Volume 4, Issue 7, July 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# 2D and 3D Face Recognition using Close-Range RGB-D Camera

**Srinivas Kishan Anapu*, Dr. Peri Srinivasa Rao**
Computer Science Engineering Department,
Andhra University, Visakhapatnam, India

***Abstract— Having a depth information of object (Face) surface available along with RGB image, face recognition algorithm can be made more efficient and robust as compared to the conventional algorithms, where only RGB image is employed. Non-uniform illumination and pose deviation are major obstacles in making ideal face recognition system. In this paper, we used close-range (of depth) camera for developing face recognition system, which can provide high recognition rate even in low-resolution images. To best of our knowledge, there is no experimental study available for the face recognition system with close range RGB-D (Creative Labs 3D RealSense) in literature so far. We analysed different feature representations like PCA, LDA, SIFT, Gabor, LBP and HOG for RGB and depth images. Experiments were performed on the in-house RGB-D dataset, captured by Creative Labs 3D camera with 20 individuals. The experimental result shows that depth image can also be used for recognition. It is also found that the fusion of RGB and Depth images improve the recognition rate. HOG and LBP feature descriptors are superior to the other feature subspaces. With HOG as a feature descriptor, face recognition rate for frontal image dataset in proper illumination has reached to 96.21%, while for pose-deviated samples, it is 75.12 %.***

***Keywords— 2D Face; 3D Face, RGB-D camera, Face Recognition, Fusion***

## I. INTRODUCTION

Face Recognition systems are of very importance and being successfully included into mobile, desktop and dedicated hardware machines installed at indoor or outdoor locations. The applications ranging from gaming to surveillance have been evolved and their performances are increasing day by day. The most of the face recognition systems are based on the texture image captured by color camera. At times, when there is improper lighting conditions and/or pose variation, texture based face recognition fails. Non-uniform illumination and pose deviation are important covariates in the face images that degrade the recognition performance significantly. It is interesting to see how effect of covariates can be overcome by using new technologies of hardware and algorithms.

The 3D cameras, with depth information along with color image, are becoming popular and easily available at cheaper cost now days. Recently, 3D cameras with RGB-D information are launched from Microsoft (Kinect Sensor), ASUS and more recently from Creative Labs (RealSense). Though, the conventional camera uses stereo matching calculation for determining depth, the cameras, mentioned above, use technique of TOF (Time of Flight) using IR camera as shown in figure 1. The depth of image is created based on the physical principal that depth, distance between surface point and camera, is proportional to the time taken by pulse wave to travel from source-to-object-to-sensor. There are several algorithms successfully applied to 2D color based face recognition. The additional depth information to the 2D color images can improve the performance of face recognition system. Though, the overall structure of face of all human is same, there is intra-subject variations across the human subjects in terms of depth pattern at local regions of face. Due to depth information of object (Face) surface available in addition to the RGB image, face recognition algorithm can be made more efficient and robust as compared to the algorithm, where only RGB image is employed. The depth range of Creative Labs Camera has been specified as a close range of 6 inch to 3.25 ft. Its frame rate is up to 30 fps and there is synchronization between depth and RGB image capture. There are two main advantages with 3D (RGB-D) camera:

1. Depth image is insensitive to lighting conditions and hence, more weightage can be given to the depth image while fusing information for face recognition in case of low light.

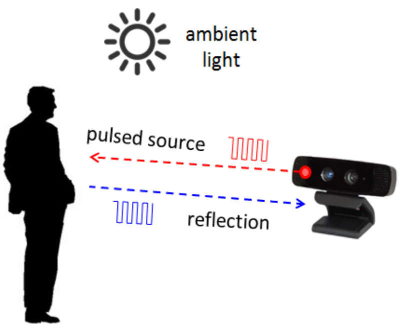


Figure 1: TOF (Time of Flight) technique for depth

 

2. Biometric spoofing created by image print can be identified and avoided.
3. At times, the 3D model required to correct pose-deviation of face images, construction of 3D model can efficiently use depth info.

Most of the study and results presented in existing literature is based on the image captured by Kinect sensor. Our study is different from this, in the sense we used close range RGB-D camera of Creative Labs 3D. To the best of our knowledge, there is no experimental study so far published in the literature with data samples acquired by Intel3D camera. In close range camera, small subject motion would be more visible in the images than that in mid-range images. Another difference with Creative Labs 3D camera is that it has lower resolution as compared to the Kinect images. However, advantage with Creative Labs 3D camera lies in its small size and hence more suitable for tablet and mobile devices. Applications like online banking, online registration, loyalty programs and call-centre security QA to continue the call; are few of lot many applications where Creative Labs 3D camera can be employed.

The performance analysis with different feature representations like PCA, LDA, SIFT, Gabor, HOG and LBP is done in this paper. These features were applied to both, RGB and depth images and explored the applicability of feature in different scenarios. Exploring the way to use the depth image along with RGB is the main objective of this paper. The reaming part of the paper is organized as follows. In section II, the literature review is presented. In next section III, face recognition system and different features are elaborated. Experimental results are presented in section IV and discussion over the results is covered in this section. Finally, paper is concluded with remarks in section V.

## II. RELATED WORK

The work done in [12] presents an algorithm applied to a low resolution 3D sensor for robust face recognition under challenging conditions. This system involves a preprocessing algorithm which employs the facial symmetry at the 3D point cloud level to obtain a canonical frontal view, shape and texture, of the faces irrespective of their initial pose. The smoothening is applied to noisy depth data captured from low resolution camera in order to fill up holes and remove the noise from depth info. The RGB and Depth images are approximated by Sparse Representation using pre-trained dictionary. Experiments performed over 5000 facial images obtained from a publicly available database of RGB-D images with varying poses, expressions, illumination and disguise, acquired using the Kinect sensor records the recognition rates are 96.7% for the RGB-D data and 88.7% for the noisy depth data alone.

Another interesting work [13] 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 in [14] 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 recent work done in [15] introduces 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.

## III. METHODOLOGY

In order to evaluate the feature representation techniques, various methods have been used in the face recognition system. The generalized system function diagram is shown in figure 2. The face region detected in RGB and depth images are represented using feature vector by particular feature representation. Euclidian distance is calculated between probe and gallery samples. Person is identified using nearest neighbour classifier.

### *A. Features:*

The different feature representation techniques that are considered in our work are described below.

PCA: Principal component analysis (PCA) is a popular unsupervised statistical method to find useful image representations [1].

This method for facial recognition was a global PCA scheme in which the facial region was cropped for all images and resized to size 128x128. For the training set, the cropped facial images were used to calculate the PCA. PCA was used to reduce the dimensionality of the feature vector by projecting onto eigenvectors. In order to reduce the dimensionality of face representation, principle components/eigenvectors, corresponding to the higher Eigen values that models the around 95-98% of feature variance, were chosen. The number of principle eigenvectors dependent on the samples used for calculating the PCA and approximately was in the range of 100-200. The each of the training was represented by PCA coefficients equal to the number of principle components chosen and projecting each of the images onto those eigenvectors. The test image is also projected on the same principle components and represented by PCA coefficients. Test PCA coefficients are compared by using appropriate classifier.

LDA: Unlike PCA, LDA (Linear Discriminant Analysis) is a supervised method [2, 3]. While calculating for LDA subspace, each sample's associated class is also considered. LDA maximizes the between-class variance, while minimizing the intra-class variance. Once LDA subspace is calculated, training and testing images are represented as explained in the PCA.

Gabor: Gabor wavelet based feature extraction is proposed for face recognition in [4] and is robust to small-angle rotation. Here, we used 7 landmarks and each landmark was processed by Gabor filter bank composed of 7 angular direction and 5 frequency scales. Thus, each landmark represented by 35 Gabor coefficients. The representations for all landmarks are concatenated to form the feature vector for give face image.

HOG: Among many, HOG is one of the local descriptors that have given promising performance in variety of problems of computer vision [5, 6]. The image is decomposed into local regions and from each local region gradient orientation and its magnitude are calculated. In each bin of gradient orientation of histogram, corresponding magnitudes are accumulated for the local region. It is believed that HOG is robust to illumination variation for recognition problems [7].

LBP: After using linear binary pattern (LBP) first time for measuring the local image contrast [8], it has been applied in several pattern classification problems [9, 10]. To calculate LBP, each pixel is assigned with a label by a type of binary pattern obtained in 3x3-neighborhood pixels by thresholding neighborhood pixel intensity with center pixel. The distribution of these binary patterns in local region is used as a feature representation, describing the nature of texture exist in that region.

SIFT: Lowe [11] has introduced the shift invariant feature transform to describe the image globally. Its invariance nature to rotation, scaling and translation has been successfully used in several applications to get improved performance over other features. SIFT features calculated as a difference of Gaussian (DoG) filtered images for two different scales for few numbers of octaves (down sampled versions of images). Key points are extracted where extrema of DoG is found.

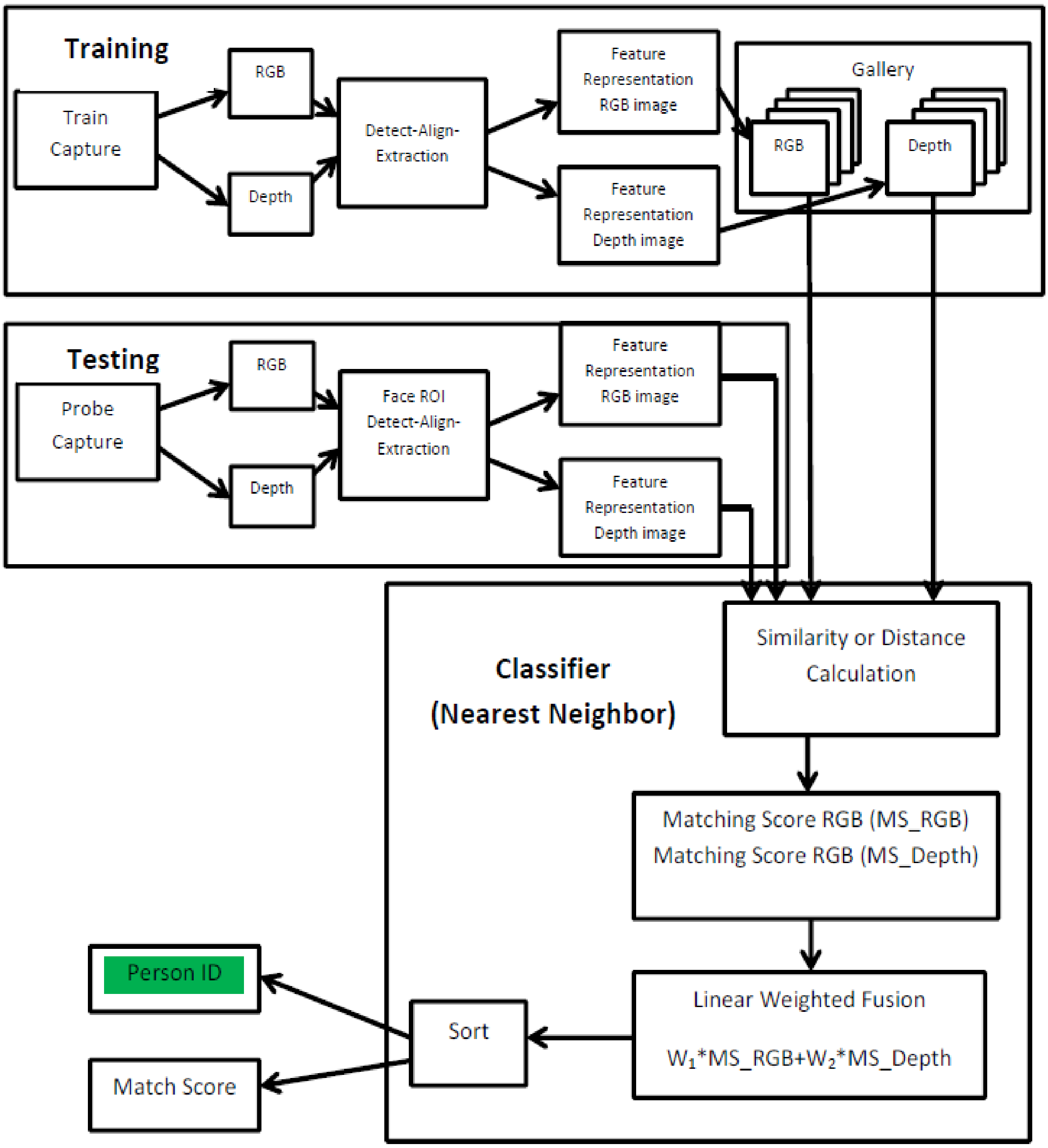


**Figure 2: System Functional Diagram for RGB-D Face Recognition**

 

**Table 1: Face Recognition Methods based on different representations for RGB or Depth images**

| Color Image Representation | Depth Image Representation |
| --- | --- |
| PCA | -- |
| -- | PCA |
| LDA | -- |
| -- | LDA |
| LBP | -- |
| -- | LBP |
| HOG | -- |
| -- | HOG |
| Gabor | -- |
| -- | Gabor |
| SIFT | -- |
| -- | SIFT |

Each of the feature technique used to represent the RGB image or depth image or both in case of fusion. Additionally in case of fusion, two different feature techniques used for RGB and depth image representation. The different combinations of feature techniques used for RGB or depth images that are applied for face recognition experiments are shown in table 1.

### *B. ROI Alignment in RGB and Depth Images:*

Ideally, face ROI in face recognition system should be well-centred and with right dimensions. The well catered property ensures the alignment in samples while calculating matching similarity between probe and gallery samples. With rights dimension of ROI, it becomes possible to have appropriate inclusion of face region while excluding background. Due to different properties of devices used in colour and IR sensors in Creative Labs 3D camera, face ROI in RGB is found to be not in alignment with face ROI in depth image. Thus, it is very critical to have alignment of both ROIs in terms of centre and dimensions, both, before using them for recognition. The algorithm, we employed in determining centre and dimensions of face ROI in depth image, is dependent on the face ROI in localization in RGB image. First, the face ROI in depth images is approximately mapped as it is from face ROI of RGB image. Then, it location is refined by locating nose-tip. The nose-tip is located at the point where a depth maximum is detected in approximated face ROI region. This is valid for the assumptions that face in images is frontal or near-frontal. Further, approximated depth ROI is passed through canny edge detection. Edge map locates the boundary of face due to large depth difference between face surface and background. From nose-tip location, boundary edges are searched on right and left side in edge map to find the distances between nose-tip and those edges. The dimensions of ROI face in depth are decided with nose-tip to edge distances.

## IV. EXPERIMENTAL RESULTS

### *A. Datasets:*

There are different subsets of Creative Labs 3D dataset. Few samples from each of them are shown in following figure 3. In pose subset of dataset, there are large pose variations in the range of +15 to -15 degree. Few samples from pose subset dataset are shown in figure 4.

### *B. Experiments:*

For training the face recognition system, 17 persons and 8 instances for each person were used in the gallery. There are 14 or 10 samples in test subjects of Frontal (10 samples), illumination-01 (14 samples), illumination-11 (14 samples) and pose (10 samples). Once face ROIs extracted and aligned, various methods of face recognition were applied and recognition rates (in %) are tabulated in table 2.

We also performed recognition experiment to examine the effect of fusion of RGB and Depth information. The results obtained with individual RGB or Depth info and fusion of both are shown figure 5.

### *C. Discussion:*

As compared to the PCA subspace, LDA has been more effective in RGB images. The local features Gabor and SIFT have not given significant performance neither in RGB nor in depth images. This may be due to the low resolution images we are dealing with in case of RGB. Another reason behind the failure of Gabor features in this experiment is that Gabor features are calculated only for 7 landmarks. Additionally, the precise localization of each of the landmark is very much dependent on the landmark localization step, which is not so accurate and reliable in this case. The landmark points in RGB image are determined by using publicly available “flandmark” API [16]. In depth image texture property is insignificant and hence local features become ineffective to pick up discriminative representation.

In contrast to global and local point features, semi-local features based on histogram of block segments of the image are able to pick up discriminative properties of texture from RGB images. HOG and LBP are the examples of these semi-local features. These features doesn’t dependent on the landmarks localization and still captures the texture property of local region of face images. This enables feature representation to be insensitive to the slight deviation in face ROI localization and also makes insensitive to the aspect ratio deviation of face ROI due to slight pose –deviation. The best performance is obtained with this kind of feature representation. It can be observed that with HOG features, recognition rate reaches to 96.21 %, while with LBP features it is 85% for frontal test dataset with good illumination. For slight pose-deviated samples, HOG and LBP gives recognition rate of 75.12 % and 69.50 % respectively. Since HOG feature is based on the gradients at each pixel and gradient is less sensitive to illumination and poses, HOG gives better performance than LBP.

In case of depth image representation, it is found that LDA and HOG are better performed features. LDA gives better results in case of varying illumination and pose deviation that HOG. In case of LDA, the depth range of face region is not normalized in the experiment performed here. However, with normalization of depth range in face region, performance with LDA can be further boosted up.

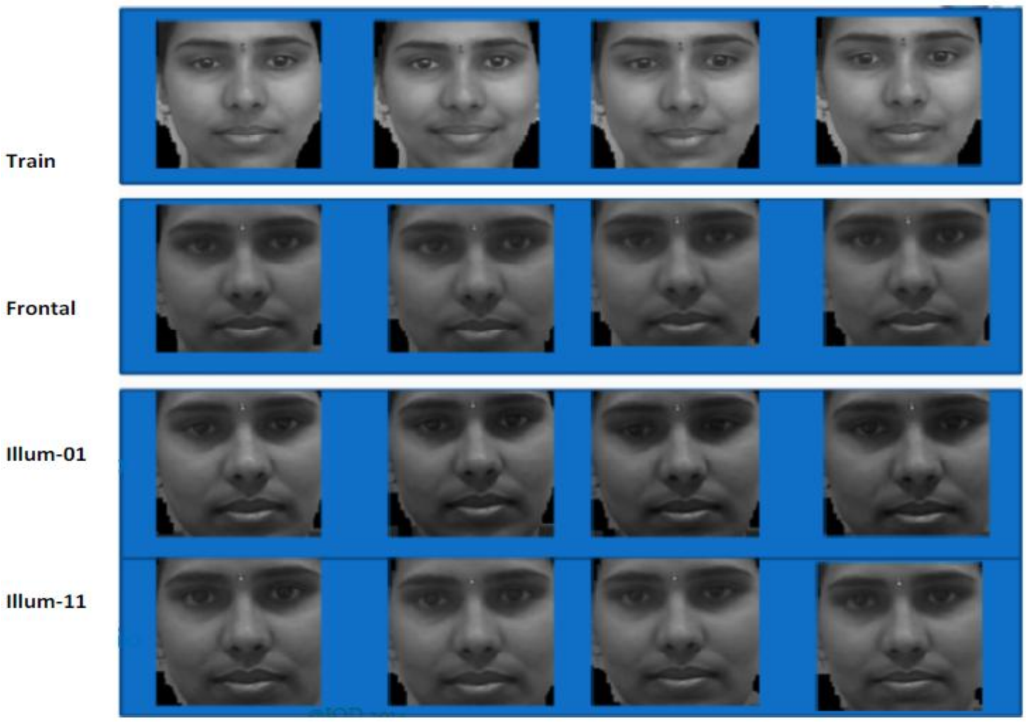


**Figure 3 Subsets of Creative Labs 2D dataset**

The effect of fusion is very much visible in the figure 4. Even with simple feature representation like PCA and LDA, the properly weighted fusion of RGB and Depth images can improve the result than RGB image alone.

**Figure 4 Pose subset of Creative Labs 3D Dataset**

**Table 2: Recognition Performance for Face Recognition Methods based on different representations for RGB or Depth images**

| Color Image Representation | Depth Image Representation | Recognition Rate (%) | | | |
|---|---|---|---|---|---|
| | | Illum-01 | Illum-11 | Pose | Frontal |
| PCA | -- | 38.23 | -- | -- | -- |
| -- | PCA | 19.74 | -- | -- | -- |
| LDA | -- | 87.81 | 73.10 | 65.88 | 78.23 |
| -- | LDA | 56.30 | 54.62 | 37.05 | 41.17 |
| LBP | -- | 85.50* | 74.00* | 69.50* | 77.50* |
| -- | LBP | -- | -- | -- | -- |
| HOG | -- | 96.21 | 80.25 | 75.12 | 81.17 |
| -- | HOG | 56.30 | 57.98 | 28.23 | 32.94 |
| Gabor | -- | 10.92 | -- | -- | -- |
| -- | Gabor | -- | -- | -- | -- |
| SIFT | -- | 15 | -- | -- | -- |
| -- | SIFT | -- | -- | -- | -- |

*No of classes in experiment are 20 and 10 samples in gallery for each person class

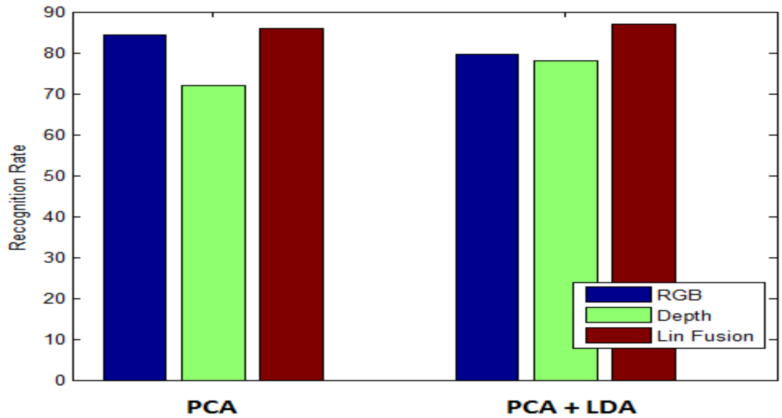


**Figure 5 Effect of Fusion of RGB and Depth images**

## V. CONCLUSIONS

Though, the overall structure of face of all human is same, there is intra-subject variation in the across the subjects in terms of depth pattern. Due to depth information of object (Face) surface available in addition to the RGB image, face recognition algorithm can be made more efficient and robust as compared to the algorithm, where only RGB image is employed. The performance analysis with different feature representations like PCA, LDA, SIFT, Gabor, HOG and LBP is done in this work. The experimental result shows that depth image can also be used for recognition. The fusion of RGB and Depth images improve the recognition rate. HOG and LBP feature descriptors are superior to the other subspaces. With HOG as a feature descriptor, face recognition rate for frontal images in proper illumination is reached to 96.21%, while for pose-deviated samples, it is 75.12 %.

## REFERENCES

[1] M. Turk and A.P. Pentland. Face recognition using eigenfaces. IEEE Conf. Computer Vision and Pattern Recognition, 1991.
[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):711–720, July 1997.
[3] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative common vectors for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1):4–13, 2005.
[4] Manjunath, B.S.; Chellappa, R.; von der Malsburg, C.; A feature based approach to face recognition Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on , 15-18 Jun 1992
[5] P. Rybski, D. Huber, D. Morris, and R. Hoffman. Visual classification of coarse vehicle orientation using histogram of oriented gradients features. In IEEE Intelligent Vehicles Symposium (IV), 2010, pages 921–928, June 2010.

 

[6] A. Satpathy, X. Jiang, and H.-L. Eng. Extended histogram of gradients features for human detection. In 17th IEEE International Conference on Image Processing (ICIP), 2010, pages 3473–3476, Sept. 2010.

[7] L. Hu, W. Liu, B. Li, and W. Xing. Robust motion detection using histogram of oriented gradients for illumination variations. In 2nd International Conference on Industrial Mechatronics and Automation (ICIMA), 2010, volume 2, pages 443–447, May 2010.

[8] T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1):51–59, 1996.

[9] Y. Fang and Z. Wang. Improving LBP features for gender classification. In International Conference on Wavelet Analysis and Pattern Recognition, 2008. ICWAPR '08., volume 1, pages 373–377, Aug. 2008.

[10] S. Junding, Z. Shisong, and W. Xiaosheng. Image retrieval based on an improved CS-LBP descriptor. In The 2nd IEEE International Conference on Information Management and Engineering (ICIME), 2010, pages 115–117, April 2010.

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–110.

[12] Li, B.Y.L.; Mian, A.S.; Wanquan Liu; Krishna, A., "Using Kinect for face recognition under varying poses, expressions, illumination and disguise," Applications of Computer Vision (WACV), 2013 IEEE Workshop on , vol., no., pp.186,192, 15-17 Jan. 2013

[13] Mauricio Pamplona Segundo, Sudeep Sarkar, Dmitry B. Goldgof, Luciano Silva, Olga Regina Pereira Bellon: Continuous 3D Face Authentication Using RGB-D Cameras. CVPR Workshops 2013: 64-69

[14] Goswami, G.; Bharadwaj, S.; Vatsa, M.; Singh, R., "On RGB-D face recognition using Kinect," Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on , vol., no., pp.1,6, Sept. 29 2013-Oct. 2 2013

[15] Nikisins, Olegs et al. "RGB-D-T based Face Recognition". International Conference on Pattern Recognition. Sweden: IEEE Computer Society Press.

[16] M. Uricar, V. Franc and V. Hlavac, Detector of Facial Landmarks Learned by the Structured Output SVM, VISAPP '12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, 2012.

 