Journal of Engineering Sciences
Vol 10, Issue 12, DEC/2019
ISSN NO:0377-9254

SALIENCY BASED OBJECT IDENTIFICATION USING SPATIO TEMPORAL OPTIMIZATION PROCESS

N. Santhosh Kumar[1],Sk. Taj Mahabhoob[2]
[1]M. Tech Student JNTU College of Engineering, Pulivendula, E-mail Id: santhosh0436@gmail.com.
[2]Associate Professor JNTU College of Engineering, Pulivendula, E-mail Id: shaiktajmahaboob@gmail.com.

ABSTRACT:

The appearance of moving targets is a challenging visual tracking problem. We present in this paper a new spatio temporal based visual object tracking technique containing optimization method. To achieve an enhancement of local saliency as a regional constraint. We suggest a new approach to moving signals from the optical stream sector, the saliency of previous video framing and position record of change recognition, which might discern moving objects from various changing background regions for robust movement of outstanding objects. In addition, an effective object measurement with intuitive geometric interpretation is proposed to extract some stable target and environment regions, that provides o the basis of defining the foreground and background context and constraint to help the distribution of saliency. In addition to the appearance template updating, we introduce a new outlier removing approach that helps to prevent accumulation of errors output of the experiment on difficult video segments show that the proposed spatial and temporal visual monitoring method includes optimization mechanism and higher performance of movement energy than current tracking algorithms.

Key words: saliency mapping, spatio-temporal method.

I. INTRODUCTION

Image recognition is an area where images are collected, interpreted, processed and understood. Object detection is among the core aspects in computer vision. It is a method of recognizing and locating targets from a series of images or videos. In several applications, that is a vital part like optical character recognition, facial recognition, automated car parking technologies, visual targeting and monitoring, etc. It is a complicated process related to objects ' multi-class variation and other visual characteristics. Different methods have been accomplished to restrain these challenges, but identification of targets with high efficiency has always been difficult.

saliency based object identification using spatio temporal optimization process shows the main concern about continuous tracking of objects. Visual tracking, an important computer vision research subject, has wide-ranging applications in different fields that inculcate self driving cars, safety and monitoring technologies, and sight based controls. Visual monitoring continuously suggests the condition of a target object in a video frame that is annotated (manually marked or identified during the first frame). Despite of being studied for several decades visual tracking has made much progress in recent years. Developing a robust monitoring algorithm continues a complicated task due to severe changes in the visibility of tracked object due to massive variation in pose, complicated background clutter, dramatic variance in lighting, etc.

An efficient model of appearance guarantees a monitoring system's robustness; in recent years this has received considerable attention [1], [2]–[16]. various active representations have been proposed, in generative models usually learn a presentation model to describe the behavior of the target and then the method to search with full similarity for the object field. In general, GMMs(Gaussian mixture model) [17], histograms [18], subspace illustration [19] and sparse recognisation [13], [14], [20], [21] are representations for the creation of generative appearance models.

In GMM based recognition with a maximization algorithm for online expectations to overcome variations in target appearance during trackingTo identify a targeted object, a collection of regional image patch histograms were used in [18], Adam et al.In [19], Ross et al. Suggested an iterative subspace learning technique for learning a subspace recognition which can evolve to adjustments in the destination appearance. In [22], an observation model for Kwon and Lee Was broken down into many basic analysis models built and use the sparse major component methodAvidan [23] initially proposed image monitoring as a dual identification

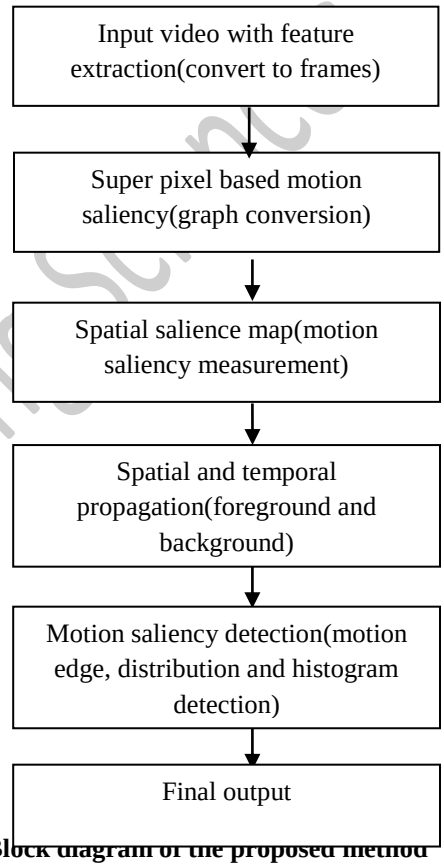
trouble, Collins et al.[24] proposed a design method for selecting the most unequal features to differentiate the target object from the background by integrating an SVM-based off-line algorithm into a visual flow screen. Grabner et al.[25] suggested a method to track the Babenko et al. range of online boosting features[26].suggested good and bad bags to acquire a visual monitoring identification for multiple training instances. By considering structured unmarked data to visual monitoring, Kalal et al. [27] developed the binary classifier. In [28], Hare et al. used a robust monitoring online constructed output SVM Classifier that may relieve the impact of samples being incorrectly labeled. A dual-expert restore scheme was suggested in monitoring Zhang et al.[29] to resolve the drift issue.

A visual monitoring as a binary categorization issue was developed by Avidon to integrate off-line algorithm based on SVM into an optical flow-based tracking system. In order to separate the target item from the context. Babenko et al. [30] proposed using positively and negatively bags to develop a visual monitoring identification for several learning instances. Zhang et al.[31] suggested a dual-expert detection system to solve the issue of track drifting.Henriques et al. [32] presented a simple monitor that uses the kernel matrix's circulating framework for kernelized correlation filters (KCF) which can be effectively solved using the fast Fourier algorithm. Through incorporating a level adaptive scheme and color naming functions, Li and Zhu enhanced the KCF monitor. Ma et al. [33] used characteristics of centralized layers of neural convolution neural networks (CNN)'learning effective visual monitoring or efficient KCF representation.

Sparse conceptions have recently been widely used in visual tracking, which can be categorized as holistic and regional sparse appearance features of representation. A specified performance monitoring has been exploited in Mei and Ling, which is heard by optimizing a' 1 minimization problem. This work was extended by Li et al. using the orthogonal matching pursuit algorithm to solve the optimization problem efficiently and further improved in the efficiency by Bao et al. via the accelerated Proximal solution to gradient. Nevertheless, these sparse trackers focused on recognition takes to consider the dynamic models of the target that are prone to extreme spatial occlusion and varying positions. This tracking system relies with the first frame on a fixed regional dictionary and also has high likelihood of failure complex scenes.

II. METHODOLOGY

In this section, we show how the object is identified by using saliency using spatio temporal process. For this we have to convert into frames ,after that saliency mapping propagation is to be applied and at last spatio temporal propagation is achieved to track the image, here we shows the hierarchal procedure of this paper



Block diagram of the proposed method

A. Saliency mapping Model

Video clearly explain that motion detection has wide application chances. Saliency mapping method was mainly used to track the moving object and to identify the behavior on moving objects. There are three ways for tracking for finding the moving objects, those are: Background parametric difference method, Optimization flow field method and Frame extraction difference method.

Each and every method has its specialty and advantages of its own but on other side the methods have disadvantages

For example: -

*Background Modeling is necessary for the 1st method that is background difference method

* Huge Calculation part is required for the optical flow methods &impact of noise is more in the frame differential method

In recent times, some motion detection methods came into existence which are based on visual notice model, to detect the importance of visual these methods are applied to image video among all those methods, the system of the Itti's visual attention is amazing. This design is made up of three sets of visual features like colours, brightness, and location. All of these are taken, and saliency maps are constructed to each saliency map and finally three saliency maps are fused.Change at brought up another visual attention model. This method includes motion information additionally with brightness &color featuresBut this approach has no time features and Spatial-Integration subsequently modified Integrated spatio-temporal features by integrating light effects with one another. But this technique is used for static background and it requires massive calculation. Combine the motion features to classify the moving object in both dynamic and static context

As a texture feature, they obtain the gray Grade matrix of angular second co-occurrence motion and combine it with Itti's saliency model to obtain the saliency mapping. Temporal salient map is produced by Observation of regional motion, local movement and absolute motion between discrete time field frames and afterwards motion difference between moving object and context. The map of space and time is eventually fused by a active fusion process. This paper propose a innovative method for Identification of spatio-temporal object.To expand the detection capability, the edge data is included in the design. The temporal saliency is obtained by the introduction of kernel density sequence video signals and Shannon entropy. This was created to calculate the object's movement on the temporal domain. Then after the two saliency maps have been combined. Throughout the video sequences, the changing targets are marked.

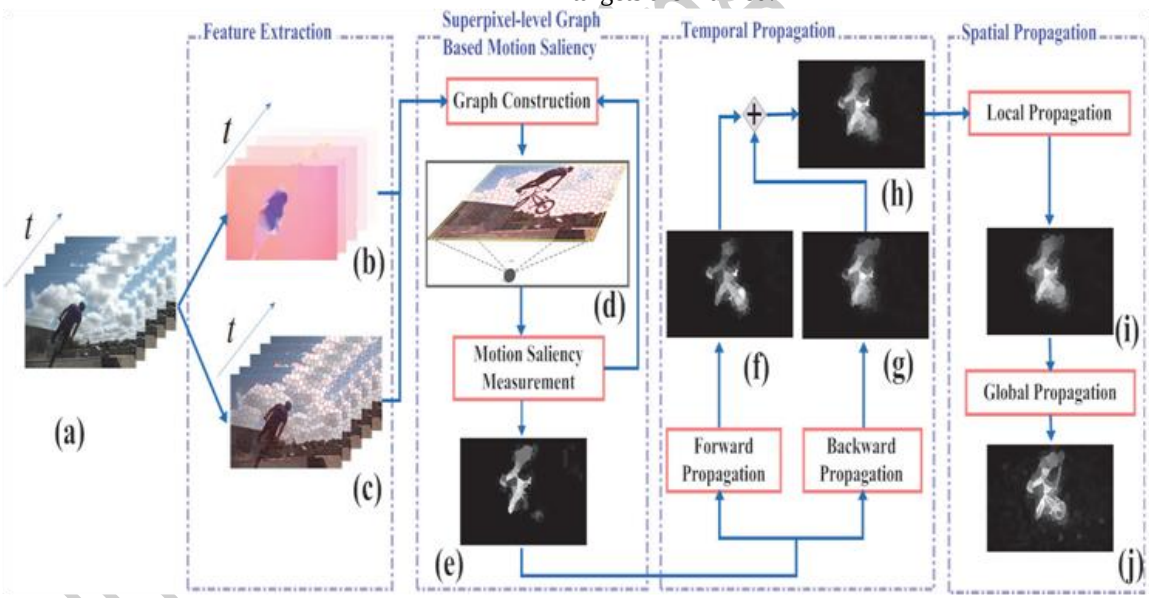


Fig. 1. Illustration of the proposed SGSP model. (a) Original video frames; (b) motion vector fields; (c) super pixel segmentation results; (d) super pixel-level graph; (e) motion saliency map; (f) forward temporal saliency map; (g) backward temporal saliency map; (h) integrated temporal saliency map; (i) initial spatiotemporal saliency map; (j) final spatiotemporal saliency map.

B. Spatial salience map

Spatial saliency shows the attractiveness of essential spatial data in the video scene. In the absence of movement characteristics, it emulates a certain movement of a scene as a salient feature. It also demonstrates the mechanism of human visual attention globally, this spatial saliency map includes features of brightness, color and direction in the method of itti, and is blurring due to lack of global object or obstacle contours. Then the visual processing operator surrounds the centre system conducts differential operations on the four types of functionality maps by obtaining the four main visual parameters in the different-scale system We merge feature maps of different levels and get four functionality maps. Finally, after we get the

final spatial saliency map by iterating differently with Gaussian. The proposed spatiotemporal restricted optimization is specified under the constraint that reliable labels are originally assigned to some super pixels and this comprises of three potentials, foreground, background and smoothness potential

$$\min\ E(S) = \sum_{i-1}^{N} \Phi\,(s_i) + \sum_{i-1}^{N} \Gamma(s_i) + \sum_{i,j\in N}^{N} \psi\,(s_i, s_j) \quad \text{s.t. } \Theta(S) = k \qquad (1)$$

To accomplish a superior abuse of neighbourhood foundation data of the objective, we use the idea of setting (for example the area or encompassing district of the objective article). In every way that really matters, we set setting as a square shape territory enveloping the target centre with fourfold size of the target skipping box. Specifically, we create a model to delineate the particular situation. With no authentic data of objective area perception, this replica is a worldwide earlier model of the setting territory. The setting area set and z belongs to X as a self-assertive area in the specific circumstance. The setting highlight at area z is signified by c (z). For an objective item in the setting district, its real area o can't be straightforwardly watched. We will probably appraise the objective area. The conceivable objective area in the setting can done by portrayed with shared likelihood P(c(z),o). This is the likelihood detailing utilized for future system inference. Here we examine about the displaying of this likelihood. Basically, an objective is commonly obvious contrasted with its neighbourhood foundation. With a saliency strategy, we can draw near to the surmised objective position in the unique situation. It is believed to be one of visual processing follow-up management systems. We implement a lighting model for appropriation P (c(z),o) of the target setting locale to Co(z) in view of the picture saliency. The Co(z) system is defined as a combination of two sections:

$$Co\,(z) = \gamma Cf\,(z) + (1-\gamma)\,Cs\,(z), \qquad (2)$$

The foreground capacity is defined in the statement that certain stable image regions O, that are certain to be part of the outstanding image, can be accomplished by ways of spatial temporal visual analysis. here,we define the capacity of the foreground in a frame of each super-pixel Ft as follows,

$$\Phi(si) = F(ri)(1 - si)^2 \qquad (3)$$

Here γ is a simple equation of weighting of the two allocation sections, specifically set by inquiries to 0.75. The Cf (z) conveyance models the probability of setting presence at the pixel level. Cs (z) shows the probability of propagation of prominent frontal region problems in surroundings.

C. Spatio-Temporal Context Integration

Spatio-temporal background information for effective monitoring is very essential. The goal description is that changes will happen gradually between two adjacent frames given the effective frame rate of the exact video sequence and from the previous presentation differences are inserted into the current appearance state calculation. Whereas the target object travels efficiently from one position to another, there is a high correlation between the target object and its valued contexts in the spatial context. In this section, we presented how to integrate the spatio-temporal background data into our robust tracking appearance model.

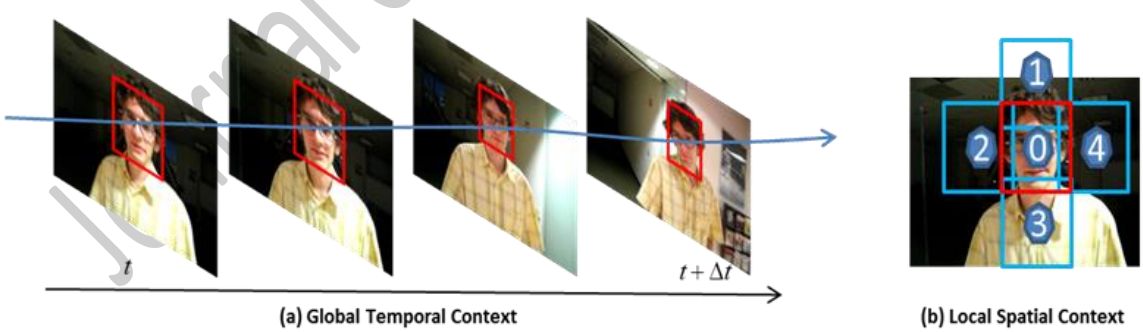


Fig. 2: Illustration of Saliency mapping Model and Spatial salience map.

1) Global Temporal Context:

We need to obtain target object piling results and use gradual subspace optimization techniques as not only to preserve the most common analysis obtained, but also to evolve well to appearance differences.We need to have a linear combination of PCA-based vectors and extra trivial model templates to estimate the target.We obtain the target object's tracking results and then implement the Increased efficiency subspace learning system as it preserves not only usual findings produced, and also responds well with the presentation variations.

The approximate goal is constructed by merging static PCA base functions with extra trivial parameters p = U.q + e = [U I] [q e] T (4)

Here p symbolizes the approximate goal function, U is the Eigen-based vector matrix, q is the Eigen-based parameter coefficient and e represents the compromised representations in p. In, both q and e's sparsely defined projections were enforced. Though, due to the Hierarchical clustering-based matrix U orthogonally, the enforcement of sparse restriction on variables q may lead to the loss of valuable confirmation. Therefore here we present a new template modification method or formulae in trying to impose sparse restriction since occlusion and noise-related mistake often has a sparse distribution..

2) Local Spatial Context:

The regional spatial context data is extracted from the target object's neighboring regions (here, the five adjacent patches were used as spatial context information). Local context data, which includes supporters and distractors to improve the tracker's robustness, is partially hidden from the target. However, there may be a difference between the relative movement model of the construct and the target as well as the auxiliary objects and using the Strengthening Process for Creating Strong Classifiers to overcome the bimodal distribution between each object and the target, they suggest a simple and effective process for analyzing the similarities between the target exterior features in the frame bufferand the ith Applicant context with region appears as a feature set of local spatial constraints data.

f i = e− ("vt−1−vi, 0"1−Σ4"vt−1−v "). (5)

To measure the low associated region between the object and its surrounding background area presence, which results in a higher range when we increase the similarities between the targets and their object and surrounding background region while minimizing similarity with the background zone surrounding them.

D. Proposed Tracking Algorithm

The projected foreground and context potential were described on the basis of accurate target regions O and reliable context regions B, respectively. in this we resent how to attain accurate regions O and B which play a role in determining the quality of relevant object tracking. The proposed concept of motion potential can collect the rough location of the contextual objectThus, they conduct a binary optimization on the movement energy map M using Otsu thresholding accompanied by a dilation procedure to create feature-like regions that are required to cover the entire salient target and some neighboring reference regions. The discovery that super-pixels are usually grouped together through object-like regions K and these super pixels are much more probable to be artifacts nearer to the cluster core. Thus, in feature-like regions K, they suggest the cluster frequency to calculate the present artifact for every super pixel ri. To measure the size of the swarm, We layer in a standard vector the mean spatial location P(ri) 2 R21, the movement distribution frequency Md(ri) and the motion power M(ri).

$$\mathfrak{T}(r_i) = \sum_{r_i, r_j \in K} \delta \left(\| V(r_i) - V(r_j) \| - dc \right) \quad (6)$$

Where dc is a cutoff value has wide range of [0:05; 0:5] in our approach.
The every point is marked with the saliency tag by optimizing the spatiotemporal constrained optimization in Eq. (1) . The restriction element will help the distribution of salience from stable regions O and B by determining the relationship between super pixels. For just the saliency diffusion centered on the stable target regions, we establish an aversion matrix Woi 2 RKN from K super pixels ro 2 O to all N super pixels ri 2 S,

$$W_{oi} = [. . . , w_{oi}(r_o, r_i), . . , w_{K\,N}(r_K, r_N)\ ]$$

where

$$w_{oi}(r_o, r_i) = \exp\left(-dist_c^2\ (r_o, r_i) / 2\sigma^2\right), (r_o, r_i) \in N$$

The constraint function (S) = form of matrix as follows

$$\begin{bmatrix} .., D_{oi} & -\ \alpha W_{oi,\cdot} \\ .., D_{bi} & -\ \beta W_{bi,\cdot} \end{bmatrix}_{(K+M)\times N} \begin{bmatrix} s_1 \\ . \\ . \\ s_i \\ . \\ s_N \end{bmatrix}_{N\times 1}$$

$$= \mathbf{b}\ \Theta \begin{bmatrix} \boldsymbol{I}\ (\boldsymbol{r_o}) \\ \boldsymbol{I}\ (\boldsymbol{r_i}) \end{bmatrix}_{(K+M)\times 1}$$

where Θ represents the element-wise multiplication; [s1; s2; s3;:::; sNT is the response vector that each component can be a saliency tag to be projected in S; [I (ro); I (rb)] is the group strength matrix; b is a (K M) - dimensional measurement vector and each item is

present at either 1 (target) or 0 (surrounding) by the specified thresholds.

III.EXPERIMENTS

A. Experimental Setup

The setup for this object tracking by saliency mapping by spatio temporal propagation is designed in MATLAB operating on an Intel Core i5 CPU processor with 8 G RAM at 5 frames per second. We define an affinity matrix Woi 2 RKN from K super pixels ro 2 O to all N super pixels ri 2S. γ is a simple equation of weighting of the two allocation sections, specifically set by inquiries to 0.75. The Cf (z) conveyance models the probability of setting presence at the pixel level. Cs (z) shows the probability of propagation of prominent frontal region problems in surroundings. This two-section combination gives an estimate of the combined potential flow P (c (z), o) in two points of view.and Θ represents the element-wise multiplication; [s1; s2; s3;:::; sNT is the response vector that each component can be a saliency tag to be projected in S; [I (ro); I (rb)] is the group strength matrix; b is a (K M) - dimensional measurement vector and each item is present at either 1 (target) or 0 (surrounding) by the specified thresholds.

Throughout this paper Results of the experiment on difficult video sequences show that the proposed spatiotemporal visual tracking algorithm incorporates mechanisms for optimization and movement energy achieves better performance than current related tracking algorithms. The figures show how the object was monitored by spatio-temporal approach in the mapping of salience.

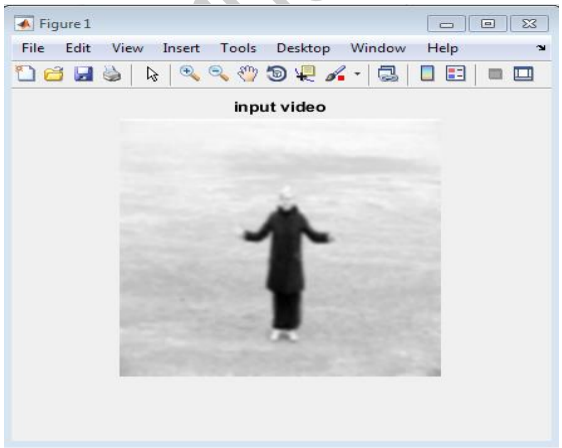


Figure-1 input video frame.

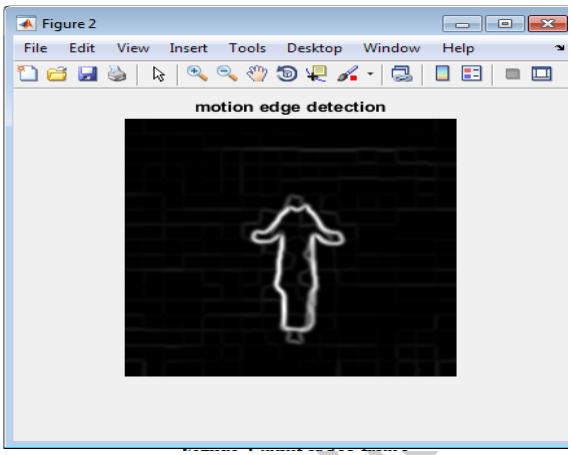


Figure-2: Motion Edge Detection

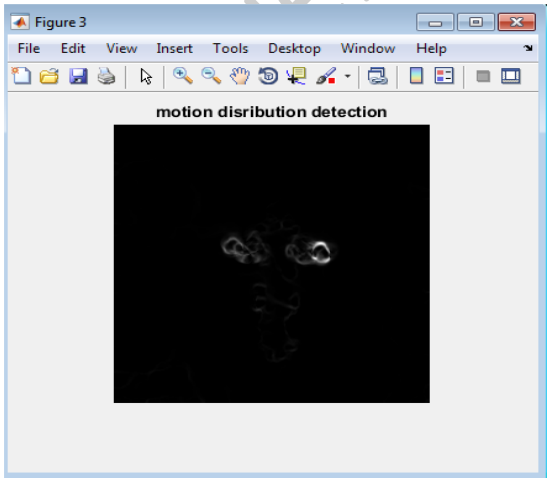


Figure-3: Motion distribution detec

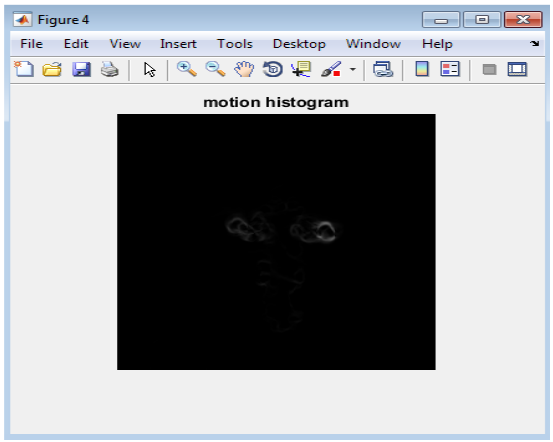


Figure-4: Motion Histogram.

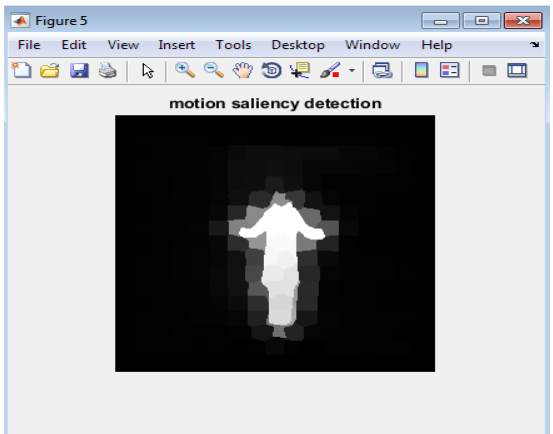


Figure-5: Motion saliency detect

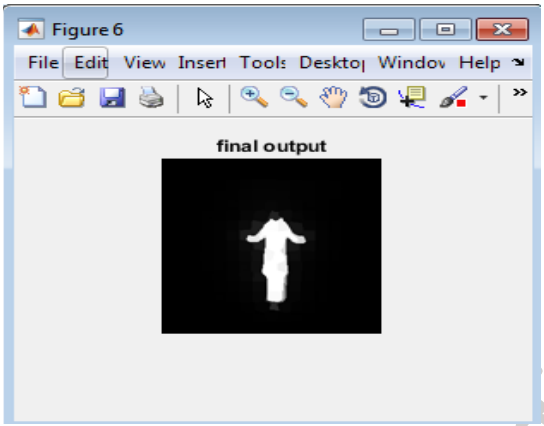


Figure-6: Final output.

IV. CONCLUSION

We also presented a new transfer-based learning monitoring algorithm with minimal spatial and temporal optimization. The spatiotemporal limited optimization formulates the definition of salience as energy minimisation in a graph centered on super pixel inside a picture and comprises of both the foreground potential, context potential, prospective smoothness and regional constraints A new objectness method is to project the identify such accurate region starting salient targets along with context to enable saliency distribution for energy potential modeling and restriction extraction. In addition, a new movement tracking method is suggested as the basis of both the objectness benchmark to extract the movement of the salient objectness measure, A new movement recognition approach is projected for perfect movement of marked tergets inside a frame from both static and evolving context regions. Both theoretical and practical comparison experiment shows the better output of the proposed visual object identification algorithm based on motion energy toward state-of - the-art tracking algorithms by delivering more stable and reliable results on a complicated video sequence.

REFERENCES

[1] J. Tang, L. Jin, Z. Li, and S. Gao, "Rgb-d object recognition via incorporating latent data structure and prior knowledge," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 1899–1908, 2015. 1

[2] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in Proceedings of European Conference on Computer Vision, pp. 188–203, 2014. 1, 5

[3] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization, "inProceedings of European Conference on Computer Vision, pp. 188–203, 2014. 1, 5

[4] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in Proceedings of the IEEE International Conference on Computer Vision, pp. 3074–3082, 2015. 1, 2

[5] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in Computer Vision-ECCV 2014 Workshops, pp. 254–265, 2014. 1, 2

[6] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in Proceedings of European Conference on Computer Vision, pp. 127–141, 2014. 1

[7] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 10, pp. 2002–2015, 2014. 1

[8] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1779–1792, 2016. 1, 4, 5

[9] H. Song, Y. Zheng, and K. Zhang, "Robust visual tracking via self-similarity learning," Electronics Letters, vol. 53, no. 1, pp. 20–22, 2016. 1

[10] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," International Journal of Computer Vision, vol. 77, no. 1-3, pp. 125–141, 2008. 1, 4, 5

[11] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in Proceedings of European Conference on Computer Vision, pp. 234–247, 2008. 1

[12] H. Song, "Robust visual tracking via online informative feature selection," Electronics Letters, vol. 50, no. 25, pp. 1931–1933, 2014. 1

[13] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity based collaborative model," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838–1845, 2012. 1

[14] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1822–1829, 2012. 1, 2, 3, 4, 5
[15] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in Advances in Neural Information Processing Systems, pp. 809–817, 2013. 1, 5
[16] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 11, pp. 2259–2272, 2011. 1, 2, 4
[17] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, "Robust online appearance models for visual tracking," IEEE Transactions on Pattern AnalysisandMachineIntelligence, vol.25, no.10, pp.1296–1311,2003. 1
[18] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 798–805, 2006. 1
[19] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," International Journal of Computer Vision, vol. 77, no. 1-3, pp. 125–141, 2008. 1, 4, 5
[20] K. Zhang, Q. Liu, H. Song, and X. Li, "A variational approach to simultaneous image segmentation and bias correction," IEEE Transactions on Cybernetics, vol. 45, no. 8, pp. 1426–1437, 2015. 1
[21] K. Zhang, Q. Liu, J. Yang, and M.-H. Yang, "Visual tracking via Boolean map representations," Pattern Recognition, vol. 81, pp. 147 – 160, 2018. 1
[22] J. Kwon and K. M. Lee, "Tracking of a non-rigid object via patch-based dynamic appearance modelling and adaptive basin hopping Monte Carlo sampling," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1208–1215, 2009. 1
[23] S. Avidan, "Support vector tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp. 1064–1072, 2004. 1
[24] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1631–1643, 2005. 1
[25] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via online boosting," in Proceedings of British Machine Vision Conference, pp. 47–56, 2006. 1
[26] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1619–1632, 2011. 1
[27] Z. Kalal, J. Matas, and K. Mikolajczyk, "Pn learning: Bootstrapping binary classifiers by structural constraints," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 49–56, 2010. 1
[28] S. Hare, A. Safari, and P. H. Torr, "Struck: Structured output tracking with kernels," in Proceedings of the IEEE International Conference on Computer Vision, pp. 263–270, 2011. 1
[29] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization, "inProceedings of European Conference on Computer Vision, pp. 188–203, 2014. 1, 5
[30] K. Zhang and H. Song, "Real-time visual tracking via online weighted multiple instance learning," Pattern Recognition, vol. 46, no. 1, pp. 397– 411, 2013. 1
[31] Y. Yuan, H. Yang, Y. Fang, and W. Lin, "Visual object tracking by structure complexity coefficients," IEEE Transactions on Multimedia, vol. 17, no. 8, pp. 1125–1136, 2015. 1
[32] J. Tang, X. Shu, Z. Li, G.-J. Qi, and J. Wang, "Generalized deep transfer networks for knowledge propagation in heterogeneous domains," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 12, no. 4s, p. 68, 2016. 1
[33] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, "A survey of appearance models in visual object tracking," ACM Transactions on Intelligent Systems and Technology, vol. 4, no. 4, p. 58, 2013. 1