

International Conference on Information and Communication Technologies (ICICT 2014)

RGB-D Face Recognition System Verification Using Kinect And FRAV3D Databases

Poornima Krishnan^{a,*}, Naveen.S.^b

^aPG Scholar, Department of ECE, LBSITW, Poojappura, Trivandrum-695012, India

^bAssistant Professor, Department of ECE, LBSITW, Poojappura, Trivandrum-695012, India

Abstract

This paper deals with a facial recognition system and its verification using the RGB-D data obtained from the Kinect and FRAV3D database. The FRAV3D database contains 106 subjects, which involves approximately one woman after every three men. The Kinect database has 17 images per 31 persons. The proposed algorithm computes a descriptor based on the entropy of RGB-D faces along with the saliency feature obtained from a 2D face which is used as input to a tree bagger classifier to establish the identity.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Facial Recognition system; Kinect; entropy; saliency

1. Introduction

The main objective in this paper is to recognize human faces from RGB-D images using FRAV3D and Kinect RGB-D database. Face recognition is a challenging problem which suffers not only from the general object recognition challenges such as illumination and viewpoint variations but also from distortions or covariates specific

* Corresponding author. Tel : +0-949-615-3683
E-mail address: poorniskrishnan@gmail.com

to faces such as expression, accessories, and high inter-class similarity of human faces. In the FRAV3D database the data were obtained using Minolta VIVID 700 scanner, that provides a VRML file (3D image) along with the texture information (2D image). VIVID 700 is a high speed, portable and easy to use 3D laser scanning system. Here in total of 16 captures per person were taken in every session, with different poses and lighting conditions, trying to cover all possible variations, including turns in different directions, gestures and lighting changes. The depth map is constructed here using laser scan principle. Each session, involves different poses and lighting conditions, thus covering almost all possible variations, including gestures, turns in different directions and lighting changes. The depth map is constructed here using laser scan principle.



Fig. 1. Minolta VIVID scanner

The Kinect RGB-D database² had been obtained using Microsoft Kinect sensor. Kinect sensor consists of a horizontal bar which is attached to a base using a pivot that is motorized and could be positioned with respect to the display. It includes a depth sensor along with RGB camera and a multi-array microphone, which has 3D motion capture, voice and face recognition capabilities. The depth sensor includes IR laser projector combined with a CMOS sensor, that captures video data in 3D under any lighting conditions. The Kinect uses infrared laser light, with a speckle pattern. The depth map is formed by analyzing the speckle pattern of IR light based on the structured light principle. The Kinect RGB-D database contains 1581 images (and their depth counterparts) taken from 31 persons in 17 different poses and facial expressions using a Kinect device. The faces in the images are not extracted neither in the RGB images nor in the depth, hence it could be used for both recognition and detection.



Fig. 2. Kinect sensor

Each pixel in Kinect's depth map has a value indicating the relative distance of that pixel from the sensor at the time the image depth maps were captured. It exhibits very high inter-class similarity (due to noise and holes), and therefore may not be able to differentiate among different individuals. However, it has low intra-class variation which can be utilized to increase the robustness to covariates such as expression and pose. Further, color images can provide inter-class differentiability which the depth data lacks. Therefore, it is essential to utilize both RGB and depth data for feature extraction and classification. The algorithm first computes entropy maps corresponding to RGB and depth information and a visual saliency map of the RGB image. The Histogram of Oriented Gradients (HOG) descriptor⁴ is then used to extract features from the entropy and saliency maps. Then the concatenation of these HOG descriptors provides the final feature descriptor which is used as input to the tree bagger classifier for establishing the identity.



Fig. 3. (a) FRAV3D database ; (b) Kinect database

2. Proposed Method

The proposed method first computes the entropy maps and saliency maps for both the texture and the depth maps and then HOG descriptors are derived for each of these maps to extract features. These maps are then combined to form the final descriptor, which is then fed to the tree bagger classifier. The detailed block diagram of the method is shown in Fig. 4. Entropy is defined as the measure of uncertainty in a random variable. Entropy of an image of an image r_k is defined as

$$H = -\sum_{k=0}^{L-1} P(r_k) \log(P(r_k)) \quad (1)$$

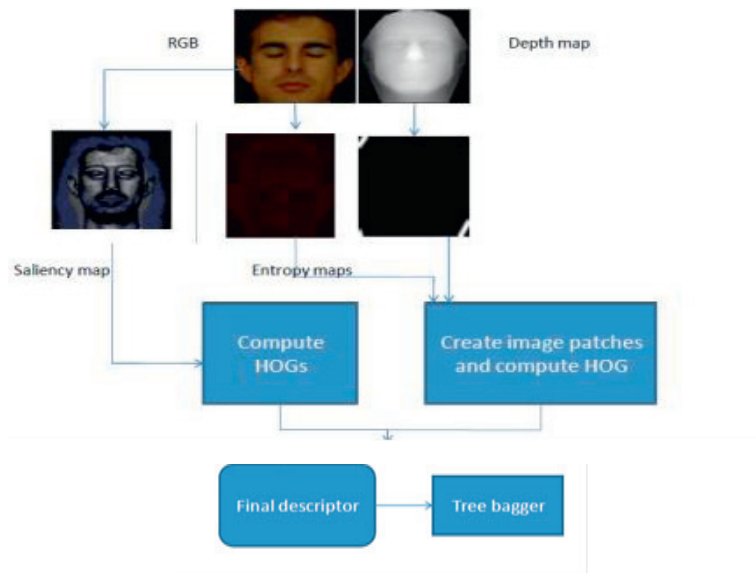


Fig. 4. Steps in proposed method

Such entropy maps are derived for both texture and depth images. But the saliency maps are derived only for texture images since saliency refers to something which catches visual attention or that stands out. The saliency attended locations for an image is shown in Fig. 5. There are bottom up approaches and top down approaches for saliency detection. Here we use top down approach for saliency detection. The key idea behind the saliency map is to extract local spatial discontinuities in the modalities of intensity, orientation and color.

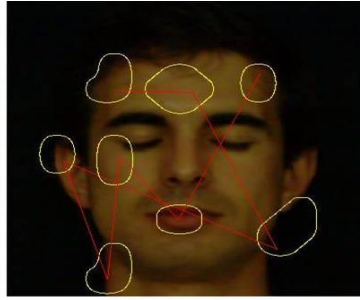


Fig. 5. Saliency attended location

Next is to compute the descriptor for these maps. HOG refers to Histograms of Oriented Gradients⁴. Its a feature descriptor that extracts features that are scale and rotation invariant. The HOG facial detector⁴ uses a “global” feature to describe a face. This detector uses detection window that could be slid over the entire image. A HOG descriptor is computed for the detection window, at each position of the detection window. The window used by HOG person detector is 64 pixels wide by 128 pixels tall. In order to obtain the final descriptor of HOG, we consider cells within the window that are 8×8 pixels in size. Here gradient vectors for each of the pixel within the cells are computed. We take the 64 gradient vectors and put them into a 9-bin histogram. The next step is to normalize the histograms. In each bin of the histogram, the value is based on the gradient magnitude of the corresponding cell. These cells are grouped into blocks and then normalized based on the histogram value. This could be carried out by concatenation of histogram of the cells within a block to a vector of 36 components.

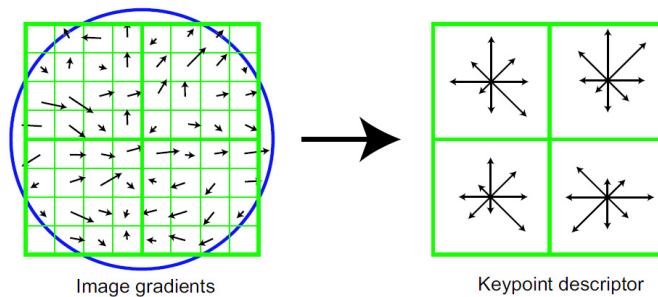


Fig. 6. HOG descriptor formation

Finally the descriptor output is fed to the classifier to recognize the identity. Here tree bagger classifier is used for this purpose to compute the descriptor for these maps. HOG stands for Histograms of Oriented Gradients. It is a type of “feature descriptor” that extracts features that find a decision criterion that best divides the multi-class training data into two groups at each node. A decision criterion consists of an attribute and a related threshold. Random trees², on the other hand, randomize the decision criteria. A random decision criterion is defined by a random attribute and a random threshold. Random trees, thus, divide the training data on completely random attributes. This randomization mainly addresses high dimensional data and generalization. The classification results from each tree are collected for an input image and typically, a simple majority voting scheme awards the resulting class label. A number of such random decision trees are combined to form a random forest.

3. Results

The proposed algorithm had been implemented using MATLAB. FRAV3D database¹² and RGB-D Kinect database had been used here. The HOG person detector uses a detection window that is 64 pixels wide by 128 pixels tall, giving a final descriptor size of 7560×1 . The experiments were conducted in the presence and absence of four components, that is texture, depth, entropy and saliency for both FRAV3D and Kinect database.

(1) Recognition rates for texture alone:

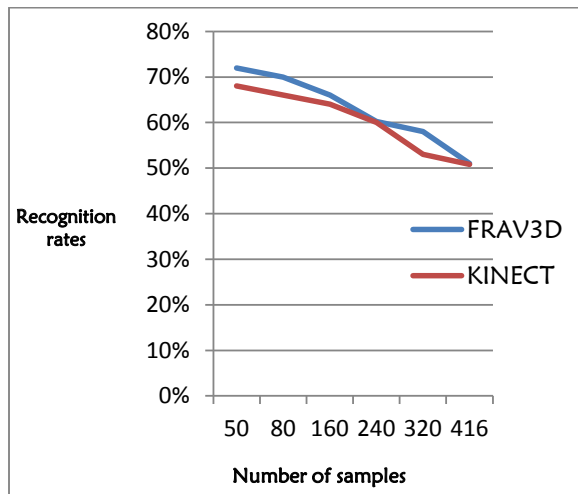


Fig. 7. Recognition rates obtained using entropy maps alone

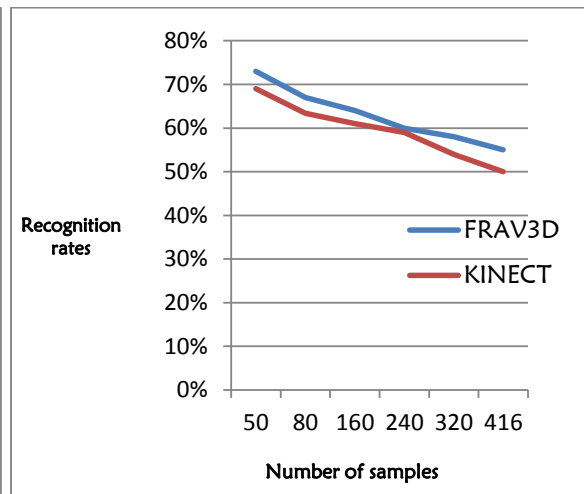


Fig. 8. Recognition rates obtained using saliency maps alone

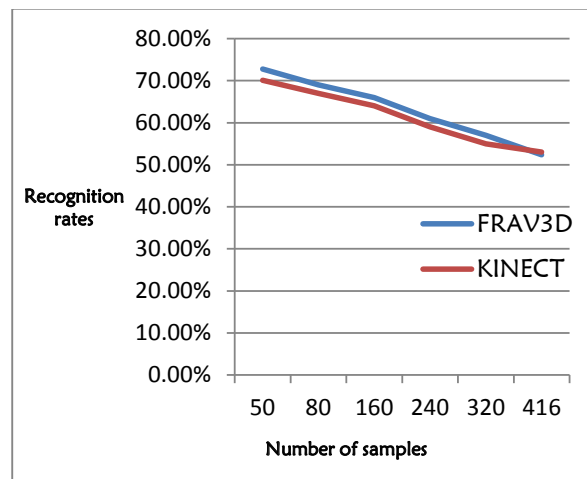


Fig. 9. Recognition rates obtained using both entropy and saliency maps

Fig. 7. to Fig. 12. shows the results obtained using the proposed algorithm that is by using HOG descriptor for feature extraction for FRAV3D and Kinect. It could be seen that increased recognition rates are obtained using FRAV3D compared to Kinect since the Kinect database were manually cropped. Moreover it could be seen that by taking the texture along with the depth maps increased the recognition to a considerable rate. This shows that the role of both the texture and depth images are significant in recognizing the face data. It could be inferred from the results of Fig. 7. to Fig. 9. that the combination of both entropy and saliency maps with the fusion of texture and saliency had obtained increased recognition rates compared to other cases

(2) Recognition rates for depth alone

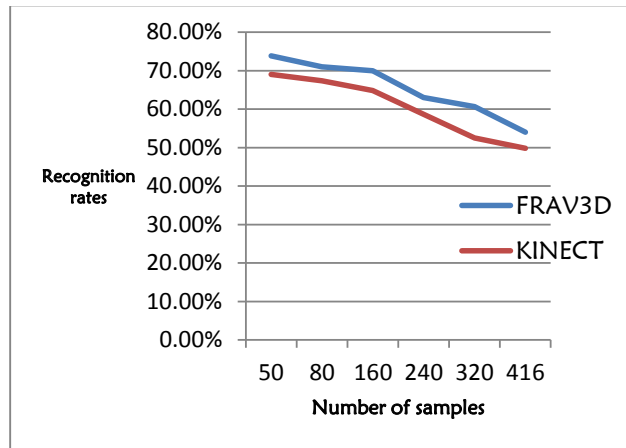


Fig. 10. Recognition rates obtained using entropy maps alone

(3) Recognition rates for fusion of texture and depth

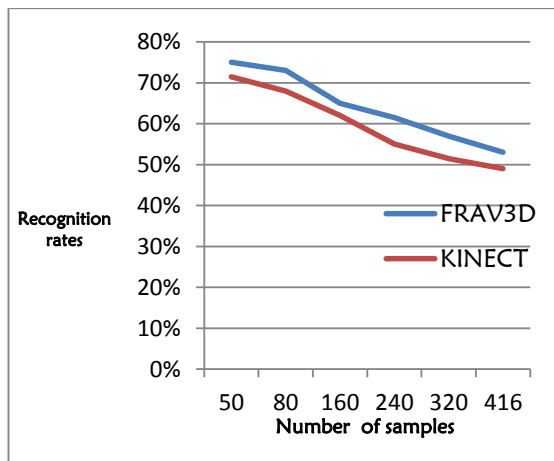


Fig. 11. Recognition rates obtained using entropy maps alone

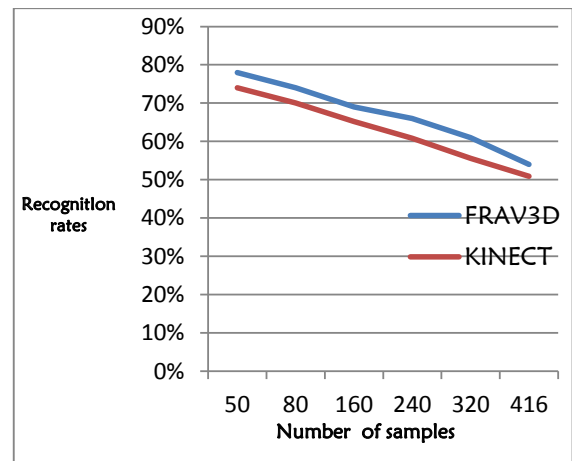


Fig. 12. Recognition rates obtained using both entropy and saliency maps

In whole the proposed approach is found to obtain better results than the existing algorithms under effect of covariates. Comparing the recognition rates of the proposed method to SIFT descriptor, it could be seen that the results of the proposed algorithm are higher by a considerable amount. Hence the method proves to have a better performance. Comparing the recognition rates of the proposed method to SIFT descriptor, it could be seen that the results of the proposed algorithm are higher by a considerable amount. Hence the method proves to have a better performance. From the Fig. 13, it could be clearly seen that the proposed method outperforms the SIFT descriptor by a considerable amount for both the FRAV3D and the Kinect databases. Table 1. and 2. shows the results of reliability testing for different combinations of texture, depth, entropy and saliency.

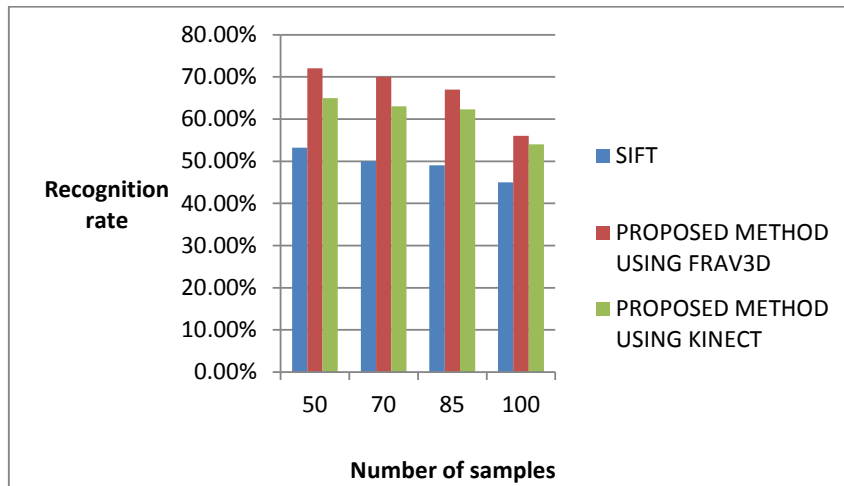


Fig. 13. Comparison of proposed method with SIFT descriptor

True Acceptance Rate (TAR) is the rate of the acceptance of the original class that is available within the training set, which is to be high. The True Rejection Rate (TRR) is the rate of rejection of an external class (that is not available in the training set), which requires to have a low value. False Acceptance Rate (FAR) is the rate of acceptance of the external class, which is to be low. False Rejection Rate (FRR) is the rate of rejection of the external class, which is to be high.

Table 1. Reliability testing using entropy alone for texture depth combinations

SAMPLES	TAR			TRR			FAR			FRR		
	USING TEXTURE AND ENTROPY	USING DEPTH AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY ALONE	USING TEXTURE AND ENTROPY	USING DEPTH AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY ALONE	USING TEXTURE AND ENTROPY	USING DEPTH AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY ALONE	USING TEXTURE AND ENTROPY	USING DEPTH AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY ALONE
100	82%	87%	86%	18%	13%	14%	2%	16%	17%	98%	84%	83%
200	100%	94%	99%	0%	6%	1%	5%	4%	14.5%	95%	96%	85.5%

From the Table 1, it could be seen that for 100 samples the proposed method using entropy alone for the fusion of texture and depth gives a TAR of 86%, TRR of 14%, FAR of 17% and FRR of 83%, that is a high rate of TAR and FRR and low rate of TRR and FAR which is desirable. As the samples are increased to 200, it is shown to give an increased TAR and FRR of 99% and 85.5% and low values of TRR and FAR of 1% and 14.5%. Similarly, from the Table 2, it could be seen that for 50 samples the proposed method using saliency along with entropy for the fusion of texture and depth gives a TAR of 80%, TRR of 20%, FAR of 0% and FRR of 100%, that is a high rate of TAR and FRR and low rate of TRR and FAR which is desirable. As the samples are increased to 200, it is shown to give an increased TAR and FRR of 96% and 92% and low values of TRR and FAR of 4% and 8%. From the above results the proposed algorithm, is shown to give a high True Acceptance Rate (TAR) and False Rejection Rate (FRR), at the same time it results in a low value of True Rejection Rate (TRR) and False Acceptance Rate (FAR). It proves the better performance of the proposed method as the percentage of true classes accepted and false rejected are high.

Table 2. Reliability testing using saliency along with entropy for texture , depth combination

SAMPLES	TAR	USING TEXTURE AND SALIENCY	USING TEXTURE FOR SALIENCY AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY AND SALIENCY	TRR	USING TEXTURE AND SALIENCY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY AND SALIENCY	USING TEXTURE AND SALIENCY	FAR	USING TEXTURE AND SALIENCY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY AND SALIENCY	USING TEXTURE AND SALIENCY	FRR	USING TEXTURE FOR SALIENCY AND ENTROPY	FUSION OF TEXTURE AND DEPTH FOR ENTROPY AND SALIENCY
50	94%	90%	80%	6%	10%	20%	2%	12%	0%	98%	76%	100%			
100	100%	93%	96%	0%	7%	4%	5%	6%	8%	95%	94%	92%			

4. Conclusion

This paper presents a novel approach for face recognition, based on HOG features and the tree bagger for classification. The proposed method is shown to give better results specially for smaller training set sizes. Moreover the proposed method is shown to outperform the SIFT descriptor .The proposed algorithm uses a combination of entropy, saliency and depth information with HOG for feature extraction and tree bagger for classification. Moreover random tree is one of the best classification techniques available. The only drawback of using Random tree for classification is the need for large training set per subject. The experiment results were obtained using five poses per class which shows a better performance of our approach as the training set is improved. It is also seen that the method is highly reliable from the reliability results obtained with a high TAR and low FAR.

References

1. A. F. Abate, M. Nappi , D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. PRL, 28(14):1885–1906, 2007.
2. B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna. Using kinect for face recognition under varying poses, expressions, illumination and Disguise, 2013
3. Y. Amit and D. Geman. Shape Quantization and Recognition with Randomized Trees. Neural Computation, 9(7):1545–1588, 1997.
4. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, volume 1, pages 886–893, 2005.
5. Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In CVPR, pages 2707–2714, 2010
6. J. Beis and D.G. Lowe. Shape Indexing using Approximate Nearest- Neighbour Search in High-Dimensional Spaces. In Conference on Computer Vision and Pattern Recognition, pages 1000–1006, Puerto Rico, 1997.
7. L. Breiman , J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and Regression Trees. Chapman Hall, New York, 1988
8. Leo Breiman. Bagging predictors. Machine Learning, 24(2):123–140, 1996.
9. Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In Conference on Computer Vision and Pattern Recognition, pages 111–119, 2000.
10. Shuai Tang, Xiaoyu Wang¹, Xutao Lv, Tony X. Han¹, James Keller, Zhihai He, Marjorie Skubic¹, and Shihong Lao -Histogram of Oriented Normal Vectors for Object Recognition with a Depth Sensor.
11. William Robson Schwartz, Huimin Guo, Jonghyun Choi and Larry S. Davis, Fellow. Identification Using Large Feature Sets .IEEE Transactions On Image Processing ,Vol. 21, No. 4, April 2012.
12. <http://www.frav.es/databases/FRAV3D>.
13. Lowe D.G. *Distinctive Image Features From Scale-Invariant Keypoints* , International Journal of Computer Vision. Vol. 60, pp. 91–110, 2004.