ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

A Methodology Based on HMRF Model for Human Actions Tracking and Labeling

K. Ravi Kumar [1], A. M. Gunashekar[2]

M.Tech Student, Dept. of Electronics & Communication Engineering, Sree Rama Engineering College, Tirupati, India[1]

Associate Professor, Dept. of Electronics & Communication Engineering, Sree Rama Engineering College, Tirupati, India [2]

ABSTRACT: We present a unified framework to track multiple people, as well localize, and label their activities, in complex long-duration video sequences. To do this, we focus on two aspects: 1) the influence of tracks on the activities performed by the corresponding actors and 2) the structural relationships across activities. We propose a two-level hierarchical graphical model, which learns the relationship between tracks, and their corresponding activity segments, as well as the spatiotemporal relationships across activity segments. Such contextual relationships between tracks and activity segments are exploited at both the levels in the hierarchy for increased robustness. An L1-regularized structure learning approach is proposed for this purpose. While it is well known that availability of the labels and locations of activities can help in determining tracks more accurately and vice-versa, most current approaches have dealt with these problems separately. Inspired by research in the area of biological vision, we propose a bidirectional approach that integrates both bottom-up and top-down processing, i.e., bottom-up recognition of activities using computed tracks and top-down computation of tracks using the obtained recognition.

KEYWORDS: Wide-area activity recognition, hierarchical Markov random field, context modelling

I.INTRODUCTION

A continuous video consists of two inter-related components: 1) tracks of the persons in the video and 2) localization and labels of the activities of interest performed by these actors. Activity analysis of continuous videos involves solving both the tracking as well as recognition problems. In the past, most research on video analysis has treated these two problems separately. However, in the context of continuous videos, such as surveillance or sports videos, the solution to one problem can help in finding the solution to the other. Knowing the tracks can help in better detection and recognition of activities. Similarly, information about the location and labels of activities in a scene can help in determining the movement of people in the scene. Therefore, we propose a method which performs the two tasks in an integrated framework, modeling contextual relationships between tracks as well as activities using graphical models.

Research in the area of biological vision has shown that, the human visual system employs bi-directional (top-down as well as bottom-up) reasoning in analyzing and interpreting data of multiple resolutions [2]. This has been found to be particularly helpful in correcting errors due to false detections or noise. Applying these concepts to the analysis of continuous videos, we consider the task of obtaining recognition scores using tracks as a bottom-up (or feed-forward) approach, while the task of correcting tracks using obtained recognition labels is treated as top-down (or feedback) processing. We alternate between both these steps resulting in a bi-directional algorithm that can help in increasing the accuracy of both these tasks.

In applications such as activity recognition, the structure of the graph is difficult to determine. Prior approaches have either used fixed graphical models such as in [6] and [8] or built graphs of known structures such as and-or graphs [10]. Recent approaches such as [11] have tackled this problem by using a greedy forward search to determine the best possible graph, thereby making the learning and inference very intensive. We learn an optimal set of structural relationships along with the parameters automatically in an L1-regularized learning framework.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

The rest of this paper is organized as follows. Section II first reviews the current state-of-the-art techniques. Our proposed method is described in Section III. Then experimental results are reported in Section IV to demonstrate the superior performance of our framework. Finally, conclusions are presented in Section V.

II.EXISTING METHOD

The ultimate goal of activity modelling is to understand behaviour, i.e. the meaning of activity in the shape of a semantic description. Clearly, action/activity/behaviour analysis entails the capability of spotting objects that are of interest for the given surveillance task. The existing work presented here[1] is the visual objects of interest occurring in video streams are the preys to be chased and handled by the visual forager. Decision at the finer level of a single stream concern which object is to be chosen and analyzed (prey choice and handling, depending on task), and when to disengage from the spotted object for deploying attention to the next (prey leave).

The reformulation of visual attention in terms of foraging theory is not simply an informing metaphor. What was once foraging for tangible resources in a physical space became, over evolutionary time, foraging in cognitive space for information related to those resources, and such adaptations play a fundamental role in goal-directed deployment of visual attention. Under these rationales, we present a model of Bayesian observer's attentive foraging supported by the perception/action cycle. Building on the perception/action cycle, visual attention provides an efficient allocation and management of resources. The cycle embodies two main functional blocks: the perceptual component and the executive control component.

The perceptual component is in charge of "What" to look for, and the executive component accounts for the overt attention shifts, by deciding "Where and How" to look at, i.e., the actual gaze position, and thus the observer's Focus of Attention (FoA). The observer's perceptual system operates on information represented at different levels of abstraction (from raw data to task dependent information); at any time, the currently sensed visual stimuli depend on the oculomotor action or gaze shift performed either within the stream (within-patch) or across streams (between-patch). Based on perceptual inferences at the different levels, the main feedback information passed on to the executive component, or controller, is an index of stream quality formalized in terms of the configuration complexity of potential objects sensed within the stream.

III. PROPOSED MITIGATION SCHEME

The illustration of our proposed method is shown below in Figure 1. We have available a set of annotated training data with labelled tracks and activities, and a test video, for which the tracks as well as activities need to be discovered. We assume that we have with us a set of tracklets, which are short-term fragments of tracks with low probability of error. Tracklets have to be joined to form long-term tracks. In a multi person scene, this involves tracklet association. Here, we use a basic particle filter for computing tracklets as mentioned. For the test video, it is assumed that each tracklet belongs to a single activity.

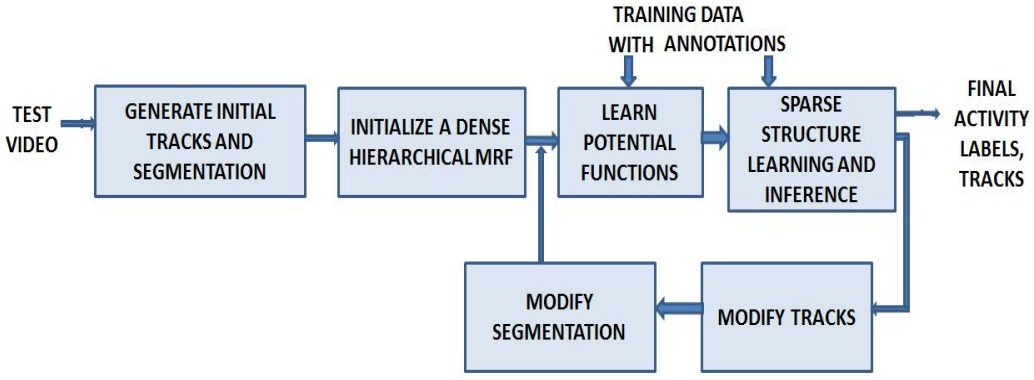

Figure.1.Block diagram shows the illustration of our proposed method.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

Pre-processing consists of computing tracklets and computing low level features such as space-time interest points in the region around these tracklets. Tracking involves association of one or more tracklets to tracks. Activity localization can now be defined as a grouping of tracklets into activity segments and recognition can be defined as the task of labeling these activity segments. To begin with, we generate a set of match hypotheses for tracklet association and a likely set of tracks. An observation potential is computed for each tracklet using the features computed at the tracklet. Tracklets are grouped into activity segments using a standard baseline classifier such as multiclass SVM or motion segmentation.

Next, we construct a two-level Markov random field using the tracklets and activity segments. The first level nodes correspond to the tracklets and the second level nodes correspond to the activity segments. One or more tracklets can correspond to the same activity segment. Edges model relationships between nodes of the same level as well as nodes at different levels. This structure incorporates the context information between adjacent tracklets as well as across activity segments. The dense HMRF has edges connecting each node to all other nodes within a certain spatiotemporal range. This gives us the initial graph on which we perform the learning and recognition. The node features and edge features for the potential functions are computed from the training data. There are two tasks to be performed on the graph - choosing an appropriate structure and learning the parameters of the graph. Both these steps can be performed simultaneously by posing the parameter learning as an L1-regularized optimization.

The sparsity constraint on the HMRF ensures that the resulting parameters are sparse, thus capturing the most critical relationships between the objects. The parameters which are set to zero denote the edges which have been deleted from the graph. The non-zero parameters denote the parameters of edges retained after automatic structure discovery. Inference on this graph provides the posterior probabilities for all nodes using information available at two resolutions. The activity labels are used in a top-down fashion to recomputed the tracks. Activity segmentation on the recomputed tracks gives us a new set of nodes on which structure learning and inference is then repeated. Convergence is said to be achieved when the node labels and tracks do not change from one iteration to the next. The output of the algorithm is a set of tracks, segments and the labels assigned to each segment.

Applications:

1. Surveillance videos
2. Sports videos
3. Tracking and activity recognition
4. Airports
5. Military applications
6. Assisting the sick and disabled
7. Home-based rehabilitation

IV. RESULTS

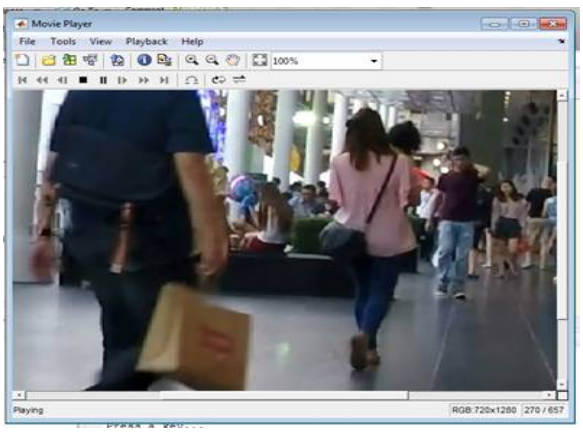

Figure.1: Input video

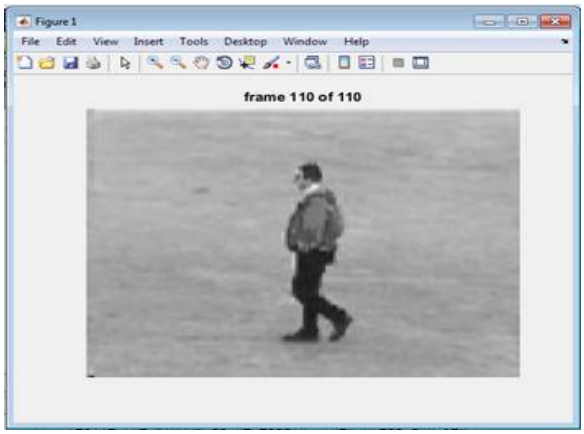

Figure2: Training of first sequence

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

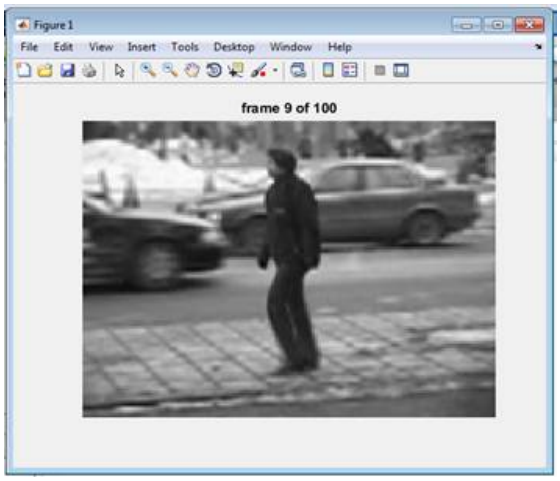

Figure 3: Training of second sequence

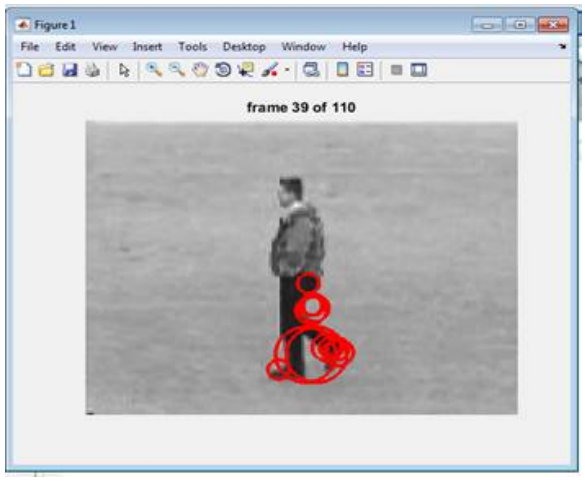

Figure 4: Extraction of features from the first sequence

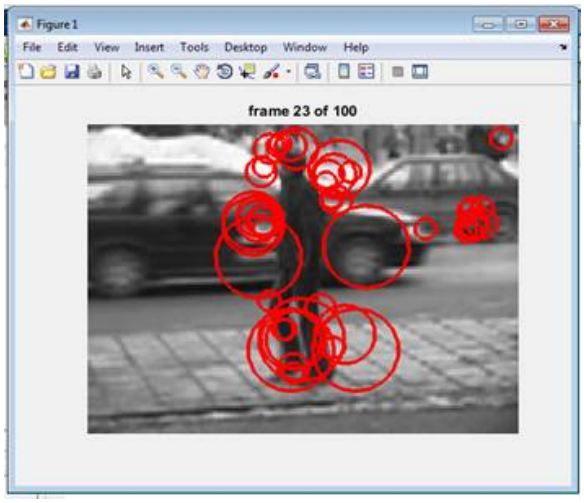

Figure 5: Extraction of features from the second sequence

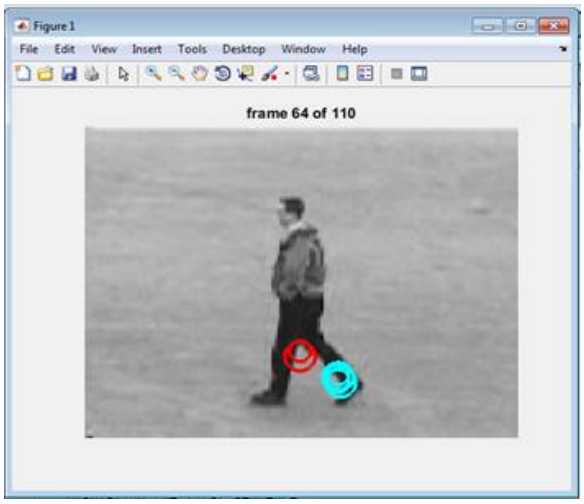

Figure 6: Feature clusters from the first sequence

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

Figure 7: Tracking at fourth iteration

Figure 8: Tracking at 23rd iteration

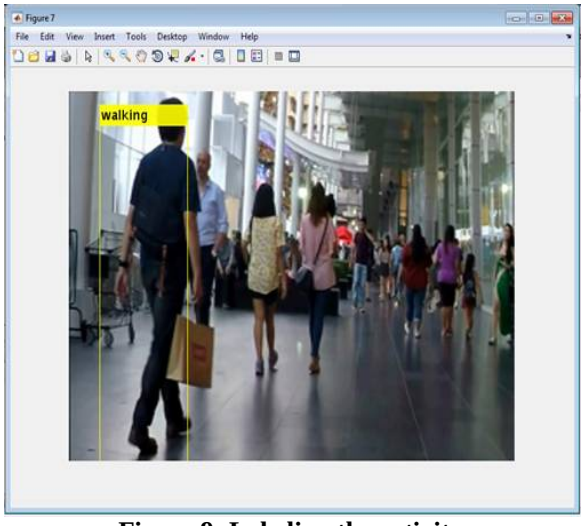

Figure 9: Labeling the activity

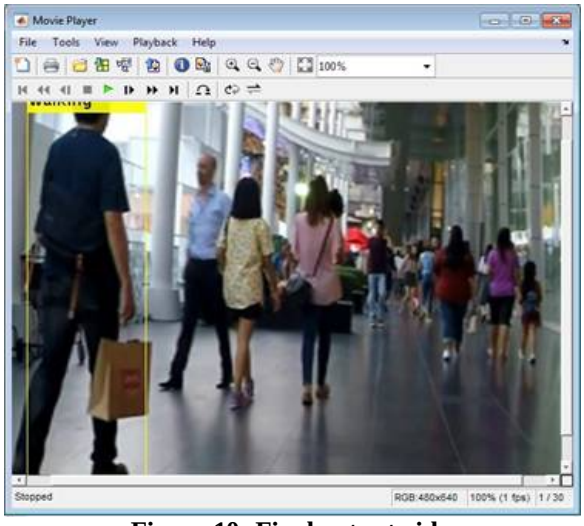

Figure 10: Final output video

V.COMPARISION

A. *B. **Methods***	*C.* *D. **Accuracy*** *E.*	*F.* *G. **PSNR***
H. *I. **Proposed Method***	*J.* *K. **99.03***	*L.* *M. **20.1439***
N. *O. **Existing Method***	*P.* *Q. **97.43***	*R.* *S. **15.9014***

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

VI.CONCLUSIONS

In this paper, we have presented a method which can perform tracking, localization and recognition of activities in continuous sequences in an integrated framework. The proposed framework uses an initial set of tracks for analysis of activities using a two-level hierarchical Markov random field. The activity labels obtained in the bottom-up processing are in turn used to correct the errors in tracking in a top-down approach. We have demonstrated that the L1-regularized learning of parameters is a good substitute to alternate methods such as greedy forward search. The biologically inspired bi-directional processing is shown to be effective in improving the accuracy of tracking as well as activity recognition. This method can be extended to other applications which utilize graphical models for context representation.

REFERENCES

[1] Attentive Monitoring of Multiple Video Streams Driven by a Bayesian Foraging Strategy Paolo Napoletano, *Member, IEEE*, Giuseppe Boccignone, and Francesco Tisato, IEEE transactions on image processing, vol. 24, no. 11, november 2015.
[2] S. Khamis, V. I. Morariu, and L. S. Davis, "A flow model for joint action recognition and identity maintenance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1218–1225.
[3] V. Chandrasekaran, N. Srebro, and P. Harsha, "Complexity of inference in graphical models," in *Proc. 24th Annu. Conf. Uncertainty Artif. Intell.*, 2008, pp. 70–78.
[4] C.-Y. Chen and K. Grauman, "Efficient activity detection with maxsubgraph search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1274–1281.
[5] W. Choi and S. Savarese, "A unified framework for multi-target tracking and collective activity recognition," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 215–230.
[6] W. Choi, K. Shahid, and S. Savarese, "Learning context for collective activity recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 3273–3280.
[7] U. Gaur, Y. Zhu, B. Song, and A. Roy-Chowdhury, "A 'string of feature graphs' model for recognition of complex activities in natural videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011,
pp. 2595–2602.
[8] V. I. Morariu and L. S. Davis, "Multi-agent event recognition in structured scenarios," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 3289–3296.
[9] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2012.
[10] M. Pei, Y. Jia, and S.-C. Zhu, "Parsing video events with goal inference and intent prediction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 487–494.
[11] Y. Zhu, N. M. Nayak, and A. K. Roy-Chowdhury, "Context-aware modeling and recognition of activities in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2491–2498.
[12] A. Dore, A. F. Cattoni, and C. S. Regazzoni, "Interaction modeling and prediction in smart spaces: A bio-inspired approach based on autobiographical memory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 6, pp. 1191–1205, Nov. 2010.
[13] S. Chiappino, L. Marcenaro, and C. Regazzoni, "Selective attention automatic focus for cognitive crowd monitoring," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2013, pp. 13–18.
[14] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, Jan. 2013.
[15] C. de Leo and B. S. Manjunath, "Multicamera video summarization and anomaly detection from activity motifs," *ACM Trans. Sensor Netw.*, vol. 10, no. 2, p. 27, Jan. 2014.