Smart Thumbnail: Automatic Image Cropping by Mining Canonical Query Objects

Yang Yang[1], Linjun Yang[2], and Gangshan Wu[1]

[1] State Key Laboratory for Novel Software Technology, Nanjing, 210046, P.R. China
[2] Microsoft Corporation, One Microsoft Way, Redmond, WA, 98052, USA
charlie.yang.nju@gmail.com, linjuny@microsoft.com, gswu@nju.edu.cn

Abstract. In this paper, we present a query-dependent thumbnailing approach for web image search. Motivated by the fact that uniform down-sampling cannot emphasize query objects while saliency-based methods may present incorrect foreground objects, we propose to employ common object discovery (COD) algorithms to mine the underlying canonical query objects from the result image collection and adopt the detected object regions of interest (ROIs) as a guide for image cropping. To make the employed COD approach more adaptive to our scenario, we enhance it by introducing text-based search rankings. We then decide for each image whether it should be cropped and determine the final cropping boundary by expanding the detected bounding box, so that the produced thumbnails are of proper appearances. The experimental results demonstrate that our method can outperform down-sampling and saliency-based methods on both object localization accuracy and general thumbnail quality.

1 Introduction

Web image search, which serves as an important business for most search engines, enables the users to access relevant images using key words, and thumbnailing techniques aim to adapt large quantity of result images into a limited space such as a web page. In this paper, we focus on a special category of queries, by which the users attempt to search for objects such as "BMW Q5", "Pearl Tower" and "Avocado". For these queries, only some parts of the image are relevant to the query. Therefore, they have specific preference for thumbnails that the relevant parts should be emphasized in the thumbnail while the other parts should be either abandoned or suppressed.

Fig. 1. An example of "Sydney opera". (a) thumbnail by down-sampling, (b) region of the opera, (c) thumbnail by cropping.

B. Huet et al. (Eds.): PCM 2013, LNCS 8294, pp. 337–349, 2013.

Uniformly down-sampling the image to thumbnail size has low time consumption but may probably give up object details. For example, Fig. 1.(a) shows a thumbnail of Sydney opera produced by down-sampling. After resizing the object to thumbmail size, the building of Sydney opera only occupies a very small region and can hardly be recognized by the users (see Fig. 1.(b)). Recently, some saliency-based methods [1, 2] propose to find the foreground object of each image and crop the image based on the detected bounding box. However, saliency-based cropping may probably produce incorrect thumbnails. First, saliency fails to indicate the correct object region when the background content is cluttered. We can see from Fig. 2.(a) the detected bounding box using saliency also includes the trees of Champs Elysee. Second, saliency-based methods are query-independent, thus may produce irrelevant thumbnail when the image multiple foreground objects. For example, Fig. 3.(a) shows an image of multiple sport brand logos where the "Adidas" logo is the most salient. Since this image is also relevant to the queries such as "Nike" and "Reebok", saliency-based algorithms still presents the "Adidas" logo as an irrelevant thumbnail.

We argue that the problem is caused by the fact that existing cropping methods only consider the appearance of each single image and do not figure out what the query object is. For object queries, we believe that the query's meaning is reflected by a few object categories underlying the result images, called "Canonical Query Objects". In this paper, the canonical query objects are mined using common object discovery (COD) with the assumption that the relevant object regions are visually consistent. We employ a COD algorithm based on Link-Analysis (LA) due to its time efficiency, which meets the requirements of web image search. Because the algorithm is not specifically designed for result image collection, we modify it by introducing text-based search ranking as a priori so that the algorithm can discover query-relevant object categories from noisy image search results. Then, for each image, we decide whether it needs to be cropped according to image relevance and object appearance. Finally, we determine the cropping boundary for each image by expanding the detected bounding box so as to include some background context that better fits the user's expectation. In convenience, the term "canonical query object" is called "query object" in the following descriptions for short. The proposed approach is regarded as query-dependent because it produces different thumbnails for one image when it is retrieved by different queries. The main contributions of this paper are summarized as follows:

- We propose to discover the underlying canonical query objects as a guide of result image cropping.
- We enhance the common object discovery algorithm in [3] to improve the relevance and completeness of the detected objects.
- We propose to crop the image based on an expanded object bounding box for a proper appearance of thumbnail.

The proposed method is tested on a dataset comprising 50 web queries and 5000 images, where an evaluation of object localization is applied as well as an

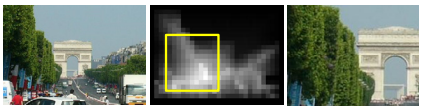

Fig. 2. An example of "Champs Elysee": (a) original image, (b) saliency map with localized foreground object, (c) produced thumbnail

Fig. 3. An example of multiple logos": (a) original image, (b) produced thumbnail in the result list, (c) saliency map with localized object

user study. The experimental results suggest that our method can outperform other baseline methods on the quality of thumbnails.

The rest of this paper is organized as follows: After reviewing the related approaches in Sect. 2, we describe the details of the proposed method in Sect. 3. In Sect. 4, we present two experiments settings and analyze the experimental results. Finally, we conclude this work in Sect. 5 with some remarks on the future work.

2 Related Work

The studied approaches on thumbnails can be generally classified into two categories: resizing-based approaches and cropping-based approaches. Other than uniform down-sampling, advanced resizing-based approaches [4, 5, 6, 7, 8] apply different sampling rates on different parts of the image according to their visual importance such as saliency, which is also known as "image re-targeting". For example, [4] removes eight-connected pixels with the lowest energy in each iteration while in [5] the authors decompose the image into curve-edge grids and optimize their shapes using max-flow-min-cost model. Image re-targeting approaches do not explicitly drop any image content but non-uniform pixel sampling may probably destroy the image structure.

Cropping-based approaches only keep the foreground object in the thumbnail, while the other parts of the image are abandoned. For example, [1] expands the foreground bounding-box by applying a greedy algorithm on the saliency map. Instead of focusing on saliency, [9] proposed an attention driven foreground detection method in localizing foreground objects. Recently, some advanced salient object detection methods [10, 11, 12, 2, 13] succeed to localize foreground objects accurately, and we believe these approaches can also be adopted in producing thumbnails. In [12], the author generates saliency maps by analyzing the frequency domain while [11] integrates multiple psychological evidence in saliency calculation, such as color, contrast, visual frequency, layout and human faces. In [2], the saliency map is calculated by comparing the content inside and outside the hypothesis bounding box, meanwhile in [13], the author generates saliency map using background priori on boundary and connectivity, and calculates saliency using geodesic distance. As mentioned on Sect. 1, saliency-based methods only consider the object appearance in a single image, leading to incorrect thumbnails.

Unsupervised object detection approaches are also known as "Common Object Discovery". According to different forms of output, they can be generally divided into two categories. The approaches in the first category provide pixel level segmentation of objects, called "co-segmentation". For example, in [14] the author simultaneously segments the foreground objects from multiple images using an expansion function analogue to heat spreading while [15] categorizes the pixels from different images through a discriminative clustering framework. The second category represents an image as a bag of ROIs and discovers the objects by finding common ROIs in different bags. For example, in [16] a conditional random field (CRF) is built for all the candidate ROIs and the common object discovery problem is transformed as finding an optimal configuration in the CRF. Guided by salient object detection result, [17] trains the object detection model through bottom-up multi-class learning and discovers object categories using discriminative Expectation-Maximization (EM) framework. In this paper, we adopt an efficient COD approach [3], where the author alternately detects the foreground ROI from each image and builds foreground model by finding representative ROIs using link analysis method. Since this method is not specifically designed for result image collection, we modify it to fit our application.

3 Approach

The proposed thumbnailing approach consists of the following three steps. We first apply the modified common object discovery algorithm to find the canonical query objects from the result image collection. Then, we decide for each image whether it should be cropped. At last, we expand the detected foreground bounding box to include some parts of background.

3.1 Canonical Query Object Discovery

At the very beginning, 100 ROIs with the highest saliency in each image are detected by the salient object detection method proposed in [2]. Each image is thus represented as a bag of ROIs. Since some images contain only foreground, the image itself is also added into the bag as an ROI. The employed COD algorithm is of an alternative optimization framework. In each iteration, the algorithm first finds a set of diverse and representative ROI exemplars. Then, the foreground ROI of each image is refined based on these selected exemplars. The process of exemplar discovery and foreground refinement are alternatively executed until convergence.

ROI Exemplar Discovery. We denote the foreground ROI of the i_{th} image selected in the last step as $r_i^{(t-1)}$, and denote $\boldsymbol{R}^{(t-1)} = \{r_i^{(t-1)}\}$, where t denotes current number of iterations. The initial ROI set $\boldsymbol{R}^0$ is selected as the ROIs presenting the entire image. Our goal of this step is to find N most representative ROIs called "exemplars". Other than simple object detection, the application of thumbnailing requires the exemplars to be as relevant to the query as possible so as to avoid the ambiguity of multiple foregrounds mentioned in Sect. 1.

With $\boldsymbol{R}^{(t-1)}$ and a similarity measure s, we apply Affinity-Propagation (AP) [18] to vote for qualified exemplars. Initially, all $r \in \boldsymbol{R}^{(t-1)}$ are considered as potential exemplars and represented as nodes in the network, and the voting process is done by recursively passing messages between all node pairs. There are two kinds of messages being transmitted over the network. 1) "Responsibility". Let $p(r_i, r_j)$ denote the "responsibility" transmitted from r_i to r_j. $p(r_i, r_j)$ reflects how well-suited for r_j to represent r_i as its exemplar, considering other potential exemplars for r_i. 2) "Availability". Denoting $a(r_i, r_j)$ as the "availability" transmitted from r_j to r_i, it indicates how appropriate for r_i to choose r_j as its exemplar, taking into account other ROIs that may choose r_j as their exemplar. $p(r_i, r_j)$ and $a(r_i, r_j)$ are recursively updated by follows:

$$\begin{aligned} p(r_i, r_j) &\leftarrow s(r_i, r_j) - \max_{j' != j}\{s(r_i, r_{j'}) + a(r_i, r_{j'})\}, \\ a(r_i, r_j) &\leftarrow \min\{0, p(r_i, r_j)\} + \sum_{i' \notin \{i,j\}} \max\{0, p(r_{i'}, r_j)\}. \end{aligned} \tag{1}$$

The "self-availability" is updated as follows:

$$a(r_j, r_j) = \sum_{j' \notin j} \max\{0, p(r_j, r_{j'})\}. \tag{2}$$

In order to improve the relevance of the exemplars, we assume that ROIs from top images are more likely to be exemplars than those from the bottom. In Affinity Propagation, the above assumption can be read as follows: for a pair of ROIs, r_i and r_j, r_j should have higher "Responsibility" to be the exemplar of r_i if r_j's parent image has higher ranking than r_i's. Here, we achieve this assumption by modifying the similarity matrix s. Originally, matrix s is calculated by comparing visual feature vectors of image pairs, such as color histogram and histogram of oriented gradient (HOG) [19]. Therefore, s is a symmetric matrix, $s(r_i, r_j) = s(r_j, r_i)$. To implement the ranking preference, we embed a non-symmetric term s_r to matrix s, according to image rankings. The enhanced similarity matrix s' is defined as follows:

$$\begin{aligned} s'(r_i, r_j) &= s(r_i, r_j) + \eta s_r(r_i, r_j), \\ s_r(r_i, r_j) &= \log(rpos(r_i) + 1) - \log(rpos(r_j) + 1), \end{aligned} \tag{3}$$

where $rpos(r)$ denotes the ranking position of the images where r is from. If r_j got a higher ranking than r_i, thus $s'(r_i, r_j) > s(r_i, r_j)$ and r_j obtains higher "responsibility" to serve as the exemplar of r_i than using the original measure of s.

The above messages keep passing over the network until convergence. Finally the exemplar collection E^t is selected by solving

$$\arg \max_{\substack{R' \subset R^{(t-1)} \\ \|R'\| = N}} \sum_{r \in R'} (p(r, r) + a(r, r)). \tag{4}$$

Foreground ROI Refinement. The goal of this step is to chose an ROI from each image bag, which is regarded as its foreground in the condition of the current

Fig. 4. Two cases where images shall not be cropped: (a) the foreground object occupies most of the image area, (b) there's no recognizable foreground object

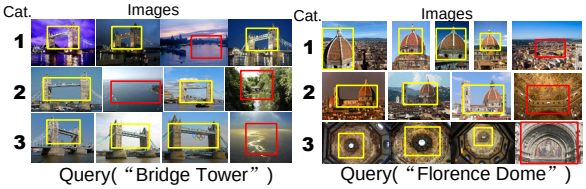

Fig. 5. The object categories from the query "Tower Bridge" and "Florence Dome" discovered by the proposed COD approach, where the correctly localized foreground ROIs are marked yellow. We can see that the localized ROIs of "foreground-less" images are playing as outliers in each category.

exemplars. The chosen foreground ROIs from all the images are then adopted as the input of exemplar selection process of the next iteration. The result of the final iteration is consequently taken as the query-relevant object for each image. In this paper, we follow the ROI selection step described in [3], where for image i, the author constructs an augmented bipartite graph W_i^* between $\boldsymbol{G}^i$ and $\boldsymbol{E}^t$. The matrix form of W_i^* is expressed as follows:

$$W_i^* = \begin{bmatrix} \alpha W_i & (1-\alpha)W_i' \\ W_i' & 0 \end{bmatrix}, \tag{5}$$

where W_i is the self-similarity graph constructed by pairwise comparison on $\boldsymbol{G}^i$, and W_i' is a bipartite graph constructed between $\boldsymbol{G}^i$ and $\boldsymbol{E}^t$. After that, we apply PageRank [20] algorithm on the constructed bipartite graph and obtain a score vector $\boldsymbol{p}$, then the foreground ROI for current iteration is selected by $\arg\max_{g_j \in \boldsymbol{G}^i} \boldsymbol{p}(j)$.

3.2 Cropping Image Selection

In our experiment, we observe that cropping is not necessary for two kinds of images. Motivated by this, we make a decision for each image whether it should be cropped after the query-relevant foreground object is localized. Figure 4.(a) shows some examples of the first kind, whose foreground objects occupy most of the image area and appear as high-quality thumbnails without any cropping. In this paper, we ignore those images in the cropping step whose foreground objects occupy more than $\frac{3}{4}$ of the image size.

The second kind of images are of an opposite situation, in which there's no recognizable foreground object. As shown in Fig 4.(b), these images only present simple colors or textures such as "blue sky" or "green grass". Although the employed COD approach assigns each image with a foreground ROI no matter there is a foreground object or not, we can observe from Fig. 5 that the so-called foreground ROIs of the "foreground-less" are usually playing as outliers within the object collection. To identify these outliers, we first cluster all foreground

ROIs by assigning each ROI to its nearest exemplar and then eliminate the ROIs with low visual densities. The evaluation of visual density is done by applying the PageRank algorithm within each category, where we calculate the self-similarity matrix by pairwise comparing the belong ROIs. Assume we obtain a vector $\boldsymbol{p}^c$ of PageRank scores from categroy c, the outliers of c are selected as $\{i|\boldsymbol{p}^c(i) < \delta \max_j\{\boldsymbol{p}^c(j)\}\}$. In our experiment, δ is set to 0.5.

3.3 Bounding Box Expansion

The second row of Fig. 6 shows some localized objects from the query "Triumphal Arch" and "Big Ben", from which we observe that most of the objects are tightly bounded by the detected bounding boxes. The thumbnails produced by cropping the ROIs straightly are illustrated in the third line of Fig. 6, and obviously, no user could accept such kind of thumbnails because the image content is incomplete without any background context. Besides, the thumbnail looks more attractive when the object is set off by some background. Motivated by this, we first expand the detected bounding box according to a conducted user study and then refine the expanded bounding box if it exceeds the image boundary or it is of an unacceptable aspect ratio.

Study of Expansion Distances. To investigate how much background should be included from the users' perspectives, 500 images of various topics are crawled from the Bing[21] search engine. We invite 5 people including 4 males and 1 female in the study, and each participant is asked to manually crop the images into thumbnails which he thinks are the best. For each image, we calculate the distances between the detected object bounding box and the manually labeled thumbnail boundary, and the average expansion distances of all samples on the four directions are shown in Fig 7. To make it fair for bounding boxes of different sizes, the horizontal and vertical expansion distances are respectively normalized by the widths and heights of the foreground ROIs. According to Fig. 7, expanding the detected bounding boxes by 41% horizontally and 36% vertically is the best expansion strategy from the users' perspectives. The last line of Fig. 6 shows the thumbnails based on the expanded bounding box. We can see that the query objects are of better appearances with the presence of proper background.

Bounding Box Refinement. The expansion distances from the above study are just a rough guide for image cropping. In the following cases, the bounding box should be refined after the expansion so as to produce the correct thumbnail. First, the expanded bounding box sometimes exceeds the image boundary as demonstrated in Fig. 8. Since the query object is away from the image center after cropping, the thumbnail may have a weird appearance. To deal with this problem, we suggest to uniformly scale-down the expanded bounding box until all edges are inside the image boundary, such that the expansion distances on the four directions remain consistent. Second, some of the object ROIs are of very high/low aspect ratios, such as the "Big Ben" towers shown in Fig. 6. In such circumstances, the expanded bounding box would follow the shape of the

tower and produce a narrow thumbnail if the above expansion distances are adopted. To control the non-reasonable expansion, we bring in a threshold σ for the thumbnail aspect ratios, and truncate the expanded bounding box if the aspect ratio is higher than σ or lower than $\frac{1}{\sigma}$. In this paper, σ is empirically set to $\frac{5}{3}$.

4 Experiment

In this section, we first demonstrate an experiment on object localization to test whether our approach can present the correct objects. Then, we present an user study to show the effectiveness of our approach on the quality of produced thumbnails.

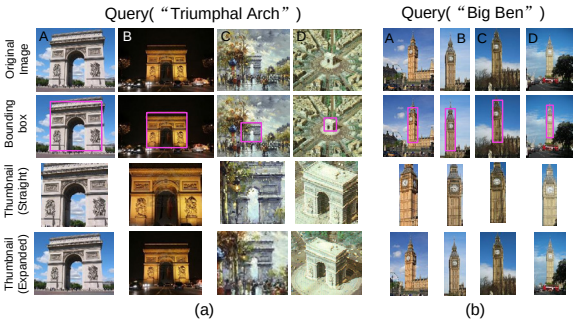

Fig. 6. Two examples of the query "Triumphal Arch" and "Big Ben". Lines from top to bottom: the original images, the detected object bounding boxes, the thumbnails by straight cropping and the thumbnails based on the expanded bounding box.

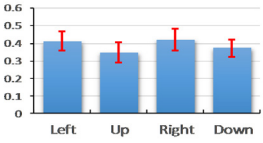

Fig. 7. The average expansion distances on the 4 directions. The corresponding variances are marked red.

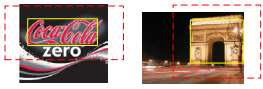

Fig. 8. Two examples when the expanded bounding box exceeds the image boundary

4.1 Dataset and Settings

The proposed approach is tested on 50 queries selected from the "Web Queries" dataset[1] [22], comprising object queries such as landmarks, vehicles, logos and productions. Since "Web Queries" dataset only provides result images of very low resolutions, we re-issue these queries to the Bing search engine [21] and download the top 200 result images. For all the approaches evaluated in this paper, including the proposed one, the visual similarity between the ROIs are calculated using a visual feature called "Pyramid Histogram Of visual Words (PHOW)" [23], which is a histogram of densely sampled visual words. The total dimensionality of this feature is 16000.

[1] The dataset is available at
`http://lear.inrialpes.fr/~krapac/webqueries/webqueries.html`

4.2 Object Localization Evaluation

Evaluation Steps. The motivation of this experiment is to test whether the proposed approach can present the correct objects in the produced thumbnails. Before the evaluation, each image is manually labeled with an object bounding box as the ground-truth of object location. For each image i, the detected bounding box is denoted as D_p^i while D_t^i denotes the ground-truth. The localization accuracy is calculated based on the overlapping area of the two bounding boxes, expressed as follows:

$$Accuracy = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{B}} \frac{Area(D_p^i \cap D_t^i)}{Area(D_p^i \cup D_t^i)}, \tag{6}$$

where $\mathcal{C}$ stands for the result image collection.

The proposed is compared with 3 baseline methods:

- ***FREQ***[12] and ***SAL_COMP***[2]. These two baselines are typical methods in salient object detection, in which *SAL_COMP* serves as the state-of-the-art. They are regarded as the representation of query-independent methods.
- ***LA***[3]. This is the original version of the employed COD approach. It is adopted to show the effectiveness of the modification we made in this paper.

Towards the query dependency problem illustrated in Fig 3, we pick up a subset of 120 images from the dataset with multiple distinct foreground objects. For these images, only the locations of the query-relevant objects are labeled as ground-truth.

Result Analysis. The localization accuracies of the 3 evaluated methods are shown in Table 1. We can observe from Table 1 that all COD-based methods achieve higher localization accuracy than saliency based methods. Especially for those images with multiple objects, the improvement is more obvious. For example, the basic *LA* method outperforms *SAL_COMP* by 6.53% while the proposed approach outperforms it by 19.19%. It suggests that the problems mentioned in Sect. 1 can be better handled by mining the canonical query objects. We can also observe that the proposed methods outperforms *SAL_COMP* by 8.64%, implying that the foreground objects can be more accurately localized with the help of text-based rankings.

Table 1. The localization accuracies of the proposed method and some baseline methods

Accuracy	*FREQ*[12]	*SAL_COMP*[2]	*LA*[3]	*Proposed*
All	41.11%	46.43%	52.96%	60.60%
Multi-Foregrounds	37.96%	39.93%	45.42%	59.12%

4.3 User Study

Evaluation Steps. In order to evaluate the proposed approach using quantified image quality, we invite 9 participants to take part in our user study, including 6 males and 3 females with various professional backgrounds. Due to the workload limit of the participants, the participants are only asked to rate the top 50 images in the search result list. The thumbnails produced by 4 different approaches are compared, including 1) uniform down-sampling (*DW_SAM*), 2) *SAL_COMP*[2], 3) *LA* and 4) the proposed approach. To make it close to a real search engine, all thumbnails are scaled to a height of 120 pixels. In the user study, each image is labeled into 3 quality levels: **Good**, **Fair**, **Bad**, according to the users' own perspectives. After above process, the ratings from the users on each level are accumulated and the proportion of each quality level by the different approaches are shown in Fig. 9.

Result Analysis. From Fig. 9, we can see that the proposed approach produces 38.84% more good thumbnails than down-sampling and 11.79% than *SAL_COMP*. It indicates that mining the canonical objects is generally effective in producing high-quality thumbnails. However, our approach produces 2.2% more bad thumbnails than down-sampling. These low-quality thumbnails should be regarded as the risk of adopting automatic cropping techniques. Once the foreground object of an image is incorrectly detected, the thumbnail probably appears to be content-less. On the other hand, down-sampling trends to produce thumbnails with small foreground objects thus obtains large quantity of "Fair" ratings. We are still glad to see that our approach produces 3.82% less low-quality thumbnails than saliency-based cropping because of the raise on localization accuracy.

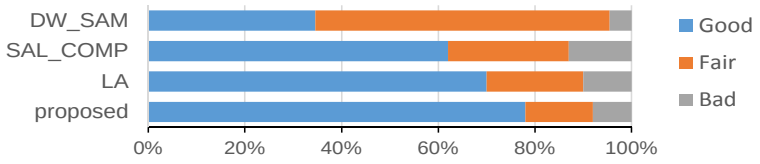

Fig. 9. The comparison of the five evaluated approaches on user ratings: 1) uniform down-sampling, 2) *SAL_COMP*, 3) *RNK_PREF*, 4) *SZ_PREF*, 5) proposed method

Fig. 10 shows a dozen of thumbnail examples produced by down-sampling, *SAL_COMP* and the proposed method. We can observe that *SAL_COMP* detects the incorrect foreground ROIs in many images, while the proposed method finds the right ones. For example, in image **A** of "Pizza Tower" and image **D** of "Arsenal Logo", *SAL_COMP* presents incorrect foreground objects while in image **A** of "Forbidden City" and image **C** of "Hollywood Sign", it presents some

[2] To make it fair for *SAL_COMP*, the detected bounding boxes are also expanded by our expansion method.

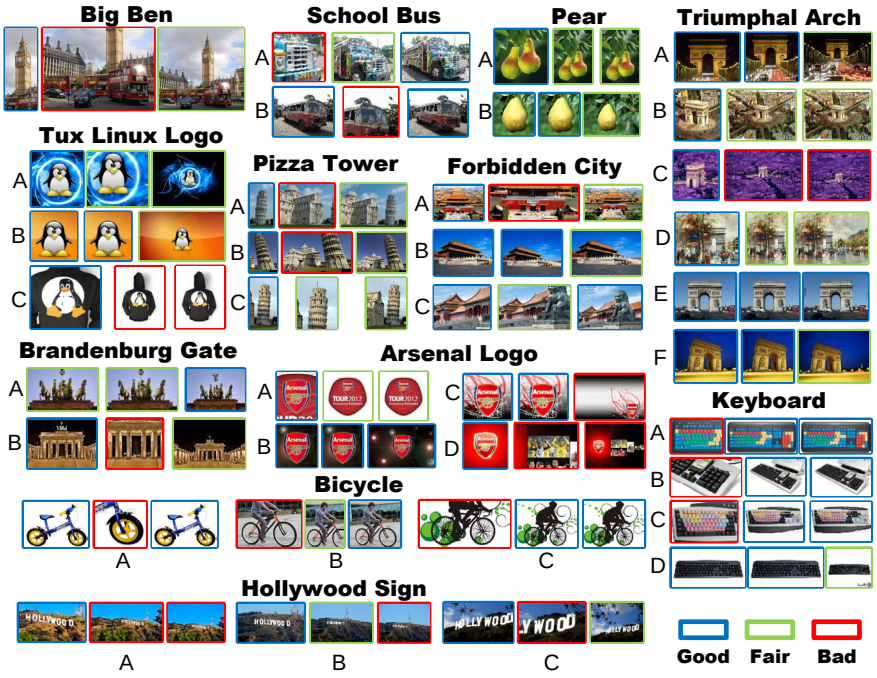

Fig. 10. Comparison of thumbnails produced by the three approaches. Each group from left to right: 1) proposed method, 2) *SAL_COMP*, 3) uniform down-sampling.

incomplete object parts. Our approach performs the worst of all in the query "Keyboard" By exploring the result image collection, we find that none of the keyboards are of a consistent gesture, causing the COD approach fail to discovery correct canonical query objects. For image **B** and **C** in the query "Bicycle", all participants give "bad" ratings to our approach, because it only discovers the region of bikes and ignores the people riding them. Although the cyclists on the bikes are not relevant to the query, all users insist them to be kept in the thumbnails. This problems is rare but typical, we decide to investigate it in our following research.

5 Conclusion and Future Work

This paper presented a novel thumbnailing approach for web image search. Motivated by the fact that saliency-based cropping can not always present correct objects in thumbnails, we suggested to mine the canonical query objects via common object discovery and use them as a guide for image cropping. We then brought in two measurements to test whether cropping is necessary for each image and proposed to obtain the cropping boundary by expanding the detected bounding box so as to include some background. The evaluation on

object localization suggested that the proposed method is more likely to detect correct objects than the baseline methods while the user study reported that our approach can provide the higher thumbnail quality. Our future work will concentrate on following two directions: First, the paper only presents a simple strategy for bounding box expansion. In the future, we are going to consider some other information such as image layout and object composition. Second, we need to further improve the efficiency so that it can be integrated into an existing search engine as an online component.

Acknowledgement. This work is supported by the NSFC of China (No. 61021062), the 863 Program of China (No. 2011AA01A202) and National Special Fund (No 2011ZX05035-004-004HZ).

References

[1] Suh, B., Ling, H., Bederson, B., Jacobs, D.: Automatic thumbnail cropping and its effectiveness. In: ACM Symposium on User Interface Software and Technology, pp. 95–104. ACM (2003)
[2] Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J.: Salient object detection by composition. In: ICCV, pp. 1028–1035. IEEE (2011)
[3] Kim, G., Torralba, A.: Unsupervised detection of regions of interest using iterative link analysis. In: NIPS (2009)
[4] Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. In: TOG, vol. 26, p. 10. ACM (2007)
[5] Ren, T., Liu, Y., Wu, G.: Rapid image retargeting based on curve-edge grid representation. In: ICIP, pp. 869–872 (2010)
[6] Ren, T., Liu, Y., Wu, G.: Image retargeting based on global energy optimization. In: ICME, pp. 406–409. IEEE (2009)
[7] Ren, T., Liu, Y., Wu, G.: Image retargeting using multi-map constrained region warping. In: ACM Multimedia, pp. 853–856. ACM (2009)
[8] Liu, F., Gleicher, M.: Automatic image retargeting with fisheye-view warping. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, pp. 153–162. ACM (2005)
[9] Amrutha, I., Shylaja, S., Natarajan, S., Murthy, K.: A smart automatic thumbnail cropping based on attention driven regions of interest extraction. In: ICIS, pp. 957–962. ACM (2009)
[10] Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: CVPR, pp. 73–80. IEEE (2010)
[11] Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. PAMI 34(10), 1915–1926 (2012)
[12] Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR, pp. 1597–1604. IEEE (2009)
[13] Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
[14] Kim, G., Xing, E., Fei-Fei, L., Kanade, T.: Distributed cosegmentation via submodular optimization on anisotropic diffusion. In: ICCV, pp. 169–176 (2011)

[15] Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image co-segmentation. In: CVPR, pp. 1943–1950. IEEE (2010)
[16] Deselaers, T., Alexe, B., Ferrari, V.: Localizing objects while learning their appearance. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 452–466. Springer, Heidelberg (2010)
[17] Zhu, J.Y., Wu, J., Wei, Y., Chang, E., Tu, Z.: Unsupervised object class discovery via saliency-guided multiple class learning. In: CVPR, pp. 3218–3225. IEEE (2012)
[18] Frey, B., Dueck, D.: Clustering by passing messages between data points. Science 315(5814), 972–976 (2007)
[19] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 1, pp. 886–893. IEEE (2005)
[20] Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30(1-7), 107–117 (1998)
[21] Bing, `http://www.bing.com`
[22] Krapac, J., Allan, M., Verbeek, J., Juried, F.: Improving web image search results using query-relative classifiers. In: CVPR, pp. 1094–1101 (2010)
[23] Bosch, A., Zisserman, A., Muoz, X.: Image classification using random forests and ferns. In: CVPR, pp. 1–8 (2007)