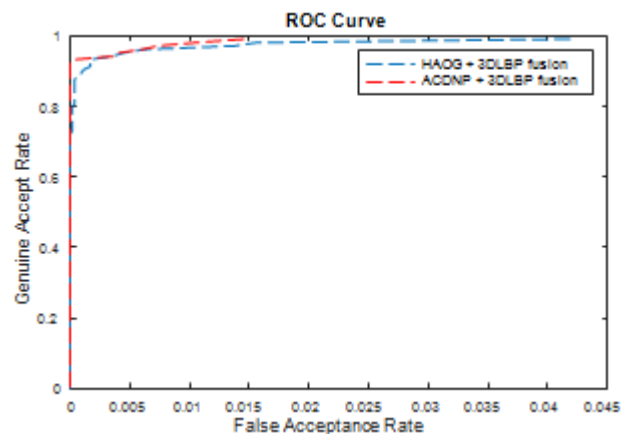


Fig. 7. The ROC Curve of the 3DLBP + ACDN+P and 3DLBP + HAOG

VIII.CONCLUSION

The experimental results show that the Symmetric Filling in combination with the Iterative Closest Point and the Savitzky-Golay filter greatly increase the recognition accuracy for both features (ACDN+P and 3DLBP). This is probably because they increase the density of the point cloud and corrects the noises generated with very small poses in the scanning process. The fusion between both low level local features helped to greatly increase the Rank 1 recognition rate and, even though there is great variation of expression and occlusion in the database, the methods performed well. In a future work it would be interesting to evaluate the process on profile images and explore, even further, the level of resistance of the proposed method to obstruction, pose and occlusion.

REFERENCES

- [1] R. Bolle and S. Pankanti, *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*, A. K. Jain, Ed. Norwell, MA, USA: Kluwer Academic Publishers, 1998.
- [2] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition, 2nd Edition*. Springer, 2011. [Online]. Available: <http://dblp.uni-trier.de/db/books/daglib/0027896.html>
- [3] A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2007.1017>
- [4] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3d face recognition approach based on the fusion of novel local low-level features." *Pattern Recognition*, vol. 46, no. 1, pp. 24–37, 2013.
- [5] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3d face recognition," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2006, pp. 90.1–90.10, doi:10.5244/C.20.90.
- [6] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987
- [7] R. W. Schafer, "What is a savitzky-golay filter? [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, July 2011
- [8] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964. [Online] Available: <http://dx.doi.org/10.1021/ac60214a047>
- [9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. New York, NY, USA: Cambridge University Press, 2007.
- [10] B. Li, A. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, 2013, pp. 186–192.
- [11] J. B. Cardia Neto and A. N. Marana, "3dlbp and haog fusion for face recognition utilizing kinect as a 3d scanner," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, ser. SAC '15. New York, NY, USA: ACM, 2015, pp. 66–73. [Online]. Available: <http://doi.acm.org/10.1145/2695664.2695807>.