EPiC Series in Computing

Volume 69, 2020, Pages 393–402

Proceedings of 35th International Conference on Computers and Their Applications

Age-group Classification Using 3DHOG Descriptor Applied to Depth Maps

Nabila Mansouri[1,2], Bougueddima Hana[3], and Ben Jemaa Yousra[3]

[1] Ha'il University, Ha'il, KSA
[2] Sfax University , ReDCAD Laboratory, Sfax, Tunisia
nabila.elmansouri@gmail.com
[3] National School of Engineers-Tunisia, U2S Laboratory, Tunisia.
Yousra.BenJemaa@enis.rnu.tn

Abstract

Age estimation has lots of real-world applications, such as security control, biometrics, customer relationship management, entertainment and cosmetology. In fact, facial age estimation has gained wide popularity in recent years. Despite numerous research efforts and advances in the last decade, traditional human age-group recognition with the sequence of 2D color images is still a challenging problem. The goal of this work is to recognize human age-group only using depth maps without additional joints information. As a practical solution, we present a novel representation of global appearance of aging-effect such as wrinkles' depth. The proposed framework relay, first-of-all, on an extended version of Viola-Jones algorithm for face and region of interest (most affected by aging) extraction. Then, the 3D histogram of oriented gradients is used to describe local appearances and shapes of the depth map, for more compact and discriminative aging effect representation. The presented method has been compared with the state-of-the-art 2D-approaches on public datasets. The experimental results demonstrate that our approach achieves a better and more stable performances.

1 Introduction

Face analysis problems have been extensively studied using conventional RGB cameras at visible light. However, this makes some face analysis tasks, such as age estimation, a challenging problem. Indeed, despite numerous research efforts and advances in the last decade, traditional human age-group estimation using 2D color images is still a challenging problem [1].

Human face appearance is in essential continuous evolution caused by aging effect like drooping upper eyelids and cheeks, darkened skin, wrinkless,etc . Although, 2D representation of such as effects can be effectively described by shape, texture,etc of face, depth evolution of these parameters can't be clearly shown. For instance, wrinkles' depth can't be clearly shown only by three-dimensional face representation. Furthermore, face images acquired using such conventional sensors may have inherent restrictions that hinder the inference of some specific information in the face. Indeed, due to the difficulty in dealing with pose and illumination and aging effects evolution on 2D face images, 3D age-group estimation methods can be the trend

in the last years. 3D data are less sensible to illumination changes and, more important, is very useful to correct face spoofing, illumination and pose changes [1].

The main drawback of using 3D based methods is the high cost of the traditional 3D sensors. One alternative to these expensive scanners are the Kinect devices (Figure 1) [2,3]. That, besides being considerable cheaper, are able to capture quite precisely the depth information, together with RGB color images (Figure 2). The recent introduction of low-cost depth cameras (such as Microsoft Kinect) provides exciting new opportunities for computer vision and face analysis research. That can clearly underscore the aging effect depth especially the wrinkles' depth.

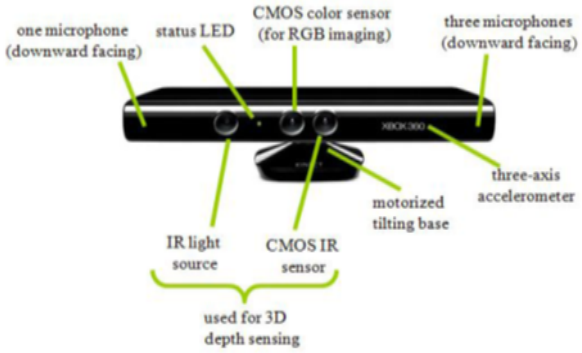

Figure 1: Presentation of the camera-Kinect's sensors.

Figure 2: Exemples of images captured by the Kinect: a) RGB image, b) Depth map.

The main goal of this work is to propose and assess the performance of a new 3D descriptor for age-group recognition based on depth maps generated by Kinect devices. Another goal is to investigate the face regions in which the depth maps contribute most to highlight the aging effect.

Paper remainder is as follows. In section 2 we presente a quick galence of age estimation methods literature and we review some works related to the use of Kinect depth images for face analysis. Then, in section 3 we detailled the proposed algorithm for depth aging features extraction. Beside, in order to illustrate the performances of the proposed method, an experimental and a comparative studies are presented in the section 5. Finally the paper is ended by

a conclusion and the futur work.

2 Related work

In the past few years manifold approaches have been proposed for studying age group recognition problem based on RGB images. There are two categories of approaches: Cranio-Facial Approaches [4–9] and Behavioral Approaches [10, 11]. Although, behavioral approaches are very promising research axe that give good performances when person is far from the camera, face remains the most affected body region by aging effect. Face appearence is widely described by 2D descriptors that aim to charecterize the shape, texture or both of them. Such as instances, [4] proposes a hybrid descriptor based on the Active Appearance Model (AAM) in order to extract the shape's features based on the 64 landmarks detected manually. Also, the Local Phase Quantization (LPQ) was used as a face's texture descriptor. The work proposed in [8] uses the Active Shape Model ASM to characterise the face's shape by 68 landmarks. Then, the Radon Transformation RT was used to extract the texture's features in the polar coordinates. Finally, the Discret Cosinus Transformation (DCT) is applied to extract just the 25% frequency's features. The experimental results have condacted on the FG-NET database give 4.18 years as MAE value. An other study presented in [9] propose shape descriptors based on the Delauney triangulation to segment the image into regions in order to locate the eyes, the nose and the mouth. Then, the Hilbert transformation [9] is used to extract the high frequency. The color histogram is then used, in order to eliminate the shady. The texture features is computed based on Local Binary Pattern (LBP) descriptor. The LBP means to divide the face's image into m regions and for each region, spatial histograms of local binary patterns were produced. The concatenation of these regional histograms builds a global descriptor of the image. Then, the Principal Component Analysis (PCA) has been applied in order to reduce the features vectors dimensionality. The experimental results on the FG-NET database and the proposed $INDIAN - DB$ one, show a satisactory performances with an MAE value equal to 5.780 years and 3.950 years. Moreover the work presented in [7] proposed another novel approach to charecterize the face's appearence by applying the HOG descriptor which is tried to locate the wrinkles contours and orientations. The experimental results have shown a good accuracy for age estimation to the order of 80%.

Face-based age estimation problems have been mainly extensively studied using conventional RGB cameras at visible light. However, this makes some aging features extraction a challenging problem. Furthermore, face images acquired using such conventional sensors may have inherent restrictions that hinder the inference of some specific aging information in the face such as wrinkles' depth.

Microsoft Kinect is introduced in 2010. It is widely adopted by the computer vision research community in various applications [2] of face analysis such as face [12–16], gender [17, 18] and ethnicity [19] recognition.

It appears that most of the few attempts on using Kinect in face analysis are mainly devoted to the face recognition problem, gender recognition and ethnicity [1], hence overlooking and ignoring other face analysis tasks such as age estimation. Moreover, most of the proposed works focused on the fusion of Kinect depth information and RGB images but did not explicitly explore how much information Kinect facial depth data alone can reveal about the faces [1]. Some of the results are also reported on size-limited and/or private Kinect databases. Also problem of age labeled subjects in the available Kinect databases.

3 Proposed Algorithm overview

The proposed algorithm consists of 3 steps (Figure 3): Preprocessing, Features extraction and Age-group recognition. The preprocessing phase consiste of the depth and RGB image alignment in order to extract face and regions of interest. Thus features are extracted from these faces' regions. Finally the classifieur is trained by the gallery images and evaluated on the probe ones.

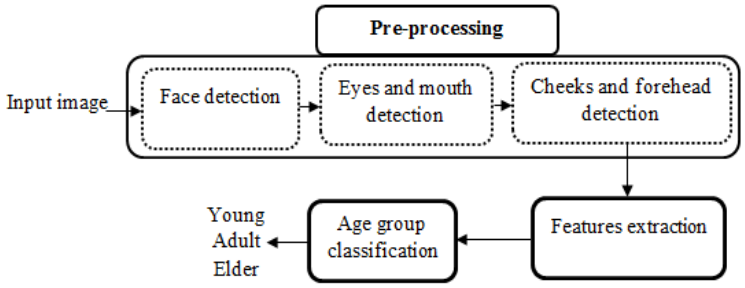

Figure 3: Steps of the proposed algorithm

3.1 Preprocessing phase

The first step is the preprocessing phase which consists on the face detection and the interest face's regions segmentation. Indeed, to locate these regions in the initial input image we rely on the Viola-Jones algorithm [20]. In fact, all human faces share some similar properties such as the eye region is darker than the upper-cheeks and the nose bridge region is brighter than the eyes. These regularities may be matched using Haar Features to recognize in the first time the face based on the RGB image. Then, a mapping, between RGB and RGB-D images, by the conservation of the same dimensions is carried out in order to localize face on the RGB-D map (Figure 4 b)).

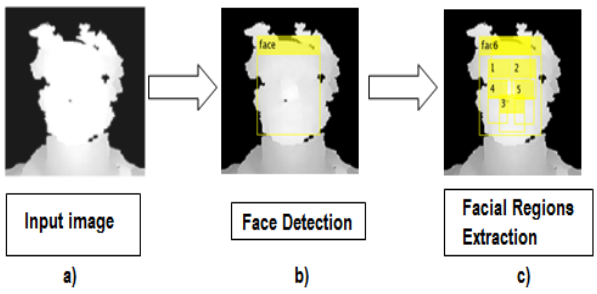

Figure 4: The preprocessing flowchart.

In the progress of aging, variations of textural features like fine wrinkles and other skin artifacts are manifested on the face skin [21]. These textural variations are clear on the forehead, eye-corners, below the eyes and near the cheekbones and the mouth. In order to define these regions, we detected the eye's regions (regions 1 and 2 in Figure 4 c)) and the mouth's region

(region 3 in Figure 4 c)) based on the Viola-Jones algorithm [20] which considered that eye's region similar to a dark region on the face while the mouth's region is detected based on the location of the eyes. Since the RGB images and their corresponding depth maps are registered, corresponding area is segmented from the depth map as well. Further, to detect the cheekbones ((regions 4 and 5 in Figure 4 c)) and the forehead ((region 6 in Figure 4 c)) regions we based on the eyes location. In fact, when the eyes are detected, we can define forehead and the cheekbone areas relatively to eyes positions especially using the distance D_{eye} between the two detected eyes as follows: Forehead's region width and height are $\frac{3}{4} \times D_{eye}$ and $\frac{2}{3} \times D_{eye}$ respectively. However the Cheekbone's region dimensions are $\frac{1}{3} \times D_{eye}$ and $\frac{3}{5} \times D_{eye}$ respectively for width and height [21].

3.2 Extracting depth features using HOG descriptor

HOG [22] is a robust descriptor that has been successfully used in many applications such as object detection and recognition [7, 8]. This descriptor is known as a robustness descriptor significantly outperforms compared with the others on pateren recognition.
The HOG technique figures out the distribution of intensity gradients or edge directions in a local patch of an image [22]. Since the descriptor operates on localized cells, the method upholds invariance to geometric and photometric transformations.

Beside these advantages mentioned above, the basic motivation behind using HoG descriptor in the age-group recognition is that the object appearance and shape can be characterized by the distribution of local intensity gradients or edge directions. Such as reprisentation can properly detecte aging effect espicily the wrinkles. However, considering that 2D face images acquired using the conventional sensors (traditional RGB cameras) may have inherent restrictions that hinder the inference of some specific information in the face. Thus, the classical 2D HOG descriptor describes properly the face's appearence and detect the first primary appearance of aging effect but it can not deal with their accentuation and digging evolution. To overcome this limitation we propose the 3DHOG method to describe the local distribution of the gradient intensity, direction and depth evolution of the depth map. Indeed, the new approach of the 3DHOG applies the HOG features extraction process on the depth maps. These depth maps are extracted from face's interst regions R1,etc and R6. The proposed algorithm computes HOG of depth maps by following these steps:

- Gamma correction is applied to the original image's window to overcome the variation of brightness and contrast.
- Respectively vertical and horizontal gradient are computed and modeled to form the image contours.
- Create histograms of orientation gradients. Referring to [22], is done in the square cells of small size (3×3 pixels). Each pixel of the cell vote for one class of the histogram, depending on the pixel's gradient orientation. The vote is weighted by the pixel's gradient intensity. Histograms are unsigned $[0; \Pi]$ or signed $[0; 2 \times \Pi]$. In the original work [22], non-signed orientations have been found to perform better. Therefore, a histogram of 9 bins (orientations) evenly spaced over 0 to 180 degrees is used.
- Block construction by grouping a 2×2 cells and perform a L1-racine normalization.

The 3DHOG feature vector extraction process is shown in Figure 5.

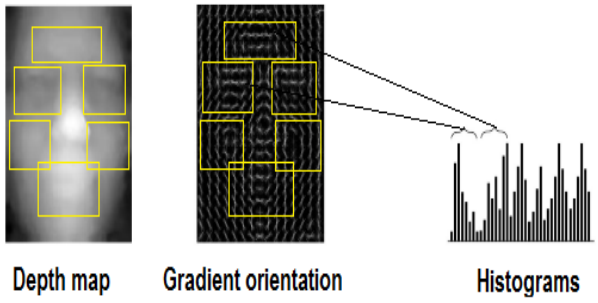

Figure 5: Extraction process of HOG feature.

4 Experimental study

4.1 RGB-D face databases

4.1.1 Existing databases

Kinect is a new hardware device that is recently used in computer vision applications. Indeed, there exist a few RGB-D face databases with subject age annotation which are publicly available [23] [24]. Two existing databases filmed with Kinect such as,

EURECOM database The EurecomKinect face database [23] contains both RGB and depth facial images of 52 subjects acquired using Kinect sensor. The people in the database belong to 2 different age-groups (young and adult). The data is captured in two sessions separated by two weeks. In each session, the facial images of each person are captured under 9 different facial variations (neutral, smile, open mouth, strong light, eyes occlusion, mouth occlusion, paper occlusion, left profile and right profile. Face image samples from this database are shown in Figure 6 a).

Superface-Kinect-Face-dataset The Superface-Kinect-Face-dataset [24] (Figure 6 b)) contained simultaneously a various sequence of different positions of the 2D and 3D facial image captured by the Kinect camera. It contained 920 images pertaining to 20 subjects for the young and the adult age.

4.1.2 Proposed database: Seniors Help Center of Sfax dataset: SHCS-Kinect-dataset

The age range available in EURECOM and Superface-Kinect-Face databases is that of young and adult. Thus we filmed a new database (Figure 6 c)) of elder subject using the camera Kinect in the Seniors Help Center of Sfax. The elder subjects in our dataset are added to extend already presented bases by adding extra age-group. The proposed RGB-D face database contains 13 subjects where their age is varied between 60 years and 80 years. Thus working corpus is composed of 3 age-group.

Figure 6: Datasets samples: a) EURECOM-Kinect dataset [23], b) Superface-Kinect-Face dataset [24] c) SHCS-Kinect-dataset.

4.2 Test protocol

In this subsection, we present the adopted test protocol. First-of-all a learning phase is done based on 60% of the already presented 3 databases. Then the remain 40% of experimental data are used as the probe images. All of the galery and probe images are resized to have the same size. Second age-group recognition purpose is to lock the correct age-class (Young, Adult, Elder) of the query face. Age-groups are defined as follows:

- Young class: person's age $<$ 30 years
- Adult class: person's $>=$ 30 age $<$ 60 years
- Elder class: person's age $>=$ 60 years

The SVM is used as classifier and the True recognition Rate (TR), the Recall and the Precision are used as a metrics to quantify the proposed descriptor's performances in terms of age group estimation.

While, precision evaluates a system's accuracy based on returned correct answers and real data labels, the recall describes a system's ability to correctly classify items to their related classes. CCR is used to evaluate descriptor general accuracy. Based on matrix confusion composition [10], Recall, precision are measured as averages of three tested classes (Young, Adult, Elder).

4.3 Results and discussion

Tables 1 summarizes the classification performance of the the proposed 3DHOG descriptor on the 3 Kinect face databases for face based age-group recognition. These results point out several findings:

Referring to Table 1 we can highlight the performances of the proposed process in term of age-group recognition. In fact we reach 83.33% as recognition rate. A close look at the results in Table 1 indicates that both recall and precision values are good and closed. This proves that the 3DHOG descriptor applied on the face's region of interrest has a good ability to discriminate

Table 1: 3DHOG's performances in terms of age-group estimation.

	Classification rate CCR (%)	Recall (%)	Precision (%)
3DHOG	83.3	63.4	82.5

between age-groups (young, adult, elder).
This experiment provided evidence that depth map images can enhance the aging effect representation in order to considerably improve the age-group classification performances.

Depth map performances will be strengthen by a comparative study with some 2D descriptors among the best existing ones in face-based age estimation.

4.4 Comparative study

A comparative study with literature work using the standard 2D descriptors is also made to highlight the robustess of our approach especially in term of recognition rate, recall and precision, as presented in Table 2.
In fact, we present here a comparative study with 3 works among the state-of-the art descriptors (HOG [7], LPQ [4], LBP [9])in age-group estimation application which reach the best performances.

Table 2: Depth features vs 2D performances in terms of age-group estimation.

		Classification rate CCR (%)	Recall (%)	Precision (%)
4*2D	HOG [7]	75	36.1	64.5
	LPQ [4]	70.8	28.7	44.4
	LBP [9]	66.6	30.7	51.3
3D	3DHOG	83.3	63.4	82.5

Of all features, only the 3DHOG performs as well as standard HOG for age-group recognition. Depth image HOG, has about 8% more RC than HOG, which is significantly highlight the performances of the depth gradients.

This good enhancement recognition rate is justified by the optimization of HOG features competing process that will be directly focalized on the wrinkles' orientation and depth evolution. This method reduces considerably the false alarms do to using only RGB image which represent only the aging effect appearance.

Although, appearance and shape features computed from 2D images can properly describe target object especially the change of face's color and/or texture ones. The important idea behind the 3DHOG descriptor is that the appearance and the local shape variation caused by aging especially the wrinkles' accentuation and digging can be better designed by the distribution of the gradient intensity or direction of the depth map. Indeed the 3DHOG features enhance conciderably the age-group classification results.

Moreover, the result is achieved with relatively small feature sets: only the RI are concidered and not all the face. This fact indicates that our aging features representation method is highly discriminative as well as computationally efficient.

5 Conclusion

The main contributions of this work include two aspects. First, we propose the Depth description of the face appeareance as a new way of describing the global 3D shape of a face's aging. It is a 3D depth map which represents a face's regions of interest. Second, the proposed approach yields the best accuracy when compared with many previous state-of-the-art age recognition methods based on 2D face representation.

This paper attempts to engineer a new, generalizable class of depth features based on appaerence distributions and uses a simple category recognition framework to evaluate their performance in contrast to more traditional approaches. The results suggest that the discriminative power of appaerance distributions in the context of depth maps is in fact very important; indeed, the 3DHOG out-performs standard HOG also the LBP and LPQ. Thus, demonstrating the importance of depth data.

There remains much exploration to do in this area in order to extend the work to other face analysis related tasks including age estimation as a classification and/or regression problem combining RGB and depth facial information.

References

[1] E. Boutellaa, M. Bengherabi, S. Ait-Aoudia and A. Hadid.: European Conference on Computer Vision, Zurich, Switzerland, pp 725-736, 2014.

[2] J. Han, L. Shao, D. Xu, J. Shotton.: Enhanced computer vision with Microsoft Kinect sensor: A review. Cybernetics, IEEE Transactions on vol 43, no. 5 pp. 1318–1334, 2013.

[3] M.Andersen,T. Jensen, P. Lisouski, A.Hansen, T. Gregersen, P.Ahrendt.: Kinect depth sensor evaluation for computer vision applications. Technical report, Department of Engineering, Aarhus University, Denmark, 2012.

[4] Sung. E. Ch., Youn. J. L., Sung. J. L. and Jaihie. K. Hierarchical age estimation from unconstrained facial images, Pattern Recognition pp.1262-1281, 2014.

[5] Song.Z. Bingbing,Ni. , Dong. G., Terence. S., and Shuicheng. Y. Learning Universal Multi-view Age Estimator by Video Context, In proc. International Conference on Computer Vision, Barcelona, Spain. pp. 1-8, 2010.

[6] Yanchao. S., Haizhou. A. and Shihong. L. Real-time face alignment with tarcking in video. In Proc. International Conference on Image Processing, San Diego, CA, USA. pp. 1632-1635, 2008.

[7] Hajizadeh. M. A. and Ebrahimnezhad. H. Classification of age groups from facial image using histograms of oriented gradients, In Proc. Machine Vision and Image Processing. Tehran, Iran. pp. 1-5, 2011.

[8] Jing-Ming G., Yu-Min. L., Hoang-Son. N. Human face age estimation with adaptive hybrid features. In proc. International conference on system science and engineering. Macau, China, pp. 55-58, 2011.

[9] Selvi. V. T. and Vani. K. An efficient age estimation system based on multi linear principal component analysis. Journal of Computer Science, pp. 1497-1504, 2011.

[10] N. Mansouri, M. Aouled Issa, Y. Ben Jemaa.: Gait-based human age classification using a silhouette model, IET Biometrics vol 7, no. 2, pp: 116-124, 2018.

[11] N. Mansouri, M. Aouled Issa, Y. Ben Jemaa.: Gait features fusion for efficient automatic age classification, IET Computer Vision vol. 12, no. 1, pp: 69-75, 2018.

[12] B. Li, A. Mian, W. Liu, A. Krishna: Using Kinect for face recognition under varying poses, expressions, illumination and disguise. IEEE Workshop on Applications of Computer Vision, Florida, USA, pp. 186–192, 2013.

[13] G. Goswami, S. Bharadwaj, M. Vatsa, R. Singh: On RGB-D face recognition using kinect. International Conference on Biometrics: Theory, Applications and Systems, pp. 1–6, 2013.

[14] R.Min, J.Choi, G. Medioni, J.Dugelay: Real-time 3D face identification from a depth camera. International Conference on Pattern Recognition, Tsukuba, Japan, pp. 1739–1742, 2012.

[15] M. Pamplona Segundo, S. Sarkar, D. Goldgof, L.Silva, O.Bellon: Continuous 3D face authentication using rgb-d cameras. In: IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Portland, Oregon, pp. 64–69, 2013.

[16] B. Li, A. Mian, W. Liu and A. Krishna, "Face recognition based on Kinect", Pattern Analysis and Applications, pp. 1-11, 2015.

[17] G. Fanelli, T. Weise, J. Gall, L. Gool: Real time head pose estimation from consumer depth cameras. In: Pattern Recognition. Vol. 68, no. 35., pp. 101–110, 2011.

[18] Y. Huang, Y. Wang, T. Tan: Combining statistics of geometrical and correlative features for 3D face recognition. In: British Conferenceon Machine Vision, pp. 879–888, 2006.

[19] L. Xiaoguang , A. K. Jain: Ethnicity Identification from Face Images. Biometric Technology for Human Identification, Orlando, Florida, United States, 2004.

[20] Vikram. K. and Padmavathi. S. Facial parts detection using Viola Jones algorithm, In proc. International Conference on advanced Computing and communication Systems. Coimbatore, India, pp. 137-154, 2017.

[21] Moghadam. H. and fard. S. K: Human age-group estimation based on ANFIS using the HOG and LBP FEATURES. Electrical and Electronics Engineering: An International Journal, Vol. 2, No. 1, pp. 21-29, 2013.

[22] N. Dalal, B. Triggs: Histograms of Oriented Gradients for Human Detection, International Conference on Computer Vision and Pattern Recognition, San Diego, United States, pp.886-893, 2005.

[23] Min. R, Kose. N and Dugelay. J. L. KinectFaceDB: A Kinect database for face recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp.1534-1548, 2014.

[24] Stefano. B. , Alberto. D. B. and Pietro. P. Superfaces: A Super-Resolution Model for 3D Faces. In proc. Eropean Conference on Computer Vision, Workshops, Firenze, Italy, pp. 73-82, 2012.