Journal of Multidisciplinary Engineering Science Studies (JMESS)
ISSN: 2458-925X
Vol. 3 Issue 4, April - 2017

A Novel Approach For Feature Extraction From RGB-D Data

Sajid Ur Rehman
Experts Vision Engineering and Technology Innovations
Riyadh, Saudi Arabia
expertsvision@live.com

www.eveati.com

Hafsa Asad
Experts Vision Engineering and Technology Innovations
Islamabad, Pakistan
hfsasad.eveati@gmail.com

www.eveati.com

***Abstract*— With the emergence of RGB- D image acquisition techniques, there has been a lot of research about various kinds of feature extractions that can be used for the purpose. In this paper, we propose a novel way of feature extraction for RGB-D images. Its strength can be gauged from the fact that it gives excellent results (95.4% on category recognition and 86% on instance recognition) without the use of any complex classification technique. Our proposed method of feature extraction, combines Local Binary Pattern, Wavelet transform and Color Auto Corellogram features from RGB data and Principal Component Analysis (PCA) features from its corresponding depth data. This is the first time that PCA has been used for extraction of depth features to give good results.**

Keywords— Principal Component Analysis; K Nearest Neigbor; RGB, LBP, RGB-D, Wavelet Transform, PCA, Color Auto Corellogram features from RGB data and Principal Component Analysis, feature extraction, classification, computer vision

I. INTRODUCTION

In today's digital age the size of data has increased massively. There are billions of photos and video clips floating on the websites like Flicker and YouTube. The numbers are even higher for Google Image database. In order to organize, retrieve or index this huge data, plenty of vigorous models and algorithms are proposed. These approaches aid in interacting with huge datasets. However, proper solution to organizing and utilizing such data is yet to be discovered. New technologies are emerging such as Kinect cameras [1], which are capable of capturing high quality videos. Kinect sensors use RGB color and depth channels to deliver good quality synchronized videos. The Microsoft Kinect [2] is one of the most advanced devices of today, having an amazing range of sensing abilities. This technology provides the opportunity to enhance the capabilities of the robots in terms of recognition, navigation and interaction with the surrounding environment.

The techniques discussed use two levels for evaluation of dataset.

- Category level Recognition: This type of recognition involves classification of different objects into their corresponding categories (e.g., phone, ball etc).

- Instance level Recognition: This type of recognition involves classification of different objects into their corresponding instances (e.g., phone of type 1, phone of type 2, red ball, green ball etc).

The instance and category level recognition are important in robotics or recognition systems where it is necessary to be able to differentiate between a general coffee mug and between various types of mug for example.

The dataset used here consists of 51 different items spread in the room. The images of each object are taken at different orientations or angles. Each object belongs to a category having subcategories known as instances. For example, 'phone' is a category and subcategories consist of phones of different brands or shapes.

The whole dataset is used for the categorizing the scenes and detecting the objects along with the classification of objects. However, if object classification is desired alone then 'evaluation database' [3] is used rather than the main database. This database consists of cropped RGB and depth images for all objects. These objects are organized in different folders and subfolders. Each object category is in a different folder and the subfolders contain the instances. We have 51 categories and 300 instances in total.

II. RGB DATSET AND SEGMENTATION

The RGB-D camera records the color as well as the depth images (640x480 resolution) simultaneously. That is the kind of data we have in the considered database. In RGB images, each pixel contains the information about the values of Red, Green and Blue color while the depth images holds the information about the depth of each pixel. In other words, it tells you the distance of each pixel's element from the sensor. To create depth images an invisible infrared light is continuously projected and then stereo tranquilization is performed by RGB-D camera. The active projection approach (particularly from the texture less regions) brings in a much more reliable depth reading as compared to the passive multi camera stereo

technology. Fig. 1 shows an RGB image and its corresponding depth image. The RGB-D camera driver software ensures the time-synchronous and alignment of RGB frame as well as depth frame.

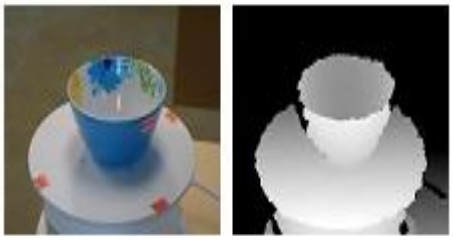

Fig. 1. RGB-D image data

The category hierarchy of some objects is represented in Fig. 2. Vegetable and fruits are in hierarchy at top level. In instrumentation category the man-made objects are added which includes the containers and devices etc.

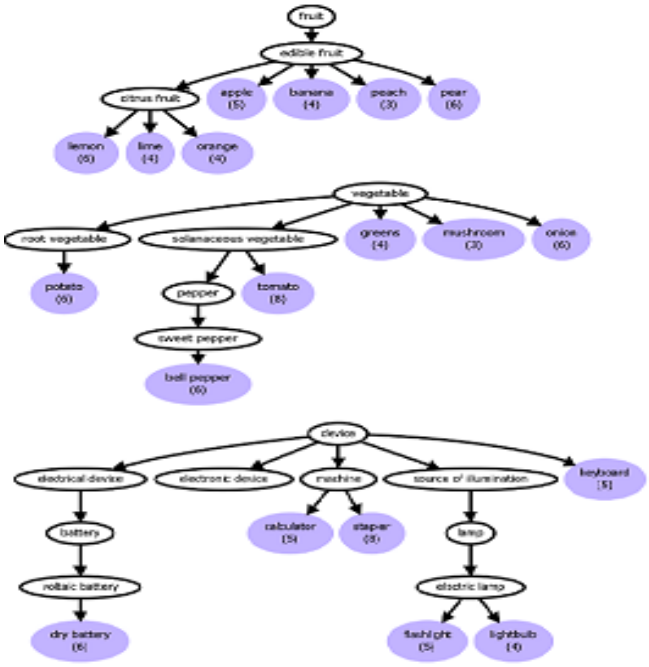

Fig. 2 Hierarchy diagram of some objects

In the hierarchy, there are a total of 51 leaf nodes that covers all the 300 instances. The number of instances for each category are given in leaf nodes by the help of parenthesis.

Fig. 3 Objects from RGB-D dataset

Some sample images from the dataset are shown in fig. 3.

The dataset also contains corresponding segmentation masks of each RGB image. Few samples of these are shown in Fig. 4.

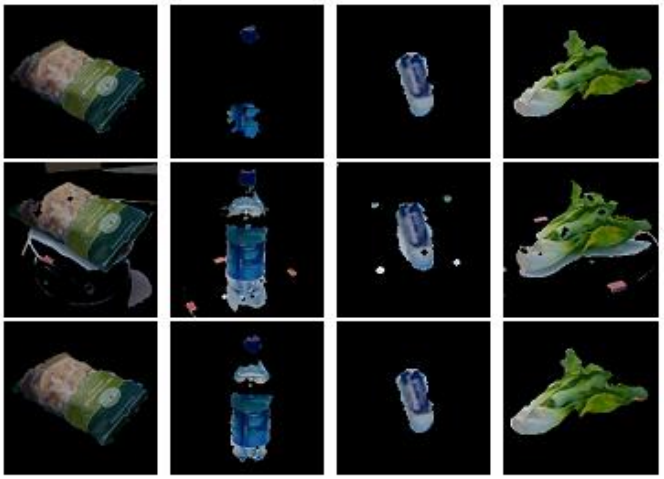

Fig. 4 Segmentation masks. From left to right bag of chips, water bottle, eraser and leaf vegetable.

In this evaluation dataset, for each sample, we have these kinds of images:

1. RGB image
2. Segmentation mask
3. Depth image

III. LITERATURE REVIEW

Each image in the dataset represents a particular view of an item. In literature, two types of features are usually required for classification. First set of features contains the information related to the shape and the second feature set consists of visual appearance of the

object. The shape retrieval community provides the most up-to-date features such as spin images in [3] and computer vision community provides SIFT descriptors [4]. Efficient Match Kernel (EMK) features are extracted by taking set of local features and then producing a fixed length feature vector. This operation is similar to bag of words (BOW) feature extraction. A continuous similarity measure is obtained by approximating the Gaussian kernel between the local features set. Random Fourier sets are used on spin images [5] for computing EMK features. The linear support vector machine (LibSVM), Gaussian kernel support vector machine (KSVM) [6] and random forest (RF) [7] are used for evaluating the performance of classification. In evaluation of these objects in category and instance classification the visual features are more suitable as compared to the shape features. However, shape features play important role in category classification and visual features are more useful in instance classification. In [8] a sliding window technique is used which evaluates the score function for all scales and positions for a particular image. The estimated scores are then used for getting a bounding box by threshold. In [9] a background subtraction algorithm is implemented in OpenCV library. Each pixel in the frame of the video is updated with mixture of k Gaussians. Each pixel is categorized as foreground if the value of σ is outside the standard deviations from all Gaussian mixtures. In [10], an alignment based on shape is done using ICP or Iterative Closest Point and error metric known as point-to-plane. Henry in [11] presented a system for handling circumstances in which alignment is not possible. Loop closure is done by 3D SIFT by comparing the frames with set of previous frames. Furthermore, pose graph optimization tool [12] is used for getting continuous alignments using TORO. The whole scene is made by small colored surface patches known as surfels [13]. This illustration allows for effective reasoning about obstructions and color for each fragment of location and resulting model has good visuals.

IV. METHODOLGY

Basic steps in any object classification method are the same:

- Image is obtained
- Features are extracted
- Image is assigned a label depending upon resemblance with training data samples

In our case, there are two target sets for complete data. One target set represents the instance labels while the second target set represents the category labels.

The strength of our proposed method, lies in its inherent simplicity. For feature extraction in RGB images we used a normalized combination of three feature extraction techniques:

1. Color Auto Correlogram
2. Wavelet Transform
3. Local Binary Pattern (LBP)

For depth images, we used Principal Component Analysis (PCA).

The assignment of targets can be explained by the example below:

TABLE I: ASSIGNMENT OF TARGETS

Sample Number (Feature Vector)	Targets	
	Instance	***Categories***
Sample Number 1	1	1
Sample Number 2	1	1
Sample Number 3	2	1
Sample Number 4	2	1
Sample Number 5	3	1
Sample Number 6	3	1
Sample Number 7	4	2
Sample Number 8	4	2
Sample Number 9	5	2
Sample Number 10	5	2
Sample Number 11	6	2
Sample Number 12	6	2

It can be seen from the above table there are two targets for each sample. From sample 1 to 12 there are two categories and 6 instance.

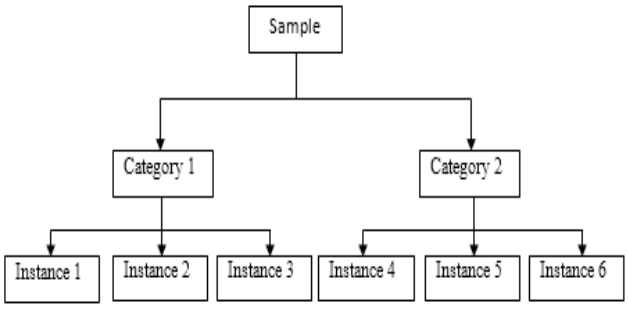

Fig. 5 Category flowchart

V. RESULTS

TABLE II ACCURACIES OBTAINED ON CATEGORY RECOGNITION

	Classification type	Accuracy (%)
1	RGB Category	90.4
3	Depth Category	82.8
5	RGB- D Category	95.41

TABLE III ACCURACIES OBTAINED ON INSTANCE RECOGNITION

	Classification type	Accuracy (%)
2	RGB Instances	83.8
4	Depth Instances	57.6
6	RGB- D Instances	85.9

Classification is done six times by varying the features used (RGB/ Depth/ RGB- D) and labels considered (category labels/ instance labels). For RGB-D, we simply concatenated the feature vectors of the corresponding RGB and depth data. We randomly selected 70% of total samples to train our classifier and remaining 30% to test it. Every time we run the program, accuracies obtained can differ by +-3% because of the random 70/ 30 division. KNN classifier is used for classification. Number of neighbors considered is 1. Lowest accuracy is 57.6% when we consider instance classification on depth data. It can easily be accounted for, because all instances of a category are almost exactly same in shape (and hence, depth features), only color varies.

It can be seen from above tables that accuracy is best when features of RGB image and depth image are combined to form a single, more powerful feature vector.

VI. CONCLUSION

The main contribution of this research is to propose a new feature descriptor for RGB- D images. The said feature descriptor combines LBP, Wavelet transform and color auto corellogram features from RGB data and PCA features from its corresponding depth data. This is the first time that PCA has been used on depth data for object recognition, and so the feature descriptor we proposed is unique to the best of our knowledge. We tested our method on the simplest available classifier to prove strength of our proposed feature extraction. As a future extension of this work, different other classifiers can be tested and compared. On RGB- D category recognition we get 95.4% accuracy, and on instance recognition we get 85.9%

The accuracy attained by combination of RGB-D information is more than that of individual classifications on RGB and depth data alone, hence proving that the combination of features considered is an excellent choice to represent RGB- D data.

REFERENCES

[1] [Online]. Available: PrimeSense. http://www.primesense.com.

[2] "Microsoft Kinect," [Online]. Available: http://www.xbox.com/en-us/kinect.

[3] [Online]. Available: http://rgbd-dataset.cs.washington.edu/dataset/rgbd-dataset_eval/.

[4] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1999.

[5] G. David and Lowe, "Object recognition from local scale-invariant features (ICCV)," in *IEEE International Conference on Computer Vision*, 1999.

[6] L. Bo and C. Sminchisescu, "Efficient match kernel between sets of features for visual recognition," in *Advances in neural information processing systems*, 2009.

[7] C.-c. Chang and C.-j. Lin, ""LIBSVM: a library for support vector machines," 2001," 2012. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[8] L. Breiman, "Random forests. Machine Learning," 2001.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human".

[10] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection.," in *European Workshop on Advanced Video Based Surveillance Systems*, 2001.

[11] Chen and M. Gerard, "Object modelling by registration of multiple range images," in *Image Vision Computer*, 1992.

[12] P. Henry, M. Krainin, Herbst, R. E and D. Fox, "RGB-D Mapping: Using depth cameras for dense 3D modeling of indoor environments," in *12th International Symposium on Experimental Robotics (ISER)*, 2010.

[13] G. Grisetti, S. Grzonka, C. Stachniss and Burgard, "Estimation of accurate maximum likelihood maps in 3d," in *International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[14] H. Pfister, M. Zwicker, J. van Baar and G. Surfels, "Surface elements as rendering primitives," in *ACM Transactions on Graphics*, 2000.