International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 3, March 2016

Collective Activity Recognition Using Naïve Bayes Classifiers

[1]Salini. S, [2] Prof.V.S. Mahalle
[1] M.E (C.E), Final year, [2] Professor, CSE dept,
Shri Sant Gajanan Maharaj College of Engineering, Shegaon- 444203
[1] salinicse23@gmail.com, [2] vsmahalle@gmail.com

Abstract: **Collective activity recognizing is the action of collecting the individual person's activity and it is distinguished as an atomic activity in isolation. In this paper, our work is based on the object detection and object recognition (ie) tracking is done for a given input video frame. The collective activity computes the class-specific person to person interaction patterns. The multi-interaction response proposed a activity – specific pattern for each interaction at the same time Naive bayes classifier is used to track the object and the tracked result is feed back into the recognition part to find category of an object and it is the final result.**

***Index Terms*— classifier, Collective activity recognition, Interaction modeling, Multi-interaction response, Naive Bayes.**

I. INTRODUCTION

Collective activity is a collective behaviour of a group of people in a frame because the interactions between people are important features. Therefore, collective activity recognition consists of the action of individual person in isolation. The scenario where two people are standing still and facing to each other, we may decide that they are talking to each other where those are standing still and facing to the same direction.

In this paper, we call the action of a single person as atomic activity and the collective activity recognition should differ from the atomic action recognition, where the actions performed by individuals are the main focus. And it should also be distinguished from the crowd activity recognition in a way that collective activity scenario to infer collective person-person interactions between several people. Comparing the crowd activity recognition is mainly based on discovering regular and common moving patterns of a large public scene often containing a crowd of more than tens or hundreds of people or objects in a single view such as in a train station.

In contrast, the people in collective activity are less occluded and the action of each person can be much more clearly observed. In addition, the number of people in collective activities is usually much less than that in crowd. The intrinsic and discriminative person-person interaction patterns may not be well exploited. The main problem is to develop a model or framework with increasing complexity to jointly learn more subtasks simultaneously of a collective activity recognition.

In this work, a different perspective and focus on one particular task of automatically learning person-person interactions. A learning-based approach to automatically mine the intrinsic person-person interaction patterns between atomic activities. In particular, it assumes that two atomic activities in a collective activity are connected.

In most cases, two connected atomic activities in one collective activity are either

1) quite similar and spatially close to each other to form a meaningful collective activity or

2) not quite similar but are strongly interacting to each other when participating a collective activity.

In order to learn the connection, we propose to formulate such a connection into the form of a inner product and to describe the collective information of the atomic activity in a clip. The collective activity is then aggregated together to generate a final response score for further prediction, also called as an interaction response (IR) model.

II. PERSON-PERSON INTERACTION MODELLING

A. Overview

In this person to person interactions modelling, an input video frame is given and it is given to object detection and object tracking method in which the persons are detected and tracked by using Interaction Response (IR) and multi-task Response (MIR). The output from object tracking is given to the classifier called naïve bayes classifier, this classifier done the effective than existing classifier method. This block diagram is shown in Fig no.1.

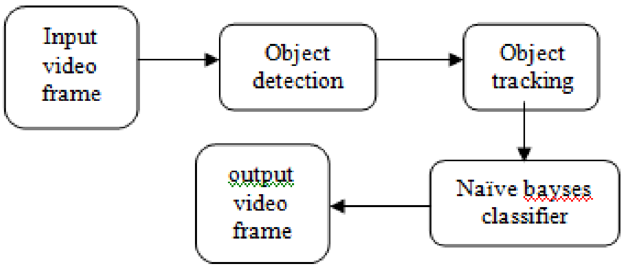

Fig no1. Block diagram of person to person interaction model

B. *Atomic Activity*

The first task is to detect and track all the people in each video clip for each collective, which could be represented by following terms. Number of people in a frame is denoted as Nq are detected and tracked in a video clip Vq. To describe the atomic activity associated to each person Pn,q, n = 1,..., Nq,where Pn,q consists of its action and its influence of a person associated to atomic activity. And fn,q is composed of two types of features: the motion-based features and spatial distribution of the other people around this person.

C. *Interaction Modeling for Atomic Activities*

Suppose there are M collective activities and we denote its label set as γ. In order to directly model the interaction between atomic activities associated to any two people (i.e. person-person interaction) in the video clip Vq, our collective activity interaction response Rm,q, is defined as follows:

$$R_{m,q} = \sum_{i,j=1,i\neq j}^{Nq} \Omega m\, f'_{i,q} f_{i,q}\ , \qquad \textbf{(1)}$$

Where m is denoted as one specified collective activity class and $f_{i,q}$ is the motion based and spatial distribution of the ith people around the person. And $f'_{i,q}$ is the transpose of$f_{i,q}$. The interaction matrix is indicated by a person-person interaction pattern for class, m ∈ R. Then$f_{i,q}\Omega m$ measures the connection between the two atomic activities associated to person i and person j acting for collective activity class m in video clip Vq.

The global collective activityconsists of sum of the effects of all person-person interactions in the video clip. Therefore, Rm,q is the response that measures the contributions of all the person-person interactions in the context of collective activity class m in the video clip Vq, it is called as the interaction response (IR) model.

In this work, we assume there is only one collective activity instance in each video clip and expect that if this class-specific person-person interaction is appearing in the mth activity class, Rm,q should output a higher score, otherwise a small value. Consequently, the inference of the collective activity class of a video clip Vq's can be casted as the following optimization problem:

$$\hat{} \text{ lq} = \text{argmax } m \in \gamma \text{ Rm,q} \qquad , (2)$$

where γ is the set of all possible activity class labels and ˆ lq is the predicted collective activity class label of Vq. It is obvious that for a given video clip Vq, the prediction of its class label depends only on the person-person interaction responses Rm,q.

D. *A Multi-Task Interaction Response (MIR)*

The main aim of the multi-task interaction response is to find better discriminative information in each collective activity, for example the collective activities of chasing and gathering could share the common element of walking, though facing to different directions. We exploit the idea of multi-task learning in order to better find out the discriminative information in each collective activity. The multi-task learning is designed to tackle different but related learning tasks in one framework, aiming to give better performance.

In addition, the collective activities could further share the element of people's spatial distribution. Multi-task interaction response (MIR) model is proposed by introducing a shared component among interaction matrices by learning each collective activity's interaction matrix as a task.

The MIR model jointly learns all the interaction matrices of different collective activities and the shared component. The learned interaction matrix for each collective activity can preserve more distinctive information for the corresponding collective activity.

III. NAIVE BAYES CLASSIFIER FOR OBJECT TRACKING

After the multi-task interaction response (MIR) of object detection then the object is moved to tracking system, inorder to the track the moving object from one frame to another frame. For this tracking naive bayes algorithm is used and it involves matching objects in consecutive frames using features such as points, lines or blobs. The object tracking can be viewed as a problem of probabilistic inference from ambiguous sensor measurements.

The tracking can be divided into in to four major categories:

- Region-based
- Active- contour based
- Feature based and
- Model based tracking.

A classifier separates the frame pixels into two classes: foreground (target object) and background, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.

Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting.

From Bayes theorem this posterior probability, a naive Bayesian classifier assumes that features are independent so the equation (1) can be rewritten as:

$$p(C|f) \propto p(C)\prod_{i=1}^{N} p(fi|C), \qquad \textbf{(3)}$$

The main advantage of the naïve Bayesian classifier is that it requires *relatively little* training data.

The necessary parameters for the classification are the *averages* and the *variances* of the various variables.

Indeed, the hypothesis of independence of variables does not require knowing more than the variance of each variable for every class, without having to calculate a *covariance* matrix.

The naive Bayes classifier has several properties that make it surprisingly useful. In particular, the decoupling of the class *conditional feature distributions* means that each distribution can be independently estimated as a *one-dimensional* distribution.

This helps alleviate problems stemming from the difficulties of dimensionality, such as the need for data sets that scale exponentially with the number of features.

IV. EXPERIMENTS

A. Datasets And Setting

Collective Activity Dataset (CAD) [1]:

It contains 44 video clips labeled with 5 different collective activities (*crossing, waiting, queuing, walking and talking*). There are eight facing directions (right, right-front,..., right-back) of people presented in this dataset. We choose the experimental setting, which splits one fourth of this dataset for testing and the rest for training. We have observed that with a limited number of splits, the averaged overall accuracies were not stable.

Unfortunately, most of the existing methods did not clarify how many number of splits were used in their settings. To compensate this ambiguity, we tested our algorithm by increasing the number of splits until the averaged results having no significant change. We observed that the averaged results become stable when the number of splits is larger than 20, at which the results of IR and MIR are reported in Table I.

Choi's Dataset [2]:

It consists of 32 video clips with 6 collective activities: gathering, talking, dismissal, walking together, chasing, and queuing. There are eight poses similar to the CAD dataset. We follow the standard experimental protocol of the 3-fold cross validation, suggested by Choi et al. [2]. This is a challenging dataset due to the large inter- and intra-class variations.

V. RESULTS

The result of the multi-interaction response based on naive bayes classifier is shown below, the video is given as input which is taken from the standard dataset.

Fig.No. 2 A. The Tracking results contain the detection of a moving as well how their motions are being tacked and classified (i.e) categorized. This result is how the multiple object are being recognized.

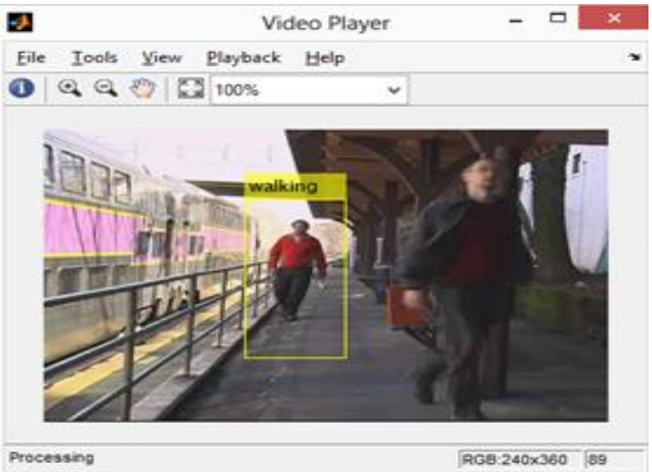

Fig No. 2 B Tracking next object

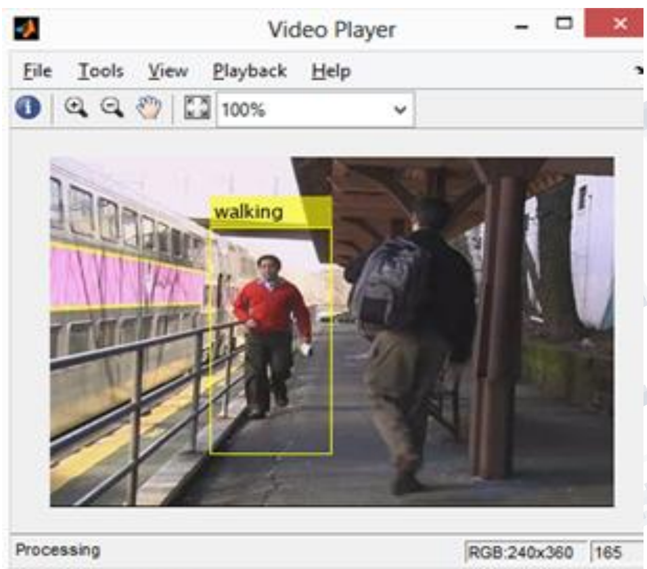

Fig no. 2 C Tracking target object

VI. CONCLUSION

In this paper, we focus on object detection and object tracking using the methods multi-interaction response and naive bayes classifier. The important characteristics is to formulate the person-person interaction of two atomic activities and multi- interaction patterns of different classes of collective activities are captured by different interaction matrices. The classification is only relative because there are still many tracking algorithms which combine different approaches of tracking. Naive Bayes provides the better classification by its class probability feature.

If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data. And even if the NB assumption doesn't hold, a NB classifier still often performs surprisingly well in practice. A good bet if you want to do some kind of semi-supervised learning, or want something embarrassingly simple that performs pretty well. The hypothesis of independence of variables does not require knowing more than the variance of each variable for every class, without having to calculate a *covariance* matrix.

.

REFERENCES

[1] M. R. Amer and S. Todorovic. A chains model for localizing partici- pants of group activities in videos. In Proc. IEEE Int. Conf. Comput. Vis. 2011 Nov, 786–793.

[2] M. R. Amer, S. Todorovic, A. Fern, and S.-C. Zhu. Monte Carlo tree search for scheduling activity recognition. In Proc. IEEE Int. Conf. Comput. 2013 Dec. 1353–1360.

[3] S.-H. Bae and K.-J. Yoon. Robust online multiobject tracking with data association and track management. IEEE Trans. Image Process. 2014 jul. 23(7), 2820–2833.

[4] W. Choi and S. Savarese. A unified framework for multi-target tracking and collective activity recognition. In Proc. 12th Eur. Conf. Comput. 2012, 4, 215–230.

[5] SERBY, D., KOLLER-MEIER, S., AND GOOL, L. V. 2004. Probabilistic object tracking using multiple features. In IEEE International Conference of Pattern Recognition (ICPR). 184-187.

[6] Yilmaz, a., li, x., and shah, m. 2004. Contour based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Trans. Patt. Analy. Mach. Intell. 26, 11, 1531–1536

[7] Silveira, g. and malis, e. 2007. real-time visual tracking under arbitrary illumination changes. ieee international conference on computer vision and pattern recognition, 1–6. 6.

[8] WU, Y. AND FAN, J. 2009. Contextual flow. IEEE International Conference on Computer Vision and Pattern Recognition, 33–40.

[9] Learning person–person interaction in collective activity recognition, xiaobin chang, wei-shi zheng, and jianguo zhang, ieee transactions on image processing, vol. 24, no. 6, june 2015.

[10] "Multiple Object Tracking and Segmentation in Video Sequences" .K. Meenatchi PG Scholar, Dept. of ECE CMS College of Engineering Namakkal, India , P. Subhashini Assistant Professor, Dept. of ECE CMS College of Engineering Namakkal, India.

[11] "Cost-Sensitive Top-down/Bottom-up Inference for Multiscale Activity Recognition"Mohamed R. Amer1, Dan Xie2, Mingtian Zhao2, Sinisa Todorovic1, and Song-Chun Zhu21Oregon State University, Corvallis, oregonamerm,sinisa}@onid.orst.edu 2University of California, Los Angeles, California {xiedan,mtzhao,sczhu}@ucla.edu.

[12] "Learning Latent Constituents for Recognition of Group Activities in Video"Borislav Antic and Bj¨ornOmmerHCI& IWR, University of Heidelberg, Germany borislav.antic@iwr.uni-heidelberg.de, ommer@uni-heidelberg.de

[13] *Collective Activity Dataset*. [Online]. Available: http://www.eecs. umich.edu/vision/activity-dataset.html, accessed Sep. 2013.

[14] *Dataset: A Unified Framework for Multi-Target Tracking and Collective Activity Recognition. [Online]. Available: http://wwwpersonal*