

# OBJECT DETECTION AND TRACKING IN THERMAL VIDEO USING DIRECTED ACYCLIC GRAPH (DAG)

Supriya Mangale, Ruchi Tambe and Madhuri Khambete

*Department of Electronics and Telecommunication Engineering, Cummins College of Engineering for Women, India*

## Abstract

*This paper suggests an incipient approach to perform target detection as well as tracking for single and multiple moving objects in thermal video sequences. Thermal imaging is complimentary to visible imaging as it has capability to detect object in low light or dark conditions by detecting the infrared radiation of an object and creating an image which contains temperature information. The extracted regions are then used for performing the segmentation of targets in thermal videos. In projected method first, Directed Acyclic Graph (DAG) is used for segmentation in thermal videos. Second, to enlarge the set of target proposals, DAG is initialized with an incremented object proposal set in which, from adjacent frames motion based predictions are used. Last, in this paper for selection of the specific object motion scoring function is used, which is having high optical flow gradient between the edges of the object and background is presented. After segmentation of object, centroid based object tracking is performed to track the objects in thermal videos. The proposed method is evaluated on different thermal videos and found to be robust compared with standard background subtraction method.*

## Keywords:

*Thermal, Directed Acyclic Graph, Segmentation, Moving Objects*

## 1. INTRODUCTION

Thermal imaging amends Thermal imaging amends overtness of the objects in a dark environment by identifying the objects infrared radiation and then engendering an image predicated on temperature information. Near infrared illumination, low light imaging is withal most commonly used night vision technologies. Contrary to these two methods thermal imaging has ability to work in darkness. Thermal imaging works better in conditions as haze, fog and smoke.

In thermal imaging all targets emanate heat as an element of their temperature. The warmth radiated by question is kenneed as its heat signature. As a rule one can verbalize, the sultrier a Question is the more radiation it emanates. A thermal camera is basically a warmth sensor that is fit for recognizing infinitesimal contrasts in temperature. The contrivance accumulates the heat from objects in the scene and engenders an image predicated on information about the temperature differences. Since items are rarely an indistinguishable temperature from different protests around them, a thermal imager can identify them and they will show up as various in a thermal picture.

Thermal images are generally grayscale in nature where ebony targets are arctic, white objects are sultry and the depth of gray shows variations between the two. Some thermal cameras, integrate color to images so it can avail users identify objects at sundry temperatures. Due to the ability of having better contrast and to detect targets at extremely long distances, in low

illumination and in extreme weather conditions, thermal imaging is proper for various applications.

Also one can say that thermal images are visual displays of emitted, reflected, and transmitted thermal radiation within a given area. When presented to an operator, color maps are often used to map pixel intensity values in order to conceptualize details more explicitly. Discovery in thermal imaging was chiefly utilized as a part of military applications. Due to the better quality of image and better resolution with reduced cost and size during recent years opened up new application areas.

Thermal cameras are used in many applications because they have the ability to see in total darkness, their robustness to illumination changes, and less intrusion on privacy. Due to various sources of thermal radiation, thermal imaging is challenging depending on the properties of the object and its surroundings. The amount of radiation taken by the object is based on emissivity. Also, thermal radiation from other objects is reflected on the surface of the object. Therefore, the reflectivity of the object is also important to be known.

The quantity of radiation that reaches the detector is overdone by the atmosphere. Some is transmitted, some is absorbed, and some is also emitted from the atmosphere itself. Moreover, the imager itself emits thermal radiation during operation. To measure thermal radiation and temperatures as accurately as possible, all these effects are need to be considered. At short distances, atmospheric effects can be ignored. But for greater distances, e.g. from aircrafts, it is important to consider atmospheric effects if temperatures are to be measured correctly. The Fig.1 shows the example of thermal image. However, if one is only interested in an image that looks good to the eye and not temperatures, these effects are not taken into account.



Fig.1. Example of Thermal Image

Materials have not identical properties in the thermal spectrum compared to the visual spectrum. Some materials that are reflective and/or transparent in the visual spectrum are not the same in thermal spectrum and vice versa, e.g. glass. From the point of measuring temperatures, thermal imaging is beneficial compared to point-based methods as temperatures over a large area can be compared. Compared with visual cameras, thermal

cameras are complimentary as soon as there is a temperature difference connected to the object or phenomena that is to be detected. For example, emerging fires, humans, animals, increased body temperatures, or differences in heat transfer ability in materials. When applications are considered, thermal cameras are especially complimentary to visual cameras in outdoor applications. Thermal imager can produce an image with no or few distortions during darkness and/or difficult weather conditions such as fog, haze, snow. This is because of the way that a thermal imager is touchy for transmitted radiation, even from generally cool items, as opposed to a visual camera that measures reflected radiation and therefore relies on light.

Pedestrian identification in thermal imaging is different than that of visible imaging. In all the seasons output of visible camera remains same as it does not take temperature information as its input. As visible cameras are dependent on light information, it does not perform well at night. In opposite to this thermal cameras make use of temperature information. Because of extremity reversal amongst foundations and people on foot saw in summer, infrared division calculations accepting hotter walkers against colder foundation should not be used for a system that needs to operate reliably in summer scenario.

Most of the approaches including Directed Acyclic Graph (DAG) [19] were used for segmentation in visible imaging. In this paper the projected approach is used for segmentation in thermal video sequences. Generally the observation is that thermal images have better contrast as compared to visible also the shape and appearance of any object varies slowly frame wise. From this observation one can say that the object which is having the highest score of similarity and objectness is the object of interest. To calculate the similarity of the target, optical flow method is used and the variation between the actual target shape and detected target shape is calculated. At this end for tracking, the centroid is calculated of each segmented object in thermal video and using the Euclidean distance that object is tracked throughout the video. When the objects are having high optical flow gradients between their boundaries and background then based on the motion of the object the 'objectness' is known. If the time domain sequence of the object is not taken into consideration [17] then there will not be any object related with the first object in no. of frames. The projected technique uses the object proposal expansion criteria. This method was applied on various challenging thermal videos. The Fig.2 shows object proposals for visible as well as thermal video frames.

In this paper, projected approach is used for segmentation in thermal video sequences. Section 2 gives the literature survey of various methods used for segmentation as well as tracking. Section 3 gives the DAG based approach for video object segmentation. Section 4 describes the dataset and predicted results. And at last, section 5 concludes the paper.

## 2. LITERATURE SURVEY

Aim of this paper is to perform the segmentation as well as tracking of single and multiple objects from the thermal video sequences. Target segmentation from the video is well known research area. Segmentation is performed for various purposes like image retrieval, video summarization, etc. There are various methods to perform video object segmentation. For the techniques

given in [1]-[3] for specified frames the notation is required to initialize the algorithm. After this, the methods that are using motion as well as appearance scoring are used to generate the object segments in any frame or in each and every frame. Alternative techniques [4], [5] solely needs correct target region notation in the primary frame and then by using the region tracking remaining frames are segmented in foreground region and background region. Generally the methods which are semi-automatic typically provide sensible results for segmentation. But most of the applications require large data due to which manual low level formatting becomes price prohibitory. Consequently, many automatic techniques were projected to perform the object segmentation in video sequences. These techniques use motion grouping [4], [6], [7] to perform segmentation of objects. Alternative techniques [8]-[10] use appearance of an object for performing segmentation in every frame. After this they use appearance as well as motion of an object for final results of segmentation. Techniques as [9], [11]-[13] present economic improvement structure for spatial and temporal set of pixels for segmentation purpose.

Now a days there are many techniques [14]-[16] which provide a precise idea about how the specific target appears. From above specified three methods the method [14] separates the targets from given images. Lee [17] and Latecki [18] engineered this work and used the separated targets for segmentation purpose. Lee [17] projected that for observing the 1st object by forming the group of the objects from given video sequence and applied the spectral graph clusters for getting the objects inliers and outliers. From this they observed that every inlier is similar with the selected target region. For each selected object from the cluster, the objectness is measured by appearance and motion related ideas. Cluster having high objectness is probably having the first target in the video sequence. The disadvantage of this method is that, it does not consider the sequence of the objects within the video. Therefore it is unable to recognize the shape and location of target over time. Latecki [18] tried to ease this problem by employing the relation between the objects by considering neighboring frames.



Fig.2. Object proposals from video frames

The drawback of choosing the region of the object is sculptured as maximum weighted cliques problem, for finding out the true region of the object from given video frames at same time. As this problem is non-deterministic polynomial time hard (NP-hard) problem [18], the improvement technique is employed for knowing the answer. Also the segmentation techniques based on object proposals [17], [18] are having to extra disadvantages when these techniques were compared with the projected technique. 1st disadvantage is that, for each method the generated target

proposal is not dependent on the target proposal from neighboring frame. Second disadvantage is that the results of segmentation lowers as each method don't actually determines the shape of the target in neighboring frames, once calculating the similarity of the region. The proposed method overcome the drawbacks and performs better on various thermal datasets.

In following sections, the actual framework of DAG for segmentation and tracking in thermal video sequences is given in detail. Also the results of the projected method shown are shown on various thermal video datasets.

### 3. DAG BASED APPROACH FOR VIDEO OBJECT SEGMENTATION

In arithmetic and applied science, the graph which does not contain any directed cycle is known as Directed Acyclic Graph (DAG). That means this graph contains several edges as well as vertices in such a way that it starts from any vertex and follows the directed path and does not go back to vertex from which it was started. The Fig.3 shows an example of Directed Acyclic Graph. In general DAG should follow the topological ordering.

The Fig.4 shows the overall block diagram of the paper. In this, first the thermal video sequence is taken. The moving object in the video is segmented using Directed Acyclic Graph. After segmenting the object the centroid is calculated of the object and by calculating the Euclidean distance the object is tracked throughout the video.

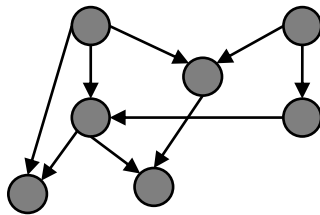


Fig.3. Directed Acyclic Graph

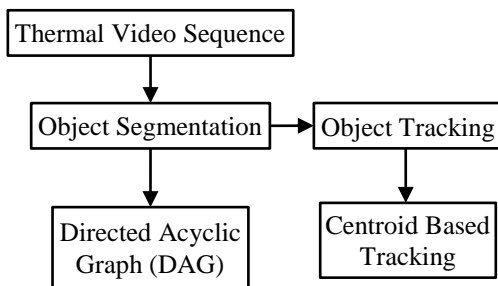


Fig.4. Flowchart of the Proposed System

### 3.1 FRAMEWORK

The projected method works in four stages as shown in Fig.5.

1<sup>st</sup> stage: Create the target proposal for each and every frame using optical flow.

2<sup>nd</sup> stage: Once the target proposals are generated the Directed Acyclic Graph is formed for each target. The objectness is satisfied by the criteria of the longest path. This improves the similarity and objectness of the target. This gives first target in the thermal video.

3<sup>rd</sup> stage: The pixel wise segmentation is done i.e. per pixel video object segmentation by using graph cuts based optimization and the target and background models are build up using Gaussian Mixture Models.

4<sup>th</sup> stage: Calculation of centroid of each segmented object for tracking in single as well as multiple objects using Euclidean distance.

The detail description is given further.

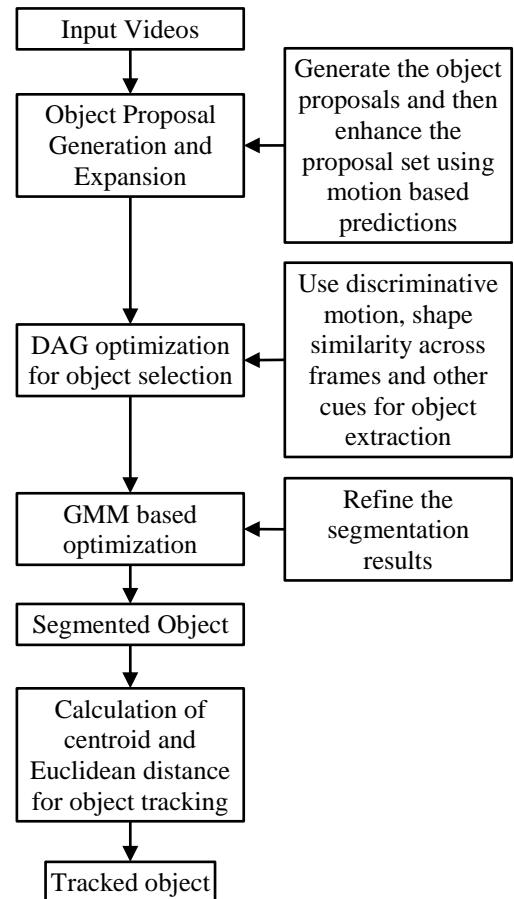


Fig.5. Framework of Proposed System

### 3.2 STRUCTURE OF DAG

We desire to obtain target proposal which are having large objectness probability, more similar appearance also swimmingly changing shape of the target proposals in thermal video. The structure of the DAG approach is formed as given below:

Every target in video frame is portrayed using 2 nodes and 2 edges. The nodes are starting nodes and end nodes and the edges are unary and binary edges. The weights on the unary edges gives 'objectness' of target. There is layered structure of starting as well as ending nodes for a particular frame. Therefore each frame in thermal video has to layers. Unary edges starts at starting node and ends at ending node. The binary edges starts from any ending node and are directed to all of the starting nodes in next frame [19]. The appearance and shape similarity are given by the weights on binary edges.

In Fig.6 [19] the graph structure is shown. In this graph 3 frames are shown as  $i-1$ ,  $i$ ,  $i+1$ . In Fig.5 the nodes are given by blue and green color where blue are starting and green are ending

nodes. Like the nodes edges are given by red and green color where green represent unary and red represent binary edges. It is important to know that for some applications binary edges can be given to next 3 frames.

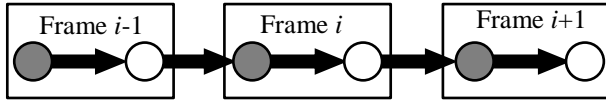


Fig.6. DAG structure

### 3.3 UNARY EDGES

As given above unary edges gives the objectness of the target proposal. To know the objectness motion is essential, therefore the scoring function is given by following formula:  $S_{unary} = Mot(o)$ , where  $Mot(o)$  represents motion score and  $o$  is the target proposal.  $Mot(o)$  is given by average Frobenius Norm [19] which is given as:

$$\|G_x\|_{frob} = \left\| \begin{matrix} g_x & g_y \\ v_x & v_y \end{matrix} \right\|_{frob} = \sqrt{g_x^2 + g_y^2 + v_x^2 + v_y^2} \quad (1)$$

where,  $g_x$  and  $v_x$  are optical flow gradients in  $x$  direction and  $g_y, v_y$  are optical flow gradients in  $y$  direction. The motivation of calculating the motion score is that the motion between object and background are different, since the boundaries of the targets are not having continuous motion. The Frobenius Norm is used to know the magnitude and it should be more at the boundaries of the moving targets. Generally if the gradient will have high magnitude then most probably the area is of the moving target.

For motion analysis in image processing optical flow is widely used method. This method has two assumptions as: First, the brightness of an object remains constant over time. Second, nearby points in an image move in same way which is known as constraint of velocity smoothness. Brightness constancy is very similar to the illumination changes which are available in thermal imagery along with the smoothness constraint as thermal images have good contrast.

### 3.4 BINARY EDGES

As given above weights on the binary edges represent the similarity score of the targets from frames. When DAG is used for segmentation in visible imaging the color feature is also used. As we are performing segmentation for thermal videos the color feature is not used as thermal images are gray scale images. To calculate the similarity between the target regions the properties like location, size and shape are used.

If any 2 regions are having same shape and size also this regions are close to each other than the value will be more and vice-versa. As we are using optical flow to recognize the similar areas or regions this method performs better than [17], [18] to get the similarity of fast moving targets.

### 3.5 PER PIXEL SEGMENTATION

After getting the segmented targets the results are more cleared by using the per pixel segmentation in thermal videos. In this the space and time based graphs are defined by joining the frames temporally with optical flow displacement [19]. In per pixel segmentation the nodes are the pixels in an image as well as

the edges are the neighbors in frame. It has eight neighbors in single frame and eighteen forward and backward neighbors in neighboring frames. For assigning the labels the energy is given as:  $g = [g_1, g_2, \dots, g_n]$  for  $n$  pixels having some knowledge of  $h$ .

$$F(e, h) = \sum_{i \in D} S_i^h(g_i) + \lambda \sum_{(i,j) \in D} U_{i,j}(g_i, g_j) \quad (2)$$

where,  $D = \{s_1, \dots, s_n\}$  set of pixels,  $E$  have adjacent pixels and index of this pixels is given by  $i, j$ .  $s_i$  represents the pixel set containing the values zero and one which means foreground and background pixels. The labeling cost is given by  $S_i^h$ , which is calculated using Gaussian Mixture Model (GMM) for location.

### 3.6 EXPANSION OF TARGET PROPOSAL

Our aim is to get the region of the target in thermal video. To fulfill this goal it is required that the target proposal corresponding with actual target must be present in every frame. This phenomenon is not compulsory but it is preferable. But only the ideas of optical flow or appearance are not sufficient for this purpose. The example is shown in Fig.6 where there is no object proposal present as the actual target although no. of proposals were generated by method [14]. In projected method the assumption is made that the location and shape of an object varies slowly from frame to frame. The enhancement of the target proposal set is made by using the target proposals generated in the neighboring frames. The overall concept is shown in Fig.6 [19]. The objects are clubbed together using the overlap ratio, which is given by,

$$O = \frac{|wrap_{i-1, i}(r_i^k - 1) \cap r_i^j|}{|r_i^j|} \quad (3)$$

For the target wrapped by optical flow in previous frame we look for the target proposal in current frame which is having high overlap ratio. The proposals which all are having the high ratio will be clubbed together to form the new proposal. This will give the correct proposal if it will not be generated by [14]. The Fig.7 shows the example of object proposal generation and expansion using thermal frames.

It is important to notice that the old proposals are not replaced; they are kept as it is so this is the process of expansion. This procedure is an iterative process i.e. it is done for forward and backward in time. The advantage is that even though the useful target proposal is missing it can be produced using this expansion procedure.

### 3.7 OBJECT TRACKING

In the conditions like changing illumination conditions, different shapes of the targets or in case of occlusion it is very challenging to perform tracking. Object tracking is defined by assigning the path to the target in number of frames.

The basic idea of tracking is to segment the interested target and locate that object by keeping the track. In this paper for tracking the objects first the centroid is calculated for each segmented object and using the Euclidean distance the object is tracked throughout the thermal video. The Fig.8 shows the overall framework with an example.

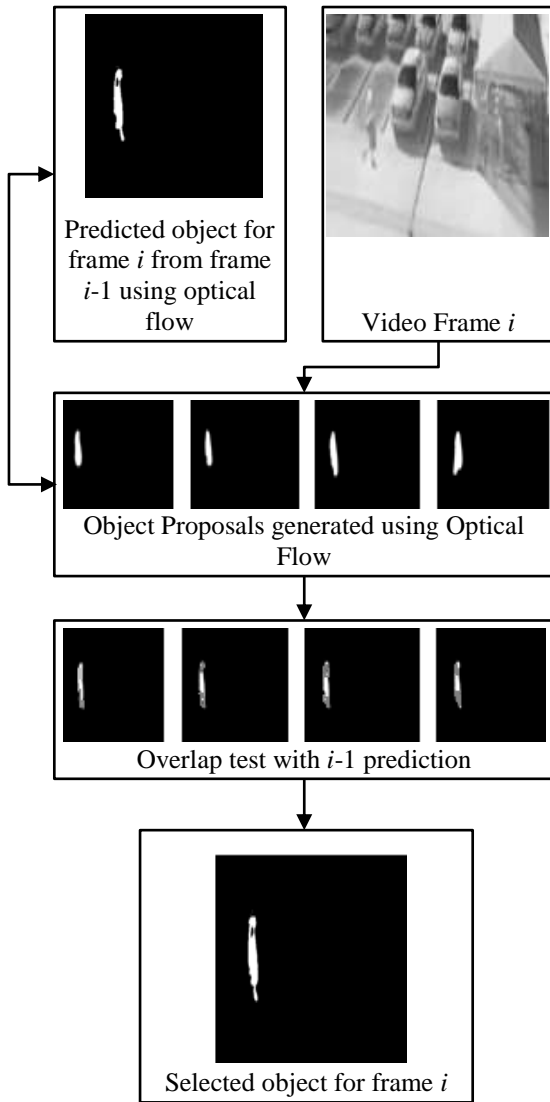


Fig.7. Example of Expansion of target proposal

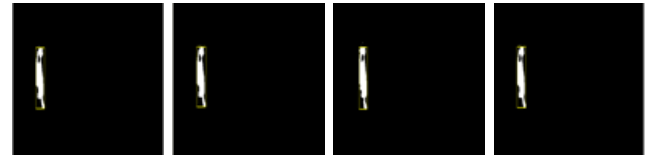
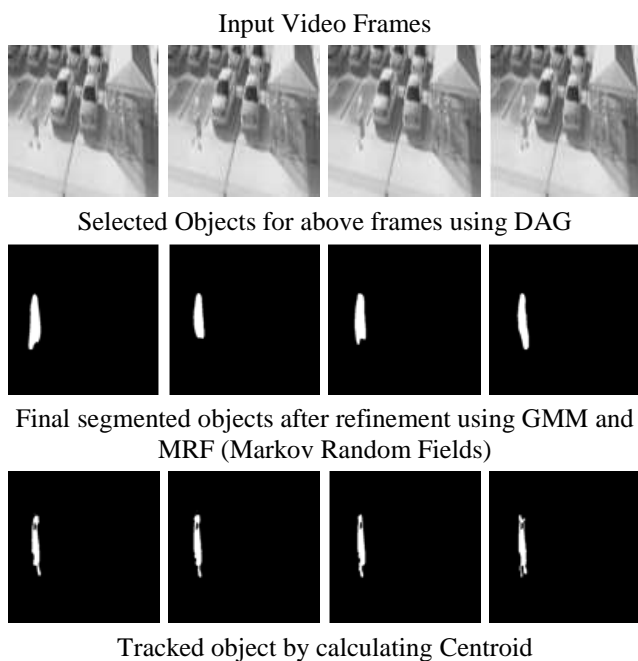


Fig.8. Framework with Example

#### 4. RESULTS

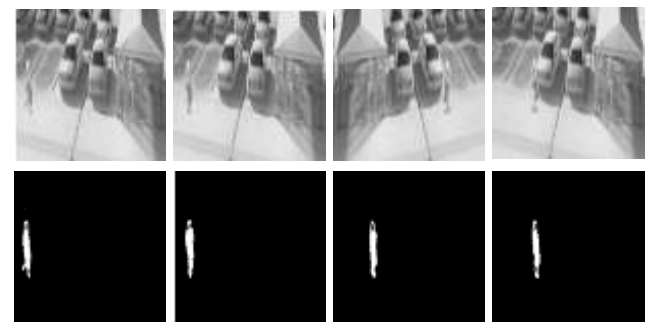
Different thermal datasets are taken from the INO Video Analytics Datasets [20].

In this section Fig.9(a) - Fig.9(h) shows the segmented output using Directed Acyclic Graph (DAG). The Fig.10 and Fig.11 shows the tracking of single objects and tracking of multiple objects by calculation of centroid. The Fig.12 shows percentage error graphs of DAG compared with Background Subtraction (BGS) technique.

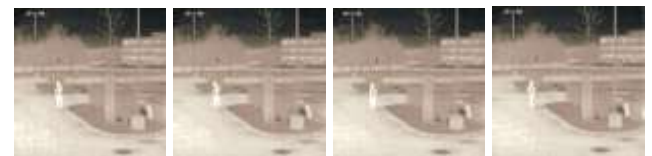
The Table.1 - Table.5 gives the statistical parameters of different DAG video outputs. The Table.6 - Table.8 gives the statistical parameters of different BGS video outputs. In Fig.9(a) - Fig.9(h) the first row show the original video frames from the thermal video sequence where the person is the object to be segmented. The second row shows the segmentation results calculated using Directed Acyclic Graph (DAG). In Fig.10 the centroid of each segmented object is calculated and by finding the Euclidean distance the object is tracked in the video. The tracked object is shown using the bounding box around the object.

The Table.1 - Table.5 gives the statistical parameters calculated for each video output using DAG. The parameters are accuracy, precision, percentage error, geometric mean, f measure, sensitivity and specificity. The Table.6 - Table.8 gives the statistical parameters calculated for each video output using background subtraction method.

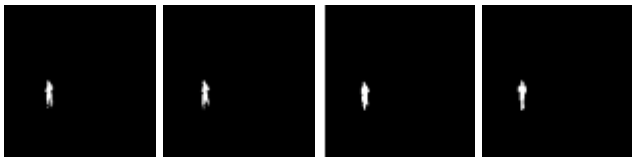
The comparison of proposed method is done with the background subtraction method using percentage error graphs shown in Fig.12. From which we can say that the proposed method performs better than background subtraction method.



(a) Video frames of close person (cp) with frame number: 146,151,156,161







(b) Video frames of trees and runners (tr) with frame number: 448,449,450,451



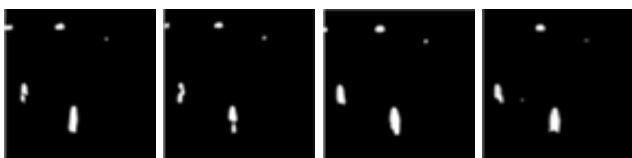
(c) Video frames of irw1 with frame number: 184,185,186,187



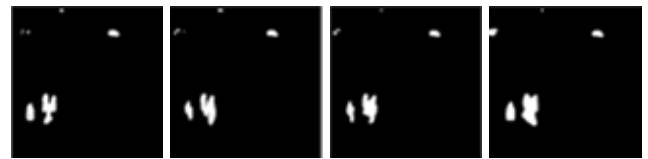
(d) Video frames of irw2 with frame number: 184,185,186,187



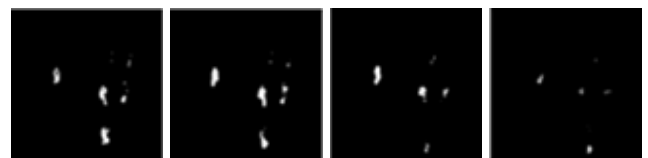
(e) Video frames of irw3 with frame number: 184,185,186,187



(f) Video frames of parking snow (ps) with frame number: 751,851,951,1051



(g) Video frames of group fight (gf) with frame number: 378,428,478,528



(h) Video frames of 2a video taken from OCTBVS with frame number: 184,185,186,187

Fig.9. Segmentation results of single and multiple objects for different thermal videos where first rows show the actual video frames and second row shows the segmented output using DAG

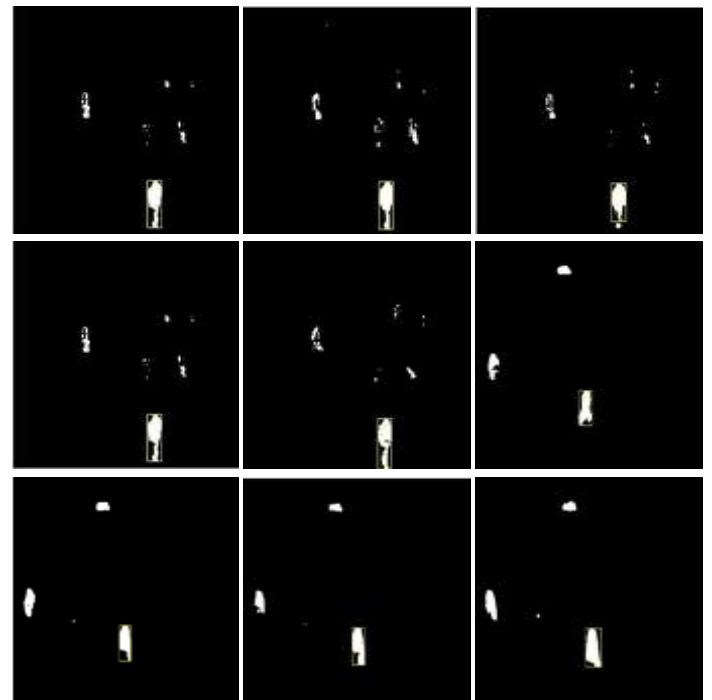


Fig.10. Tracking results of single object

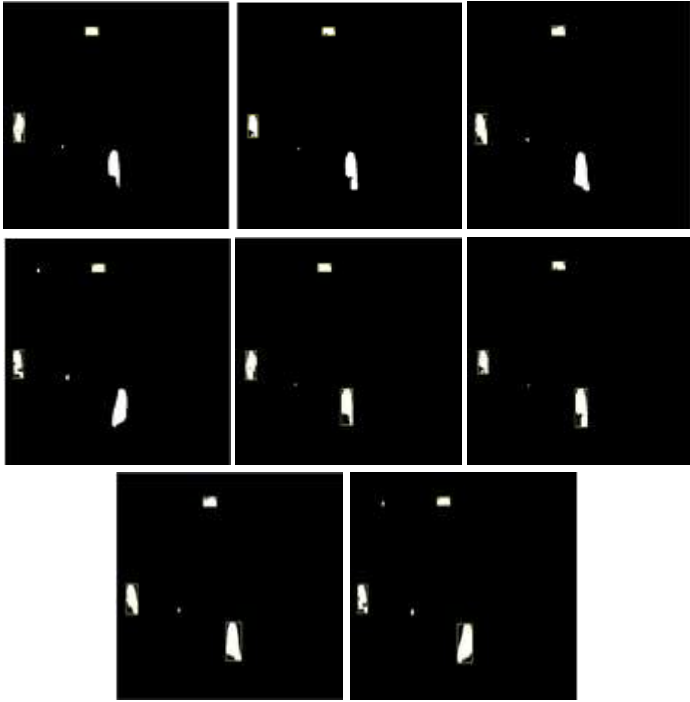


Fig.11. Tracking results of multiple objects

Since the database is not same, direct comparison of results will not make justice to the algorithm. Results presented here are therefore nearly indicative which proves superiority of the proposed algorithm.

Table.1. Statistical parameters of close person video using DAG

Frame number	146	151	159	161	166	171
Accuracy	0.986264	0.993748	0.993461	0.994671	0.994555	0.995977
F-measure	0.545646	0.719123	0.722272	0.783434	0.751092	0.815392
G-mean	0.946518	0.931242	0.840633	0.865332	0.849125	0.880908
Percentage Error	0.09229	0.128324	0.29115	0.249587	0.277311	0.222841
Precision	0.39006	0.612013	0.736213	0.819495	0.781818	0.857582
Sensitivity	0.90771	0.871676	0.70885	0.750413	0.722689	0.777159
Specificity	0.986985	0.994879	0.996917	0.997849	0.997681	0.998507

Table.2. Statistical parameters of trees and runners video using DAG

Frame number	146	151	159	161	166	171
Accuracy	0.999352	0.99922	0.999412	0.999448	0.999496	0.998992
F-measure	0.928	0.913679	0.934579	0.936639	0.943089	0.92953
G-mean	0.988426	0.993871	0.996873	0.998262	0.99832	0.954387
Percentage Error	0.022472	0.011494	0.005682	0	0	0.088816
Precision	0.883249	0.849383	0.881612	0.883117	0.894602	0.94863
Sensitivity	0.977528	0.988506	0.994318	0.997067	0.997135	0.911184
Specificity	0.999445	0.999265	0.999433	0.999458	0.999506	0.999637

Table.3. Statistical parameters of irw1 video using DAG

Frame number	146	151	159	161	166	171
Accuracy	0.96138	0.963646	0.96293	0.966107	0.964219	0.965951
F-measure	0.604428	0.625737	0.622364	0.645609	0.630347	0.647051
G-mean	0.813233	0.823604	0.818277	0.84159	0.825862	0.836976
Percentage Error	0.32115	0.304943	0.313834	0.274924	0.301431	0.283194
Precision	0.544712	0.568991	0.569417	0.58184	0.574265	0.589668
Sensitivity	0.67885	0.695057	0.686166	0.725076	0.698569	0.716806
Specificity	0.974218	0.975927	0.975825	0.976826	0.97635	0.977293

Table.4. Statistical parameters of irw2 video using DAG

Frame number	146	151	159	161	166	171
Accuracy	0.981094	0.98082	0.982174	0.981016	0.982747	0.981992
F-measure	0.796866	0.795615	0.812901	0.802064	0.822648	0.819711
G-mean	0.821242	0.81614	0.831213	0.818514	0.836345	0.834404
Percentage Error	0.324798	0.333566	0.308694	0.330007	0.300478	0.303654
Precision	0.972014	0.986919	0.986401	0.998985	0.998376	0.996198
Sensitivity	0.675202	0.666434	0.691306	0.669993	0.699522	0.696346
Specificity	0.99887	0.999476	0.999434	0.999959	0.999931	0.999834

Table.5. Statistical parameters of irw3 video using DAG

Frame number	146	151	159	161	166	171
Accuracy	0.99543	0.995378	0.996133	0.996146	0.996901	0.996901
F-measure	0.881778	0.880592	0.901165	0.90153	0.915121	0.915121
G-mean	0.970058	0.969689	0.965693	0.966377	0.961945	0.961945
Percentage Error	0.055556	0.056236	0.064917	0.06358	0.072977	0.072977
Precision	0.826911	0.825347	0.869621	0.869147	0.903521	0.903521
Sensitivity	0.944444	0.943764	0.935083	0.93642	0.927023	0.927023
Specificity	0.996367	0.996327	0.997306	0.997293	0.998183	0.998183

Table.6. Statistical parameters of close person video using BGS

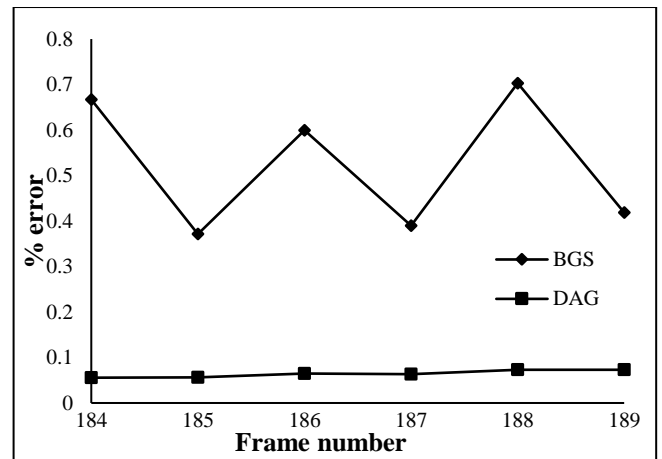
Frame number	146	151	159	161	166	171
Accuracy	0.60255	0.578603	0.618705	0.58166	0.626465	0.605172
F-measure	0.042697	0.0406	0.058304	0.056453	0.054083	0.053922
G-mean	0.764481	0.747227	0.777486	0.749519	0.764895	0.768964
Percentage Error	0.275947	0.183394	0.236189	0.230365	0.305409	0.261991
Precision	0.027943	0.026361	0.031096	0.029069	0.027843	0.02772
Sensitivity	0.977148	0.975342	0.983761	0.97438	0.939309	0.984215
Specificity	0.600619	0.576364	0.614523	0.57655	0.622867	0.600788

Table.7. Statistical parameters of trees and runners video using BGS

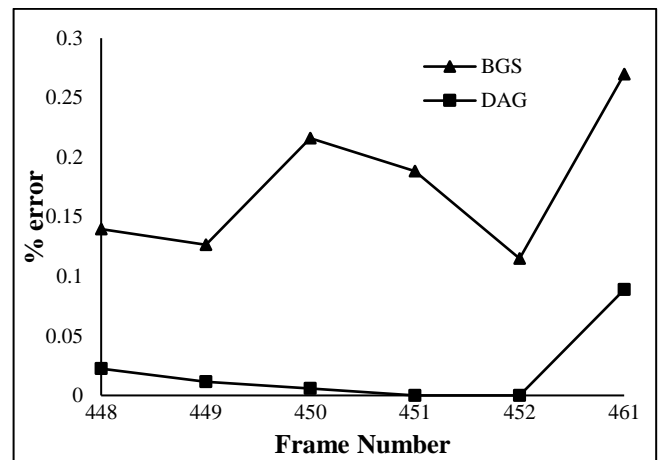
Frame number	146	151	159	161	166	171
Accuracy	0.25845	0.268029	0.268629	0.264788	0.261067	0.263419
F-measure	0.010253	0.009872	0.008978	0.008931	0.009907	0.015813
G-mean	0.469043	0.481582	0.457072	0.461656	0.47827	0.458619
Percentage Error	0.139785	0.126437	0.215909	0.188235	0.114943	0.189145
Precision	0.005157	0.004964	0.004515	0.00449	0.004981	0.007985
Sensitivity	0.860215	0.873563	0.784091	0.811765	0.885057	0.810855
Specificity	0.255751	0.265489	0.266442	0.262546	0.258449	0.259395

Table.8. Statistical parameters of irw2 video using BGS

Frame number	146	151	159	161	166	171
Accuracy	0.445911	0.102266	0.427096	0.10418	0.438112	0.10625
F-measure	0.021252	0.024671	0.025687	0.025027	0.018692	0.022918
G-mean	0.386428	0.241261	0.413858	0.24009	0.361764	0.238171
Percentage Error	0.666667	0.371305	0.599448	0.389772	0.703035	0.418353
Precision	0.010976	0.012582	0.013269	0.012775	0.00965	0.011689
Sensitivity	0.333333	0.628695	0.400552	0.610228	0.296965	0.581647
Specificity	0.44798	0.092584	0.427606	0.094462	0.440702	0.097526



(b) Percentage Error for irw3 DAG



(c) Percentage Error for tr DAG

Fig.12. Comparison graphs of percentage error of proposed method with Background subtraction method

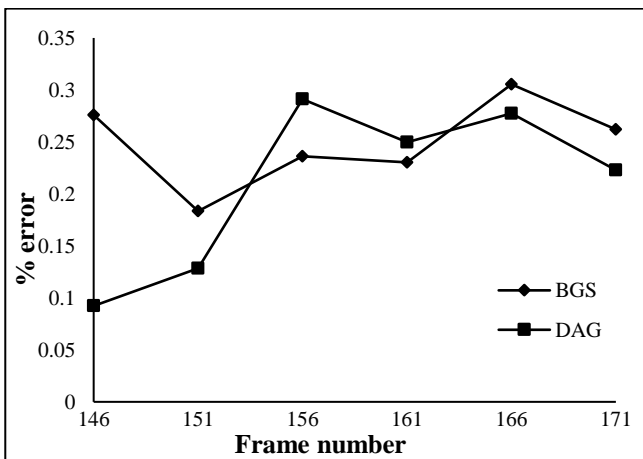
### 5. CONCLUSION

We have projected a different and useful technique of Directed Acyclic Graph for segmentation and tracking of the objects in thermal video sequence. The technique additionally makes use of new approach for calculation of the objectness of an area and compute homogeneous attribute between target proposals across neighboring frames. The projected method is proved to be robust in terms of F-measure, sensitivity and precision compare to background subtraction technique. The projected method performs well for different thermal datasets.

In our proposed method, we have used motion cues as a feature vector input to DAG. In future we can use appearance model along with motion model to generate different object proposals and fed it to multilayered DAG which can make system Robust against different challenging conditions.

### REFERENCES

[1] X. Bai, J. Wang, D. Simons and G. Sapiro, "Video Snapshot-Robust Video Object Cutout using Localized Classifiers", *ACM Transactions on Graphics*, Vol. 28, No. 3, pp. 1-11, 2009.



(a) Percentage Error for cp DAG



- [2] X. Ren and J. Malik, "Tracking as Repeated Figure or Ground Segmentation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [3] J. Yuen, B. Russell, C. Liu and A. Torralba, "Labelme Video: Building a Video Database with Human Annotation", *Proceedings of IEEE 12<sup>th</sup> International Conference on Computer Vision*, pp. 1451-1458, 2009.
- [4] Y. Sheikh, O. Javed and T. Kanade, "Background Subtraction for Freely moving Cameras", *Proceedings of IEEE 12<sup>th</sup> International Conference on Computer Vision*, pp. 1219-1225, 2009.
- [5] D. Tsai, M. Flagg and J. Rehg, "Motion Coherent Tracking with Multi Label MRF Optimization", *International Journal of Computer Vision*, Vol. 100, No. 2, pp. 190-202, 2010.
- [6] J. Shi and J. Malik, "Motion Segmentation and Tracking using Normalized Cuts", *Proceedings of 6<sup>th</sup> International Conference on Computer Vision*, pp. 1154-1160, 1998.
- [7] T. Brox and J. Malik, "Object Segmentation by long term Analysis of Point Trajectories", *Proceedings of European Conference on Computer Vision*, pp. 282-295, 2010.
- [8] Y. Huang, Q. Liu and D. Metaxas, "Video Object Segmentation by Hyper graph Cut", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1738-1745, 2009.
- [9] W. Brendel and S. Todorovic, "Video Object Segmentation by Tracking Regions", *Proceedings of IEEE 12<sup>th</sup> International Conference on Computer Vision*, pp. 833-840, 2009.
- [10] A. Vazquez-Reina, S. Avidan, H. Pfister and E. Miller, "Multiple Hypothesis Video Segmentation from Super Pixel Flows", *Proceedings of European Conference on Computer Vision*, pp. 268-281, 2010.
- [11] M. Grundmann, V. Kwatra, M. Han and I. Essa, "Efficient Hierarchical Graph based Video Segmentation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2141-2148, 2010.
- [12] J. Wang, B. Thiesson, Y. Xu and M. Cohen, "Image and Video Segmentation by Anisotropic Kernel Mean Shift", *Proceedings of European Conference on Computer Vision*, pp. 238-249, 2004.
- [13] C. Xu, C. Xiong and J. Corso, "Streaming Hierarchical Video Segmentation", *Proceedings of European Conference on Computer Vision*, pp. 626-639, 2012.
- [14] I. Endres and D. Hoiem, "Category Independent Object Proposals", *Proceedings of European Conference on Computer Vision*, pp. 575-588, 2010.
- [15] J. Carreira and C. Sminchisescu, "Constrained Parametric Min Cuts for Automatic Object Segmentation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3241-3248, 2010.
- [16] B. Alexe, T. Deselaers and V. Ferrari, "What is an object?", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 73-80, 2010.
- [17] Y. Lee, J. Kim and K. Grauman, "Key Segments for Video Object Segmentation", *Proceedings of International Conference on Computer Vision*, pp. 1995-2002, 2011.
- [18] T. Ma and L. Latecki, "Maximum Weight Cliques with Mutex Constraints for Video Object Segmentation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 670-677, 2012.
- [19] Zhang D, Javed O and Shah M, "Video Object Segmentation through Spatially Accurate and Temporally Dense Extraction of Primary Object Regions", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 628-635, 2013.