RobustVideo Labeling with Voice Recognition System

K.PRIYANKA[1]P.SABARINATHAN[2]
PG Scholar, Dept. of CSE, Pavendar Bharathidasan College of engineering and technology
Trichy, Tamilnadu, India[1]
Assistant Prof, Dept. of CSE, Pavendar Bharathidasan College of engineering and technology
Trichy, Tamilnadu, India[2]
priyankakamaraj@gmail.com[1]
sabarime09@gmail.com[2]

Abstract:**The main theme is focusing on the video labeling with voice labeling with voice announcement. The voice announcement concept is used to recognize the accuracy of the human action. It is used to declare the two level hierarchical graphical structure. It is used to learning the relationship between the tracks and corresponding labeling activities. The level of hierarchy in the relationship between the tracks and activity labeling segments are used to exploit the increasing levels of robustness. An L1 regularization structure play a major role in the activity segments. The voice announcement is used for the voice action. It using the bidirectional approach and used for blind peoples. The labelling concept is used for the defundum peoples. The text should be composed the utteranceformat. The converted into the utterance composing of phonemes and wave form generation whole process completed it produces the voice announcement. Speech synthesis is used for the human speech by artificial production.**

Keywords: L1 regularization, HMRF, Text to speech, Speech synthesizer.

I.INTRODUCTION

The framework focused to the simultaneous tracking and activities labeling in the continuous video sequence with voice announcement. It is used to integrating the bottom up and top down approaches in the learning structure. It introduces the two factors are feed forward and feed backward approaches.A continuous video has lot of activities in that particular video sequence. The activities are first identified and detects the tracks in each and every moment in the video sequences. The activity analysis is used to solving the tracking and recognition problems. In ancient days, video tracking analysis faced two problems. Continuous video or any cartoon videos, the find out the solution and previous solution is used to get the new solution and it gives first preference in tracks. It helps to detecting and recognizing the activities. In video sequence, the location of the movement used to find human activities or any other activities. In two important process are used in the labeling activities. A bidirectional approach is used to integrate the correcting the noise and detecting the false. In bidirectional approach refers the two terms bottom up approach and top down approach. The bottom up approach perform the adjusting the task and top down approach is used to improve the task. The bidirectional algorithm is used to increase the task accuracy process. The main theme in the process is to tracking and labeling of the activities in the video sequence.The nodes in which the lower level of hierarchical graphical model is used to identifying the relationships over the track lets.

The nodes in which the higher level finding the information related activities. The both process like detection and recognition is used in HMRF with L1 regularization algorithm. The main intension of thebidirectional approach is assuming the particular tracks and labeling the activities. A unified framework is used to track the multiple people; it localizes and labels their individual activities, in complex their long duration video sequences. It perform two important role first, the influence of tracks on the activities. Second, the structural relationships based on the activities. If the relationship between the tracks, and their same activity segments. An L1 regularized structure learning approach is used in the tracking and labeling process. A HMRF is another approach used for detects and track the activities. A bidirectional approach used for integrating the bottom up and top down processing. The bottom up processing recognition of activities using computed tracks top down approach is used to improve the tracks. The two level of graphical structure defined the inter activity context. The activity recognition and labeling the activities used for hierarchical markov random field (HMRF). The speech synthesizer is used to convert the text into the voice process. Theanother name of speech synthesizer is also known as tokenization. Theconsists of two ends like front end and back end. The text like verb, phrase clause and sentences.

The front end of the synthesizer used to denote the gather the information of the text. The target is should envelope the pitch, phenomenon speech. It is used to develop by the software and hardware. The speech can be transcriptions of the speech conversion. In the speech can be stored by the database and lots of speech

can be combined to form the synthesized speech. The output range uses the phone and diaphones during the contribution of the voice tracks. The incorporate model of the voice tracking process in synthetic voice output identification. The quality of the speech is the ability of understood the human voice clearly. A text to speech program is used to available for the visual impairments of the disabilities works on the personal computer and any other computer. Lots of operating system is used for the speech synthesizer. In the blind people does not able to see the materials or any videos. The blind peoples does not able to check the email and any other websites. The using the process GPS has to maintain the hand held devices giving the portable directions. The provide the robotic sound process like the natural and very high speed. The Computation technique is developed by the text to speech and other contribution in the process of blind peoples. The tracks forming the track lets and measuring the accuracy of the frame and playing the video sequences, the contribution process envelope to the viewing hearing the process. The process should be able to envelope the process of labeling activity.

II. **RELATED WORK**

In the related work focus on probably two main functions like tracking and recognition. In the ancient days localization and classifications are researched. But it is used for only single person or any activities. A system used to combine the labeling activities during the lack of contextual information.In the structure learning is used to both bottom up and top down approaches. The bidirectional approach is used to varying the labelling activities. The tracking of framework is mainly perform the multiple video sequences using classification and localization. Activity recognition and tracking is used for object interaction. The related dynamic Bayesian network is used for tracking process between the in object in interaction from activity recognition. Large college of activity and tracking involving the multiple sequences. The video analysis sequences is encoding by the graphical models. The complex activities in the video sequences is using the techniques stochastic and context free grammar. The hierarchical MRF act as a major role in image segmentation HMRF is used for variation of labeling and activities. Spatio temporal relationship is a used for detecting the complex activities in the video sequences.

The graphical models was captured by bidirectional approaches. The spatio temporal graph is representing the multi scale video sequences. The MRF technique is built by spatial temporal context and labeling activity is founded. In higher and lower level of representation is not explained in their approach. The well-known activity location assumes the context of hierarchical model and it is a corporate tracking. It can notify the particular single action compared to group activity. It is used to develop the tracks using recognition scores. The graph structure is challenging to explain. It also used the greedy forward approach. The main role of the approach is used to find out the possible solutions graph structure. It is used to learning the optimal structural relations with parameter. An L1-regularized relations with parameter. An L1-regularization is used to learning the optimal structural is used to create the graph structure and reduce the noise.The spatio temporal configurations in the primitive actions is declared by the complex human activities videos. It is used to estimating the activity recognition for relatively significance.

The main contribution of the process is properly identified the human activity and related videos. It is used to capturing the hierarchical structure of the graph. The spatio temporal is used for video parsing, detecting the learning activities. The connected graph of the pixel is inference due to the process as well as the focusing the neighboring the unary computation.The inference in fully connected with graphical method and CRF methods corresponding to the process in which the exploring the voice announcement process. The testing videos of activities by the generating the globally optimization labels. It is used to view unconstrained optimization of convex in the problem and predict the view of correcting the labels of instance labeling of multiple activities. The learning model and generating labels are used to detect the detection of anomaly video process. The collect dataset from promising results from the VIRAT ground dataset of joint modeling and recognition process in the activities of the wide area scenes of effectiveness of anomaly detection. A network flow is used to optimize for data association of multiplesobject tracking method. The maximum a posteriori (MAP) is defined the mapping the cost- flow network with non-overlap constraint.

A minimum cost flow algorithm is associated due to the network and anomaly detection process. The explicit occlusion model (EOM) including the network augmentation of network track with long term inter object occlusions. The solution of the EOM is based on the network to define by an interactive built up approach. Initialization and termination process act as a major role in the false observations and intrinsically formulation of the process. The efficient method does not requiring the pruning hypotheses method. The performance should be compared to be previous results on the public pedestrian datasets and used to improve the process.

III. PROPOSED ALGORITHM HIDDEN MARKOV MODEL

A hidden markov model is defined as the markov chain and it is used for observed the partial state of the system process. The observations are used to relate the state of the system and the insufficient state of the following aspects of the contribution between the algorithms. The process are completed the hidden markov model can be existed. The observation of the Viterbi algorithm is used to compute by the most-likely the state of the sequences. The feed forward algorithm is used to applying the probality of the sequence in the particular observation. It is also observed the transition function and observation function of the hidden markov model. The speech of recognition is used for the observed data in the audio waveform and the hidden state of the spoken text. The Viterbi algorithm is used to identifying the sequence of the voice announcement.

L1 Regularization

It is used to maintain the graph structural and parameters. The L1 regularization algorithm consist of parameter sets. There are three important parameter graph edges. The parameters are concatenating the weighted vectors. All the parameter are connected the graphical model. All the nodes are connected to each and every node of the graph. It is basically build on spatio temporal method. A parameter of a sparse set contains the node parameter. The important contextual information are encoding the non-zero edge. It is accepted by the L1-algorithm of the given instruction. The nodes are represented by n and edge is denoted by edge parameter. The joint distribution of node can be identified by parameter edge. Inference algorithm is used to find out the hidden variables. The two important steps in the inference algorithm, EM framework and bottom up inference strategy. The top down approach is used for re-computation of tracks. It have two important process are bottom up activities and top down activities.

$$\min_{f} \sum_{i-1}^{n} V(f(\hat{x}_i), \hat{y}_i) + \lambda R(f)$$

The bottom up inference is used for estimate the activity labeling. A consecutive actions are getting the same labeling activity. The HMRF is used to create the track let formation of labeling activity.

The R (f) is used to generating the loss of function and V is used for the cost of predicting, f(x) when label is y and F is used to control the restrictions of the smoothness and bonds vectors. Regularization is used to the learning the simplier the models and inducing the models to be sparsing introducing the group of structure into the learning problem description.A theoretical justification used for regularization is attempts to impose Occam's razor on the solution. Many regularization techniques corresponding to the learning the algorithm.The main goal of the bidirectional approach is labelling the activities and activity recognition in the video sequences. It select the challenging video sequences or any other experimental video sequences. It analyse the full length of the video. It is used to viewing the sequence of the activity.

It is represent the graphical structure. The graph consists of two levels of nodes. Each level having individual values. The HMRF is act as a major role in video sequences. The nodes are denoted by n and the parameter of the node is denoted by n2. The every node is connected to the potential observation. In the video sequence, the images can recognized in each movement represented by the track let. The lower level node link is connected to the higher level node. If the long duration video sequences are compressed by the size and the video sequence small activity can be easily identified the labeling. . The bidirectional algorithm is used to increase the task accuracy process. The main theme in the process is to tracking and labeling of the activities in the video sequence.

IV. PSEUDOCODE

choose initial point x^0
$S \Leftarrow \{\}, Y \Leftarrow \{\}$
for $k = 0$ **to** MaxIters **do**
  Compute $v^k = -\diamond f(x^k)$ (1)
  Compute $d^k \Leftarrow \boldsymbol{H_k} v^k$ using S and Y
  $p^k \Leftarrow \pi(d^k; v^k)$ (2)
  Find x^{k+1} with constrained line search (3)
  if termination condition satisfied **then**
    Stop and return x^{k+1}
  end if
  Update S with $s^k = x^{k+1} - x^k$
  Update Y with $y^k = \nabla\ell(x^{k+1}) - \nabla\ell(x^k)$ (4)
end for

Step 1: Calculate the total length of the video sequences.
Step 2: Each tracks should be identified.
Step 3: Tracks are formed known as track lets.
Step 4: Identify the tracks and labeling the activities.
Step 5: Identify the activities using bidirectional approach
Step 6: An HMRF algorithm is used to creating graphical structure.
Step 7: An L1 regularization is used to reduce the noise
Step 8: Where $Xk+1_t$ denote the lower level of node and Xk_a denote the each segment of the activity. T represent the tracks.
Step 9: End.

A unified framework is used to track the multiple people; it localizes and labels their individual activities, in complex their long duration video sequences. It perform two important role first, the influence of tracks on the activities. Second, the structural relationships based on the activities. If the relationship between the tracks, and their same activity segments. An L1 regularized structure learning approach is used in the tracking and labeling process. A HMRF is another approach used for detects and track the activities. A bidirectional approach used for integrating the bottom up and top down processing. The bottom up processing recognition ofactivities usingcomputed tracks top down approach is used to improve the tracks. A bidirectional approach is used to integrate the correcting the noise and detecting the false. In bidirectional approach refers the two terms bottom up approach and top down approach. The bottom up approach perform the adjusting the task and top down approach is used to improve the task The nodes in which the lower level of hierarchical graphical model is used to identifying the relationships over the track lets. The nodes in which the higher level finding the information related activities. The both process like detection and recognition is used in HMRF with L1 regularization algorithm. The main intension of the bidirectional approach is assuming the particular tracks and labeling the activities.

A continuous video has lot of activities in that particular video sequence. The activities are first identified and detects the tracks in each and every moment in the video sequences. The activity analysis is used to solving the tracking and recognition problems. In ancient days, video tracking analysis faced two problems. Continuous video or any cartoon videos, the find out the solution and previous solution is used to get the new solution and it gives first preference in tracks. It helps to detecting and recognizing the activities. In video sequence, the location of the movement used to find human activities or any other activities. In two important process are used in the labeling activities.

The related dynamic Bayesian network is used for tracking process between the in object in interaction from activity recognition. Large college of activity and tracking involving the multiple sequences. The video analysis sequences is encoding by the graphical models. The complex activities in the video sequences is using the techniques stochastic and context free grammar. The hierarchical MRF act as a major role in image segmentation HMRF is used for variation of labeling and activities. Spatio temporal relationship is a used for detecting the complex activities in the video sequences. The graphical models was captured by bidirectional approaches. The spatio temporal graph is representing the multi scale video sequences. The MRF technique is built by spatial temporal context and labeling activity is founded. In higher and lower level of representation is not explained in their approach. The well-known activity location assumes the context of hierarchical model and it is a corporate tracking. It can notify the particular single action compared to group activity. It is used to develop the tracks using recognition scores. The graph structure is challenging to explain. It also used the greedy forward approach. The main role of the approach is used to find out the possible solutions graph structure. It is used to learning the optimal structural relations with parameter. An L1-regularized relations with parameter. An L1-regularization is used to learning the optimal structural is used to create the graph structure and reduce the noise.

V. SYSTEM DESIGN AND IMPLEMENTATION

The implementation is defined the labeling concept and voice announcement of the process. The video concept is used to the defundum people and voice announcement is used for the blind peoples. A video generate the HMRF and the labeling of the video length. Finally testing the video and playing the video. If the testing the voice announcement with the speaker. It is used to eliminate the noise and also correcting the tracks.The main goal of the bidirectional approach is labelling the activities and activity recognition in the video sequences.

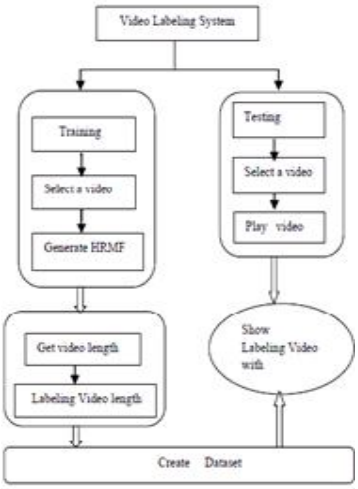

Fig1.System Design

VI. **CONCLUSION**

The evaluation of tracking using the bidirectional approach and no priority of the research has been provide the result on labeling the activity. It gives the formation of track and opposite the ground truth (GT) and compiling the tracking results. The tracks are measured the tracking accuracy using the metrics. Single track should be divided into the multiple tracks. It is denoted the performing the track and bidirectional approach. The technique will be used in the future framework and it is used to store the graph by using representing the list of representation.it can be self-explained by the pseudo code of the self-explanation. The sub graph framework can be used for labeling the edges.If take two data set like VIRAT and UCLA, VIRAT representing the classification results and it faces the many challenging activities.it involve the full dense of depth in the sequences. UCLA represent the unique identification of the sequence. It used to localize the object foreground and labeling activities. If the larger video sequence should be divided into small videos. It represent the joint activity labeling sequences. It analyse the full length of the video. It is used to viewing the sequence of the activity. It is represent the graphical structure. The graph consists of two levels of nodes. Each level having individual values. The HMRF is act as a major role in video sequences.

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 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 svideos," 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 ACM Int. Conf. Image Video Retr., 2007, pp. 494–501.

Author Profile

K.PRIYANKA, PG scholar, Dept. of CSE, Pavendar Bharathidasan College of engineering and technology Trichy, Tamilnadu, India
P.SABARINATHAN, Assistant Prof, Dept. of CSE, Pavendar Bharathidasan College of engineering and technology Trichy, Tamilnadu, India.