Thematic Saliency Detection using Spatial-Temporal Context

Ye Luo Gangqiang Zhao Junsong Yuan
School of Electrical and Electronic Engineering
Nanyang Technological University
{luoy0009, gqzhao, jsyuan}@ntu.edu.sg

Abstract

We propose a new measurement of video saliency termed thematic video saliency. Video saliency is detected in terms of finding the thematic objects that frequently appear at the salient positions in the video scenes. By representing all image segments in the video as the spatial-temporal context, we build an affinity graph among them, and formulate the thematic object discovery as a novel cohesive sub-graph mining problem. A trust region algorithm is also proposed to solve the challenging optimization problem. Unlike individual image saliency or co-saliency analysis, our proposed video saliency fully incorporates the whole spatial-temporal video context. Experiments on our newly developed eye tracking dataset as well as other two datasets further validate the effectiveness of our method on video saliency detection.

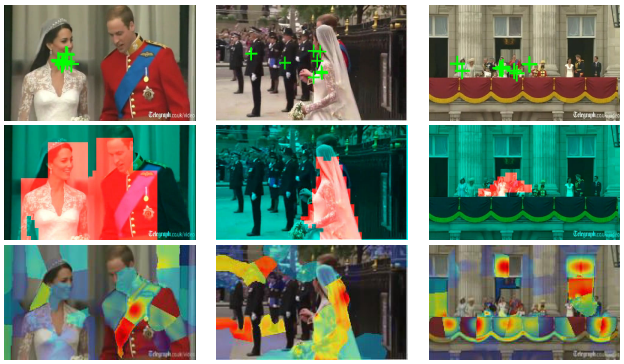


Figure 1. Sample results of thematic video saliency discovery. The first row shows the original key-frames with the gaze points overlapped. Each green cross '+' represents one gaze point. The second row shows our detection results and the discovered video thematic salient region is rendered in red. The last row shows the heat map results of the co-saliency method [14]. Hot colors correspond to large salient values.

1. Introduction

Like the salient object in an image, many videos contain the thematic object, such as the bride and the groom in a wedding ceremony video (Fig. 1), the birthday girl in a birthday party video, or a product logo in a commercial video. As the key object to be highlighted, such a thematic object appears frequently and occupies salient positions in the video scenes, thus retain our impression after watching the video. Our thematic video saliency is estimated by finding the relevance of the locations to the thematic objects in a scene. Similar to the task-relevance map [1], locations which have high relevance to the thematic objects are highlighted and large values are set to these locations. In practice, finding such a thematic object is of great interests as it can help to better understand and summarize the video contents.

Although visual saliency has been extensively studied in psychology, neuroscience and computer vision literature, unfortunately, it remains a challenge to estimate thematic saliency in a video clip. Different from the existing task driven based saliency estimation methods, which predict humans' attention with the task of searching a known object (e.g. a red wedding car), there is no prior information about the thematic objects before the search. Moreover, the leap from image saliency to video saliency analysis is non-trivial. Although a video is composed of a sequence of images, most often than not, not every single frame is important, e.g., contains the thematic object. A wedding car could be the salient object at a specific frame, but it may not be the thematic object given the whole wedding video. It thus requires the help of the video contexts to determine the thematic object. Despite various types of contexts have been explored to estimate image/video saliency in either the bottom-up or the top-down manner [7, 29], few of them apply the whole video sequence as a context for saliency estimation. On the contrary, some methods explore the spatial-temporal context but they did not use it for video saliency estimation [30].

We propose a new way to estimate thematic saliency, which takes the whole spatial-temporal video context into consideration. In order to detect the thematic object, a video is firstly decomposed into salient image segments

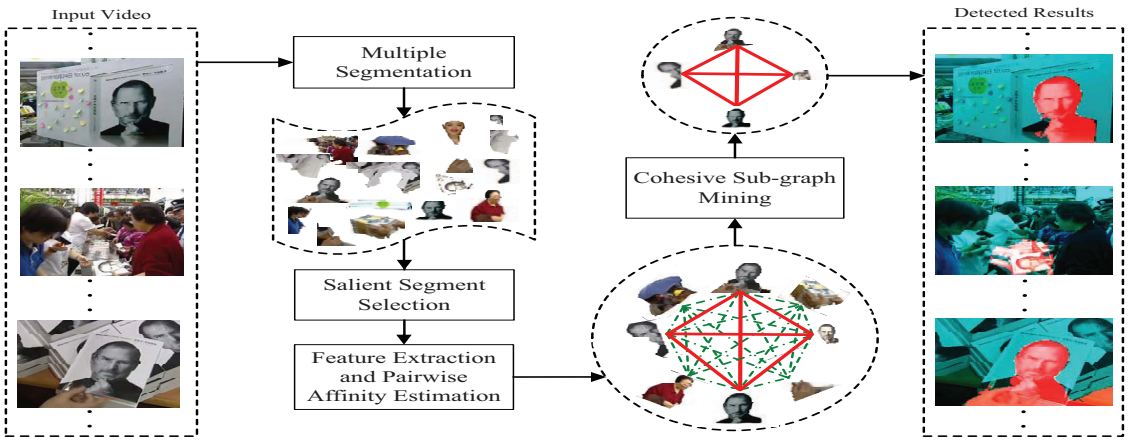


Figure 2. Illustration of the main steps in our method.

and their similarity relationships are presented in an affinity graph. By discovering the cohesive sub-graph, i.e., a group of image segments of high similarity, we can discover the thematic object. The likelihood of each segment belonging to the subgraph identifies the thematic saliency of the video. Fig. 1 shows our results and the comparison with the ground truths using eye tracking, and the co-saliency method in [14] which only emphasizes the common parts between consecutive frames.

2. Related Work

Regarding to how context has been employed to estimate saliency, saliency estimation methods are reviewed from the following three aspects. At first, context is broadly used in image saliency estimation. A center-surround feature extraction-integration scheme is used to estimate saliency at each pixel point in [10]. In [7], image saliency defined as the best parts to depict the image content is obtained by detecting salient object and its near background. In [29], image saliency estimation is modeled as anomaly detection with regard to different context, such as patch saliency to patch dictionary and image saliency to image dictionary. In order to add the global effects to the saliency estimation, features characterizing the global effects are extracted from images in [27, 24]. A global center-surround technique is implemented by using Difference-of-Gaussian (DoG) band pass filters to estimate image saliency uniformly in [2]. Secondly, by taking another image as comparison context, co-saliency as a new concept is proposed to estimate pairwise image saliency. Methods in [11, 21] aim to detect difference such as the novel signal or the newly changed positions as saliency while methods proposed in [15, 6] tend to detect the common objects in the two images as the saliency. At last, video saliency estimated from consecutive frame has also been explored recently [9]. However, most of them start from finding a wise combination scheme for different salient clues [23, 22]. Some supervised methods are also proposed to estimate video saliency [16]. Different from them, our target here is to automatically detect the video saliency from the understanding of its themes.

Common object discovery is also related to our work. It targets at finding the sharing objects among images. In [20], common object discovery is formulated as the common subgraph discovery problem in image pairs. There are also works to discover common objects in image sequences [19, 32]. Considering the computation complexity of graph-based representation of image collections, Yuan et al. [31] proposed to detect thematic objects by gradually pruning uncommon patches, but it does not leverage the visual saliency to identify the thematic object. In this paper, we propose to employ visual saliency as a primary way to prune unimportant segments and propose a new solution based on sub-graph mining which is different from any of the previously proposed methods [19, 32, 31].

3. Our Proposed Method

Fig. 2 illustrates the main steps of our method. First, the key-frames are sampled from the input video and over-segmented into superpixels. Then candidate segments are selected based on the image saliency. The appearance features are extracted for the selected segments consequently. After that the pairwise relations between segments are characterized by an affinity graph, where the red bold lines indicate the positive affinity value, while the green dashed lines show the negative affinity value. The segments as nodes on the cohesive sub-graph are obtained by maximizing the overall affinity score. The detection results are presented in corresponding key-frames at last.

3.1. Segment Selection and Representation

To obtain the segments, we perform a superpixel segmentation method [3] per key-frame, with the expecta-

Figure 3. Thematic video object discovery result comparisons with bottom-up saliency filtering (the third row) and without saliency filtering (the second row).

tion that some segments could correspond to object regions while some may fail to agree with object boundaries. Several types of bag-of-features histograms are used to describe each segment and the details can be found in Sec. 4.2.

In order to distinguish candidate segments from background ones, a simple saliency estimation is performed to select image segments. To obtain the simple saliency, we use a linear combination of spatial saliency [8], motion saliency [18], and a face mask [28] to represent our simple video saliency. The face detector is also employed to emphasize human being's particular interests to human faces. Before the fusion, each saliency map is normalized into $[0, 1]$ by min-max normalization. Similar to [2, 4], a segment is kept as a salient object region if its average saliency is larger than its saliency density. Fig. 3 shows the benefits of saliency filtering for segment selection. After the segment selection, more thematic objects rather than the background have been correctly detected. It is worth noting that, even if some segments of the thematic objects are occasionally missed by the bottom-up saliency, our method can still discover them considering their frequent occurrences in the whole video.

3.2. Affinity Graph of Image Segments

Before giving the affinity estimation, we explain the concept of support segment set firstly. For a specific segment S_i in a key-frame I_m, we select segments which have similar appearance as S_i from all other key-frames. All these selected segments are the support segment set of S_i which is defined as $\{\Psi_{mi}\}$, where $\Psi_{mi} = \{S_{mi}\}$ represents all support segment of S_i in key-frame I_m, and S_{mi} is one support segment in key-frame I_m. To reduce the size of the support segment set, we only select one support segment which has a smaller distance with S_i. In other words, the size of Ψ_{mi} is set to be 1. The support segment set formulation integrates not only the spatial information of the salient object, but also the temporal trajectory information. Therefore, by comparing the support segment sets of two segments, we can obtain the affinity relationship of these two segments. If two support sets have an intersection of large size, then the two segments have high affinity relationship. Otherwise, these two segments have weak affinity relationship.

Given the support set of each segment, the pairwise segment affinity can be estimated spatial-temporally. Based on the Jaccard similarity coefficient, the affinity value of two segments is defined as:

$$A_{i,j} = \begin{cases} \frac{|\{\Psi_{mi}\} \bigcap \{\Psi_{mj}\}|}{|\{\Psi_{mi}\} \bigcup \{\Psi_{mj}\}|} & \text{if } |\{\Psi_{mi}\} \bigcap \{\Psi_{mj}\}| > 0 \\ \tau & \text{else} \end{cases}, \quad (1)$$

where τ is a negative value and $|.|$ represents the cardinality of one set. If S_i and S_j have strong affinity, the value of $A_{i,j}$ is positive and vice versa. If the support segment sets of S_i and S_j don not have any intersection, $A_{i,j}$ is set to be a constant τ.

3.3. Cohesive Sub-graph Building and Mining

According to our observation, the segments belonging to the same thematic object share the similar appearance and have strong mutual affinity relationship. On the other hand, they tend to have weak affinity relationship with the segments from the background or other objects. Therefore, we represent the thematic saliency by the cohesive sub-graph and denote this sub-graph by using its vertices set Ω, where elements of Ω are the segments belong to the same thematic saliency. In other words, the thematic saliency can be represented as the spatial-temporal collocated segment group $\Omega \subseteq \mathit{\Pi}$, where all segments $S_i \in \Omega$ belong to the same thematic saliency. The affinity potential function of the sub-graph is defined as Ω as $f(\Omega) = \sum_{S_i, S_j \in \Omega} A_{i,j}$ and the solution to the following optimization problem gives the maximum cohesive sub-graph:

$$\Omega^* = \underset{\Omega \subseteq \mathit{\Pi}}{arg\max}\, f(\Omega), \quad (2)$$

i.e., the sub-graph that has the largest affinity potential is the maximum cohesive sub-graph. Thematic segment presented by a sub-graph can be discovered one by one with erasing the corresponding features belonging to previously found segment.

When obtaining the affinity matrix A for all segment pairs, the subset optimization problem in Eq. 2 can be converted to a binary optimization problem. Given a sub-graph Ω, let $\mathbf{x} = \{x_i\}_{i=1}^N$ with $x_i \in \{-1, 1\}$ represents its indicator vector. When $x_i = 1$, segment S_i belongs to sub-graph Ω, and vice versa. As the indicator vector $\mathbf{x}$ and the sub-graph Ω correspond to each other, Eq. 2 can be rewritten as:

$$\begin{aligned} \mathbf{x}^* &= arg\max_{\mathbf{x}} f(\mathbf{x}) = \tfrac{1}{4}(\mathbf{1}+\mathbf{x})^T A(\mathbf{1}+\mathbf{x}), \\ &s.t. \quad x_i \in \{-1, 1\}, \quad i = 1, ..., N, \end{aligned} \quad (3)$$

where $f(\mathbf{x}) = \frac{1}{4}(\mathbf{1}+\mathbf{x})^T A(\mathbf{1}+\mathbf{x})$ is the objective function. Eq. 3 is a binary quadratic programming (BQP) problem. Since A may not be the positive definite matrix, the

objective function $f(\mathbf{x})$ can be non-convex and the solution is difficult to be obtained. Inspired by the methods in [32] and [30], we can obtain our solution as follows.

To solve Eq.3, we relax a binary constraint $x_i \in \{-1, 1\}$ to an equilibrium constraint, i.e.,$-1 \leq x_i \leq 1, (1 + x_i)(1 - x_i) = 0$, and this equilibrium constraint can be implied by the nonlinear complementarity problem (NCP) function $\psi(1 + x_i, 1 - x_i) = 0$ [5]. In the implementation, we select the popular Fischer-Burmeister function $\psi(a, b) = \sqrt{a^2 + b^2} - (a + b)$ and obtain the differentiable constraints: $\psi(1 + x_i, 1 - x_i) = \sqrt{2 + 2x_i^2} - 2 = 0$.

For simplification, the Fischer-Burmeister function $\psi(1 + x_i, 1 - x_i) = 0$ is written as $\psi(x_i) = 0$. To deal with the constraint $\psi(x_i) = 0$, we introduce the quadratic penalty $\sum_{i=1}^{N} \psi^2(x_i)$ into the objective function and Eq. 3 can be reformulated as the following continuous optimization problem:

$$\begin{aligned} \mathbf{x}^* &= arg\max_{\mathbf{x}} F(\mathbf{x}) = f(\mathbf{x}) - \tfrac{\beta}{2} \textstyle\sum_{i=1}^{N} \psi^2(x_i), \\ s.t. &\quad -1 \leq x_i \leq 1, \quad i = 1, ..., N, \end{aligned} \tag{4}$$

where $\beta > 0$ is a penalty parameter. By adding the $-\frac{\beta}{2} \sum_{i=1}^{N} \psi^2(x_i)$ into the objective function, it not only incorporates the constraint but also obtains a concave objective function when the penalty parameter β is one large positive value (the supplementary material provides the proof).

To solve the optimization problem of Eq. 4 with a specific penalty parameter β, we employ the trust region newton method [17]. After all iterations, the global and superlinear convergence of the trust region Newton method leads to an efficient implementation for the cohesive sub-graph mining, and the maximum cohesive sub-graph is accordingly obtained as the best solution of our method.

4. Experimental Results

In this section, two public datasets and one self-built eye-tracker dataset are used to evaluate our approach on the thematic video saliency detection.

4.1. Datasets

Eye tracker dataset To validate that the estimated video saliency can represent human's understanding of a video theme, we designed an experiment via asking 10 participants to find thematic objects for 5 videos and recording their gaze data at the time by the ASL eye tracker system. Five videos are all from Youtube.com with averaged length around 40 seconds. Each video has at least one theme such as Steve Jobs' biography, or Prince William with his bride Kate. Ten participants (5 males, 5 females) are with normal sight or corrected normal sight. We use the recorded gaze data as the ground truth to indicate the appearance of thematic saliency in each video. In particular, we convolve gaze points by a Gaussian filter to generate the ground truth of the thematic video saliency as [12] did. All videos and the corresponding ground truth will be released soon.

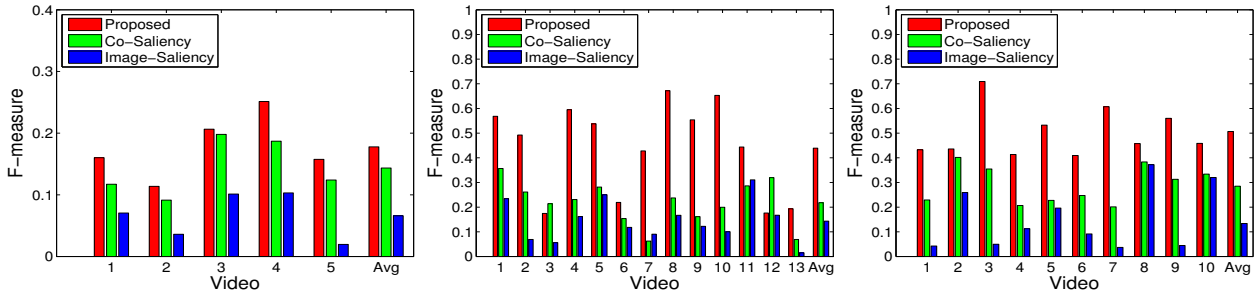
RSD saliency dataset [16] The dataset contains videos from six genres: documentary, ad, cartoon, news, movie and surveillance. Twenty-three subjects are assigned to manually label the rectangular regions to indicate the binary saliency maps. Since this dataset is built for regional video saliency detection which is different from our aim to find the thematic video saliency, we only choose ad and news videos for evaluation in that the labeled regions are coincident to the video themes.

Commercial video dataset [32] In this dataset, there are ten commercial advertisement video sequences with the length of videos ranging from 30 to 40 seconds. The ground truth, which are the manually labeled bounding boxes for the thematic object masks, are provided.

4.2. Experimental Setting

Parameters To obtain the segment representation of a video, we first uniformly sample key-frames at 2 frames per second from the video. Then each key-frame is segmented multiple times using normalized cut [26] with different number of segments K (K = 3, 5, 7, 9, 11 and 13). To handle the scale problem, image segmentation is performed into two scales: original key-frames and a half size of the original ones. Several types of bag-of-features histograms are used to describe a segment: SIFT Histograms, Texton Histograms (TH), Color Histograms (CH), and pyramid of HOG (pHOG) [13]. After extracting SIFT from each key-frame, all SIFT features in a video are quantized into 1000 visual words by k-means. For TH, we use a filter bank with 18 bars and edge filters and quantize them to 400 textons via k-means. For CH, we use Lab color space, with 23 bins per channel. For pHOG, we use 3 pyramid levels with 8 bins. The concatenation of four types of feature histograms is used to describe each segment. We empirically set the negative value τ in Eq. 1 to be -0.05 and the penalty parameter $\beta > 0$ to be 10. Other parameters of the trust region methods are set similar with the modern trust region methods [17]. All these parameters are fixed in our experiments. All the experiments are performed on an Xeon 2.67GHz PC. After the segmentation and feature extraction, the proposed method can process the 30-second length video in one minute.

Evaluation criteria To quantify the performance of the proposed approach on the employed video datasets, a precision-recall based measurement is employed. Let DR and GT be the discovered thematic object region and the bounding boxes of ground truth, respectively. Precision and recall can be calculated as: $P = \frac{|GT \cap DR|}{|DR|}$ and $R = \frac{|GT \cap DR|}{|GT|}$. To generally measure P and R, we employ $F\text{-}measure = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R}$ ($\alpha = 0.3$ as in [2, 25]) as the



(a) (b) (c)

Figure 4. The performance evaluation of the proposed video saliency detection algorithm, the co-saliency algorithm [14], and the single image saliency algorithm [2]. (a) shows the results of the eye tracker dataset, (b) shows the results of the RSD dataset. (c) shows the results of the commercial video dataset.

only evaluation criterion. To obtain $F\text{-}measure$ of a video, $F\text{-}measure$ for each key-frame is first calculated and then the average value of all key-frames is used for the whole video.

4.3. Comparisons with Eye Gaze Data

To validate that the estimated video saliency can represent human's understanding of a video theme, we compare our results with the ground truth obtained by the eye tracker system. Fig. 5(a) and (b) show some sample results of thematic saliency discovery on the eye tracking dataset. In these videos, the thematic saliency is subject to variations caused by partial occlusions, scales, viewpoints and lighting condition changes. But our results still get the best overlap with the gaze points. The objective measurements by $F\text{-}measure$ for our method on this dataset are shown in Fig. 4(a). The high value of $F\text{-}measure$ demonstrates that the obtained thematic video saliency gets the best fit to human's understanding of a video theme.

4.4. Comparisons with Other Approaches

In order to show the superiority of the proposed approach on video saliency estimation, we compare with several closely related saliency estimation methods: the co-saliency method [14] and the image saliency model[2]. To fairly compare our method with [2] and [14], we follow [2] to select the smallest rectangular region containing at least 95% salient points in each frame as the target salient region. Due to the thematic objects only appearing in several frames and every frame will have a bottom-up saliency by both of the saliency detetion methods, for those key-frames that the thematic saliency is lost by our method but having image saliency/co-saliency, we select the top 30% salient pixels from image saliency/co-saliency map as their detection results.

Fig. 5(a) and (b) show the sampled detection results on the eye tracker dataset. From the comparisons, we can see that our method can correctly avoid the false alarm when there is no thematic object instance on a frame while the co-saliency and image-saliency based methods are not so discriminative in finding thematic objects due to the lack of spatial-temporal context information, e.g. some background are wrongly detected as the thematic objects in the third and the forth rows of Fig. 5(a) and (b). More results on the commercial video dataset are shown in Fig. 5(c). The objective measurements by $F\text{-}measure$ for three methods on three different datasets are shown in Fig. 4(a), (b) and (c), respectively.

It is also worth mentioning that we did not provide the comparisons with the common object discovery method [32] in that we have different objectives.

5. Conclusion

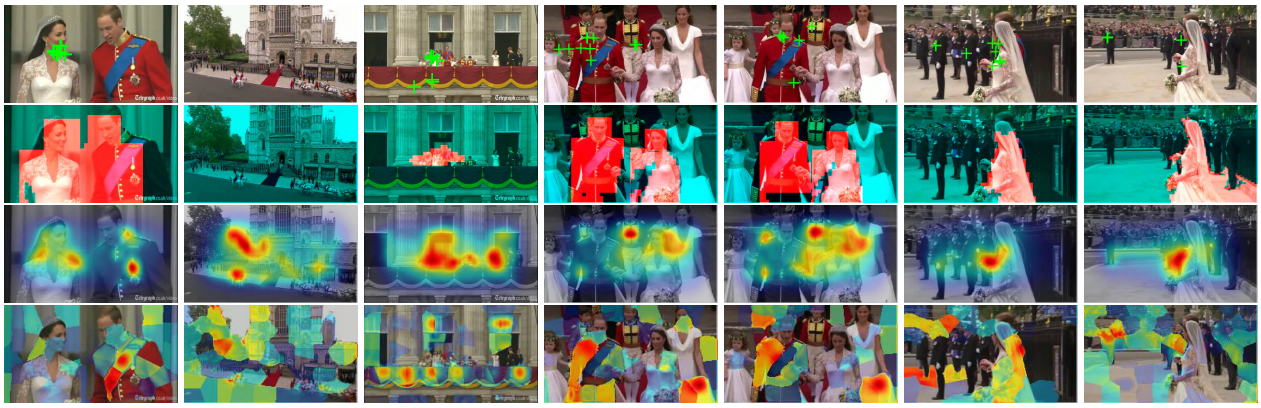
Thematic saliency detection in videos is a challenging problem due to the possibly large visual pattern variations of the thematic object and the prohibitive computational cost to explore the candidate set without *a priori* knowledge of the thematic object. By representing the relations of all frame segments in the video as an affinity graph, we formulate the thematic object discovery problem as a novel cohesive sub-graph mining problem. Our approach has the ability to identify the thematic saliency and accurately locate its regions in the cluttered and dynamic video scenes. Experiments on challenge video datasets show that our method is efficient, robust and accurate.

References

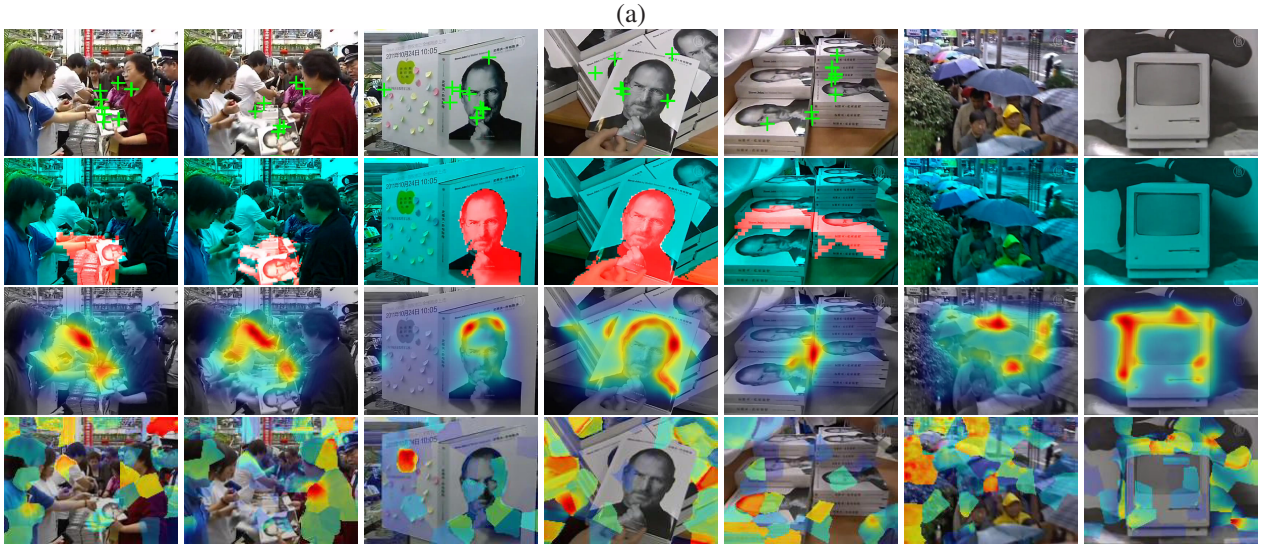
[1] Modeling the influence of task on attention. *Vision Research*, 45(2):205 – 231, 2005. 1

[2] R. Achanta, S. Hemami, F. Estraday, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009. 2, 3, 4, 5

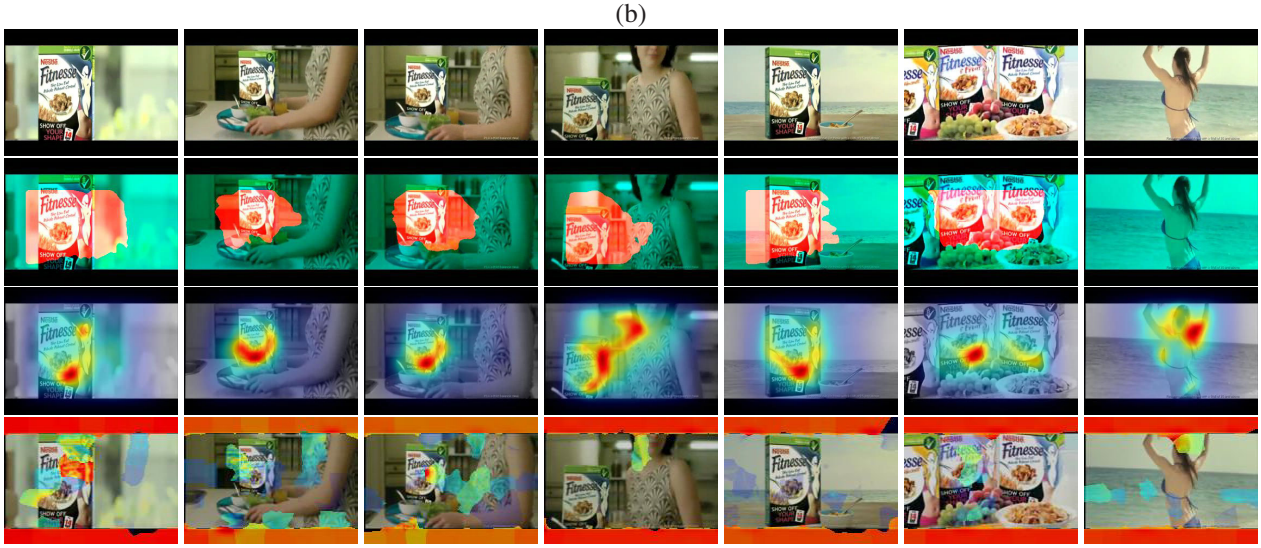
[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to



(a)



(b)



(c)

Figure 5. Sample results of video saliency discovery. (a) and (b) show two videos of our eye tracker dataset. (c) shows one commercial video. The first rows of all three sub-figures show the original key-frame and the gaze points obtained by the eye tracker are also rendered for (a) and (b). Each green cross '+' represents one gaze point. The second rows show our results. The heat map results of image saliency and co-saliency are shown in the third and fourth rows respectively. Hot colors correspond to large salient values.

state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012. 2

[4] K.-Y. Chang, T.-L. Liu, and S.-H. Lai. From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model. In *CVPR*, pages 2129–2136, 2011. 3

[5] C. Chen and O. L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Comput. Optim. Appl.*, 5:97–138, March 1996. 4

[6] D. Y. Chen and C. Y. Lin. Efficient co-salient video object detection based on preattentive processing. In *ICME*, pages 2097–2104, 2008. 2

[7] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, pages 2376–2383, 2010. 1, 2

[8] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552. MIT Press, 2007. 3

[9] L. Itti and P. Baldi. Bayesian surprise attracts human attention. In *NIPS*, volume 19, pages 547–554, 2006. 2

[10] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11):1254–1259, 1998. 2

[11] D. E. Jacobs, D. B. Goldman, and E. Shechtman. Cosaliency: Where people look when comparing images. In *Proc. UIST*, pages 219–228, 2010. 2

[12] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, 2009. 4

[13] Y. J. Lee and K. Grauman. Object-graphs for context-aware visual category discovery. *TPAMI*, 99, 2011. 4

[14] H. Li and K. N. Ngan. A co-saliency model of image pairs. *IEEE Trans. on Image Processing (TIP)*, 110(3):346 – 359, 2011. 1, 2, 5

[15] H. Li and K. N. Ngan. A co-saliency model of image pairs. *TIP*, 110(3):346 – 359, 2011. 2

[16] J. Li, Y. Tian, T. Huang, and G. Wen. Probabilistic multi-task learning for visual saliency estimation in video. *IJCV*, 90:150–165, 2010. 2, 4

[17] C.-J. Lin and J. J. More. Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9:1100–1127, 1998. 4

[18] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, MIT, 2009. 3

[19] D. Liu, G. Hua, and T. Chen. A hierarchical visual model for video object summarization. *TPAMI*, 2010. 2

[20] H. Liu and S. Yan. Common visual pattern discovery via spatially coherent correspondences. In *CVPR*, pages 1609–1616, 2010. 2

[21] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *CVPR*, pages 2097–2104, 2011. 2

[22] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *TPAMI*, 99:353–367, 2010. 2

[23] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. *TPAMI*, 32:171–177, 2010. 2

[24] F. Perazzi, P. Krhenbhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, pages 733–740, 2012. 2

[25] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, pages 853–860, 2012. 4

[26] J. Shi and J. Malik. Normalized cuts and image segmentation. 22:888–905, 2000. 4

[27] R. Valenti, N. Sebe, and T. Gevers. Image saliency by isocentric curvedness and color. In *ICCV*, pages 2185–2192, 2009. 2

[28] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004. 3

[29] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley. Image saliency from intrinsic to extrinsic context. In *CVPR*, pages 417–424, 2011. 1, 2

[30] J. Xu, J. Yuan, and Y. Wu. Learning spatio-temporal dependency of local patches for complex motion segmentation. *Computer Vision and Image Understanding(CVIU)*, 115(3):334 – 351, 2011. 1, 4

[31] J. Yuan, G. Zhao, Y. Fu, Z. Li, A. K. Katsaggelos, and Y. Wu. Discovering thematic objects in image collections and videos. *TIP*, 21(4):2207–2219, 2012. 2

[32] G. Zhao and J. Yuan. Discovering thematic patterns in videos via cohesive sub-graph mining. *ICDM*, pages 1260–1265, 2011. 2, 4, 5