2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Salient Object Detection in Tracking Shots

Karthik Muthuswamy
School of Computer Engineering
Nanyang Technological University
Singapore 639758
Email: kart0028@e.ntu.edu.sg

Deepu Rajan
School of Computer Engineering
Nanyang Technological University
Singapore 639758
Email: asdrajan@ntu.edu.sg

***Abstract*—Tracking shots have posed a significant challenge for salient region detection due to the presence of highly competing background motion. In this paper, we propose a computationally efficient technique to detect salient objects in a tracking shot. We first separate the tracked foreground pixels from the background by accounting for the variability of the pixels in a set of frames. The focus of the tracked foreground pixels is utilized as a measure of saliency of objects in the scene. We evaluate the performance of this method by comparing the salient region detection with ground truth data of the location of the salient object that are manually generated. The results of the evaluation show that the proposed method is able to achieve superior salient object detection performance with very low computational load.**

I. Introduction

Saliency refers to the degree of conspicuity of a region compared to other regions in an image or video. This conspicuity could be attributed to the contrast in some feature(s). While salient object detection algorithms in images use features such as intensity, colour and orientation, video saliency needs to exploit the additional motion information. Detecting salient object(s) in videos is challenging due to competing motion from other moving objects or the background motion resulting from ego-motion of the camera.

In [1], saliency is identified in terms of the discriminating power of features, modelled as dynamic textures, in a centre-surround formulation. Although this method performs well in videos shot with static cameras, it fails to identify the salient object accurately in tracking shots (in which the camera tracks an object). Gopalakrishnan et al. [2] utilize the control theory concept of observability of a linear dynamic system, modelling pixels that have predictable motion as salient. In our previous work [3], we extended [2], by modelling background motion via controllability of a linear dynamic system. Although it outperformed [2] and performed on par with [1], tracking shots still posed a problem with low performance metrics. The main reason why previous methods fail to identify the salient object in tracking shots is the inability to suppress pixels in the background that compete for saliency.

Tracking shots are important from a saliency viewpoint. A tracked object can be deemed salient for it is the cameraman's intention (by virtue of tracking the object) to focus the viewer's visual attention on it. Tracking shots are shot with a camera mounted on a stationary pod or held in hand such as in home videos. Examples of such shots can often be found in sports broadcasts or chase sequences. Identification of the foreground object from freely moving cameras have garnered a lot of interest owing to its applicability in a variety of situations. Li et. al. [4] proposed a framework that utilizes conditional random fields (CRF) to automatically identify the foreground objects from freely moving cameras. Although this framework does not require any prior knowledge about the foreground, their method is computationally expensive due to the usage of CRFs. Sun et. al. [5] identify foreground objects through SIFT correspondence across frames of the video, using their trajectories to identify the moving object via template matching. However, SIFT is known for its lack of robustness to illumination changes and would be expensive when computed for every frame in a video. A similar approach of using temporally coherent trajectories to detect salient objects is proposed in [6]. However, this approach requires a learning step for removing inconsistent motion trajectories and depends on long-term motion trajectories to detect salient objects. Kim et. al. [7] propose an optical flow based object detection framework which utilizes corner feature points where optical flow vectors are extracted. These optical flow features are clustered using RANSAC based on the scatteredness of the optical flow vectors. Finally, the detected objects are merged using Delaunay triangulation. This method is also expensive, as the optical flow vectors are extracted and then clustered. The background is subtracted from the motion compensated frame. Motion vectors are employed by [8] to obtain a temporal dissimilarity measure of the motion along with a spatial dissimilarity measure. Depending on motion vectors alone to identify the saliency in motion in tracking shots would be largely error-prone owing to the dominance of background motion.

In this paper, we formulate saliency measure of an object in a tracking shot in terms of the degree of focus that its pixels garner when compared to the background. Blur has been used as a cue for saliency in images in [9] and [10]. Baveye et al. [9] propose a saliency model based on wavelets wherein the Difference of Gaussian (DoG) on each wavelet sub-band is combined to obtain the final saliency map. They employ a blur measure that produces a blur map, which is multiplied with a saliency map to get the final saliency map. Thus, blur is used only as a refinement for an already generated saliency map. Khan et al. [10] conducted experiments to study the effect of blur on human visual attention in images. They concluded that blur information should be integrated in visual attention

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

models to facilitate efficient extraction of salient regions. In their comparison of four different saliency algorithms, they found the graph-based visual saliency framework proposed by [11] (GBVS) and spectral residual [12] (SR) to perform the best for identifying salient regions in blurred images.

II. Saliency from Focus

A. Identifying Foreground Pixels

The first of two steps in the proposed method is to identify the foreground pixels. The most relevant work to the proposed framework is that of Sheikh et. al. [13] where they detect salient objects by exploiting a rank constraint on the trajectories on the background. However, the main drawback of their method is that they require a large number of frames to identify the salient object, an assumption which will fail when the salient object appears for just a few frames.

Attention cues in tracking shots are influenced by (a) local motion of the object and (b) global motion induced by the ego-motion of the camera. Even when the foreground regions move very fast, e.g. in a car race, the salient object would be identifiable since it is being tracked by the camera. We wish to exploit this condition in order to detect foreground regions. Consider a buffer of τ frames. Our objective is to reconstruct the foreground regions in the centre frame, based on other frames in the buffer. We use the well-known eigenimages framework [14] for this purpose. The frames in the buffer are vectorized and stacked into a data matrix V. The eigenvectors corresponding to the N largest eigenvalues of the covariance matrix of V are used to determine the projection of the center frame onto the eigen space according to $\Omega = \mathbf{U}^T(V - \Psi)$, where $\mathbf{U}$ is the matrix of eigenvectors and Ψ is the mean of the frames in the buffer. The centre frame is reconstructed as $\Phi_t = \mathbf{U}\Omega + \Psi$. Fig. 1(a) shows a centre frame from a video sequence and Fig. 1(b) shows the reconstructed frame in which the foreground region corresponding to the tracked car can be seen while the other cars with inconsistent motion with respect to the camera are poorly reconstructed. As we would like to obtain consistent amount of information from each video matrix V, N is determined as the number of eigenvectors corresponding to $\epsilon\%$ of the energy present in the eigenvalues. We set $\epsilon = 80\%$ in our experiments.

B. Measuring Saliency

The second step in identifying the salient object is to determine the blur in the reconstructed frame. Regions that have been reconstructed well will be relatively less blurred than other regions, and are able to successfully identify the foreground pixels as we account for the pixels with the most variability in the video matrix V. Focus measures have been utilized to ensure the sharpness of the image captured by a camera.

Focus measures have been classified into three main groups [15] namely, Gradient-based, Laplacian-based and Statistics-based operators. Gradient based approaches use the first derivative of the image in order to measure the degree of focus. The sum of squares of the gradient magnitude from

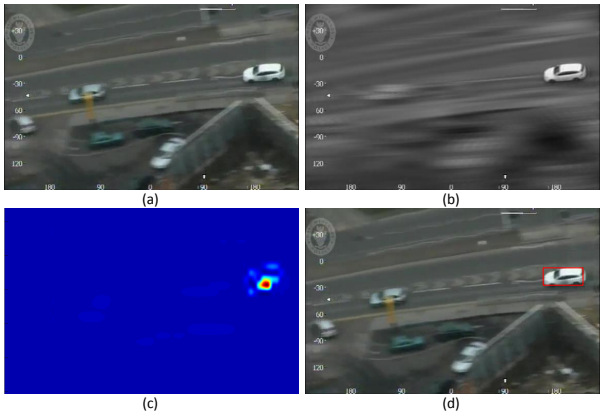

Fig. 1. (a) Original frame (b) Reconstructed frame from a buffer of $\tau = 7$ frames (c) Saliency map (colour coded for better visualization) (d) Bounding box showing salient object.

both directions has been proposed as a focus measure in [16], [17]. However, one of the popular measures of focus is computed from the magnitude of the image gradient obtained by convolving the image with the Sobel operator [16], [18]. Laplacian-based approaches utilize the second derivative of the image to measure the focus of the pixels in a given area. The energy of the Laplacian of an image has shown promising results as a focus measure [16], [17]. A faster version of this method was proposed by modifying the Laplacian calculation, by Nayar et al. [19]. Vertical variations of the image are included in [20] in order to compute a modified Laplacian of the image. Statistics-based focus measures are obtained by measuring various image features. Chebyshev moments have been shown to be a good measure of the focus, based on the ratio of the energy of the high-pass band and the low-pass band in [21]. The trace of the matrix of Eigenvalues calculated from the covariance of the image has been shown to be a good measure of sharpness in [22]. The interested reader is referred to [15] for a detailed study of the performances of various other focus measures. Thelen et al. [20] proposed a variation of the Laplacian which included vertical variations of the image while computing the Laplacian of the image, utilizing the energy of the Laplacian as a focus measure.

Pertuz et al. [15] show that the measure proposed by Nayar et al. [19] and Thelen et al. [20] have the best overall performance from among 36 different focus measures. We employ [19] as a measure of focus (or blur) owing to its superior performance and low computation cost. The focus at location (x, y) in the reconstructed frame is determined by [19]

$$\zeta(x, y) = \sum_{(i,j)\in\omega(x,y)} \Delta_m\Phi'(i, j), \tag{1}$$

where $\Delta_m\Phi'$ is the modified Laplacian of Φ' computed as $\Delta_m\Phi' = |\Phi' \star \eta_x| + |\Phi' \star \eta_y|$ where $\eta_x = [-1, 2, -1]$ and $\eta_y = \eta_x^T$, and ω is the neighbourhood of (x, y) pixels using which the focus measure is evaluated. The focus measure is

normalized in each frame to the range [0,1] as the measure of saliency.

Fig. 1(c) shows the saliency map for the frame in Fig. 1(a). The saliency map is colour coded (blue indicates low saliency while red indicates high saliency). As seen, the region corresponding to the car that is tracked by the camera has been correctly marked as the salient region. We avoid marking logos and watermarks as salient by by setting to zero in the saliency map, those pixels whose intensity variance across the entire (original) video is less than a threshold. In order to obtain the bounding box for the salient object in the saliency map, we binarize it and apply an erosion operation followed by dilation with a kernel of size 5×5 to remove isolated patches. Connected component labelling identifies the region with the largest area around which the bounding box is drawn. Fig. 1(d) shows the salient object detected in the sequence. As the bounding box is calculated for every frame, we maintain temporal coherence by ensuring that the frame-to-frame bounding box area does not exceed or reduce by more than 10%. If it does, we reduce the threshold for binarization, thereby increasing the area of the salient object at the expense of introducing stray binary regions.

III. Experimental Results

In the first experiment, we demonstrate that applying the focus measure directly on the raw image will not yield the correct salient object. Fig. 2(a) shows the original frame in which the car in the foreground is in focus and the background is largely blurred. When the focus measure of Eq. (1) is applied on it, the resulting saliency map consists of regions from the background that are also marked as salient as seen in Fig. 2(b). However, when the focus measure is applied on the reconstructed frame, it suppresses the background and correctly marks only the car as the salient object (Fig. 2(c)). As mentioned in Sec. I, current methods for salient object

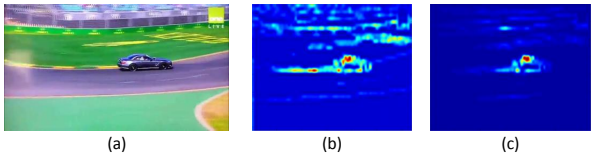

Fig. 2. (a) Original frame (b) Saliency map from raw frame (c) Saliency map from reconstructed frame.

detection fail in tracking shots like those in the dataset of [1]. In the next experiment, we test our method on tracking shots available from the dataset and compare the performance quantitatively with other methods in terms of equal error rate (EER). EER is the value at which the false alarm rate is equal to the miss rate. Table I compares the EER measures of the proposed method (FM) with that of Discriminant Saliency (DS) [1], Monnet et al (MO) [23], Itti et al. (IT) [24], Modified Gaussian Mixture Model (GMM) [25] and Sustained Observability (SO) [2], Compressed Domain Motion Saliency (CDMS) [26] and State Controllability (SC) [3]. Although the methods compared are generic motion saliency detection techniques, none of the methods are able to model the highly dynamic background motion present in such sequences. The simplicity of the proposed method does not preclude it from outperforming all the other methods.

Along with the comparison of the objective measures described in [1] and [2], we provide a comparison of two spatio-temporal saliency measures we had proposed in [3] and [26] to show the improvement in the performance of saliency detection using the proposed method. The videos in the dataset provided by [1] are challenging sequences of objects moving in a dynamic background. The 'Hockey' video is an example where an ice hockey player is skiing towards a camera while the background is randomly moving audience. The 'Cyclists' and 'Landing' videos are examples of tracking shots where the objects are moving while the background has competing salient regions. The camera tracks the salient object surfing on a highly dynamic wave background in the 'Surf' and 'Surfers' videos. Table I provides the performance comparison of the EER measures of the five tracking videos with that of Discriminant Saliency (DS) [1], Monnet et al. (MO) [23], Itti et al. (IT) [24], Modified Gaussian Mixture Model (GMM) [25] and Sustained Observability (SO) [2]. We have reproduced the objective performance measures of the various methods, reported by the authors of [1] and [2]. We also include the results of our previous frameworks [26] Compressed Domain Motion Saliency (CDMS) and [3] State Controllability (SC) to show the improvement in performance of the proposed framework, Focus Measure (FM). MO [23] uses an on-line auto-regressive model to predict the dynamic background motion. Salient motion is detected by comparing the predicted dynamic background and the observed frame with an assumption of the availability of background only scenes. Both SO [2] and SC [3] utilize an ARMA model to model the video as a dynamic texture and identify the salient pixels by using control theory concepts. CDMS [26] uses compressed domain motion vectors and colour information to identify regions of saliency. However, all these methods fail to model salient foreground motion when there is competing motion present in the video. Although DS performs well on videos shot using stationary cameras, the tracking motion of the camera distorts the foreground-background centre-surround saliency measure of the pixels, resulting in higher false positives.

Fig. 3 shows the results of salient object detection for the 'Surfer' and 'Cyclist' videos. These are particularly challenging sequences due to the highly dynamic background of the waves and of the bushes, respectively. One of the original frames, the corresponding salient regions and the bounding box around the identified salient object are shown in Figs. 3 (a), (b) and (c), respectively.

We also show the results of salient object on the videos provided in [1], in Fig. 3. Fig. 3(a) shows the frame extracted from the videos while Fig. 3(b) shows the saliency map calculated using the proposed saliency framework while Fig. 3(c) shows the bounding box calculated using the generated saliency map.

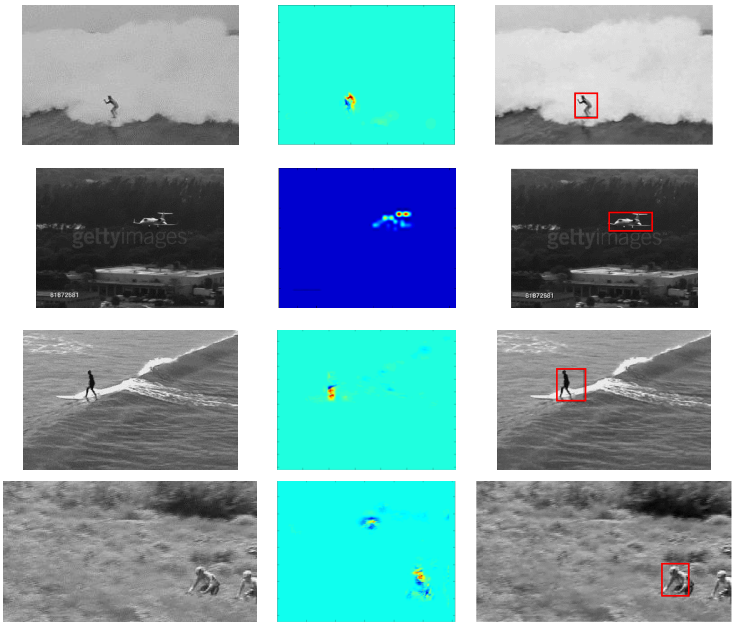

Fig. 3. Qualitative performance: (a) Original frame, (b) Saliency map (shown as hot spots) (c) Bounding box around salient object.

The top row shows a frame from the 'Surf' video sequence while the bottom row shows a frame from the 'Cyclists' video sequence. In both these videos, it can be seen that the proposed saliency framework is able to neglect the highly dynamic background motion to identify the salient object, indicated by the red box.

TABLE I
COMPARISON OF EER MEASURES (IN %) OF TRACKING SHOTS OF DS [1], MO [23], IT [24], GMM [25], SO [2], SC [3] AND CDMS [26] FROM THE DATASET PROVIDED BY [1] WITH THE PROPOSED METHOD, FM.

	Other Motion saliency algorithms							
Video	DS	MO	IT	GMM	SO	SC	CDMS	FM
Cyclists	8	28	41	36	17.9	8.8	14.5	**4.6**
Hockey	24	29	28	39	27	18.1	25.4	**9.1**
Landing	3	16	31	40	27	16.3	11.5	**2.3**
Surf	4	10	30	23	7.6	7.5	4.5	**3.6**
Surfers	7	10	24	35	9	5.2	6.4	**4.8**
Average	7.6	16	26.2	29.7	12.6	9.2	12.4	**4.9**

As the number of tracking shots available in the dataset provided by [1] are very few in number, we obtained tracking shots available in the public domain and manually marked the salient object in each frame. As tracking shots are a commonality in sport videos, we identified 25 different tracking videos from *YouTube*. We marked the salient object in each frame of these 25 videos in order to construct the ground truth information. In order to obtain the ground truth, we manually marked a bounding box (rectangle) around the tracked salient object on each frame of the video. In order to obtain the bounding box from the saliency maps, we binarized the saliency maps obtained from each frame and applied an erosion operation followed by a dilation with a kernel of size 5×5 to remove noisy neighbours in the binary image. Subsequently, we employ connected component labelling on the binary image to identify the region with the largest area as the region around which we calculate the bounding box coordinates.

Fig. 4 shows the bounding boxes calculated for three different shots from the videos we collected, where Fig. 4 (a) shows the marking scheme used to obtain the location of the bounding boxes on each frame while Fig. 4 (b) shows the bounding boxes identified by the proposed framework (as a red rectangle). We compare the performance of the proposed

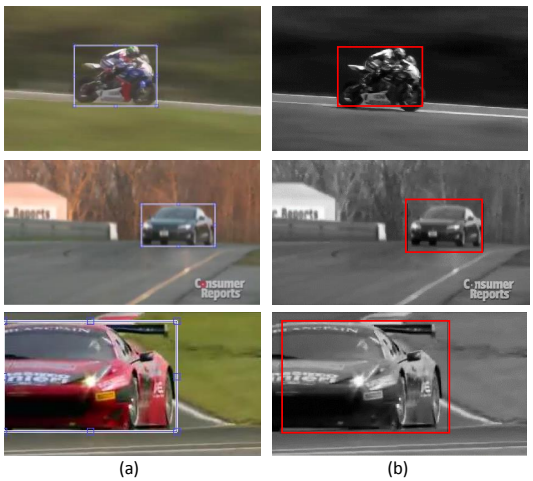

Fig. 4. Qualitative comparison of the bounding boxes: (a) Ground-truth bounding box; (b) Bounding box identified by the proposed framework

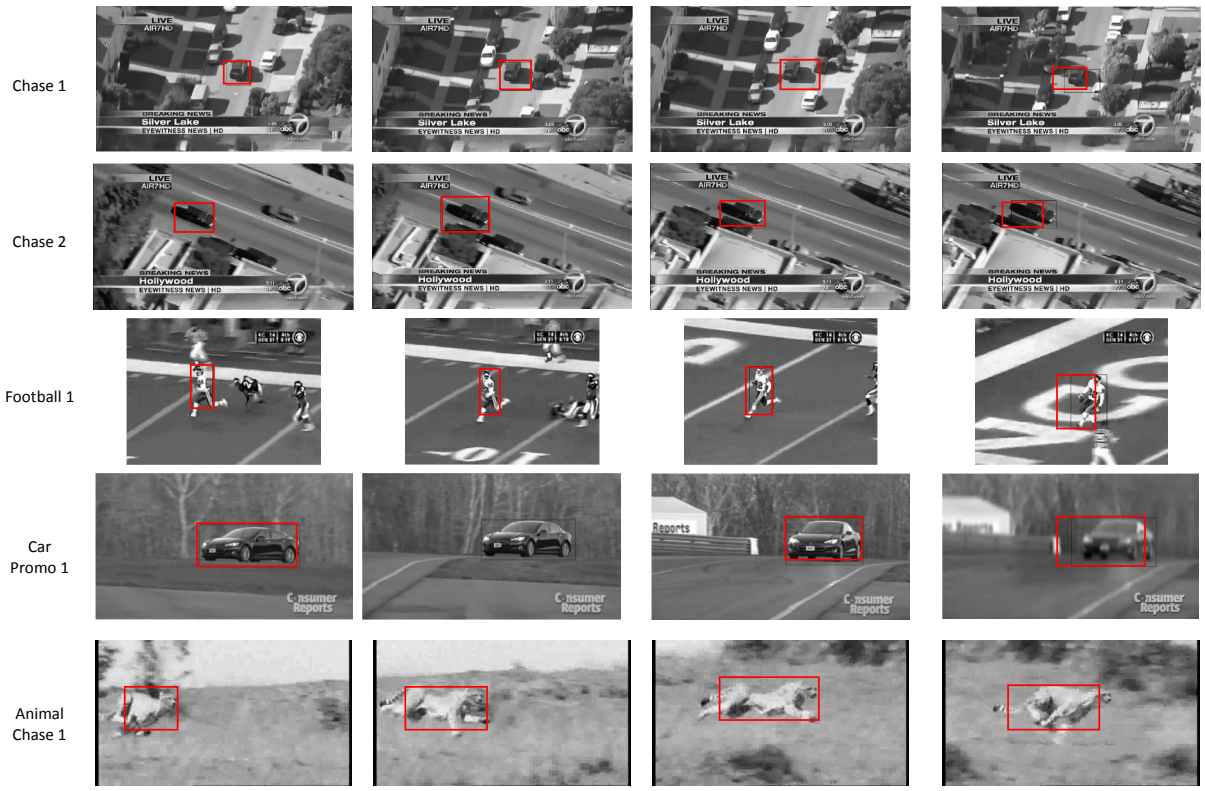

Fig. 5. Bounding box showing salient regions on five different videos.

saliency measure with six other spatial saliency measures namely IT [27], AC [28], RCS [29], SS [30], GBVS [11] and SR [12]. For our comparison we grouped the 25 videos we collected under 6 genres of tracking such as 'Racing', 'Chase', 'Landing', 'Hunting', 'Football' and 'Promos'. The videos under the 'Racing' genre has 5 different videos that were shot at Motorbike, Nascar and F1 racing events. These are quite challenging sequences primarily owing to the speed of the vehicle and the presence of distracting regions in the background. The videos categorized under the 'Police Chase' genre has 4 different car chase sequences that were shot from a helicopter. These sequences are challenging owing to a number of occluding buildings and cars that would impede the identification of the salient object in the scene. The videos categorized under the 'Landing' consists of 4 different plane landing sequences shot from different angles around an airport. 'Hunting' consists of 4 videos of animals hunting for their food. The challenge in such types of videos lies in the demarcating the animal from the background grass and trees. Videos under the 'Football' are shot by both professionals and amateurs. This category consists of 2 videos shot by professional cameramen while the other 2 were amateur videos shot at a school game. Videos under 'Promos' consists of 4 videos of cars shot by professional car video makers. The videos in this category are typically shot at very high resolution wherein the foreground car is clearly demarcated from the background by using a depth of field effect.

Table II reports the average f-measures of the videos categorized under these 6 groups. It can be seen that the proposed method is able to consistently localize the tracked object as shown by the high average f-measure for each video category. Even though the study conducted by [10] reports GBVS and SR to perform quite well with blurred images, it can be seen that the proposed framework is able to perform better than these two methods consistently across different categories. Figs. 4 and 5 corroborates the f-measures reported for different videos from the collected dataset.

TABLE II
F-MEASURES OF IT [27], AC [28], RCS [29], SS [30], GBVS [11], SR [12], PROPOSED METHOD FM ON DIFFERENT VIDEOS. THE NUMBER OF VIDEOS ARE IN PARENTHESIS.

Video	IT	AC	RCS	SS	GBVS	SR	FM
Racing (5)	0.47	0.37	0.48	0.56	0.47	0.64	**0.83**
Chase (4)	0.16	0.53	0.07	0.12	0.08	0.04	**0.74**
Landing (4)	0.47	0.33	0.03	0.65	0.65	0.64	**0.86**
Hunting (4)	0.14	0.37	0.22	0.59	0.35	0.57	**0.68**
Football (4)	0.16	0.5	0.18	0.45	0.28	0.52	**0.71**
Promos (4)	0.03	0.2	0.42	0.49	0.45	0.61	**0.94**
Average (25)	0.24	0.38	0.23	0.48	0.38	0.50	**0.80**

We show the results of salient object detection in Fig. 5 for five different videos from the dataset we collated. Additional results and detailed analysis are available on the supplementary material provided with the paper. In Fig. 5, Rows 1 and 2 are two chase sequences with highly competing background pixels consisting of large buildings and vehicles that travel alongside the salient object. The proposed method can accurately identify the salient car even when it is occluded over a number of frames as the proposed framework depends largely on pixels with the least intensity variance around a set of frames which it can easily locate the salient pixels even under partial occlusion. Row 3 consists of frames from a video shot at an American football game. Even though there are a number of players with similar motion and confusing background pixels from the field markings, the proposed saliency method is able to successfully identify the tracked player as the framework relies on pixels that have the least variance in pixel intensities across a set of frames and the focus measure accurately localizes the salient object in the reconstructed scene. The results of this video have also been compared with a number of state-of-the-art spatial saliency techniques in the supplementary material provided with this paper. Row 4 shows a video where there are regions with large contrast differences that might otherwise be marked salient if a method similar to a centre-surround formulation is employed. The last row in Fig. 5 shows a cheetah chasing its prey. Even at such high speeds, the cheetah is clearly separated from the background clutter. A centre-surround formulation would fail in identifying the salient pixels in such sequences as the contrast difference between the background and the salient object is very low. The results of the saliency detection for a number of other videos in the dataset are available as supplementary material[1]. Our experiments were conducted on a 2.4 GHz Intel Core2 Duo processor with 4 GB RAM on MATLAB 2012b running on a Windows 8.1 operating system. On a video with a resolution of 640×480, the average time to process a frame by the proposed method is 0.57 seconds.

IV. Conclusion

In this paper, we proposed a novel technique to identify salient objects in tracking shots, using focus as a measure of saliency. We showed its superior performance by providing qualitative and quantitative comparisons with different methods and videos. As a future work, we would like to explore the possibility of combining the proposed framework with metadata information such as camera focal length, aperture, ISO speed, available from exchangeable image format (Exif) data stored in images, to improve accuracy and speed of saliency detection.

References

[1] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE PAMI*, 2010.

[2] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Sustained observability for salient motion detection," in *ACCV*, 2011.

[3] K. Muthuswamy and D. Rajan, "Salient motion detection through state controllability," in *ICASSP*, 2012.

[4] Q. Li, S. Chen, and B. Zhang, "Predictive video saliency detection," in *Pattern Recognition*. Springer, 2012, pp. 178–185.

[5] S.-W. Sun, Y.-C. Wang, F. Huang, and H.-Y. Liao, "Moving foreground object detection via robust sift trajectories," *Journal of Visual Communication and Image Representation*, 2013.

[6] C.-R. Huang, Y.-J. Chang, Z.-X. Yang, and Y.-Y. Lin, "Video saliency map detection by dominant camera motion removal," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 8, pp. 1336–1349, 2014.

[7] J. Kim, X. Wang, H. Wang, C. Zhu, and D. Kim, "Fast moving object detection with non-stationary background," *Multimedia Tools and Applications*, 2013.

[8] L. Duan, T. Xi, S. Cui, H. Qi, and A. C. Bovik, "A spatiotemporal weighted dissimilarity-based method for video saliency detection," *Signal Processing: Image Communication*, vol. 38, pp. 45–56, 2015.

[9] Y. Baveye, F. Urban, and C. Chamaret, "Image and video saliency models improvement by blur identification," in *Computer Vision and Graphics*, 2012.

[10] R. A. Khan, H. Konik, and E. Dinet, "Enhanced image saliency model based on blur identification," in *IVCNZ*, 2010.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006.

[12] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007.

[13] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1219–1225.

[14] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, 1991.

[15] S. Pertuz, D. Puig, and M. Angel Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, 2012.

[16] M. Subbarao, T.-S. Choi, and A. Nikzad, "Focusing techniques," *Optical Engineering*, 1993.

[17] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, 2007.

[18] H. N. Nair and C. V. Stewart, "Robust focus ranging," in *CVPR*, 1992.

[19] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE TPAMI*, 1994.

[20] A. Thelen, S. Frey, S. Hirsch, and P. Hering, "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation," *IEEE TIP*, 2009.

[21] P. Yap and P. Raveendran, "Image focus measure based on chebyshev moments," *IEE Proceedings-Vision, Image and Signal Processing*, 2004.

[22] C.-Y. Wee and R. Paramesran, "Measure of image sharpness using eigenvalues," *Information Sciences*, 2007.

[23] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *ICCV*, 2003.

[24] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Advances in Neural Information Processing Systems*, 2005.

[25] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *ICPR*, 2004.

[26] K. Muthuswamy and D. Rajan, "Salient motion detection in compressed domain," *IEEE SPL*, 2013.

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

[28] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[29] V. Narayan, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognition*, 2012.

[30] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE TPAMI*, 2012.

[1] https://sites.google.com/site/saliencytrackingshots/home/resultspage