A survey of image retargeting techniques

Daniel Vaquero[a], Matthew Turk[a], Kari Pulli[b], Marius Tico[b], Natasha Gelfand[b]

[a]University of California, Santa Barbara
Department of Computer Science
Santa Barbara, CA, 93106-5110, USA

[b]Nokia Research Center
955 Page Mill Road
Palo Alto, CA, 94304-1003, USA

ABSTRACT

Advances in imaging technology have made the capture and display of digital images ubiquitous. A variety of displays are used to view them, ranging from high-resolution computer monitors to low-resolution mobile devices, and images often have to undergo changes in size and aspect ratio to adapt to different screens. Also, displaying and printing documents with embedded images frequently entail resizing of the images to comply with the overall layout. Straightforward image resizing operators, such as scaling, often do not produce satisfactory results, since they are oblivious to image content. In this work, we review and categorize algorithms for content-aware image retargeting, i.e., resizing an image while taking its content into consideration to preserve important regions and minimize distortions. This is a challenging problem, as it requires preserving the relevant information while maintaining an aesthetically pleasing image for the user. The techniques typically start by computing an importance map which represents the relevance of every pixel, and then apply an operator that resizes the image while taking into account the importance map and additional constraints. We intend this review to be useful to researchers and practitioners interested in image retargeting.

Keywords: Image retargeting, automatic cropping, saliency measures, content-aware image resizing

1. INTRODUCTION

With the recent advances in imaging technology, digital images have become an important component of media distribution. Images are frequently used in news stories, and people post their pictures online to be seen by

Authors' e-mail addresses: {daniel,mturk}@cs.ucsb.edu, {kari.pulli,marius.tico,natasha.gelfand}@nokia.com. Corresponding author: Daniel Vaquero. This is a preprint version of the paper. Citation: "Daniel Vaquero, Matthew Turk, Kari Pulli, Marius Tico, and Natasha Gelfand. "A survey of image retargeting techniques," Applications of Digital Image Processing XXXIII, Andrew G. Tescher, Editor, Proc. SPIE 7798, 779814 (2010)." http://dx.doi.org/10.1117/12.862419

Figure 1. An example of retargeting. From left to right: original image, resized using seam carving,[1] scaling, and cropping.

family and friends. Images, however, are typically authored once, but need to be adapted for consumption under varied conditions. As an example, pictures are often printed on paper that can vary in size, or the area available for the picture may have a different aspect ratio than the original image has for layout reasons. Dynamically changing the layout of web pages in browsers should take into account the distribution of text and images, resizing them if necessary. The use of thumbnails that faithfully represent the image content is important in image browsing applications. In addition, a variety of displays can be used for image viewing, ranging from high-resolution computer monitors to TV screens and low-resolution mobile devices.

This diversity of image consumption conditions introduces a new problem: images must be resized for optimal display or use in different applications. The process, also known as image retargeting or image resizing (Fig. 1), consists of modifying the image's aspect ratio and size in order to best satisfy the new requirements. However, straightforward image resizing operators, such as scaling, often do not produce satisfactory results, since they are oblivious to image content. To overcome this limitation, a class of techniques attempt to resize the images in a content-aware fashion, i.e., taking the image content into consideration to preserve important regions and minimize distortions. This is a challenging problem, as it requires preserving the relevant information while maintaining an aesthetically pleasing image for the user.

Mobile phones and personal digital assistants (PDAs) typically have limited resolution due to their small form factor. Their increase in popularity in recent years makes the image retargeting problem be very relevant, due to the need to convert high resolution images for displaying on small screens. Even if technological advances allow for their resolution to increase, their physical area will still be small. Hence, rearranging the relative sizes of different objects in the image could still provide an improved viewing experience, despite the availability of more pixels. Retargeting techniques can also be useful in photography. Liu et al. (2010)[2] proposed an approach for changing the composition of objects in a given image in order to improve its aesthetic value, based on rules of thumb from photography such as the rule of thirds.

Motivated by the compelling applications and the challenges related to the problem, researchers have proposed several techniques for automatic retargeting of images, and the topic is still a subject of ongoing investigation. Solutions have been contributed by the computer vision, computer graphics, and human-computer interaction communities. The detection of interesting or salient areas in an image is an important part of computer vision research; a considerable body of work in graphics focuses on creating more compelling pictures; and the human-computer interaction community has interest in exploring novel types of interaction for retargeting images, as well as evaluating the effectiveness of retargeting algorithms in different tasks. All these points are relevant to the retargeting problem.

In this paper, we summarize and categorize recent work from the image retargeting literature. We start by reviewing the problem formulation and presenting the general sequence of steps shared by most approaches. We then summarize and discuss recent papers, categorizing them according to the methodology used. We do not intend to cover every single published paper in the area; however, we aim to provide a comprehensive view of the approaches by sorting them into groups of similar techniques. This review refers to techniques that are applicable to the retargeting of still images; while some approaches have been proposed to deal with video retargeting, those are outside of the scope of this document. However, we point out that some of the techniques we cover extend to both images and video. We intend this review to be useful to researchers and practitioners interested in image retargeting.

2. THE IMAGE RETARGETING PROBLEM

A digital image of size $m \times n$ can be represented by a 2D discrete grid of pixels with m rows and n columns, where each pixel has a value that encodes its color or intensity information. For example, in the case of RGB color images, each pixel is represented by a triplet $[R, G, B]$ corresponding to its red, green, and blue channels. Pixels in gray-level images are represented by a single value that corresponds to an intensity level. The image retargeting problem can be stated as follows. Given an image I of size $m \times n$ and a new size $m' \times n'$, the goal is to produce a new image I' of size $m' \times n'$ that will be a good representative of the image I. As also pointed by Shamir and Sorkine (2009),[3] there is no clear definition or measure to date as to the quality of I' being a good representative of I. In loose terms, they define the three main objectives for retargeting as:

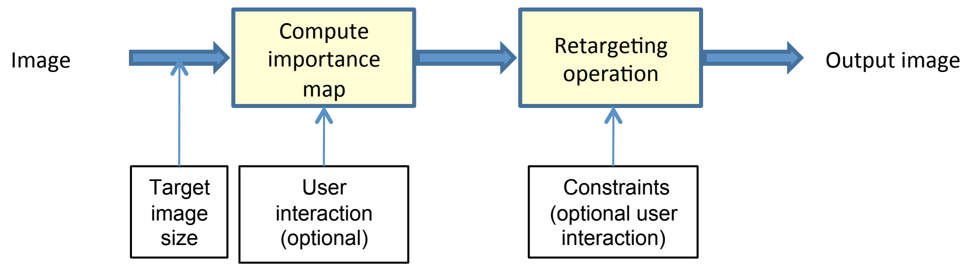

Figure 2. The image retargeting steps. Most techniques follow this general flow.

1. The important *content* of I should be preserved in I'.
2. The important *structure* of I should be preserved in I'.
3. I' should be free of visual *artifacts*.

This is a very challenging problem due to a number of reasons. Real-world scenes exhibit tremendous variability, and the techniques are expected to handle different kinds of imagery, such as barren landscapes and complex scenes with no clear foreground-background segmentation. Images taken outdoors have different overall characteristics from images taken indoors, and the presence of elements such as faces and text also may bring different meanings to specific image regions. Crowded and complex scenes are typical failure cases of current algorithms, due to the issues on automatically determining what is considered important in the image.

The definition of *important* can depend on the specific application being considered. Different papers define different importance measures that specify the level of importance of pixels in the image. Those will be described in Section 3. Also, the definition of what is important and what is unimportant is clearly subjective. While automatic retargeting methods may sometimes lead to impressive results, there are situations where user interaction is unavoidable. Many techniques support the specification of important areas as an input provided by the user, in order to overcome this issue.

Influenced by the three objectives stated above, the majority of image retargeting techniques follows a similar flow (Fig. 2). The inputs are an image and the desired size for the retargeted image. In most works, the target size $m' \times n'$ is smaller than the original size. This is the prevalent situation in applications such as display on mobile devices and thumbnailing. However, a few approaches are also suitable for image enlargement, as we will discuss in Section 4. The first step is the computation of an importance map, which quantifies the importance of every pixel in the image. This typically involves extracting low-level features or applying object detectors, or a combination of the two. Since estimating important regions can be subjective and prone to errors, optional user interaction may be employed for manually specifying the saliency map. The methods may also define constraints, such as mapping lines to lines, and structure preservation, to be satisfied during the retargeting process. A few techniques support user interaction at this point, enabling users to, for example, set the size and scale of important objects in the target image. Finally, given the importance map and the constraints, a retargeting operator is applied to the image, altering its size while taking into consideration the importance map and the constraints.

Keeping the important areas (given by the saliency map) in the result while simultaneously satisfying structural constraints are often contradictory objectives, and the retargeting operators try to find a balance between both. Some techniques[4,5] also proposed an initial classification of the images into categories (such as "containing faces" and "landscape") before initiating the retargeting process. Different algorithms for computing the importance map are then applied depending on the image category being considered.

Real-time algorithms for automatic retargeting are crucial in interactive applications such as dynamic adaptation of web page layout. However, solutions can be computationally expensive. The computation of importance measures may involve the extraction of several features and the use of complex object detection schemes. As the quality of the importance map estimation directly influences the final retargeting result (Fig. 3), there exists a trade-off between choosing a simple importance measure (such as gradient energy) that can be quickly computed, but may sometimes not be accurate (e.g., in low-contrast areas or noisy images), and a more complex measure that takes longer time to calculate. Also, the retargeting problem is often formulated as an optimization problem.

Figure 3. The importance map directly influences the retargeting result. From left to right: input image, gradient-based and context-aware[6] importance maps, and outputs from seam carving using the gradient-based and the context-aware maps. Adapted from Goferman et al. (2010)[6] (©2010 IEEE).

Several different formulations have been proposed, and additional requirements on structure preservation and minimization of artifacts can also differ between methods, requiring solutions with varying levels of complexity.

3. IMPORTANCE MEASURES

A vital component of retargeting techniques is the estimation of where the important regions of the image are located, in order to preserve them in the final image. This is done by assigning a value in the $[0, 1]$ range to every pixel in the image, where higher values mean higher importance. The importance map is typically a *saliency map*, which represents the image areas that draw human attention. Saliency estimation is an important area of research in computer vision — see Goferman et al. (2010)[6] for a good overview of previous methods. In this section, we review the main saliency detection approaches used by retargeting techniques. In Section 4, we classify the retargeting approaches according to their retargeting operators, and refer to the techniques in this section when mentioning the saliency estimation procedure used in each paper.

There are two categories of approaches to automatically estimate saliency: bottom-up methods, and top-down methods. Bottom-up methods are based on low-level features such as edge orientation, color, and intensities, while top-down methods make use of semantic information, such as the locations of important objects (e.g., faces, bodies, and text), structures, and symmetries. Face detectors[7] are popular among retargeting approaches that use top-down methods. Fan et al. (2003)[8] also used a text detector as a component of their top-down saliency. Top-down approaches are often combined with bottom-up saliency to generate the importance map.

A popular approach for computing bottom-up saliency was proposed by Itti et al. (1998).[9] It is inspired by the human visual system, and is based on low-level features: color, intensity, and orientation. A multi-resolution pyramid of the image is built, and significant changes in the features are searched for and combined into a single high-resolution map. Itti's method is so popular that from now on if we simply say some authors used saliency map without giving other references, they used that method.

Stentiford (2003)[10] proposed a method for computing saliency based on dissimilarities between neighborhoods in the image. Their method usually determines larger and smoother salient regions than Itti's method, which tends to result in more focused peaks. Ma and Zhang (2003)[11] introduced a heuristic-based method that analyzes contrast and is more efficient than Itti's method, while leading to similar results for image retargeting.[12] Achanta and Süsstrunk (2009)[13] proposed a saliency measure based on comparing pixels in a blurred version of the image to the average color of the original image in the *Lab* color space, which is useful when salient objects differ in color from the rest of the image.

Harel et al. (2006)[14] proposed a graph-based visual saliency model. It consists of two steps: first forming activation maps on certain feature channels, and then normalizing them in a way which highlights conspicuity and admits combination with other maps. Goferman et al. (2010)[6] proposed a method that, besides finding salient areas, also includes regions near the salient objects that are important to give them context. The method is supported by four basic principles of human visual attention: local low-level considerations, including factors such as contrast and color; global considerations, which suppress frequently-occurring features, while maintaining features that deviate from the norm; visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized; and high-level factors, such as human faces.

Edge maps are also widely used with the goal of preserving prominent objects in the image. The idea is to give high importance to strong contours, and low importance to smooth regions. For this purpose, the L1-norm and the L2-norm (normalized to $[0, 1]$) of the gradient vector at a single pixel can be used, as well as the Canny edge detector.[15] However, strong edges may appear in non-salient areas due to the presence of noise. Other possible measures include Harris corners,[16] histograms of gradients (HoG),[17] and entropy, which are computed from local neighborhoods around a pixel. Wang et al. (2008)[18] demonstrated that combining a gradient map with Itti's saliency map by multiplication has advantages, since Itti's approach eliminates noisy gradients by filtering.

A few methods rely on segmentation to assign saliencies to different regions in the image. Liu et al. (2007)[19] first segment the image into regions and then assign saliencies to each region by considering heuristics such as the region size, position in the image, and relationships between neighboring regions. Hasan and Kim (2009)[4] follow a similar approach for images without faces. Ma and Guo (2004)[20] perform initial segmentation using fuzzy k-means, and compute saliency based on entropy and the relative position and area of the regions with respect to the original image. Setlur et al. (2005)[21] segment the image using mean-shift and assign saliencies to the obtained regions by a combination of bottom-up and top-down features (saliency map and a face detector).

The definition of *important* is subjective, and automatic techniques may fail. To overcome this limitation, some approaches support user interaction to specify the saliency map. Santella et al. (2006)[22] proposed the use of a gaze tracker to estimate the regions where users focus their attention, marking them as high-saliency areas. Avidan and Shamir (2007)[1] suggested that users could scribble on salient areas. Golub (2007)[23] and Gal et al. (2006)[24] proposed manual specification of points of interest, and other papers (for example,[12,21,25]) mention that the saliency map could be manually specified to avoid failures of the automatic algorithms.

4. CONTENT-AWARE IMAGE RETARGETING TECHNIQUES

Straightforward automatic resizing operators such as scaling, fixed-window cropping, and letterboxing (padding a uniformly scaled image with black margins to fit the target display area) work in a content-oblivious way, and often fail to produce satisfactory results. Scaling introduces distortions when the aspect ratio changes, can generate artifacts such as blockiness and aliasing, and can make important objects unrecognizable due to the change in size. Cropping using a predefined criterion (e.g., cropping the center of the image, or the top-left corner, etc.) may fail to include important areas in the result, and letterboxing does not make optimal use of the new image's real estate. To address these issues, techniques that take into account the image's content while resizing attempt to preserve important regions while maintaining an aesthetically pleasing image. In this section, we describe and categorize recently proposed methods for content-aware image retargeting. See Fig. 1 for an illustration of the differences between seam-carving, which is a content-aware operator, cropping, and scaling.

4.1 Content-Aware Cropping

Several techniques for content-aware cropping of images have been proposed. Although these approaches usually do not constrain the size of the output image, allowing the crop window size to vary, they can still be made compatible with the retargeting problem statement (Section 2) by rescaling the cropped image to the target size, or constraining the search for the cropping window to be of the size of the desired output.

A class of techniques search for the cropping window that contains the most important areas (given by an importance map) and satisfies a few constraints. Suh et al. (2003)[26] propose automatically cropping images before scaling them to create thumbnails, with the objective to avoid important image contents from being scaled beyond recognition. Their approach computes a saliency map and searches for the cropping window that maximizes the percentage of salient points inside the window, using a greedy approach. In the case of images with people, they simply crop the regions detected by a face detector. Ma and Guo (2004)[20] present a technique based on initial segmentation by fuzzy k-means, followed by selection of the cropping region that optimizes a cost function based on entropy and the relative position and area of the candidates with respect to the original image. Zhang et al. (2005)[27] propose cropping by optimizing an objective function containing three terms: a composition term based on heuristics motivated by rules of thumb in photography, a conservative term to avoid the picture from being cropped too aggressively, and a penalty term to prevent faces and regions of interest from being cropped. The location of faces and salient areas are the inputs to the model, and the solution is found using particle swarm optimization. Ciocca et al. (2007)[5] first classify the image into a semantic type (landscape, close-up, and "other") using a CART classifier, and then apply different algorithms for cropping the

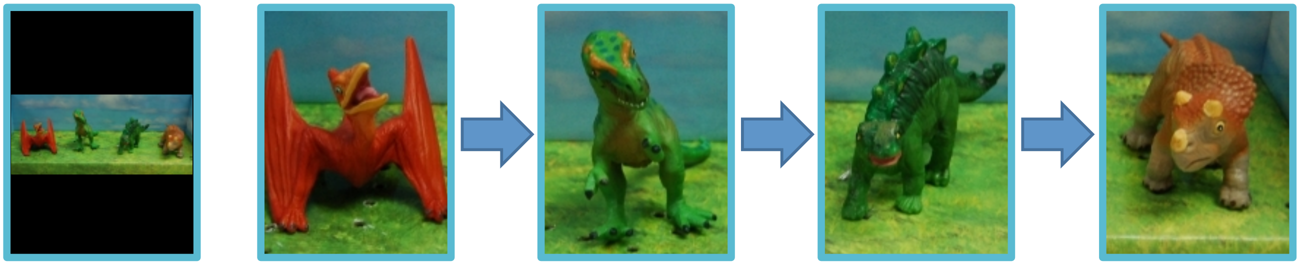

Figure 4. A large image can be presented as a sequence of important subimages.

image depending on the semantic type. Landscape images are not modified, close-up images are cropped based on a saliency map, and "other" images first undergo the application of a face detector, and are then cropped based on saliency if no faces are detected, or based on saliency, skin color, and face regions otherwise. Stentiford (2007)[28] crops images based on a saliency map computed by analyzing similarities between neighborhoods in the image.[10] The cropping window at a given zoom factor and aspect ratio that maximizes the average saliency is chosen as the result. Amrutha et al. (2009)[29] find the best crop based on regions of interest obtained from the combination of Itti's and Stentiford's saliency models. Nishiyama et al. (2009)[30] propose a brute-force search of subwindows for the window that maximizes the output of a quality classifier, which measures the aesthetic value of a crop. A technique for quantifying aesthetics is presented, based on calculating a saliency map and clustering the values into regions, extracting features from these regions, and designing a classifier based on photography rules of thumb, along the lines of Ke et al. (2006)[31] and Datta et al. (2006).[32]

Other techniques suggest including user interaction in the process to facilitate the estimation of salient areas. Santella et al. (2006)[22] argue that it is often subjective and difficult to automatically find the important regions in an image, and propose a semi-automatic method for cropping images based on gaze interaction. The image is segmented into regions by color similarity, and an importance map is built based on the output of a gaze tracker. To determine the crop, a cost function is minimized, with the objectives of including regions of interest, avoiding cuts through background objects, maximizing important content area, and placing content at particular locations (such as the center or according to the rule of thirds). Golub (2007)[23] describes a system for semi-automatic image cropping, where the user selects a point of interest in an image and the system suggests a few cropping candidates that place the point of interest according to photography rules-of-thumb. The user then picks the desired cropping and can adjust the zoom level while retaining the interest point at the selected location.

4.2 Scaling Variants

A disadvantage of the traditional scaling operator is that detail is often lost in the resizing process. In order to overcome this, techniques that try to preserve important details during the scaling process have been proposed. Muñoz et al. (2001)[33] introduce an optimal spline-based algorithm for image resizing with arbitrary scale factors. The algorithm tries to minimize loss of information in the least-squares sense. The complexity of the approach is independent of the scale factor used and the method outperforms the standard interpolation technique for image resizing. Samadani et al. (2007)[34] deal with the problem of creating thumbnails from images, where typically a large image is scaled to generate a small image. The authors point out that the traditional process of thumbnail creation (low-pass filtering followed by subsampling) does not preserve noise and blur, and propose a method for creating thumbnails that retains blurry and noisy characteristics from the original images by estimating both and adding them to the final thumbnail.

4.3 Rapid Serial Visual Presentation (RSVP)

Current mobile devices have limited display resolution, and visualizing large images is challenging. In this scenario, a few approaches have been inspired by the Rapid Serial Visual Presentation (RSVP)[35] methodology from the human-computer interaction community. To effectively visualize the image in its entirety on a small display, it trades space for time, by browsing through subareas of the high-resolution image as time progresses (see Fig. 4). While this approach allows more information to be visualized by increasing the number of displayed images over time, techniques that output a single image enable complete visualization at a single instant of time.

Figure 5. Two original images modified by scaling, cropping the main object of interest, and retargeted to include all objects of interest. Adapted from Setlur et al. (2005)[21] (©2005 HIT Lab NZ, University of Canterbury).

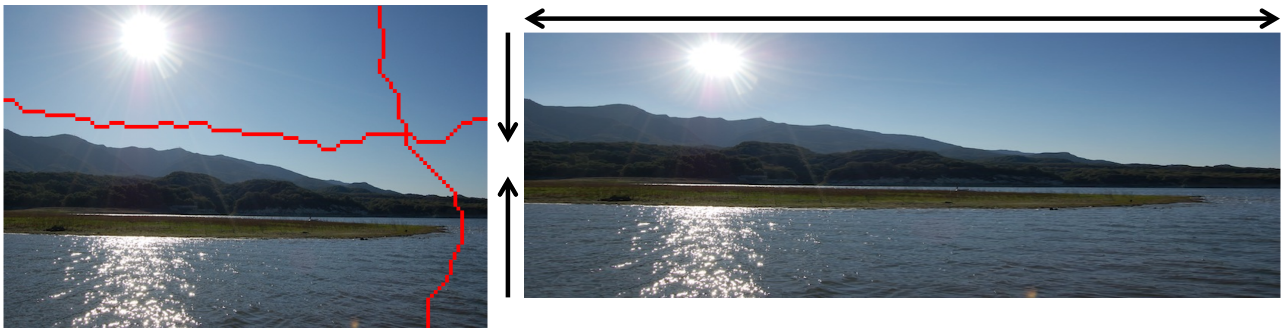

Figure 6. A seam is a connected path of low-energy pixels in an image (left). The output (right) carves out horizontal seams and inserts vertical seams to change the aspect ratio.

Fan et al. (2003)[8] estimate interesting regions by bottom-up (saliency map) and top-down analysis (face and text detectors) and then determine a path for browsing through the image contents. It is given by a sequence of pan and zoom operations, inspired by the RSVP technique. Liu et al. (2003)[36] extends the work to determine an optimal path to maximize the information displayed in the minimum amount of time. Liu et al. (2007)[19] detect regions of interest (ROIs) and sequentially display them, either cropped or rescaled to fit the size of the device. They do not pan while browsing as in the previous works; they simply display the ROIs sequentially, claiming that panning slows down the process. However, this comes at the expense of smoothness and context. To find the ROIs, the saliency of regions determined by segmentation is computed considering heuristics such as size, position in the image, and relationships between neighboring regions. Hasan and Kim (2009)[4] begin with the application of a face detector. If faces are present, they are set as ROIs and every face has the same saliency value. Otherwise, a segmentation-based method similar to the one at Liu et al. (2007)[19] is used. A browsing path of panning and zooming is then determined using a nearest-neighbor approach.

4.4 Segmentation-Based Approach

Setlur et al. (2005)[21] propose a non-photorealistic method for retargeting. It first assigns saliency values to regions obtained by mean-shift segmentation, by combining saliency map and face detection. Regions of interest are determined from the saliency map and the image is either cropped (if all the objects of interest are contained within the target size) or retargeted (otherwise) as illustrated in Fig. 5. The retargeting procedure first removes the ROIs from the image and inpaints the resulting holes to generate a "background" image. The background is scaled to the target size, and the cropped objects are placed back at the same locations. If necessary, they are rescaled to fit inside the target image without overlapping, and so that the background around each object does not differ in color from the corresponding background in the original image (to avoid things like objects that were originally on the grass to be floating in a blue sky). This method relies on accurate segmentation of important objects, and generates distortions. However, it has the ability to retain important areas while discarding unimportant background for scenes with two or more scattered regions of interest.

4.5 Seam Carving

The seam carving technique by Avidan and Shamir (2007)[1] is a popular approach for content-aware image resizing. The general idea is to decrease the image width (or height) one pixel at a time, by removing a seam of minimal importance. A seam is defined as an 8-connected path of pixels (from top to bottom, or from left to right of the image, depending on which dimension is being reduced) that contains only one pixel per row (or column)

(Fig. 6, left). Intuitively, if the importance map is based on gradient energy, the first removed seam will be in a homogeneous area. The image is then readjusted by shifting pixels left or up to compensate for the removed seam, resulting in an image which is one pixel smaller, either on width or height. The image changes only at the seam region, while the other areas remain intact. The authors observe that using gradient energy as the importance map gives satisfactory results, but other importance measures could be used, such as saliency map, entropy, and histograms of oriented gradients.[17] The optimal seams are computed using dynamic programming, and an algorithm for resizing in both dimensions by choosing between optimal vertical or horizontal seams is also presented. The technique can be used for enlarging the image, by finding seams to be removed and duplicating them (Fig. 6, right). It produces impressive results when there are enough low-importance seams to be removed, but creates distortions and artifacts when seams cut through important areas.

After the publication of the seam carving paper, others proposed improvements to the method. Rubinstein et al. (2008)[37] introduce a *forward energy* criterion to deal with the fact that seam carving, despite being an energy removal operation, may actually introduce more energy into the importance map due to previously non-adjacent neighbors becoming neighbors, causing artifacts. The criterion specifies that the optimal seam to be removed is the one whose removal re-introduces a minimum amount of energy. The authors also formulate the problem of finding the optimal seam as a graph cut optimization. The formulation is then extended to video by finding a 2D monotonic and connected manifold in the 3D cube given by width, height, and time dimensions. Cho et al. (2009)[38] propose the use of importance diffusion, which increments the importance of pixels in the importance map that are adjacent to seams being removed. The intuition is that removed seams still provide context information, and importance diffusion helps to preserve this information by adding it to the neighbors of removed seams. Experiments show that the method results in less distortion than seam carving, and even using simple column or row removal with importance diffusion (instead of general seams) can lead to better results.

Other works introduce alternative ways for computing saliency, and demonstrate that seam carving performs better when using the proposed saliency map instead of the gradient energy (L1 norm) used in Avidan and Shamir (2007). Achanta and Süsstrunk (2009)[13] calculate distances of pixels in a blurred version of the image to the average color of the original image in the *Lab* color space. The computation of the forward energy is also modified to consider color (as opposed to gray-scale gradient energies). The authors claim that this saliency measure tends to result in contiguous salient regions, and is more robust to noise due to smoothing. However, in most of their examples the important regions significantly differ in color from the rest of the image. Goferman et al. (2010)[6] propose a method that, besides finding salient areas, also includes regions near the salient objects that are important to give them context. First, a saliency map is determined at a single scale considering local and global distinctiveness, based on color differences and distances between image patches; second, the saliency map is enhanced by analyzing patches for consistency across multiple scales. Finally, pixels are weighted according to proximity to most salient areas to preserve context. Seam carving is presented as an application for the saliency map, and the experiments show that context-aware saliency maps lead to better results due to the preservation of context and the generation of contiguous regions that prevent seams from cutting through important objects.

4.6 Warping-Based Methods

Warping-based methods, sometimes also referred to as continuous methods,[3] perform nonlinear distortion to obtain the resized image. The local distortion of important areas is constrained to be as small as possible, while unimportant regions are allowed to distort more. This way, both important and unimportant areas are kept in the final image, which can be useful for preserving context for the relevant objects. However, depending on the amount of distortion, unimportant areas can even disappear, effectively resulting in content removal. Several methods have been proposed in this category, which make use of different constraints and optimization methods, and can produce smoother results when compared to methods that explicitly remove pixels such as seam carving and cropping.

Liu and Gleicher (2005)[12] introduce a technique that warps the image in a way similar to a fisheye lens, by employing a piecewise linear warping scheme that has more distortion in uninteresting areas and less distortion in the ROI. This method assumes that there is only one ROI per image, and the importance map is computed using a contrast-based method.[11] Gal et al. (2006)[24] propose a method for mapping textures into different surfaces, which avoids distortion of important features. The user manually specifies important areas, and their

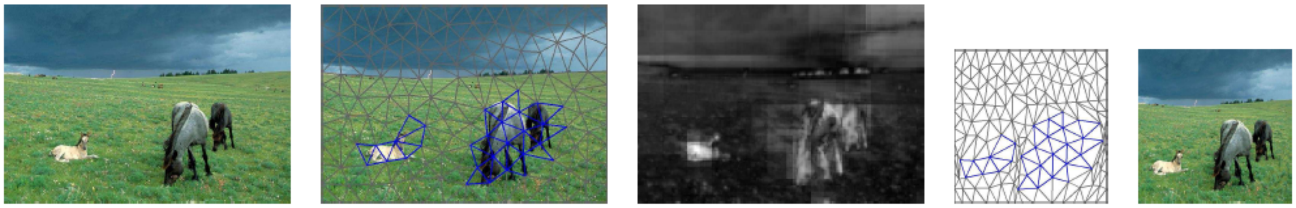

Figure 7. A mesh is built from the input and associated with saliency information (blue edges). The mesh is then resized, and the output image is rendered using texture mapping. Adapted from Guo et al. (2009)[25] (©2009 IEEE).

deformation is constrained to be a similarity transformation, within a Laplacian image editing optimization framework. Wolf et al. (2007)[39] retarget videos to a smaller width or height (images come as a particular case). A sparse system of linear equations is solved to determine the new pixel locations. It is built from constraints that specify where output pixels must be with respect to their neighbors, weighted according to an importance map that combines the L2-norm of the gradient, face detection, and motion detection. In the video case, additional constraints enforce smoothness across adjacent frames. Wang and Lai (2009)[40] extend the method of Wolf et al. (2007)[39] for images by adding constraints to prevent lines from distorting in the final result.

Ren et al. (2008)[41] sample the image at different rates in different regions. Important information is defined by an edge map, computed from the boundaries of the result of mean shift segmentation, the Canny operator,[15] and artificial edges determined using the relative positions of salient regions. The problem of finding the sampling rates that minimize edge distortion is formulated as finding the minimum cost flow in a graph. Kim et al. (2009)[42] first partitions an image into vertical (horizontal) strips based on the complexity of each region, computed as a sum of gradient energy. Each strip is then downsampled based on its frequency content (via Fourier analysis) and scaled according to the target image width (height) using an adaptive scaling method formulated as a constrained optimization problem. The approach is extended to video by using a partitioning scheme intended to maintain temporal coherence.

Other methods explicitly formulate the problem as a label assignment to pixels, which determines the contribution of each pixel to the final result. Ren et al. (2009)[43] formulate retargeting as an integer programming problem, and solve it by linear programming relaxation. The resulting labels are in $[0, 1]$, and are used as weights to combine the pixels into the final result. Their importance map is a combination of bottom-up saliency and face detection. Brand (2009)[44] formulates image retargeting as a problem of integer dynamic programming on trellises of possible image edits. The approach crafts a set of indicator variables defining pixel displacements associated with costs to be optimized; these serve to minimize artifacts and distortions while making the optimization problem convex. It can be used for stretching as well as reducing image size, and it is applicable to video sequences. Kim et al. (2009)[45] retarget images and videos to smaller sizes (either width or height). Importance is given by a weighted average of gradient energy, saliency, and motion. Scaling values in $[0, 1]$ are determined for every pixel such that they sum to the target width (or height). The result is then generated by a weighted average of the input image pixels, similarly to Ren et al. (2009).[43] In the video case, drastic variations in the scaling factors between frames are suppressed to avoid jittering.

Another category of warping-based methods comprises techniques that represent the image as a mesh, and find a nonlinear warping function that resizes the mesh to the target size, typically through local deformations. Constraints are set on the mesh's control points and lines to emphasize important content and minimize distortion. The final image is then rendered based on the warped mesh (Fig. 7). Wang et al. (2008)[18] map the image to a quad grid, which is deformed to match the target image size. The importance map multiplies, for each pixel, an image gradient magnitude map and a saliency map. The optimization process is initialized by a homogeneous resizing, after which a solver iterates two steps that first solves quad and edge scaling factors and then vertex locations. The method retains vertical and horizontal large features pretty well, but diagonal features may get distorted. Pavić and Kobbelt (2010)[46] introduce the concept of two-colored pixels (TCP), where an image is tessellated into quads, and each quad is split by an edge into two regions of constant color such that the TCP best approximates the colors of the original pixels within the quad. For image resizing, the pixels define a quad mesh, which is otherwise similar to the mesh of Wang et al. (2008),[18] but it also has the TCP split edges. The importance function comes from the contrast of a two-colored pixel. The TCP edges align with important feature

Figure 8. Retargeting the input image (left) to a smaller image using the patch transform preserves the global context. Adapted from Cho et al. (2008)[51] (©2008 IEEE).

edges, preserving slanted line features better than Wang et al. (2008), and can be extended to video by using two-colored voxels.

Zhang et al. (2009)[47] estimate a nonlinear warping by minimizing a quadratic distortion energy function defined over a set of control points. The control points include the vertices of a regular mesh grid and a number of selected edge points, and are grouped into small local groups called handles, which are warped using a linear similarity transformation. Constraints are given by the importance map and preservation of edge structure. The total distortion energy function is quadratic and has a closed form minimization solution, avoiding the need for iterative methods. Ren et al. (2009)[48] represent the image as a mesh of curve-edge trapezoids built from mean-shift segmentation boundaries, and warp the mesh's control points. The amount of warping for each trapezoid is determined by an optimization procedure that uses two energy maps: the first, responsible for emphasizing important content, is based on saliency and face detection, while the second, which attempts to minimize visual distortion, is the weighted gradient map from Wang et al. (2008).[18] Guo et al. (2009)[25] represent the image by a mesh associated with importance information, and then transform it to the target image size via a stretch minimizing parameterization scheme. Mesh edge lengths around salient objects are constrained to be rigid, and the remaining edge lengths are computed via a constrained mesh parameterization which uses the multi-dimensional Newton's method with a multi-grid solver at each iteration. The importance map is a combination of a contrast-based method,[11] face detection and body estimation, and optional user input.

Jin et al. (2010)[49] want to retain both linear and curved features of important areas, and use a triangle mesh. They first find important strong geometric elements using the Canny edge detector[15] and the Hough transform, and allow additional feature curves to be specified by the user. Triangles with a large saliency value (Harel et al. (2006)[14]) are also marked as important. The warp is found at interactive rates by solving a sparse linear system, constructed to retain the aspect ratio of salient regions, avoid discontinuities between neighboring triangles, and preserve the shape of the sampled feature points. Laffont et al. (2010)[50] address the problem of interactive zooming by introducing a content-aware zooming operator for high resolution image visualization on small displays. Through interactivity, this method allows the user to control the trade-off between maximizing the amount of visual information on the screen and minimizing the amount of distortion. This method uses an adaptive view-dependent mesh representation. When zooming out, a global view is generated by nonlinearly warping the mesh. When zooming in, the distortions caused by the nonlinear warping are progressively removed to provide a more accurate view of the zoom region.

4.7 Patch-Based Methods

Methods based on image patches achieve retargeting through the manipulation of patches. The algorithms use distances between image patches, aiming to minimize a distance measure between the input image and the retargeted image. Patches are then rearranged to form the final image. The creation of a metric that captures the nuances of similarity between two arbitrary images is still a very challenging problem in computer vision.[52] However, in retargeting the output is a resized version of the input, and this makes the problem more tractable.

Figure 9. Individual retargeting operators have advantages and disadvantages, and combining multiple operators can lead to better results. The original image is retargeted using cropping, scaling, seam carving, and a combination of cropping, seam carving, and scaling.

Simakov et al. (2008)[53] define a bidirectional similarity measure between images, which contains both completeness and coherence measures computed from image patches. Completeness measures whether the target image contains all the visual features present in the source image, while coherence checks that the transformation has not created new visual artifacts that are not present in the source image. For example, cropping creates a fully coherent but not complete target image. To scale an image to a different size or aspect ratio, they scale the image in small steps, and at each step iterate to minimize the error. The method can incorporate importance functions to retain features such as faces or remove objects from the image.

Cho et al. (2008)[51] introduces the patch transform and illustrates its application to a few image editing tasks. The general idea is to break the image into patches and then reconstruct it by rearranging the patches similarly to a jigsaw puzzle, given constraints specified by the user regarding locations of some patches and the size of the final image. In the case of retargeting, this is equivalent to breaking the image and trying to fit the pieces inside a smaller palette without changing the patches' size, and leaving a few patches out, as illustrated in Fig. 8. The process is formulated within a Markov Random Field framework, and solved by belief propagation. Drawbacks of the method are the high computational cost and the inability to preserve locally salient structures well, but an advantage is the preservation of the global context.

Barnes et al. (2009)[54] propose a randomized algorithm for quickly finding approximate nearest neighbor matches between image patches. The algorithm consists of two steps in each iteration: *random search* among a sequence of randomly selected candidates, and *propagation* of the best matches found among the neighborhood patches. It is very fast and works as a good approximation of the nearest neighbor field when few iterations are used, enabling the development of a real-time interactive interface for patch-based image editing applications. For retargeting, the authors extend Simakov's approach[53] by including constraints such as line preservation and new locations of objects and lines, which are implemented by constraining the nearest neighbor search.

Pritch et al. (2009)[55] employ a shift map, specifying the relative shift of every pixel in the output from its source in an input image, effectively removing image segments. They perform a global optimization on the discrete graph representing the output image, where the graph labels indicate the shifts, using a cost function that incorporates data and smoothness terms. The data term includes explicit mapping constraints (e.g., at image boundaries) and saliency information, while the smoothness term penalizes discontinuities in the shift map, and both color and gradient differences. A hierarchical heuristic optimization method is used, which improves speed by several orders of magnitude, though not guaranteeing global optimality. The shift map can be considered as a generalization of seam carving, adding the flexibility to remove larger strips in a single step.

4.8 Multi-Operator Methods

As presented so far, retargeting operators have advantages and disadvantages, and there is no single operator that performs well in every case. A recent trend consists of sequentially applying different operators to the image, in order to capture the best aspects of each (see Fig. 9). In fact, Rubinstein et al. (2009)[56] present a user study that concludes that users generally prefer to combine different retargeting operators (such as cropping, scaling and seam carving) to obtain more pleasing results, rather than relying on the use of a single operator. In this direction, a few approaches have been published. Hwang and Chien (2008)[57] combine seam carving and scaling. Seams are sequentially removed until the energy of the removed seam becomes larger than a threshold. Scaling is then applied to reduce the image to the target size. The importance map used is a combination of low-level saliency, face detection, and the L1-norm of the gradients.

Rubinstein et al. (2009)[56] propose an approach to combine multiple operators (seam carving, cropping and scaling in their implementation). The goal is to find the combination of operators that maximizes the similarity between the input image and the final retargeted image. The optimization is performed using dynamic programming, and it is guided by a patch-based bidirectional image similarity measure. For this purpose, the authors introduce Bi-Directional Warping, a variation of the Dynamic Time Warping algorithm,[58] widely used in speech recognition systems; however, other measures, such as Simakov's,[53] could be used. Both measures are compared, and Bi-Directional Warping is more computationally efficient and order-preserving. An interactive version of the algorithm also allows for users to explore the space of operator combinations. The technique is extended to video by using key-frames and interpolation.

Dong et al. (2009)[59] also combine seam carving and scaling in this order to obtain the final result. An image distance function formulated as a combination of a patch-based bidirectional image Euclidean distance (IMED),[52] dominant color descriptor (DCD)[60] similarity, and seam energy variation is proposed. The technique then searches for the best combination of seam carving and scaling. The optimization is performed by removing one seam at a time, scaling the image to the target size, and computing the distance between the final result and the original image. The result that minimizes this distance is chosen, and it determines the number of seams to be removed and the amount of scaling to be performed after seam carving. The approach is suitable for resizing in both dimensions, as well as for image enlarging, similarly to seam carving.

Liu et al. (2010)[61] introduce the continuous seam carving (CSC) operator and combine it with uniform scaling to generate the final retargeted result. The importance map is first determined by a multiscale contrast-based procedure based on color features. Seam carving is then applied to extract a sequence of seams, and their energy profile is analyzed to determine how many seams are actually to be removed, leaving the remaining reduction in size to uniform scaling. A "reserving ratio map" is constructed based on the relationship between the energy of individual seams and the sum of the energy removed by all seams. This map indicates the contribution of each pixel to the final result, which is created in a continuous space and interpolated to integer coordinates.

5. DISCUSSION AND CONCLUDING REMARKS

While several approaches to image retargeting have been proposed, no single operator completely solves the problem, and automatic retargeting is still a subject of active research. Methods vary in advantages and disadvantages, and the choice of which operator to use depends on the requirements posed by the application at hand. Cropping neither introduces artifacts nor distorts the original structure of the image. However, it may fail to retain all important objects when the image contains scattered important regions or important objects larger than the desired size. Pixels are always discarded, and the image composition can be damaged (or improved, in some cases). In addition, cropping can only eliminate unimportant content at the periphery of the image. Uniform scaling also does not introduce distortions, since the aspect ratio is preserved. However, it can generate artifacts such as blockiness and aliasing, and key objects become less visible, or even unrecognizable due to the change in size. RSVP-based methods are an interesting alternative for visualizing images without distortion on small screens, but they trade space for time, being inappropriate in scenarios where this is not acceptable.

Other approaches nonlinearly manipulate separate regions of the image. While they are able to fit multiple important areas into the target image with minimal distortion, this comes at the expense of greater distortion or removal of less relevant areas, possibly changing the relative distances between objects and the overall photo composition. The segmentation-based method of Setlur et al. (2005)[21] gives interesting non-photorealistic results, but it relies on accurate segmentation and inpainting steps. Seam carving generalizes cropping, and is suitable for preserving important areas free of distortion when there are enough low-importance seams to be removed. However, it produces distortions and artifacts when seams cut through areas of high-frequency content. Warping-based approaches usually do not create discontinuity artifacts like seam carving does. However, important objects can be distorted, smearing artifacts can appear, and the relative proportions of objects can change. Patch-based methods usually preserve the global context of the image, but may introduce local discontinuity artifacts. Some of them can be seen as a generalization of seam carving that allow for removal of segments larger than seams. Given that operators have their advantages and disadvantages, a recent trend consists of combining multiple operators. The idea is to find the best way of combining different operators, aiming to keep the best aspects of each.

Retargeting algorithms can also vary in terms of computational complexity. As discussed in Section 2, there exists a trade-off between speed and quality in the saliency map computation. The techniques may also resort to different formulations of an optimization problem, which require solvers with different levels of complexity. Discrete methods such as seam carving can be solved by dynamic programming or graph cuts. The requirements of warping-based and patch-based approaches range from the solution of linear systems to iterative searches.

Most papers evaluate the results qualitatively on a small and restricted set of images, by visual comparison against previously published techniques with respect to presence of artifacts, distortions, and preservation of important content. A few papers presented user studies that demonstrate the users' preference for the specific approach being proposed. However, objective evaluation of methods remains an open problem. A step toward quantitative assessment was the introduction of bidirectional similarity measures between images (e.g., the works of Simakov et al. (2008)[53] and Rubinstein et al. (2009)[56]). Still, the area is lacking a quantitative assessment of image characteristics, such as artifacts and distortions, that tend to receive better evaluations from users. Finding a similarity metric that takes into account the presence of artifacts and distortions in a perceptual way would allow for a more objective evaluation of the different approaches. This would also support the creation of a standard image dataset for testing, which would incorporate perceptual considerations. However, preferences for some output results over others are generally difficult to quantify, due to the subjective nature of aesthetics.

The subjective element also motivated the development of semi-automatic retargeting techniques where the user is allowed to manipulate the importance map. Other methods allow setting constraints in the final result, such as the scale or size of important objects, or the specification of lines that should be preserved. This is also an important research area, although completely automatic techniques are still required in applications such as dynamic change of layout for web pages. Specific requirements for different kinds of imagery are also not well understood. Some approaches[4,5] attempt to perform a pre-categorization of images into semantic classes, before performing automatic cropping using a different algorithm for each class. Designing retargeting operators using domain-specific information could also be a promising area for investigation.

REFERENCES

[1] Avidan, S. and Shamir, A., "Seam carving for content-aware image resizing," *ACM Trans. on Graphics (Proc. of SIGGRAPH)* **26**(3), 10 (2007).

[2] Liu, L., Chen, R., Wolf, L., and Cohen-Or, D., "Optimizing photo composition," *Computer Graphics Forum (Proc. of Eurographics)* **29**(2), 469–478 (2010).

[3] Shamir, A. and Sorkine, O., "Visual media retargeting.," in [*SIGGRAPH ASIA Courses*], ACM (2009).

[4] Hasan, M. A. and Kim, C., "An automatic image browsing technique for small display users," in [*Intl. Conf. on Advanced Communication Technology*], 2044–2049 (2009).

[5] Ciocca, G., Cusano, C., Gasparini, F., and Schettini, R., "Self-adaptive image cropping for small displays," *IEEE Trans. on Consumer Electronics* **53**(4), 1622–1627 (2007).

[6] Goferman, S., Zelnik-Manor, L., and Tal, A., "Context-aware saliency detection," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], (2010).

[7] Viola, P. and Jones, M., "Rapid object detection using a boosted cascade of simple features," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], (2001).

[8] Fan, X., Xie, X., Ma, W.-Y., Zhang, H.-J., and Zhou, H.-Q., "Visual attention based image browsing on mobile devices," in [*IEEE Intl. Conf. on Multimedia and Expo*], 53–56 (2003).

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

[10] Stentiford, F., "An attention based similarity measure with application to content based information retrieval," in [*SPIE Storage and Retrieval for Media Databases*], **5021** (2003).

[11] Ma, Y.-F. and Zhang, H.-J., "Contrast-based image attention analysis by using fuzzy growing," in [*ACM Intl. Conf. on Multimedia*], 374–381 (2003).

[12] Liu, F. and Gleicher, M., "Automatic image retargeting with fisheye-view warping," in [*ACM Symposium on User Interface Software and Technology*], 153–162 (2005).

[13] Achanta, R. and Süsstrunk, S., "Saliency Detection for Content-aware Image Resizing," in [*IEEE Intl. Conf. on Image Processing*], (2009).

[14] Harel, J., Koch, C., and Perona, P., "Graph-based visual saliency," in [*Proc. of Neural Information Processing Systems*], **19**, 545–552 (2006).

[15] Canny, J., "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **8**(6), 679–698 (1986).

[16] Harris, C. and Stephens, M., "A combined corner and edge detector," in [*Proc. of the 4th Alvey Vision Conference*], 147–151 (1988).

[17] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], (2005).

[18] Wang, Y.-S., Tai, C.-L., Sorkine, O., and Lee, T.-Y., "Optimized scale-and-stretch for image resizing," *ACM Trans. on Graphics (Proc. of SIGGRAPH ASIA)* **27**(5) (2008).

[19] Liu, H., Jiang, S., Huang, Q., Xu, C., and Gao, W., "Region-based visual attention analysis with its application in image browsing on small displays," in [*ACM Intl. Conf. on Multimedia*], 305–308 (2007).

[20] Ma, M. and Guo, J. K., "Automatic image cropping for mobile device with built-in camera," in [*IEEE Consumer Communications and Networking Conference*], 710 – 711 (2004).

[21] Setlur, V., Takagi, S., Raskar, R., Gleicher, M., and Gooch, B., "Automatic image retargeting," in [*ACM Intl. Conf. on Mobile and Ubiquitous Multimedia*], 59–68 (2005).

[22] Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., and Cohen, M., "Gaze-based interaction for semi-automatic photo cropping," in [*ACM Conf. on Human Factors in Computing Systems*], 771–780 (2006).

[23] Golub, E., "Photocropr: A first step towards computer-supported automatic generation of photographically interesting cropping suggestions," tech. rep., University of Maryland HCI Lab (2007).

[24] Gal, R., Sorkine, O., and Cohen-Or, D., "Feature-aware texturing," in [*Proc. of Eurographics Symposium on Rendering*], 297–303 (2006).

[25] Guo, Y., Liu, F., Shi, J., Zhou, Z.-H., and Gleicher, M., "Image retargeting using mesh parametrization," *IEEE Trans. on Multimedia* **11**(5), 856–867 (2009).

[26] Suh, B., Ling, H., Bederson, B. B., and Jacobs, D. W., "Automatic thumbnail cropping and its effectiveness," in [*ACM Symposium on User Interface Software and Technology*], 95–104 (2003).

[27] Zhang, M., Zhang, L., Sun, Y., Feng, L., and Ma, W., "Auto cropping for digital photographs," in [*IEEE Intl. Conf. on Multimedia and Expo*], (2005).

[28] Stentiford, F. W. M., "Attention based auto image cropping," in [*Workshop on Computational Attention and Applications, ICVS*], (2007).

[29] Amrutha, I. S., Shylaja, S. S., Natarajan, S., and Balasubramanya Murthy, K. N., "A smart automatic thumbnail cropping based on attention driven regions of interest extraction," in [*Intl. Conf. on Interaction Sciences*], 957–962 (2009).

[30] Nishiyama, M., Okabe, T., Sato, Y., and Sato, I., "Sensation-based photo cropping," in [*ACM Intl. Conf. on Multimedia*], 669–672 (2009).

[31] Ke, Y., Tang, X., and Jing, F., "The design of high-level features for photo quality assessment," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], 419–426 (2006).

[32] Datta, R., Joshi, D., Li, J., and Wang, J. Z., "Studying aesthetics in photographic images using a computational approach," in [*European Conf. on Computer Vision*], 7–13 (2006).

[33] Muñoz, A., Blu, T., and Unser, M., "Least-squares image resizing using finite differences," *IEEE Trans. on Image Processing* **10**(9), 1365–1378 (2001).

[34] Samadani, R., Lim, S., and Tretter, D., "Representative image thumbnails for good browsing," in [*IEEE Intl. Conf. on Image Processing*], (2007).

[35] de Bruijn, O. and Spence, R., "Rapid serial visual presentation: a space-time trade-off in information presentation," in [*Proc. of Advanced Visual Interfaces*], 189–192 (2000).

[36] Liu, H., Xie, X., Ma, W.-Y., and Zhang, H.-J., "Automatic browsing of large pictures on mobile devices," in [*ACM Intl. Conf. on Multimedia*], 148–155 (2003).

[37] Rubinstein, M., Shamir, A., and Avidan, S., "Improved seam carving for video retargeting," *ACM Trans. on Graphics (Proc. of SIGGRAPH)* **27**(3), 1–9 (2008).

[38] Cho, S., Choi, H., Matsushita, Y., and Lee, S., "Image retargeting using importance diffusion," in [*Intl. Conf. on Image Processing*], 977–980 (2009).

[39] Wolf, L., Guttmann, M., and Cohen-Or, D., "Non-homogeneous content-driven video-retargeting," in [*IEEE Intl. Conf. on Computer Vision*], (2007).

[40] Wang, S.-F. and Lai, S.-H., "Fast structure-preserving image retargeting," in [*IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*], 1049–1052 (2009).

[41] Ren, T., Guo, Y., Wu, G., and Zhang, F., "Constrained sampling for image retargeting," in [*IEEE Intl. Conf. on Multimedia and Expo*], 1397–1400 (2008).

[42] Kim, J.-S., Kim, J.-H., and Kim, C.-S., "Adaptive image and video retargeting technique based on Fourier analysis," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], 1730–1737 (2009).

[43] Ren, T., Liu, Y., and Wu, G., "Image retargeting based on global energy optimization," in [*IEEE Intl. Conf. on Multimedia and Expo*], 406–409 (2009).

[44] Brand, M., "Image and video retargetting by darting," in [*Intl. Conf. on Image Analysis and Recognition*], 33–42 (2009).

[45] Kim, J.-H., Kim, J.-S., and Kim, C.-S., "Image and video retargeting using adaptive scaling function," in [*European Signal Processing Conference*], (2009).

[46] Pavić, D. and Kobbelt, L., "Two-colored pixels," *Computer Graphics Forum (Proc. of Eurographics)* **29**(2) (2010).

[47] Zhang, G.-X., Cheng, M.-M., Hu, S.-M., and Martin, R. R., "A shape-preserving approach to image resizing," *Computer Graphics Forum (Proc. of Pacific Graphics)* **28**(7), 1897–1906 (2009).

[48] Ren, T., Liu, Y., and Wu, G., "Image retargeting using multi-map constrained region warping," in [*ACM Intl. Conf. on Multimedia*], 853–856 (2009).

[49] Jin, Y., Liu, L., and Wu, Q., "Nonhomogeneous scaling optimization for realtime image resizing," *The Visual Computer (Proc. of CGI)* **26**(6–8) (2010).

[50] Laffont, P.-Y., Jun, J. Y., Wolf, C., Tai, Y.-W., Idrissi, K., Drettakis, G., and Yoon, S.-E., "Interactive content-aware zooming," in [*Proc. of Graphics Interface*], 79–87 (2010).

[51] Cho, T. S., Butman, M., Avidan, S., and Freeman, W. T., "The patch transform and its applications to image editing," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], (2008).

[52] Wang, L., Zhang, Y., and Feng, J., "On the euclidean distance of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**, 1334–1339 (2005).

[53] Simakov, D., Caspi, Y., Shechtman, E., and Irani, M., "Summarizing visual data using bidirectional similarity," in [*IEEE Conf. on Computer Vision and Pattern Recognition*], (2008).

[54] Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D., "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. on Graphics (Proc. of SIGGRAPH)* **28**(3) (2009).

[55] Pritch, Y., Kav-Venaki, E., and Peleg, S., "Shift-map image editing," in [*IEEE Intl. Conf. on Computer Vision*], 151–158 (2009).

[56] Rubinstein, M., Shamir, A., and Avidan, S., "Multi-operator media retargeting," *ACM Trans. on Graphics (Proc. of SIGGRAPH)* **28**(3), 1–11 (2009).

[57] Hwang, D.-S. and Chien, S.-Y., "Content-aware image resizing using perceptual seam carving with human attention model," in [*IEEE Intl. Conf. on Multimedia and Expo*], 1029–1032 (2008).

[58] Sakoe, H. and Chiba, S., "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. on Acoustics, Speech, and Signal Processing* **26**, 43–49 (1978).

[59] Dong, W., Zhou, N., Paul, J.-C., and Zhang, X., "Optimized image resizing using seam carving and scaling," *ACM Trans. on Graphics (Proc. of SIGGRAPH ASIA)* **28**(5), 1–10 (2009).

[60] Manjunath, B., Salembier, P., and Sikora, T., [*Multimedia Content Description Interface*], Wiley, Chichester (2002).

[61] Liu, Z., Yan, H., Shen, L., Ngan, K. N., and Zhang, Z., "Adaptive image retargeting using saliency-based continuous seam carving," *Optical Engineering* **49**(1) (2010).