ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 6, Issue 8, August 2017

Detection of the Motion of A Person Where The Activities Were Done In The Video

M. Lakshmi Tulasi and Dr.P. Rajesh Kumar

Abstract—this is a multi-functionality system that proposes in front of you a framework which is discriminative and coherent in nature. With the help of this framework, multiple people can be traced simultaneously. Whatever activities are being performed by the human beings is estimated in a collective manner. Unlike other systems where each person is treated separately, our model has the power to calculate out the motions of two or more people together. There is an intuition within the system that between motions of two people, a strong correlation does exist. Their activities relate to each other in some or other manner. The person's motion, activities, behaviour etc are being correlated with other nearby person's activities and motions. In spite of directly dealing with the respective solutions to the problems, the system aims to work out in a hierarchical manner. A proper and systematic hierarchy of the activities is being created where a natural progression is being generated. With the help of this progression, the motion of a particular person is dealt out with that of a complete group. A graphical based model is being proposed out which works out in a two level hierarchy. With the help of this hierarchy, the relationship between the tracks is recognised and along with this, simultaneous activity segments arealso tracked out. An algorithm is also being proposed out which helps in solving out the entire mechanism. The entire inference joint related problem is being eloped out with the combination of the propagation scenario. The different versions of the branch and bound algorithms are being used along with the integer based programming. Unlike other papers where modeling of the activities was done individually in the videos, this research paper works on joint model working. The related activities are being recognised in a scene with the help of the present features in the context and the motion. This task is performed on the basis of the fact that activities related in space and time rarely occur independently and these activities can widely serve out as important context for each other. A conditional random model is being introduced which is a two layer architecture. In this model, the activities along with the segments are represented in a hierarchy. With the help of this model, different motions and features can be integrated at different levels and the statistics can be learnt in an automatic manner.

Index Terms— natural progression, activity segments, joint related problem, random model, statistic

1 INTRODUCTION

The surveillance videos that are being present within an unconstrained environment are having sequences which are available with activities of long sequences. These all activities tend to occur at different- different locations which are spatio temporal in nature. There are multiple people which are being involved within the entire activity sequence. All the activities are being related to each other in a contextual manner. A novel method is being proposed out in this paper which helps in capturing out the context within the activities with the help of the Markov random field. During the gradual test time, the structure of the Markov random field is being improvised and this structure is not predefined. When a collection of the videos is being provided along with the set of classifiers which are weak in nature for each and every activity, the spatio-temporal relationship that exists within the activities is represented with the help of probabilistic edge weights in the Markov random field. A generic representation is being provided by the model for the entire sequence of the activity. With the help of this model, we are trying to show that activity recognition within a video can be depicted out over the graph as a form of interface problem.

The experiments are being conducted over the UCLA official dataset which is available publicly. This is done so as to demonstrate out the improvements that need to be done in order to recognise the accuracy with the help of the proposed model. The natural scenes which are realistic, noise within the images, motion of the object, entertainment and high dimensionality of the pose are some sort of challenges that are to be addressed out. A class of approach is being proposed out in this thesis where the objects are being modelled out with the help of state graphical models which are continuous in nature. It is being shown that with the help of these approaches, the complex objects can be effectively modelled out. This is done with the help of the inference algorithms which are tractable and robust in nature. These algorithms are able to infer out the exact pose of the object within the presence of the variations. To model out the rigid as well as the articulated objects, continuous state graphical models are being used. In these models, nodes represent the part of the objects and the constraints between the parts are represented with the help of the edges. In comparison to the traditional methods, there are numerous advantages to this model. The very first is that inference algorithms are allowed withinthese.

models. Thus, it becomes easy to scale out in a linear motion with the parts of the body. It is done by breaking up the high dimensional search into many small and minute low dimensional searches. Another advantage is that the partial occlusions can be easily and robustly dealt out with the help of the propagation of the spatial information within the parts.

Images and video help in providing the cues related to the scenes which are low level and rich in quality. Also the cues are offered in regards with the objects associated with the scenes. Major goal that relates out with the vision of the machine is that an approach can be developed out which is helpful in extraction of semantic knowledge that proves out to be meaningful with the help of these cues. This proves out to be challenging as the variability exist within the objects in a very high ratio. Different objects vary in shape, size, colour etc. Therefore, these objects have motion which is highly complex and is governed out with the help of environment having many physical interactions. Due to all these

challenges, it becomes difficult to find out the exact regions of the image for an object and also the various parts of the object.

We will take into consideration the issue of tracking out the people in an automatic mode. This process will be done first by inferring out the pose of a person. The next step will be to incorporate constraints within an inference framework which is highly collaborative in nature. The contribution is made up within all the aspects depicting the issues by addressing out the choice of modelling, priorities and the inferences.

2 RELATED WORKS

The very first approach within this work was the use of the video structures. It was done so as to properly evacuate out the scene boundary candidates from the shot boundaries. The MCMC method was then implied so as to get the scene boundaries which are true out of these candidates. Due to this, the segmentation of the highly accurate scenes becomes easy and possible. One thing to note out is that when the when the probability of the total number of scenes is given correctly prior to the process, then the MCMC process can propose out more exact results.

Thus, the second approach goes with the same tactics. With the help of Multiple Regression Analysis (MRA), the parameter that is being used for the task of prior probability is already set to some optimal value. In this approach, the graphical methods are being used so as to properly encode out the relationships within the analysis of the video. So as to recognize the activities that are complex, spatio-temporal relationships play a very important role. A general framework has been proposed out in this paper for the segmentation of the temporal video. Markov chain Monte Carlo (MCMC) technique has been used to accomplish this task.

The methods that have been developed in the past are mostly dependent over the global thresholds which are already fixed. This is not required in many of the cases. Many times due to fix value of thresholds, the problem of over-segmentation or under-segmentation can be generated. Moreover some of these methods tend to use knowledge which relates to a particular domain which thus in some or other way is not helpful for the use with other domains. This is the reason why it is not easy to use these methods as the generalised ones over some other domains. We are not going to use any fixed value of threshold and also no information about the structure will be used from the video. An iterative method has been devised so as to evaluate the parameters related to segmentation. It includes the total number of segments related with the scene along with the respective locations of those scene segments. In our proposed system, if in any case the scene segment number changes then vector dimensions need to be changed completely that holds the boundary location. For the task of direct analytical computation, the solution space that withholds these two parameters becomes very complex.

Thus, a statistical fashion approach is being used to get out the estimate of these two parameters. It is done with the help of the Markov chain Monte Carlo (MCMC) technique. There are many applications which are using the Markov chain Monte Carlo (MCMC) technique for the various tasks like image processing, video content analysis and computer vision. The system applied the concept of MCMC technique first time for the task of image analysis. It was done with the of the Gibbs sampler. Grenander et al was the person who introduced the concept of jump and diffusion method within the MCMC Technique. Then the new concept of reversible jumps was further proposed by Green. This concept has been used for the operations like learning and sampling.

Phillips et al introduced the concept of change point problem for the tasks associated with 1- D signal segmentation. An EM based technique was being proposed out by the existing system to solve out the problems called as structure-from-motion (SFM) problem where the correspondences are not known.

To generate out the samples of the assignment vectors, MCMC algorithm with symmetric transition probabilities was used. This was done so as to get the vectors in relation to the feature points within each and every frame. To get the estimate of the posterior distribution of the disparity, process of MCMC sampling was applied. The data-driven Markov chain Monte Carlo (DDMCMC) has been applied by the MCMC sampling process till the range image and the optical image segmentations.

There are three different types of updates presented by the Markov chain method being proposed out. These are shifting of boundaries, merging of two adjacent scenes and the splitting of one scene into two scenes. With the help of these updates, along the different parameter spaces you can find out the solution jumping. At this time, the parameter vector dimension can be changed. Also, the solution can get diffused within the same space and hence in this case the elements present within the parameter vector can be changed and also in this case the dimension of the vector won't change in any case.

There is an assumption that for each shot that is present within the boundary must be declared out as a scene boundary. The shots that are having likelihood higher than others are able to coincide with the boundaries that are true. In the initial stage, two random segments are been taken and both of them are being separated with the help of a shot which is selected randomly. Then the next step is the updation in the process of MCMC. During this, several shots are taken into consideration and they are been declared as the boundaries of the scenes. So as to avoid the misdetections that can occur possibly due to the single chain, several Markov chains are executed independently. Then the next task is to collect out all the samples from the chain. This is done so as to compute out the likelihoods of the shots. At last, the shots that are computed out as the one with highest likelihood are given the task of the scene boundary locations.

The major advantage of using the concept of sampling is that it is easy to find out both the weak and strong boundaries. Also, you need not detect any fixed value of threshold. This framework has been applied over the home videos and the feature films and it gives out the results which are highly competitive and very accurate with no chance of errors.

3 OVERVIEW

A two level hierarchical model which is graphical in nature is being presented out in the proposed system. With the help of this model, all the forms of spatio-temporal relationships within the activity segments can be tracked out easily. The relationship within the tracks can also be represented with the help of the proposed model. Over these activity space segments and the traced tracks, the HMRF are being constructed.

The tracks can be easily recomputed and the cost matrix can be updated with the help of the labels that are being obtained from the task of recognition. With the help of the modified tracks, the algorithm keeps on iterating. Due to all this, STIP feature is being used by us in the performance of various experiments.

We have proposed the hierarchical model approach for the unifying framework because it helps in smooth integration of the low level and the high level activities of the human within a video. The graphical models being used over here as an approach are the data structures which are pervasive in nature. They are frequently used in the field of computer science and engineering within the algorithms being introduced. There are thousands of problems in computational world which are being defined and introduced in terms of graph only.

The application areas of the graphical models are wide in nature and there are multiple fields where they are used in a high ratio. So as to represent and depict out a complicated system, graphical models prove out to be a very efficient tool. This complicated system is being devised from a lot of variables. Within a graphical model, whatever the variables are to be focused upon and are of user's interest is represented with the help of nodes. The relationship within these nodes is depicted with the help of links which are also called as edges. These edges help in connecting the corresponding nodes with each other, forming a relation within these nodes. The graphical models are mainly divided into two parts: the directed graphs and the undirected graphs. In the directed graphs, you will find that a direct link is present which shows the cause effect relation between the various nodes. From the cause variable, a direct link is originated and it is being directed towards the effect variable. In case when there is no cause effect relationship to be depicted out, then the concept of undirected graphs is being entertained. Based on the type of problems being characterised, different graphical based models can be formulated out. So as to represent a complicated problem, different classes of graphs can be combined possibly in a systematic manner. Through this dissertation, a hierarchy of graphical model is being presented so as to recognize the actions of the human and interactions within the video. The methods being incorporated undertake the entire processing which spans over the level of pixels. Other than the pixel level, object level, event level and the blob levels are also being entertained. To represent the variables and different relations within them at low levels, undirected graphs are being used. The directed graphs are used at the higher levels like the event level and the object level to show a relation between them.

4 OUR METHODOLOGY

The method we have proposed is a hierarchical graphical model. This model helps in dealing with the related tasks and a proper framework is been presented so as to understand the human interaction in general within a video that is coloured. Some of the main works involved within the research are the integration of the algorithms which are very low level and the ones which are having knowledge of high level. This integration is done in terms of graph related model which follow a proper hierarchy. Here are some of the major contributions that are being provided by the research. [1] To recognize the interaction of two people, a new framework has been proposed out which works on a graphical model format. This gives the recognition details in a proper manner. The network being used is a hierarchical Bayesian network. [2] With the help of the Bayesian network, the ambiguities in the interaction of humans can be easily handled. It is done by the interference of the Bayesian network to remove out the occlusion. [3] For the task of high level descriptive events, a vocabulary is being generated which is very human friendly. [4] A model is being entertained which is stochastic in nature. This is done so as to work out on the task of recognition of the interactions between the human beings. The highest level of representation within the hierarchical framework is the event level. This is done so as to understand interaction of humans with each other in the video sequences.

Within the level of the event, the major focus of the system was to get the full understanding of the semantic measurements of the images within the video frames. If we consider the object level, the major focus was to get the description of the video which is of high quality. For example, the pose of the body can be taken under this. Event semantics is taken into consideration when the event level understanding is talked about. It is required to make an association between the natural language verbs and symbols along with the visual features. This is done so as to make the semantics of the events at the time of the interaction of two people.

A comprehensive framework has been developed out so as to recognize all the actions of the human beings and the interactions done between them. This is done within a colour video with the help of a graphical model. There are four levels at which the activities performed by a human can be represented. These are: the pixel level, the blob level, the object level, and the event level. If we talk about the pixel level, here the subtraction of the background is done and the classification of the colours is performed. Eeach and every pixel are merged into a blob at blob level and the tracking of each and every blog is done at every sequence. At the level of object, the blobs are being associated with the objects being present in the real world. Thus, the approach is basically the appearance based scenario. The entire human body is divided into different parts at the object level. These parts are head, upper body and the lower body. There are further subdivisions into the non-skinny and the skinny areas. On the basis of frame, the estimation of the pose is been done.

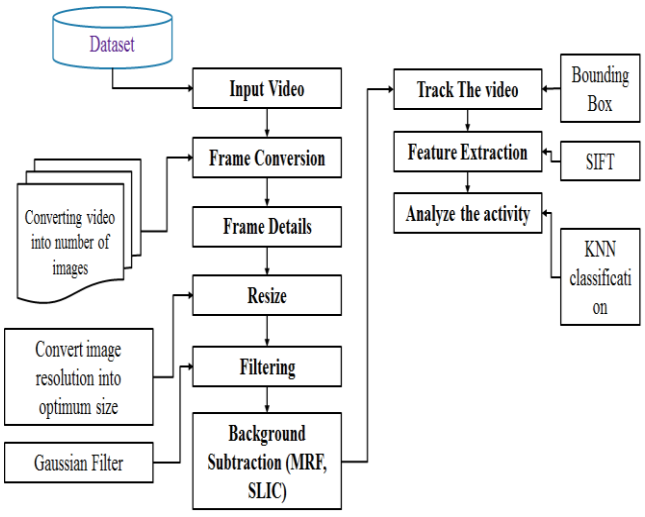
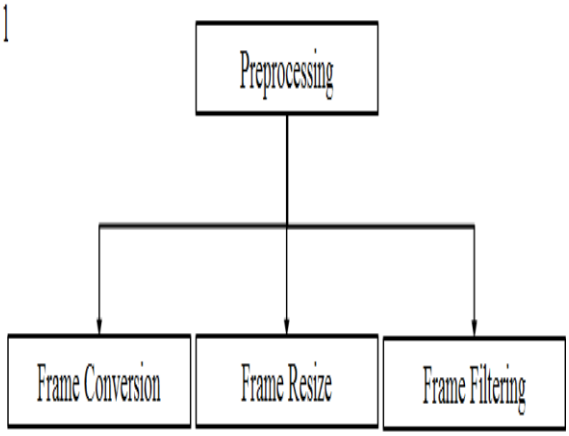
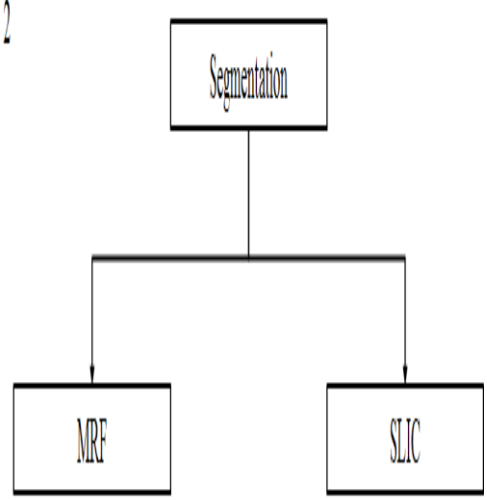


Fig1. Flow diagram showing the complete work flow of the proposed model

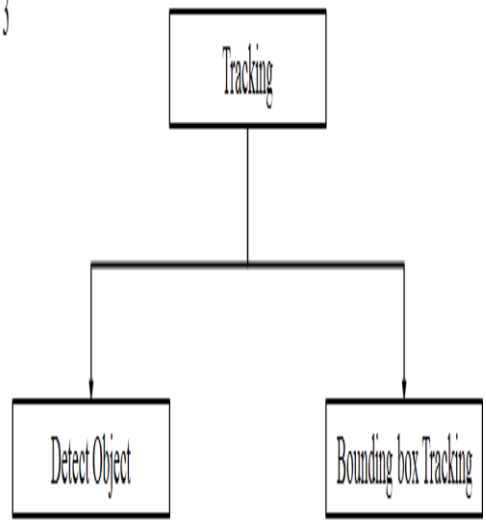
MODULE : 1



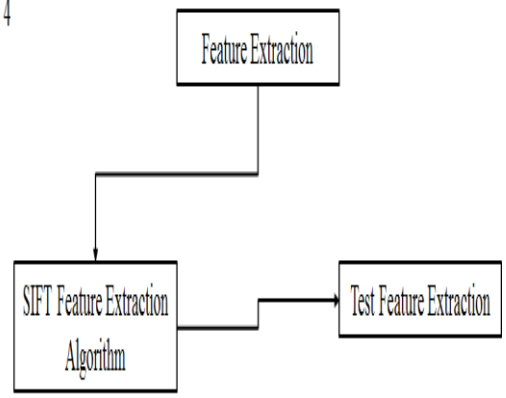
MODULE : 2



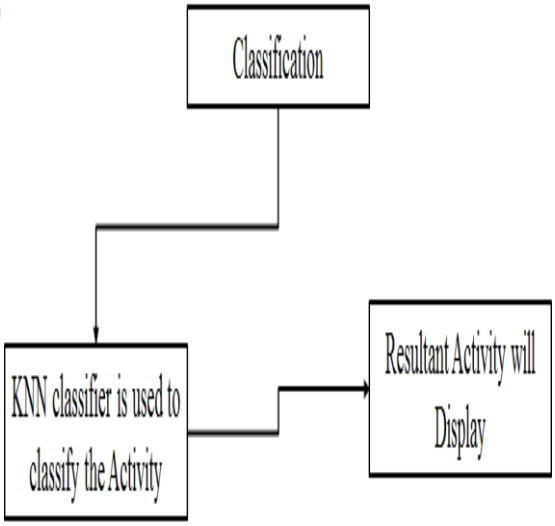
MODULE : 3



MODULE : 4



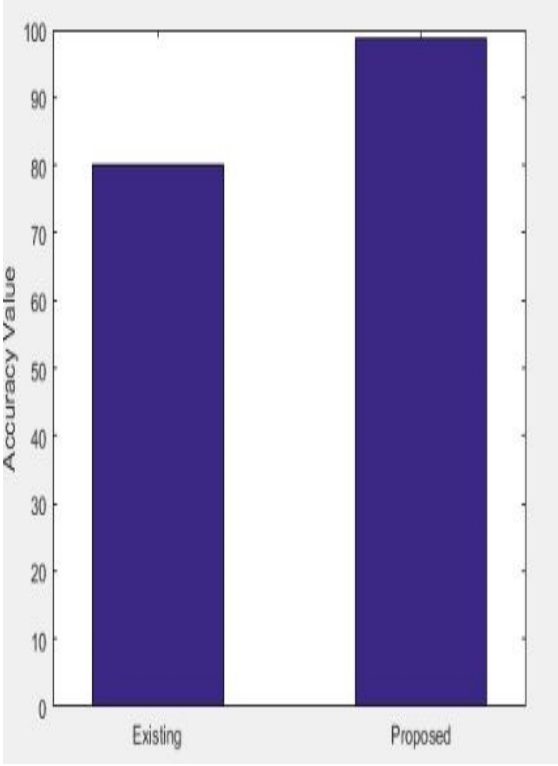
MODULE : 5



All these module figures clear depict the work process being followed within each and every step.

5 RESULTS

The following are the real time results associated with our proposed work.



6 CONCLUSION

With the help of these graphs based approach, the contextual relationships which are spatio - temporal in nature are being depicted between the various activities. Also the total influence of the tracks on these activities is being shown out. When a top down approach is being used, whatever errors have been raised during the bottom- up processing are properly rectified. It has been clearly demonstrated in the paper that one of the best substitutes to the alternative methods like the greedy search etc is the L1-regularized learning of parameters. Through this paper, we have proposed out the methods like tracking of a person, localization done along with the task of recognition of the various human activities with a framework that in integrated to the continuous sequences. To recognise the activities of a human being, gesture representation is very important. The event of the evolution of the instant poses is known as gesture. Interaction hierarchy has been introduced at the event level. Within this hierarchy, the interaction of two persons is being defined in terms of whatever actions they perform. A Bayesian network has been developed out so that all the gestures can be together combined into poses. A rule based decision tree is used to classify all the interactions. There are several factors over which the performance of the entire system depends. On the basis of the robustness of the low level based operations, the high level operations are being performed.

REFERENCES

[1] M. R. Amer and S. Todorovic, "Sum-product networks for modelingactivities with stochastic structure," in Proc. IEEE Conf. Comput. Vis.PatternRecognit., Jun. 2012, pp. 1314–1321.
[2] W. Brendel and S. Todorovic, "L earning spatiotemporal graphs ofhuman activities," in Proc. IEEE Int. Conf. Comput. Vis. , Nov. 2011,pp. 778–785.
[3] V. Chandrasekaran, N. Srebro, and P . Harsha, "Complexity of inferencein 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 trackingand 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 offeature 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 andclassification 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 networksfor 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-featuresfor 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 anGool, 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 forhuman 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 hierarchicalinvariantspatio-temporal features for action recognition with indepen-dent subspace analysis," in Proc. IEEE Conf. Comput. Vis. PatternRecognit. , 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 structure for wide-area activity analysis in unconstrainedenvironments," 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 fieldanalysis 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.
[21] S. Oh et al., "A large-scale benchmark dataset for event recognition insurveillance video," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. ,Jun. 2011, pp. 3153–3160.
[22] S. H. Park, S. Lee, I. D. Yun, and S. U. Lee, "Hierarchical MRF of glob-ally consistent localized classifiers for 3D medical image segmentation,"PatternRecognit., vol. 46, no. 9, pp. 2408–2419, 2013.
[23] M. Pei, Y. Jia, and S.-C. Zhu, "Parsing video events with goal inferenceand intent prediction," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011,pp. 487–494.
[24] M. S. Ryoo and J. K. Aggarwal, "Recognition of composite humanactivities through context-free grammar based representation," in Proc.IEEEComput. Soc. Conf. Comput. Vis. Pattern Recognit. , 2006,pp. 1709–1718, Jun. 2006.

AUTHORS

\M. Lakshmi TulasiStudying M.Tech, Final year in VLSI & ES Specialization, ECE Department, VelagapudiRamakrishna Siddhartha EngineeringCollege,Kanuru, Vijayawada-7, AffiliatedtoJawaharlalNehru Technological University Kakinada, Andhra Pradesh.

Dr. P. Rajesh Kumari,Ph.D(IITM), Head of Department, ECE, PVP Siddhrtha Institute of Technology, Kanuru,Vijayawada-7, AffiliatedtoJawaharlalNehru Technological UniversityKakinada, Andhra Pradesh.