International Journal of Electrical and Computer Engineering (IJECE)
Vol. 5, No. 6, December 2015, pp. 1227~1233
ISSN: 2088-8708

An Efficient Face Recognition Using SIFT Descriptor in RGB-D Images

M. I. Ouloul*, Z. Moutakki*, K. Afdel, A. Amghar***
* Department of Physics, LMTI, Ibn Zohr University, Morocco
** Department of Computer Science, LabSIV, Ibn Zohr University, Morocco

Article Info

Article history:

Received Jul 3, 2015
Revised Aug 25, 2015
Accepted Sep 13, 2015

Keyword:

Face Recognition
Identification
Keypoints
RGB-Depth
SIFT

ABSTRACT

Automatic face recognition has known a very important evolution in the last decade, due to its huge usage in the security systems. The most of facial recognition approaches use 2D image, but the problem is that this type of image is very sensible to the illumination and lighting changes. Another approach uses the 3D camera and stereo cameras as well, but it's rarely used because it requires a relatively long processing duration. A new approach rise in this field, which is based on RGB-D images produced by Kinect, this type of cameras cost less and it can be used in any environment and under any circumstances. In this work we propose a new algorithm that combines the RGB image with Depth map which is less sensible to illumination changes. We got a recognition rate of 96, 63% in rank 2.

Corresponding Author:

Mohamed Imad Ouloul,
Department of Physics,
Ibn Zohr University,
B.P 8106, Ibn Zohr University, Agadir, Morocco.
Email: md1.ouloul@gmail.com

1. INTRODUCTION

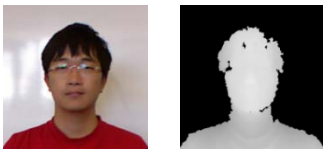
Facial recognition becomes nowadays one of the most useful methods of individual identification, due to its ability to identify from distance. This technology has a lot of applications in several domains, such as sensitive locations surveillance (airport, banks...), access-control, E-commerce. There are two approaches: one that uses 2D images, and the other that uses 3D. The first one gives good results, but they are too sensitive to illumination changes and to pose variation. On the other hand the 3D approach is less useful because it requires some sort of special sensors and a relatively long processing duration, which limits this approach from being functioned in real time's applications.

A new approach has recently appeared which uses RGB-D images (see figure 1); these types of images were produced by a Kinect camera, which was principally developed for sample usage in a computer game environment [1]. The RGB-D image combines two types of data; a 2D color image captured with a RGB camera and a depth map captured by an infrared camera that function in parallel with an infrared emitter. Due to its invariance to illumination and lightning changes, depth map can be used robustly in a non-controlled environment. However the kinect camera has a high speed of image capturing and cost less, which makes kinect an alternative to 3D sensors [2], [3], see Table 1.

In recent years, numerous methods on face recognition have appeared. The work presented in [4] uses RGB-D image produced by Kinect for facial recognition. The algorithm proposed in this work calculates entropy map and visual saliency from RGB-D image, and uses HOG [5] (Histogram of Oriental Gradients) to extract the features from these images; the different features obtained are concatenated in a single descriptor which is used to train a RDF (Random Decision Forest) classifier.

In [6], another work that consists on using only the depth maps for face recognition. The proposed algorithm is based on three steps, it starts with the segmentation of the face from the depth map, after that the face is divided into 64 regions. Then the application of 3DLBP (3D Local Binary Patterns) allows extracting the features. Finally in the third step, the face descriptor resulted is used to train the SVM (Support Vector Machines) classifier.

The advantage that lies in [4] is the simultaneous utilization of RGB images and depth map for facial recognition. However the HOG (Histogram of Oriental Gradients) descriptor used in this work is non-robust to rotation and scale changes. This problem limited the application of this work in non-controlled environments. In [6], the utilization of depth map could be insufficient to accomplish the identification in case of large similarity between faces.



(a) (b)

Figure 1. RGB-D images produced by Kinect (a): RGB (b): Depth map

Table 1. Comparison Of 3d Sensors [3]

Device	Speed (sec)	Charge (s)	Size (inch3)	Price (USD)	Acc (mm)
3dMD	0.002	10 sec	N/A	> $50k	< 0.2
Minolta	2.5	no	1408	> $50k	~ 0.1
Artec Eva	0.063	no	160.3	> $20k	~ 0.5
3D3 HDI R1	1.3	no	N/A	> $10k	> 0.3
SwissRanger	0.02	no	17.53	> $5k	~ 10
DAVID SLS	2.4	no	N/A	> $2k	~ 0.5
Kinect	0.033	no	41.25	< $200	1.5-50

To develop a performing facial recognition system that is robust to illumination and scale changes, we propose, in this paper, a new algorithm based on the application of SIFT (Scale Invariant Feature Transform) descriptor on RGB-D images produced by Kinect. The rest of the paper is organized as follows: a presentation of the proposed method in section 2, the section 3 contains an explanation of the Research method, the experiment results and analysis are in section 4, and we finish with the conclusion in section 5.

2. PROPOSED METHOD

The procedure involved in our algorithm is generally presented in five steps (Figure 2): at first we begin with face detection in RGB-D images. Then computing the saliency map and LTP (Local Ternary Patterns) corresponding respectively to RGB and Depth map. After that we use SIFT descriptor to extract the features detected in RGB, saliency Map, and depth map. Then we apply K-means algorithm to normalize the features retained in each image. At last we concatenate all features in a single vector named Face descriptor which will be used in the classification.

3. RESEARCH METHOD

In this section, we present the essential elements that are used to compose our proposed method introduced in section 2.

3.1. Scale InvariantFeature Transform (SIFT)

Scale Invariant Feature Transform Descriptor, proposed by David Lowe in [7], permits the local matching between different images by using the invariants Keypoints which are robust to scale and rotation changes [8]. The SIFT Descriptor's calculation could be accomplished in four steps:

1) Detecting the potential Keypoints in the image by using the difference of Gaussian (DoG) function presented by (Equation (1))

$$DoG(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{(x^2+y^2)}{2\sigma^2}} \quad (2)$$

Where G (Equation (2)) is the Gaussian kernel, k is the scale factor, and $I(x, y)$ is the source image.

2) The Keypoints that present a maximum or minimum are stable so we keep them. The other points are instable and they're rejected.

3) An orientation and magnitude is assigned to each keypoint.

4) Each keypoint is coded into a vector with a 128 dimensions which is invariant to scale, rotation and illumination changes.

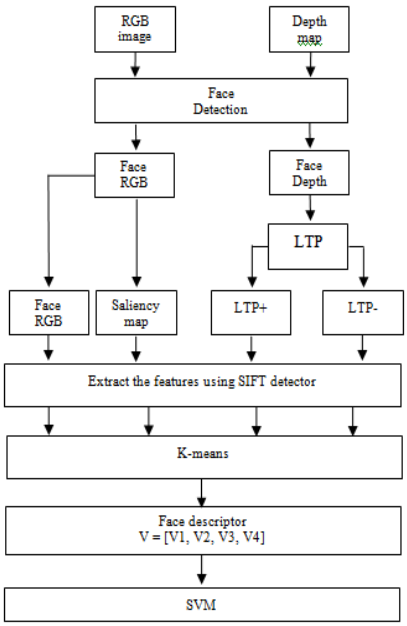
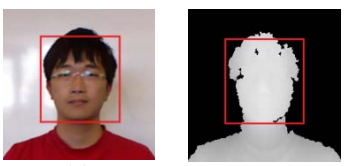


Figure 2. The proposed algorithm

3.2. Face Detection in RGB-D Images

The face detection is the first step of the Facial recognition process. It permits to localize the face in an image that can contain several objects. In our proposed algorithm we start by detecting the face in a RGB image (Figure 3(a)) using the Viola & Jones method [9]. After that we then get to localize the face in the depth map (Figure 3(b)) by using the correspondence of coordinates between RGB image and depth map.



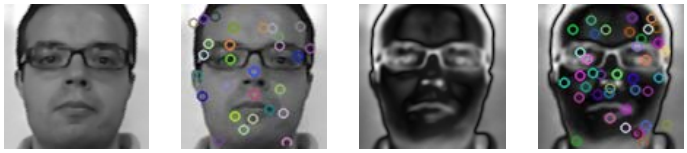
(a) (b)

Figure 3. Face detection in (a): RGB (b): Depth map

3.3. Saliency Map

Saliency is a very important technique in the computer vision domain. It is generally used in case of segmentation and detection problems. It permits to localize the most important regions in an image due to the

combination of certain characteristics such as color, intensity and orientation [10], [11]. The visual Saliency map (Figure 4(c)), produced during the application of the saliency technique on the RGB image (Figure 4(a)), presents a new and efficient source of data that helps to increase the different inter-class between images. Saliency orient SIFT descriptor to detect new keypoints (Figure 4(b) and 4(d)) in the important regions of the face (mouth, nose, eyes).



(a) (b) (c) (d)

Figure 4. (a) RGB image, (b) keypoints detected in RGB image, (c) Saliency map, (d) new keypoints detected in Saliency map

3.4. Local Ternary Patterns Images

The depth map produced by Kinect is composed of a number of pixels where each pixel presents the distance between the camera and a corresponding point in the face [1]. The problem we confront during the application of SIFT descriptor on depth map is that the number of keypoints detected is insufficient for the matching between images (Figure 5(a)). To overcome this problem, we use a pretreatment by the LTP descriptor proposed by X.Tan and B.Triggs in [12]. By applying the LTP descriptor, the depth image will be transformed into two images named the positive (Figure 5(b)) and the negative image (Figure 5(c)). The number of Keypoints detected by SIFT descriptor in each of the two images is much important than the number detected in the depth map.

(a) (b) (c)

Figure 5. Keypoints detected in (a): Depth map, (b): LTP+, (c): LTP-

3.5. SIFT Matrix

When we apply the SIFT descriptor on the image I(x,y), we detect a certain number of Keypoints N that describes the image. On one hand the number of keypoints depends on the SIFT parameters (number of octaves, edge threshold, kernel Gaussian), and on the other hand on the image type (RGB, gray-scale, depth map, binary …). All keypoints are gathered in a matrix named SIFT matrix, in which the number of columns is set on 128, and the number of lines equals N. After that the K-means algorithm transforms the SIFT matrix of RGB, Saliency map and LTB images into vectors. These vectors are then concatenated in a single vector which will be used in the classification.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Database

To evaluate the performance of our algorithm we use the EURECOM database. It is composed of 52 subjects: 38 males and 14 females from different ethnic groups. Database images are taken from two separated sessions. Each session contains faces of different expressions and positioning (see Figure 6) (Neutral, lightning changes, smiling , opened mouth, occlusion eyes, occlusion mouth, occlusion paper, left profile, right profile) [13]. To overcome the problem of face detection, we only focus on the first four images.

Figure 6. Variation in expression illumination and pose

4.2. Results and Analysis

A facial recognition system can have two different operations: whether in verification mode, in which the system decides if the identity proclaimed by the person true or false, or in identification mode where the system have to find the person's identity within the existed identities in the Database [14], [15]. For our system can be functioned in non-controlled environments, we use it in the identification mode.

The proposed algorithm in our work is based on the feature extraction from different types of images (RGB, Saliency map, LTP image). To evaluate the utility of the extracted features, we tested our system in three different cases: (1) RGB only, (2) RGB + Saliency map, (3) RGB + Saliency map+ LTP images. Each case presents the functioning of our system with one or several types of images. From the curve presented in Figure 7, we can analyze the evolution of our system depending on the types of the used images. Thus, the functioning of the system in case (2) RGB + Saliency map is effective compared to the functioning of the same system in case (1) RGB only. However, the functioning in case (3) RGB + Saliency map+ LTP is the most efficient.

The improvement of our recognition system is composed of two steps. The first step consists in using Saliency map to detect a new keypoints, the increase of the detected keypoints by SIFT involves the increase of inter-class differences between images of persons during the learning phase. This first step allows us to increase the recognition rate by 6% to rank 2. In the second step, we add depth image. This type of image is characterized by its resistance against the illuminations changes [16]. This feature reduces the intra-class differences between the images of the same person due to the illumination changes. The addition of the depth image in the system has allowed reaching a recognition rate equal to 96.63% to rank 2.

To validate this work, we have compared the results obtained using our approach to other existing methods: Local binary patterns (LBP), Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). Based on the comparison results presented in Table 2 and Figure 8, we notice that the recognition rate in the methods (LBP, LDA and PCA) has not exceeded 90% to rank 2. However, the recognition rate in our method has reached 96.63% to rank 2. We can deduce that our approach is significantly more efficient.

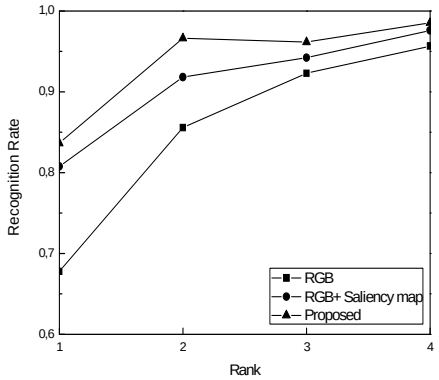


Figure 7. Cumulative Match Characteristics curve illustrating performance of the proposed algorithm with each data type used

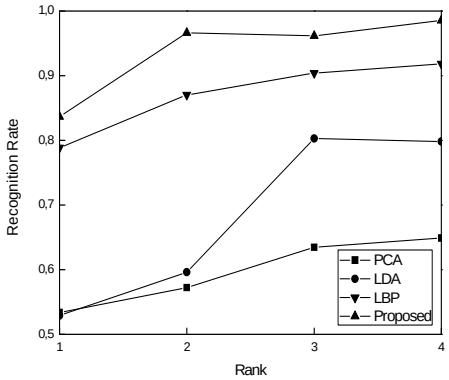


Figure 8. Comparing the results obtained using our proposed algorithm to other existed methods in literature

Table 2. Comparing the Recognition Rate (%) of our Method to other existing Methods

Method	Rank1	Rank2	Rank3	Rank4
LBP	78.84%	87.01%	90.38%	91.82%
LDA	52.88%	59.61%	80.28%	79.80%
PCA	53.36%	57.21%	63.46%	64.90%
Proposed	**83.65%**	**96.63%**	**96.15%**	**98.55%**

5. CONCLUSION

The approach that uses the RGB-D images, produced by kinect, are suitable for the real time facial recognition systems in non-controlled environments; where lightning and illumination are variants.

In this work we have proposed a new facial recognition algorithm that uses the RGB-D images, It is based on the extraction and the concatenation of the SIFT descriptors from these data sources (RGB, Saliency map, LTP images).The performance of our algorithm has been validated by testing it with the EURECOM database. The algorithm we have proposed in this paper can be developed in a future work and maybe will be used in case of images with occlusion and pose variations.

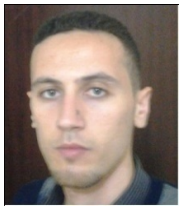
ACKNOWLEDGEMENTS

This work was supported by the National Center for Scientific and Technical Research (CNRST).

REFERENCES

[1] K. Khoshelham, S. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
[2] Anapu *et al*., "Fusion of RGB and Depth Images for Robust Face Recognition using Close-Range 3D Camera," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 7, 2014.
[3] Li, B. Y *et al*., "A. Using kinect for face recognition under varying poses, expressions, illumination and disguise," *in IEEE Workshop Applications of Computer Vision*, pp. 186 – 192.
[4] G. Goswami, *et al*., "On RGB-D Face Recognition using Kinect," *in IEEE Sixth International Conference Biometrics: Theory, Applications and Systems 2013*, pp. 1 - 6.
[5] Dalal N. and Triggs B., "Histograms of oriented gradients for human detection," *in IEEE Computer Society Computer Vision and Pattern Recognition CVPR*, 2005, pp. 886 – 893.
[6] Neto J. B. C. and Marana, A. N., "Face Recognition Using 3DLBP Method Applied to Depth Maps Obtained from Kinect Sensors", *in x workshop computer vision WVC*, 2014. pp. 168 – 172.
[7] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, n. 2, pp. 91-110, 2004.
[8] Luo J., Ma Y., Takikawa E., Lao S., Kawade M., Lu B. L. "Person-specific SIFT features for face recognition," *in IEEE International Conference Acoustics, Speech and Signal Processing 2007*, pp. 1520-6149.
[9] Viola P., and Jones M. "Rapid object detection using a boosted cascade of simple features," *in Proceedings of the 2001 IEEE Computer Society Conference Computer Vision and Pattern Recognition CVPR 2001*. Vol.1 pp. 511-518.
[10] L. Itti, C. Koch, E. Niebur, *et al*., "A Model of SaliencyBased Visual Attention for Rapid Scene Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, n. 11, pp.1254–1259, 1998.
[11] Liu, R., Cao, J., Lin, Z., Shan, S., "Adaptive partial differential equation learning for visual saliency detection", in IEEE conference Computer Vision and Pattern Recognition 2014. pp. 3866 – 3873.
[12] Tan X., and Triggs B., "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.
[13] Rui Min, Neslihan Kose, Jean-Luc Dugelay, "KinectFaceDB: A Kinect Database for Face Recognition", *IEEE Transactions on systems man and cybernetics*, vol. 44, no.11, pp.1534-1548, 2014.
[14] Anouar Mellakh, "Reconnaissance des visages en conditions dégradées," Ph.D. Thesis, Institut National des Télécommunications, France, 2009.
[15] S. G. ABABSA, "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D," Ph.D. Thesis, Université Evry Val d'Essonne, France, 2008.
[16] HAN Jungong, SHAO Ling, XU Dong, *et al*., "Enhanced computer vision with Microsoft kinect sensor: A review." *Cybernetics, IEEE Transactions on, 2013*, vol. 43, no 5, p. 1318-1334.

BIOGRAPHIES OF AUTHORS



Ouloul Mohamed Imad was born in Taza, Morocco, in 30/07/1989, he got the Master thesis Instrumentation and Telecommunication in 2013 from the University of Ibn Zohr, Agadir, Morocco. Since November 2013, he prepared his Ph.D thesis on computer vision and embedded systems. His main scientific interests are face detection/recognition and complex systems based on FPGA board. He is current interests lie in face recognition using RGB-Depth images produced by Kinect.

Moutakki Zakaria was born in casablanca, Morocco, in 25/07/1988, he got the master thesis Instrumentation and Telecommunication in 2011 from the University of Ibn Zohr, Agadir, Morocco. Since October 2011, he prepared his Ph. D thesis on computer vision and embedded systems. His main interests are the video surveillance systems applied on road safety, traffic management and the embedded systems using FPGA boards. he is currently interested on the optimization of the detection accuracy in road traffic surveillance systems.

Afdel Karim is a Professor in the Computer science Department, Faculty of Science, University Ibn Zohr , Morocco. He received the Doctorat (French Ph.D) in Computer Engineering and Medical Image Processing from the University of Aix Provence France.
His areas of research interests include Image Processing and Analysis, computer Vision, and Medical Image

Amghar Abdellah is a Professor in the Physics Department, Faculty of Science, University Ibn Zohr ,Morocco. He received his DEA and DES degree in 1994 from Department of Physics , Faculty of Science, University Hassan II , Morocco. In January 2002, he has Ph.D degree in microelectronic from Department of Physics, Faculty of Science, University Ibn Zohr ,Morocco. His areas of research interests include Cryptography, DNT, embedded systems and microelectronic.