INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 6, Issue 2 NOV 2017

PARTICLE FILTER FRAMEWORK FOR SALIENT OBJECT DETECTION IN VIDEOS

Hafsa Zareen(M.TECH.) [1]

Ms.G.Srilatha (Assistant professor and M.TECH) [2]

[1]Shadan Womens college of engineering and Technology for women, Hyderabad, Telangana, (500004), INDIA

zareenhafsa@gamil.com[1] srilathawins@gmail.com

ABSTRACT

Distinguishing object of enthusiasm for video is testing a direct result of the constant development in the backend, coming about because of camera recognizing a question of intrigue, or development of items in the frontal area. The specialists introduce a focused method to discover question of enthusiasm for video utilizing haphazardly created numbers, which are guided by space and time essential maps and shading highlight with the capacity to rapidly recuperate from false location. The displayed technique for producing space and movement imperative maps depends on contrasting neighborhood highlights and exceptionally includes show in the edge. An area is stamped particularly if there is an expansive contrast amongst nearby and profoundly includes. For space and time significance, shaded and immersion highlights are utilized, while for movement significance objects, optical stream vectors are utilized as highlights. Exploratory outcomes on standard datasets for video isolating and for sastate-of-the-art techniques.

I.INTRODUCTION

The New pattern in everyday life has extremely incredible measure of data accessible and grew consistently. This improvement in image information has prompted different errands of achieving them quick and separating precise information, in order to encourage diverse undertakings from image inquiry to image packing and change over system. One particular issue of PC image handling calculations utilized for extricating data from images, is to remove a substances of enthusiasm for a picture. Human vision framework has a superb component to distinguish correct data from a scene. Such awesome vision arrangement of people permits to use their assets on the most widely recognized gathering of the accessible information, permitting new strategies and leaving in everyday exercises. This element is known as sight vital. Thus, the machine PC it is important to discover element of intrigue with the goal that the different things can be used appropriately to process redress data. Applications run from element finding or image Separating (IR), face or human re-distinguishing proof and video following.

1.1 Objective and Scope

The goals of the venture are to outline an essential substance discovering technique that can be made as a quantum figuring advance for a considerable lot of the already specified assignments. Further, the technique must be unattended with the goal that it can identify any substance. Further, it is to be tentatively certain to guarantee speed arranging, looking at the vast amount of obtained information. As of now talked about base up element extraction can be characterized by the capacity turn out in a scene. Subsequently, most fundamental element discovering techniques in artistic writings (Achanta et al. (2009); Gofer man et al. (2010); Cheng et al. (2011); Peruzzi et al. (2011); Li et al. (2013); Yang et al. (2013); Jiang et al. (2013)) proposed a model by making utilization of little measure of highlights. Yet, as likewise said by Wei et al. (2011); Zhu et al. (2014) the element subordinate model doesn't finish to extricate substance of enthusiasm from unique pictures of different scene circumstances.

We identify this shortcoming in the rarity of feature based approach and exploit border prior as a signal to implement our rest method of entity finding.

Further, class independent entity dividing has recently gained importance in the Computer Sight community (Carrier and Sminchisescu (2010); Enders and Hiram (2010)). In this context, Alexei et al. (2011) had addressed the problem of finding common entities and they need entityless properties as likelihood of a region belonging to an entity. Hence it provides concatenated boxes compared than pixel value segmented outcome. However their precision is very low.

Lower regions are implemented as entities. The expansion is dependent on segmenting of complete criteria (Enders and Hiram (2010)) to priorily maps sequence, though the outputs view that the top-most pointings almost have the half of the complete frame .

Hence, we implemented an algorithm to generate a single map segmenting only the entities of interest, using important and entityless on a conditional random led (CRF). The objective of this project is on one of the vision abilities of human - ending entities of interest in images.

2.LITERATURE SURVEY

In look into writings there have been many methodologies taken by various writers beginning from space innovation to recurrence area examination. Highlight uniqueness and outskirt past has turned out to be fruitful recommendation in imperative finding. Graphical model based methodologies indicate great spread of low-level priors and results in better execution.

2.1 Important Using Low-Level Perceptual Suggestion

An appropriate mix of novel measure of imperative finding is actualized utilizing uncommonness of highlights and backend earlier. Uncommonness of highlight is caught by measuring ghastly component based irregularity and space nearly of shading. Backend prior,on the other hand, is displayed utilizing Gaussian blend show (GMM) of fringe picture components. These two recommendation have turned out to be reciprocal and similarly vital, contending cutting edge strategies.

2.1.1 Important Entity Dividing

In Common Pictures Element level vital in characteristic pictures is fined to deliver class autonomous element isolating. Vital alongside entitles recommendation is as one utilized on a graphical model to speak to a restrictive arbitrary drove (CRF). Chart cut based correct surmising produces the separating, as the vitality is demonstrated as sub modules. CRF parameters are found out utilizing organized max-edge technique. This shows preferable execution over numerous imperative and class free separating calculations.

3.Proposed work

3.1 SALIENCY OBJECT DETECTION

We utilize a molecule channel for its capacity to change the before dispersion of a framework in light of a limited arrangement of weighted specimens. Haphazardly delivered numbers are likewise very strong to halfway conclusion and tentatively require less cost. The heaviness of every molecule is produced utilizing a uniform circulation. The main arrangement of particles is introduced around the focal point of the primary edge of the video. Weights are consequently figured as the weighted aggregate of separation measures of the hopeful districts to the reference dispersion. The space and time imperative and the shading maps are utilized to compute the heaviness of the examples, which enables consequent cycles to draw the particles nearer to the most vital element. In the recommended structure, we recognize just a single substance of intrigue. Fig. 1 delineates a work process of how the molecule channel structure is utilized to recognize the critical substance. Shading renditions of the considerable number of Figures utilized as a part of this paper are accessible on the web.

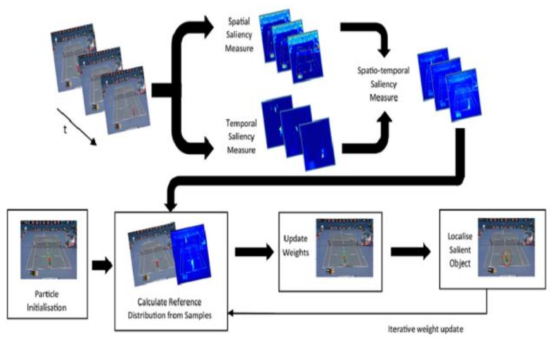

Fig. 5.1 Illustration of the work flow of the suggested framework

3.1.1 Randomly generated numbers

In this segment, we give a concise survey to haphazardly created numbers. It would be ideal if you allude to Through net al. [43] and Arulampalam et al. [44] for more subtle elements. A molecule channel is a constant technique that iteratively approximates the before circulation from a limited arrangement of given specimens.

Xi(t),wi(t)… … ..i=1,… ..N

where each specimen x I t speaks to a speculative condition of the framework with a comparing weight w I t . Considering that we have the perceptions Yt=y0 , … , yt of the framework up to time t, the objective is to appraise the condition of the framework xt from the later dispersion p(xt |Yt). The state-space model of the framework can be spoken to as

$$\begin{aligned} x_t &= Ax_{t-1} + \eta_{t-1} \\ y_t &= Cx_t + \epsilon_t \end{aligned} \qquad (1)$$

Where A and Care the state transition and measurement matrices andηt–1 and εt are the system and measurement noises. Similar to any linear extraction technique, a particle filter employs an accurate and update approach. During the prediction step, the after probability density at time t is calculated using the state transition model

$$p(x_t|Y_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})\,\mathrm{d}x_{t-1} \qquad (2)$$

The correction step updates the posterior using Bayes'rule as

$$p(x_t|Y_t) = \frac{p(y_t|x_t)p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} \qquad (3)$$

5.1.2 Extracting feature distribution

The suggested approach utilizes color and space and time maps as features to detect the important entity. At this point, we consider that both feature maps are given to us. Our experimentally efficient method to generate space and time maps is presented in Section 4. We employ color for its robustness against structural changes in the entity while the space and time important measure allows the process to quickly compress on the important entity. We calculate the identified similarity of color as a feature and consequently use it together with the space and time measure of the samples to remodify the weights. The feature distribution(x)=pu(x)u=1, 2,…, move a region centered at location is given by Numeral et al.

Where Norm is a normalizing constant, δ is the Koneke function, kis a kernel with bandwidth h, Np is the number of pixels in the region and b(xi) is a function that assigns the bin number to the feature obtained from location xi. The histograms are calculated from the HSV color space to produce the algorithm less sensitive to lighting conditions, where in the number of binary values m is set to 8 × 8 × 4. The kernel is a spatial weighting function that gives lower weight to pixels more away from the center xi and is defined as

$$k(r) = \begin{cases} 1 - r^2, & \text{if} \quad r < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

In order for the particles to be temporally coherent, the feature distribution sampled from frame t must be similar to the distribution sampled from framet–1. Although a variety of methods such as histogram intersection, entropy and more have been used to compute the similarity between features, we adopt the Bhattacharyya distance [45] as it is experimentally inexpensive. Let us define p r ={pr (x)} r=1,…,m as the distribution sampled from the previous model and p(xt) as a candidate model sampled from the current frame. The distance between the distributions r and p(xt) is calculated as

$$\rho[p^r, p(x_t)] = \left[1 - \sum_{u=1}^{m} \sqrt{p_u^r(x) p_u(x_t)}\right]^{1/2} \quad (6)$$

Where m is the number of binary values in the histogram.

3.1.3 Particle initialization and weight computation

Unlike a tracking framework, where the filter is initialized with a bounding box containing the entity to be tracked, our method initializes the first set of particles at the center of the frame, with each particle carrying the same weight. The first samples are obtained within an ellipse whose two values are set to half the width and height of the frame size of the window, respectively. Fig. 2 shows the particle initialization and the result generation of important entity extraction. Fig.2a shows the frame in which the particles are activated. Once the particles are generated, the subsequent steps iteratively update the weights of the objects based on the color and space and time maps to detect the important entity. Fig. 2b shows the result of the entity detection (as a red ellipse). The values of the particles are computed using space and time important and the color maps as features. We acquire N samples x It,I=1,…, N from the frame t.

$$w_t^i = \bar{\psi}_s e^{-\lambda(\rho_c^i)^2} \quad (7)$$

(Note that the important maps in this paper are represented as hot spots for better visualization, but in the computations, they are represented as grey-scale images.) The parameter lies set to 20 in our experiments similar to [46]. According to (7), if the Bhattacharya distance of the candidate and the reference distribution is low, implying high color similarity and the mean of the space and time samples are high, then the corresponding sample gets a high weight. This ensures that new particles are obtained based on high color similarity and a high space and time important, thus removing particles that have low space and time important. This also ensures that the particles move closer to the most important entity over a period of time through particle removal and addition of new ones.

3.1.4 PARTICLE FILTER IMPLEMENTATION

A particle filter is driven by the state vector and the dynamic model of the system. We sample an ellipse whose state is defined as x = [x, y, x˙, y˙ , Hx, Hy, a˙], where x, y provides the location of the ellipse, x˙ , y˙ are the velocity components, Hx, Hy are the length of the half axes of the ellipse and a˙ indicates the scale change. The state propagation is done using first-order auto-regressive (AR) process as xt = Axt–1 + ηt–1. The state transition matrix used in the dynamic model is a constant velocity and scale model. The observation similarity for each example is updated using (7). Once the weights of the examples are modified, the mean state or the position of the important entity is calculated as

$$E[w_t] = \sum_{n=1}^{N} w_t^n x_t^n \quad (8)$$

The location of the ellipse in frame t is calculated using the state information calculated from (8). Fig. 3 shows the results of entity extracting on two various videos. The first row of Fig. 3 shows a tracking shot of a lynx moving on a snow covered region. Even though the background consists of large movement, the space and time important map (to be described in the next section) is able to clearly identify the important entity allowing the particles to dissolve and extrct the lynx reliably. The bottom row is also an example of a tracking shot of a skier in action. The shot demonstrates the possibility of the framework to detect different important entities at different times. Initially, the pine trees in the background are less important as shown in Fig. 3b. However, the skier is detected and tracked successfully in the consequent frames as shown in Figs. 3d and f.

The weights of the particles are reinitialized every t frames in order to avoid problems on a particular entity for the entire duration of the video sequence. In our experiments, we set t= F/2, where F is the frame rate of the video. We described how the suggested framework avoids fixation on a particular entity through an example shown in Fig. 3 second row. We provide another example in Section 5 describing how the suggested method might erroneously detect the wrong entity as important, but with particle weight reinitialization, there is quick recovery from the error. It can be noted that the suggested model can be used to detect multiple entities of interest by clustering and segmenting the video frame based on the space and time important values.

This would allow pixels with similar important measures and those that are spatially close to the clustered together. The cluster, which represents a region/ entity is associated with a rank based on the average important value of the cluster. Multiple particle trackers can be initialized around each of the cluster cancroids as seed points, thus, building a framework that would consider multiple hypothesis based on different color models that are obtained from each cluster and whose weights are calculated based on the color composition of the segment.

The important entity detection approach presented assumed the availability of a space and time important map. In the next section, we describe how such a map can be obtained at the pixel level with accurate efficiency.

3.2 SPACE AND TIME SALIENCY

Important region detection approaches in measure important at the patch-level resulting in important maps at reduced resolutions. The importance of a pixel in the provided resolution is obtained by intercepting, which introduces spacious regions in the important map. Our approach calculates important at pixel-level without proper computation above the object. The occurrence of a region is measured by computing the contrast difference present in the region to the remaining of the frame. Important at a pixel location can be measured as the calculated distance of the feature at the pixel to the features extracted from the rest of the pixel locations in the frame. For a frame consisting of n pixels, this needs the calculation of the differences for n –1 pixels for every pixel, acquiring a total of n(n –1) calculations per frame. This would acquire calculations of the order of O(n2) for every frame of a video, which is inefficient especially for videos.

In the produced framework, we decrease the number of computations required for generating important, by measuring the contrast of a feature at a given pixel location to that of its set of most important values. Thus, we measure the occurrence of a pixel by accumulating the distance of the feature value if at pixel I to the most important feature values d ∈ D, weighted by the event occurrence of an the action most important feature value. The occurrence of it is hence formulated as

$$\mathrm{Sal}(f_i) = \sum_{d \in D} p(d) || f_i - d ||_1 \qquad (9)$$

Where p(d) denotes the probability of occurrence of the most important feature d and ||. ||1 is the L1 norm. The most important feature values are calculated by constructing a 50-bin histogram from the set of feature values. This histogram is smoothed to avoid secondary peaks, using a moving average filter with a 5-bin window. The local maxima values present in the smoothed histogram form the set of most important feature values. The suggested measure reduces the maximum number of computations required per pixel to the number of most important feature values. As the provided conspicuity measure is able to accurately measure the difference of features with the less number of calculations, we employ it to calculate the time important measure, discussed in Section 5.1.1 and the space important measure, discussed in Section 5.1.2. Subsequently, we combine the two measures to calculate the space and time important measure in Section 5.1.3.

3.2.1 MOTION SALIENCY MAP

Studies have shown that movement plays a vital role in collecting attention and to outweigh low-level features such as a specific position, intensity and texture in videos. Yantis et al. have shown that the HVS is particularly sensitive to identify accurate stimulus and relative motion of entities which refers to the difference in motion within a space neighborhood. In a similar spirit, we identify important motion from the contrast or the relative movement of entities in the scene calculated as

$$\mathrm{SMot}(m_k) = \sum_{\forall m \in M} p(m) \| m_k - m \|_1 \qquad (10)$$

Where M is the set of most significant motion arrangements computed from optical flow vector velocities that are obtained according to movement and m is the motion magnitude at pixel k.

3.2.2 SPATIAL SALIENCY MAP

Space attention results influence the attention process when similar motion overcomes the frame, for example, movement of trees in the background or when the foreground entities have same motion, for example, a crowd moving in the same direction. The space important map of a frame is calculated depending on the color difference according to the formula in (9). Unlike motion, color contrast is affected by the robustness of similar color pixels [50]. We adopt the method suggested by Morse et al. [51] to calculate the most important colors using shade and saturation maps from the HSV image. The effect of the threshold values are included in the shades histogram which is hence defined as H (b) = (x, y)[I S(x, y), where H(b) is the histogram value in the both bin, S(x, y) is the saturation value at location (x, y) and I is the set of image coordinates having shades corresponding to the both bin. The most important shades are calculated in a similar manner as most important movement by obtaining the local maximum values from a smoothed histogram H(b). The spatial important measure of a pixel k is

$$\mathrm{SSp}(c_k) = \sum_{\forall c \in C} p(c) \| c_k - c \|_1 \qquad (11)$$

Where C is the set of most important shades and ck is the shade value at pixel k. Fig. 5.2 shows retrived images from three different videos and their corresponding space, movement and space and time saliency maps. The first row shows a frame from a video of a moving rhinoceros, shot with an ordinary camera while the middle and bottom rows are videos of a wolf moving in the forest and a skier executing a stunt, respectively. The last two videos are of extracting shots

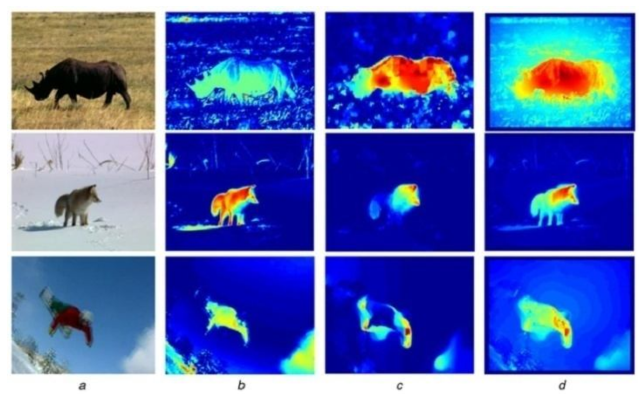

Fig. 5.2 Saliency maps

Important maps

a) Original frame
b) Spatial important map
c) Motion important map
d) Space and time important map

The location of the ellipse in frame t is calculated using the state information calculated from (8). Fig. 3 shows the results of entity extracting on two various videos. The first row of Fig. 3 shows a tracking shot of a lynx moving on a snow covered region. Even though the background consists of large movement, the space and time important map (to be described in the next section) is able to clearly identify the important entity allowing the particles to dissolve and extrct the lynx reliably. The bottom row is also an example of a tracking shot of a skier in action. The shot demonstrates the possibility of the framework to detect different important entities at different times. Initially, the pine trees in the background are less important as shown in Fig. 3b. However, the skier is detected and tracked successfully in the consequent frames as shown in Figs. 3d and f.

The weights of the particles are reinitialized every t frames in order to avoid problems on a particular entity for the entire duration of the video sequence. In our experiments, we set t= F/2, where F is the frame rate of the video. We described how the suggested framework avoids fixation on a particular entity through an example shown in Fig. 3 second row. We provide another example in Section 5 describing how the suggested method might erroneously detect the wrong entity as important, but with particle weight reinitialization, there is quick recovery from the error. It can be noted that the suggested model can be used to detect multiple entities of interest by clustering and segmenting the video frame based on the space and time important values.

This would allow pixels with similar important measures and those that are spatially close to the clustered together. The cluster, which represents a region/ entity is associated with a rank based on the average important value of the cluster. Multiple particle trackers can be initialized around each of the cluster cancroids as seed points, thus, building a framework that would consider multiple hypothesis based on different color models that are obtained from each cluster and whose weights are calculated based on the color composition of the segment.

The important entity detection approach presented assumed the availability of a space and time important map. In the next section, we describe how such a map can be obtained at the pixel level with accurate efficiency.

In the first row, the position of the rhinoceros is high when compared to the background movement in the grass. Hence, the animal's motion is assigned a huge important than that of the moving grass in the background. In the video with the wolf (between row), the values present in the background have high motion mainly providing to the camera extracting the wolf while the motion values of the pixels on the wolf are very low. The motion important map provides a large confidence to the movement of the wolf owing to the large difference in the magnitudes between the identified wolf and the most important background movement. The frame shown in the bottom row of Fig. 4 is from a video of a camera tracking a skier performing a stunt. Similar to the case above, the skier is executing a stunt when his hand moves faster than the rest of his body, producing a large important pixel values along his arm. Thus, it can be seen that the provided motion important measure is able to successfully measure the important motion present from videos shot with a ordinary camera or an advanced camera.

3.2.3 SPATIO TEMPORAL SALIENCY MAP

The space and time saliency map combines the space and movement important maps produced for each frame in such a way that the movement map gets a larger value if there is high motion difference in the order while the spatial important map gets a larger weight if the motion contrast is lesser. This is formulated as

$$\mathrm{STSal}(I) = \alpha \times \mathrm{SMot}(I) + (1-\alpha) \times \mathrm{SSp}(I) \qquad (12)$$

Where α is an adaptive weight given by

$$\alpha = \frac{\mathrm{median}(\mathrm{SMot}(I))}{\max(\mathrm{SMot}(I))} \qquad (13)$$

If a huge amount of pixels have large motion important, then the value is near to the limit. On the other hand, if there are lesser pixels the art are nearer to the maximum motion important measure, the value returns a lower value and α evaluates to a lower value indicating that the impact of motion is not huge enough for the motion important to overcome the main process.

The location of the ellipse in frame t is calculated using the state information calculated from (8). Fig. 3 shows the results of entity extracting on two various videos. The first row of Fig. 3 shows a tracking shot of a lynx moving on a snow covered region. Even though the background consists of large movement, the space and time important map (to be described in the next section) is able to clearly identify the important entity allowing the particles to dissolve and extrct the lynx reliably. The bottom row is also an example of a tracking shot of a skier in action. The shot demonstrates the possibility of the framework to detect different important entities at different times. Initially, the pine trees in the background are less important as shown in Fig. 3b. However, the skier is detected and tracked successfully in the consequent frames as shown in Figs. 3d and f.

The weights of the particles are reinitialized every t frames in order to avoid problems on a particular entity for the entire duration of the video sequence. In our experiments, we set t= F/2, where F is the frame rate of the video. We described how the suggested framework avoids fixation on a particular entity through an example shown in Fig. 3 second row. We provide another example in Section 5 describing how the suggested method might erroneously detect the wrong entity as important, but with particle weight reinitialization, there is quick recovery from the error. It can be noted that the suggested model can be used to detect multiple entities of interest by clustering and segmenting the video frame based on the space and time important values.

This would allow pixels with similar important measures and those that are spatially close to the clustered together. The cluster, which represents a region/ entity is associated with a rank based on the average important value of the cluster. Multiple particle trackers can be initialized around each of the cluster cancroids as seed points, thus, building a framework that would consider multiple hypothesis based on different color models that are obtained from each cluster and whose weights are calculated based on the color composition of the segment.

The important entity detection approach presented assumed the availability of a space and time important map. In the next section, we describe how such a map can be obtained at the pixel level with accurate efficiency.

The space and time important map generated for the three video sequences discussed earlier are shown in Fig. 4 d. The motion provided weights allow for a accurate estimate of the space and time important measure when the video is effected by a large motion contrast. The first row in Fig. 4 d is an example where the motion evaluation plays a major role in the space and time important measure as the most important shades are not very separate for the foreground important entity and background.

The frames shown in the middle and lower row of Fig. 4 d are examples where the camera tracks an entity of interest. In this case, the saliency motion values are available in the background even though the motion important is not huge enough until the entity had its own local value. In these two tasks, the space important map takes over the maximum amount of the space and time important measure as the motion difference is not large enough for the motion map to get control.

The location of the ellipse in frame t is calculated using the state information calculated from (8). Fig. 3 shows the results of entity extracting on two various videos. The first row of Fig. 3 shows a tracking shot of a lynx moving on a snow covered region. Even though the background consists of large movement, the space and time important map (to be described in the next section) is able to clearly identify the important entity allowing the particles to dissolve and extrct the lynx reliably. The bottom row is also an example of a tracking shot of a skier in action. The shot demonstrates the possibility of the framework to detect different important entities at different times. Initially, the pine trees in the background are less important as shown in Fig. 3b. However, the skier is detected and tracked successfully in the consequent frames as shown in Figs. 3d and f.

The weights of the particles are reinitialized every t frames in order to avoid problems on a particular entity for the entire duration of the video sequence. In our experiments, we set t= F/2, where F is the frame rate of the video. We described how the suggested framework avoids fixation on a particular entity through an example shown in Fig. 3 second row. We provide another example in Section 5 describing how the suggested method might erroneously detect the wrong entity as important, but with particle weight reinitialization, there is quick recovery from the error. It can be noted that the suggested model can be used to detect multiple entities of interest by clustering and segmenting the video frame based on the space and time important values.

This would allow pixels with similar important measures and those that are spatially close to the clustered together. The cluster, which represents a region/ entity is associated with a rank based on the average important value of the cluster. Multiple particle trackers can be initialized around each of the cluster cancroids as seed points, thus, building a framework that would consider multiple hypothesis based on different color models that are obtained from each cluster and whose weights are calculated based on the color composition of the segment.

The important entity detection approach presented assumed the availability of a space and time important map. In the next section, we describe how such a map can be obtained at the pixel level with accurate efficiency.

4.RESULT SCREENS

SIMULATION RESULTS

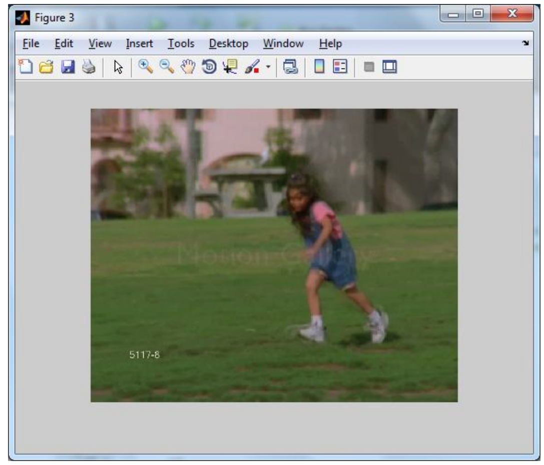

Figure 1.Input video of a girl running

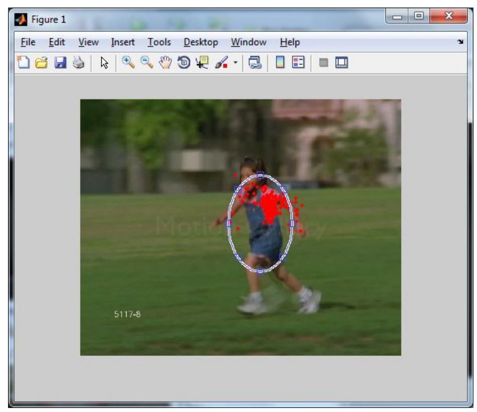

Figure 2.Comparing the pixel values of the given input video with the defined parameters

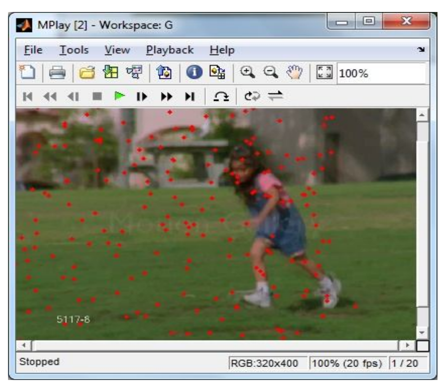

Figure 3. Creating the frames of the video

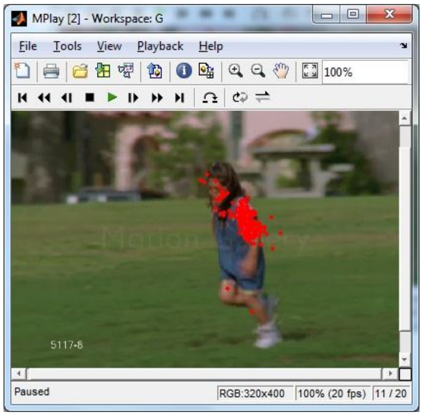

Figure 4.Calculating the smoothing factor,sensitivity parameter of the video

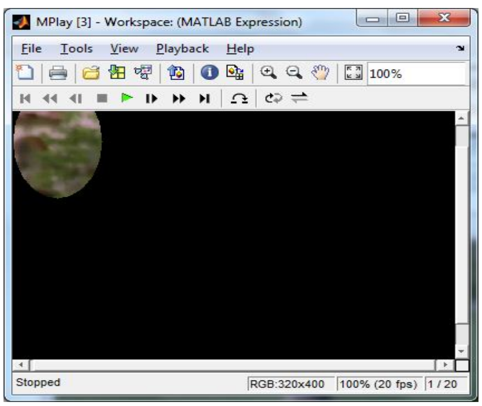

Figure 5. Extracted video of only the girl ignoring the background

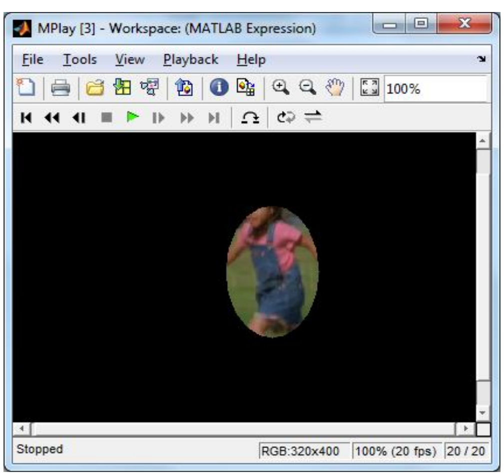

Figure 6. Extracted girl running in the garden

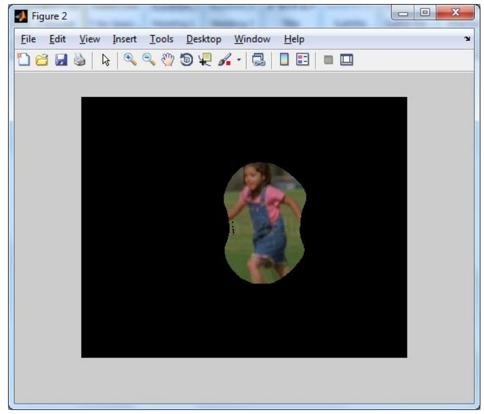

Figure 7.Image extracted from the given input video

5. conclusion

We have proposed an algorithm for important entity detection in videos based on saliency that uses space and time important maps and color as a solution. The performance is calculated on dividing set of data. We also develop a simple algorithm to produce space and time important map that can performs many state-of-the-art methods. As a future work, we extend the results of the important entity detection framework to be made useful for larger frame size of the video.

REFERENCES

[1] 'Co-salient object detection from multiple images' H Li, F Meng, KN Ngan - IEEE Transactions on Multimedia, 2013

[2] 'Salient object detection via global and local cues' N Tong, H Lu, Y Zhang, X Ruan - Pattern Recognition, 2015 – Elsevier

[3]'Prognostics of PEM fuel cell in a particle filtering framework' M Jouin, R Gouriveau, D Hissel, MC Péra… - International Journal of …, 2014 – Elsevier

[4]' A survey of computer vision-based human motion capture' TB Moeslund, E Granum - Computer vision and image understanding, 2001 - Elsevier

[5]'Introductory digital image processing' A remote sensing perspectiveJohn R. Jensen &Dr. Kalmesh Lulla 2008

[6]'Image processing: dealing with texture'M Petrou, PG Sevilla– 2006

[7]'Spatiotemporal saliency detection and its applications in static and dynamic scenes'

W Kim, C Jung, C Kimon Circuits and Systems for Video 2011

[8]'A principled approach to detecting surprising events in video'L Itti, P Baldi - Computer Vision and Pattern Recognition, 2005

[9]'Summarizing visual data using bidirectional similarity' D Simakov, Y Caspi, E Shechtman, 2008