Noname manuscript No.
(will be inserted by the editor)

Scale and Object Aware Image Thumbnailing

Jin Sun · Haibin Ling

Received: date / Accepted: date

Abstract In this paper we study effective approaches to create thumbnails from input images. Since a thumbnail will eventually be presented to and perceived by a human visual system, a thumbnailing algorithm should consider several important issues in the process including thumbnail scale, object completeness and local structure smoothness. To address these issues, we propose a new thumbnailing framework named *Scale and Object Aware Thumbnailing* (SOAT), which contains two components focusing respectively on saliency measure and thumbnail warping/cropping. The first component, named *Scale and Object Aware Saliency* (SOAS), models the human perception of thumbnails using visual acuity theory, which takes thumbnail scale into consideration. In addition, the "objectness" measurement (Alexe et al) is integrated in SOAS, as to preserve object completeness. The second component uses SOAS to guide the thumbnailing based on either retargeting or cropping. The retargeting version uses the Thin-Plate-Spline (TPS) warping for preserving structure smoothness. An extended seam carving algorithm is developed to sample control points used for TPS model estimation. The cropping version searches a cropping window that balances the spatial efficiency and SOAS-based content preservation.

The proposed algorithms were evaluated in three experiments: a quantitative user study to evaluate thumbnail browsing efficiency, a quantitative user study for subject preference, and a qualitative study on the RetargetMe dataset. In all studies, SOAT demonstrated promising performances in comparison with state-of-the-art algorithms.

Jin Sun · Haibin Ling
Dept. of Computer & Information Sciences, Temple University
1805 N Broad St., Philadelphia, PA
Tel.: +1-215-204-6973
Fax: +1-215-204-5082
E-mail: jinsun@cs.umd.edu, hbling@temple.edu

Keywords Image Thumbnailing · Image Retargeting · Thumbnail Cropping · Contrast Sensitivity Function · Seam Carving · Thin Plate Spline

1 Introduction

With the increasing popularity of image capturing and displaying devices, effective ways for presenting and browsing image datasets are drawing a significant amount of new research attention. In image browsing, tiny thumbnails provide the basic function for a user to quickly explore an image dataset visually, such as personal photo albums or scientific image collections. In this paper we use the term *image thumbnailing* to indicate the process of creating a thumbnail from an input image. A straightforward thumbnailing approach is to simply shrink the original image. Such a solution, despite being widely used in image browsing and management systems, has been shown to be less effective than smarter solutions such as thumbnail cropping (Chen et al 2003; Lam and Baudisch 2005; Suh et al 2003) and retargeting (Liu and Gleicher 2005; Avidan and Shamir 2007). This is particularly true for tiny thumbnails displayed on devices with small sized screens just as a smart phone.

Thumbnailing can be viewed as a special case of image resizing, where the basic philosophy is to preserve important content as much as possible while changing the image size. Many image resizing methods generate aesthetically impressive results when the target image size is comparable to size of the original image (Rubinstein et al 2010b). By contrast, insufficient attention has been paid to image thumbnailing scenarios where the target image is much smaller than the input one. Several important issues need to be addressed by an effective thumbnailing algorithm:

- *Thumbnail scales*. In a browsing task, thumbnails usually have much smaller scales/sizes than the origi-

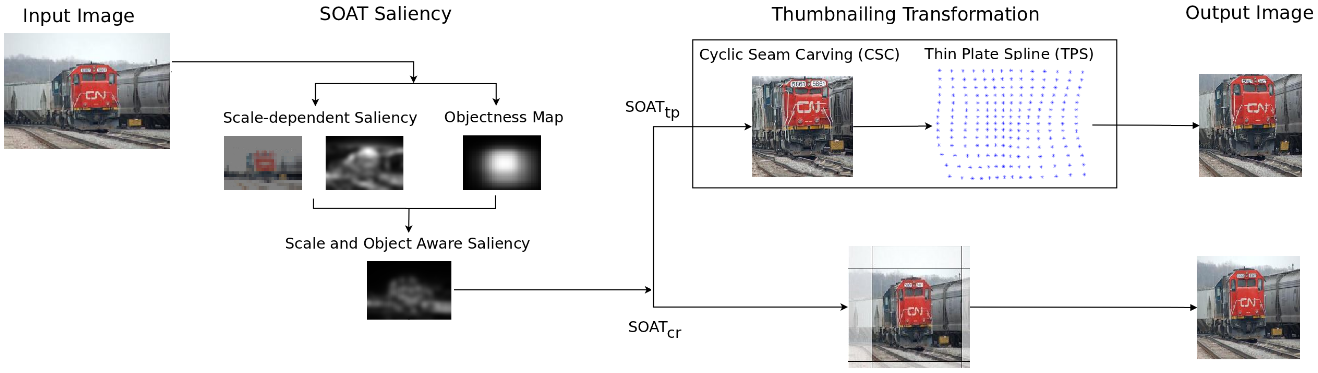


Fig. 1 Flowchart of the proposed method. SOAT contains two main components: (1) *Scale and Object Aware Saliency* (SOAS) and (2) thumbnailing transformation. Given an input image, SOAS is calculated as a measurement of pixel-level importance. It guides the thumbnailing transformation method (either SOAT$_{\mathrm{tp}}$or SOAT$_{\mathrm{cr}}$) to create a thumbnail output image. See text for details.

nal images. Studies have shown that image scales can significantly affect the human visual perception process (Judd et al 2011; Mannos and Sakrison 1974; Van Nes and Bouman 1967). Such scale information is under-explored in existing retargeting algorithms.

- *Object completeness*. Preserving completeness of objects is crucial yet challenging in thumbnailing. A major difficulty lies in the explicit measurement of object completeness. Consequently, many retargeting methods address this issue implicitly by preserving low-level image information such as gradients. The object-level completeness, by contrast, is seldom considered.
- *Structure smoothness*. The contamination of structure smoothness caused by pixel-removal type of retargeting methods usually creates only minor visual artifacts when the retargeted images are relatively large. This is unfortunately not true for thumbnails: removing a large amount of pixels often creates serious image structure discontinuities which human visual systems are not comfortable with.

To address these issues, we propose a new image thumbnailing framework, *Scale and Object Aware Thumbnailing* (SOAT) as shown in Figure 1. SOAT contains two main components: (1) *Scale and Object Aware Saliency* (SOAS) and (2) thumbnailing transformation. As a measurement of pixel-level importance, SOAS guides the thumbnailing transformation to create a thumbnail from the input image.

SOAS. We propose a novel scale-dependent saliency to encode information of thumbnail scales. Inspired by the study in visual acuity (Mannos and Sakrison 1974; Van Nes and Bouman 1967), we calculate such saliency on the image perceived by the human vision system when the original image is observed. Specifically, the perceived image is estimated by using a scale-dependent contrast sensitivity function to filter out high frequency structures that are difficult to perceive at the smaller thumbnail scales. The resultant scale-dependent saliency is further enhanced by preserving object-level object completeness, which is measured by the recently proposed *objectness* (Alexe et al 2012). In particular, the saliency values are tuned down in regions of weak objectness, e.g., backgrounds.

Thumbnailing transformation. We propose two SOAT algorithms based on image retargeting and thumbnail cropping respectively. The retargeting method we proposed, named SOAT$_{\mathrm{tp}}$, uses the *Thin-Plate-Spline* (TPS) (Bookstein 1989) model to balance warping smoothness and matching accuracy. The control points for TPS model estimation are carefully traced and sampled from an extended *Seam Carving* (Avidan and Shamir 2007) algorithm guided by SOAS. Alternatively, the cropping method we proposed, named SOAT$_{\mathrm{cr}}$, extends the previously proposed thumbnail cropping algorithm (Suh et al 2003) by using SOAS as the saliency measure.

We evaluated the proposed SOAT algorithms both quantitatively and qualitatively. The quantitative study is of higher priority as a human's perception of image thumbnails can be very subjective. For this purpose, we have designed user studies with two tasks: 1) *Image browsing task*. Each user was asked to search a target thumbnail from a screen fully tiled by thumbnails. The time cost and accuracy of the searching procedure were recorded and analyzed statistically using ANOVA and Tukey's significance test. The results show that our methods compare favorably in both efficiency and accuracy with previously proposed methods. 2) *Subjective preference task*. Each user was asked to pick a preferred thumbnail from a pair of thumbnails generated by different methods. The preference ratio for method comparison pairs were recorded and analyzed statistically using t-test. The results show that SOAT$_{\mathrm{cr}}$ is preferred against other methods.

For the qualitative study, we applied SOAT to the RetargetMe dataset (Rubinstein et al 2010a) on which the results from many state-of-the-art algorithms have been previously collected and made public. The study shows that, our methods, despite emphasizing on thumbnail browsing effectiveness rather than aesthetic effects, generate images that are visually as good as previously reported results.

In the rest of the paper, we first summarize related work in Sec. 2. Then, we introduce the proposed SOAT framework in Sec. 3, followed by the scale and object aware saliency in Sec. 4 and the two SOAT algorithms in Sections 5 and 6, respectively. The experiments are described and discussed in Sec. 7. Finally, conclusion is drawn in Sec. 8.

2 Related work

2.1 Image retargeting and thumbnail cropping

Automatic image thumbnailing can be viewed as a special case of content aware image resizing: reducing the size of an input image to generate a much smaller thumbnail. Previous studies can be roughly divided into two classes: image retargeting and thumbnail cropping.

Image retargeting (Liu and Gleicher 2005) has been actively studied fairly recently. A comparative evaluation of existing retargeting algorithms can be found in (Rubinstein et al 2010b). They can be further sorted into two categories: discrete methods that remove unimportant pixels and continuous methods that reorganize image content with a continuous warping function.

A typical example of discrete image retargeting methods is the popular *Seam Carving* (SC) proposed by Avidan and Shamir (2007). The idea is to selectively and iteratively remove continuous pixel segments, termed *seams*, while preserving the image structure as much as possible. Several notable works have extended and improved the original SC algorithm. Grundmann et al (2010) introduced discontinuous seams in spatial-temporal space for video retargeting task. Mansfield et al (2010) defined scene consistency and used user-provided image depth map to guide seam carving. Rubinstein et al (2008) presented an improved version of SC with a new type of energy criterion to remove seams that introduce the least amount of energy into the retargeted result. There are other discrete retargeting methods such as (Simakov et al 2008), which uses a bi-directional similarity to guide the retargeting process.

Continuous methods warp the original image to the target image through a continuous transformation, driven by image content. Liu and Gleicher (2005) used non-linear fisheye-view warping that emphasizes parts of an image. Guo et al (2009) constructed a mesh-based image representation and obtained the retargeting result by finding a homomorphous target mesh with the desired size. Karni et al (2009) presented an energy minimization-based shape deformation for image resizing. An adaptive image and video retargeting algorithm was proposed by Kim et al (2009). It solves a constrained optimization problem on strip-based scaling and distortions formulated in the frequency domain. Similarly, Wolf et al (2007) conducted transformations to shrink unimportant regions more than important ones with multiple criteria including local saliency, motion detection and object detection. Wu et al (2010) analyzed image semantics such that the main content of an image could be summarized. Ding et al (2011) pursued the trade-off between keeping important parts and reducing visual distortions by using importance filtering. Niu et al (2012) proposed using non-homogeneous warping by restricting distortion in weakly noticeable regions.

Thumbnail cropping algorithms have been shown to be effective for thumbnail browsing and recognition (Suh et al 2003). These algorithms usually first estimate the spatial distribution of saliency or importance and then find a cropping window that best balances the content preservation and window size efficiency. The main challenges are: (1) how to define the saliency, and (2) how to find the "best" cropping window. In (Suh et al 2003), the classic saliency map (Itti et al 1998) is combined with a greedy window searching algorithm for automatic thumbnail cropping. The similar idea was also explored in (Chen et al 2003). Lam and Baudisch (2005) developed a technique to display web pages on small screen devices with the combination of thumbnails and readable text. Luo et al (2010) proposed a searching strategy to find the maximum saliency density. In (Marchesotti et al 2009), a visual saliency detection mechanism based on an annotated image database was applied to thumbnailing. Learning-based thumbnail cropping has been explored in (Li and Ling 2009; Kennedy et al 2011). Cropping and warping based approaches have also been used for video summarization such as in (Wang et al 2010; El-Alfy et al 2007).

Thumbnail cropping can be viewed as a special case of the discrete retargeting method: it removes all pixels outside the cropping window while keeping the internal ones untouched. It is worth mentioning that in the evaluation of retargeting algorithms by Rubinstein et al (2010b), manual cropping outperforms all other retargeting algorithms in terms of user preferences. Aside from the above methods, Rubinstein et al (2009) presented a multi-operator approach that smartly combines several retargeting and cropping operations for image resizing.

Our method is different from these previous studies mainly in several aspects. First, we encode scale information explicitly with a perception model simulating the thumbnail observation process. Second, we model object completeness by the objectness measurement (Alexe et al 2012). Third, we use the TPS model as the warping function in the retargeting-based thumbnailing. The proposed SOAT$_{\mathrm{tp}}$ algorithm can also be viewed as a combination of continuous and discrete retargeting schemes where extended seam carving is used to guide TPS model estimation. Fourth, we introduce the scale and object aware saliency (SOAS) to both retargeting and cropping for thumbnailing. As shown in the experiments (Section 7), all components contribute to the effectiveness of the proposed SOAT framework.

A preliminary version of SOAT_{tp} appeared in ICCV'11 (Sun and Ling 2011). In this paper, we improve SOAT_{tp} in not only the saliency computation (frequency domain filtering) but also the warping function (better control points sampling for TPS). Furthermore, we present a new thumbnail cropping approach, SOAT_{cr}, which in fact performs the best in the user study. Last but not least, the evaluation in this paper is more thorough by involving more evaluation criteria, methods, user studies and subjects.

2.2 Saliency calculation

The scale dependent saliency in our method is motivated by the studies of human visual acuity (Mannos and Sakrison 1974; Van Nes and Bouman 1967; Peli 2001), which describes how well human can distinguish visual stimuli at various spatial frequencies. Given the image size and an observation distance, the theory can be used to determine the minimum scale at which an image structure is perceivable. In particular, we apply the contrast sensitive function to the input image as a frequency filter to inhibit high frequency components, which reflects tiny structures that are barely visible at the smaller scales. In addition, a recent study (Judd et al 2011) has shown that image scales do affect visual fixations, which support our advocation of explicitly modeling scale information for thumbnailing.

Our study is also related to the multi-scale saliency computation proposed by Liu and Gleicher (2006), which constructs a *scale-invariant* saliency map from an image, segments the image into regions, and enhances the saliency map with region information. Our proposed saliency, SOAS, is different in several aspects: 1) SOAS calculates saliency map in a single scale rather than a scale pyramid. This scale is determined by the viewing distance, the thumbnail size and the display resolution. This renders the proposed method *scale-aware*. 2) SOAS uses objectness as object localization measurement rather than the image segmentation used in (Liu and Gleicher 2006). 3) SOAS obtains low-level saliency information from multiple channels (colors, intensity and orientations) with the same nature of (Itti et al 1998). By contrast, in (Liu and Gleicher 2006) the pixel level saliency is calculated on neighborhood color difference. In addition to (Liu and Gleicher 2006), our saliency also shares some philosophy with the work on perceptual scale space (Wang and Zhu 2008).

Another important component used in our saliency computation is the objectness measurement proposed in (Alexe et al 2012), which estimates the likelihood that a given bounding window contains an object. A short description of objectness is given in Sec. 4.2.

3 Overview

3.1 Problem Formulation

Let I be an input image of size $m_0 \times n_0$ and $S(I)$ be its saliency map[1], such that $S(I)$ is an $m_0 \times n_0$ matrix. We formulate the thumbnailing process as a thumbnailing transform $\mathcal{T}$:

$$J = \mathcal{T}(I, S(I)) \ , \tag{1}$$

where J is the result of size $m_1 \times n_1$ such that $m_1 \ll m_0$ and $n_1 \ll n_0$. In this formulation, a thumbnailing method is characterized by two components: the saliency computation $S(.)$ and the thumbnail transformation $\mathcal{T}(.)$.

Note that the function $\mathcal{T}(.)$ can be either continuous or discrete. For example, the *seam carving* (SC) algorithm (Avidan and Shamir 2007), when carving a vertical seam $\{(i, seam(i)), 1 \leq i \leq m_0\}$, can be defined as

$$J(i,j) = \mathcal{T}^{sc}_{i,j} = \begin{cases} I(i,j), & \text{if } j < seam(i) \\ I(i,j+1), & \text{if } j \geq seam(i) \end{cases} \ , \tag{2}$$

where $i \in [1, m_0]$ and $j \in [1, n_0]$ are row and column indices respectively; $seam(i)$ indicates the seam position at row i calculated by the seam searching strategy to minimize the carving energy defined over saliency $S(I)$. Horizontal seam removal can be defined similarly. SC employs a discrete assignment because the pixels along the path of seams, i.e. $\{I(i, seam(i))\}_{i=0}^{m_0}$, are eliminated and the information they carried is discarded.

Our goal is to design an image thumbnailing algorithm in the context of tiny thumbnail browsing task. A naive solution is to directly resize the original image towards the thumbnail size. This is however too aggressive since thumbnails are usually much smaller than the original images. Instead, we first use the proposed thumbnailing methods to reduce I to an intermediate image J of a smaller yet comparable size, and then shrink J to create the final thumbnail. For retargeting-based thumbnailing, i.e. SOAT_{tp}, we fix the size of J as 20% of that of I; while for cropping-based thumbnailing, i.e. SOAT_{cr}, we use the cropping results as J. For conciseness, in the rest of the paper we treat image J as the final result by ignoring the trivial shrinking step.

[1] We use saliency to indicate the importance measurement used in general, which is not limited to the visual attention-based saliency.

3.2 Framework Overview

To achieve the above goal, we propose a new image thumbnailing framework, named *Scale and Object Aware Thumbnailing* (SOAT) and denoted as $\mathcal{T}^{so}$, with a novel saliency, named *Scale and Object Aware Saliency* (SOAS) and denoted as S^{so}, which captures the scale and objectness information.

Given an input image I and the target size of the output thumbnail, SOAT generates a thumbnail $J = \mathcal{T}^{so}(I, S^{so}(I))$ in two steps: saliency calculation and thumbnailing transformation. The flowchart of SOAT is summarized in Figure 1.

In the saliency calculation step, we address issues of thumbnail scale and object completeness. Inspired by the study of visual acuity in human vision, we propose the scale dependent saliency, denoted as $S^{scale}(I)$, to reflect the relative scale variation between the original image and the image thumbnail. $S^{scale}(I)$ is further augmented by combining with object completeness. The resulting *scale and object aware saliency* $S^{so}(I)$ then guides thumbnailing transformation, i.e., the second step.

In the thumbnailing transformation step, we investigate two different approaches based on retargeting and cropping respectively. The retargeting approach, named SOAT$_{\mathrm{tp}}$, uses the *thin-plate-spline* (TPS) model as the warping function and uses S^{so} to estimate the TPS parameters. The cropping approach, named SOAT$_{\mathrm{cr}}$, uses S^{so} to guide cropping window searches with a greedy algorithm.

In the following sections, we detail each component of SOAT separately.

4 Scale and Object Aware Saliency

4.1 Scale-dependent Saliency

When a thumbnail is presented on a digital display device to an observer, three different images are involved in the visual perception process: the original image of size s_o in pixels, the displayed images of size s_d in inches, and the perceived images on the retina of size s_p. The relationship between these sizes are bonded by two distances: the distance D between the eye of the observer and the display device and the distance D_{rp} between the human retina and pupil. Figure 2 illustrates the relations between the these variables, which is summarized below:

$$s_d = \frac{s_o}{\delta}, \quad s_p = \frac{s_d \cdot D_{rp}}{D} \;, \tag{3}$$

where δ is the screen resolution in DPI (Dots Per Inch).

Since a thumbnail is eventually presented to and perceived by the human visual system, it is critical to explore how well the system preserves the image information. In particular, we want to know which regions of the image become indistinguishable at the target thumbnail scale. This has been studied in psychology and vision sciences in terms of *Contrast Sensitivity Function* (CSF) (Mannos and Sakrison 1974; Van Nes and Bouman 1967; Peli 2001). According to the study, not all patterns in an image are recognizable by humans. The perceived image, denoted as I_p, is determined by both the frequency of the stimuli and its contrast.

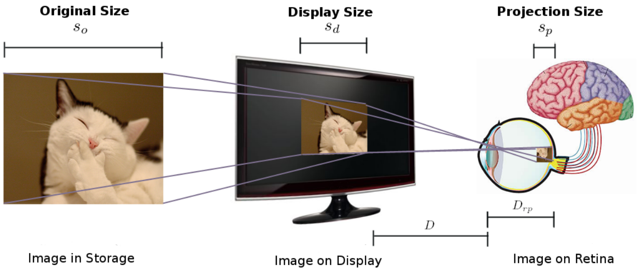


Fig. 2 Demonstration of image sizes in different stages.

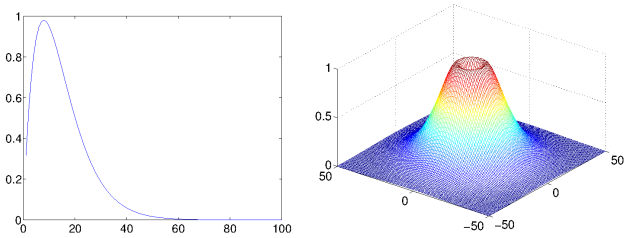


Fig. 3 Left: a 1D CSF curve. Right: a 2D CSF filter.

We apply the following CSF defined in (Mannos and Sakrison 1974) as a filter in the frequency domain.

$$\mathrm{CSF}(\nu) = 2.6 \cdot (0.0192 + 0.114 \cdot \nu) \cdot \exp(-(0.114 \cdot \nu)^{1.1}), \tag{4}$$

where ν is the stimulus frequency in CPD (Cycles Per Degree). An illustration of CSFs is given in Figure 3.

Since the center of the frequency domain is the DC component with frequency zero, we maintain whatever value it is in the filtering process. Furthermore, since the domain of CSF is defined in CPD with a visual angle of one degree, we need to adjust the filter according to the particular visual angle where the image is perceived. Specifically, when an observer looks at an image with of s_o pixels, sitting D inches away in front of the monitor of resolution δ, the corresponding CSF is calculated as:

$$\mathrm{CSF}^*(\nu) = \mathrm{CSF}\left(\frac{\nu}{2 \cdot \arctan\left(s_o/(2\delta D)\right)}\right) \;. \tag{5}$$

Finally, the perceived image I_p can be obtained by:

$$I_p = \mathcal{F}^{-1}(\mathrm{CSF}^* \otimes \mathcal{F}(I)), \tag{6}$$

where $\otimes$ stands for convolution and $\mathcal{F}/\mathcal{F}^{-1}$ are the Fourier transform pair.

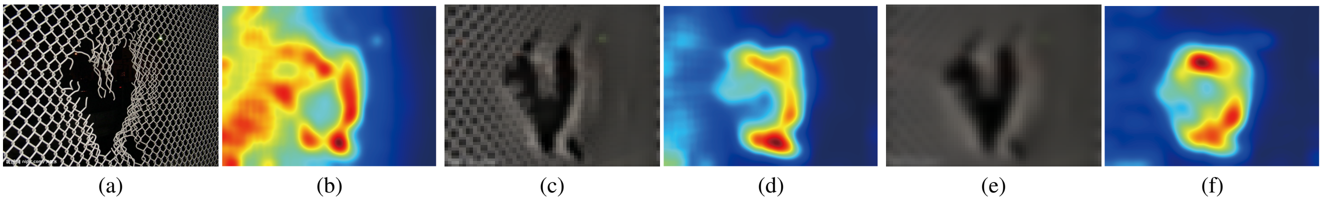


Fig. 4 Scale dependent saliency. (a-b) An input image and its saliency map. (c-d) The image filtered by the CSF filter with viewing distance of one meter and corresponding scale dependent saliency. (e-f) The image filtered by the CSF filter with viewing distance of two meters and corresponding scale dependent saliency.

According to Eqn. 4, for a small display size (thumbnail size in our case), a pixel may become indistinguishable from its neighbors to a human observer. Consequently, an image structure that is salient in the original image may not appear salient to a human observer when the structure is displayed in a small thumbnail. Inspired by this observation, we propose using scale dependent saliency to encode the scale information in the final thumbnails.

The calculation of the scale-dependent saliency goes as follows:

- First, the original image I is resized homogeneously into the target thumbnail size[2], i.e. the display size, 60×60 pixels in our experiment. Denote the resized image as I_0.
- Then, the perceived image I_p is derived from I_0 according to Eqn. 6. Specifically, in our experiment we use a monitor of 1680×1050@65hz and 120 DPI, and all subjects are requested to sit 0.5 meter away from the monitor.
- Finally, the scale dependent saliency $S^{scale}(I)$ is defined as

$$S^{scale}(I) = \uparrow (S(I_p)) \ , \tag{7}$$

where $\uparrow$ denotes upsampling and $S(.)$ is a saliency measure, which in our study is defined according to the visual attention model in (Itti et al 1998).

Figure 4 illustrates the effect of scale dependent saliency. When the image scale is large enough (a–b), fine details in the image are perceivable thus the saliency map shows significant responses in such detail regions, e.g., the texture of the barb wire fence. However, when the image scale becomes smaller (or view from further distance), details fade out while the main structures survive (c–d). This effect is more obvious when the image scale gets even smaller (e–f). The scale dependent saliency successfully identifies the important structures of the image (e.g., the heart shape) by assigning high saliency values to them. Meanwhile, low saliency values are given to regions of over-fined details. More examples of scale dependent saliency are shown in Figure 5.

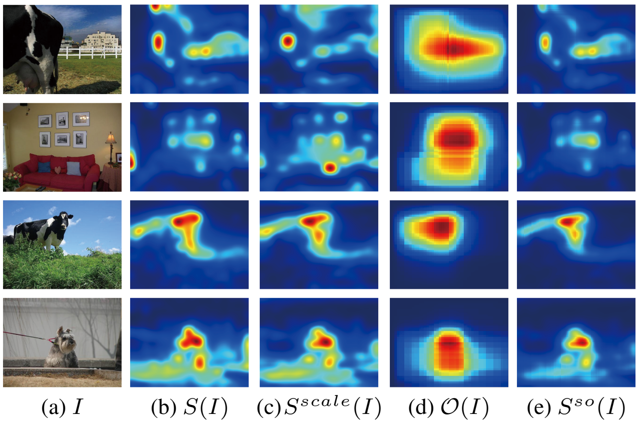


Fig. 5 Saliency computation. (a) Input image I. (b) Original saliency (Itti et al 1998) $S(I)$. (c) Scale dependent saliency $S^{scale}(I)$. (d) Objectness map (Alexe et al 2012) $\mathcal{O}(I)$. (e) Scale and object aware saliency $S^{so}(I)$.

The proposed scale dependent saliency is modified from the one in our preliminary study (Sun and Ling 2011). Both approaches are biologically plausible: the new one works in the frequency domain while the old one in the spatial domain. By working in the frequency domain, the new version gains some robustness to local noises, which could mislead the saliency computation and kill important structures due to shrinking artifacts. In general the two approaches generate similar saliency patterns. Some examples on which they disagree are shown in Figure 6, which illustrates that the new version picks more local structures than so does the old version.

4.2 Scale and Object Aware Saliency

It is desirable for a thumbnailing algorithm to preserve object completeness when removing or distorting parts of an input image. One difficulty lies in explicitly defining such completeness. Recently, Alexe et al (Alexe et al 2012) pro-

[2] Rigorously speaking, we should use the resulting image J instead of I to shrink into the thumbnail size. This however requests the size of J to be known beforehand which may not be true for some thumbnailing algorithms. In addition, though smaller than I, J is still much larger than the final thumbnail. Therefore the approximation using I does not bring significant difference in practice.

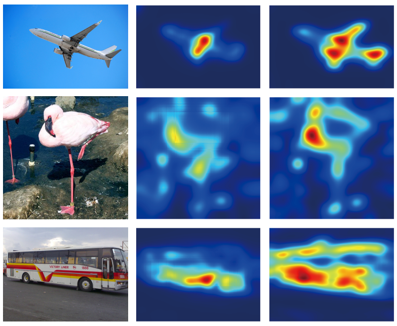


Fig. 6 Scale dependent saliency map comparison. Left: input image I. Middle: scale dependent saliency in (Sun and Ling 2011). Right: proposed scale dependent saliency.

posed a novel *objectness* measure, which is trained to distinguish object windows from background ones. The objectness measure combines several image cues, such as multiscale saliency, color contrast, edge density and superpixel straddling, in order to predict the likelihood of a given window containing an object. Specifically, for a rectangular window $\mathbf{w} = (x, y, w, h)$ with a top-left corner at (x, y), width w and height h, its objectness is defined as the likelihood $\mathrm{Pr}_{obj}(\mathbf{w})$ that $\mathbf{w}$ contains an object.

To get the objectness distribution over pixels, we first sample n_w windows $W = \{\mathbf{w}_i\}_{i=1}^{n_w}$ for an input image and then calculate the *objectness map* $\mathcal{O}$ as the expected objectness response at each pixel,

$$\mathcal{O}(i,j) = \frac{1}{\Gamma} \sum_{\mathbf{w} \in W \wedge (i,j) \in \mathbf{w}} \mathrm{Pr}_{obj}(\mathbf{w}) \ , \tag{8}$$

where $\Gamma = \max_{i,j} \mathcal{O}(i,j)$ is used for normalization; $(i,j) \in \mathbf{w}$ means pixel (i,j) falls in $\mathbf{w}$; $n_w = 10000$ is used in our experiments; and we use the objectness code implemented by the authors of (Alexe et al 2012)[3].

The map $\mathcal{O}(I)$ is then combined with the scale-dependent saliency $S^{scale}(I)$ to inhibit regions containing weak objectness:

$$S^{so}(i,j) = S^{scale}(i,j) \cdot \mathcal{O}(i,j) \ . \tag{9}$$

$\mathcal{O}(I)$ is computed directly on I while $S^{scale}(i,j)$ is obtained from Eqn. 7. We name the augmented saliency S^{so} as *scale and object aware saliency* (SOAS). Figure 5 illustrates some examples.

The proposed SOAS is independent of thumbnailing transformation. This enables it to be combined with different image resizing approaches, such as retargeting and cropping described in the following sections.

5 Scale and Object Aware Thumbnail Retargeting

The first proposed SOAT method, SOAT$_{\mathrm{tp}}$, is a retargeting method guided by SOAS. To encourage structural smoothness, we choose a continuous model, TPS, for thumbnail warping. TPS provides a natural balance between warping smoothness and matching accuracy. A crucial issue is to find the appropriate control points to harness the warping artifacts. To solve the problem, we use a discrete retargeting approach, cyclic seam carving combined with SOAR (Sec. 5.1), to guide the sampling of control points. These control points are then used to estimate the TPS model. Consequently, our approach can also be viewed as a combination of discrete and continuous retargeting approaches.

5.1 Cyclic Seam Carving

To guide the TPS warping in SOAT$_{\mathrm{tp}}$, we use seam carving (SC) (Avidan and Shamir 2007) with two extensions: *cyclic seams* and SOAS integration.

Cyclic Seams. In the standard seam carving, sometimes the "best" seam has no choice but to cut across the extent of objects, which is largely due to the original definition of "seam": a continuous polyline from one boundary of the image to its opposite boundary. In this scenario, the resulting images suffer from damage to well-structured objects. With these broken object structures in the thumbnails, the thumbnail browsing efficiency and accuracy can be seriously hurt.

To reduce the possibility of a seam trespassing objects, we introduce the cyclic seams as following: an input image is first virtually warped into a cylinder shape by sticking its left and right (or top and bottom) margins together. Then a cyclic seam is defined as the original continuous seam but on this new virtual "cylinder" image. An illustration is shown in Figure 7. We name this the extended SC algorithm *Cyclic Seam Carving* (CSC). Intuitively, CSC allows a seam that crosses image boundaries to stay away from highly salient regions. On the other hand, a cyclic seam is still continuous in most of its segments. Being allowed to cross image boundaries, a cyclic seam reduces its chance to cut across foreground objects that usually reside far apart image boundaries.

SOAS Integration. Our second extension to the original SC is to augment the energy function with SOAS. We denote E^{seam} as the original energy used in SC which is based on the distribution of histogram of gradients, our scale and object aware energy E^{so} is defined as

$$E^{so} = \rho \cdot E^{seam} + (1 - \rho) \cdot S^{so} \ , \tag{10}$$

where ρ is the weight and empirically set to 0.3 throughout our experiments. The improved energy is then integrated in the CSC algorithm to provide control points needed for estimating TPS warping in the next step.

[3] http://www.vision.ee.ethz.ch/~calvin/objectness/

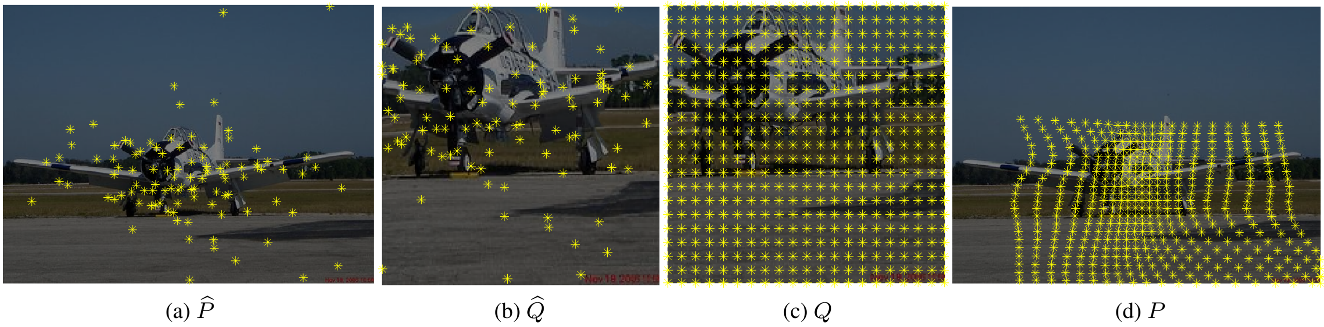

(a) $\widehat{P}$ (b) $\widehat{Q}$ (c) Q (d) P

Fig. 8 Control points sampling. (a) Control points in the original image; (b) Control points traced in the seam carved image; (c) New control points sampled from the seam carved image; (d) Control points traced in the original image.

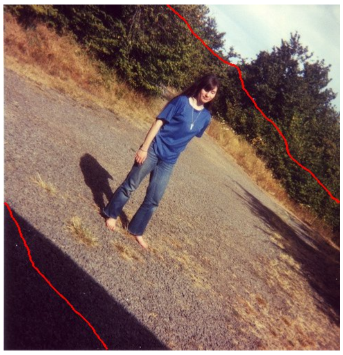


Fig. 7 Cyclic seam carving (CSC): a cyclic seam (in red line) is shown on an image picked from the PASCAL VOC 2008 dataset. The photo is originally taken this way (rotated). Indeed, this image inspired our idea of CSC.

5.2 Thin Plate Spline Warping

Discrete retargeting methods often generate serious artifacts when the target image has a size much smaller than the original image. In (Simakov et al 2008) a bidirectional similarity is used to alleviate the problem. We address this issue differently by combining a continuous warping model with a discrete retargeting guidance.

We use the thin-plate-spline (TPS) (Bookstein 1989) for our continuous warping model. TPS has been widely used in vision tasks, such as registration and matching, due to its attractive property in balancing local warping accuracy and smoothness. Specifically, given two sets of control points $P = \{\mathbf{p}_i \in \mathbb{R}^2, i = 1, 2, \cdots, n_l\}$ and $Q = \{\mathbf{q}_i \in \mathbb{R}^2, i = 1, 2, \cdots, n_l\}$, where $\mathbf{p}_i$ corresponds to $\mathbf{q}_i$, the TPS transformation $\widehat{f}$ is defined as the transformation from P to Q that minimizes the regularized bending energy $\mathcal{E}(f)$

$$\mathcal{E}(f) = \sum_i \| \mathbf{q}_i - f(\mathbf{p}_i) \|^2 + \lambda \iint (\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2 dxdy \ , \tag{11}$$

where λ is the weight parameter. Then $\widehat{f}$ is estimated as $\widehat{f} = \arg\min_f \mathcal{E}(f)$. In our image warping problem, $\widehat{f}$ is the displacement mapping from the input image to the target image.

To estimate the TPS model, we need to provide the control point sets P (in the original image) and Q (in the target image). This can be achieved from the CSC retargeting algorithm. A natural solution is to first define P in the input image and then locate its corresponding Q in the target image by tracing point shifting during the CSC process. However, the control point pairs obtained this way can be unstable: if most pixels in a region are eliminated by CSC, control points who originally resided in this area will be unevenly clustered in the target image. As a consequence, the estimated TPS model will be intensively affected by these singular points. Instead, we design a solution to get a uniformly sampled point set Q in the target image. The solution combines two steps described as following (both steps are used in estimating the TPS model):

- **Step One.** First, we sample a control point set $\widehat{P}$ (Figure 8(a)) based on the original image's saliency distribution. In particular, the density of $\widehat{P}$ is proportional to the saliency values. Then, we trace the displacement of $\widehat{P}$ during the CSC process. If a control point is eliminated during a CSC iteration, it will be re-assigned to its closest neighbor. Finally, after the CSC process, we acquire the corresponding control point set $\widehat{Q}$ (Figure 8(b)).
- **Step Two.** Instead of using $\widehat{Q}$, the control point set Q (Figure 8(c)) is sampled uniformly in the target image generated by CSC. Then, the control point set P (Figure 8(d)) is generated by mapping Q to the original image using the TPS model estimated by $\widehat{Q}$ and $\widehat{P}$. The control point sets P and Q are then used to estimate the warping used in the final retargeting.

Figure 14 shows some examples illustrating the effectiveness of each component in SOAT$_{\mathrm{tp}}$.

6 Scale and Object Aware Thumbnail Cropping

Retargeting methods can be viewed as inhomogeneously shrinking the input image I to the output image J. They intend to preserve important information over the *whole* image with some potential distortion of local geometric structures. Such local distortion may however confuse the human perception system, especially when the distortion is as severe as it is in thumbnailing.

An alternative solution, *cropping*, on the other hand, discards all information outside a cropping widow while keeping local structures untouched inside the window. In the recent evaluation (Rubinstein et al 2010b), manually cropping is considered to be one of the most favorite retargeting methods by human users. However, manual cropping can be tedious and time consuming, and thus impractical for real world applications. To design an automatic thumbnail cropping algorithm, two issues need to be addressed: the criterion to evaluate a cropping window and the method to search a good window according to the criterion. For example, in (Suh et al 2003) the classic saliency map (Itti et al 1998) is used as the solution to the first issue and a greedy searching strategy to the second one.

In this paper we focus on the effectiveness of encoding scale and object awareness in thumbnailing. As a result, we choose to extend the cropping algorithm in (Suh et al 2003) by using SOAS for saliency measurement. Given an image I defined on a grid Ω and its saliency S^{so}, the thumbnail cropping problem is to find a window $\mathbf{w}^*$ that contains as much as possible important pixels with a reduced size. In the following we describe the greedy search algorithm in (Suh et al 2003), with the proposed saliency S^{so}. We first define $\widehat{\mathbf{w}}(\lambda)$ as the minimum window that contains at least certain fraction (i.e. λ) of the total saliency in the whole image, that is

$$\widehat{\mathbf{w}}(\lambda) = \arg\min_{\mathbf{w}\in\mathcal{W}(\lambda)} (\text{area}(\mathbf{w})) \,, \tag{12}$$

where λ is the fraction threshold and

$$\mathcal{W}(\lambda) = \left\{ \mathbf{w} : \frac{\sum_{(x,y)\in\mathbf{w}} S^{so}(x,y)}{\sum_{(x,y)\in\Omega} S^{so}(x,y)} > \lambda \right\}$$

is the set of windows whose saliency is no less than λ of the saliency in the whole image.

The exhaustive search for $\widehat{\mathbf{w}}(\lambda)$ is computationally very expensive. This is because, despite using integral images, optimizing over λ is still very time consuming. Instead, an efficient greedy solution is used: $\widehat{\mathbf{w}}(\lambda)$ is initialized as the image center and then iteratively merged with the most salient point outside the current $\widehat{\mathbf{w}}(\lambda)$ until it falls into $\mathcal{W}(\lambda)$. After that, the "best" λ is found by

$$\lambda^* = \arg\max_{\lambda\in[\lambda_0..\lambda_1]} a'(\lambda), \tag{13}$$

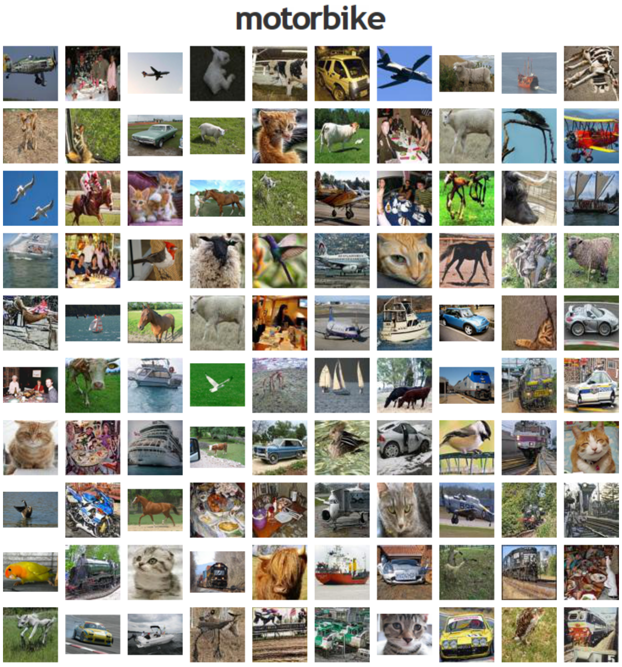


Fig. 9 The user interface used in the user study.

where $a(\lambda) \triangleq \text{area}(\widehat{\mathbf{w}}(\lambda))$ is the area of $\widehat{\mathbf{w}}(\lambda)$ and $[\lambda_0..\lambda_1]$ defines the search range ($\lambda_0 = 0.6$ and $\lambda_1 = 1$ are used throughout our study). Details about λ can be referred to (Suh et al 2003). Finally, the cropping window is chosen as $\mathbf{w}^* = \widehat{\mathbf{w}}(\lambda^*)$.

We name the proposed approach SOAT$_{\text{cr}}$. It demonstrates excellent performance in the user study (Section 7.1). Figure 15 shows some examples illustrating the effectiveness of each component in SOAT$_{\text{cr}}$. In addition, it is worth noting that the use of SOAS is not limited to the window searching algorithm used in this paper.

To facilitate future studies and evaluations related to the proposed SOAT algorithms, we have made the Matlab code implementation of SOAT$_{\text{cr}}$ and the result images in our experiment available for research usage at http://www.dabi.temple.edu/~hbling/code_data.htm#SOAR.

7 Experiments

Objective evaluation of thumbnailing algorithms is not easy: there is no "ground truth" thumbnails for an input image; and humans' feeling about the quality of a thumbnail can be very subjective. For this reason, carefully designed experiments are conducted to assess the SOAT algorithms. We first perform rigorous quantitative user studies with two different tasks: an *Image Browsing Task* and a *Subjective Preference Task*, in the context of thumbnailing. We then evaluate SOAT qualitative in the context where target scales are significantly larger than thumbnails.

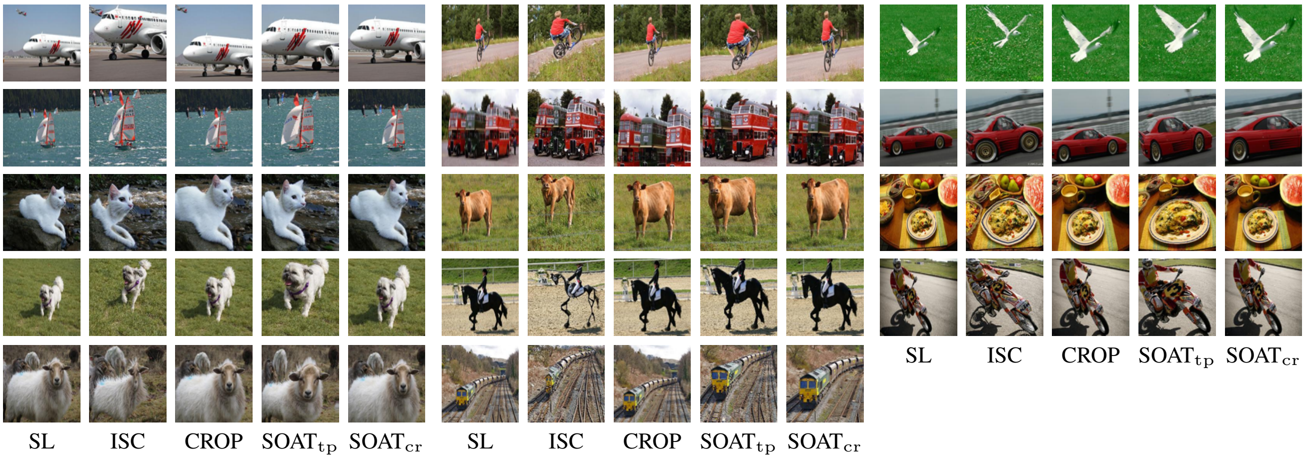


Fig. 10 Thumbnails used in the quantitative study. One thumbnail is shown per class per method.

7.1 Quantitative User Study

Our qualitative study contains two different tasks: (1) the *image browsing task* similar to the one used in (Suh et al 2003), in which participants are asked to find a thumbnail that matches given description, and (2) the *subjective preference task* similar to the one used in (Liu and Gleicher 2005), in which participants are asked to select thumbnails that they prefer to receive.

Design of study

The *image browsing task* is designed to measure the effectiveness (in terms of three criteria) of thumbnails generated by different methods in image search. This task is a within-subjects design with thumbnail method as the independent variable. Browsing accuracy, browsing efficiency and browsing rank (defined in the paragraph Evaluation criteria) are dependant variables. In this task, a subject is required to browse and search a target image, described verbally, from a set of image thumbnails randomly tiled in one page. For example, in the sample page shown in Figure 9, a subject is asked to visually search one particular thumbnail of “motorbike” from a screen of thumbnails of various classes: “bike”, “train”, “bird”, etc. There exists only one correct thumbnail (i.e., containing a “motorbike” in the example) in each page to avoid ambiguity.

The *subjective preference task* is designed to determine which of the compared methods is preferred by users. It is also a within-subjects design with thumbnail method as the independent variable. In this task, subjects are shown two thumbnails at each page and asked to choose their favorite one. The two thumbnails are generated from different thumbnailing methods from the same input image. The test page interface is similar to the one used in image browsing task except with less thumbnails.

Participants

In the image browsing task, 30 college student volunteers have been recruited and none of them has previous research experience related to image thumbnailing[4]. The display device is a monitor of 1680 × 1050@65 hz and 120 DPI. Participants sit 0.5 meters away from the monitor with normal indoor illumination. The test takes about one hour per subject.

In subjective preference task, 15 college student volunteers have been recruited with no related research experience in image thumbnailing. The hardware and environment setups are the same as the image browsing task.

Image datasets

Experiments have been done using a carefully prepared dataset. The image set used in the study contains 210 images randomly selected from the PASCAL VOC 2008/2009 database (Everingham et al 2008). Because the objectness measurement (used in the SOAS calculation) is also trained using a subset of PASCAL VOC 2008, we carefully checked our image selection to prevent any overlap. The selected images are divided into 14 classes, each with 15 images. These classes are: *aeroplane*, *bicycle*, *bird*, *boat*, *bus*, *car*, *cat*, *cow*, *dining table*, *dog*, *horse*, *motorbike*, *sheep* and *train*. Images are also checked inter-class to exclude those multi-labeled in different classes. For example, an image containing a cat and a bicycle may be included in both class *bicycle* and *cat*; such images are removed to avoid ambiguity. The image browsing task uses all 210 images while the subjective preference task uses a subset of randomly selected 105 images.

Thumbnailing methods

For each image in the data set, five versions of thumbnails are generated beforehand using different methods including scaling (SL), improved seam carving (ISC) (Rubinstein et al 2008), automatic cropping (CROP) (Suh et al

[4] There is no overlap between these 30 subjects and the 20 subjects in the experiment of our previous study (Sun and Ling 2011).

Comparison	Mean	Std.	p-value
$\mathrm{SOAT_{cr}}$ vs. SL	0.5572	0.1899	0.2801
$\mathrm{SOAT_{cr}}$ vs. ISC	0.7568	0.0871	5.6867×10^{-8}
$\mathrm{SOAT_{cr}}$ vs. CROP	0.5235	0.0476	0.0874
$\mathrm{SOAT_{cr}}$ vs. $\mathrm{SOAT_{tp}}$	0.7098	0.0745	9.7937×10^{-8}
$\mathrm{SOAT_{tp}}$ vs. SL	0.3931	0.1753	0.0400
$\mathrm{SOAT_{tp}}$ vs. ISC	0.6584	0.0583	1.4931×10^{-7}
$\mathrm{SOAT_{tp}}$ vs. CROP	0.3235	0.07826	1.2266×10^{-6}

Table 1 Preference ratios in the subjective preference user study.

2003), and the proposed $\mathrm{SOAT_{tp}}$ and $\mathrm{SOAT_{cr}}$. SL is included because it is straightforward and widely used in current real world applications. ISC is an improved version of SC and was shown to outperform SC in our preliminary study (Sun and Ling 2011). Furthermore, the proposed $\mathrm{SOAT_{tp}}$ can also be viewed as an extension of SC since CSC is used to guide the model estimation in $\mathrm{SOAT_{tp}}$. CROP is included because cropping is considered to be one of the users' favorite methods in a recent study (Rubinstein et al 2010b) and $\mathrm{SOAT_{cr}}$ borrows the window searching algorithm from CROP.

Though there exists many other excellent image resizing algorithms, we limit our selection to the above ones for several reasons: First, the limitation in the number of subjects prevents us from including too many methods to perform valid statistical analysis. Second, the selected methods are most related to ours as described above. Third, the source codes of ISC[5] and CROP[6] are publicly available from the original authors, which makes it easier for fair comparison.

Test procedure

Image Browsing Task: Each subject is requested to browse a total of 210 pages. In each page, the subject sees a word describing the class of the target and 100 image thumbnails aligned in a 10 × 10 grid. Among the 100 thumbnails, there is only one that matches the description and it is placed randomly. 99 other "filler" images are chosen randomly from the rest of the classes with restrictions: the potential class conflict due to appearance ambiguity is taken into consideration. For example, motorbike images and bicycle images will not appear in the same page since they are highly similar in appearance, especially when displayed in small scales. For each page, we set up a mandatory 60 second time-out to limit the time duration of the experiment. In failure to find the correct image in 60 seconds from the current page, the user will be redirected automatically to next page. For every thumbnail position on the screen, one of the five versions of thumbnails is randomly picked to be presented. The subject needs to click on the correct one among the 100 thumbnails. The time cost to find the result and the selection of each page are recorded for every user. The records of the first 10 pages are discarded to allow the subjects get familiar with the system. Example thumbnails from different methods are shown in Figure 10. The order of the pages is randomized for each subject.

Subjective Preference Task: Each subject is requested to browse a total of 105 images × 7 comparisons = 735 pages. Each page shows two different thumbnails aligned side by side, with a word describing the class of the thumbnails shown on the top. The two thumbnails are generated from the same input image with two different algorithms that randomly selected from one of the seven pairs listed in Table 1. Then, one of the two thumbnails is randomly placed on the left while the other on the right. A subject is forced to choose one of the two thumbnails depending on subjective preference. The selection of left/right side and name of the two methods are recorded for every user. Order of pages is randomized individually.

Evaluation criteria

Image Browsing Task: Three criteria, *browsing accuracy*, *browsing efficiency* and *browsing rank*, are used in performance evaluation and defined below.

First of all, we denote n_T as the number of thumbnailing methods and n_u the number of subjects in the experiment ($n_T = 5$ and $n_u = 30$ in this study). For the i^{th} user, denote N_i^k as the number of pages in which thumbnails generated by the k^{th} algorithm are used as the searching target. Among these N_i^k pages, suppose there are M_i^k pages in which the user correctly identifies the target. The time cost the i^{th} user took in each correctly identified page j in M_i^k is $t_{i,j}^k$.

The ***browsing accuracy*** for the k^{th} algorithm over the whole experiment is defined as:

$$\frac{1}{n_u} \sum_{i=1}^{n_u} \frac{M_i^k}{N_i^k} \,. \tag{14}$$

The ***browsing efficiency*** for the k^{th} algorithm over the whole experiment is defined as:

$$\frac{1}{n_u} \sum_{i=1}^{n_u} \frac{1}{M_i^k} \sum_{j=1}^{M_i^k} t_{i,j}^k \,. \tag{15}$$

Browsing efficiency may be affected by many subjective factors such as browsing habit and mood. To address this issue, we include a less user-sensitive measure, browsing rank. For the i^{th} user, we rank the thumbnailing methods according to the average searching time for his/her records. Specifically, we denote r_i^k as the rank of the k^{th} thumbnailing algorithm, such that $r_i^k = 1$ is the best and $r_i^k = n_T$ the worst. Then, the ***browsing rank*** for the k^{th} algorithm over all users is defined as:

$$\frac{1}{n_u} \sum_{i=1}^{n_u} r_i^k \,. \tag{16}$$

[5] http://people.csail.mit.edu/mrub/index.html#code_seamcarving

[6] http://www.dabi.temple.edu/~hbling/code/auto_thumb.zip

Method	SL	ISC	CROP	SOAT$_{\mathrm{tp}}$	SOAT$_{\mathrm{cr}}$
Browsing accuracy (%)	90.47±8.1	88.40±6.8	93.92±6.5	92.56±7.3	**95.04±5.2**
Browsing efficiency (sec.)	13.70±3.5	14.79±3.4	12.27±3.3	11.44±3.0	**11.20±2.8**
Browsing rank ([1..5])	3.69±1.1	4.44±1.0	2.91±1.1	2.13±1.2	**1.84±0.9**

Table 2 Average browsing accuracies, time costs, and ranking of methods in the quantitative study. SL, ISC, CROP stand for Scaling, Improved Seam Carving, Automatic Cropping respectively.

The three criteria capture different but correlated characteristic aspects of a thumbnailing algorithm. A major reason to include browsing accuracy is to ensure that the thumbnails generated by SOATs do not bring extra difficulties for recognition. Consequently, we are more interested in browsing efficiency and browsing rank.

Subjective Preference Task: In this task we have $n_P = 7$ pairs, $n_u = 15$ participants and use preference ratio as the criterion. For the i^{th} user, denote N_i^p as the number of pages presented to him or her containing thumbnails from the p^{th} pair, and denote M_i^p as the number of pages voted in favor of the first thumbnailing algorithm in the p^{th} pair. The ***Preference ratio*** of p^{th} pair is then defined as:

$$\frac{1}{n_u}\sum_{i=1}^{n_u}\frac{M_i^p}{N_i^p}\ . \tag{17}$$

Results and Analysis

Image Browsing Task: The browsing accuracies, efficiencies and ranks of tested methods are summarized in Table 2. The box plots of the three criteria are given in Figure 11. The results show in general the effectiveness of the proposed SOAT framework for thumbnailing.

From Table 2, we can see that all methods have achieved high accuracies (around or above 90%), which confirms our assumption that these thumbnailing methods do not bring extra difficulty in recognition. Amongst all methods, the proposed SOAT$_{\mathrm{cr}}$ performs the best. Before drawing conclusions about browsing efficiency, which is our focus in evaluation, we first perform the following statistical analysis to validate the significance between different approaches.

For a rigorous evaluation, we have conducted an one-way ANOVA analysis on the browsing efficiency for all five methods. The F-value is 23.14 and the p-value is 5.20×10^{-19}, implying that the five methods are significantly different in browsing efficiency. Furthermore, a multiple comparison test using the information from ANOVA has been performed to distinguish if our method is significantly different in pair-wise comparison with other methods. Results are given in Table 3. The 95% confidence intervals for all compared mean differences have rejected the hypothesis that the true differences are zero. In other words, the differences between our methods (SOAT$_{\mathrm{cr}}$ and SOAT$_{\mathrm{tp}}$) and other methods (SL, ISC and CROP) are significant at 0.05 level. The SOAT$_{\mathrm{cr}}$ and SOAT$_{\mathrm{tp}}$ themselves do not show a significant difference statistically though.

The Kruskal-Wallis test has been used to analyze the browsing rank. The p-value is 3.34×10^{-15}, indicating that the five methods are significantly different in browsing rank. Results from pair-wise comparison tests are reported in Table 3, which show again significant differences between our methods and others at 0.05 level.

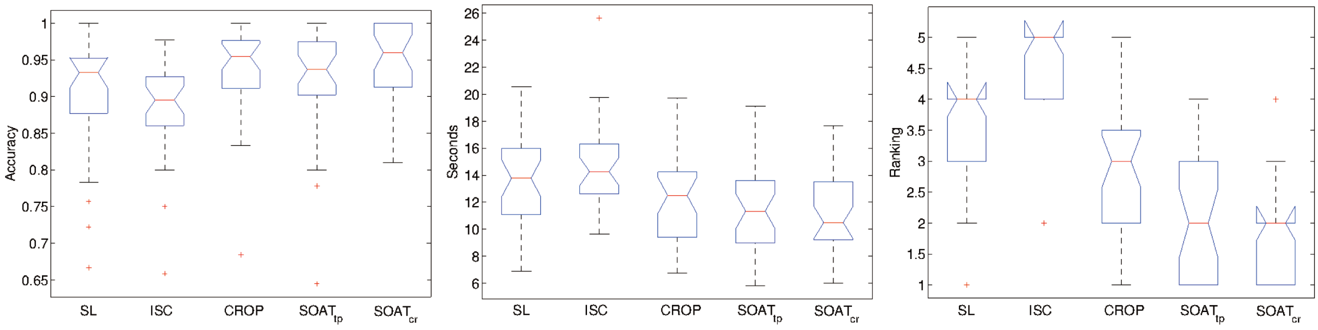
From the results and analysis we have the following observations:

- The proposed SOAT$_{\mathrm{cr}}$ algorithm performs the best in all three criteria. In particular, it significantly outperforms the straightforward scaling approach. This is consistent with the previous discovery (Suh et al 2003).
- The retargeting-based SOAT (i.e. SOAT$_{\mathrm{tp}}$) beats CROP in browsing efficiency and rank, but not in browsing accuracy. This suggests that the browsing efficiency and browsing accuracy are not necessarily consistent.
- Since SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$ are extended from SC and CROP respectively, the superiority of SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$ over SC and CROP validates the benefit of encoding scale and object awareness in the SOAT algorithms.
- It is not surprising that ISC performs not as good as other approaches. Because a typical retargeting method like ISC is not designed for dealing with targets with extremely small scales. In contrast, SOAT algorithms perform better by explicitly taking into account thumbnail relevant factors.

In addition to the above analysis, we plot accuracy versus time cost in Figure 12. From the figure we find that browsing accuracy in general decreases as the time cost increases. That is to say, even with much more time, a user is unlikely to find the correct answer if he or she did not find it earlier.

Subjective Preference Task: The means and standard deviations of preference ratios of tested method pairs are summarized in Table 1. We have done a t-test on each pair with the null hypothesis that the mean of the distribution is 0.5, meaning that users have no preference between the two methods. The p-values are reported in Table 1. Note that all pairs are significantly different at 0.05 level except (SOAT$_{\mathrm{cr}}$, SL) and (SOAT$_{\mathrm{cr}}$, CROP). We have following observation based on the results and analysis:

- The proposed SOAT$_{\mathrm{cr}}$ algorithm outperforms all other algorithms. In particular, the preferences of SOAT$_{\mathrm{cr}}$ over



(a) Browsing accuracy (b) Browsing efficiency (c) Browsing rank

Fig. 11 Box plots for the quantitative results. (a) Browsing accuracy (in [0..1]). (b) Browsing efficiency (in seconds). (c) Browsing rank (in [1..5]).

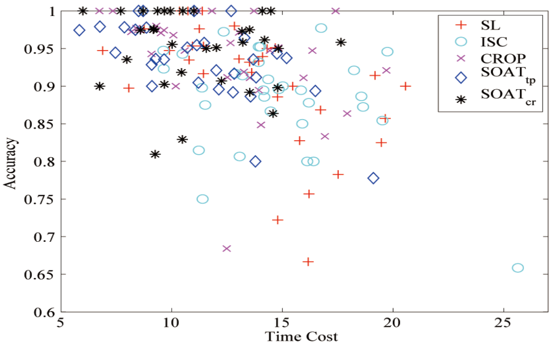


Fig. 12 Browsing efficiency versus browsing accuracy.

SOAT$_{tp}$ and ISC are statistically significant. This superiority should be attributed to the thumbnail specific saliency and the distortion resistance of cropping.

- The preference of SOAT$_{tp}$ against ISC confirmed our efforts on reducing distortion and maintaining structure smoothness in algorithm design.
- Interestingly, SOAT$_{tp}$ is less preferred than CROP and SL, though it beats both in the image browsing task. This suggests that humans may use different mechanisms for the two tasks. In particular, a visually preferable thumbnail is not necessarily efficient for image browsing.

7.2 Qualitative Experiments

While we are interested mainly in tiny thumbnails, it is also worth investigating how well the proposed method performs when the size change is less drastic. Recently, Rubinstein et al (2010a) released a benchmark dataset, *RetargetMe*, together with the results from many state-of-the-art image retargeting methods. Thanks to their efforts, we run the proposed SOAT algorithms on the dataset for qualitative inspection. For a fair comparison, our methods are tuned to create target images in the same scales used in (Rubinstein et al 2010b), i.e., 75% in the horizontal dimension of the original image sizes. Some example results are shown in Figure 13. We visually check the results rather than designing another subjective preference user study because the user study is only appropriate in the context of thumbnailing. Other retargeting methods in this dataset are not designed for thumbnailing problems hence it will be unfair to get them involved.

By checking the results, we feel it is hard to visually pick out a method that is better than all others. In particular, the results from SOAT algorithms look similar to the results from others. The best and worst results for different input images often come from different methods.

7.3 Analysis and Discussion

Analysis of components in SOAT. To further understand the role of each components introduced in the SOAT framework, we design different “intermediate” versions of SOAT algorithms with various combinations of components. We then apply these algorithms to the dataset used in the quantitative study and visually check the results.

Figure 14 shows some results from various combinations of the components in SOAT$_{tp}$. First, the results suggest that the use of TPS model drastically improves the visual quality of resulting thumbnails. This is mainly due to the local structure preservation by TPS. Second, the integration of scale and object aware information helps in some cases. The use of cyclic seams, in contrast, affects only marginally to the results, indicating that CSC is a conservative extension of SC. Finally, in average the thumbnails by SOAT$_{tp}$ are visually better than those from other combinations.

Figure 15 shows some results from various combinations of the components in SOAT$_{cr}$. Similar to the cases for SOAT$_{tp}$, the integration of scale and object aware information in general helps improving the visual quality of resulting thumbnails. However, the improvement from CROP to SOAT$_{cr}$ is less impressive than that from SC to SOAT$_{tp}$. This is because the CROP algorithm, which preserves per-

	Browsing efficiency			Browsing rank	
Comparison	Lower	Difference in means	Upper	Difference in means	p-value
SL vs. SOAT$_{\mathrm{cr}}$	1.4327	2.3317	3.2306	1.84	5.0883×10^{-08}
ISC vs. SOAT$_{\mathrm{cr}}$	2.6861	3.5851	4.4841	2.59	2.3585×10^{-10}
CROP vs. SOAT$_{\mathrm{cr}}$	0.0550	0.9540	1.8530	1.06	1.4202×10^{-04}
SL vs. SOAT$_{\mathrm{tp}}$	1.4232	2.3222	3.2212	1.56	3.5359×10^{-06}
ISC vs. SOAT$_{\mathrm{tp}}$	2.6767	3.5756	4.4746	2.31	1.6757×10^{-09}
CROP vs. SOAT$_{\mathrm{tp}}$	0.0456	0.9445	1.8435	0.78	0.0108
SOAT$_{\mathrm{cr}}$ vs. SOAT$_{\mathrm{tp}}$	-0.9084	-0.0095	0.8895	-0.28	0.4054
SL vs. ISC	-2.1524	-1.2534	-0.3545	-0.75	7.3706×10^{-04}
SL vs. CROP	0.4787	1.3777	2.2766	0.78	0.0041
ISC vs. CROP	1.7321	2.6311	3.5301	1.53	1.2831×10^{-06}

Table 3 Browsing efficiency: Tukey's least significant difference (LSD) test for multiple comparisons. The 95% confidence intervals are [Lower, Upper]. Browsing rank: Kruskal-Wallis test.

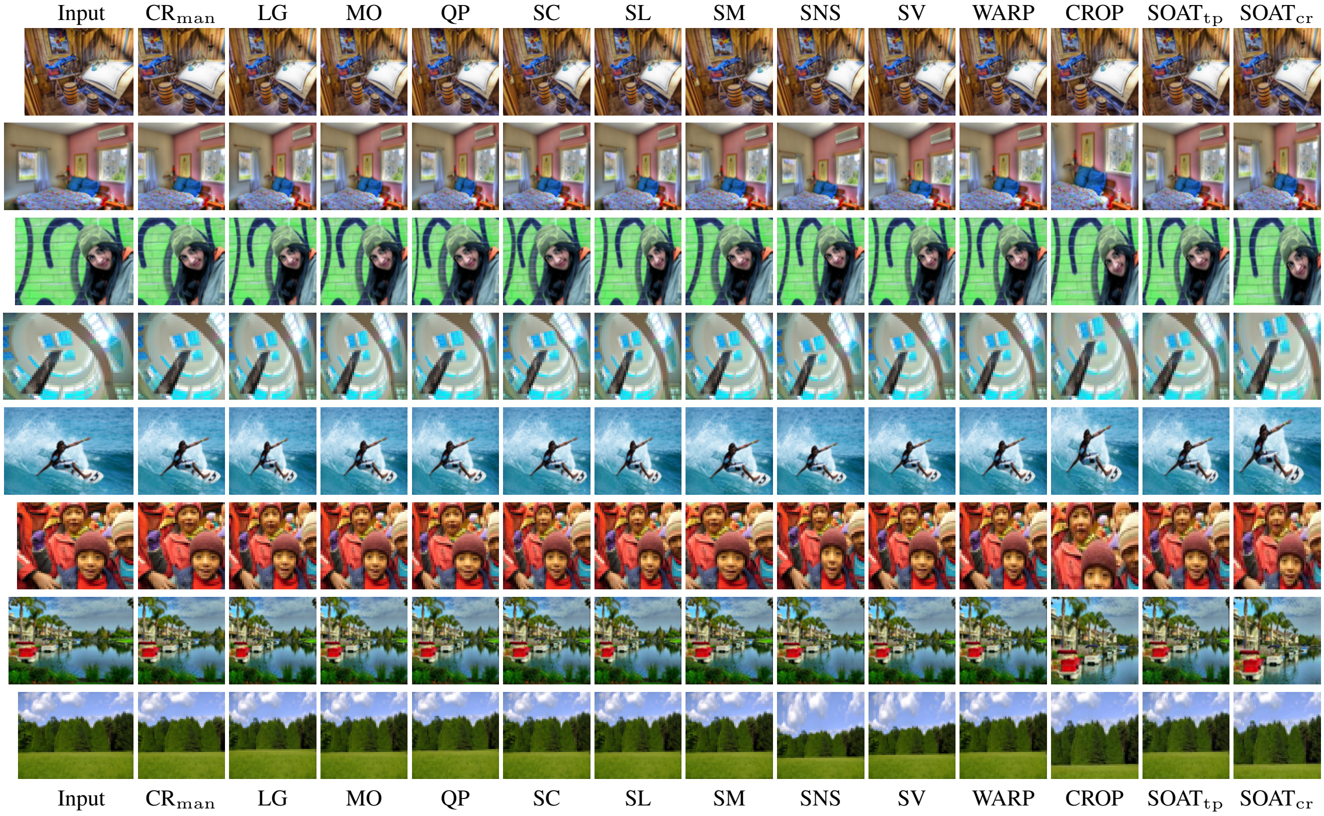


Fig. 13 Example results on the RetargetMe dataset. From left to right: the input image, manual cropping (CR$_{\mathrm{man}}$), energy-based deformation (LG) (Karni et al 2009), multi-operator media retargeting (MO) (Rubinstein et al 2009), quadratic programming (QP) (Chen et al 2010), seam carving (SC) (Avidan and Shamir 2007), scaling (SL), shift-maps (SM) (Pritch et al 2009), optimized scale-and-stretch (SNS) (Wang et al 2008), streaming video (SV) (Krähenbühl et al 2009), nonhomogeneous warping (WARP) (Wolf et al 2007), automatic cropping (CROP) (Suh et al 2003), SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$. All results, except those of CROP, SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$, are from (Rubinstein et al 2010b).

fect structure smoothness inside the cropping window, is to some extent similar to the role of TPS in SOAT$_{\mathrm{tp}}$.

In summary, both SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$ benefit from addressing the three issues: thumbnail scale, object completeness, and structure smoothness.

Discussion and error analysis. We summarize the number of failures in the image browsing task in Table 4, where a *failure* means a user failed to correctly identify the target thumbnail in a given page. From the table we have several observations: (1) There is not a single method that beats all others over all classes. This phenomenon is consistent with the observations in our qualitative study. (2) SOAT$_{\mathrm{cr}}$ works the best in half of the classes, which confirms the results in our quantitative study. (3) It is interesting that the recogni-

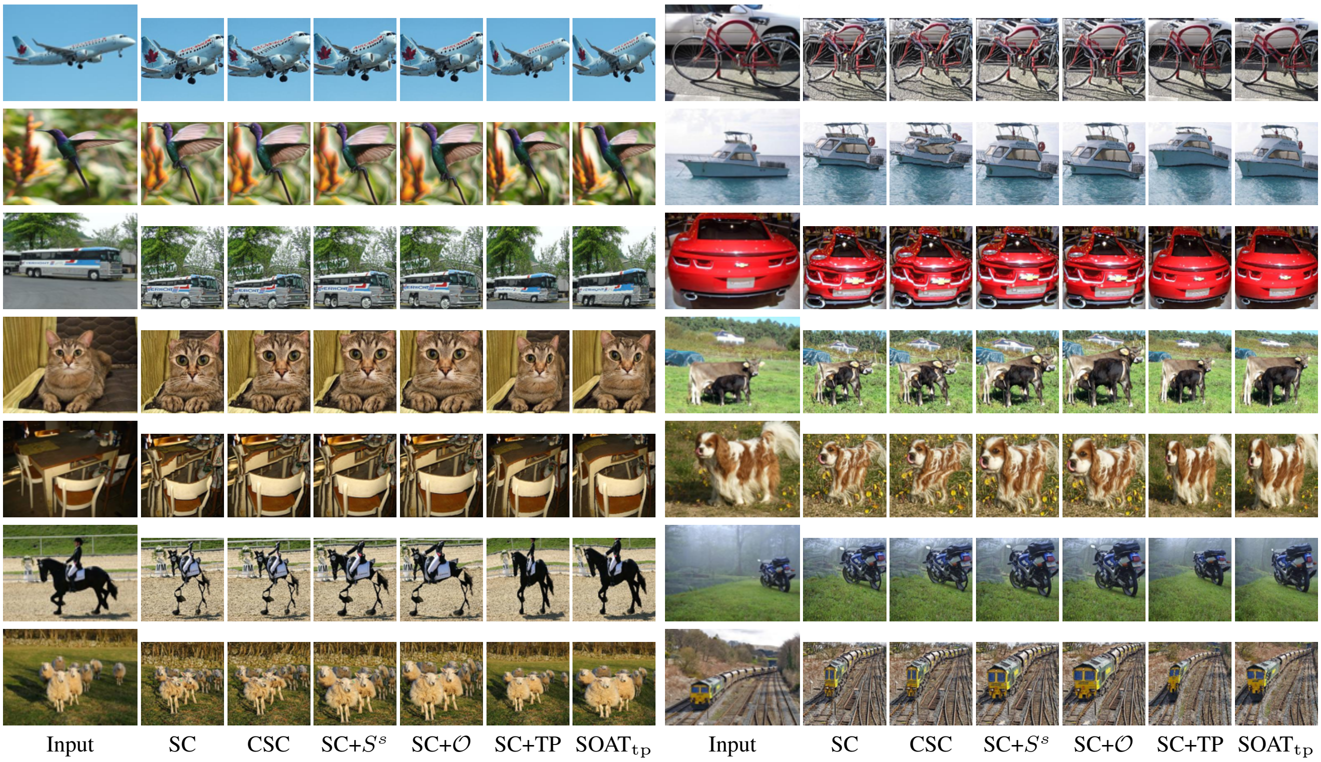


Fig. 14 Effectiveness of different components in SOAT$_{\text{tp}}$, details in Sec. 7.3. Notations: SC+S^s denotes for SC with scale-dependent saliency, SC+$\mathcal{O}$ for SC with objectness, SC+TP for SC followed by the TPS warping. One example is chosen per class per method.

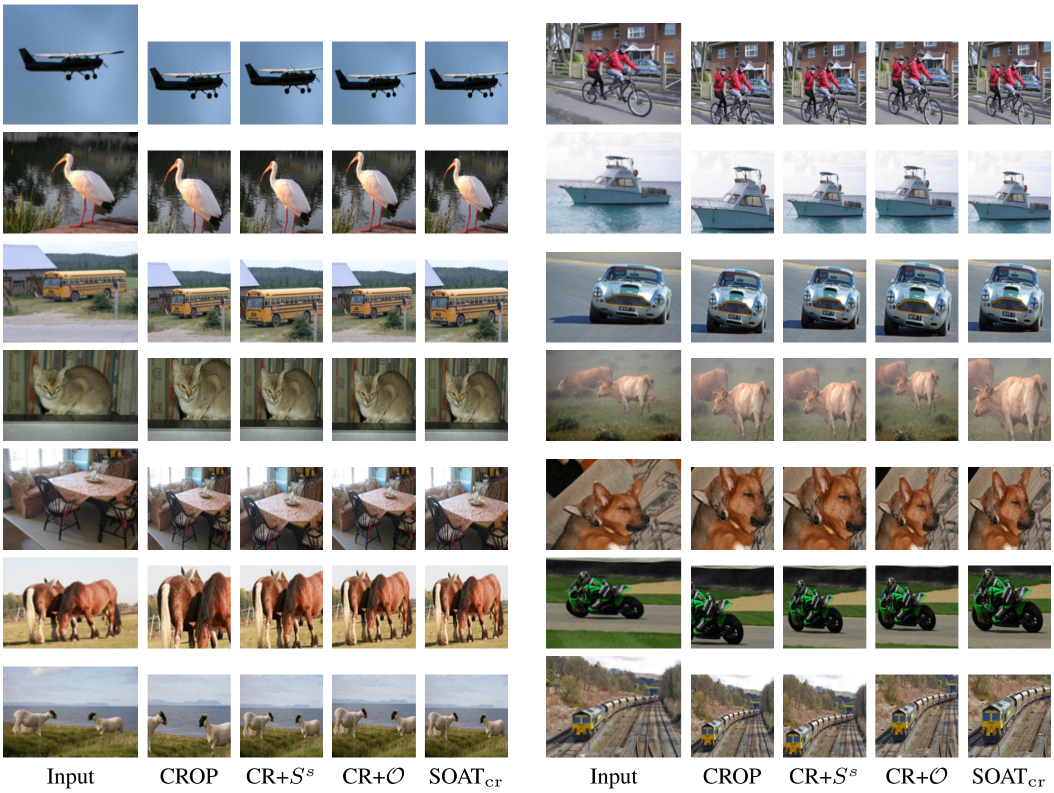


Fig. 15 Effectiveness of different components in SOAT$_{\text{cr}}$, details in Sec. 7.3. Notations: CR+S^s denotes for CROP with scale-dependent saliency, CR+$\mathcal{O}$ for CROP with objectness. One example is chosen per class per method.

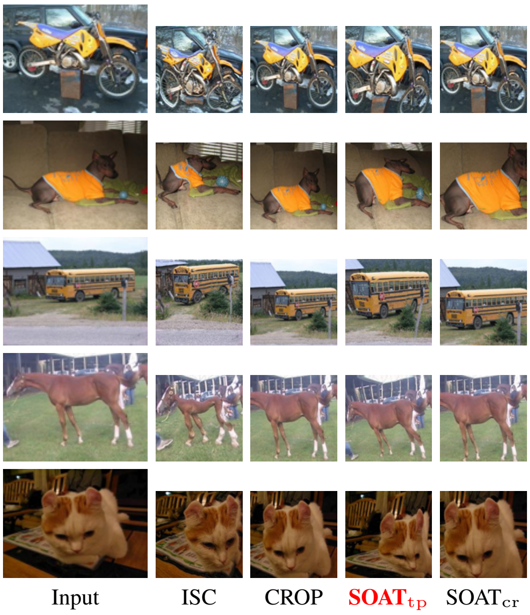


Fig. 16 Examples where the SOAT$_{\mathrm{tp}}$ thumbnails fail to be recognized. The classes from top to bottom are: *bike*, *dog*, *bus*, *horse*, and *cat*.

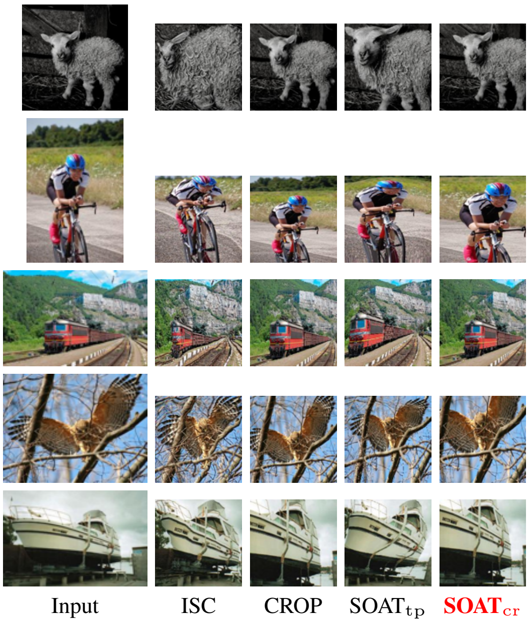


Fig. 17 Examples where the SOAT$_{\mathrm{cr}}$ thumbnails fail to be recognized. The classes from top to bottom are: *sheep*, *bicycle*, *train*, *bird*, and *boat*.

Category	SL	ISC	CROP	SOAT$_{\mathrm{tp}}$	SOAT$_{\mathrm{cr}}$
aeroplane	7	**2**	**2**	7	**2**
bicycle	1	**0**	**0**	1	3
bird	10	11	**5**	6	7
boat	3	5	8	6	**2**
bus	2	4	4	**1**	2
car	8	8	4	**3**	6
cat	13	**6**	**6**	9	9
cow	20	24	**6**	16	**6**
dining table	7	2	3	5	**1**
dog	15	33	11	19	**10**
horse	7	18	4	7	**2**
motorbike	5	5	7	6	**4**
sheep	16	24	12	**7**	10
train	4	3	5	**1**	3
Average	8.4	10.4	5.5	6.7	**4.8**

Table 4 Numbers of failure cases in the browsing user study for different categories.

tion of images from “natural” classes (e.g., *dog* and *cow*) are in general more difficult than those from “artificial” classes (e.g., *train* and *bicycle*).

Figures 16 and 17 show some failure examples in the quantitative study for SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$ respectively. Thumbnails from other methods are also shown for reference. The failure of SOAT$_{\mathrm{tp}}$ is mainly due to large geometric distortion, e.g., the *dog* example, which causes trouble for human perception. The failure of SOAT$_{\mathrm{cr}}$, on the other hand, is mainly due to the removal of some background area (context information) that helps users to understand the image content at a small scale. One interesting example is the *bicycle* example in Figure 17. The cropping window focuses on the rider rather than the bicycle itself. The window seems to be accurate but not consistent with the class label, which is *bicycle*, hence resulting in a recognition failure.

It is worth noting that one challenge for cropping-based thumbnailing appears when the original image has distributed saliency (e.g., multiple foreground objects). In such a case, retargeting-based solution may perform better by squeezing out unimportant regions.

8 Conclusion

In this paper we proposed encoding scale and object aware information for thumbnail generation, with a new framework named scale and object aware thumbnailing (SOAT). Two thumbnailing algorithms, namely SOAT$_{\mathrm{tp}}$ and SOAT$_{\mathrm{cr}}$, have been designed to combine the scale and object aware saliency with image retargeting and thumbnail cropping respectively. To objectively evaluate the proposed algorithms, we conducted user studies with an image browsing task and a subjective preference task. The statistical analysis on the study strongly suggests the effectiveness of the SOAT algorithms, especially the cropping-based version, i.e., SOAT$_{\mathrm{cr}}$. The algorithms were further tested qualitatively on the RetargetMe benchmark and demonstrated state-of-the-art performances.

There are two directions that we are interested in the future. First, for static images, we expect more effective ways to utilize scale and object awareness for thumbnailing. In particular, the currently used heuristic cropping algorithm can be replaced with more principled solutions. Second, for video sequences, we are interested in investigating the scale and object awareness in the spatio-temporal domain.

Acknowledgment We thank B. Alexe for making the objectness code available, M. Rubinstein for seam carving code, G. Teodoro for proofreading the manuscript, and all anonymous reviewers for valuable suggestions which help significantly improve the manuscript. This work was supported in part by NSF Grant IIS-1218156.

References

Alexe B, Deselaers T, Ferrari V (2012) Measuring the objectness of image windows. Pattern Analysis and Machine Intelligence, IEEE Transactions on to appear

Avidan S, Shamir A (2007) Seam carving for content-aware image resizing. ACM Transactions on Graphics 26(3)

Bookstein F (1989) Principal warps: thin-plate splines and the decomposition of deformations. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11(6):567 –585

Chen LQ, Xie X, Fan X, Ma WY, Zhang HJ, Zhou HQ (2003) A visual attention model for adapting images on small displays. Multimedia Systems 9:353–364

Chen R, Freedman D, Karni Z, Gotsman C, Liu L (2010) Content-aware image resizing by quadratic programming. In: Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on, pp 1 –8

Ding Y, Xiao J, Yu J (2011) Importance filtering for image retargeting. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp 89 –96

El-Alfy H, Jacobs D, Davis L (2007) Multi-scale video cropping. In: Proceedings of the 15th international conference on Multimedia, ACM, New York, NY, USA, MULTIMEDIA '07, pp 97–106

Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2008) The PASCAL Visual Object Classes Challenge 2008 (VOC2008) Results. http://www.pascal-network.org/challenges/VOC/voc2008/workshop/index.html

Grundmann M, Kwatra V, Han M, Essa I (2010) Discontinuous seam-carving for video retargeting. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, pp 569 –576

Guo Y, Liu F, Shi J, Zhou ZH, Gleicher M (2009) Image retargeting using mesh parametrization. Multimedia, IEEE Transactions on 11(5):856 –867

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20(11):1254 –1259

Judd T, Durand F, Torralba A (2011) Fixations on low-resolution images. Journal of Vision 11(4)

Karni Z, Freedman D, Gotsman C (2009) Energy-based image deformation. In: Proceedings of the Symposium on Geometry Processing, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, SGP '09, pp 1257–1268

Kennedy L, van Zwol R, Torzec N, Tseng B (2011) Learning crop regions for content-aware generation of thumbnail images. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ACM, New York, NY, USA, ICMR '11, pp 30:1–30:8

Kim JS, Kim JH, Kim CS (2009) Adaptive image and video retargeting technique based on fourier analysis. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, pp 1730 –1737

Krähenbühl P, Lang M, Hornung A, Gross M (2009) A system for retargeting of streaming video. ACM Transactions on Graphics 28(5):126:1–126:10

Lam H, Baudisch P (2005) Summary thumbnails: readable overviews for small screen web browsers. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA, CHI '05, pp 681–690

Li X, Ling H (2009) Learning based thumbnail cropping. In: Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, pp 558 –561

Liu F, Gleicher M (2005) Automatic image retargeting with fisheye-view warping. In: Proceedings of the 18th annual ACM symposium on User interface software and technology, ACM, New York, NY, USA, UIST '05, pp 153–162

Liu F, Gleicher M (2006) Region enhanced scale-invariant saliency detection. In: Multimedia and Expo, 2006 IEEE International Conference on, pp 1477 –1480

Luo Y, Yuan J, Xue P, Tian Q (2010) Saliency density maximization for object detection and localization. In: Proceedings of the 10th Asian conference on Computer vision - Volume Part III, ACCV'10, pp 396–408

Mannos J, Sakrison D (1974) The effects of a visual fidelity criterion of the encoding of images. Information Theory, IEEE Transactions on 20(4):525 – 536

Mansfield A, Gehler P, Van Gool L, Rother C (2010) Scene carving: scene consistent image retargeting. In: Proceedings of the 11th European conference on Computer vision: Part I, Springer-Verlag, Berlin, Heidelberg, ECCV'10, pp 143–156

Marchesotti L, Cifarelli C, Csurka G (2009) A framework for visual saliency detection with applications to image thumbnailing. In: Computer Vision, 2009 IEEE 12th International Conference on, pp 2232 –2239

Niu Y, Liu F, Li X, Gleicher M (2012) Image resizing via non-homogeneous warping. Multimedia Tools Appl 56(3):485–508

Peli E (2001) Contrast sensitivity function and image discrimination. Journal of the Optical Society of America A 18(2):283–293

Pritch Y, Kav-Venaki E, Peleg S (2009) Shift-map image editing. In: Computer Vision, 2009 IEEE 12th International Conference on, pp 151 –158

Rubinstein M, Shamir A, Avidan S (2008) Improved seam carving for video retargeting. ACM Transactions on Graphics 27(3):16:1–16:9

Rubinstein M, Shamir A, Avidan S (2009) Multi-operator media retargeting. ACM Transactions on Graphics 28(3):23:1–23:11

Rubinstein M, Gutierrez D, Sorkine O, Shamir A (2010a) A benchmark for image retargeting. http://people.csail.mit.edu/mrub/retargetme/

Rubinstein M, Gutierrez D, Sorkine O, Shamir A (2010b) A comparative study of image retargeting. ACM Trans Graph 29(6):160:1–160:10

Simakov D, Caspi Y, Shechtman E, Irani M (2008) Summarizing visual data using bidirectional similarity. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, pp 1 –8

Suh B, Ling H, Bederson BB, Jacobs DW (2003) Automatic thumbnail cropping and its effectiveness. In: Proceedings of the 16th annual ACM symposium on User interface software and technology, ACM, New York, NY, USA, UIST '03, pp 95–104

Sun J, Ling H (2011) Scale and object aware image retargeting for thumbnail browsing. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp 1511 –1518

Van Nes FL, Bouman MA (1967) Spatial modulation transfer in the human eye. Journal of the Optical Society of America 57(3):401–406

Wang Y, Zhu SC (2008) Perceptual scale-space and its applications. International Journal of Computer Vision 80(1):143–165

Wang YS, Tai CL, Sorkine O, Lee TY (2008) Optimized scale-and-stretch for image resizing. ACM Transactions on Graphics 27(5):118:1–118:8

Wang YS, Lin HC, Sorkine O, Lee TY (2010) Motion-based video retargeting with optimized crop-and-warp. ACM Transactions on Graphics 29:90:1–90:9

Wolf L, Guttmann M, Cohen-Or D (2007) Non-homogeneous content-driven video-retargeting. In: Computer Vision (ICCV), IEEE 11th International Conference on, pp 1 –6

Wu H, Wang YS, Feng KC, Wong TT, Lee TY, Heng PA (2010) Resizing by symmetry-summarization. ACM Transactions on Graphics 29(6):159:1–159:10