Parallel Processing Of Visual and Motion Saliency from Real Time Video

Vaishali R. Khot
Dept. of Comp. Sci. and Engg.
ADCET Ashta, Shivaji University,
Kolhapur, India.
vaishalikhot25@gmail.com

A. S. Tamboli
Asst. prof. Dept.of Info. Tech. and engg
ADCET Ashta, Shivaji University,
Kolhapur, India.
shikalgar.arifa@gmail.com

Abstract— Extracting moving and salient objects from videos is important for many applications like surveillance and video retargeting .The proposed framework extract foreground objects of interest without any user interaction or the use of any training data(Unsupervised Learning).To separate foreground and background regions within and across video frames, the proposed method utilizes visual and motion saliency information extracted from the input video. The Smoothing filter is very helpful in characterizing fundamental image edges, i.e. salient edges and can simultaneously reduce insignificant details, hence produces more accurate boundary information. Our proposed model uses smoothing filter to reduce the effect of noise and achieve a better performance. Proposed system uses real time video data input as well as offline data to process using parallel processing technique. A conditional random field can be applied to effectively combine the saliency induced features. To evaluate the performance of saliency detection methods, the precision-recall rate and F-measures are utilized to reliably compare the extracted saliency information.

Keywords— *Video Object Extraction, Visual Saliency, Motion Saliency, Smoothing filter, Conditional Random Field.*

*****

I. INTRODUCTION

Visual saliency is a term refers to the meaningful region of an image. It is the perceived quality ,which is used for evaluating whether an object attracts viewers attention. Saliency of a region depends on its uniqueness, unpredictability, or rarity, and is highly related to the edge ,colour, boundaries , and gradient, which are the main sources of visual stimuli[13]. It is known that there are two stages of visual processing: i.e. bottom up and top-down. Bottom up factors highlight image regions that are different from their surroundings. To detect the salient region from an image it should have the following characteristics :

- Highlight whole salient regions from an image.
- Emphasize the whole salient object.
- Establish boundaries of extremely salient objects.
- Efficiently output large resolution saliency maps.
- Disregard high frequencies acquired from noise, blocking artifacts and texture.

Object segmentation in image processing is very important to application areas such as human computer interaction, content-based video coding, and multi-object tracking[14]. To robustly differentiate independently moving objects in a video sequence, the strategy to integrate motion(temporal) and visual(spatial) information from the video sequence is a key issue throughout the image segmentation process. To address these issues in a video sequence the proposed strategy uses visual and motion based saliency features for extracting foreground object in the scene. Saliency detection in static image as well as in video sequences has attracted much attention in recent years[15].This method is different from conventional segmentation problem of separating the whole scene into discrete parts, saliency detection finds semantic regions and filter out the unimportant area from the scene. The idea of saliency detection comes from human visual system, in that case the first stage of human vision is a fast but simple process. Saliency detection is an important research topic in computer vision, since it provides a fast pre-processing stage for many computer vision applications.

It is needed to detect the object from a rapid scene without training sequences Proposed system focus on Video Object Extraction (VOE) problems for single concept videos (i.e., videos which have only one object category of interest presented), proposed method is able to deal with multiple object instances (of the same type) with pose, scale, etc. variations. Proposed system follows the visual saliency and motion saliency technique for video object extraction because of the following reason:

1. Detection of visually salient image regions is useful for applications like object segmentation from an image, object recognition from video sequences , and adaptive compression.
2. Used in security applications like monitoring people and vehicle activities captured by CCTV systems.

Related work

1) *Visual and Motion Saliency:*

General saliency detection or image segmentation tasks are solved in an unsupervised setting. Based on spectrum analysis, Hou and Zhang [9] utilized the spectral residual as saliency information. This method is limited to static images. Achanta [10] computed the saliency by taking symmetric surrounding pixels into consideration and averaging the color differences between pixels within each region. Goferman et al. [11] applied multi-scale patches and calculated both color differences and locations between different patches. Evaluated the contribution of context aware saliency in only two applications such as retargeting and summarization.

2) *Video Object Extraction and Image Segmentation:*

In [2] decomposed an object shape model in a hierarchical way to train object part detectors, and these detectors are used to describe all possible configurations of the object of interest.Can not capture the details of the objects. Another type of supervised methods requires user interaction for annotating candidate foreground regions. Image segmentation algorithms proposed in [3], [4] focused on an interactive scheme and required users to manually provide the ground truth label information. The major difficulty with energy minimization lies in the enormous computational costs. Turbopixels proposed by [12] for segmentation, and the resulting image segments (superpixels) are applied to perform saliency detection. The use of Turbopixels allows to produce edge preserving superpixels with similar sizes.Some unsupervised approaches aim at observing features associated with the foreground object for VOE. Graph-based methods [5] identify the foreground object regions by minimizing the cost between adjacent hidden nodes/pixels in terms of colour, motion, etc. information. This works on only single concept video.

3) *Conditional Random Field:*

For videos captured by a monocular camera, methods such as Criminisi et al., Yin et al. [6] applied a conditional random field (CRF) maximizing a joint probability of color, motion, etc. models to predict the label of each image pixel.It has Limitations with respect to stereo-based segmentation. Although the color features can be automatically determined from the input video, these methods still need the user to train object detectors for extracting shape or motion features.Proposed to use some preliminary strokes to manually select the foreground and background regions, and they utilized such information to train local classifiers to detect the foreground objects [7]. In principle, this can be solved with enough user interactions, but could be tedious. Some unsupervised approaches aim at observing features associated with the foreground object for VOE. Leordeanu and Collins [8] proposed to observe the co-occurrences of object features to identify the foreground objects in an unsupervised setting. Errors in the matching of interest point increase as the window size increases. Graph-based Method [5] identifies the foreground object regions by minimizing the cost between adjacent hidden nodes/pixels in terms of color, motion, etc. information.

II. Proposed System

In this paper, we present the video object extraction from real time video using parallel processing technique. With our system, fast video object extraction can be performed without the need for any user interaction or the use of any training data. In past, no online video data were considered. But here, work is going on the online video data. This paper uses parallel processing techniques to process the online video frames.

A. *System Overview*

The system described in this paper follows the block diagram shown in fig. 1. It consists of five modules: The online video buffering and framing where online video is captured through web camera simultaneously frames are generated for the further task, the pre-processing, where noise reduction takes place, the visual saliency extraction where image segmentation is performed, the motion saliency extraction where colour and shape information is retrieved by doing forward and backward processing of frames , and the conditional random field combines the extracted features.

The proposed system can be implemented using parallel processing technique.In parallel processing, we are using CUDA platform to perform master slave type of parallelization. In which the master will divide the work among all other slaves and work can be performed , in this CUDA cores are used to do parallelization. The main idea is to divide the problem into simple tasks and solve them concurrently, using CUDA as parallelization technique. Each CUDA core will process per second frames i.e. 30 frames.

Proposed System Works as follows :

- Capture live video from system web camera or CCTV camera
- Extract frames from video and store it sequentially in buffer with sequence no.
- Frames per second will be given to different cores using CUDA platform at each core threads will perform the further task.
- Each thread will perform filtering , Visual feature and Motion feature extraction.
- CRF will integrate all these features of and final output will be an image

B. *Pre-processing*

Each video frame contains some sort of noise in terms of blurred edges or pixels throughout the individual frame of respective video data.We are using smoothing filter to reduce the effect of noise and achieve a better performance. The filter is extremely helpful in characterizing salient edges of an object of interest from an image/video sequences and can simultaneously reduce insignificant details, hence produces more accurate boundary information.

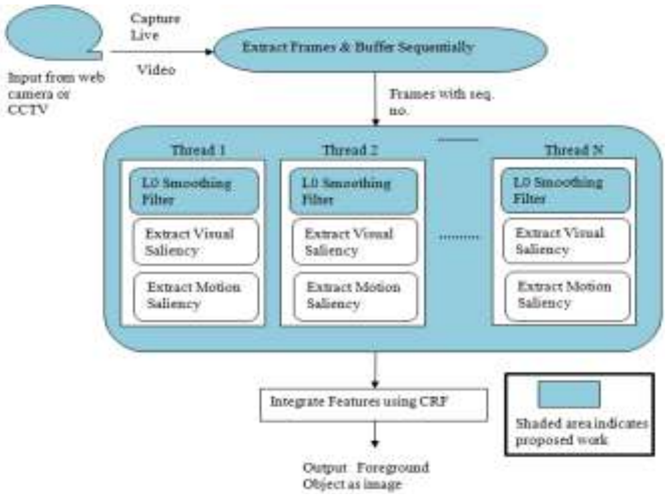
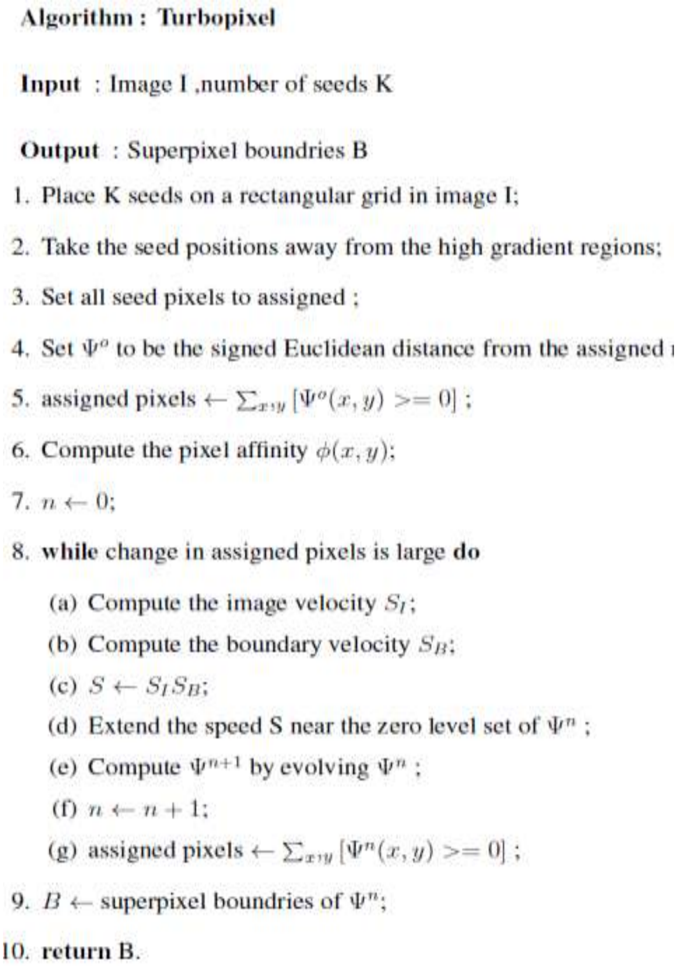

Capture Live
Video
Input from web camera or CCTV
Extract Frames & Buffer Sequentially
Frames with seq. no.
Thread 1
Thread 2
Thread N
L0 Smoothing Filter
Extract Visual Saliency
Extract Motion Saliency
....
Integrate Features using CRF
Shaded area indicates proposed work
Output Foreground Object as image

Fig 1. A block diagram of the proposed system

C. *Extraction of visual saliency:*

To extract visual saliency of each frame, we have to perform image segmentation on each video sequence. In our work, we Turbo pixels before segmentation, and the processed image segments (super pixels) are applied to perform saliency detection of an object. Turbopixel algorithm is used to generate superpixels the algorithm is as follows :



```
Algorithm : Turbopixel
Input  : Image I ,number of seeds K
Output  : Superpixel boundries B
1. Place K seeds on a rectangular grid in image I;
2. Take the seed positions away from the high gradient regions;
3. Set all seed pixels to assigned ;
4. Set $\Psi^o$ to be the signed Euclidean distance from the assigned
5. assigned pixels $\leftarrow \sum_{x,y} [\Psi^o(x, y) >= 0]$ ;
6. Compute the pixel affinity $\phi(x, y)$;
7. $n \leftarrow 0$;
8. while change in assigned pixels is large do
    (a) Compute the image velocity $S_I$;
    (b) Compute the boundary velocity $S_B$;
    (c) $S \leftarrow S_I S_B$;
    (d) Extend the speed S near the zero level set of $\Psi^n$ ;
    (e) Compute $\Psi^{n+1}$ by evolving $\Psi^n$ ;
    (f) $n \leftarrow n + 1$;
    (g) assigned pixels $\leftarrow \sum_{x,y} [\Psi^n(x, y) >= 0]$ ;
9. $B \leftarrow$ superpixel boundries of $\Psi^n$;
10. return B.
```

Turbo pixel segments an image into a lattice-like structure of compact regions (set of pixels) called superpixels. For the kth super pixel r_k, we are going to calculate its saliency score $S(r_k)$ as follows

$$S(r_k) = \sum_{r_k \neq r_i} exp(D_s(r_k, r_i)/\sigma_s^2)\omega(r_i)D_r(r_k, r_i)$$

$$\approx \sum_{r_k \neq r_i} exp(D_s(r_k, r_i)/\sigma_s^2)D_r(r_k, r_i)$$

Where D_s is the Euclidean distance between the centroid of r_k and that of its surrounding superpixels r_i,width of the kernel is controlled by σ_s. The parameter ω (r_i) is the weight of the neighbor super pixel ri ,which is proportional to the number of pixels in r_i. (r_i) can be as [1- max(r_i / r_k)].

D. Extraction of motion saliency :

At extracting motion salient regions based on the retrieved optical flow information. To detect each pixels from moving part , perform dense optical-flow forward at each frame of a video sequence and backward propagation. A moving parts pixel q_t at each frame t is determined by:

$$q_t = \hat{q}_t, t-1 \cap \hat{q}_t, t+1$$

Where $\hat{q}_t$ denotes the pixel pair detected by forward or backward optical flow propagation. Derived optical flow results to calculate the motion saliency M (i, t) for each pixel i at each frame t, and the saliency score is normalized to the range of [0, 1] at every frame of video sequence.To describe the motion salient regions, we will convert the motion saliency image into a binary output and extract the shape information from the motion salient regions. We first binaries and divide each video frame into disjoint 8*8 pixel patches.

E. Conditional Random Field :

The extracted features are combined in buffer sequentially and images are constructed by motion-induced shape information, color information of the foreground object by following
equation.

$$\hat{X}_t^s = \sum_{n\epsilon I_t}\sum_{k=1}^{K}(a_n, k.M_k)$$

Where $\hat{X}_t^s$ serves as the likelihood of foreground object at frame t, where n; k is the weight for the nth patch using the kth Codeword.

. The Gaussian mixture model estimates probability density functions (PDF) for each class, and then performs classification on that class based on Bayes rule :
Where $P(X_j/C_j)$ is the PDF of class j,evaluated at X, $P(C_j)$ is the prior probability for class j, and P(X) is the overall probability distribution function ,evaluated at X.

It is not like the unimodal Gaussian model, which assumes $P(X_j/C_j)$ to be in the form of a Gaussian, the Gaussian mixture model estimates $P(X_j/C_j)$ as a weighted average of multiple Gaussians.

$$P(X|C_j) = \sum_{k=1}^{Nc} W_k G_k$$

Where w_k is the weight of the k-th Gaussian G_k and the weights sum to one .One such probability distribution function for each class. Free parameters of the Gaussian mixture model consist of the means and covariance matrices of the Gaussian components and the weights indicating the contribution of each Gaussian to the approximation of $P(X_j/C_j)$. Conditional random field is a technique to estimate the structural information of a set of variables with the associated observations. CRF has been applied to predict the label of each observed pixel in an image I.With energy minimizatin task and foreground and background colour model CRF performs Video Object Extraction in form of image.

III. CONCLUSIONS

In this paper, we presented the video object extraction from a real time video using parallel processing techniques.This system is advantageous in several vision applications e.g. surveillance etc. Fast preprocessing stage for many vision application is provided through saliency detection which includes Visual and Motion saliency.Here both types of input data is used i.e. offline video as well as online video data. Proposed work uses smoothing filter to remove the noise from the extracted frames of the video data.

To detect the important part(salient) from the video both visual and motion saliency can be used hence we can detect the object of interest from input data effectively.We have worked on smoothing filter and the segmentation of the video frames, which gives the salient image from the respective frames on CUDA platform . Further work can be carried out on Extraction of Motion saliency ,and Conditional random field is applied to give the final output.

ACKNOWLEDGMENT

Vaishali R. Khot Author thanks Ms. A. S. Tamboli, project guide, Department of Information Technology, Kolhapur, India, for her precious guidance, encouragement and continuous monitoring throughout the presented work.

REFERENCES

[1] Wei-Te Li, Haw-Shiuan Chang, Kuo-Chin Lien, Hui-Tang Chang, and Yu-Chiang Frank Wan,"Exploring Visual and Motion Saliency for Automatic VideoObject Extraction ," IEEE Transactions on image processing, vol. 22, no. 7, july 2013.

[2] B. Wu and R. Nevatia," Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses," Int. J. Comput. Vis.,vol. 82, no. 2, pp. 185204, 2009.

[3] Y. Boykov, O. Veksler, and R. Zabih, " Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 23, no. 11, pp. 12221239, Nov. 2001.

[4] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts ," ACM Trans. Graph., . 23, no. 3, pp. 309314, 2004.

[5] K.-C. Lien and Y.-C. F.Wang, "Automatic object extraction in single concept videos," Proc.IEEE Int. Conf. Multimedia Expo, Jul. 2011, pp. 16.

[6] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, "Bilayer segmentation of live video," Proc. IEEE Conf. Comput. Vis. Pattern Recognit,Jun. 2006, pp. 5360.

[7] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting ," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2007, pp. 18.

[8] M. Leordeanu and R. Collins, "Unsupervised learning of object features from video sequences," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 11421149.

[9] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach ," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,Jun. 2007, pp. 18.

[10] R. Achanta and S. Ssstrunk, "Saliency detection using maximum symmetric surround ," in Proc. IEEE Int. Conf. Image Process., Sep. 2010, pp. 26532656.

[11] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection ," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 23762383.

[12] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, and S. J. Dickinson, "TurboPixels: Fast superpixels using geometric flows," IEEE Trans. Pattern Anal. Mach. Intell.,l.vol. 31, no. 12, pp. 22902297, Dec. 2009.

[13] S.B.Choudhari and N.M.Shahane ,"Novel Approach for salient region detection",in IJARCCE ,vol.3,Issue6,June 2014.

[14] N.P.Varkey and S.Arumugam,"Automatic video object exteaction by using generalized visual and motion saliency",IJISER,vol 1,Issue 4.Apr 2014.

[15] X.Ciu,Q.Liu,S.Zhang,F.Yang and D.M.Metaxas,"Temporal Spectral Residual for fast salient motion detection", Neurocomputing 86 (2012) 24–32