ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

An Efficient Object Tracking and Detection In Videos

S.Himagiri Sudha, PG Scholar
Department of Computer Science and Engineering,
Indra Ganesan College Of Engineering
Trichy, Tamilnadu
sudhasridhar594@gmail.com

S. Vimalathithan, Associate Professor
Department of Computer Science and Engineering,
Indra Ganesan College Of Engineering
Trichy, Tamilnadu
athi_svimal@hotmail.com

Abstract— **The system present a coherent, discriminative framework for simultaneously tracking multiple people and estimating their collective activities. Instead of treating the two problems separately, our model is grounded in the intuition that a strong correlation exists between a person's motion, their activity, and the motion and activities of other nearby people. Instead of directly linking the solutions to these two problems, Introduce a hierarchy of activity types that creates a natural progression that leads from a specific person's motion to the activity of the group as a whole. The system propose a two-level hierarchical graphical model, which learns the relationship between tracks, relationship between tracks, and their corresponding activity segments, as well as the spatiotemporal relationships across activity segments.We also propose an algorithm for solving this otherwise intractable joint inference problem by combining belief propagation with a version of the branch and bound algorithm equipped with integer programming.**

***Keywords*—MRF, SVM, Background Subtraction.**

I. INTRODUCTION

Surveillance videos in unconstrained environments typically consist of long duration sequences of activities which occur at different spatio-temporal locations and can involve multiple people acting simultaneously. Often, the activities have contextual relationships with one another. Although context has been studied in the past for the purpose of activity recognition, the use of context in recognition of activities in such challenging environments is relatively unexplored. In this paper, It propose a novel method for capturing the spatio-temporal context between activities in a Markov random field. The structure of the MRF is improvised upon during test time and not predefined, unlike many approaches that model the contextual relationships between activities. Given a collection of videos and a set of weak classifiers for individual activities, the spatio-temporal relationships between activities are represented as probabilistic edge weights in the MRF. This model provides a generic representation for an activity sequence that can extend to any number of objects and interactions in a video. We show that the recognition of activities in a video can be posed as an inference problem on the graph. We conduct experiments on the publicly available UCLA office dataset and the VIRAT dataset, to demonstrate the improvement in recognition accuracy using this proposed model as opposed to recognition using state-of-the-art features on individual activity regions.

Reasoning about pose and motion of objects, based on images or video, is an important task for many machine vision applications. Estimating the pose of articulated objects such as people and animals is particularly challenging due to the complexity of the possible poses yet has applications in computer vision,

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

medicine, biology, animation, and entertainment. Realistic natural scenes, object motion, noise in the image observations, incomplete evidence that arises from occlusions, and high dimensionality of the pose itself are all challenges that need to be addressed. In this thesis propose a class of approaches that model objects using continuous-state graphical models. It show that these approaches can be used to effectively model complex objects by allowing tractable and robust inference algorithms that are able to infer pose of these objects in the presence of realistic appearance variations and articulations. It use continuous-state graphical models to model both rigid and articulated object structures; where nodes correspond to parts of objects and edges represent the constraints between parts encoded as statistical distributions. For rigid objects, these constraints can model spatial and temporal relationships between parts; for articulated objects kinematic, inter-penetration and occlusion relationships. Localization, pose estimation, and tracking can then be formulated as inference in these graphical models. This has a number of advantages over more traditional methods. First, these models allow inference algorithms that scale linearly with the number of body parts by breaking up the high-dimensional search for pose into a number of lower-dimensional collaborative searches. Secondly, partial occlusions can be dealt with robustly by propagating spatial information between parts. Thirdly, "bottom-up" information can be incorporated directly and effectively into the inference process, helping the algorithm to recover from transient tracking failures. Show that these hierarchical continuous-state graphical models can be used to solve the challenging problem of inferring the 3D pose of the person from a single monocular image.

Images and video provide rich low-level cues about the scenes and the objects in them. The goal of machine vision is to develop approaches for extracting meaningful semantic knowledge from these low-level cues; for example, in the case of robotics, allowing direct interaction of the computer with the real world. This is challenging because of the large variability that exists in imaging conditions and objects themselves. Objects that belong to same semantic classes can appear differently, image differently, and even act differently. Objects like cars vary in size, shape and color; people in weight, body shape and size/age. Motion of these objects is often complex and is governed by physical interactions with the environment (e.g. Balance, gravity) and higher order cognition tasks like intent. All these challenges make it impossible to determine the regions of the image that belong to a particular object, or part of the object, directly. Computer vision algorithms must propagate information both spatially and temporally, to effectively resolve ambiguities that arise, by inferring globally plausible and temporally persistent interpretations. Statistical methods are often used for these tasks, to allow reasoning in the presence of uncertainty. Graphical models provide a powerful paradigm for intuitively describing the statistical relationships precisely and in a modular fashion. These models effectively represent statistical and conditional independence relationships between variables, and allow tractable inference algorithms that make use of encoded conditional independence structure. In computer vision, inference algorithms for these graphical models need to be developed to handle the high-dimensionality of the parameter-space, complex statistical relationships between variables and the continuous nature of the variables themselves. This thesis will concentrate on localizing, estimating the pose of and tracking rigid and articulated objects (most notably people) in images and video. Estimating the pose of people is particularly interesting because of a variety of applications in rehabilitation medicine, sports and the entertainment industry. Pose estimation and tracking can also serve as a front end for

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

higher level cognitive reasoning in surveillance or image understanding. Localizing and tracking articulated structures like people, however, is challenging due to the additional degrees of freedom imposed by the articulations (compared with rigid objects). In general the search space grows exponentially with the number of part s and the degrees of freedom associated with each joint connecting these parts, making most straight forward search algorithms intractable. The recurring theme of this thesis will be the merge of Monte Carlo sampling and non-parametric inference methods with graphical models, resulting in tractable and distributed inference algorithms for localizing and tracking objects in 2D and 3D. Advocate the use of a hierarchical inference approach for mediating the Complexity of harder inference problems. We will first describe the problem of pose estimation and tracking as it applies to rigid and articulated objects. Describe a kinematic model and the corresponding Monte Carlo sampling methods, which have successfully been applied to track articulated objects given an initial pose (often supplied manually at the first frame). Consider a more general problem of tracking people automatically, by first inferring the pose of the person and then incorporating temporal consistency constraints in a collaborative inference framework. Show that contributions in all aspects of this problem by addressing modeling choices, inference, likelihoods and priors.

A prevalent view in the biological community on the role of feedback among cortical areas is that of selective attention modeled by biased competition. Vision is still considered to be accomplished by a feed forward chain of computations. Although these models give an apparently complete explanation of some experimental data, they use the sophisticated machinery of feedback rather impoverished way and they persist in viewing the computations in each visual area as predominantly independent processes. However, some of this recent neuro physiological evidence cannot be fully accounted for by biased competition models. Instead, believe that they reflect underlying cortical processes that are indicative of a generative model. This link data to the proposed framework and explain how ideas of resonance and predictive coding can potentially be reconciled and accommodated in a single framework.

In proposed system the system propose a two-level hierarchical graphical model, which learns the relationship between tracks, relationship between tracks, and their corresponding activity segments, as well as the spatiotemporal relationships across activity segments. The HMRF is constructed on these track lets and activity segments. Using the obtained labels from recognition, the cost matrix is updated and the tracks are re-computed. The algorithm is repeated with the modified tracks. Therefore, This utilize STIP features in This experiments. Similarly, This choose the bag-of-words against other approaches such as String of feature graphs for the baseline classifier because feature relationships are not as prominent in a distant scene and graph matching can be computationally expensive. A radial basis function kernel has been used for the SVM.

As a unifying framework to integrate the low- and high- level representations of human activity in video, This propose a hierarchical graphical model for recognizing human activities. Graphical models (also called graphs) are a pervasive data structure in computer science and engineering, and algorithms for working with them are fundamental to these fields. Hundreds of interesting computational problems are defined in terms of graphs. Graphical models have been widely and successfully used in many application areas. In general, a graphical model is an efficient tool to represent a complicated system that is composed of multiple variables. In graphical models, the

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

variables of interest are represented as the nodes and the relations between the variables are represented as links (or edges) that connect the corresponding nodes. Basically, graphical models (or graphs) are classified into two classes: directed graphs and undirected graphs. Directed graphs contain directed links thatre present cause-effect relations between the nodes; a directed link denoted by an arrow originates from a cause variable and is directed toward an effect variable. Undirected graphs are used when no cause-effect relations are involved between the variables. Depending on the problem characteristics, various graphical models can be formulated. It is also possible to combine different classes of graphs in a systematic way to represent a complicated problem domain. In this dissertation, This present a hierarchical graphical model for recognizing human actions and interactions in video. This method encompasses the whole processing that spans the pixel level, blob level, object level and the event level computation of video. At low levels (i.e., the pixel level and the blob level), undirected graphs are used to represent the variables and their interrelations. At high levels(i.e., the object level and the event level), directed graphs are used to represent the variables and possible causal relations between the variables.

This propose a hierarchical graphical model to deal with these tasks and present a framework for understanding generic human interaction in a color video. The main contributions of this research include the integration of low-level vision algorithms and high-level knowledge representation in terms of a hierarchical graphical model, and a probabilistic treatment of uncertainty due to occlusion inhuman interaction situations. The contributions of this research are also as follows:(1) A new hierarchical framework is proposed for the recognition of two-person interactions at a detailed level using a hierarchical Bayesian network. Ambiguity in human interaction due to occlusion is handled by inference with the Bayesian network. (3) A human-friendly vocabulary is generated for high-level event description. (4) A stochastic graphical model is proposed for the recognition of two-person interactions in terms of 'subject + verb + object' semantics. In This hierarchical framework, the event level is the highest level of representation for understanding human interactions in video sequences. At the event level, the focus of This system is to achieve semantic understanding of video imagery, whereas at the object level the focus was to obtain a qualitative description of video objects such as body pose.

II. Proposed Methodology

In proposed system the system propose a two-level hierarchical graphical model, which learns the relationship between tracks, relationship between tracks, and their corresponding activity segments, as well as the spatiotemporal relationships across activity segments. The HMRF is constructed on these track lets and activity segments. U sing the obtained labels from recognition, the cost matrix is updated and the tracks are re-computed. The algorithm is repeated with the modified tracks. Therefore, This utilize STIP features in This experiments. Similarly, This choose the bag-of-words against other approaches such as String of feature graphs for the baseline classifier because feature relationships are not as prominent in a distant scene and graph matching can be computationally expensive. A radial basis function kernel has been used for the SVM.

As a unifying framework to integrate the low- and high- level representations of human activity in video, This propose a hierarchical graphical model for recognizing human activities. Graphical models (also called graphs) are a pervasive data structure in computer

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

science and engineering, and algorithms for working with them are fundamental to these fields. Hundreds of interesting computational problems are defined in terms of graphs. Graphical models have been widely and successfully used in many application areas. In general, a graphical model is an efficient tool to represent a complicated system that is composed of multiple variables. In graphical models, the variables of interest are represented as the nodes and the relations between the variables are represented as links (or edges) that connect the corresponding nodes. Basically, graphical models (or graphs) are classified into two classes: directed graphs and undirected graphs. Directed graphs contain directed links that represent cause-effect relations between the nodes; a directed link denoted by an arrow originates from a cause variable and is directed toward an effect variable. Undirected graphs are used when no cause-effect relations are involved between the variables. Depending on the problem characteristics, various graphical models can be formulated. It is also possible to combine different classes of graphs in a systematic way to represent a complicated problem domain. In this dissertation, This present a hierarchical graphical model for recognizing human actions and interactions in video. This method encompasses the whole processing that spans the pixel level, blob level, object level and the event level computation of video. At low levels (i.e., the pixel level and the blob level), undirected graphs are used to represent the variables and their interrelations. At high levels(i.e., the object level and the event level), directed graphs are used to represent the variables and possible causal relations between the variables.

This propose a hierarchical graphical model to deal with these tasks and present a framework for understanding generic human interaction in a color video. The main contributions of this research include the integration of low-level vision algorithms and high-level knowledge representation in terms of a hierarchical graphical model, and a probabilistic treatment of uncertainty due to occlusion inhuman interaction situations. The contributions of this research are also as follows:(1) A new hierarchical framework is proposed for the recognition of two-person interactions at a detailed level using a hierarchical Bayesian network. (2) Ambiguity in human interaction due to occlusion is handled by inference with the Bayesian network. (3) A human-friendly vocabulary is generated for high-level event description. (4) A stochastic graphical model is proposed for the recognition of two-person interactions in terms of 'subject + verb + object' semantics. In This hierarchical framework, the event level is the highest level of representation for understanding human interactions in video sequences. At the event level, the focus of This system is to achieve semantic understanding of video imagery, whereas at the object level the focus was to obtain a qualitative description of video objects such as body pose. The event level understanding involves the semantic description of video events, called event semantics. A gap exists between geometric information obtained from images and semantic information contained in natural language. It is necessary to associate visual features with natural language verbs and symbols to build the event semantics of two-person interactions.

This have developed a comprehensive framework for recognizing human actions and interactions in color video using a hierarchical graphical model. This represented the human activity captured in video at four levels: the pixel level, the blob level, the object level, and the event level. At the pixel level, background subtraction and pixel-color classification are performed. At the blob level, individual pixels have been merged into blobs and the blobs are tracked along the sequence. The blob-level

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

representation of video is domain-independent and does not depend on world-object models. At the object level, the blobs are associated with world objects (i.e., a human body in This research) with the introduction of minimal domain knowledge about the objects. Therefore, This approach is an appearance-based method. At the object level, the human body has been segmented into individual body parts: head, upper body, and lower body, each of which has been subdivided into skin vs. non-skin areas. This developed a Bayesian network to estimate poses based on the observed appearance of the human body. Pose estimation has been performed on a frame-by-frame basis.

III. SUPPORT VECTOR MACHINE

Support vector machine (SVMs) are used for data categorization. It is used for mapping facts into high dimensional gap and for the maximal margin.

IV. MARKOV RANDOM FIELD (MRF)

The concept of MRF is a generalization of that of Markov processes (MPs) which are widely used in sequence analysis. An MP is defined on a domain of time rather than space.
Let Z = {*Z*1, *Z*2, ..., *Zm*} be a family of random variables defined on the set *S*, in which each random variable *Zi* takes a value *zi* in *L*. The family *Z* is called a *random field* [12]. We use the notion *Zi* = *zi* to denote the event that *Zi* takes the value *zi* and the notion (*Z*1 = *z*1, *Z*2 = *z*2,.... *Zm* = *zm*) to denote the joint event. For simplicity a joint event is abbreviated as *Z* = *z* where, *z* = {*z*1, *z*2, ...} is a configuration of *z*, corresponding to realization of a field. For a discrete label set *L*, the probability that random variable *Zi* takes the value *zi* is denoted *P*(*Zi* = *zi*),abbreviated *P*(*zi*), and the joint probability is denoted as *P*(*Z* = *z*) = *P*(*Z*1 = *z*1, *Z*2 = *z*2,.... *Zm* = *zm*) and abbreviated *P*(*z*).

F is said to be a Markov random field on *S* with respect to a neighborhood system *N* if and only if the following two conditions are satisfied,
P(*Z* = *z*) > 0, *z Z* (Positivity) (6) □ □
P(*zi*|*zS*−*i*) = *P*(*zi*|*zNi*) (Markovianity)

V. IMPLEMENTATION AND RESULT

Motion detection is a process of confirming a change in position of an object relative to its surroundings or the change in the surroundings relative to an object. Results of moving object detection in the continuous frames

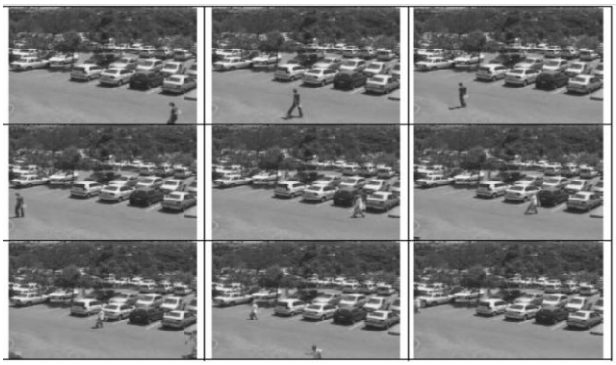

FIG 1. RESULTS OF MOVING OBJECT DETECTION IN THE CONTINUOUS FRAMES

Background subtraction is the process of separating out foreground objects from the background in a sequence of video frames.

a) Background image

(b) Foreground image

V (A) FRAME CONVERSION

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

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.

V (B) FILTERING

The system 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.

V (C) TRACK THE VIDEO

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.

V (D) FEATURE EXTRACTION

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 labeling, the recognition is assumed to be correct. Some examples of data which were correctly identified using our approach while incorrectly identified using a dense graphical model are shown. The lower nodes of the graph denote tracklets and the upper nodes denote activities.

VI. CONCLUSION

Spatio-temporal contextual relationships between activities and the influence of tracks on them has been modeled using the graph. The activity labels obtained in the bottom-up processing are in turn used to correct the errors in tracking in a top-down approach. It demonstrated that the L1-regularized learning of parameters is a good substitute to alternate methods such as greedy forward search. In this paper, it presented a method which can perform tracking, localization and recognition of activities in continuous sequences in an integrated framework. The recognition of human activity requires the representation of gesture. The gesture is defined to be the event of temporal evolution of the instantaneous poses. At the event level, introduced the notion of interaction hierarchy. In the interaction hierarchy, two-person interaction is defined in terms of the combination of two single-person actions. The single person action is represented in terms of the torso gesture and arm/leg gesture(s) using the operational triplets proposed in this research. It developed a dynamic Bayesian network to combine the poses into gestures, and a rule-based decision tree to classify human interactions occurring between two persons. The performance of the system depends on several factors. The high-level processing relies on the robustness of the low level processing. For example, accurate background subtraction at the pixel level is crucial to each subsequent level. This experiments showed that pixel-color

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 2, April 2017

classification is robust and reliable with rare exceptions. The experiments showed that the background subtraction is affected by illumination change and by the noise due to fast motion of objects. The shadow removal step leaves some residue, which affects the appearance of the fore-ground area. These issues are common problems in appearance-based approaches. It is desired to enhance the performance of the low level processing.

REFERENCES

[1] M. R. Amer and S. Todorovic, "Sum-product networks for modeling activities with stochastic structure," in Proc. IEEE Conf. Comput. Vis.Pattern Recognit., Jun. 2012, pp. 1314–1321.
[2] W. Brendel and S. Todorovic, "Learning spatiotemporal graphs of human activities," in Proc. IEEE Int. Conf. Comput. Vis. , Nov. 2011,pp. 778–785.
[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 max-subgraph 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 collec-tive activity recognition," in Proc. IEEE Conf. Comput. Vis. PatternRecognit. , 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 naturalvideos," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011,pp. 2595–2602.
[8] M. Hoai, Z.-Z. Lan, and F. De la Torre, "Joint segmentation and classification of human actions in video," in Proc. IEEE Conf. Comput.Vis. Pattern Recognit. , Jun. 2011, pp. 3265–3272.
[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] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in Proc. 6th ACMInt. Conf. Image Video Retr. , 2007, pp. 494–501.
[11] D. Kuettel, M. Breitenstein, L. V an Gool, and V. Ferrari, "What's goingon? Discovering spatio-temporal dependencies in dynamic scenes,"in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010,pp. 1951–1958.
[12] T. Lan, L. Sigal, and G. Mori, "Social roles in hierarchical models for human activity recognition," in Proc. IEEE Conf. Comput. Vis. PatternRecognit. , Jun. 2012, pp. 1354–1361.
[13] Q. V. Le et al., "Building high-level features using large scale unsuper-vised learning," in Proc. Int. Conf. Mach. Learn., 2012, p. 103.
[14] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with indepen-dent subspace analysis," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. , Jun. 2011, pp. 3361–3368.
[15] Y. Li and R. Nevatia, "Key object driven multi-category object recogni-tion, localization and tracking using spatio-temporal context," in Proc.10th Eur. Conf. Comput. Vis. , 2008, pp. 409–422.
[16] V. I. Morariu and L. S. Davis, "Mu lti-agent event recognition in struc-tured scenarios," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,Jun. 2011, pp. 3289–3296.
[17] N. M. Nayak, Y. Zhu, and A. K. Roy-Chowdhury, "Exploiting spatio-temporal scene

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

structure for wide-area activity analysis in unconstrained environments," IEEE Trans. Inf. Forensics Security, vol. 8, no. 10,pp. 1610–1619, Oct. 2013.

[18] N. M. Nayak, A. T. Kamal, and A. K. Roy-Chowdhury, "Vector field analysis for motion pattern identification in video," in Proc. 18th IEEEInt. Conf. Image Process. , Sep. 2011, pp. 2089–2092.

[19] N. M. Nayak and A. K. Roy-Chow dhury, "Learning a sparse dictionaryof video structure for activity modeling," in Proc. IEEE Int. Conf. ImageProcess., Oct. 2014, pp. 4892–4896.

[20] N. M. Nayak, Y. Zhu, and A. K. Roy-Chowdhury, "Vector field analysisfor multi-object behavior modeling," Image Vis. Comput. , vol. 31,nos. 6–7, pp. 460–472, 2013.