Detection of salient regions in crowded scenes

M.K. Lim, C.S. Chan, D. Monekosso and P. Remagnino

The increasing number of cameras and a handful of human operators to monitor the video inputs from hundreds of cameras leave the system ill equipped to fulfil the task of detecting anomalies. Thus, there is a dire need to automatically detect the regions that require immediate attention for more effective and proactive surveillance. A framework that utilises the temporal variations in the flow field of a crowd scene to automatically detect salient regions is proposed, while eliminating the need to have prior knowledge of the scene or training. The flow fields are deemed to be a dynamic system and adopt the stability theory of dynamic systems, to determine the motion dynamics within a given area. In this context, the salient regions refer to the areas with high motion dynamics, where the points in a particular region are unstable. The experimental results on public, crowd scenes have shown the effectiveness of the proposed method in detecting salient regions which correspond to an unstable flow, occlusions, bottlenecks, and entries and exits.

Introduction: Conventional CCTV monitoring by human operators becomes increasingly demanding as the average number of the cameras deployed grows. The research findings have shown that besides fatigue and boredom, human attention tends to decline after 20 min. Therefore, a high percentage of questionable activities are often overlooked. This is made even more challenging when monitoring crowded scenes such as the footage of a pilgrimage shown in Fig. 1*a*. Anomalous activity or behaviour in a crowded scene can be very subtle and imperceptible to a human operator [1]. Thus, an automated detection of the suspicious regions is critical to direct the attention of the security personnel to the areas that require further investigation. It is useful in numerous applications, such as identifying bottlenecks, which may help in avoiding congestion or evacuation planning.

Fig. 1 *Sample shots of the different scenarios of crowded scenes*

a Pilgrimage
b Train station and
c Marathon

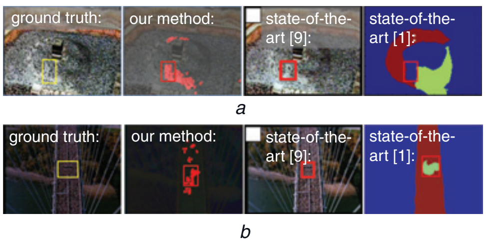

Fig. 2 *Sample comparison results for pilgrimage and marathon sequences (with synthetic noise)*

a In addition to ground truth unstable region (as enclosed in red and yellow bounding boxes), our method detected salient regions caused by bottlenecks (as highlighted in red blobs)
b Our method detects salient regions that may be caused by sudden slow down or potential danger due to high densities and instability

Most work in saliency detection is focused on the detection of the salient regions in an image, where the saliency originates from a visual uniqueness and is often deciphered from the image attributes such as colour, gradient and edges [2]. The saliency in an image differs from the saliency in a video sequence and using the image attributes alone is not sufficient to infer the motion dynamics of crowded scenes. Boiman and Irani in [3] proposed a graphical inference algorithm to detect irregularities in videos. Although their method works well in detecting the irregularities in both the images and the videos, it is not utilising the benefit of the motion information from videos and does not cope with large-scale crowd scenes.

Research into motion dynamics in dense crowds [4–6] is limited to learning the coherent motion patterns or dominant crowd flows, where the regions with a similar motion information are grouped into the same cluster. In contrast, our method ignores the dominant flows and instead is focused on the regions with high motion dynamics or hat are unstable, to infer the salient regions. The closest work to ours, thus far, is by Loy *et al.* in [7] where dominant flows are suppressed, while focusing on the motion flows that deviate from the norm. However, their method which is based on spectral analysis of the motion flows is only reliable when detecting obvious saliency such as crowd instability and counter flow detection. They do not deal with the subtler scenarios of saliency such as bottlenecks. In [8], a set of rules is applied to the eigenvalue map to discover the different motion behaviours. Although their method is able to discriminate the different types of saliency, it is restricted by the pre-defined conditions and requires the characteristic flows. Our method on the other hand is not restricted by the set of the rules, and assumes an anomaly when a particular region exhibits high motion dynamics.

This Letter extends the definition of the salient regions to include a subtle anomaly which corresponds to the bottlenecks and the occlusions. In addition, we introduce simple, yet effective idea of amplifying regions with an unstable motion instead of disregarding them as noise. This alludes to the social behaviour of humans in crowds. In a dense crowded scene, the motions of individuals tend to follow the dominant flow of a particular region due to the physical constraints of the environment (i.e. path and junction) and the social conventions of the crowd dynamics. We can therefore consider the possibility of irregularities or anomalies occurring, when the motion dynamics of individuals differs from its close neighbours. In our Letter, we first magnify and then examine the unstable regions by performing a two stage segmentation process to infer the salient regions. Our method does not rely on tracking each object or on prior learning, thus it can adapt to the environment over time and is more practical for real-time applications.

Magnification of the unstable flow: We estimate the velocity field at each point, $V(p) = (u, v)$ by employing the dense optical flow algorithm in [9], and estimate the average velocity within an interval of τ frames

$$\overline{V} = \{\overline{u}, \overline{v}\} = \left\{ \frac{1}{\tau} \sum_{t}^{t+\tau} u_{\mathrm{p}}, \ \frac{1}{\tau} \sum_{t}^{t+\tau} v_{\mathrm{p}} \right\} \tag{1}$$

Although the mean velocity field may be a good indicator of the global flow of the individuals in a crowd, it is unstructured and may change over time. A particle advection process is implemented to keep track of the velocity changes for each point, p along its velocity field, (u, v)

$$\frac{\mathrm{d}\vec{x}_{\mathrm{p}}}{\mathrm{d}t} = u_{\mathrm{p}}(x_{\mathrm{p}}, \ t; \ t_0, \ x_0) \tag{2}$$

$$\frac{\mathrm{d}\vec{y}_{\mathrm{p}}}{\mathrm{d}t} = v_{\mathrm{p}}(y_{\mathrm{p}}, \ t; \ t_0, \ x_0) \tag{3}$$

subject to

$$p = p_0 \quad \text{at} \quad t = t_0 \tag{4}$$

The suffix p indicates the motion of a particular particle or point, p. Assuming that the initial position of p_0 is the mean velocity fields, $(\overline{u}, \overline{v})$, we deem the dynamic system as an initial value problem. Thus, the pathlines which trace the points from their x_0 and y_0 positions at time, t_0 to their positions, x_t and y_t at time, t can be solved by using the fourth-order Runge–Kutta scheme (RK4) in [10]. We adopted the Jacobian method in [11], to measure the separation between the particle's pathlines which are seeded spatially close to a point, p, within a time instance, τ. The Jacobian of the flow map is computed by the partial derivatives of d$\boldsymbol{x}$ and d$\boldsymbol{y}$, where

$$\nabla F_t(p) = \begin{bmatrix} \dfrac{\partial\, \mathrm{d}\vec{x}}{\partial x} & \dfrac{\partial\, \mathrm{d}\vec{x}}{\partial y} \\ \dfrac{\partial\, \mathrm{d}\vec{y}}{\partial x} & \dfrac{\partial\, \mathrm{d}\vec{y}}{\partial y} \end{bmatrix} \tag{5}$$

According to the theory of linear stability analysis, the square root of the largest eigenvalue, $\lambda_t(p)$ of $F_t(p)^{\mathrm{T}} F_t(p)$ indicates the maximum displacement, if the particle's seeding location is shifted by one unit as it satisfies the condition that $\ln\lambda_t(p) > 0$. In the context of this Letter, a large eigenvalue indicates that the query point is unstable, and vice versa for a small eigenvalue. Note that in contrast to the existing solutions, where the high motion dynamics are regarded as noise and thus removed, our method exploits these unstable regions. We estimate the dynamics of a point within its spatially close neighbouring points by its stability by using equation

$$\phi_t = \frac{1}{|\tau|} \log \sqrt{\lambda_t(p)} \tag{6}$$

We propose two stages of segmentation that combine the outputs of fine and the coarse segmentation obtained from the local and the global flow segmentation steps, followed by a flow magnification of the regions with a high motion instability to synthesise the signal, where β is the magnification factor and α is the segmentation threshold.

$$\hat{\phi}_t = \begin{cases} \beta\, \phi_t, & \text{if } \phi_t \geq \alpha \\ (1-\beta).\, \phi_t, & \text{otherwise} \end{cases} \tag{7}$$

Experiment: instability detection: A set of four test sequences which comprises of large-scale crowd scenes was used for evaluation. The first sequence is obtained from the National Geographic documentary, 'Inside Mecca', whereas the second depicts a marathon scene. Synthetic noise was injected into both the scenes to simulate instability in the motion of the crowd. A comparison between our Letter, Loy *et al.* [7] and Ali and Shah [4], is performed. It is observed that all the three methods are able to detect instability successfully as indicated by the red bounding boxes in Fig. 2. However, our method identified additional regions as salient. After a thorough investigation of the original sequence by three operators, we noted that these regions correspond to areas where there strong interactions and motion dynamics within the crowd. It is worth noting that a manual annotation of the ground truth salient region due to bottlenecks or turbulence is an open issue because these types of salient regions are considered subjective. In the pilgrimage sequence, we noted that the additional salient regions detected by our method in fact do correspond to the regions where there are strong interactions and motion dynamics. Owing to the structure of the scene, or the physical constraints of the Kaaba which is situated at the centre of the scene, the crowd tended to slow down their pace during the turning. In addition, the salient region detected near the synthetic instability is caused by high motion dynamics near the entry and exit points. Thus, we argue that it is unfair to deem these detections as false positive. Instead, we presuppose that the detected regions can aid us in investigating and understanding the non-obvious motion dynamics of a scene.

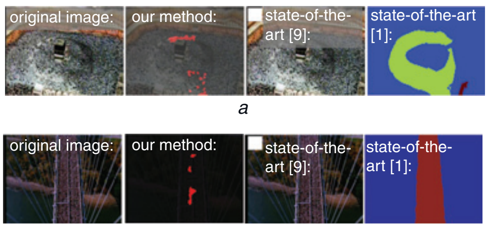

Fig. 3 *Sample comparison results (without synthetic noise)*

a Subtle saliency due to bottlenecks are detected by our method, whereas state-of-the-art methods fail to detect these variations of saliency
b Subtle saliency due to high densities and stop-and-go waves state-of-the-art methods fail to detect such saliency

Bottleneck detection: We further validated the capability of our method in detecting subtle saliency by using the original sequences, where no synthetic instability is introduced, as shown in Fig. 3. The detections of the bottlenecks have tremendous potential as an indication of impending danger such as a stampede or overcrowding taking place, due to the stop-and-go waves or sudden build up in the crowd motion.

Occlusion and turbulence detection: We further tested the robustness of the proposed method by using other scenarios of large-scale crowds; the school of fish and the marathon sequence (where there is a lamp post obstructing the flow); the results are shown in Fig. 4.

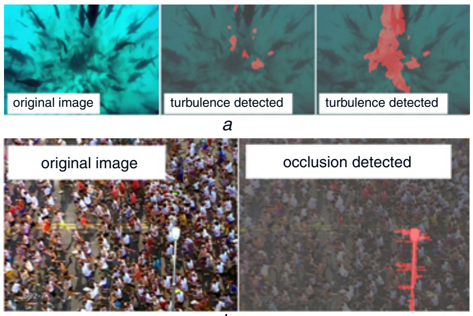

Fig. 4 *Qualitative results on other scenarios of saliency by our method*

a Detected regions grow across frames as motion dynamics of school increases. This sequence comprises of school of fish manoeuvring towards centre of scene
b Street light which simulates scenario of occlusion or barrier, is detected

Conclusion: We have proposed a framework that detects the salient regions by observing the flow activities in a given scene with minimal observations. In addition, the proposed method eliminates the need to track each object individually or prior learning of the scene, which is critical for a real-time operation. The experimental results show that the proposed method is not only able to detect the salient regions that correspond to a clear instability, but also the bottlenecks and the occlusions which is often difficult to be noticed by the naked eye. The promising results obtained are definitely worthy of future investigation since it is able to detect the regions that would otherwise go unnoted by the human operator. The capability of the proposed method in spotting the patterns of crowd activities that are subtle play a very important role in triggering a real-time alarm to alert of potential danger such as stampedes, failed evacuations and crushes for operational decision making.

Acknowledgment: This work is supported by the University of Malaya Program Rakan Penyelidikan UM (PRPUM) under Grant CG065-2013.

9 December 2013
doi: 10.1049/el.2013.3993
One or more of the Figures in this Letter are available in colour online.

M.K. Lim and C.S. Chan (*Center of Image and Signal Processing, University of Malaya, Lembah Pantai, Malaysia*)

E-mail: imeikuan@siswa.um.edu.my

D. Monekosso and P. Remagnino (*Computing and Information Systems, Kingston University, Kingston Upon Thames, UK*)

References

1 Challenger, R., Clegg, C.W., and Robinson, M.A.: 'Understanding crowd behaviours: supporting evidence', in Leigh, M. (Ed.): 'Understanding Crowd Behaviours' (Crown, 2009), pp. 1–326
2 Xu, D., and An, J.: 'Attribute based salient image extrema detection algorithm', *Electron. Lett.*, 2005, **41**, (1), pp. 13–14
3 Boiman, O., and Irani, M.: 'Detecting irregularities in images and in video', *Int. J. Comput. Vis.*, 2007, **74**, (1), pp. 17–31
4 Ali, S., and Shah, M.: 'A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis'. CVPR, Minneapolis, MN, USA, June 2007, pp. 1–6
5 Zhou, B., Tang, X., and Wang, X.: 'Coherent filtering: detecting coherent motions from crowd clutters'. ECCV, 2012, vol. 7573, pp. 857–871
6 Rodriguez, M., Sivic, J., Laptev, I., and Audibert, J.-Y.: 'Data-driven crowd analysis in videos'. ICCV, Barcelona, Spain, November 2011, pp. 1235–1242
7 Loy, C., Xiang, T., and Gong, S.: 'Salient motion detection in crowded scenes'. ISCCSP, Rome, Italy, May 2012, pp. 1–4
8 Solmaz, B., Moore, B.E., and Shah, M.: 'Identifying behaviors in crowd scenes using stability analysis for dynamical systems', *PAMI*, 2012, **34**, (10), pp. 2064–2070
9 Liu, C., Freeman, W.T., Adelson, E.H., and Weiss, Y.: 'Human-assisted motion annotation'. CVPR, Anchorage, AK, USA, June 2008, pp. 1–8
10 Kennedy, C.A., and Carpenter, M.H.: 'Additive Runge–Kutta schemes for convection-diffusion-reaction equations', *ANM*, 2003, **44**, pp. 139–181
11 Haller, G.: 'Finding finite-time invariant manifolds in two-dimensional velocity fields', *Chaos*, 2000, **10**, pp. 99–108