Applied Computer Science, vol. 17, no. 3, pp. 73–81
doi: 10.23743/acs-2021-22

Submitted: 2021-06-13 | Revised: 2021-09-08 | Accepted: 2021-09-21

RGB-D, Kinect, Local Binary Pattern, Pattern Recognition, Feature Extraction, Histogram, Face Recognition

Sunil Kumar B L [0000-0002-7406-036X]*, *Sharmila Kumari M* [0000-0002-0707-0891]**

RGB-D FACE RECOGNITION USING LBP-DCT ALGORITHM

Abstract

Face recognition is one of the applications in image processing that recognizes or checks an individual's identity. 2D images are used to identify the face, but the problem is that this kind of image is very sensitive to changes in lighting and various angles of view. The images captured by 3D camera and stereo camera can also be used for recognition, but fairly long processing times is needed. RGB-D images that Kinect produces are used as a new alternative approach to 3D images. Such cameras cost less and can be used in any situation and any environment. This paper shows the face recognition algorithms' performance using RGB-D images. These algorithms calculate the descriptor which uses RGB and Depth map faces based on local binary pattern. Those images are also tested for the fusion of LBP and DCT methods. The fusion of LBP and DCT approach produces a recognition rate of 97.5% during the experiment.

1. INTRODUCTION

For the authentication of the individuals, biometric verification using signature, iris, retina, fingerprints, voice and face is being used. Face recognition is considered amongst these techniques as one of the most popular, collectable and easily accessible systems. The method of biometric identification consists of collecting the biometric data of the individual and matching it with biometric data of the specific person. Some of the problems involved in this process are changes in posture and illumination, facial background etc. Recognition of the human face extracts features from the face and compares them with all facial images stored in the database. Local binary pattern (LBP) extraction technique can be used for clustering, classification, and segmentation of features. Using LBP and other extended methods, the texture classification (Song et al., 2015; Yu et al., 2014) is achieved which involves easy calculation and also produce good results. RGBD is the image obtained with the help of Kinect are the combination of red, green, blue color information and the depth information (Abebe & Hwang, 2019). These red green blue color uses 8-bit representation and depth uses 16-bits. To reduce the space 16-bit is transformed into 11-bits. Depth is measured in terms of millimeters and the value ranges from "1" to "10.000". RGB along with the depth can be used to get the improved results in digital image processing.

* Canara Engineering College, India, blsuny@gmail.com
** PA College of Engineering, India, sharmilabp@gmail.com

These images can be obtained with the low price Kinect camera. These images are used in 3D image reconstruction, Augmented Reality, Robotics, Pattern Recognition and Image Processing.

2. RELATED WORK

The RGBD datasets by Microsoft Kinect have motivated improvements in the areas from reconstruction to gesture recognition. Extracting relevant information helps to obtain proper data for the needs. In recent years, there has been increased focus on usage of RGB-D cameras for development of 3D scene understanding and object detection algorithms (Lin, Fidler & Urtasun, 2013). Kinect sensor based surveillance systems have also been deployed for border control (Chowdhury & Vatsa, 2016). Usage of Kinect sensors for indoor surveillance systems is an interesting research problem due to its capability of capturing RGB, Depth and NIR footage from a single camera unit. Face is highly analyzed biometric (Zhao et al., 2003) and face recognition gives better performance on 3D images compared to 2D images in the presence of pose and illumination covariates. Additional discriminative information provided by the depth map, increases the rate of recognition (Silberman et al., 2012; Wang et al., 2012). Recently, decreased cost of depth sensors has made it feasible to use the images in different applications and has consequently led to increased curiosity in RGB-D face detection (Hg et al., 2012) and face recognition. (Han et al., 2013) introduced the RGB and Depth based images for face-recognition (Huynh, Min & Dugelay, 2012) when depth sensors are started. Because of the difficulty in managing pose and lighting on 2D face recognition, 3D face recognition approaches have been suggested. The 3D image is less sensitive to changes in lighting and therefore more useful for correcting pose variations. The downside to using facial recognition approaches based on 3D is the high cost (Goswami, Vatsa & Singh, 2013) of the conventional 3D sensors. Kinect (Cruz, Lucio & Velho, 2012) tools are the alternative to these costly scanners and are capable of collecting the depth information very precisely. The depth information shows the possibility of accurate 3D face reconstruction which is used in the recognition of faces (Hsu et al., 2014). The performance improvements can be seen using local binary pattern descriptor due to RGB and depth images (Min, Kose & Dugelay, 2014; Zohra, Rahman & Gavrilova, 2016).

2.1. LBP based face descriptors

Ojala proposed Local Binary Pattern in 1990, which uses an effective and easy method to extract the local characteristics from the given face data. The two important LBP properties are (a) Simple computational features and (b) Working for various conditions of illumination. With the support of the neighborhood, LBP feature vector calculation can be performed by first dividing the entire image into cells. Then the middle pixel of the cell is compared with the eight neighboring pixels in the clockwise or anticlockwise direction around it. If (middle pixel-value > neighbor pixel value), then return "0" otherwise "1". The resulting 8-digit binary number transforms into decimal number. The histogram is computed for the features collected.

To measure the LBP code, the cell size of 3×3 is used. Each pixel in the image is compared to its 8 neighbors and encoded to "0" or "1" based on the comparative values as shown in the figure 1.

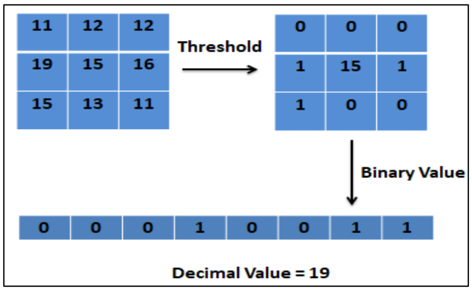
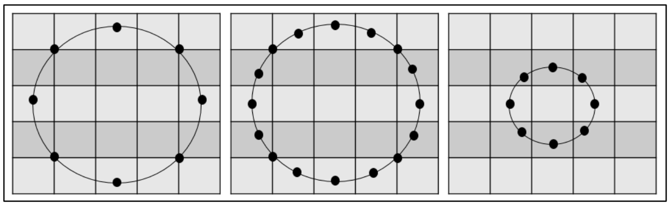


Fig. 1. LBP code calculation

Concatenate all encoded binary values in either clockwise or anticlockwise direction. This binary number consisting of 8 bits is translated to decimal number that is assigned to the cell's central pixel.



(a) r = 2 and p = 8 (b) r = 2 and p = 16 (c) r = 1 and p = 8

Fig. 2. Neighbour Pixels

The same process for the full image is continued. The cell size can be increased, and the extended LBP operator uses the notation (p, r), as shown in Figure 2, to denote a 'p' pixel neighbourhood on a 'r' radius sphere. Circular (8, 2) showing 8 neighbours for the central pixel in the circle with a radius of 2 and for circular (16, 2) showing 16 neighbours for the central pixel in the circle with a radius of 2 and also (8, 1) reveals that around the central pixel there are 8 neighbours with a radius of 1.

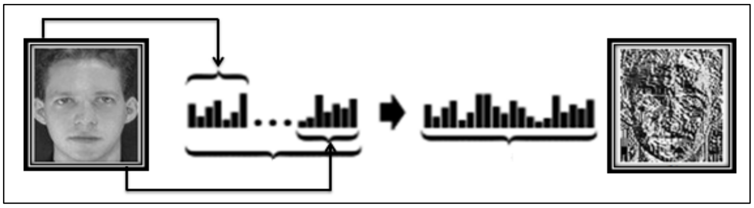


Fig. 3. Extracting feature histogram from the face image

Figure 3 demonstrates the exact workings of LBP. At first divide the window of the image into cells. Then, the histogram is determined for every cell. Concatenation of the LBP histogram obtained for each cell gives a resultant histogram feature.

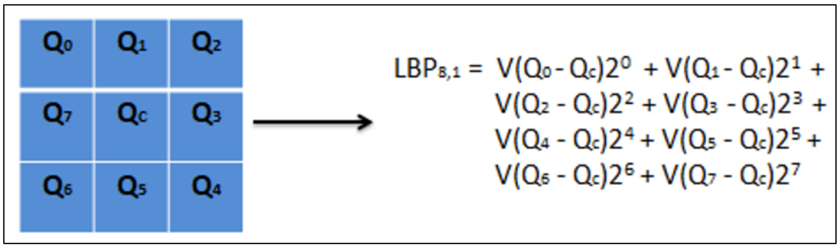


Fig. 4. LBP Calculation Code

LBP characteristics of the circular neighbours of a central pixel are indicated by LBPp,r, here p is the number of circular neighbouring points with r radius. Figure 4 shows one such calculation when p=4 and r=1. The following code is obtained in LBPp,r

$$LBP_{p,r} = \sum_{Q=0}^{Q-1} V(Q_i - Q_c)\, 2^i \,, V(x) = \begin{cases} 1 \; if \; x \geq 0 \\ 0 \;\; otherwise \end{cases} \tag{1}$$

Where Q_c gives the central pixel grey-level value, and Q_i is the value of the surrounding pixel in the circular neighbour. To reduce the number of bins, the idea of uniform pattern is used. It is possible to distinguish uniform patterns if there are at most two transitions from "1" to "0" or "0" to "1". Consider pattern 1111111111 with no transition or pattern 00110000 with two transitions as examples of a uniform transition, whereas pattern 10101011 with six transitions is not uniform. This restriction reduces the LBP patterns automatically from 256 to 58.

2.2. DCT based face descriptors

Using DCT, Invertible linear transforms image into feature vectors with low and high frequency coefficients. 2D-DCT converts the face image into a frequency domain, and reverse processes can also be done using invert 2D-DCT (Chen & Chen, 2010; Shermina 2011). The 2D-DCT on an image of dimension r x c:

$$\mathrm{F(u,v)} = \alpha(u)\alpha(v) \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} f(i,j) \cos\left[\frac{u(2i+1)\pi}{2r}\right] \cos\left[\frac{\mathrm{v(2j+1)\pi}}{\mathrm{2c}}\right] \tag{2}$$

$$\text{where, } \alpha(u) = \sqrt{\frac{1}{r}} \text{ for u} = 0 \tag{3}$$

$$\alpha(u) = \sqrt{\frac{2}{r}} \text{ for u} = 1 \text{ to r} - 1 \tag{4}$$

$$\text{and } \alpha(v) = \sqrt{\frac{1}{c}} \text{ for v} = 0 \tag{5}$$

$$\alpha(v) = \sqrt{\frac{2}{c}} \text{ for v} = 1 \text{ to c} - 1 \tag{6}$$

The important few, low frequency DCT components of the training and testing images are extracted. Then the more relevant extracted information also called the feature vectors of training and testing images are compared. The Euclidean distance measure is used for the classification.

3. PROPOSED METHOD

Face recognition algorithm uses 2D or 3D input images for the recognition. Usually, 3D faces are used to create advanced algorithms to improve the performance, but these images are difficult to get because of the cost. This paper uses RGB-D images alternate to 3D images to implement certain face recognition algorithms.

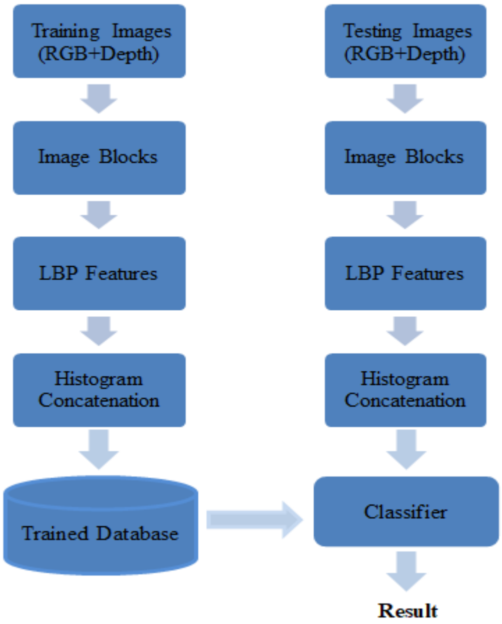


Fig. 5. Block diagram of face recognition using LBP

Local binary pattern is an extremely effective method to describing a digital image's texture, and it was ideal for obtaining features for face recognition systems as shown in Figure 5. First of all, split the image for extracting the LBP histograms and then stored in a vector. This vector represents the features of face and can be used for further processing. Much of the current face recognition work using RGB-D has focused on controlled environments. But in a real-world environment, the sensors can also capture images from a distance. The accuracy of the depth images is strong when the images are taken at close ranges. Yet, with the camera's large distance to the actual face, depth sensors struggle to capture images of good quality depth. It may not be desirable to use poor-quality pictures in these circumstances. So, the proposed approach uses both RGB and depth images for the better recognition of faces.

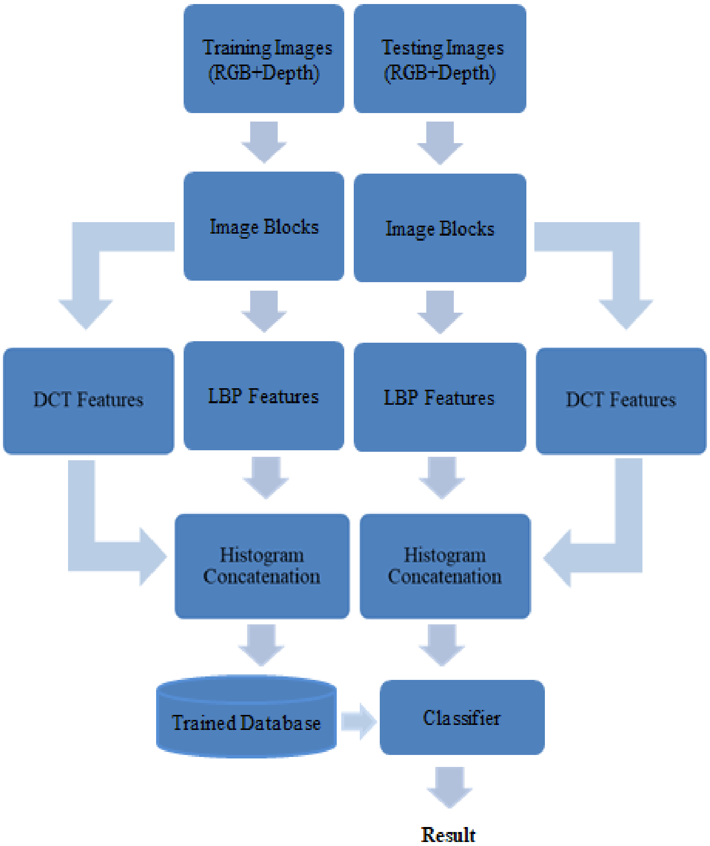


Fig. 6. Block diagram of face recognition using LBP and DCT

The proposed algorithm calculates an LBP descriptor and DCT features for the RGB images as well as for the depth images. The combined histograms are computed based on the collected features from LBP and DCT. The steps as shown in Figure 6 is repeated for all the training datasets. Finally, features of testing data are extracted and compared with all the features of the trained data set to recognize the image. The series of experiments were performed using the combined effect of features LBP and DCT. To compare and recognize the images, Euclidian distance is used in this experiment.

4. RESULTS AND DISCUSSIONS

The above methods are tested on IIITD RGB-D face images. This data-set consists of 106 subjects captured using the sensor version 1 of Microsoft Kinect. It has a large number of photos, ranging from 254 to 11 pictures per subject, per fold. The RGB and the depth images are recorded separately as 24-bit images. For these images the 640×480 resolution is used.

Fig. 7. Sample RGB-D images of a person from IIITD RGBD face dataset

The face recognition algorithms are executed by changing the count of training samples for each subject. For all such cases, the accuracy of recognition is noted as shown in the table 1 and Figure 8 shows the results on RGB-D face image data collection using conventional LBP and LBP+DCT algorithms.

Tab. 1. Recognition rate using LBP and LBP+DCTmethods

	Recognition Rate (%) (LBP)			Recognition Rate (%) (LBP+DCT)		
	RGB	Depth	RGB+ Depth	RGB	Depth	RGB+ Depth
1	85.0	75.0	92.5	85.0	82.5	92.5
2	87.5	80.5	95.0	87.5	80.5	95.0
3	90.0	77.5	92.5	90.0	82.5	95.0
4	92.5	77.5	95.0	95.5	90.0	97.5
5	92.5	75.0	95.0	95.5	90.0	97.5
6	87.5	60.0	87.5	90.0	67.5	95.0

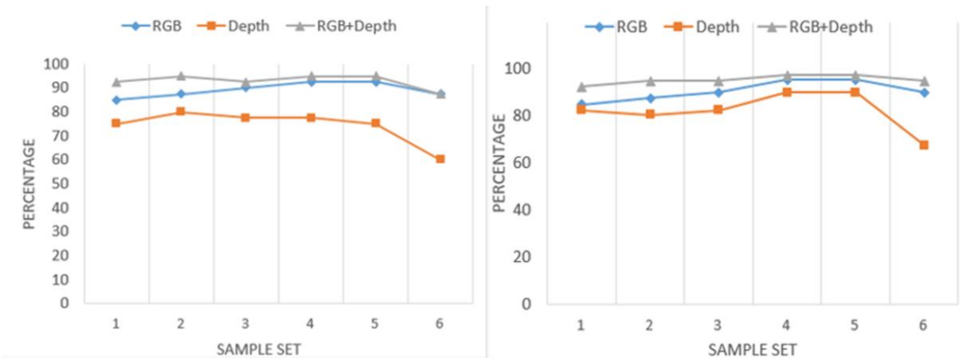


Fig. 8. Results of LBP and LBP+DCT on RGB, Depth and RGBD images

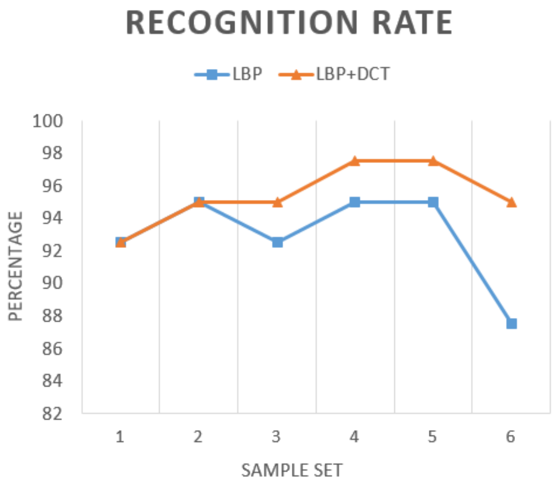


Fig. 9. Comparison of LBP and LBP+DCT

Figure 9 clearly shows combining LBP and DCT algorithms gives better result than only LBP.

5. CONCLUSION

Algorithms for face recognition generally use features of 2D or 3D images. This paper compares the effect of algorithms for extraction of features such as LBP and the fusion of LBP+DCT using both RGB and depth images. IIT-D RGB-D face image set is used for the analysis of above algorithms. For these images proposed algorithm shows 97.5% of recognition. As a future work different combinations of the feature extracting algorithms can be used on RGB-D dataset.

REFERENCES

Abebe, H. B., & Hwang, C. L. (2019). RGB-D face recognition using LBP with suitable feature dimension of depth image. *IET Cyber-Physical Systems: Theory & Applications*, *4*(3), 189–197. https://doi.org/10.1049/iet-cps.2018.5045

Chen, P. Z., & Chen, S. L. (2010). *A new face recognition algorithm based on dct and lbp. In Quantitative Logic and Soft Computing 2010* (pp. 811–818). Springer. https://doi.org/10.1007/978-3-642-15660-1_82

Chowdhury, A., & Vatsa, M. (2016). *RGB-D face recognition in surveillance videos* (Doctoral dissertation). Retrieved from https://repository.iiitd.edu.in/jspui/handle/123456789/440

Cruz, L., Lucio, D., & Velho, L. (2012). Kinect and rgbd images: Challenges and applications. In *2012 25th SIBGRAPI conference on graphics, patterns and images tutorials* (pp. 36–49). IEEE. https://doi.org/10.1109/SIBGRAPI-T.2012.13

Goswami, G., Vatsa, M., & Singh, R. (2014). RGB-D face recognition with texture and attribute features. *IEEE Transactions on Information Forensics and Security*, *9*(10), 1629–1640. https://doi.org/10.1109/TIFS.2014.2343913

Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics*, *43*(5), 1318–1334. https://doi.org/10.1109/TCYB.2013.2265378

Hg, R. I., Jasek, P., Rofidal, C., Nasrollahi, K., Moeslund, T. B., & Tranchet, G. (2012). An rgb-d database using microsoft's kinect for windows for face detection. In *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems* (pp. 42–46). IEEE. https://doi.org/10.1109/SITIS.2012.17

Hsu, G. S. J., Liu, Y. L., Peng, H. C., & Wu, P. X. (2014). RGB-D-based face reconstruction and recognition. *IEEE Transactions on Information Forensics and Security*, *9*(12), 2110–2118. https://doi.org/10.1109/TIFS.2014.2361028

Huynh, T., Min, R., & Dugelay, J. L. (2012). An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In *Asian Conference on Computer Vision* (pp. 133–145). Springer. https://doi.org/10.1007/978-3-642-37410-4_12

Lin, D., Fidler, S., & Urtasun, R. (2013). Holistic scene understanding for 3d object detection with rgbd cameras. In *Proceedings of the IEEE international conference on computer vision* (pp. 1417–1424). IEEE. https://doi.org/10.1109/ICCV.2013.179

Min, R., Kose, N., & Dugelay, J. L. (2014). Kinectfacedb: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(11), 1534–1548. https://doi.org/10.1109/TSMC.2014.2331215

Shermina, J. (2011). Illumination invariant face recognition using discrete cosine transform and principal component analysis. In *2011 International Conference on Emerging Trends in Electrical and Computer Technology* (pp. 826–830). IEEE. https://doi.org/10.1109/ICETECT.2011.5760233

Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. In *European conference on computer vision* (pp. 746-760). Springer. https://doi.org/10.1007/978-3-642-33715-4_54

Song, K., Yan, Y., Zhao, Y., & Liu, C. (2015). Adjacent evaluation of local binary pattern for texture classification. *Journal of Visual Communication and Image Representation*, *33*, 323–339. https://doi.org/10.1016/j.jvcir.2015.09.016

Wang, J., Liu, Z., Chorowski, J., Chen, Z., & Wu, Y. (2012). Robust 3d action recognition with random occupancy patterns. In *European Conference on Computer Vision* (pp. 872–885). Springer. https://dl.acm.org/doi/10.5555/2964398.2964463

Yu, W., Gan, L., Yang, S., Ding, Y., Jiang, P., Wang, J., & Li, S. (2014). An improved LBP algorithm for texture and face classification. *Signal, Image and Video Processing*, *8*(1), 155–161. https://doi.org/10.1007/s11760-014-0652-5

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, *35*(4), 399–458. https://doi.org/10.1145/954339.954342

Zohra, F. T., Rahman, M. W., & Gavrilova, M. (2016). Occlusion detection and localization from Kinect depth images. In *2016 International Conference on Cyberworlds (CW)* (pp. 189–196). IEEE. https://doi.org/10.1109/CW.2016.40