



Fusion of RGB and Depth Images for Robust Face Recognition using Close-Range 3D Camera

Srinivas Kishan Anapu, Dr. Srinivasa Rao Peri

Department of Computer Science, Andhra University, Visakhapatnam, India.

HOD, Department of Computer Science, Andhra University, Visakhapatnam, India.

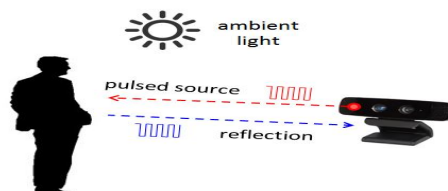
ABSTRACT: Face recognition systems are becoming popular in most of the applications ranging from gaming to surveillance. Conventionally, 2D cameras are used in face recognition applications and hence 2D face recognition algorithms are employed. However, it has become very easy to obtain 3D camera at cheaper cost, now days. It is interesting to see how 3D camera information can be used to overcome the challenges faced by conventional 2D face algorithms. Recently Creative Labs launched Senz3D camera which outputs the RGB and depth images with misalignment in object present in them. To overcome this, we proposed two algorithms in this paper. It is found that these algorithms have improved the recognition performance significantly. We also explored the various feature representations for the fusion of RGB and depth face ROIs. The fusion of RGB and depth ROIs representation is achieved at matching score level with nearest neighbor classifier. It is observed that among all combinations of feature representations for RGB and depth images, two methods, HOG (RGB) +LDA (Depth) and LBP (RGB) +LDA (Depth), have performed better for fusion system. Additionally, the fusion can play more critical role in the situations where light is not present in vicinity by adaptively increasing the fusion weight associated with depth information.

KEYWORDS: 2D Face; 3D Face; RGB-D camera; Face Recognition; Fusion; Depth image.

I. INTRODUCTION

The importance of face recognition systems is well known in various areas of applications such as e-governance, surveillance, border-security, office-attendance systems etc. Recently, face biometric systems are embedded into mobile, desktop and dedicated hardware machines installed at indoor or outdoor locations. The applications ranging from gaming to surveillance can use camera based face recognition systems. Seeing the importance of these systems, it is very important to have reliable and robust performance from face recognition system. Covariates like non-uniform lighting and pose variations are challenges in making application commercially viable. To remove the effects of these covariates, researchers across the world are focusing towards making biometrics flawless. More recently, there is an easy availability of 3D cameras is observed. Thus, it is worth to study the effectiveness of 3D cameras to overcome the undesired effects due to illumination changes and pose variation.

Fig1: TOF (Time of Flight) technique for depth (Range)





International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

The 3D camera which outputs RGB and range (depth) images is now easily available at cheaper cost [17]. Recently, 3D cameras with RGB-D information are launched from Microsoft (Kinect Sensor), ASUS and more recently from Creative Labs (Senz3D). Though, the conventional camera uses stereo matching calculation for determining the 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 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. 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 Senz3D. To the best of our knowledge, there is no experimental study so far published in the literature with data samples acquired by Senz3D 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 3D camera is that it has lower resolution as compared to the Kinect images. However, advantage with this 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 Senz3D camera can be employed.

The fusion of RGB and depth face images for the face recognition is the main objective of this paper. While fusing the two sources, it is important to ensure face images are registered properly. In Senz3D camera, face ROI is not aligned between depth and RGB images. Thus, it is essential to find the face region in both images separately in order to have aligned face across them. This paper presents the study on the different approaches of face ROI align-and-extract step. The performance analysis of fusion system at matching score level with different feature representations like PCA, LDA, SIFT, Gabor, HOG and LBP is performed in this paper. These features were applied to both, RGB and depth images and explored the applicability of feature in different scenarios. The remaining part of the paper is organized as follows. In section II, the literature review is presented. In next section III, possible approaches of face ROI align-and-extract are described. In section IV, face recognition system and different features are elaborated. Experimental results are presented in section V and discussion over the results is covered in this section. Finally, paper is concluded with remarks in section VI.

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 smoothing 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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

images by introducing a database of 51 persons including facial images of different rotations, illuminations, and expressions.

In [18], authors have worked towards reliable face detection system using RGB and depth data together. 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 [19], 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 in [20]. Various global and local features applied to represent face region and fuse the data from RGB and depth images. The work described in [21] deals in process of face synthesis by image morphing from less expensive 3D sensors such as kinect that are prone to sensor noise. This synthesis can be used to make 3D dataset for the study of face recognition methods.

III. FACE ROI ALIGN-AND-EXTRACT APPROACHES

Ideally, face ROI in face recognition system should be well-centred and with right dimensions. The well centered 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 3D camera, face ROI in RGB is found to be not in alignment with face ROI in depth image. This is illustrated in figure 2. However, since binary depth mask outputted by this 3D camera is aligned with RGB image, it can be used to remove the background in the case where object is close to camera and background is far from it. This step helps in making face detection computationally faster in RGB image. But, as depth image is not aligned with RGB image, it is very critical to have alignment of ROIs in terms of centre and dimensions, both, before using them for recognition.

In order to overcome this problem, we developed the different approaches of faceROI alignment and extraction modules and studied their performances. We used three approaches as described below:

Baseline Approach: In this approach no alignment module is employed. Face ROI detection is easier in the RGB image than in the depth image. The face ROI from RGB image is mapped as it is onto depth image. Even dimensions of ROI are kept same.

Algorithm A: In this algorithm, calculations of centre and dimensions of face ROI in depth image are dependent on the face ROI localization in RGB image. First, the face ROI in depth image is approximately mapped as it is from face ROI of RGB image. Then, its 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. Now, nose-tip point is considered to be a centre of ROI. The dimensions of depth ROI are kept proportional to that of RGB ROI.

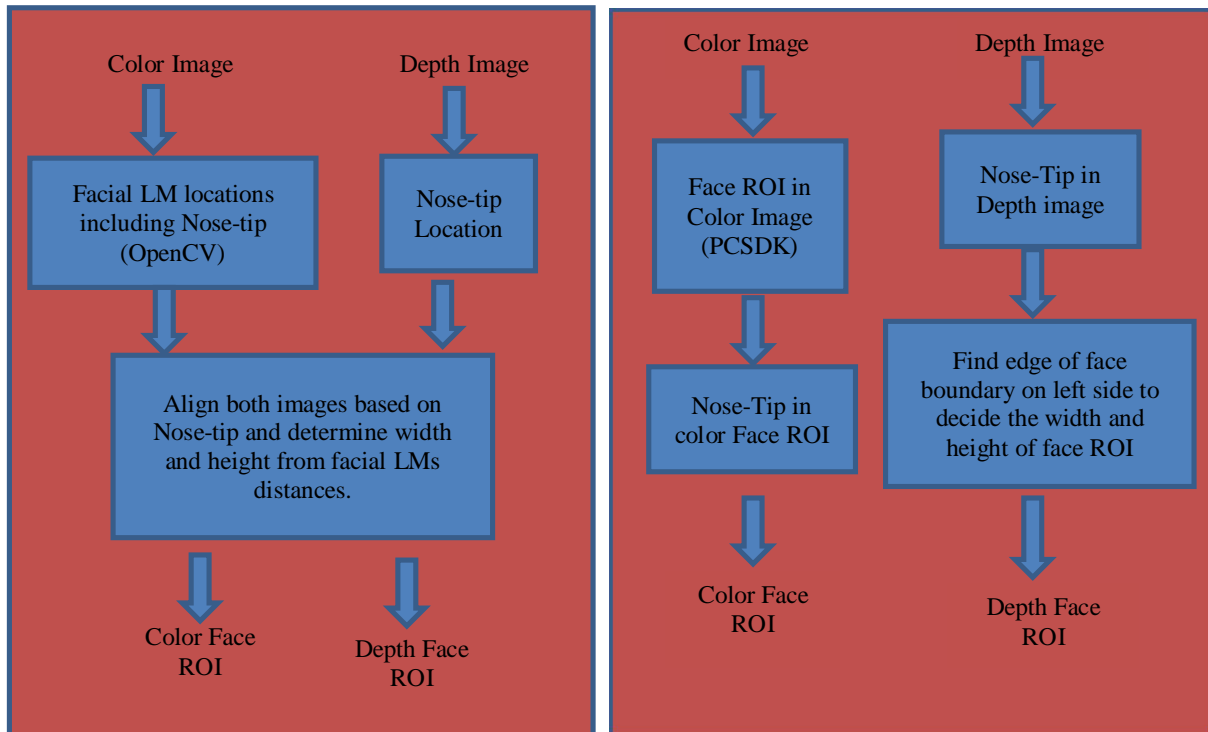


Fig 2: Misalignment of object in RGB (left) and Depth (Middle) images (Observe -Scaled down face in depth image) and Depth-Mask (right) image in align with RGB image.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014



Algorithm –A- Alignment and determination of Width and Height of Face ROI in depth image from RGB ROI.

Algorithm –B- Alignment and determination of Width and Height of Face ROI in depth image from face boundary edges.

Fig3 Proposed Face ROI detect-and-extract algorithms

Algorithm B: In this algorithm, step of finding the centre of ROI is same as that in algorithm A. However, dimensions of depth ROI are calculated differently. To determine the ROI dimensions, 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. FUSION BASED FACE RECOGNITION SYSTEM

In face Recognition systems, feature representation and classifier are two important modules that affect the performance of application. It is evident in [21] that fusion of depth and RGB face ROIs can boost the recognition performance. Depth image can also play critical role when there is no illumination. The proposed face recognition system based on fusion of RGB and depth images is shown in figure 4. In following subsection, feature representations and fusion systems are explained in detail.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

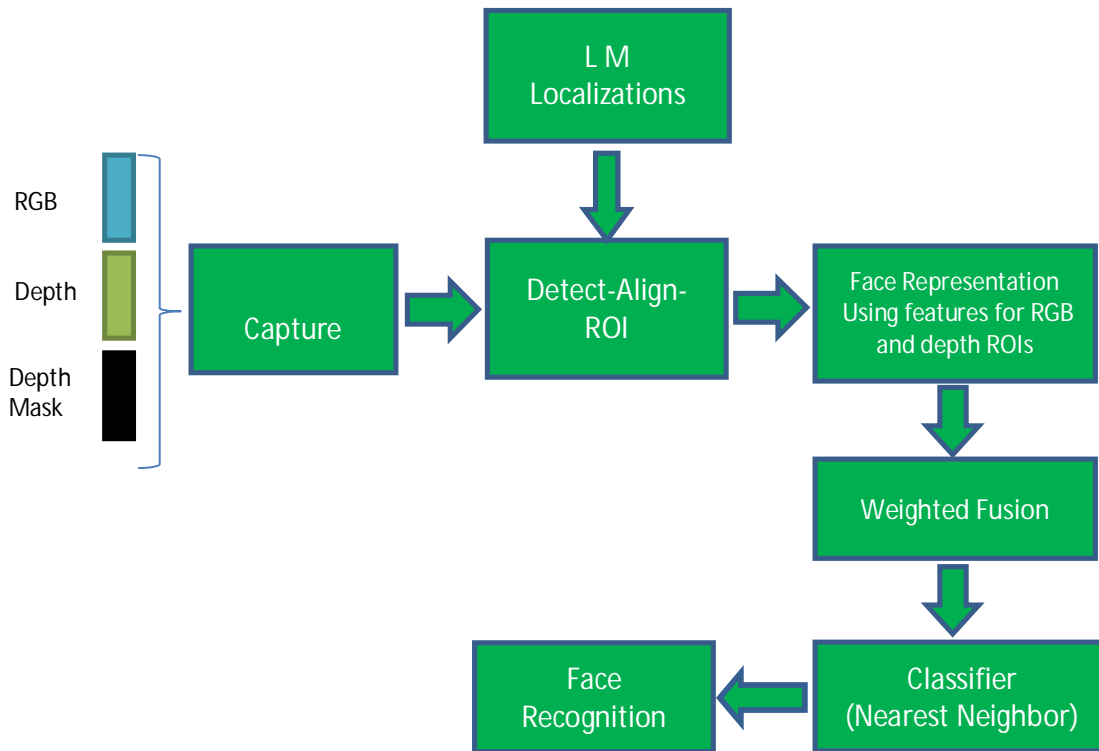


Fig 4: Fusion System for RGB and Depth based Face Recognition

A. Features:The different feature representation techniques that are considered in this 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



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

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 centre 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.

However, it is seen in [21] that Gabor and SIFT features representation doesn't provide good perfraomce in its current form. Thus, expreimetns with the features are ignored in this work.

B. Fusion: The features described earlier are used for representing the face ROI in RGB and depth images. We used various combination of these features applied to fusion of RGB and depth images. These combinations are tabulated below in table 1. The matching score between probe and gallery images are linearly weighted to calculate final matching score. Matching score is applied to nearest neighbour classifier to identify the recognized person.

Table 1: Fusion based methods based on different representations for RGB and Depth images

Sr. No.	Color Image Representation	Depth Image Representation	Weighted Fusion (W_1, W_2)
1	PCA	PCA	$W_1 * C + W_2 * D$
2	LDA	LDA	$W_1 * C + W_2 * D$
3	LBP	LBP	$W_1 * C + W_2 * D$
4	HOG	HOG	$W_1 * C + W_2 * D$
5	LBP	PCA	$W_1 * C + W_2 * D$
6	LBP	LDA	$W_1 * C + W_2 * D$
7	HOG	PCA	$W_1 * C + W_2 * D$
8	HOG	LDA	$W_1 * C + W_2 * D$

V. EXPERIMENTAL RESULTS

The dataset used here for all the experiments is explained in detail in [21]. First, we analysed the effect of face ROI align-and-extract modules. To do so, we performed the recognition experiments using two sub datasets, Frontal and Pose datasets. Since, face ROI in RGB and depth images are not aligned and not of the same dimensions, face ROIs from both images need to be detected separately. We developed two algorithms to align the face ROIs. The recognition rates with these algorithms are presented in table 2. It is obvious to see that effect of misalignment is more in pose

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

deviated dataset than in frontal dataset. Both algorithms A and B have improved the recognition performance by means of alignment.

Table 2 Recognition with Depth image only with different Detect-Align-Extract Algorithms

FaceROI-Detect-Align-Extract Algorithm	Recognition Rate (%)	
	Frontal Dataset (illum_11)	Pose Dataset
RGB-to-Depth (without alignment – (Baseline Algo)	42.01	33.52
RGB-NoseTip-Depth-ROI (Algo -A)	60.08	41.76
RGB-NoseTip-Depth-Edge-ROI (Algo-B)	60.50	37.05

Table 3: Recognition Performance for fusion based methods based on different representations for RGB or Depth images

Color Image Representation	Depth Image Representation	Weighted Fusion (W_1, W_2)	Recognition Rate (%)			
			Illum-01	Illum-11	Pose	Frontal
PCA	PCA	C (RGB)	--	--	--	--
		$W_1 * C + W_2 * D$	--	--	--	--
LDA	LDA	C (RGB)	--	--	65.88	78.23
		$W_1 * C + W_2 * D$	--	--	68.23	78.82
LBP	LBP	$W_1 * C + W_2 * D$	--	--	--	--
HOG	HOG	$W_1 * C + W_2 * D$	--	--	--	--
LBP	PCA	$W_1 * C + W_2 * D$	--	--	--	--
LBP	LDA	C(RGB)	85.50*	74.00*	69.50*	77.50*
		$W_1 * C + W_2 * D$	87.5*	79.0*	72.0*	76.0*
HOG	PCA	$W_1 * C + W_2 * D$	--	--	--	--
HOG	LDA	C (RGB)	--	80.25	75.12	--
		$W_1 * C + W_2 * D$	--	81.09	76.47	--

*No of classes in experiment are 20 and 10 samples in gallery for each person class.
-- The values of recognition rate in fusion that are smaller than 50% are ignored

We also performed recognition experiment to examine the effect of fusion of RGB and Depth information with various features representation. The results obtained with fusion are shown table3. The recognition obtained using fusion is compared with recognition obtained with RGB image only. It can be seen that recognition with fusion is slightly higher



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

than that with RGB image only. Among all combination of feature representation for RGB and depth images, HOG(RGB) +LDA (Depth) and LBP(RGB) +LDA (Depth) have shown their appropriateness for fusion system. It is evident from this that LDA feature subspace has good capability to represent the depth image. This fusion system can be even more critical in case of situation where there is no light present; consequently, adaptive fusion can give more weightage to the depth images.

VI. CONCLUSION

It is easy to obtain 3D cameras at cheaper cost now days. Thus, it is interesting to see how 3D camera information can be used to overcome the challenges faced by conventional 2D face algorithms. Creative 3D camera outputs the RGB and depth images with misalignment in object present in images. To overcome this we proposed two algorithms in this paper and performances obtained with these two algorithms are reported. It is found that these two algorithms have improved the recognition performance due to proper alignment. Further, various features, which could be a candidate for feature representation for the RGB or depth face ROIs are explored in this work. The fusion of RGB and depth ROIs representation is achieved at matching score level with nearest neighbor classifier. It is observed that among all combinations of feature representation for RGB and depth images, two methods, HOG (RGB) +LDA (Depth) and LBP (RGB) +LDA (Depth), have performed better for fusion system. Thus, LDA feature subspace proved to be good representation for the depth images.

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. Stpathy, 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.
17. M Hossny, D Filippidis, W Abdelrahman, H Zhou, M Fielding, J Mullins, L Wei, D Creighton, V Puri, S Nahavandi, "Low cost multimodal facial recognition via Kinect sensors", in proceeding of: Land Warfare Conference 2012: Potent Land Force for a Joint Maritime Strategy.
18. Loris Nanni, Alessandra Lumini, Fabio Dominio, Pietro Zanuttigh, Effective and precise face detection based on color and depth data, Applied Computing and Informatics, Volume 10, Issues 1–2, January 2014, Pages 1-13,
19. Kolb Andreas, Koch Reinhard," Fusing Time-of-Flight Depth and Color for Real-Time Segmentation and Tracking", Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2009



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 7, July 2014

20. Juan Zhou, Yongping Li and Jingyan Wang, "2D&3D-ComFusFace: 2D and 3D Face Recognition by Scalable Fusion of Common Features", IJCSNS International Journal of Computer Science and Network Security, VOL.12 No.3, March 2012
21. Ramalingam, S.; Nguyen Trong Viet, "3D Face Synthesis with KINECT," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on , vol., no., pp.4433,4438, 13-16 Oct. 2011.
22. S Kishan; Srinivasa Rao, "2D and 3D Face Recognition using Close-Range RGB-D Camera", submitted in IJARCSSE 2014.

BIOGRAPHY

Srinivas Kishan Anapu is a PhD Student in Computer Science Engineering Department, Andhra University, India. He is also the CEO of Insights of Data Pvt. Ltd. He received his BE in Civil Engineering from Andhra University, Vizag, India and Masters in Mathematics from University of Oklahoma, Norman, OK, USA and another Masters in Computer Science from Midwestern State University, Wichita Falls TX, USA. His research interests include but not limited to developing face recognition algorithms and products related to big data analytics.