SALIENT MOTION DETECTION IN CROWDED SCENES

Chen Change Loy, Tao Xiang, Shaogang Gong

School of EECS, Queen Mary University of London

ABSTRACT

To reduce cognitive overload in CCTV monitoring, it is critical to have an automated way to focus the attention of operators on interesting events taking place in crowded public scenes. We present a global motion saliency detection method based on spectral analysis, which aims to discover and localise interesting regions, of which the flows are salient in relation to the dominant crowd flows. The method is fast and does not rely on prior knowledge specific to a scene and any training videos. We demonstrate its potential on public scene videos, with applications in salient action detection, counter flow detection, and unstable crowd flow detection.

1. INTRODUCTION

The attention of CCTV operators tends to deteriorate after prolonged surveillance video monitoring, especially when they are overwhelmed with unpredictable and complex motion patterns observed in a crowded scene [1]. To increase surveillance effectiveness and reduce operator cognitive overload, it is critical to filter the displayed input streams and focus operator attention on interesting/salient events.

Finding interesting events in a scene is typically achieved through (1) training an activity model, e.g. topic models [2] or probabilistic graphical models [3] on a large amount of video data, and (2) using the learned model to detect/classify interesting actions or events on unseen video clips. From the point of view of human vision processing [4], this standard pipeline can be seen as an *attention process* of how human perceives the world, which is usually slow, complex, and specific. On the other hand, saliency detection [5] that extracts unique regions from an unknown background is regarded as a *pre-attentive process*, which is fast, simple, and general. We believe motion saliency detection, which has been largely ignored in video surveillance literature, can be a compelling way of suppressing background clutter in a video to achieve automatic focus of attention.

Most existing saliency detection methods are devoted to finding fixation points and dominant objects in static images [5, 6, 7]. These method are useful for static image understanding but not suitable for dynamic motion interpretation in video (see Fig. 1). Methods for motion saliency detection do exist [8, 9, 10, 11] but most of them either focus on *local* notion of saliency (e.g. local space time interest points, image contrast, or colour) or they are computationally infeasible for surveillance applications [8].

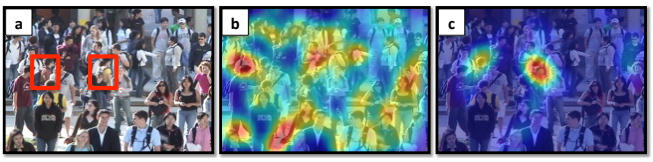


Fig. 1. (a) Two individuals walking in a different direction against the crowd flow. (b) Conventional saliency detection methods only detect object with unique appearance. (c) Global motion saliency detection detects motion patterns that are salient in relation to the dominant crowd flows.

In this paper, we propose a simple yet effective motion saliency detection method by adapting existing static image saliency detection approach [6]. The proposed method is *global* – it aims to suppress dominant crowd flows, while focusing human attention on motion flows that deviate from the norm. It is fast and independent of any forms of prior knowledge of a scene and training videos, so it is scalable and able to generalise well on different surveillance scenarios. As opposed to conventional saliency detection methods that model the appearance or motion properties of a target object, our method makes use of the *background properties*. Specifically, we summarise the properties of motion flow field in a scene using a log spectrum representation – a popular way of modelling image statistics [12]. A smoothed version of this log spectrum approximates the background or redundant motion components. Removing the redundant components from the original log spectrum highlights the statistical singularities, which inherently correspond to the salient motion patterns.

The proposed method is inspired by the popular spectral residual approach [6]. However, unlike [6] that performs static image saliency detection, we have a different goal on discovering salient motion region in video. It is worth pointing out that several studies have attempted to extend spectral residual approach, but for very different purposes than ours, such as background subtraction [13] and dominant object detection in non-crowded scenes [14]. In this paper, we show for the first time how to detect global anomalous motion flows in a video by analysing spectral singularities in the motion space.

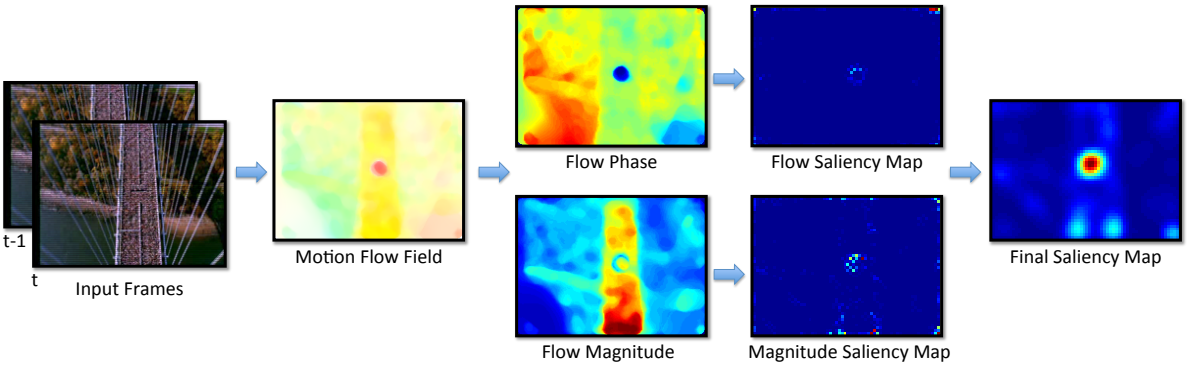


Fig. 2. Key steps in global motion saliency detection.

2. MOTION SALIENCY DETECTION

Figure 2 illustrates the key steps in our approach. The inputs are the current and previous frames. The output is a saliency map that suggests the regions of salient motion.

Our method commences with optical flow field estimation, which can be achieved using various strategies [15, 16]; here we adopted a method proposed by Liu [17]. Let the image intensity at pixel location (x, y) at time t be denoted by $I_{x,y,t}$. We represent the associated flow vector by its phase angle $-\pi \leq \varphi_{x,y,t} \leq \pi$ and the velocity magnitude $r_{x,y,t} \geq 0$

$$\varphi_{x,y,t} = \arctan\left(\frac{V_y}{V_x}\right) \quad (1)$$

$$r_{x,y,t} = \sqrt{V_x^2 + V_y^2}, \quad (2)$$

where V_x and V_y are the horizontal and vertical components of the optical flow vector of $I_{x,y,t}$. An example of the estimated flow field is depicted in Fig. 2.

At each time t we obtain $\boldsymbol{\varphi}_t$ and $\mathbf{r}_t$ that characterises the direction and motion velocity in that direction. Both $\boldsymbol{\varphi}_t$ and $\mathbf{r}_t$ have a resolution of $x \times y$. We refer them as *motion signatures*, from which we aim to detect for salient motion regions. The salient regions can be intuitively interpreted as region with irregular motion direction and velocity magnitude.

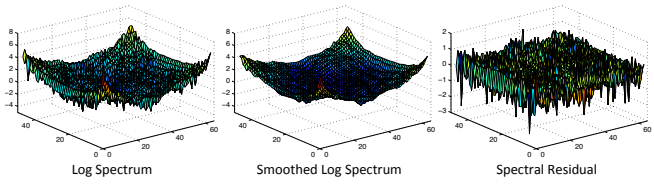


Fig. 3. From left to right: Log spectrum $\mathcal{L}_\varphi(f)$, smooth log spectrum $\overline{\mathcal{L}}_\varphi(f)$, and spectral residual $\mathcal{R}_\varphi(f)$.

In the following we explain the steps to obtain the saliency map correspond to $\boldsymbol{\varphi}_t$. Note that the following steps are similar to those of the spectral residual approach [6], but adapted for working with motion signature, allowing motion saliency detection that is not possible using the original method. First, we down-sample the motion signature $\boldsymbol{\varphi}_t$ to a lower resolution with width equals to 64 pixels. We then perform Fourier Transform $\mathfrak{F}$ on the down-sampled signature to obtain the real and imaginary components of the spectrum

$$\mathcal{A}_\varphi(f) = \Re(\mathfrak{F}[\boldsymbol{\varphi}_t]) \quad (3)$$

$$\mathcal{P}_\varphi(f) = \Im(\mathfrak{F}[\boldsymbol{\varphi}_t]). \quad (4)$$

We then compute the log spectrum (see Fig. 3) as

$$\mathcal{L}_\varphi(f) = \log\left(\mathcal{A}_\varphi(f)\right). \quad (5)$$

Next, we apply a local average filter $h_n(f)$ onto $\mathcal{L}_\varphi(f)$ to obtain an averaged log spectrum that approximates the background or redundant motion components

$$\overline{\mathcal{L}}_\varphi(f) = h_n(f) * \mathcal{L}(f), \quad (6)$$

where $h_n(f)$ is an $n \times n$ matrix ($n = 3$ is used in our study)

$$h_n(f) = \frac{1}{n^2}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (7)$$

This is followed by the computation of spectral residual $\mathcal{R}_\varphi(f)$ between the original log spectrum and a smoothed version

$$\mathcal{R}_\varphi(f) = \mathcal{L}_\varphi(f) - \overline{\mathcal{L}}_\varphi(f), \quad (8)$$

which contains the compressed information of the salient region of the motion direction $\boldsymbol{\varphi}_t$.

Finally, we transform the spectral residual back to the spatial domain to obtain the saliency map through Inverse Fourier Transform $\mathcal{F}^{-1}$

$$\mathcal{S}_\varphi(f) = g(x) * \mathcal{F}^{-1}\left[\exp(\mathcal{R}_\varphi(f) + \mathcal{P}_\varphi(f))\right], \quad (9)$$

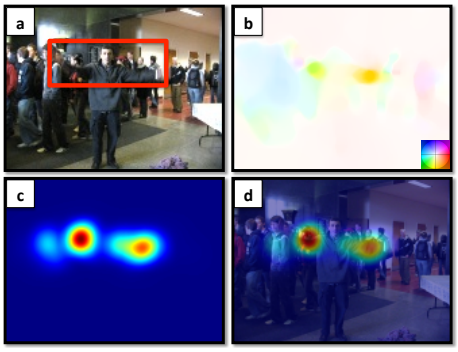
where $g(x)$ is a Gaussian filter to smooth the saliency map of better visualisation output.

The same process is applied to $\mathbf{r}_t$ to obtain $\mathcal{S}_r(f)$. We weigh $\mathcal{S}_\varphi(f)$ with $\mathcal{S}_r(f)$ to give more emphasis on phase changes caused by high velocity. To generate the final saliency map $\mathcal{S}(f)$, we combine both saliency maps, i.e. the weighted $\mathcal{S}_\varphi(f)$ and $\mathcal{S}_r(f)$ by taking the maximum values between the two.

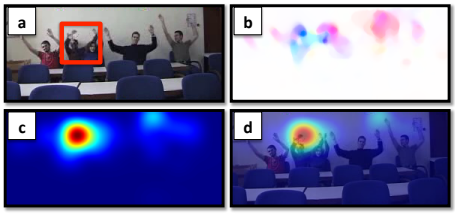
3. RESULTS

In this section, we demonstrate the effectiveness of the method on various video clips including public surveillance footage, showing its capability in salient action detection, counter flow detection, and crowd instability detection.

Salient Action Detection: Action detection in crowd is known to be a non-trivial problem [18], especially when the background consists of random and irrelevant activities. Isolating the action of interest often requires exhaustive sliding windows-based search using a classifier trained on a specific action. The search space can be drastically reduced by using the proposed saliency detector to first locate potential regions with salient motion. Example I is shown in Fig. 4: owing to the irregular motion phase and velocity of the jumping action compared to the rest of the scene, it was picked up by our saliency detector despite the large moving crowd behind the person. Example II in Fig. 4 shows a person (highlighted in red boxes) waving his hands in a opposite direction compared to others. This salient action was again detected by our method. Note that all detections were achieved without training data and any forms of prior knowledge.



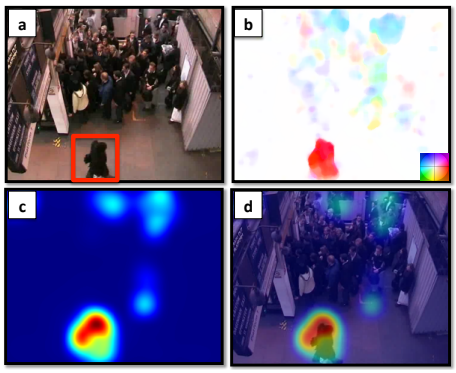
(a) Example I



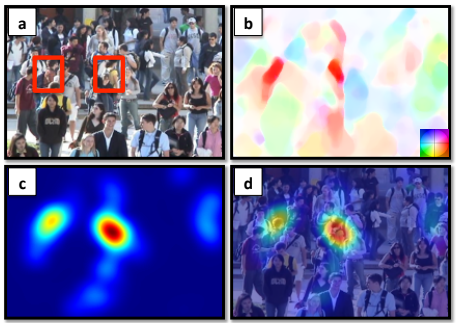
(b) Example II (video from [8])

Fig. 4. Salient action detection: (a) input frame, (b) optical flow field, (c) motion saliency map, and (d) saliency map overlaid on the input frame.

Counter Flow Detection: In many public spaces such as an underground station, one would like to detect people moving against the dominant crowd flow, which may imply a possible security threat. The method presented here is able to discover this type of behaviour on-the-fly. For instance, in Fig. 5-Example I and II, our method detected individuals walking in different direction against the dominant crowd flow.



(a) Example I

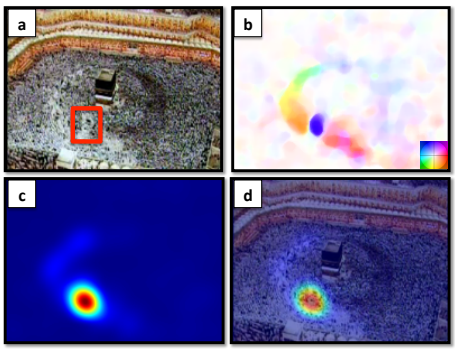


(b) Example II

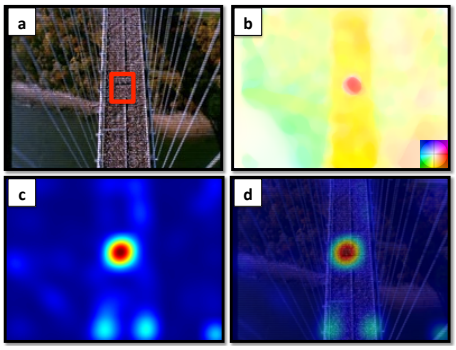
Fig. 5. Counter flow detection: (a) input frames, (b) optical flow field, (c) motion saliency map, and (d) saliency map overlaid on the input frame.

Crowd Instability Detection: We demonstrate how the motion saliency method can be employed to detect flow instability in extremely crowded scenes, such as the marathon scene and pilgrimage scene depicted in Fig. 6.

For evaluation, we employed the marathon and pilgrimage video sequences used in [19]. In these sequences, synthetic instabilities were inserted into the original videos by flipping and rotating the flow of a random location. As can be seen from Fig. 6, the proposed method was able to locate and flag the area with potential unusual flow patterns. It is worth pointing out that similar results were obtained by using the state-of-the-art method [19]. Nevertheless, our solution achieved the same goal but with a significant simplification of the formulation, therefore having a much reduced computational cost, which is critical and necessary for real-time video processing.



(a) Example I



(b) Example II

Fig. 6. Crowd instability detection: (a) input frames , (b) optical flow field, (c) motion saliency map, and (d) saliency map overlaid on the input frame.

4. CONCLUSION

We showed that spectral analysis of image spectrum [6] can be adapted to analysing motion spectrum for global motion saliency detection in crowded scenes. The presented method is computationally effective and independent of any forms of prior knowledge of a scene. It is therefore suitable for various surveillance scenarios that demand real-time focus-of-attention. We have shown its potential in various applications such as salient action detection, counter flow detection, and unstable region detection in extremely crowded scenes. The current solution assumes the existence of dominant motion between consecutive frames and captures only the short-term background motion properties. It would be interesting to extend it for summarising long-term background statistics for more robust saliency detection.

5. REFERENCES

[1] Mary W. Green, "The appropriate and effective use of security technologies in U.S. schools.," Tech. Rep. NCJ 178265, Sandia National Laboratories, 1999. 1

[2] T. Hospedales, J. Li, S. Gong, and T. Xiang, "Identifying rare and subtle behaviours: A weakly supervised joint topic model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 99, 2011, Early Access. 1

[3] C. C. Loy, T. Xiang, and S. Gong, "Incremental activity modelling in multiple disjoint cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011. 1

[4] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980. 1

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998. 1

[6] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference Computer Vision and Pattern Recognition*, 2007, pp. 1–8. 1, 2, 4

[7] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conference Computer Vision and Pattern Recognition*, 2010, pp. 2376–2383. 1

[8] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 17–31, 2007. 1, 3

[9] D. Russell and S. Gong, "Segmenting highly textured nonstationary background," in *British Machine Vision Conference*, 2007. 1

[10] David Russell and Shaogang Gong, "Multi-layered decomposition of recurrent scene," in *European Conference on Computer Vision*, 2008, pp. 574–587. 1

[11] H.S.W. Hung, *From visual saliency to video behaviour understanding*, Ph.D. thesis, University of London, 2007. 1

[12] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: Computation in Neural Systems*, vol. 14, no. 3, pp. 391–412, 2003. 1

[13] X. Cui, Q. Liu, and D. Metaxas, "Temporal spectral residual: fast motion saliency detection," in *ACM International Conference on Multimedia*, 2009. 1

[14] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 1

[15] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Understanding Workshop*, 1981, pp. 121–130. 2

[16] B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981. 2

[17] Ce Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, Massachusetts Institute of Technology, 2009. 2

[18] Parthipan Siva and Tao Xiang, "Action detection in crowd," in *British Machine Vision Conference*, 2010. 3

[19] S. Ali and M. Shah, "A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6. 3