Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 3991-3995
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6221

# Feature extraction for RGB-D cameras

Reeman Jumaa Abd Ali[a,*], Aqiel Almamori[a]

[a]*Electronic and Communication Engineering Department, University of Baghdad, Baghdad, Iraq*

*(Communicated by Madjid Eshaghi Gordji)*

**Abstract**

A proposed feature extraction method for RGB-D cameras is developed. The proposed method for feature extraction is based on a Histogram of oriented gradient HOG which is used to extract the features of RGB image alongside with Histogram of oriented depth HOD which extracts the depth features to find a new different feature vector that is better describe the image. The new feature extraction method is benchmarked by human action recognition of pause images and shows better performance than HOG and HOD.

*Keywords:* RGB-D cameras, RGB image, benchmarked.
*2010 MSC:* 54H30

## 1. Introduction

The features of the image are important in artificial intelligence and machine learning, the success rate of object detection, prediction, recognition, and classification depends mainly on the strength of the features vectors of images. In recent years, attention has been drawn to feature fusion techniques to improve the level of classification of images and detection of objects within images, combining different types of features leads to a stronger processing enhancement in machine learning [14]. Image fusion is important for improving the perception of a scene and minimizes uncertainty in the result. RGB-D is an important camera in robotics, drones, self-driving vehicles and monitoring. It gives in addition to the RGB images a depth map for the objects in the RGB image. Due to this property, RGB-D cameras attracted the researcher to improve the current algorithms [13]. One of the areas that can be improved by utilizing the features of RGB-D is feature extraction.

In paper [6], a generation of a combination of a different set of cues adaptively selects the most proper cues to better classifications of images. In paper [5] the author cascades a Gray level co-occurrence matrix (GLCM), HOG features and Hu moment features in one multi-features matrix.

*Corresponding author
*Email address:* r.ali1020@coeng.uobaghdad.edu.iq (Reeman Jumaa Abd Ali)

*Received:* November 2021 *Accepted:* January 2022

The new features matrix tackles poor classification problems and recognition of distorted texture images.

To ensure the grantee of high accuracy, paper [15] introduces a fast fusion feature algorithm by working with shallow layer network feature and large pre-trained convolution neural network feature then fused them with the traditional features like (SIFT) and (HOG). In [11] object feature detection is evaluated by the fusion of two types of images, intensity, and range images.

The author of [8] shows the power of the combination of two traditional descriptors, histogram of oriented gradient (HOG) and histogram of optical flow (HOF) for detection of human action recognition. In paper [1] a modified new descriptor is developed by adding depth features obtained by a histogram of oriented depth (HOD) to the original histogram-oriented gradient (HOG) is proposed.

Histogram of oriented gradient [12] is a convenient way to extract features from images of all kinds, and it is the most popular way in image processing. Using the histogram method in a different types of images produces multiple feature vectors and this will contribute to the success of detection and classification operations. By applying a histogram of gradient on depth image produces a histogram of oriented depth (HOD) [3]. In [4], the author tackles the problem of extracting sharp edges of colour_depth images by utilizing both depth and colour information that produce more coherent and cleaner features lines. In this work, a new feature extraction method based on both HOG and HOD is proposed. By fusing them using the linear first-order equation as a weighted sum that depends on the type of applications needed. This work differs from the research done by how the two feature vectors are added. Unlike [1, 4, 5, 6, 8, 11, 15] this work shows a new method for RGB-D sensors which has many scientific, industrial and research applications, in addition to possessing high-resolution technology in the field of image processing that enables the computer to see like a human, the working principle of the RGB-D camera depends on stereovision [10]. RGB-d has a conventional RGB (Red Green Blue), an infrared (IR) sensor and an infrared laser projector. the lenses and sensors are functioned together to make it possible to calculate the depth by detecting the infrared beam reflected from the object that is in front of the camera [13]. Two images are captured at the same moment to the same field of view (FOV).

## 2. Methodology

HOG is used to extract feature vectors from RGB images. The original $[64 \times 64]$ cells size in the HOG, the length of the feature vector is important factor when describing images. Short feature vector leads to loss of essential properties while the redundant points in feature vectors increase the rate of error and computational time [9].

HOG detects edges, corners, curves, and colour contrast in RGB images. It should be noted that the color of RBG images is real color of (FOV) while color in depth images represents a metric distance of RGB-D sensor as depicted in Fig.1. To test which type of vectors characterizes the (FOV) better, we built a human action recognition system with SVM classifier [2] with a dataset from RGB-D camera. After all, combining the HOG with HOD in a specific mathematical equation (2.1). The result is a new feature extraction method that performs better as shown in the results section.

Since the RGB-D sensor gives two matrices one for color image and the other for depth information. Histogram of the oriented gradient is performed on the color image to produce first feature vector $a$. The Histogram of oriented depth is performed on the depth matrix to produce the second feature vector $b$.

$$c = \alpha a + (1 - \alpha) b \tag{2.1}$$

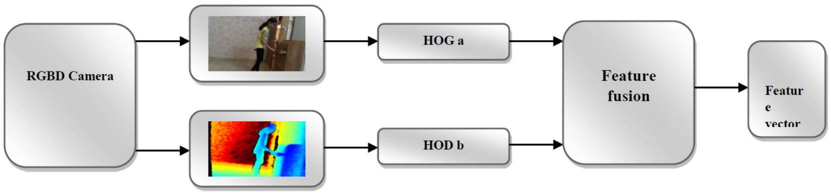


Figure 1: The proposed system for extracting features from RGB-D sensor

Where $c$ is the proposed vector that is used to train the classifier. In the test phase, $c$ is generated from the test image for the classification phase. Where $\alpha$ is the number value between 0 and 1, tuned to get the best feature vector that combines both HOG and HOD. Then the extracted new feature vector is tested by the human action recognition system. The results show that the proposed feature vector improved the classifier rate by around 5%. The system is shown in Fig.2.

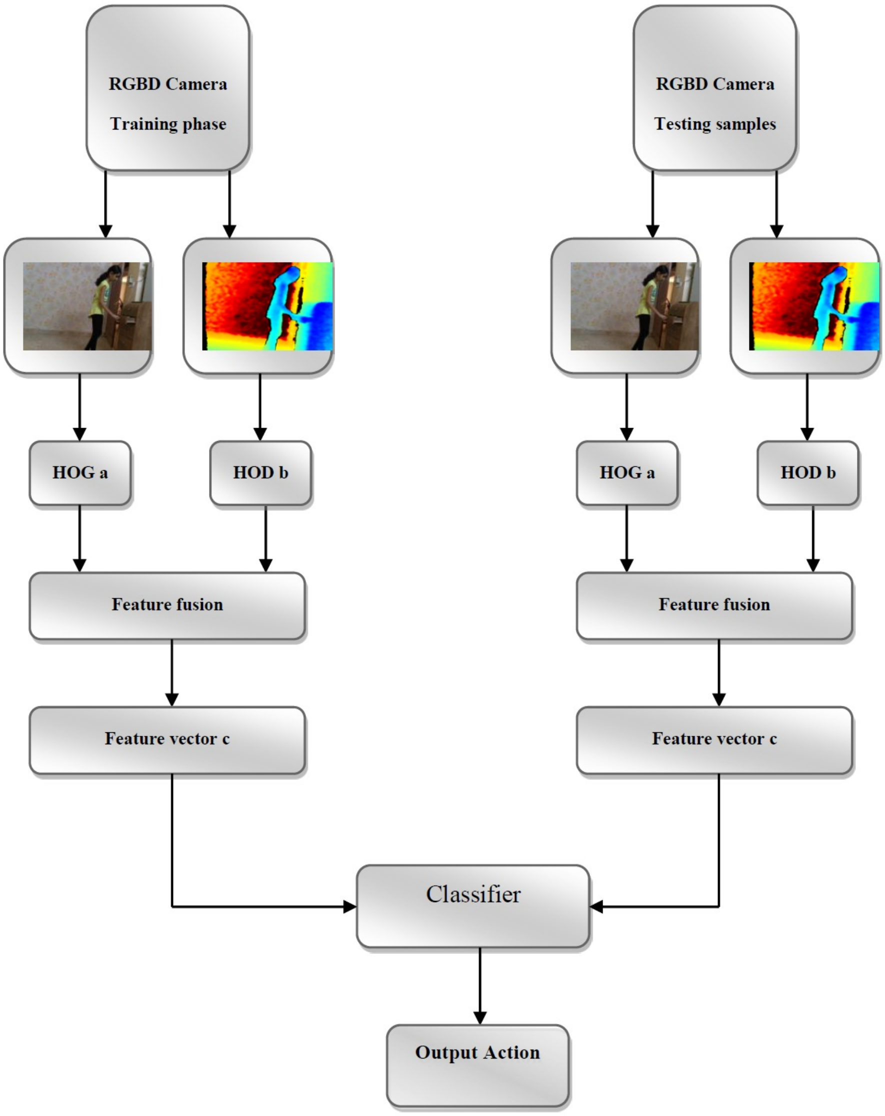


Figure 2: The human action recognition system used to benchmark the proposed feature extraction.

Table 1: SVM Classifier

| SVM Classifier | | | |
|---|---|---|---|
| HOG | | HOD | |
| Correct actions | Success rate | Correct actions | Success rate |
| 22.24 | %91.66 | 22.24 | %91.66 |

Table 2: Details about $\alpha HOG + (1-\alpha)HOD$

| $\alpha HOG + (1-\alpha)HOD$ | | |
|---|---|---|
| SVM Classifier | | |
| $\alpha$ | Correct actions | Success rate |
| 0.01 | 22.24 | %91.66 |
| 0.2 | 22.24 | %91.66 |
| 0.5 | 23.24 | %95.8 |
| 0.8 | 23.24 | %95.8 |
| 0.99 | 22.24 | %91.66 |

## 3. Simulation and Results

Intel Real sense RGBD camera d415 is used for the proposed system. The data set has 120 colour images and 120 depth images so the total dataset images is 240. As scaling law of splitting data the 48 samples images to 240 dataset images is appropriate [7], by setting the resolution of Intel real sense camera to 640 widths and 480 height during capturing. The resolution of all images is $640 \times 480$ pixels. Most of the relating works reprocess the images into $(128 \times 64)$ pixel [12] but we find that dividing the original images into $[64 \times 64]$ cells size in the HOG algorithm gives a better result and faster process.

Figure 3: Intel Real Sense camera d415

The dataset is four groups of actions (walking, running, clamping and waving) each action has 30 depth images and 30 colour images. So the total number of dataset images is 240. The images below Fig. 4 show samples of the images taken during the training process.

## 4. Conclusion

Feature extraction is essential in image processing and computer vision. Herein, a new feature extraction method for RGB-D sensors is proposed and benchmarked by SVM and human action recognition. The results show that the classifier gives a more accurate success rate after feature fusion which is depend on eq. (2.1) with feature vector length of both HOG and HOD is $[120 \times 1944]$ while that's length remains constant after a fusion because the modified feature vector length is $[120 \times 1944]$ also, that means the feature fusion improved the quality of feature vector points without increased the RGB-D feature vector length, the SVM classifier start gives the best result when the

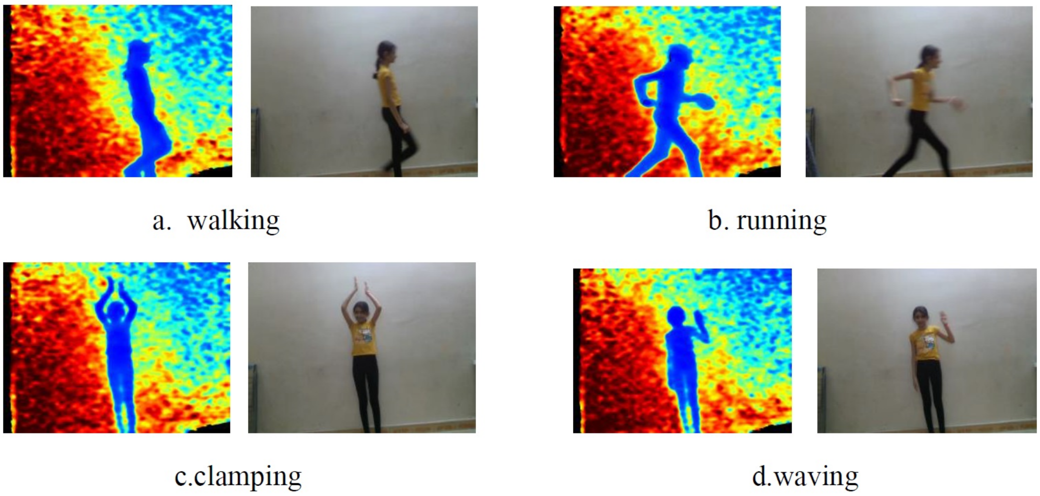


Figure 4: Samples of the images captured during the training process (a. walking b. running c. clamping d. waving)

$\alpha$ is 0.5. The feature fusion method produces a modified feature vector that has all the benefits of HOG and HOD of around 5% more than HOG.

## References

[1] E. Arıcan and T. Aydın, *Object detection with RGB-D data using depth oriented gradients*, B. Proc. Int. Conf. Eng. Nat. Sci. (2017) 876-–879.

[2] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress 2015.

[3] N.A. Bakr and J. Crowley, *Histogram of oriented depth gradients for action recognition*, arXiv preprint arXiv:1801.09477, (2018).

[4] Y.P. Cao, T. Ju, J. Xu and S.M. Hu, *Extracting sharp features from RGB-D images*, Comput. Graph. Forum 36(8) (2017) 138-–152.

[5] J. Chen, Y. Zhang and Y. Jiang, *Multi-features fusion classification method for texture image*, J. Eng. 2019(23) (2019) 8834-–8838.

[6] B. Fernando, E. Fromont, D. Muselet and M. Sebban, *Discriminative feature fusion for image classification*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2012) 3434—3441.

[7] I. Guyon, *A scaling law for the validation-set training-set size ratio*, AT&T Bell Lab. 1(11) (1997) 1—11.

[8] M.I. Khedher, M.A. El-Yacoubi and B. Dorizzi, *Human action recognition using continuous hmms and hog/hof silhouette representation*, ICPRAM 2012 - Proc. 1st Int. Conf. Pattern Recognit. Appl. Methods, 2 (2012) 503—508.

[9] C.Q. Lai and S.S. Teoh, *An efficient method of HOG feature extraction using selective histogram bin and PCA feature reduction*, Adv. Electr. Comput. Eng. 16(4) (2016) 101-–108.

[10] F. Lourenço and H. Araujo, *Intel realsense SR305, D415 and L515: Experimental evaluation and comparison of depth estimation*, VISIGRAPP 4(VISAPP) (2021) 362–369.

[11] D.U.C. Pati, P.K. Dutta and A. Barua, *Feature detection of an object by image fusion*, Int. J. Comput. Appl. 1(17) (2010) 59-–67.

[12] T. Surasak, I. Takahiro, C.H. Cheng, C.E. Wang and P.Y. Sheng, *Histogram of oriented gradients for human detection in video*, Proc. 2018 5th Int. Conf. Bus. Ind. Res. Smart Technol. Next Gener. Info. Eng. Bus. Soc. Sci. ICBIR 2018 (2018) 172-–176.

[13] V. Tadic, A. Odry, I. Kecskes, E. Burkus, Z. Kiraly and P. Odry, *Application of intel realsense cameras for depth image generation in robotics*, Wseas Trans. Comput. 18 (2019) 2224–2872.

[14] U. Thakur, S. Rai and S.K. Sahu, *A study an image fusion for the pixel level and feature based techniques*, Adv. Comput. Sci. Technol. 10(10) (2017) 3047—3055.

[15] Y. Wang, B. Song, P. Zhang, N. Xin and G. Cao, *A fast feature fusion algorithm in image classification for cyber physical systems*, IEEE Access 5 (2017) 9089—9098.