IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727
PP 37-43
www.iosrjournals.org

Visual Words for Human Activity Recognition in Surveillance Video

S.Kiruthiga[1], M.Kalaiselvi Geetha[2], J.Arunnehru[3]
[1,2,3]Department of Computer Science, Annamalai University, India

ABSTRACT : *Recognition and classification of human actions for the purpose of safety from video sequences is always a challenging problem because of the variations in its environment, different backgrounds used in videos, appearance of actors and their clothing, so in our work we propose a method for constructing effective and appropriate codebooks for action categorization. In the formation of codebook fuzzy C-means clustering algorithm and Pairwise Nearest Neighbor algorithm (PNN), is used and hence the performance of these methods are compared and analyzed on Weizmann dataset.*

Keywords - *Video Surveillance, Action Recognition, Bag of Visual Words (BoVW), frame differencing, feature extraction, Vector Quantization, fuzzy C-means clustering, Pairwise Nearest Neighbor (PNN),codewords, codebook.*

I. INTRODUCTION

Widespread use of video cameras, as powerful sensors for collecting visual data in surveillance systems, has provided the ability of monitoring wide areas. In addition, cameras provide information that is conceivable by human operators, and recognize the actions being performed by different persons on videos, in various background situations. By the extend of the number of cameras in surveillance systems, the role of computer vision for automating the process of extracting information from videos has become more important. With the continuous growth of video production and archiving, the need for automatic annotation tools that enable effective retrieval by content has accordingly gained increasing importance. In particular, action recognition is a very active research topic with many important applications such as human-computer interaction, video indexing and video-surveillance, human robot interaction, sports and video annotation. Existing approaches for human action recognition can be classified as using holistic or part-based information . Most of the holistic-based methods usually perform better in a controlled environment and are also computationally expensive due to the requirement of pre-processing the input data. In this work for recognizing the ongoing activities in video, codebook construction approach is proposed. In the formation of codebook, fuzzy C-means clustering and Pairwise Nearest Neighbor method is used, where fuzzy C-means clustering is based on fuzzy logic, and the data points may belong to more than one cluster, and associated with the points are membership grades. In Pairwise Nearest Neighbor algorithm the computation overhead is reduced since the number of category to be calculated is minimized based on the nearest neighbor found. Hence the performance the effective codebook using both the algorithms is compared and analyzed.

II. RELATED WORK

Novel codebook representation method is proposed by E. Shabaninia [1] for appearance modeling of moving vehicles. The modeling could cope with illumination changes of environment. It also was able to accommodate the different viewing angles of objects. In this model, instead of keeping the histogram bins of each object, the major colors of that object were preserved by using the related cylinders in the RGB space. A fast PNN-based multilevel thresholding algorithm is proposed by Olli Virmajoki [2],in which Pairwise Nearest neighbor is calculated with computational efficiency and the quality of codebook constructed is analyzed. A survey of fuzzy logic is done by Makhalova Elena [3] and it explains the theory applied in cluster analysis. Lamberto Ballan [4] et al proposes a new method for the formation of the codebook ,in which radius-based clustering with soft assignment is employed in order to create a rich vocabulary that may account for the high variability of human actions. It also show that the solution scores very good performance with no need of parameter tuning. A strong reduction of computation time can be obtained by applying codebook size reduction with Deep Belief Networks with little loss of accuracy. A fast variant of the exact PNN algorithm was introduced by Timo Kaukoranta et al [5]. The main idea of the algorithm is to maintain a table of nearest neighbors as in the PNN algorithm. In addition to it , postponing the updation of the closest distance information to the moment when the (old) distance becomes the new tentative minimum among the cluster distances is done.

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* dimension training vectors. The representative codebook is generated from these training vectors by the various vector quantization techniques. Fuzzy C-means clustering and Pairwise Nearest Neighbor method 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. Testing is performed on this codebook using Euclidean distance measure, to find corresponding action based on the minimum value obtained during testing.

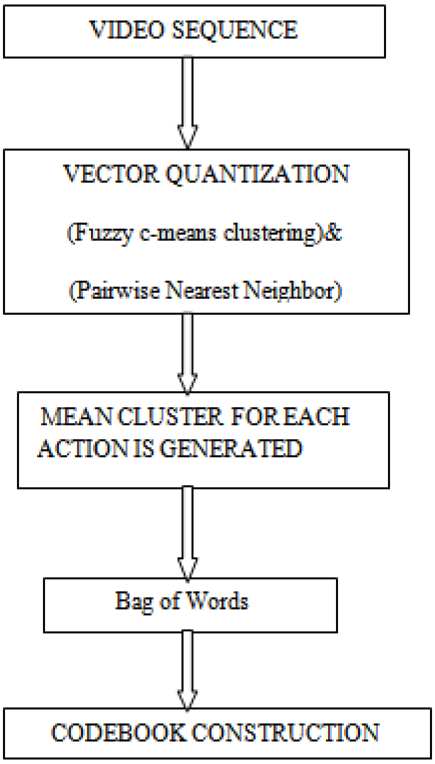
VIDEO SEQUENCE
VECTOR QUANTIZATION
(Fuzzy c-means clustering)&
(Pairwise Nearest Neighbor)
MEAN CLUSTER FOR EACH ACTION IS GENERATED
Bag of Words
CODEBOOK CONSTRUCTION

Fig 1. Process involved in codebook formation

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 k^th^ frame, h and w are the width and height of the image respectively.

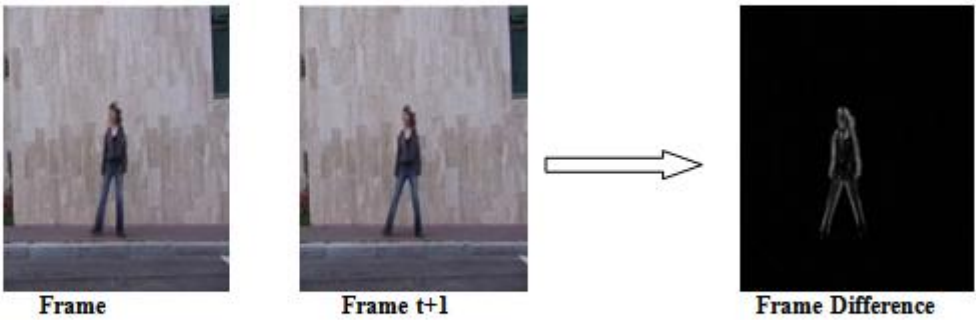

Fig 2. Frame Difference

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:

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

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 interest area extraction

III. 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, VQ has been widely used in variety of applications, such as pattern recognition of, compression of images, recognition of speech, and face detection. One of the key roles of Vector Quantization (VQ) is how to generate a good codebook such that the distortion between the original image and the reconstructed image is the minimum. Vector Quantization is lossy data compression technique and has various applications. 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. In other aspects, the main goal of VQ is the representation of vectors X ⊆ Rk by a set of reference vectors CB = {C1; C2; : : : ;CN} in R_k in which R_k is the *k*-dimension Euclidean space. CB represents a codebook which has a set of reproduction code words and Cj = {c1; c2; : : : ; ck} is the j-th codeword. The total number of codewords in CB is N and the number of dimensions of each codeword is k. In this work we propose two novel approaches for vector quantization, like fuzzy C-means clustering and Pairwise Nearest Neighbor method.

3.1.1 Fuzzy C-means clustering

It is based on fuzzy logic, in which data points may belong to more than one cluster, and associated with these points are the membership grades, which indicates the degree to which the data point may belong to different cluster. The membership grades are the points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of the cluster. It expresses how ambiguously or definitely a data point belong to a cluster, with the update of membership(M_{ij}) and cluster centers (c_j).Fuzzy partitioning is carried out through an iterative optimization of the objective function.

The algorithm is composed of the following steps:

1. Initialization

Select the following parameters:

- the required number of clusters *N*.
- measure Euclidean distance.
- fixed parameter q is assigned, Initial (at zero iteration) matrix $M^{(0)}=(c_i)^{(0)}$ object ownership x_i with the given initial cluster centers c_j .

2. Calculate the centers vectors $C(k)=[cj]$,with $M(k)$.

3.Modified membership measure M_{ij} is calculated.

4. If $\| M^{(k+1)}-M^{(k)}\| <$ □ then STOP; otherwise return to step 2.

The cluster centers of the train file and test file is compared using Euclidean distance measure and for the corresponding action category the minimum value is obtained**.**

3.1.2 Pairwise Nearest Neighbor Method

Pair wise nearest neighbour uses an opposite, bottom-up approach to the codebook generation. It is one of the simplest but widely used machine learning algorithms. An object is classified by the “distance” from its neighbors, with the object being assigned to the class most common among its k distance-nearest neighbors. If k = 1, the algorithm simply becomes nearest neighbor algorithm and the object is classified to the class of its nearest neighbor’s choosing appropriate k is an important factor. Distance is a key word in this algorithm, each object in the space is represented by position vectors in a multidimensional feature space. It uses Euclidean distance to calculate distance between two vector positions in the multidimensional space.

It starts by initializing a codebook where each training vector is considered as its own code vector. Two code vectors are merged in each step of the algorithm and the process is repeated until the codebook reduces to the desired size. A common solution is to calculate distance to all categories, like side, skip, Pjump, jack, bend and recognize the one with the smallest distance as action predicted.

Steps:

- ***Initialization***

  In the initialization phase, each training vector (T_i) is set as its own code vector (C_i), and the sizes of the clusters (n_i) are set to one. In order to generate the nearest neighbor table, we need to find nearest neighbor for every cluster. This is done by considering all other clusters as tentative neighbor and selecting the one that minimizes.
- ***Finding the Two Nearest Clusters***

  The clusters to be merged are the cluster pair minimizing.
- ***Merging the Clusters***

Merging of the two clusters causes changes to all data structures. The code vector of the combined cluster is the centroid of the training vectors in the cluster and it can be calculated as the weighted average of two nearest clusters.

- ***Updating the Nearest Neighbor Table***

  The clusters can be classified into two groups: 1) clusters whose nearest neighbor before merge and 2) all other clusters.
- **Test Results—**based on the maximum number of neighbors, the category of action to which it belong is recognized.

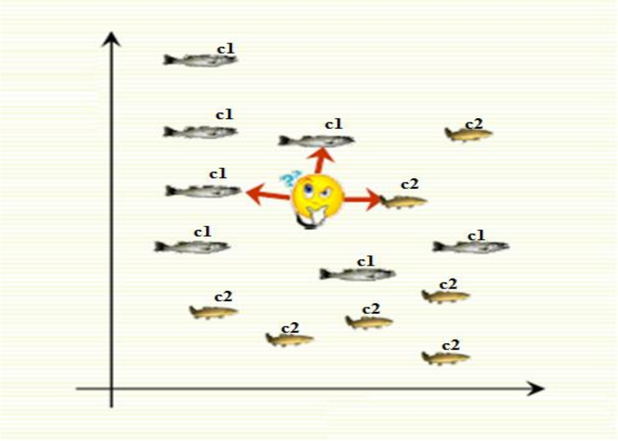

Fig 4 Classification based on nearest neighbor

In the above figure 4,C1 are the maximum neighbors to the object to be identified.
C2 is not much nearer, so it is discarded.
First, extract feature vectors in the action video; and then construct the vocabulary of codewords using fuzzy *C*-means clustering and Pairwise nearest neighbor method. 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.

IV. EXPERIMENTAL RESULTS

This approach is tested on datasets commonly used for human action recognition: like 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.

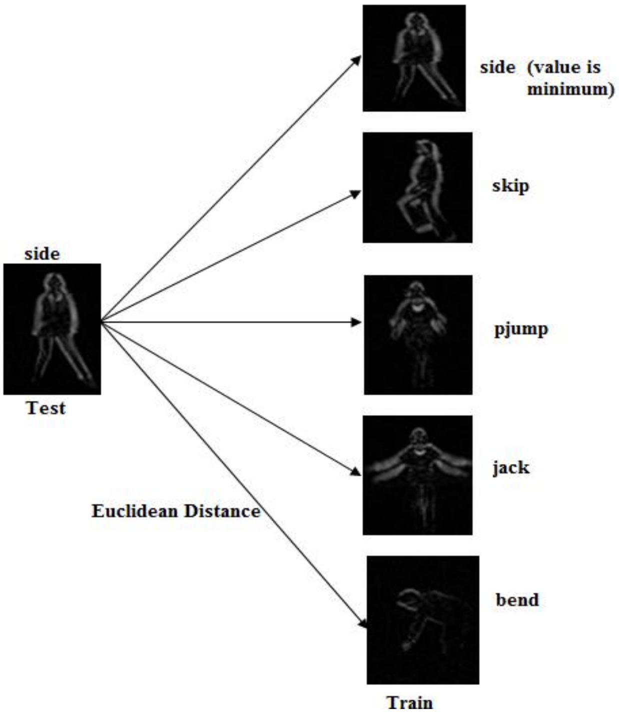

Fig 5.Testing the Performance of Codebook

After performing *Fuzzy C*-means clustering and Pairwise Nearest Neighbor on feature vector mean clusters and nearest neighbors are obtained. In Weizmann dataset totally 9 persons perform 10 actions out of which we consider only 5 actions like side, skip, Pjump, jack and bend and 9 persons are taken for training and 1 person is kept for testing. Using Euclidean distance measure, *fuzzy C*-means testing is carried out, in which clusters of one testing cluster is compared with all other five training clusters using One Versus All Rule and as a result minimum value is obtained for the corresponding action alone**,** like for action walk alone as shown in fig 5.and in Pairwise Nearest Neighbor method the minimum value is calculated based on the maximum number of neighbors identified during classification.

Table 1.Classification accuracy for fuzzy C-means testing(for various cluster sizes)

	Side(%)	Skip(%)	Pjump(%)	Jack(%)	Bend(%)
Side	**96.96**	96.92	96.91	96.91	96.95
Skip	96.92	**96.97**	92.84	96.83	96.95
Pjump	96.91	96.96	**96.99**	96.94	96.93
Jack	96.89	96.81	96.91	**96.97**	96.92
Bend	96.92	96.94	96.94	96.33	**96.98**

Table 2.Classification accuracy for Pairwise Nearest Neighbor (PNN) method (for various *k* values).

	Side(%)	Skip(%)	Pjump(%)	Jack(%)	Bend(%)
Side	**99.97**	99.87	99.95	99.94	99.77
Skip	99.90	**99.99**	99.82	99.77	99.98
Pjump	99.98	99.89	**99.99**	99.94	99.80
Jack	99.86	99.95	99.78	**99.98**	99.97
Bend	99.64	99.86	99.53	99.45	**99.89**

Performance Obtained By Effective Codebook

Performance is measured using no of clusters (*k*), and dissimilarity measure i.e., value obtained during testing in *fuzzy C*-means testing and nearest neighbor classification. The proposed approach was tested on it, which gives results that outperform the other BoW approaches.

$$\boldsymbol{Accuracy = \frac{number\ of\ clusters(k)}{dissimilarity\ measure(M)}}$$

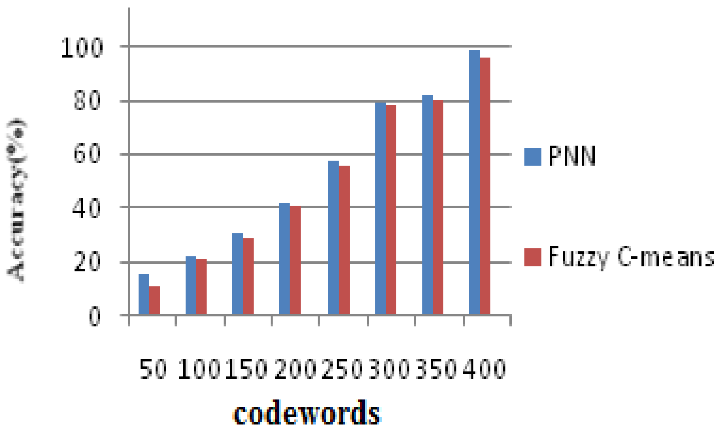

Fig 6.Performance measure

If the number of codewords increases, then accuracy also increases. Overall accuracy = 99% for PNN(Pairwise Nearest Neighbor)method and 96% for fuzzy C-means clustering algorithm.

V. 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 *fuzzy C-means* clustering algorithm and Pairwise Nearest Neighbor (PNN) method as the number of clusters increase the performance also increases, since *fuzzy C*-means clustering is an unsupervised method, training and testing is done using Weizmann dataset and for Pairwise nearest neighbor method the classification yields very good performance.

REFERENCES

[1] E. Shabaninia ,and Sh. Kasaei ,” Codebook appearance representation for vehicle handover across disjoint-view multicameras”, Scientia Iranica, Transactions D: Computer Science & Engineering ,Elsevier,(2011).

[2] Olli Virmajoki and Pasi Franti,” Fast pairwise nearest neighbor based algorithm for multilevel thresholding”, Journal of Electronic Imaging 12(4), 648–659 (October 2003).

[3] Makhalova Elena,” Fuzzy C - Means Clustering In Matlab”, The 7th International Days of Statistics and Economics, Prague, September 19-21, 2013.

[4] Lamberto Ballan, Lorenzo Seidenari, Giuseppe Serra, Marco Bertini and AlbertoDel Bimbo,” Recognizing Human Actions by using Effective Codebooks and Tracking”, IEEE Transactions on Multimedia [4],2013.

[5] Pasi Fränti, Timo Kaukoranta, Day-Fann Shen, and Kuo-Shu Chang ,” Fast and Memory Efficient Implementation of the Exact PNN” IEEE Transactions On Image Processing, Vol. 9, No. 5, May 2000.

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

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

[8] 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.

[9] 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.

[10] 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.

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