INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

EFFICIENT CODEBOOK FOR HUMAN ACTIVITY RECOGNITION IN SURVEILLANCE VIDEO

S. Kiruthiga[1], M. Kalaiselvi Geetha[2], J. Arunnehru[3]

[1]M. E. Student, Dept of Computer Science, Annamalai University, India.
[2]Associate Professor, Dept of Computer Science, Annamalai University, India.
[3]Research Scholar, Dept of Computer Science, Annamalai University, India.
[1]cskirthi946@gmail.com, [2]geesiv@gmail.com, [3]arunnehru.aucse@gmail.co

ABSTRACT—*Automatic human activity recognition methods are useful for many applications such as Video Surveillance, Video Annotation and Retrieval and Human Computer Interaction. Bag of Visual Words (BoVW) has become an important framework in many computer vision areas, due to their robustness and simplicity, as well as reported excellent performance on visual recognition tasks. The basic idea is to construct a visual codebook on the statistics of the various features in videos, so the first step is to extract feature from videos. As a pre-processing step, frame differencing is done using Weizmann dataset, to convert the video into frames, and then feature is extracted from those frames, in which feature extraction based on the intensity values of those frames obtained after frame differencing. After extracting the features, Vector Quantization method is used to generate the codebook. k-means clustering algorithm is used, using which for each action; the feature vectors are clustered to form a codebook, whereas each action cluster is called a codeword. All action clusters comprise to form a group of codewords called as Bag of Words (BoW).Using those codeword, action recognition task is performed. In this work, we focus towards the construction of a codebook of compact size, to recognize human actions.*

Keywords— *Video Surveillance, Action Recognition, Bag of Visual Words (BoVW), frame differencing, feature extraction, Vector Quantization, k-means clustering, codewords, codebook.*

1. INTRODUCTION

Video surveillance is the process of monitoring and recognizing the ongoing activities in surveillance video for the purpose of safety. Video surveillance is largely developing due to both the increasing population, especially in cities, and the exploding number of video surveillance cameras deployed. It is a complex process to recognize human action due to many factors such as variation in speed, postures and clothing. Codebook generation is one of the most important and inevitable step that has to be done efficiently for automatic human activity recognition. The basic idea is to construct a visual codebook, by extracting features from videos and comparing the features of various types of activities. The extracted features are generated as codewords, using the *k*-means clustering technique, for each action using the clustering technique, codewords are generated. The codewords of each action comprise to form a Bag of Words (BoW) model, using which codebook is formed.

1.1 RELATED WORK

Novel and effective solution to classify human actions in unconstrained videos have been proposed by Lamberto Ballan et al [1].It improves on previous contributions through the definition of a local descriptor, for codebook formation radius based clustering with soft assignment is done in order to create a rich vocabulary that may account for the high variability of human actions. Qingdi Wei et al [4] propose a novel method to address these issues by constructing a compact but, effective visual codebook using sparse reconstruction. In it a large codebook generated by *k*-means, they formulate it in a sparse manner and calculate the weight of each codebook using original visual codebook. Mingyan Jiu et al [8] presents a novel approach for supervised codebook learning and optimization for BOW models. This type is frequently used in visual recognition tasks like object class recognition and it proposes a new supervised method for joint codebook creation and class learning which learns the cluster centers of the codebook in a goal directed way using the class labels of the training set. Liu Yang et al [6] propose a novel optimization framework that unifies codebook generation with classifier training. It uses each of the image features to encode by a sequence of "Visual bits" optimized for each category.

1.2 OVERVIEW OF THE PROPOSED APPROACH

This paper deals with construction of effective codebook for activity recognition in video surveillance. This approach is evaluated using Weizmann dataset with 9 persons performing 10 actions .Frame difference and motion interest area alone is extracted as one of the preprocessing steps for feature extraction. In the codebook generation process, various images are divided into several *k i.e.,7*-dimension training vectors. The representative codebook is generated from these training vectors by the clustering techniques. *k*-means clustering is used to generate mean clusters from the feature vector, called as codewords and these codewords are grouped together to form a BoW(Bag of Words) model. *K-means* testing is performed on this codebook using Euclidean distance measure ,to find corresponding action based on the minimum value obtained during testing.

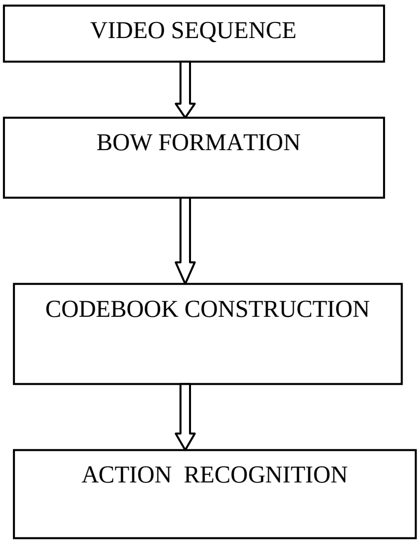


Fig 1.Process involved in codebook generation

2. FEATURE EXTRACTION

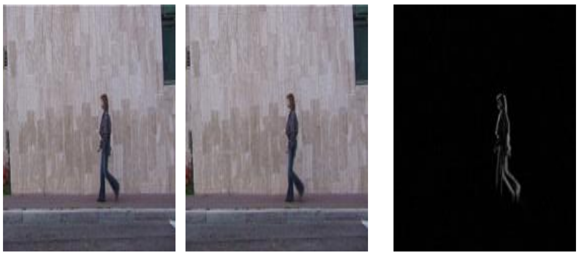
Frame differencing is defined by the differences between successive frames in time. The frame subtraction method considers every pair of frames of time t and t + 1, to extract any motion information in it. In order to locate the motion interest area, the current frame is subtracted with previous frame on a pixel by pixel basis,

The frame difference at time t is given by:

$$D_t(x,y)=|I_t(x,y)– I_{t+1}(x,y)| \qquad (1)$$

$$1 \leq x \leq w,\ 1 \leq y \leq h$$

$I_t(x,y)$ is the intensity of the pixel (x, y) in the kth frame, h and w are the width and height of the image respectively.



Frame t Frame t+1 Frame Difference

fig 2. Frame Difference

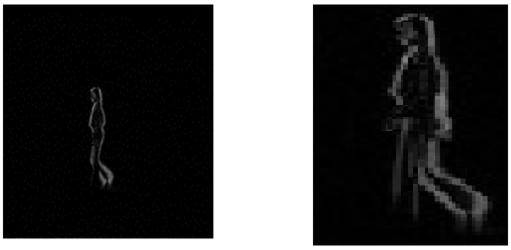
Motion is an important cue in action recognition research. This work extracts the motion information from the equation 2.

Motion information T_k or difference image is calculated using:

$$T_{k(i,j)} = \begin{cases} 1, & if\ D_{k(i,j)} > t; \\ 0, & otherwise; \end{cases} \qquad (2)$$

where t is the threshold.

The value of t = 30 has been used in the experiments. To capture the dynamic information, motion is extracted from the difference image D_t as in Eq. 1.



a) Before extracting ROI b) After extracting ROI

fig 3. Motion Area Extraction

The motion identified area is considered as Region of Interest (ROI).This ROI is extracted and only this is used for further analysis. To reduce the computational complexity, the ROI is further divided into two blocks as seen in fig.4.Further among these two blocks the maximum motion identified block is used for the extraction of the features. The block is further divided into 5x5 blocks and average intensity of each block is obtained. Thus generating a 25 dimensional feature vector.

In the proposed method the frame is divided into two blocks (35x49) and based on the maximum motion region, the block from which feature has to extracted is done. For example if the action considered is like walking and running, then feature is extracted only from block2 and if it is actions like wave1 and wave2 then feature extraction is done based on block1.

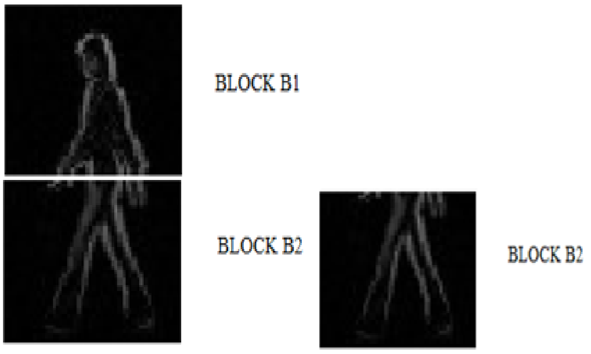


Fig 4. Feature Extraction

3. CODEBOOK GENERATION

3.1 VECTOR QUANTIZATION

Vector Quantization (VQ) is an efficient and simple approach for data compression, since it is very simple and easy to implement. For the purpose of compression of the images, the procedural operations of VQ include dividing an image into several vectors (or blocks) and each vector is mapped to the code words of a codebook to find its vectors reproduced. The main goal of VQ is the representation of vectors X ⊆ R_k by a set of reference vectors CB = {C1; C2; : : : ;CN} in R_k in which R_k is the k-dimension Euclidean space. The total number of codewords in CB is N and the number of dimensions of each codeword is k. There are three important steps carried out in VQ, they are firstly generation of codebook, encoding and decoding procedure carried out using the codebook generated.

3.1.1 *k*-MEANS CLUSTERING ALGORITHM

The features extracted during feature extraction phase can be classified (or) grouped based on *k*-means algorithm into k number of groups, where k is any positive integer. By minimizing the sum of squares of distances between data and the corresponding cluster centroid the grouping can be done. Thus, the main goal of *k*-mean clustering is to classify the data or feature extracted**.** The aim of clustering is to partition a set of objects which have associated multi-dimensional attribute vectors into homogeneous groups such that the patterns within each group are similar.

It uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters: Euclidean distance equation is used to perform centroid calculation, where distance d is given by

$$d=\sqrt{(y1-x1)^2+(y2-x2)^2} \qquad (3)$$

Where,

d->Euclidean distance between two centroids.
(x1,y1)->1st cluster point
(x2,y2)->2nd cluster point

The batch updates are used in first phase, where each iteration consists of reassigning points to their nearest cluster centroid, all at once, ensued by recalculation of cluster centroid.

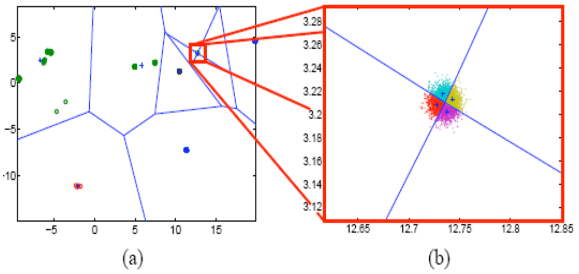


Fig 5.a)*k*-means clustering **fig 5.b)detail of a splitted dense region into 4.**

Steps:

A dataset is taken and to obtain the required number of clusters k, a set of k initial starting points is taken and k means clustering algorithm finds the desired number of distinct clusters and their centroids**.** A centroid is a point whose coordinates are obtained by computing the average of each of the coordinates of the points of the samples assigned to each cluster. Formally, the *k*-means clustering algorithm follows the following steps.

1. *Set k*: Choose a number of desired clusters, k.
2. *Initialization*: Choose k starting points to be used as initial estimates of the cluster centroids. These are the initial starting values.
3. *Classification*: Examine each point in the data set and assign it to the cluster whose centroid is nearest to it.
4.*Centroid Calculation*: Recalculate the new k centroids, when each point is assigned to a cluster.
5. *Convergence condition*: Repeat steps 3 and 4 until no point changes its cluster assignment, or until a maximum number of passes through the data set is performed. Before the clustering algorithm can be applied, actual data samples are collected. The features that describe each data sample in the database are required in advance.

First, extract feature vectors in the action video; construct the vocabulary of codewords using *k*-means clustering. This is also called feature or vector quantization. Finally, an action video is represented by the mean clusters called as code-words. These codewords are grouped together to form Bag of Words (BoW). With the codewords, action recognition or indexing can be conducted.

4. EXPERIMENTAL RESULTS

This approach is tested on datasets commonly used for human action recognition: i.e., Weizmann datasets. The Weizmann dataset contains 93 video sequences showing nine people who are different from each other and, each performing ten actions such as run, walk, gallop sideways, wave-two-hands, wave-one-hand, skip, jumping-jack, jump forward- on-two-legs, jump-in-place-on-two-legs, and bend. The video resolution is

180x144 pixels. Using *K*-means clustering algorithm training and testing is carried out.

After performing *k*-means clustering on feature vector, mean clusters are obtained. In Weizmann dataset totally 9 persons perform 10 actions out of which we consider only 5 actions like walk,run,jump,wave1,wave2 and 6 persons are taken for training and 3 persons are taken for testing. Using One Versus All rule (each test file of one action is compared with all other train files), Euclidean distance measure is employed, to carry out *k*-means testing, in which clusters of one testing cluster is compared with all other five training clusters and as a result minimum value is obtained for the corresponding action alone**,** like for action walk alone as shown in fig 6.This step is used in most of the recent works and it is most simple and suitable for direct comparison.

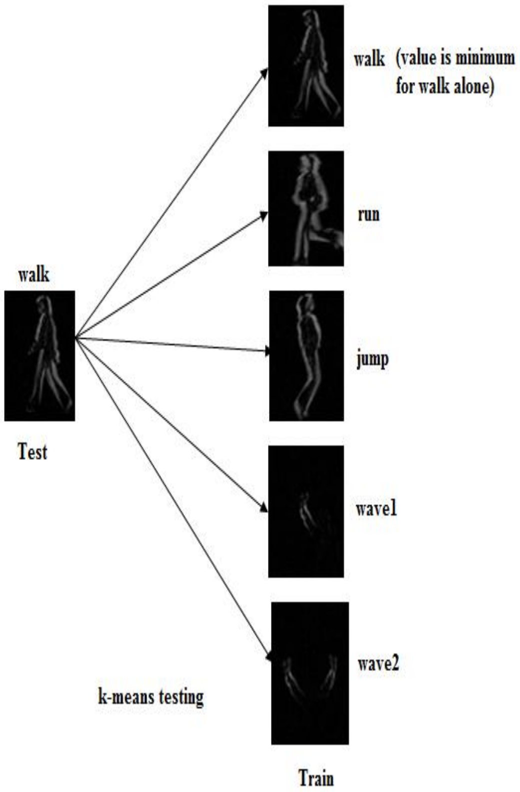


Fig 6.Testing the performance of codebook

Table 1. Average of values obtained during *k*-means testing(for various cluster sizes)

	WALK	RUN	JUMP	WAVE1	WAVE2
WALK	**0.0306**	0.0942	0.1112	0.0891	0.0760
RUN	0.0851	**0.0314**	0.1352	0.0579	0.0660
JUMP	0.0857	0.0552	**0.0113**	0.0549	0.1050
WAVE1	0.1020	0.0888	0.0801	**0.0257**	0.1611
WAVE2	0.0413	0.0692	0.0885	0.0831	**0.0247**

PERFORMANCE OBTAINED BY EFFECTIVE CODEBOOK:

In the set of experiments we use Weizmann dataset, The resolution of Weizmann dataset video is 180 × 144 pixel. The classification performance is obtained by the standard K-means approach, in which testing is carried out using One Versus All rule. Performance is measured using no of words (n) and dissimilarity measure(value obtained during testing in *k*-means testing). The proposed approach was tested on it, which gives results that outperform the other BoW approaches.

$$Accuracy = \frac{\boldsymbol{number\ of\ words}(\boldsymbol{n})}{\boldsymbol{dissimilarity\ measure}(\boldsymbol{M})}$$

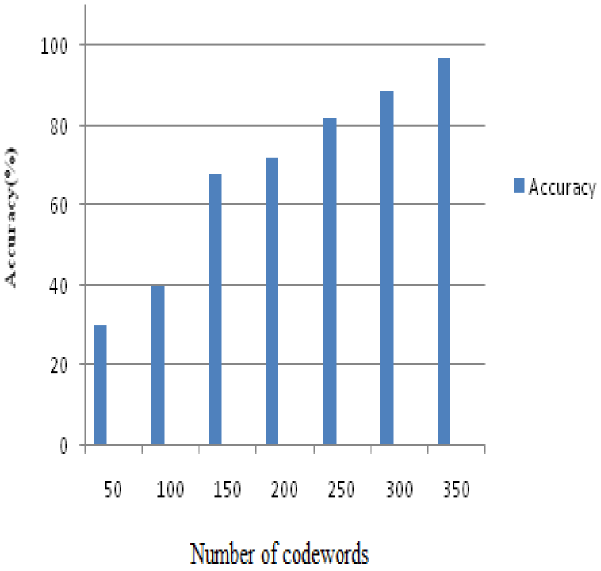


Fig 7. Performance measure

The graph reports the variation in accuracy w.r.t the number of codewords. If the number of codewords increases, then accuracy also increases, overall accuracy = 97%.

5. CONCLUSION

This paper presents, a novel method for human action recognition using the codebook constructed. The codebook is generated by extracting visual features from videos, and by using those visual features action is recognized. As a preprocessing step before extracting visual features from videos, the video is first converted into frames using frame differencing. The codebook consist of several codewords, for each action codeword is generated by using *k*-means clustering algorithm and as the number of clusters increase the performance also increases, since *k*-means clustering is an unsupervised method, training and testing is done using Weizmann dataset which yields very good performance.

REFERENCES

[1] Lamberto Ballan, Marco Berti, Alberto Del Bimbo, Lorenzo Seidenari, and Giuseppe Serra, "Effective Codebooks for Human Action Representation and Classification in Unconstrained Videos", IEEE Transactions on Multimedia,2012.

[2] Lingqiao Liu, Lei Wang, and Chunhua Shen, "A Generalized Probabilistic Framework for Compact Codebook Creation ", IEEE Conf.Comp.Vis.Pattern Recogn.2011.

[3] Lamberto Ballan, Marco Bertini, Alberto Del Bimbo, Lorenzo Seidenari and Giuseppe Serra, "Effective Codebooks for Human Action Categorization", Media Integration and Communication Center, University of Florence, 2009.

[4] Qingdi Wei, Xiaoqin Zhang, Yu Kong, Weiming Hu, and Haibin Ling," "Compact Visual Codebook for Action Recognition", National laboratory of Pattern Recognition, Institute of Automation,2010.

[5] M. S. Ryoo, "Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos", IEEE International Conference on Computer Vision(ICCV),Nov 2011.

[6] Liu Yang Rong Jin, Rahul Sukthankar,and Frederic Jurie," "Unifying Discriminative Visual Codebook Generation with Classifier Training for Object Category Recognition" IJCV(International Journal of Computer Vision),2008.

[7] Yang Wang, *Student Member IEEE*, and Greg Mori, *Member IEEE*, "Human Action Recognition by Semi-Latent Topic Models", IEEE Transactions on Pattern Analysis and Machine Intelligence,2012.

[8] Mingyuan Jiu Christian Wolf , Christophe Garcia ,and Atilla Baskurt, "Supervised Learning and Codebook Optimization for Bag-of-Words Models", Springer Science+ business Media,LLC 2012.

[9] J. Arunnehru, M. Kalaiselvi Geetha, "Automatic Activity Recognition for Video Surveillance", International Journal of Computer Applications, (0975 - 8887) Volume 75 - No. 9, August 2013.

[10] Xingxing Wang, LiMin Wang,and Yu Qiao, "A Comparative Study of Encoding, Pooling and Normalization Methods for Action Recognition", National Natural Science Foundation of China,2012.

[11] L.Torres and J.Huguet,"An improvement on codebook search for vector quantization",IEEE Trans.Commun.,vol.42,No.3,pp.208-210,Feb.1994.

[12] C.Bei,R.M.Gray,"An improvement of the minimum distortion encoding algorithm for vector quantization", IEEE Trans. Commun., Vol.33, n0.10, pp.1132-1133,oct 1985.

[13] Annu and Dr. Chander Kant," Liveness Detection in Face Recognition Using Euclidean Distances", International Journal For Advance Research In Engineering And Technology, Vol. 1, Issue IV, May 2013, ISSN 2320-6802.

[14] Tarun Dhar Diwan, Deepa Gupta , and Praveen Sahu," Application Of Gesture Recognition In Virtual Touch In Virtual Environments", International Journal For Advance Research In Engineering And Technology, Vol. 1, Issue VII, Issn 2320-6802, Aug 2013.