Generating Novel Information Salient Maps for Foreground Object Detection in Video

Chang Liu[1], Pong C Yuen[1], Guoping Qiu[1,2]
[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong
[2]School of Computer Science, University of Nottingham, UK
cliu, pcyuen, gqiu@comp.hkbu.edu.hk

Abstract

The conceptual model of visual saliency in human vision system has been employed in extracting salient features from images and multimedia data in the last decade. This paper proposes to employ the visual saliency for moving object detection. The crucial factor is to compute a saliency map such that visual attention can be performed. This paper proposes a new method for saliency map construction based on information theory and spatio-temporal model, called information saliency map (ISM). The ISM provides rich information content of the video. Moving object detection are then performed based on the ISM. Two popular and publicly available visual surveillance databases from CAVIAR and PETS are selected for evaluation. Experimental results show that the proposed method is robust for moving object detection in complex background and illumination changes. The average detection rate is 90.35% while the false alarm rate is 2.46% in CAVIAR (INRIA entrance hall) dataset with ground truth data, and it has shown merits comparing with the current state of the art.

1 Introduction

Moving object detection is the first and important step for a computer vision system. It is also one of the most active research areas in computer vision because of the wide range of applications such as visual surveillance, human identification, human behavior recognition, event recognition and traffic congestion control. Therefore, an efficient and robust object detection algorithm under different situations such as illumination change and complex background, is required. A number of algorithms have been proposed for moving object detection methods in the last decade, many of these work are based on the background subtraction method. The rationale of this approach is to build an appropriate representation (background image model) of the scene so that the object(s) in the current frame can be detected by subtracting the current frame with the background image. Based on this idea, several adaptive background models have been proposed. Stauffer and Grimson [11] developed a method to model each pixel as a Mixture Of Gaussians (MOG) and constructed a model that can be updated on-line. Along this line, other similar methods have been developed [14] [9]. But a common problem in background subtraction is that it requires a long time for estimating the background image model. Furthermore, because MOG assumes all pixels are independent, pixel correlation is not considered, so the background model based on individual pixels is sensitive to illumination and noise. Some researchers have proposed motion detection methods based on optical flow [12] [10], these methods can accurately detect motion in the direction of intensity gradient, but the motion which is tangential to the intensity gradient can not be well represented by the feature map. Moreover, optical flow methods also have problems when there is illumination changes, because it ignores temporal changes of images. Some researchers introduced level set and active contour method for moving object detection [1] [4], but these methods are computationally too expensive for real-time applications.

The conceptual model of visual saliency has been employed in (image) scene analysis[5, 7]. The crucial factor is how to compute the saliency map. The methods[13, 2, 6, 3] determine the saliency map based on the spatial and temporal information separately. The advantage is computational simple while the saliency map may be sensitive to illumination changes and may not be suitable for robust object detection. Along this line, this paper presents a new method to construct the saliency map based on information theory and spatio-temporal model. Shannon information theory shows that uniqueness or rarely happened events contain high information while common or frequently happened events imply low information. This is in line with our HVS as well as the visual saliency model. In order to overcome the illumination sensitivity, both spatial and temporal information will be considered as a 3D volume to compute saliency map. The illumination change in the volume is a much more frequently happened events than motion, so it shows lower

978-0-7695-3119-9/08 $25.00 © 2008 IEEE
DOI 10.1109/CISP.2008.406

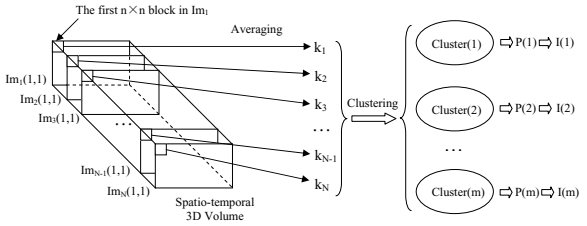


Figure 1. Computing the information saliency map of current frame

saliency, the saliency map will then be insensitive to illumination changes. In this way, moving object detection can be effectively and efficiently performed using the ISM under different circumstances.

2 Proposed Method Using Information Content

Based on the conceptual model of visual saliency, this paper proposes to employ the information theory and spatio-temporal model to compute the information saliency map (ISM) which reflects the information content on each pixel. The ISM can be computed very efficiently and moving object detection are then performed based on the ISM.

2.1 Information theory

Consider a discrete random variable $X \in 1,...,K$, suppose the event $X = k$ is observed, the Shannon's self-information content of this event $I(k)$ is defined as follows

$$I(k) = \log_2 1/p(X = k) = -\log_2 p(X = k) \qquad (1)$$

It means that the information of event k is inversely proportional to the probability of the observation of event k, a rarely happen event contains high information while an event which happen frequently contains low information.

2.2 Information saliency map

In order to compute the information saliency map, we need to calculate the information on each pixel based on Eq (1). In turn, we need to estimate the probability of each pixel. Qiu et al. [8] has proposed an information theoretic model to compute integrated spatiotemporal visual saliency features. Their model is theoretically sound, but estimation of the conditional probability of the high dimensional variables is computational expensive when the number of frames is large. In this paper, we develop a new method to generate the saliency map more efficiently and accurately.

The structure of ISM can be represented as the following:

$$ISM(t) = \begin{pmatrix} Info(1,1,t) & \dots & Info(1,w,t) \\ \vdots & \ddots & \vdots \\ Info(h,1,t) & \cdots & Info(h,w,t) \end{pmatrix} \qquad (2)$$

Where the ISM in time t can be divided into several blocks: $Info(r,s,t), r = \{1,2,...,h\}, s = \{1,2,...,w\}$. The block diagram of the proposed method in computing the ISM is shown in Figure 1. Consider a $N-frame$ (N=20 in this paper) spatio-temporal volume which consists of the current frame and the previous $N-1$ frames. Each frame Im is then divided into a number of blocks with smaller size: $\{Im(1,1), Im(1,2), ..., Im(h,w)\}$. Information saliency map of each block will be computed individually and the information content of block $Im(r,s)$ in time t will be represented by $Info(r,s,t)$, where

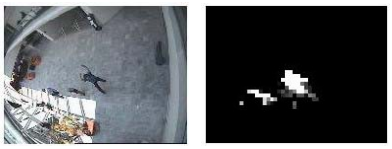
$$Info(r,s,t) = -\log[P(B(r,s,t)|V(r,s,t))] \qquad (3)$$

$B(r,s,t)$ represents block $Im(r,s)$ at time t, and $V(r,s,t)$ represents the spatio-temporal volume containing $B(r,s,t)$ as the current image block. The current block probability density function is determined by considering the DC coefficient of each block $\{k_1,k_2,...,k_N\}$ which can be calculated by the block mean value.

To compute the information I(k) from (1), the probability of variable k needs to be computed from the DC coefficients $\{k_1,k_2,...,k_n\}$. The straightforward method is to make use of histogram, but it requires a pre-defined bin (histogram) width. Since the probability is calculated on-line and the number of data (N) is 20, the fixed width histogram may introduce large error. Instead, we propose an adaptive method to construct the histogram based on the clustering technique. By clustering the DC coefficients into different clusters, each cluster can be consider as a bin with adaptive bin wide. To do so, k-mean is one of the possible choices. However, k-mean is an iterative algorithm. It would be computational expensive because we need to calculate around 200 blocks for each frame. Therefore, we propose a simple but effective method as shown in Algorithm 1. Suppose K=$\{k_1,k_2,...,k_N\}$ is a sorted DC coefficients. Two consecutive coefficients will belong to the same cluster if $(|k_i|-|k_{i+1}|)|\ /\ (|k_i|+|k_{i+1}|) < \alpha$, ($\alpha$ = 0.05 in this paper) where $i = 1,2...N-1$. The probability of each cluster is then computed by dividing the number of coefficients in each cluster by N (total number of coefficients). Then the information content of each block is calculated using (1). The information saliency map is then generated for the current frame. Figure 2(a) shows the frame 70^{th} of the video Browse1 from CAVIAR dataset. The video Browse1 consists of two people browsing around the entrance hall. Figure 2(b) shows the corresponding ISM. It can be seen that the locations of the two people are clearly indicated in the ISM.

```
Given sorted K={k_1, k_2, ..., k_N},
j=1, cluster(j)=1;
for i=1 to N-1
  if dist(k_i, k_{i+1}) < α (|k_i| + |k_{i+1}|)
    cluster(j) = cluster(j)+1;
  else
    P(j) = Number of coefficients in cluster(j)/N;
  j = j+1, cluster(j) = 1;
```

Algorithm 1. PDF computation



(a) Original image (b) ISM

Figure 2. Information saliency map for the 70^{th} frame in video Browse1 in CAVIAR database

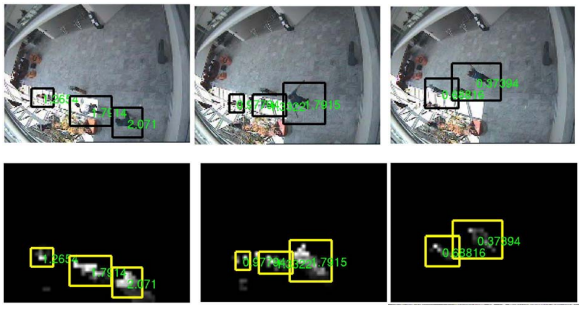


Figure 3. ISM and object detection results in Frame 20, 50, 70 from video "*Browse*1" CAVIAR database

2.3 Moving object detection using ISM

The larger the value in the ISM, the larger the information at the corresponding position. Basically, moving object detection based on the ISM can be performed by defining a threshold value. A typical result is shown in Figure 3 where the upper row shows the original video Browse1 at frame 20, 50 and 70 while the lower row shows the corresponding ISM and detection results.

However, this straightforward detection method suffers from two limitations. The first limitation is that there may have false detections because of the present of noise. In order to avoid the noise effects in the block information saliency image, we apply an averaging filter to the ISM. The rationale is that if a block within a moving object has a relatively high information saliency value, its neighbor blocks should also have high probability to contain high information. Otherwise, it must be noise. This simple process can remove noise effectively. However, the drawback is that an object with similar size may also be treated as noise.

The second limitation is that the human/object will be lost detected if he/she changes from moving to stationary. In such a case, the information content will decrease to a smaller value. This can be illustrated using the example in which an object is changed from motion to stationary and starting moving again. The object motion and the corresponding ISM value are recorded and shown in Figure 4. It can be seen that the video can be divided into three video segments, namely motion, stationary and motion. In the stationary part, the saliency value is below threshold and is closed to zero. If we only based on the saliency value, the object will be lost detected. This drawback can be solved by monitoring the falling edge of the saliency curve. When-

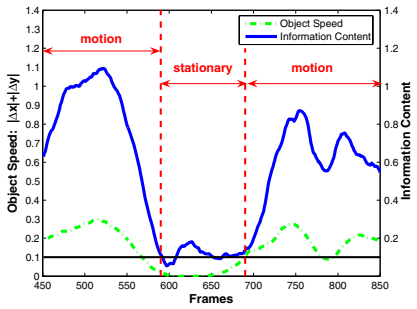


Figure 4. Relation of object speed and object information content, the horizontal black solid line represents a predefined information content threshold to identify if an object is static, it is set to be 0.1 here

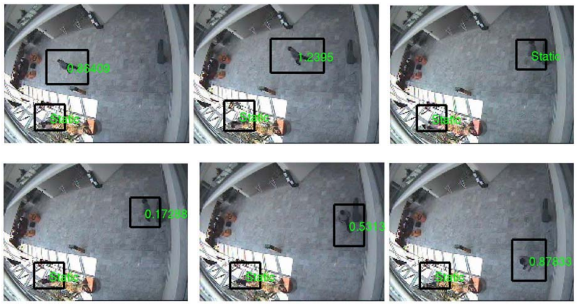


Figure 5. Moving object detection results in Frame 500, 550, 600, 700, 750, 800 from CAVIAR "*Browse*1" database, the value in the rectangle correspond to the object information content or object status

ever, there is a falling edge, the object of interest becomes stationary and the area is also classified as a region of interest. Whenever, we find that there is a rising edge of the saliency curve, that object is in motion again. The video sequence from this curve is show in Figure 5.

3 Experimental Results

The experimental results in this section is divided into two parts. First, two popular and publicly available surveillance datasets, namely CAVIAR[16] and PETS[15],are used to evaluate the performance of the proposed method. Second, the proposed method is compared with existing methods. A widely used moving object detection method namely Mixture Of Gaussians (MOG) [14], are selected for comparison.

3.1 Evaluation of the proposed method

CAVIAR database[16] consists of 3 datasets, namely INRIA entrance hall, shopping mall frontal and shopping mall side. Only INRIA entrance hall dataset is selected to evaluate our method because only this dataset has the ground truth data. The INRIA entrance hall dataset has six types of events, namely "*Browsing*", "*Fighting*", "*Groups_meeting*", "*Leaving_bags*", "*Rest*"

Figure 6. Moving object detection results in Frame 220, 270, 350 from CAVIAR "*Fight_OneManDown*" database

and "*Walking*", totally 28 video sequences. These video sequences are captured from inclined look-down camera with a wide angle. People appear in front of the camera have different sizes and body figures, which make it difficult to detect object. Moreover, the bottom left part region of the video sequences is under severe illumination condition.

The detection rate and the false detection rate of each video sequence are recorded and tabulated in Table 1. The average detection rate is 90.35% while the false detection rate is 2.46%. The results are encouraging. In particular, we would like to point out that our proposed method is able to detect moving objects under both illumination changes and small motion which can be demonstrated using the video "*Fight_OneManDown*". The experimental results are shown in Figure 6 where 3 key frames are selected to show the detection process. It can be seen that the woman standing at the left side of the image is not a visual salient region until she moves and its information content value exceed a certain value at the 270 th frame. She is kept detected under our attention even she becomes stationary again.

Database	**TP**	**FP**	**TG**	**FAR**	**DR**
Browse	6722	204	7298	2.9%	92.1%
Fight	5060	165	5625	3.2%	90.0%
Meet	7327	165	7815	2.2%	93.8%
LeftBag	7220	140	8702	1.9%	83.0%
Rest	4768	113	5322	2.3%	89.6%
Walk	5277	103	5512	1.9%	95.7%
Average	**1299**	**32**	**1438**	**2.46%**	**90.35%**

Table 1. Moving object detection results in CAVIAR database, 28 video sequences totally, TP:True Positive, FP:False Positive, TG:Total Ground truth, FAR:False Alarm Rate, FAR=FP/(TP+FP), DR:Detection Rate, DR=TP/TG

The PETS2001[15] database consists of 20 video sequences. All video were captured at outdoor environment, various objects including humans, bicycles and vehicles. Moreover, there are global illumination changes during a short period of time. Since we would like to test the video sequence with large illumination changes, 4 out of 20 video sequences are selected to evaluate our proposed method. Six frames of one of the serve change of illumination video are shown in Figure 7. In this video, the background is under a 20-second global illumination changing process (due to sunlight occluded by cloud). The result shows that our

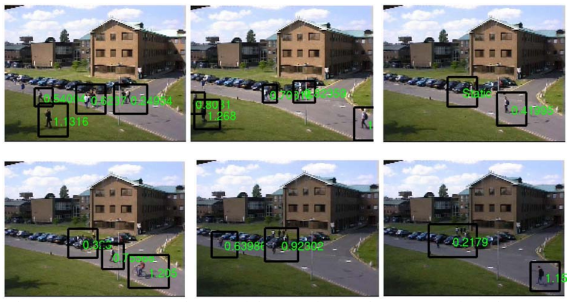


Figure 7. Moving object detection results in Frame 2720, 2800, 2880, 2930, 3080, 3150 from PETS2001 "*Dataset3_Testing_1*" database

method is robust to global illumination changes. Since no ground truth data is available in PETS2001 database, no statistic on detection rate nor false detection rate is reported.

3.2 Comparing the proposed method with existing methods

The objective of this section is to compare the proposed method with Mixture Of Gaussian (MOG) [14] using the INRIA entrance hall dataset in CAVIAR.

The improved adaptive MOG constructs a probability density function for each pixel independently and pixel-level background subtraction is performed to find the region of interest. Experimental results show that MOG is able to model the background very well, even for serve illumination changes. However, MOG suffers from a drawback. When an object is moving slowly or with relatively small motion, MOG may mis-classify that region(s) as background and update the background accordingly. As a result, the object will then be missed. This situation can be illustrated using the example in Figure 9. A human at the bottom left region has a slow motion and stationary at certain period of time in the video. The detection results using MOG are shown in the third row in Figure 9. It can be seen that MOG is not able to locate that region of interest and considers as illumination noise. As a comparison, the results of the ISM generated using our method are shown in the last row in Figure 9. Our method is able to locate the relatively small motion object most of the time. In order to give a quantitative comparison, the ROC curves for the MOG method and the proposed method are plotted in Figure 8. It can be seen that the proposed method outperforms the MOG.

4 Conclusions and Future Works

A novel moving object detection method based on information theory and spatio-temporal model has been developed and reported in this paper. The proposed method determines an information saliency map (ISM) which shows the information content of each pixel. The ISM not only provides the saliency of each pixel for moving object detection, but also gives additional higher level object information such as the object motion speed. Two publicly avail-

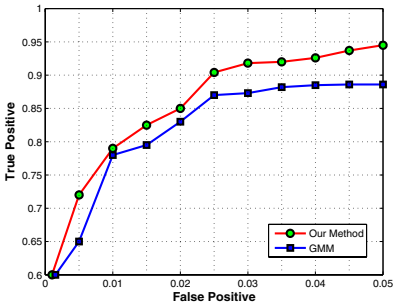


Figure 8. ROC curves for MOG[14] and our proposed method in CAVIAR INRIA entrance hall dataset

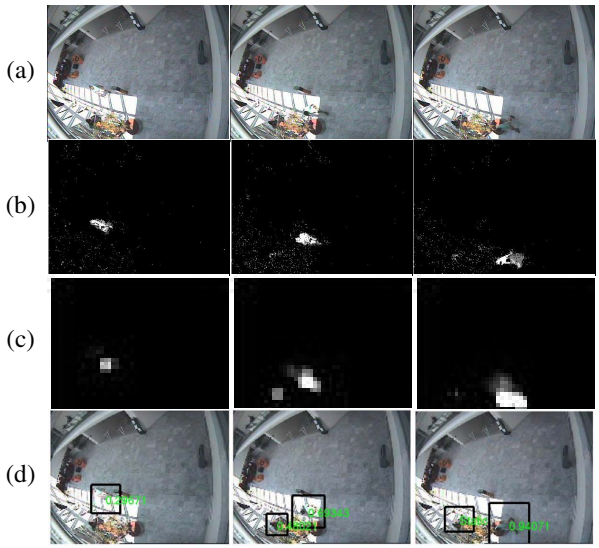


Figure 9. (a) Original frames, (b) Results on MOG [14], (c) Our proposed ISM, and (d) Moving object detection results in Frame 1005, 1030, 1050 from CAVIAR "*Browse*4" database.

able databases have been selected to evaluate the proposed method and the results are encouraging. The detection rate and false detection rate on CAVIAR INRIA entrance hall dataset are 90.35% and 2.46% respectively. Experimental results show that the proposed method is robust to illumination changes. Although this paper has successfully demonstrated the feasibility of using visual saliency for moving object detection, the computational time (on P4 3GHz personal computer using Matlab implementation) is around 2.5 fps without optimization process. Our next step is to develop a fast algorithm to compute the ISM.

Acknowledgment

This project is partially supported by the Faculty Research Grant of Hong Kong Baptist University. The authors would like to thank the EC Funded CAVIAR project/IST 2001 37540 for the contribution of the CAVIAR dataset, and thank the IEEE International Workshops on Performance Evaluation of Tracking and Surveillance for the contribution of the PETS2001 dataset.

References

[1] T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. *European Conference on Computer Vision*, pages 471–483, 2006.

[2] N. D. Bruce. Features that draw visual attention: an information theoretic perspective. *Neurocomputing*, 65-66:125–133, 2005.

[3] N. V. Dashan Gao. Integrated learning of saliency, complex features, and object detectors from cluttered scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, pages 282–287, 2005.

[4] D. L. Guidry and A. A. Farag. Using active contours and fourier descriptors for motion tracking with applications in mri. *International Conference on Image Processing*, 2:177–181, 1999.

[5] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on PAMI*, 20(11):1254–1259, 1998.

[6] V. S. James W. Davis. Fusion-based background-subtraction using contour saliency. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005.

[7] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal on Computer Vision*, 45(2):83–105, 2001.

[8] G. Qiu, X. Gu, Z. Chen, Q. Chen, and C. Wang. An information theoretic model of spatiotemporal visual saliency, to appear, international conference on multimedia and expo. 2007.

[9] A. Shimada, D. Arita, and R. Taniguchi. Dynamic control of adaptive mixture-of-gaussians background model. *International Conference on Video and Signal Based Surveillance*, Nov 2006.

[10] S. P. N. Singh, P. J. Csonka, and K. J. Waldron. Optical flow aided motion estimation for legged locomotion. *IEEE International Conference on Intelligent Robots and Systems*, pages 1738–1743, 2006.

[11] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[12] A. A. Stocker. An improved 2d optical flow sensor for motion segmentation. *Proceedings of IEEE International Symposium on Circuits and Systems*, 2:332–335, 2002.

[13] J. van de Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *IEEE Transactions on PAMI*, 28(1):150–156, 2006.

[14] Z. Zivkovic. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2006.

[15] http://ftp.pets.rdg.ac.uk/pets2001/.

[16] http://homepages.inf.ed.ac.uk/rbf/caviar/.