Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 78 (2016) 160 – 164

Procedia Computer Science

International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA

Detection of Moving Objects in Surveillance Video by Integrating Bottom-up approach with Knowledge Base

Aarthi.R[a] , Amudha.J[b],Boomika.K[c] ,Anagha Varrier[d]

[a]Department of CSE,Amrita Vishwa Vidyapeetham,Coimbatore, India
[b]Department of CSE,Amrita Vishwa Vidyapeetham,Bengaluru, India
[c,d]Student's Department of CSE,Amrita Vishwa Vidyapeetham,Coimbatore, India

Abstract

In the modern age, where every prominent and populous area of a city is continuously monitored, a lot of data in the form of video has to be analyzed. There is a need for an algorithm that helps in the demarcation of the abnormal activities, for ensuring better security. To decrease perceptual overload in CCTV monitoring, automation of focusing the attention on significant events happening in overpopulated public scenes is also necessary. The major challenge lies in differentiating detecting of salient motion and background motion. This paper discusses a saliency detection method that aims to discover and localize the moving regions for indoor and outdoor surveillance videos. This method does not require any prior knowledge of a scene and this has been verified with snippets of surveillance footages.

Peer-review under responsibility of organizing committee of the ICISP2015

Keywords: Video Surveillance; Visual Saliency; Moving Object Detection;

1. Introduction

We deal with a lot of information in our day-to-day life and there is always more to the picture than what meets the eye. Because of the sheer volume of information it often becomes challenging to demarcate between important and unimportant information. Such a separation makes a great difference in security, though, especially with surveillance cameras. Identifying an anomalous event could even help us save thousands of lives. Most of the computer vision methods are yet to catch up to the level of efficiency of the human eye, which is capable of swiftly focusing on the necessary part of a scenario. There are many applications trying to bridge this gap, though[1]. One such work is by Zhengzheng Tu et al., which proposed[2] a novel approach that combines Independent Component Analysis of optical flow with Principal Component Analysis to detect multiple moving objects in the context of complex outdoor scenes but it is restricted to traffic surveillance videos.

In case of surveillance videos, it is critical to be able to quickly identify aberrant incidents occurring in congested outdoor scenes. The method should aim to suppress dominant crowd flows, while focusing its attention on activities that deviate from the norm. For scalability and extensiveness of application on different surveillance videos it should be expeditious and need not require any preparatory knowledge. Salient motion detection becomes onerous especially when dynamic background motion obscures the movement. The visual saliency can be one of the step for many computer vision tasks to process the image without prior knowledge of their contents.

A wide spectrum of applications may benefit from separate processing of salient and non-salient regions in an image. One of those applications is providing an online solution that can detect the region of interest given a surveillance video sequence. To cater to this need the following three saliency detection algorithms are discussed

Peer-review under responsibility of organizing committee of the ICISP2015
doi:10.1016/j.procs.2016.02.026

and the performance is tested with multiple categories of images. A regional contrast based method evaluates global contrast differences and spatial coherence to extract the salient region[3]. The algorithm is simple, efficient, and yields full resolution saliency maps. The histogram-based approach targets natural scenes, and is suboptimal for extracting saliency of highly textured scenes[4]. The visual based saliency attention is inspired from the behavior of human visual system and the fact that visual attention is driven by low level stimulus. It is often found to be better in case of surveillance videos[5]. There are various applications that are existing which use the Bottom-up approach of visual saliency based detection[6]. Apart from this, various applications have used top-down knowledge obtained from images during training phase[7].Some applications have used location or colour based cue to obtain the region of interest[8].

In order to combat the difficulties of salient motion detection few methods were studied. In a State Space Controllability method exploits the dynamic background in order to estimate the salient foreground motion. The frames of the video is analyzed and depicted in form of the linear system, it generates an accurate saliency map for videos with both local and global motion. Periodic Motion Detection[9] detects the movement of multiple object by analyzing their periodic motion in space with the period p. it is treated as a collection of rigid 3D configurations reoccurring with frequency 1/p. The periodic features found in the previous runs of the algorithm and the presence of additional instances of periodic motion within the remaining features. Another method is creating a dynamic saliency map by calculating entropies in every corresponding local region between the current and the previous frames in a video. This map includes information from both the frames to detect motion.

Thus there are many studies being done where algorithms are tested with various videos with different complexities. In a comparison of performances of different algorithms it is observed that Laurent Itti's visual based saliency model provided the best results for surveillance videos. The dynamic saliency map provided accurate results and was optimized to improve efficiency. The subsequent sections of the paper will discuss how we have used Itti's model along with entropy calculation.

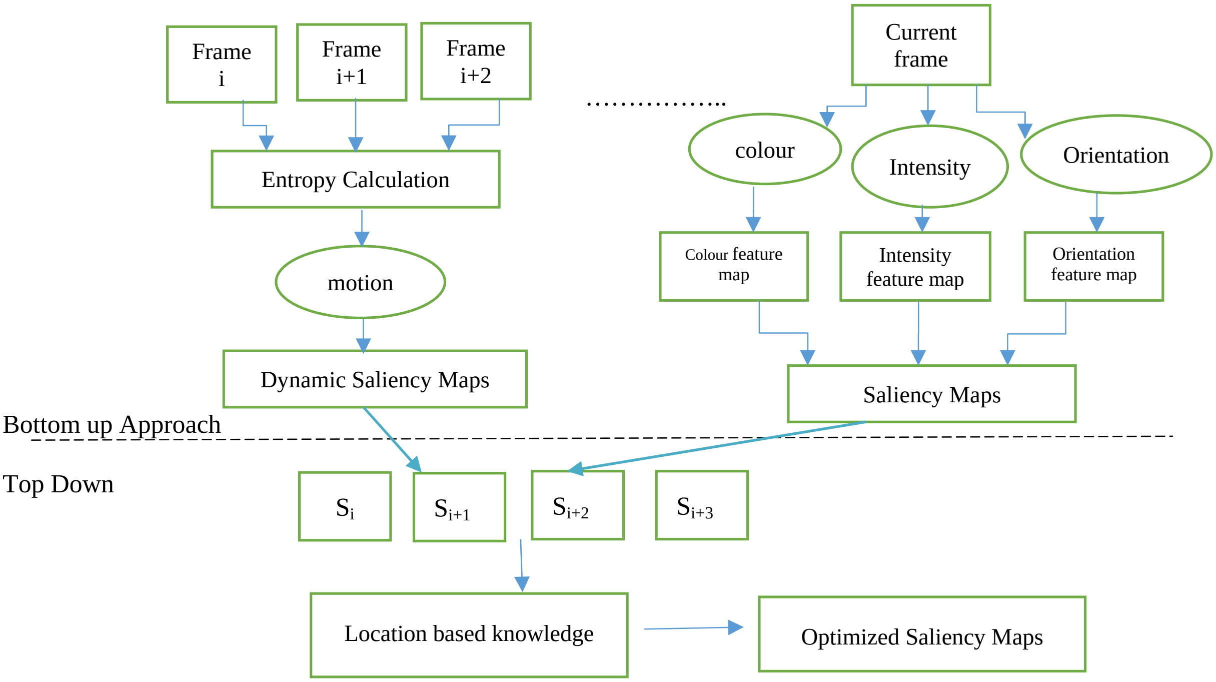


Fig. 1.Process of computing bottom up and top down saliency

2. Method

Visual saliency helps our brain to select the most salient region in real-time. Colour, intensity and orientation play an important role in determining the salient regions in images. In human vision system, moving object attracts more attention than other static object, so, while extending to video analysis - motion is an additional feature required to determine the salient region. The process of detecting moving object has been divided into two major portions calculation of saliency map with static and dynamic saliency as shown in Fig1.The unwanted background movement is later removed using top down knowledge.

2.1. Static Saliency

The Static saliency is determined by Visual based saliency detection model[10]. The model is developed based on the observation in human vision, that attention is driven by low level stimulus. Features like colour, intensity and orientation of the image are analysed in various resolutions and maps are computed on multiple levels. Colour feature maps are calculated on the basis of colour opponency of R-G and B-Y system. Intensity feature map is concerned with intensity contrast which is sensed as dark background with light surround or vice versa. Orientation information is obtained by using oriented Gabor pyramids for € $[0^0, 45^0, 90^0, 135^0]$. The colour, intensity, orientation feature maps are individually combined to give corresponding conspicuity maps. The normalised intensity map N(I) , colour map N(C) and orientation map N(O) are summed to obtain the final saliency map, as in equation (1):

$$S = (N(I) + N(C) + N(O)) \tag{1}$$

2.2. Dynamic Saliency Map

Moving objects grab more attention than stationary objects in a video. The dynamic saliency map helps to locate the moving object by analysing 'k' successive frames in the video. The map is created by calculating entropies in every corresponding local region between the current and previous 'k-1' frames in the video[11]. This map includes information not only from the current frame, but also from the previous frames.

In order to minimize the computation complexity the frames are converted to grayscale and quantised into 8 levels as colour information is not mandatory for deriving motion. Intensity of frames is divided by M to normalize the pixel values (in equation (2)) to a fixed range [0, 1], where M denotes the maximum pixel value of all frames.

$$f(x,y,z) = I(x,y,z)/M \tag{2}$$

The final saliency map S is calculated by weighted sum of dynamic saliency map (M^d) and static saliency map(M^s):

$$M = t * M^d + (1-t) * M^s \tag{3}$$

3.3 Background Noise Suppression

The human vision perceives its environment by combining top down knowledge and bottom up approach. Bottom-up approach directs the attention to the most salient region in the viewer's environment whereas top down approach focuses on locating a desired object in the environment. In day to day life a person sees many moving objects in his environment, the most prominent moving object grabs the viewer's attention. Since entropy based dynamic map detects all moving objects in a video, the equation (3) has an impact of the dynamic saliency map in final saliency map calculation. The saliency map (S) contains the prominent moving object along with the background noises like tree, waving flag, elevator, fan, etc. This affects further processing in video analysis, which must be suppressed and is not achieved by this approach.

The attention mechanism integrates bottom-up elements with a top-down mechanism that can guide our attention. The video sequences are dynamically learned to understand the location of the background noises in the frames. This knowledge is obtained by observing successive frames in the video. A background movement will have static location feature and thus can be detected by observing if the salient objects change position in a given number of successive frames. A number of successive frames are taken, 1...k, where k is the number of frames. The location feature of each moving object detected is learned and at k+1th frame the pixel values of all the detected objects are compared with the knowledge derived from the previous frames. If the object hasn't moved even after the k frames, the pixel values would not change and thus they can be nullified.

If S is the Saliency map obtained by bottom-up approach with t as the current frame number, then,

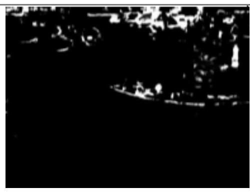
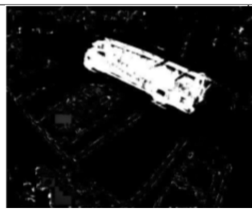
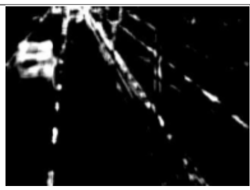
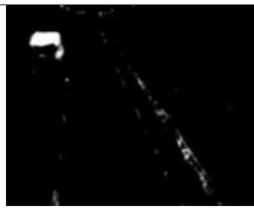
(4)

The resulting map TD(S) will then show only the prominent moving object.

3. Experimental Results

All the methods were tested on a system with configuration of operating system windows 8 platform with hardware capacity as i5 processor, 1 GB RAM, in software MATLAB version R2013a. The testing was on more than 25 short videos. Our method is evaluated on different videos taken from different environments such as shopping mall, parking lot, traffic ,.etc. Table 1. depicts the results for a frame from few surveillance videos. The first column shows the sample input frame of the surveillance video, the second column shows the number of moving objects in that video as inferred by human evaluation from different people, the next two consecutive maps show the comparison between the previous method and our method, the last column shows the number of moving objects detected after applying top down approach. In most of the cases, the proposed algorithm shows better moving object detection rate is approximately 95-100 %. Though our method gives better result, there is a reduction in detection rate which is approximately 80% in cases, such as for input fame no 5 in Table 1., where there is a digital advertisement board in the background, the screen was also detected since it showed dynamic images.

Table 1. Saliency maps and the moving objects detected before and after applying top down approach for a given frame

Sr. no.	Sample Input frame	Number of moving objects	Output of S	Output of TD(S)	Number of objects detected moving
1		1			1
2		5			5
3		1			1
4		1			1

5		10			12

4. Conclusion

Salient motion detection in surveillance videos was achieved by combining the static and the dynamic saliency maps of the frames and then optimizing it with top down knowledge obtained by dynamic learning video sequence dynamically. The mechanism was tested with many surveillance videos of different complexities. The final saliency map after using top down knowledge provided better results. The mechanism can be further exploited to boost security by detecting anomalous events so that necessary action can be taken immediately.

References

1. R.Aarthi,J.Amudha,Ushapriya, A generic bio inspired framework for detecting Humans Based on saliency detection, Springer International conference on artificial Intelligence and evolutionary algorithms in engineering Systems-2014(ICAEES-2014), 22nd and 23rd april 2014,Kuamracoil,kanyakumari. ISBN 978-81-322-2134-0 [SCOPUS]
2. Zhengzheng Tu, Aihua Zheng, Erfu Yang, Bin Luo, Amir Hussain, A Biologically Inspired Vision-Based Approach for Detecting Multiple Moving Objects in Complex Outdoor Scenes, 30 January 2015, Cognitive Computation. ISSN 1866-9964.
3. Ming-Ming Cheng, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang, Shi-Min Hu, Global Contrast based Saliency Detection. IEEE International Conference on Computer Vision and Pattern Recognition (IEEE CVPR), 2011, p. 409-416.
4. Chaun Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, Ming-Hsuan Yang, Saliency Detection via Graph based Manifold Ranking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013) , Portland, June, 2013.
5. Satyadev Sarma, Shanth, IN ini Srinivas, R. Aarthi Visual Saliency for Surveillance Images- A Brief Study, International Journal of Software and Web Sciences (IJSWS),March-may 2014 issue, vol.1(14-212),pages: 10-12,ISSN(ONLINE):2279-0071, ISSN(PRINT):2279-0063
6. R.Aarthi,J.Amudha, Saliency based modified chamfer matching method for sketch based image retrieval, DRDO Sponpored 2nd IEEE international conference on Innovations in information, embedded and communication system(ICIIECS-2015), On 19th and 20th march 2015, Coimbatore. IEEE XPLORE : ISBN:978-1-4799-6818-3 , DVD :ISBN :978-1-4799-6816-9[Submitted for SCOPUS indexing]
7. Simone Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-directed Search, in Proc of international conference on Intelligent System, January 9(2006).
8. Amudha J, Soman. K.P and Padmakar Reddy. S, A Knowledge Driven Computational Visual Attention Model, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011. ISSN (Online): 1694-0814.
9. Laptev Ivan, Belongie. J. Serge, Perez Patrick, Wills Josh, Periodic motion detection and segmentation via approximate sequence alignment. Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference Vol. 1.
10. Laurent Itti, Christof Koch, Ernst Niebur, A model of Saliency based Visual attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL 20, NO 11, NOV 1998.
11. Longsheng Wei, Nong Sang and Yuehuan Wang. A Spatio Temporal Saliency model on visual attention based on maximum entropy. *International Society for Photogrammetry and Remote Sensing*, 2011.
12. Yao Shen, ParthasarathyGuturu, ThyagarajuDamarla, Bill P. Buckles, Senior , and KameswaraRaoNamuduri, Video Stabilization Using Principal Component Analysis and Scale Invariant Feature Transform in Particle Filter Framework, IEEE Transactions on Consumer Electronics, Vol. 55, No. 3, AUGUST 2009.
13. Muthuswamy Karthik, Rajan Deepu. Salient motion detection through state controlability, Proceedings of the IEEE Workshop on Motion and Video Computing,2005.
14. P. Felzenswalb and D. HuttenLocher, Efficient Graph based Image Segmentation, IJCV, 59 (2) : 167-181 , 2004.
15. Change Loy, C., Xiang, T. and Gong, S. (2012). Salientmotion detection in crowded scenes. Communications Control and SignalProcessing, (ISCCSP), pp.1-4