

AUTOMATIC IDENTIFICATION OF ANIMAL USING VISUAL AND MOTION SALIENCY

G.N.Muruganathan¹, R.Sathish²

¹Professor, EEE Department, Kumaraguru College of Technology, Tamil Nadu, India

²PG Scholar, EEE Department, Kumaraguru College of Technology, Tamil Nadu, India

Abstract

This paper examines the problem of detecting various types of animals in video sequence taken in the wild. Animal detection is useful in prevention of animal- vehicle accidents and will increase human and wildlife safety. We propose a fully automated method for detecting large animals before they enter the road and warn the driver through audio and visual signals, this also helps in saving crops in farm from animals. The proposed framework aims to automatically extract foreground objects of interest without any user interaction or the use of any training data (i.e., not limited to any particular type of object). 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. A conditional random field is applied to effectively combine the saliency induced features, which allows us to deal with unknown pose and scale variations of the foreground object (and its articulated parts). Based on the ability to preserve both spatial continuity and temporal consistency in the proposed VOE framework, experiments on a variety of videos verify that our method is able to produce quantitatively and qualitatively satisfactory VOE results. Neural network is used to find the animal type.

Keywords: Animal detection, segmentation, Video Object Extraction (VOE), visual saliency, motion saliency

-----***-----

1. INTRODUCTION

Human can easily determine the subject of interest in a video, even though that subject is presented in an unknown or cluttered background or even has never been seen before. With the complex cognitive capabilities exhibited by human brains, this process can be interpreted as simultaneous extraction of both foreground and background information from a video. Many researchers have been working toward closing the gap between human and computer vision. However, without any prior knowledge about the subject of interest or training data, it remains as a challenging job for computer vision algorithms to automatically extract the foreground object of interest from a video. As a result, if one needs to design an algorithm to automatically extract the foreground objects from a video, tasks stated below need to be addressed.

- 1) Unknown object category and unknown number of the object instances in a video.
- 2) Complex or unexpected motion of foreground objects due to articulated parts or arbitrary poses.
- 3) Ambiguous appearance between foreground and background regions due to similar color, low contrast, insufficient lighting, etc. conditions.

In practice, it is infeasible to manipulate all possible foreground object or background models beforehand. However, if one can extract representative information from either foreground or background (or both) regions from a video, the extracted information can be utilized to differentiate

foreground and background regions, and thus the task of foreground object extraction can be addressed. As discussed later in Section II, most of the prior works either consider a fixed background or assume that the background exhibits dominant motion across video frames. These assumptions may not be practical for real world applications, because they cannot generalize well to videos captured by freely moving cameras with arbitrary movements.

In this paper, a robust video object extraction (VOE) framework, which utilizes both visual and motion saliency information across video frames is proposed. The observed saliency information allows us to infer several visual and motion cues for learning foreground and background models, and a conditional random field (CRF) is applied to automatically determine the label (foreground or background) of each pixel based on the observed models. With the ability to preserve both spatial and temporal consistency, our VOE framework exhibits promising results on a variety of videos, and produces quantitatively and qualitatively satisfactory performance.

The paper is organized as follows: Section 2 deals with visual saliency by super pixel segmentation, Section 3 gives a brief of motion saliency feature extraction, Section 4 describes the VOE by EM segmentation, simulation results are presented in section 5 and finally conclusion.

2. VISUAL SALIENCY BY SUPER PIXEL SEGMENTATION

To extract visual saliency of each frame, we perform image segmentation on each video frame and extract color and contrast information. In our work, we advance Turbopixels proposed [26] for segmentation, and the resulting image segments (superpixels) are applied to perform saliency detection. The use of Turbopixels allows us to produce edge preserving superpixels with similar sizes, which would achieve improved visual saliency results as verified later. For the k th superpixel r_k , saliency score $S(r_k)$ is calculated as follows:

$$s(r_k) = \sum_{r_i=r_i} \exp\left(\frac{D_s(r_k, r_i)}{\sigma_s^2}\right) \omega(r_i) D_r(r_k, r_i) \quad (1)$$

$$\approx \sum_{r_i=r_i} \exp\left(\frac{D_s(r_k, r_i)}{\sigma_s^2}\right) D_r(r_k, r_i) \quad (2)$$

where D_s is the Euclidean distance between the centroid of r_k and that of its surrounding superpixels r_i , while σ_s controls the width of the kernel. The parameter $\omega(r_i)$ is the weight of the neighbor superpixel r_i , which is proportional to the number of pixels in r_i . Compared to [27], $\omega(r_i)$ can be treated as a constant for all superpixels due to the use of Turbopixels (with similar sizes). The last term $D_r(r_k, r_i)$ measures the color difference between r_k and r_i , which is also in terms of Euclidean distance.

As suggested by [22], we consider the pixel i as a salient point if its saliency score satisfies the condition $S(i) > 0.8 * \max(S)$, then the collection of the resulting salient pixels will be considered as a salient point set. Since image pixels which are closer to this salient point set should be visually more significant than those which are farther away, we further refine the saliency

$S(i)$ for each pixel i as follows:

$$\hat{s}(i) = s(i) * \left(1 - \frac{\text{dist}(i)}{\text{distmax}}\right) \quad (3)$$

Where $S(i)$ is the original saliency score derived by (1), and $\text{dist}(i)$ measures the nearest Euclidean distance to the salient point set. We note that distmax in (2) is determined as the maximum distance from a pixel of interest to its nearest salient point within an image, thus it is an image-dependent constant.

3. MOTION SALIENCY BASED FEATURE EXTRACTION

Unlike prior works which assume that either foreground or background exhibits dominant motion, our proposed framework aims at extracting motion salient regions based on

the retrieved optical flow information. To detect each moving part and its corresponding pixels, we perform dense optical-flow forward and backward propagation [28] at each frame of a video. A moving pixel q_t in frame t is determined by

$$q_t = \hat{q}_{t, t-1} \hat{q}_{t, t+1} \quad (4)$$

Where \hat{q} denotes the pixel pair detected by forward or backward optical flow propagation. The frames are not ignored which result in a large number of moving pixels at this stage as [13], [14] did, and thus our setting is more practical for real-world videos captured by freely-moving cameras.

After determining the moving regions, we propose to derive the saliency scores for each pixel in terms of the associated optical flow information. Inspired by visual saliency approaches like [27], we apply our proposed algorithms in (1) and (2) on the derived optical flow results to calculate the motion saliency $M(i, t)$ for each pixel i at frame t , and the saliency score at each frame is normalized to the range of [0, 1]. It is worth noting that, when the foreground object exhibits significant movements (compared to background), its motion will be easily captured by optical flow and thus the corresponding motion salient regions can be easily extracted. On the other hand, if the camera is moving and thus results in remarkable background movements, the proposed motion saliency method will still be able to identify motion salient regions (associated with the foreground object), as verified later by our experiments.

4. VOE BY EM SEGMENTATION

Zhang Xiaoyan, Liu Lingxia, Zhuang Xuchun implemented a new scheme for performing automatic video object segmentation. First they proposed a new adaptive and reliable threshold estimation method for the change detection process. Secondly the early edge model of the video object is obtained on the basis of the adaptive change detection and canny edge detection. Thirdly the edge model has been tracked and updated in every frame to accommodate rotation and changes in shape on the tracked object and then the outer contour of the video object is determined based on the mathematical morphological operation. Lastly, an improved active contour which uses the gradient vector as the external force is applied to guide the initial outer contour moving to the actual video object contour. Experiment results show that the proposed algorithm is strong to the entire motion and local twist of object adaptable to the complex background and can get the closed and accurate video object contour [3].

As per Guo Lihua the video object segmentation is the main section of digital video representation, transmission and manipulation. For example, application including content based video retrieval, object-based video coding and many more. In this paper he proposed the fast and automatic video segmentation method with the aim at foreground and

background segmentation through efficient mixture of color and motion analysis module. He started with the watershed segmentation algorithm to provide initial regions as per the pixels luminance gradient, then regions are merged according to their color and motion connection. Finally, the semantic video article will be obtained after post processes. Advantage of this method is its fast and regular execution of video object segmentation. [6]

The key idea in our approach is to reduce superpixel computation to an efficiently solvable geometric-flow problem.

Our approach is guided by five basic principles:

(a) Uniform size and coverage: Superpixel segmentation should partition an image into regions that are approximately uniform in size and shape (compactness), minimizing region under segmentation, provided that superpixel size is comparable to the size of the smallest target region. By designing a geometric flow that dilates an initial set of uniformly distributed seeds, where each seed corresponds to one superpixel. The seeds behave initially like reaction-diffusion bubbles [24].

(b) Connectivity: Each superpixel should represent a simply connected set of pixels. Our dilation-based flow combined with its level-set implementation, ensures that this constraint is always satisfied.

(c) Compactness: In the absence of local edge information, superpixels should remain compact. Our flow begins from circular seeds and assumes no prior bias on the location of superpixel boundaries. To maximize compactness, we include a term that produces constant motion in the direction of the outward normal in regions of uniform intensity. This term maximizes the rate of area growth, while retaining the minimum possible isoperimetric ratio, which is 4 for a circular region.

(d) Smooth, edge-preserving flow: When growth stops, superpixel boundaries should coincide with image edges.

This requires a geometric-flow formulation with three properties: 1) It should slow down boundary growth in the vicinity of edges, 2) it should be attracted to edges, and 3) it should produce smooth boundaries. To do this, we borrow ideas from work on geometric active contours [1], [2], [12], [13], [14]. Such formulations provide an easy way to incorporate image-based controls on boundary growth, and include both a "doublet" term for attraction and a curvature term for shape regularization.

(e) No superpixel overlap: A superpixel segmentation should assign every pixel to a single superpixel. Therefore, boundary evolution should stop when two distinct dilating seeds are about to collide. To achieve this, we incorporate a simple skeleton-based mechanism for collision detection in the background.

These considerations lead to a geometric-flow-based algorithm, that we call TurboPixels, whose goal is to maintain

and evolve the boundary between the assigned region, which contains all pixels that are already inside some superpixel, and the unassigned region, which contains all other pixels. At a conceptual level, the algorithm consists of the following steps:

1. place initial seeds;
2. iterate over the following basic steps until no further evolution is possible, i.e., when the speed at all boundary pixels is close to zero,
 - a. evolve this boundary for T time steps;
 - b. estimate the skeleton of the unassigned region;
 - c. update the speed of each pixel on the boundary and of unassigned pixels in the boundary's immediate vicinity.

5. SIMULATION RESULTS

The following are the simulation results of the automatic animal identification [Figure 1 to 6]

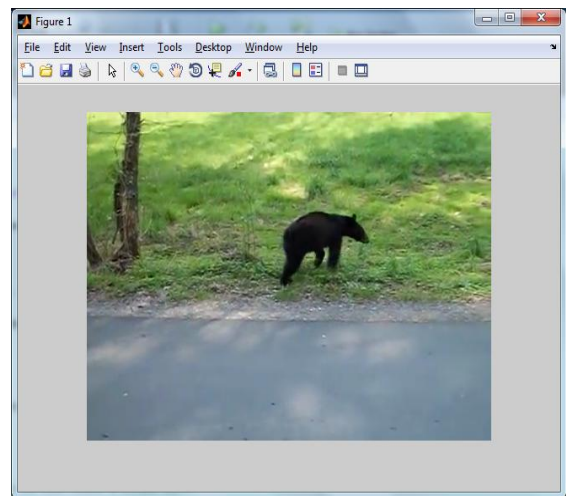


Fig-1: Input video frame

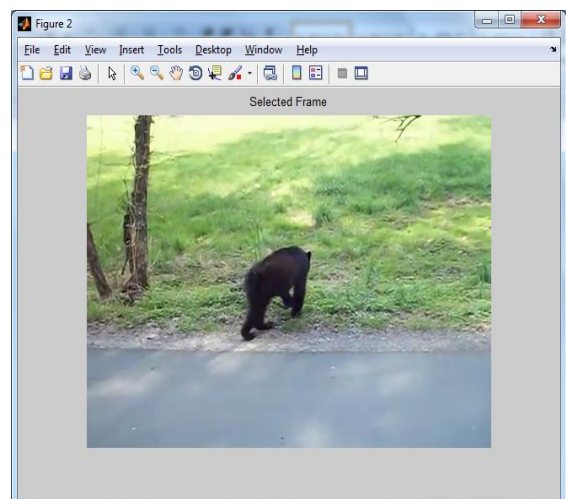


Fig-2: Selected frame from the video

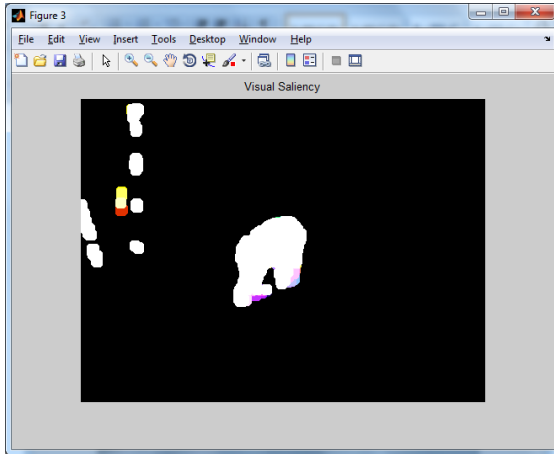


Fig-3: Result of visual saliency

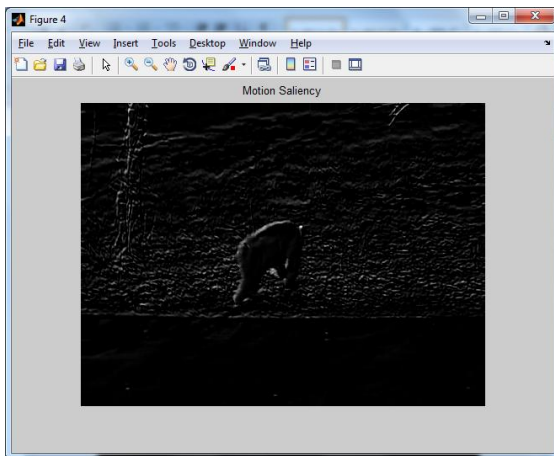


Fig-4: Result of motion saliency

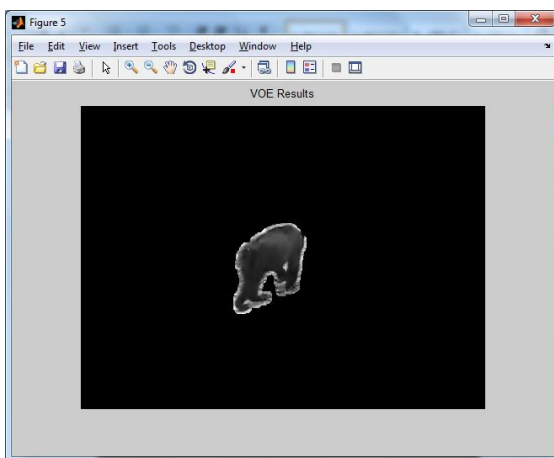


Fig-5: Result of VOE

From Figure 1 to 6 it is inferred that the animal detection pre-processing steps are done and the bear is detected in the VOE result. The performance analysis graph shows that the

proposed method gives the minimum segmentation score when compared to the conventional segmentation methods

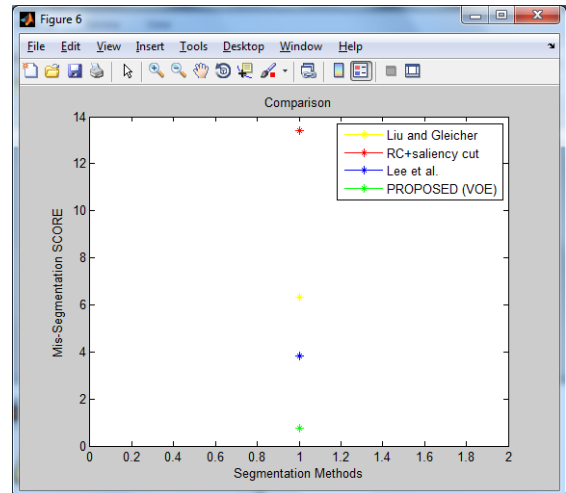


Fig-6: Performance analysis

6. CONCLUSIONS

Intelligent farm surveillance system refers to the video level processing techniques for identification of specific objects, in recorded videos of the farm. In my work, video is taken, to be a series of images and have extended the concept to identify animal from videos of the farm. Different image processing techniques have been surveyed and implemented to recognize animals in video more efficiently. Background subtraction methods like frame differencing, mixture of Gaussian and sum of absolute differences were implemented and tested on the video. Different image processing techniques have been trained and tested to recognize animals in an image efficiently. Image morphology using HSV color space worked well on all types of images compared to other color spaces. Visual and motion saliency were used followed by VOE by EM segmentation for detecting the animal. Segmentation results show that the proposed segmentation gives better result when compared to the existing methods.

REFERENCES

- [1]. 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. 185–204, 2009.
- [2]. Z. Lin and L. S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 604–618, Apr. 2010.
- [3]. Y. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[4]. C. Rother, V. Kolmogorov, and A. Blake, “GrabCut’: Interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[5]. A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, “Bilayer segmentation of live video,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 53–60.

[6]. Yin, A.Criminisi, J.M.Winn, and I. A.Essa, “Bilayer segmentation of webcam videos using tree-based classifiers,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 30–42, Jan. 2011.

[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. 1–8.

[8]. M. Gong and L. Cheng, “Foreground segmentation of live videos using locally competing 1SVMs,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2105–2112.

[9]. T. Bouwmans, F. E. Baf, and B. Vachon, “Background modeling using mixture of Gaussians for foreground detection—A survey,” *Recent Patents Comput. Sci.*, vol. 3, no. 3, pp. 219–237, 2008.

[10]. F.-C. Cheng, S.-C. Huang, and S.-J. Ruan, “Advanced background subtraction approach using Laplacian distribution model,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2010, pp. 754–759.

[11]. J. Zhong and S. Sclaroff, “Segmenting foreground objects from a dynamic textured background via a robust Kalman filter,” in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 44–50.

[12]. J. Sun, W. Zhang, X. Tang, and H.-Y. Shum, “Background cut,” in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 628–641.

[13]. F. Liu and M. Gleicher, “Learning color and locality cues for moving object detection and segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 320–327.

[14]. K.-C. Lien and Y.-C. F. Wang, “Automatic object extraction in singleconcept videos,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[15]. M. Leordeanu and R. Collins, “Unsupervised learning of object features from video sequences,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 1142–1149.

[16]. L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[17]. P. Harding and N. M. Robertson, “Visual saliency from image features with application to compression,” *Cognit. Comput.*, vol. 5, no. 1, pp. 76–98, 2012.

[18]. X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[19]. C. Guo, Q. Ma, and L. Zhang, “Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[20]. T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, “Learning to detect a salient object,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[21]. R. Achanta and S. Süssstrunk, “Saliency detection using maximum symmetric surround,” in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2653–2656.

[22]. S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2376–2383.

[23]. Y. Zhai and M. Shah, “Visual attention detection in video sequences using spatiotemporal cues,” in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.

[24]. Y.-F. Ma and H.-J. Zhang, “Contrast-based image attention analysis by using fuzzy growing,” in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 374–381.

[25]. W. Wang, Y. Wang, Q. Huang, and W. Gao, “Measuring visual saliency by site entropy rate,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2368–2375.

[26]. A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, and S. J. Dickinson, “TurboPixels: Fast superpixels using geometric flows,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

BIOGRAPHIES



G.N. Muruganathan received his M.E degree in Power system from Government college of technology, Coimbatore, India in 1987 and worked in Tamil Nadu Electricity Board, 35 and half years in various cadre as Assistant Engineer, Assistant Executive Engineer and Executive Engineer in Kundaha Hydro electric Project; lower Metro hydro electric project; Kadamparai Hydro Electric Project etc; now he is currently an Professor with Department of Electrical and Electronics Engineering, Kumaraguru College of Technology, Coimbatore, India.



R. Sathish currently doing M.E Embedded Systems in Kumaraguru College of Technology, Coimbatore, India He Received B.E Degree in Electronics and Communication Engineering in Kalaingar Karunanidhi Institute of Technology, Coimbatore, India He Received Diploma in Electronics and Communication Engineering in RMPTC, Udumalaipet, Tamil Nadu, India.