Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences 54 (4): 365–374 (2017)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Fusion of Color and Depth Information for Facial Recognition using a Multi Perspective Approach

Mudassir Aman[1], Waqar Shah[1], Ihtesham-ul-Islam[2], Bilal Jan[2], Muhammad Amir*[1], and Bilal-ur-Rehman[1]

[1]Electrical Engineering Department, University of Engineering and Technology, Peshawar, Pakistan
[2]Department of Computer Science and IT, Sarhad University, Peshawar, Pakistan

Abstract*:* Facial-recognition is an explored and demanding task. Previously, mostly color (RGB) images were used to tackle it. Recently, advances in 3D scanners have been providing extra facial information. This new information improves the performance of current facial recognition architectures. In this research, both RGB and depth image information were utilized for addressing the problem of facial-recognition and by characterizing each image with the use of multi-perspective-approach (MPA). Data were combined from different textural-image-descriptors (TIDs) while keeping the most relevant features. Feature vectors resulting from such combinations were entered into a random-forest-classifier (RFC) to obtain a comparative analysis through the EURECOM facial dataset. The outcomes of our case-studies are comprehensively elaborated in this paper.

Keywords: 2D face, 3D face, RGBD source, face recognition, feature fusion, depth image

1. INTRODUCTION

Personal identity assurances by great majority of security systems rely on passwords, access cards or both combined. These types of identification related entities are based either on something a person knows in form of a password or possesses in shape of a security card. So it is very possible for an impostor to get hold of a password or an access card which can compromise the security of an organization [1]. A stolen password or an access card of a genuine person makes it impossible for an organizational security system to differentiate between a genuine person and an impostor.

Biometric analysis of a person physically is the best method to overcome such shortcomings of security systems. Such analysis is based on what a person anatomically has not on what he or she remembers or physically carries.

Human-based biometric characteristics are related to human physiology (e.g. face, fingerprint, iris, hand veins, etc.) or they can be behavioral (e.g., gait, voice, typing dynamics, signature, etc.). The satisfaction of certain human characteristics makes it possible to utilize it for biometrics recognition [2].

Requirements based on human characteristics are:

The characteristic must occur in as many people as possible. This is known as Universality.

The characteristic must be different from a person to another. This is known as Uniqueness.

The characteristic should not vary over time known as Permanence character.

The collection of the characteristic must be easy signifying its “Collectability”.

The “Performance Characteristic” must allow high accuracy having the least processing time plus low computational overhead.

The Characteristic should be “Acceptable” for the subjects that are going to be identified.

The characteristic must be difficult to bypass known to be “difficult to Circumvent”.

Received, August 2017; Accepted, November 2017
*Corresponding author: Muhammad Amir; Email: amir@nwfpuet.edu.pk

Table 1. Biometric characteristics comparison [4].

Requirements	Faces	Fingerprints	Irises	Gait (Walk)
1. Universality	HIGH	MEDIUM	HIGH	MEDIUM
2. Uniqueness	LOW	HIGH	HIGH	LOW
3. Permanence	MEDIUM	HIGH	HIGH	LOW
4. Collectability	HIGH	MEDIUM	MEDIUM	HIGH
5. Performance	LOW	HIGH	HIGH	LOW
6. Acceptability	HIGH	MEDIUM	LOW	HIGH
7. Circumvention	LOW	MEDIUM	HIGH	MEDIUM

Table 1 highlights a comparison about the most pivotal and recognized biometric-characteristics with respect to the above elaborated requirements. From the table it is evident that face has certain advantages over the other characteristics as it portrays higher universality, higher collectability with universal acceptability. Humans are known to naturally ID another human by analyzing their facial characteristics. Such information can be obtained from a distance and in a discrete way [3]. However, this biometrics analysis is not without some limitations. Such as: in 2 dimensional facial-recognition ambient illuminations, subject occlusion and variations in pose can decrease the system's performance. As a fact, the human facial structure constitutes a 3D object. The prescribed way of dealing with illumination problem and pose changes is to utilize a 3-dimensional or 2.5-dimensional representation of the human face. However, dealing with 3-dimensional data gives rise to certain problems such as: high cost of 3D sensors.

Specific pricing of popular 3D sensors accompanied by other important characteristics are compared by Li et al. [4]. From such comparison it is evident that Kinect distinctly arises as a cost effective alternative than other expensive 3D devices.

From environment, Kinect sensor procures depth data of the 3D objects. This captured data can be used to handle pose changes, illumination & facial expression changes while doing recognition.

Apart from lower cost, Kinect also possesses another advantage that of speed e.g. if a realistic scenario is considered, it is infeasible to wait long for a device in order to get a facial scan e.g. 2.5s is the average time taken by some off the shelf sensors. On the other hand, Kinect face scanning time is only 0.033s.

In recent times a lot of applications such as object identification, surface modeling and tracking, simulating indoor locations, locating objects in video frames and machine vision have utilized RGB-D images [5, 6, 7]. The authors in [8, 9] have implemented face recognition and detection and gender classification using RGB-D information. Li et al. [4] proposed a method to recognize human faces in the presence of variations. The authors propose a method performed on data acquired by a low resolution 3D sensor for robust facial identification in varying conditions. The preprocessing involved in this method utilized symmetry property of the face at 3D point cloud level to attain a known frontal pose, shape and texture of the face regardless of the original pose. Noise is removed from depth information by applying smoothening which fills the holes present in the depth map. The method is an integration of Discriminant Color Space transform and sparse coding. Experiments have been done on more than 5000 images acquired from a publicly available database of RGB-D images with variations in poses, expressions, lighting and occlusion. The images from the Kinect sensor record the recognition accuracy of 96.7% for the RGB-D data and 88.7% for the depth information individually. An inbuilt recognition algorithm of Kinect has been discussed by Cao et al. [10].

Another interesting work [11] presents a continuous 3D face authentication system that uses a RGB-D camera to monitor the accessing user and ensure that only the authorized user uses a protected system. This system reduces the amount of cooperation required from user as compared to the other existing systems. The algorithm was evaluated with four 40 minutes long videos with variations in facial expressions, occlusions and pose, and an equal error rate of 0.8% was achieved. The proposed algorithm by Goswai et al. [12] computes a descriptor based on the entropy of RGB-D faces along with the saliency feature obtained from a 2D face. Random decision forest classifier is used over the input descriptor for identification. Experiments were performed with RGB-D face database pertaining to 106 individuals. The experimental results indicate that the RGB-D information obtained by Kinect can be used to achieve improved face recognition performance compared to existing 2D and 3D approaches. The work by Nikisinis et al. [13] introduced the facial analyzes using synchronized RGB-D-T, where T is for thermal modality image. The recognition was performed using facial images by introducing a database of 51 persons including facial images of different rotations, illuminations, and expressions.

Nanni et al. [14] proposed fusion of Depth and RGB data for a dependable facial recognition system. In Kinect sensor, RGB and depth data are well matched with the help of device drivers provided and doesn't need alignment across them. The camera used in [15], 3DV System's ZCam also gives RGB and depth images aligned with each other and thus doesn't need extra alignment module. The normalization of range data is achieved by detecting nose-tip and then face region in input image [16]. Numerous global and local features were extracted from face and the data from depth and RGB was fused. The work presented in [17] dealt with face synthesis by image morphing from cheap but noisy depth sensors such as Kinect. This synthesis can be used to make 3D dataset for the study of face recognition methods.

Fusion of techniques as HOAG and 3DLBP regarding facial-recognition from (Kinect depth data) is elaborated in [18]. The use of Scale-Invariant Feature Transform (SIFT) technique for facial-recognition in RGB-D images is laid out [19].

The proposition in this paper is based upon the fusion of feature vectors extracted from depth maps and RGB images which evidently improves the performance of systems used for facial-recognition. Our idea is tested on an RGB-D (data set) for to validate our proposed methodology.

2. EURECOM FACE DATASET

This dataset [19, 20] has a sample space consisting of 52 unique subjects (14-females and 38-males). In an interval of 15 days, two separate sessions are captured with nine variations of individual subjects. The variations include: neutral look, smiling face, open mouth, illumination difference, eyes/mouth and half face left/right profiles occlusions. Prospective images of subject detailed in the EURECOM Kinect Facial Dataset are laid out in Fig. 1. Such variations present certain hurdles in image-processing making the dataset a bit more interesting plus challenging to handle. Three different types of data remain pertinent to all subjects: (1) Best Management Practice (.bmp) and Text (.txt) file formats depths that encompass all detected Kinect intensities): (2) RGB images & 3D Object (.obj) formats: and finally: (3) Each sample is provided with the position of eyes, nose, chin and left/right corners of lips.

A sphere radius of 65 pixels attribute constitutes the depth maps and RGB images created by Kinect and are centrally cropped at nose tip for face-localization. Such pre-processing is only applied to depth images created by Kinect.

3. METHODOLOGY

Here, various portrayals of information possessed by human faces have been used for countering the issues that relate in general to facial-recognition and image-processing. Commonly known attributes relating to a human face are mostly

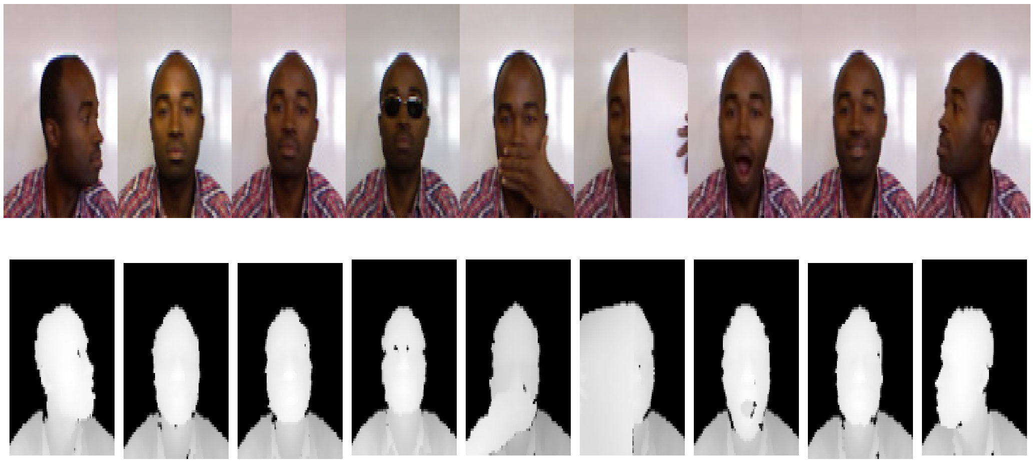

Fig. 1 EURECOM Kinect face dataset (subject images).

based on gaps between facial-landmarks, LIDs (local-image-descriptors) or facial-features. Previously, fusion of sets of facial attributes have shown enhancement in the recognition rates when compared to a singular attribute [21].

The fusion of RGB and depth maps is our proposed idea here for testing the performance of such attributes. The idea presented here focuses on various TFDs (Textural- Facial-Descriptors) possessing functionality of memorizing similarities among images and is well suited for resolving the problem at hand. Descriptors applied here procure complementary information from images. For such an approach, a multi-perspective view was deemed necessary where facial descriptors were initially studied separately and then were amalgamated for a more sturdy result.

The organization of this paper is: Section (A) elaborates the features used to represent images and how they have been utilized; Section (B) on the other hand highlights how sources such as (RGB/Depth) and attributes are combined in our proposed classification approach. The above is followed by our results and conclusions.

3.1. Feature Extraction

The following facial descriptors were used in the proposed method:

3.1.1. HOG Features

HOG stands for "Histogram of Oriented Gradients" [22]. HOG gets its evolution from SIFT descriptors [23]. This descriptor is extracted by image division into cells. Each pixel of the cell is responsible for creating Orientation based histogram channel while the individual pixel's vote is dependent on the (L2-norm) gradient. The utilization of rectangular cells is here for calculating channels of the histogram.

Furthermore, due to cells overlapping each other, the vector obtained from it includes numerous values from an individual cell. Moreover, regarding individual cells, flux in illumination is stabilized. An 81D feature-vector is obtained by fusing histograms with bins of an average value of 9 [24].

3.1.2. Local Binary Pattern Code (3 Patch)

Is based on the idea in which a specific code is assigned to each pixel. In this technique, associating values of three different patches from the image are used for extracting a single bit of code [25]. Patch w is assigned to the centralized pixel while other pixel-patches are scattered around it in a circle of radius r. Patches which are assigned as (α) are taken on the exact boundary of the circle and the values of such α-patches are compared to the value of the w-patch. Furthermore, a single bit code is considered the value of which is nearest to the central-patch w. The methodology above results in a code for each pixel with S-bits. Algorithm of "Three Patch Local Binary Pattern Code" is based on the given equation:

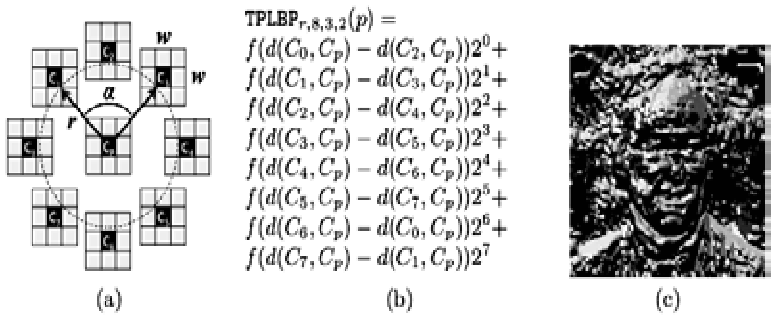

Fig. 2 TPLB-Code (α=2/S=8) (b) Computing TPLB (α=2/S=8/w=3) (c) Image generated by TPLB-Code.

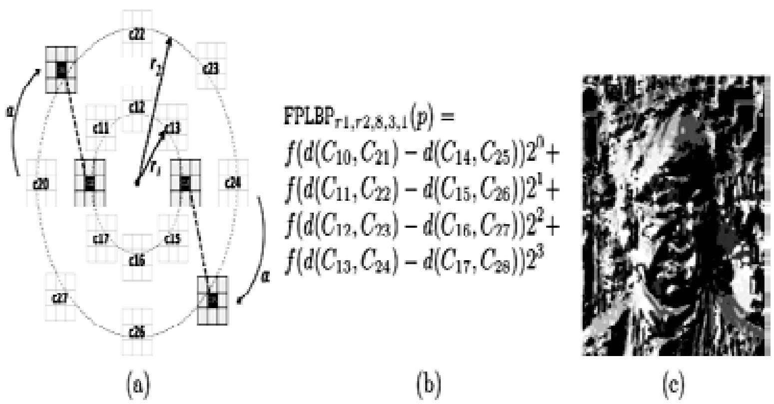

Fig. 3 Coded-FPLBP" (α=1) (b) Computing Coded-FPLBP (S=8/w=3/α=1) (c) Image generated by Coded-FPLBP.

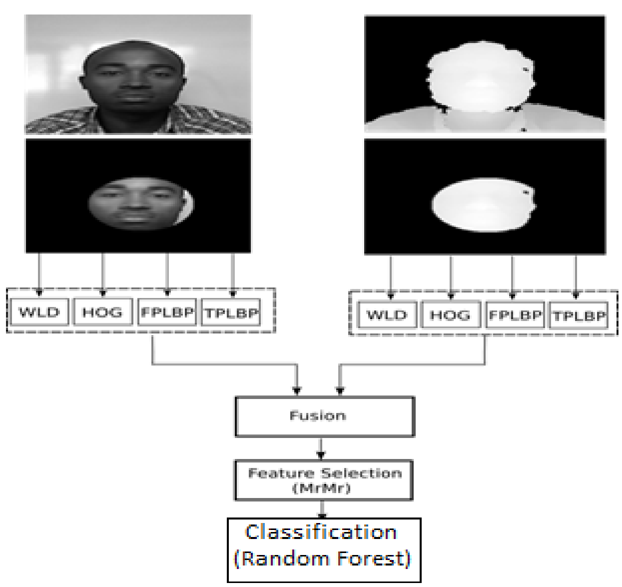

Fig. 4 Algorithm for random forest based classification.

$$\mathrm{TPLBP}_{r,S,w,\alpha}(p) = \sum_{i}^{S} f(d(C_i, C_p) - d(C_{i+\alpha \mod S}, C_p))2^i$$

[*(Ci)* and *(Ci+α mod S)*] present in the above equation comprise two patches. Respectively, both patches are along the above mentioned circle and *(Cp)* is the centralized-patch. Moreover, calculation regarding the intermediate patch-separation is done by [*d(Ci,Cp)]*.

Net value of function "*f*" is 1 when "*x*" is greater than or equal to τ (noise-level) and it is 0 when "*x*" is less than τ. Similar to [26], the value of τ is kept at 0.01 level. The considered image is uniformly spread into non-overlapping segments. A computing histogram on other hand extracts regularity of every binary code for all regions. All of these histograms are standardized to unit length. After standardization the value is reduced to 0.2 before being normalized to unit length again. Once the above is done, all of these histograms are joined together in order to form a single vector [27]. This explanation is presented in Fig. 2.

3.1.3. Four-Patch LBP (FPLB) Codes

In this depiction, every available pixel is surrounded by two concentric circles of radii [r1, r2]. Furthermore, an arbitrary number of patches (S) of W-size are uniformly distributed along each circle. To acquire the four-patch-code, two of the "central-symmetric-patches" present on the inner circle are compared to the two present on the outer circle. The comparison bears an output as the circles are α-patches away from one another. Finally, a single-bit-code for each pixel is set depending upon the highest level of similarity among the two patches. Each of the circles we have been assigned S-number of patches & S/2 central-symmetric-pairs constituting the length of the code. The FPLB algorithm is mathematically represented in Fig. 3(b).

3.1.4. Weber's Law Based Localized-descriptor

This descriptor is constituted using (Weber's law) that clearly states that the deviation between two matched objects is going to be identified as merely a difference only as long as the variance crosses a defined limit. According to [23], Weber-Local-Descriptor (WLD) accomplishes positive results as a texture- classifier because it is a compact local image descriptor. Also, [24] recently used it for facial-recognition with positive results. The algorithm focuses on parameters such as: (1) the intensity of the pixel and (2) summation of intensities of all surrounding pixels with their pixel-gradients. All pixels are defined by the ratio of the two above mentioned parameters helping in creating a histogram of 2880 elements.

3.2. Classification Method

The structure of our classification method is shown in Fig. 4. Random Decision Forrest was used for classification. Random Forest proposed by Leo Breiman is a collection of un-clipped classification or regression trees which are taken from the casual choice of samples of the training set. Random features are obtained during the training process. Estimation is done by combining (Majority-vote-for-classification/Averaging-regression) the predictions of the group. A tree is grown by sampling which is based on number of instances which are present in the training set and replacing them by the original data. This sample is then used to grow the tree. If the numbers of input variable are M, then a second variable m is selected where m is less than node. For splitting the node m is used which is kept constant during the process, and taken from M [28]. The size of the tree is kept as large as possible without trimming. Radom forest performs much better than classifiers based on a single tree. Feature selection of random forest was not used as that involves training the classifier for improved results.

The utilized samples for classification are identification-vectors of individual subjects that are extracted through the procedure which is explained as following. From RGB and depth information, a feature vector of each subject was extracted. Also, individual vectors of each of the sources were fused with one another in order to achieve the combined effect of two information sources.

Please refer to Section 4 for grouped characteristics; the identification-vectors were

Table 2. EURECOM accuracies list (FS – Feature Selection).

Descriptors	FS, k(50)			Without FS			FS, k(80)		
	"D"	"R"	"D+R"	"D"	"R"	"D+R"	"D"	"R"	"D+R"
1. "FPLBP"	"66.76"	"52.19"	"74.17"	"70.6"	"56.04	"76.92"	"67.03"	"51.65"	"75"
2. "TPLBP"	"89.01"	"88.46"	"93.68"	"None"	"None"	"None"	"90.38"	"88.19"	"92.86"
3. "HOG"	"60.99"	"67.033"	"78.02"	"77.2"	"66.34"	"85.16"	"64.84"	"68.96"	"77.47"
4. "WLD"	"60.91"	"62.08"	"71.43"	"40.39"	"29.4"	"35.99"	"64.84"	"66.21"	"73.9"
5. "HOG+FPLBP"	"79.39"	"66.48"	"82.14"	"73.07"	"56.31"	"75.27"	"77.47"	"64.01"	"83.79"
6. "HOG+WLD"	"76.92"	"73.35"	"82.41"	"56.87"	"50.82"	"60.99"	"75.55"	"77.2"	"83.24"
7. "HOG+TPLBP"	"91.48"	"88.46"	"92.03"	"None"	"None"	"None"	"90.66"	"88.19"	"91.76"
8. "WLD+FPLBP"	"67.03"	"64.56"	"76.37"	"69.51"	"56.04"	"76.1"	"68.41"	"63.19"	"76.1"
9. "WLD+TPLBP"	"89.01"	"88.46"	"93.68"	"None"	"None"	"None"	"90.38"	"88.19"	"92.86"
10. "HOG+TPLBP+WLD"	"91.48"	"88.46"	"92.03"	"None"	"None"	"None"	"90.66"	"88.19"	"91.75"
11. "HOG+FPLBP+WLD"	"80.77"	"69.78"	"83.51"	"71.98"	"57.97"	"76.37"	"78.3"	"70.6"	"83.79"

Table 3. S2 Accuracies (EURECOM).

Descriptor	FS(k=80)		
	D	R	D+R
FPLBP	66.48	47.52	69.78
TPLBP	90.66	81.59	92.86
HOG	75.55	68.96	80.77
WLD	65.66	61.81	72.53
HOG+FPLBP	71.2	59.34	81.04
HOG+WLD	79.95	72.27	85.71
HOG+TPLBP	91.76	81.59	93.13
WLD+FPLBP	70.88	57.14	75.55
WLD+TPLBP	90.66	82.14	91.21
HOG+TPLBP+WLD	91.76	82.14	93.13
HOG+FPLBP+WLD	79.4	66.48	84.07

Table 4. S1+S2 Accuracies (EURECOM).

Descriptor	FS(k=80)		
	D	R	D+R
FPLBP	59.61	46.29	65.52
TPLBP	89.29	81.18	90.52
HOG	62.09	65.25	74.73
WLD	55.63	56.04	66.35
HOG+FPLBP	72.8	62.23	78.3
HOG+WLD	72.12	70.32	78.29
HOG+TPLBP	89.29	81.18	91.34
WLD+FPLBP	64.01	58.24	70.6
WLD+TPLBP	89.29	82.69	90.52
HOG+TPLBP+WLD	89.29	82.69	91.34
HOG+FPLBP+WLD	74.18	67.17	78.85

obtained through joining the respective identification-vectors of singular-descriptors.

Furthermore, for minimizing size of the extracted facial representations and for improving the quality, FS (Feature Selection) algorithm was applied to the facial representation. Discarding redundant information for size and accuracy improvement was the idea. The proposed thinking demanded elimination of redundant information for the sake of size reduction and to retain useful information for accuracy alleviation.

Feature-Selection process in our methodology bases on Minimum-Redundancy-Maximum-Relevance (MRMR) algorithm [29]. The feature set is organized in MRMR in decreasing order of importance and then in our case the feature set's level K=80 has been selected where the highest correlation features are kept while the rest are discarded.

4. RESULTS & DISCUSSION

The classification approach proposed in this paper is tested on the RGB-D dataset. Matlab on the other hand is used for extracting the features and combining RGB-D with feature selection. The classification is completed using the RDF (Random-Decision-Forest) with 10-fold cross-validation.

The case-study was done on EURECOM dataset (both sessions) first individually and then in combination to test the sturdiness of the method. The individual and fused accuracies were duly observed and noted. Our case study had two objectives: (a) Combining depth with RGB information and (b) Feature level fusion of separate attributes. Firstly, classification based on single source attributes was evaluated then the evaluation was done for fused source. After that, the same feature extracted individual and fused sources were tested for a combination of separate attributes. Results of our case-study are respectively laid out in Table 2 to Table 4.

The results indicated the following points:

- The respective amalgamation of RGB plus Depth-features improves ability of individual and grouped attributes.
- In comparison to RGB imaging, in general, maps based on Depth possess more discriminative info. However, it further increases performance of facial-recognition system by fusion.
- Highest accuracies are achieved by TPLBP-HOG-WLD group on individual and combined dataset sessions containing (RGB images/Depth maps) plus their amalgamation. This shows the advantage of a multi-perspective approach as it captures more discriminate information compared to individual descriptor.
- When the data set is small, TPLBP performs better but more robustness is shown by grouped attributes when the size of the dataset is increased which indicates the advantage of using a multi perspective approach.
- Both individually and in combination with other features, TPLBP is more accurate as compared to FPLBP because TPLBP has a higher dimensional feature vector and contains more information as compared to FPLBP.
- Individual and grouped attribute accuracies are both improved by the selection process due to discarding redundant information.

Finally, the results exhibit an improvement of performance in FR through fusion of RGB and depth information. In addition, a multi-prong (Feature fusion) idea in combination with appropriate feature selection is exhibiting to be sturdier and highly accurate.

Furthermore, it was observed from experimentation, that the performance of attributes grouped together is more enhanced than the attributes analyzed individually. Moreover, when size of the sample space is expanded, it is observed that stability consequently increases.

5. CONCLUSIONS

A solution is presented in this paper for improving the performance of facial-recognition systems. As we know, a human face appearance is affected by intra-class/inter-class variations. For it, a solution is made possible by the advent of Microsoft Kinect, commercial off the shelf depth sensors and RGB cameras which are both cost effective. As a conclusion, it is important to note here that depth based information can be used additionally to enhance facial-recognition. The classification approach presented in this paper suggests that amalgamation of (Depth and RGB) image information improves system's performance.

6. FUTURE WORK

The preprocessing involved for facial recognition has already been done in the dataset utilized in this work. Furthermore, the face localization has also been made easy by providing vital information such as position of the eyes the nose and lips. This enables the researchers to focus their attentions on the key tasks of features extraction and classification. However, for the real time facial recognition (RTFR) systems there are varying image acquisition conditions which increase the challenges. In addition, RTFR systems time taken from image acquisition to recognition should be kept at minimal. Such considerations were not taken into account during this work. In future, the algorithm can be tested on images acquired in the wild to comment on the feasibility of this work.

7. REFRENCES

1. Jain, A., R. Bolle & S. Pankanti (Ed.). *Biometrics: Personal Identification in Networked Society*. Springer Science & Business Media, Vol. 479 (2006).
2. Maltoni, D., D. Maio, A. Jain, & S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer Science & Business Media (2009).
3. Jain, A.K. & S.Z. Li. *Handbook of Face Recognition*. Springer, New York (2011).
4. Li, B.Y. Mian, A. S. Liu, W. & A. Krishna. Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, p. 186-192 (2013).
5. Bo, L., X. Ren, & D. Fox. Depth kernel descriptors for object recognition. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, p. 821-826 (2011).
6. Park, Y., V. Lepetit, & W. Woo. Texture-less object tracking with online training using an RGB-D camera. In: *10th IEEE International Symposium on Mixed and Augmented Reality*. ISMAR, p. 121-126 (2011).
7. Ramey, A., V.González-Pacheco, & M. A. Salichs. Integration of a low-cost RGB-D sensor in a social robot for gesture recognition. In: *Proceedings of the 6th International Conference on Human-Robot Interaction*. p. 229-230. ACM (2011).
8. Hog, R. I., P. Jasek, C. Rofidal, K. Nasrollahi, T. B. Moeslund, & G. Tranchet. An rgb-d database using microsoft's kinect for windows for face detection. In: *8th IEEE International Conference on Signal Image Technology and Internet Based Systems*. Aalborg University, Denmark, p. 42-46 (2012), **DOI:** 10.1109/SITIS.2012.17.
9. Huynh, T., R. Min, & J. L. Dugelay. An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In: *Asian Conference on Computer Vision*. Springer Berlin Heidelberg. p. 133-145 (2012).
10. Cao, Z., Q. Yin, X. Tang, & J. Sun. Face recognition with learning-based descriptor. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. p. 2707-2714 (2010).
11. Pamplona Segundo, M., S. Sarkar, D. Goldgof, L. Silva, & O. Bellon. Continuous 3D face authentication using RGB-D cameras. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. p. 64-69 (2013).
12. Goswami, G., S. Bharadwaj, M. Vatsa, & R. Singh. On RGB-D face recognition using Kinect. In: *IEEE 6th International Conference on Biometrics: Theory, Applications and Systems*. p. 1-6 (2013).
13. Nikisins, O., K. Nasrollahi, M. Greitans, & T.B. Moeslund. RGB-DT based face recognition. In: *22nd IEEE International Conference on Pattern Recognition*, p. 1716-1721 (2014).
14. Nanni, L., A. Lumini, F. Dominio, & P. Zanuttigh. *Effective and precise face detection based on color and depth data*. *Applied Computing and Informatics* 10(1): 1-13 (2014).
15. Bleiweiss, A. & M. Werman. Fusing time-of-flight depth and color for real-time segmentation and tracking. In: *Dynamic 3D Imaging*. Springer, Berlin, Heidelberg. p. 58-69 (2009).
16. Zhou, J., Y. Li, & J. Wang. *2D & 3d-comfusface: 2d and 3d face recognition by scalable fusion of common features*. *International Journal of Computer Science and Network Security* 12(3): 30 (2012).
17. Ramalingam, S. & N. T. Viet. 3D face synthesis with Kinect. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, p. 4433-4438 (2013).
18. Cardia Neto, J. B. & 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*. p. 66-73 (2015).

19. Ouloul, M. I., Z. Moutakki, K. Afdel, & A. Amghar. An efficient face recognition using SIFT descriptor in RGB-D images. *International Journal of Electrical and Computer Engineering* 5(6): 1227-1233 (2015).
20. Min, R., N. Kose, & J. L. Dugelay. A Kinect database for face recognition. *IEEE Transactions on Systems, Man and Cybernetics* 44(11): 1534-1548 (2014).
21. Bottinok, A., I. U. Islam, & T. F Vieira. A multi-perspective holistic approach to Kinship verification in the wild. In: *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Vol. 2*, p. 1-6 (2015).
22. Dalal, N., & B. Triggs. Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 1*, p. 886-893 (2005).
23. Lowe, D.G. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision* 60(2): 91-110 (2004).
24. Junior, O. L., D. Delgado, V. Gonçalves, & U. Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In: *12th International IEEE Conference on Intelligent Transportation Systems*, p. 1-6 (2009).
25. Su, Y. S. Shan, X. Chen, & W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. *IEEE Transactions on Image Processing* 18(8): 1885-1896 (2009).
26. Heikkilä, M., M. Pietikäinen, & C. Schmid. Description of interest regions with center-symmetric local binary patterns. In: *Computer Vision, Graphics and Image Processing*. Springer Berlin Heidelberg. p. 58-69 (2006).
27. Wolf, L., T. Hassner, & Y. Taigman. Descriptor based methods in the wild. In: *Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition.* Marseille, France, p. 1-14 (2008).
28. Ali, J., R. Khan, N. Ahmad, & I. Maqsood. Random forests and decision trees. *IJCSI International Journal of Computer Science Issues* 9(5): 1-7 (2012).
29. Peng, H., F. Long, & C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8): 1226-1238 (2005).