Journal of Computational and Applied Mathematics 236 (2011) 704–713

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Multiple-cue saliency measurement and optimized image composition for image retargeting☆

Xiaonan Luo [a,b,c,d], Zhongming Zhao [a,b,c,d], Zhuo Su [a,d,*], Yun Liang [a,d]

[a] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, 510006, China
[b] Research Institute of Sun Yat-sen University in Shenzhen, Shenzhen, 518057, China
[c] Shenzhen Key Laboratory of Digital Living Network and Content Service, Shenzhen, 518057, China
[d] National Engineering Research Center of Digital Life, Guangzhou, 510006, China

ARTICLE INFO

Keywords:
Image retargeting
Visual saliency
Image feature
Spline interpolation
Region extraction

ABSTRACT

We present a novel image retargeting approach for high-definition displaying based on multiple-cue salient region extraction and optimized image composition, which allows resizing an image into user-specified aspect ratios while preserving prominent visual salient image information. Firstly, we sum up six properties for measuring the salient objects and apply them to region extraction. Then, according to our optimized rules of image composition, we evaluate the aesthetic scores with different extracted sub-images. Finally, we sort out the target image with the highest aesthetic score and retarget it to fit the final scale of displaying. The experimental results show the performance and effectiveness of our approach.

1. Introduction

With the popularity of high-definition display devices, such as large screen televisions, widescreen LCD monitors, MP4s and so on, the scales of media display devices are becoming more and more diversified and personalized. High-definition images or videos are usually displayed on screens of different sizes. Therefore, how to retarget images to fit such a variety of display screens while preserving and highlighting the prominent objects and their relative positions in the original image becomes a significant problem. The retargeting problem can be stated as follows [1]. Given an original image I with the size $I_w \times I_h$, we want to retarget it to a new size $I'_w \times I'_h$ such that the original image can be well represented. The goal of retargeting is to preserve maximally aesthetic effects while minimizing the distortions of the prominent objects and regions. Thus, the problem of image retargeting can be defined as

$$f : I \to I', \quad f(I_w, I_h) = (I'_w, I'_h), \tag{1}$$

where f is a mapping from the original image to the retargeted image, and it must satisfy the following constraint condition:

$$\min \sum_{i=1}^{n} D(S_i), \quad i \in N^+, \tag{2}$$

☆ This work was supported by the NSFC–Guangdong Joint Fund (U0735001, U0835004, U0935004), the National Project of Scientific and Technical Supporting Programs (Grant No. 2007BAH13B04), and the National Key Basic Research and Development Program of China (973) (Grant No. 2011CB302204).

* Corresponding author at: School of Information Science and Technology, Sun Yat-sen University, Guangzhou, 510006, China. Tel.: +86 020 39943199.
E-mail address: suzhuoi@gmail.com (Z. Su).

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.06.022

where $D(S_i)$ denotes the distortion of salient objects of S_i and N^+ denotes the positive integers. The following are the main principles currently used for retargeting [1]:

(1) The important information of the original image should be preserved.
(2) The basic structure of the original image should be preserved.
(3) Visual artifacts should be as few as possible.

Recently, a large number of non-homogeneous image retargeting approaches have appeared and the main technologies can be classified as mesh-based [1–3], seam-based [4,5], cropping-based [6–8], and patch-based approaches [9,10] (and so on). Meanwhile, the image-based video retargeting problem [11–13] is also attracting more and more attention.

In fact, either for static image or for dynamic video retargeting, the necessary operator has to distinguish the salient objects from the original image. This operator can be considered as a quantitative process, from subjective perception to objective measurement. And the result of retargeting depends heavily on the value of the saliency measurement. In our work, we present a novel saliency measurement approach based on multiple cues and use it for salient image region extraction. In addition, we present a numerical computation approach with the rules of optimized image composition based on the aesthetic principle.

On the basis of our above approaches, we convert the retargeting problem into that of extracting the salient objects of the original image while keeping the important image information as far as possible. This can be defined as follows:

$$I_{\text{sub}} = \sum_{k=1}^{m} \sum_{(i,j)\in A_k} f(i,j), \tag{3}$$

where A_k, $1 \le k \le m$, are the areas of the kth salient regions. In the next section, we use these areas to compute the best threshold for region extraction. In this way, we establish a novel aesthetic measurement for objectively measuring the aesthetic quality of the extracted sub-image compositions. To summarize, our main contributions are that we:

(1) Present some properties for measuring the salient objects for region extraction.
(2) Present and implement the rules for how to select the salient objects such that the distribution of extracted regions follows the distributions of dominant points and dominant lines.
(3) Combine multiple-cue salient region extraction with the rules for optimized image composition in order to automatically retarget a high-definition image.

The rest of the paper is organized as follows. In Section 2, we describe some properties as cues for measuring the saliency and use them for region extraction. In Section 3, we introduce the rules of image composition for calculating the aesthetic scores of extracted samples. In Section 4, we describe the implementation details of two extracting schemes for the sub-image and give our experimental results. Our conclusion and perspectives are given in Section 5.

2. Multiple-cue saliency measurement and region extraction

The retargeting operation depends on the importance of the image content for determining the information to preserve or not preserve. Actually, the importance of the image content can be considered as the degree of human visual attention. However, the visual attention belongs to a subjective factor, which cannot be computed in the computer. The saliency measurement operator uses some heuristic schemes to quantify the subjective factor to an objective numerical value. Therefore, it can divide the image content into different numerical levels according to the degree of importance in human vision. The process of measurement does not depend on user interaction; therefore, it make sense to use homogeneous image retargeting that requires more automatic computation. Combining multiple image features as cues, we present a novel saliency measurement approach and use it to extract salient image regions. Our saliency measurement includes two parts: one is region measurement and the other is line measurement. The former reflects the distribution of the salient objects and the latter describes the importance of the straight lines. Section 2.3 details how the saliency measurement works in the region extraction.

2.1. Salient region measurement

Salient region measurement has been extensively studied in the past few years, and used in face detection [14,15] and object recognition. These approaches to salient region measurement are usually based on low-level intensity and contrast information, the combination of high-level image content and scene understanding, and establishing computable models by means of heuristic rules and strategies [3]. Ma et al. [16] presented a contrast-based visual attention model using a series of heuristic rules for human vision. Itti et al. [17] presented a model by using a multi-scale filter to obtain image color, intensity and orientation information, and then detected the regions that have differences from their surroundings. Harel et al. [18] presented a novel visual saliency model based on graph theory. In our approach, we adopt an open source saliency

Fig. 1. An example of hybrid saliency detection and segmentation. (a) Input image. (b) GBVS saliency map. (c) Gradient magnitude. (d) Multiple-cue saliency measurement. (e) Region extraction via saliency measurement.

measurement toolbox called Graph-based Visual Saliency (GBVS[1]) [18] to obtain the saliency map from the original image. The algorithm first computes the feature maps M, e.g., by means of linear filtering followed by some elementary nonlinearity, and then followed by activation, normalization and their combination. In GBVS, the dissimilarity between two nodes is defined as follows [18]:

$$\text{dissim}((j,k)||(p,q)) := \left|\log \frac{M(j,k)}{M(p,q)}\right|. \tag{4}$$

Now we consider the fully connected directed graph G obtained by connecting adjacent nodes (pixels) of M which are labeled with indices (j,k) and (p,q). The directed edge from node (j,k) to node (p,q) will be assigned a weight as follows [18]:

$$w_A((j,k),(p,q)) = \text{dissim}((j,k)||(p,q)) \cdot \exp\left(-\frac{(j-p)^2+(k-q)^2}{2\sigma^2}\right), \tag{5}$$

where σ is a free parameter. Therefore, the weight of the edge from node (j,k) to node (p,q) is proportional to their dissimilarity and inversely proportional to the exponential of the square of their closeness in M. Then a Markov chain can be defined on G by normalizing the weights of the outbound edges of each node to 1, and drawing the equivalence between nodes and states, and edge weights and transition probabilities. The result is an activation measure derived from the pairwise contrast.

The goal of "normalizing" is stated as: concentrating mass on activation maps. By mass concentration, another algorithm based on the Markov chain is defined. Beginning with the activation measure B, the new dissimilarity weight of the feature map is re-defined as [18]

$$w_N((j,k),(p,q)) = B(p,q) \cdot \exp\left(-\frac{(j-p)^2+(k-q)^2}{2\sigma^2}\right). \tag{6}$$

Again, normalizing the weights of the outbound edges of each node to 1 and treating the result graph as a Markov chain in order to compute the equilibrium distribution over the nodes, the final result is the saliency measurement for the original image. An example is shown in Fig. 1(b).

In application, it is usually impossible to completely detect the boundaries of salient objects. Therefore, on the basis of a finite difference operator, we introduce the gradient magnitude in order to increase the accuracy of the boundary detection. We formulate this process as follows:

$$F(S) = \alpha I_S(w_A, w_N) + (1-\alpha) I_G(S), \tag{7}$$

where $I_S(w_A, w_N)$ denotes the saliency map of the input image with the above weights w_A and w_N, $I_G = \text{mag}(\nabla I) = ((\partial I/\partial x)^2 + (\partial I/\partial y)^2)^{1/2}$ denotes the gradient magnitude matrix, and $I_G(S) = I_G|S$ is the restriction of I_G to S. In addition, α is a weight for controlling the saliency value and the gradient magnitude. In the experiments, firstly, one must ensure the saliency value and the gradient magnitude to be of the same order of magnitude. And we normalize the α values to obtain weighting factors within 0 to 1. The simplest method is to set $\alpha = 0.5$, for this can balance the saliency map and the gradient magnitude. But in our experiment, we find that a suitable setting is $\alpha = 0.8$, and this can eliminate most of the disturbance caused by the gradient variation. Fig. 1(d) shows the hybrid saliency map which consists of the GBVS saliency map (see Fig. 1(b)) and the gradient magnitude (see Fig. 1(c)).

The recognizing of salient objects is critical to image retargeting. To quantify the salient object measurement, we require the saliency measurement to have the following properties.

2.1.1. Gradient disturbance

The gradient magnitude reflects the greatest changing rate in the image domain, but over-changing will lead to inaccuracy in the recognizing of important objects. So, we expect the gradient disturbance to be as little as possible (see Fig. 1(c) and (d)). The gradient disturbance can be measured from the scale of sets of pixel gradients with a threshold in the specific image

[1] GBVS website: http://www.klab.caltech.edu/~harel/share/gbvs.php.

domain, and the scale is defined by

$$\mathcal{G} = \bigcup_{i=1}^{n} p_i, \quad p_i \in (\nabla I|T) \subset W, \tag{8}$$

where p_i is the ith pixel with a gradient that is distinguished by a specific threshold, T is the threshold and W denotes a designated image domain.

2.1.2. Multiple-object detection

Multiple-object detection is helpful for recognizing multiple important objects and determining the quantity of objects (see Fig. 1(b) and (d)). We regard the multiple objects as a set and define as follows:

$$\bigcup_{i=1}^{n} O_i = O, \tag{9}$$

where O is an object set, O_i, $i = 1, 2, \ldots, n$, denotes the ith detective object. And $O_i \bigcap O_j = \varnothing$ for all i and j, $i \neq j$.

2.1.3. Edge sensitivity

The capacity for edge detection reflects the accuracy of boundary detection. The edge can be described by a Laplacian of a Gaussian (LoG) filter simply, and the result reflects the sensitivity to edges in the saliency measurement. ∇^2 is the Laplacian operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$, and G is the 2D Gaussian function $G(x, y) = \mathrm{e}^{-(x^2+y^2)/(2\sigma^2)}$. We use the following:

$$\begin{aligned}
\nabla^2 G(x, y) &= \frac{\partial^2 G(x, y)}{\partial x^2} + \frac{\partial^2 G(x, y)}{\partial y^2} \\
&= \frac{\partial}{\partial x}\left[\frac{-x}{\sigma^2}\mathrm{e}^{-(x^2+y^2)/(2\sigma^2)}\right] + \frac{\partial}{\partial y}\left[\frac{-y}{\sigma^2}\mathrm{e}^{-(x^2+y^2)/(2\sigma^2)}\right] \\
&= \left[\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right]\mathrm{e}^{-(x^2+y^2)/(2\sigma^2)} + \left[\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}\right]\mathrm{e}^{-(x^2+y^2)/(2\sigma^2)} \\
&= \left[\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4}\right]\mathrm{e}^{-(x^2+y^2)/(2\sigma^2)}.
\end{aligned} \tag{10}$$

2.1.4. Shape completeness

Shape completeness is a basic requirement for preserving the shapes of important objects. A completeness measurement function is defined on the basis of the similarity in color between the original shape and the target shape in a selected region. The function is the following:

$$SC_{\mathrm{sim}} = \frac{1}{N}\sum_{P \subset \mathcal{O}} \min_{Q \subset \mathcal{T}} C(P, Q), \tag{11}$$

where P and Q are extracted patches of specific size in the original shape and target shape, respectively. $C(\cdot)$ is the similarity in color of P and Q. N is the overlap patch on the pixel. In fact, $SC_{\mathrm{sim}} = 0$ means that the original shape and target shape are the same, exactly.

2.1.5. Sharp feature preservation

Sharp feature preservation reflects the recognition ability for preserving details of important objects, especially the sharp edges and corners (see Fig. 1(d)). Because the sharp features are usually reflected by edges, the quantity processing for sharp feature preservation is similar to that for edge sensitivity, but the former emphasizes the ability to preserve detail and the latter emphasizes the sensitivity of edge detection and location.

2.1.6. Location accuracy

Location accuracy reflects the precise positions of the important objects. The relative error ε reflects the difference between the original object and target object and it can be defined as follows:

$$\varepsilon = \|\mathcal{O}_{x,y} - \mathcal{T}_{x,y}\|/\|\mathcal{O}_{x,y}\|, \tag{12}$$

where $\mathcal{O}_{x,y}$ and $\mathcal{T}_{x,y}$ denote the centroids of the original object and the target object, respectively.

Table 1 compares the results obtained by using threshold segmentation, visual saliency, the gradient magnitude and our approach based on the above six properties, where Y means the approach having the corresponding property, and N means it not having that property.

Table 1
Saliency measurement properties.

	Threshold segmentation	Visual saliency	Gradient magnitude	Ours
Gradient disturbance	*N*	*N*	*Y*	*Y*
Multi-object detection	*N*	*Y*	*N*	*Y*
Edge sensitivity	*N*	*N*	*Y*	*Y*
Shape completeness	*N*	*N*	*Y*	*Y*
Sharp preservation	*N*	*N*	*Y*	*Y*
Location accuracy	*N*	*Y*	*N*	*Y*

2.2. Salient line measurement

Straight line measurement is another commonly used approach. The line detection process can be achieved through the edge detecting operator and the Hough transformation. In the process of the salient region detection, we have already run the finite difference operator and obtained the partial derivatives in the X and Y directions. The partial derivative can be expressed as the edge in a binary image style. Assume that the lines are given by [19]

$$f(\lambda, \varphi) := \varphi_1 \cos\theta + \varphi_2 \sin\theta - \rho = 0, \tag{13}$$

where the parameter is set as follows:

$$\lambda = (\rho, \theta) \in \Lambda \subset \mathbb{R}_+ \times [0, 2\pi], \tag{14}$$

$$\varphi = (\varphi_1, \varphi_2) \in \Omega \subset \mathbb{R}^2, \tag{15}$$

and $\Omega = [0, a] \times [0, b]$ is a rectangle and is usually taken as $[0, 1]^2$.

The Hough transform is defined by the integral

$$H(\lambda) = \int_{\Omega} \chi(\varphi)\delta(f(\lambda, \varphi))\mathrm{d}\varphi, \tag{16}$$

where δ is the Dirac delta function and χ is the characteristic function of Ω. The integral will respond only when the pixel φ belongs to the line defined by λ, i.e., $f(\lambda, \varphi) = 0$. The peak of $H(\lambda)$ at λ_0 corresponds to a good match with the line of parameter λ_0. Then the line detection problem is converted into a peak detection problem. Numerically, the integral equation (16) can be approximated as follows:

$$\forall \lambda \in \Lambda_d, H(\lambda) = \sum_{\varphi \in \Psi} \chi(\varphi)\Phi f(\lambda, \varphi), \tag{17}$$

where Λ_d is a discrete element of Λ and Ψ is a discrete element of Ω, and $\Phi(\cdot)$ is a discrete approximation of $\delta(\cdot)$. More specifically, Λ_d and Ψ can be respectively defined as follows ($\Omega = [0, 1]^2$):

$$\Lambda_d = \{u\Delta\rho, u \in [0, Z-1]\} \times \{v\Delta\theta, v \in [0, Z-1]\}, \tag{18}$$

$$\Psi = \{u\Delta\varphi_1, u \in [0, Z-1]\} \times \{v\Delta\varphi_2, u \in [0, Z-1]\}, \tag{19}$$

where Z is the positive integer selected by users and $\Delta\rho = 1/(\sqrt{2}Z)$, $\Delta\varphi_1 = \Delta\varphi_2 = 1/Z$.

To describe the importance of each line obtained above, for each line L_i obtained above we introduce the following number:

$$G(L_i) = (\gamma\|L_i\| + \varphi H_i)/(\gamma + \varphi), \tag{20}$$

where $\|L_i\|$ is the length of L_i restricted to Ω, γ and φ are two heuristically weighted coefficients, and

$$H_i = \sum_{p \in L_i} (|\partial I(p)/\partial x| + |\partial I(p)/\partial y|), \tag{21}$$

where $\sum_{p \in L_i}$ in Eq. (21) means the sum over all the integer points of L_i. The bigger the value of $G(L_i)$, the more important the line L_i, and the darker the color. Fig. 3(b) shows two salient lines in gray and in black.

2.3. Region extraction based on multiple-cue saliency measurement

In this section, we extract the salient regions from the original image by region extraction. This depends on the selection of some thresholds and we will dynamically compute these thresholds on the basis of the ratios between the salient regions and the entire image region. For a given threshold T, we assume that $S_i(T)$, $1 \le i \le n$, are the salient regions obtained by region extraction. Then, the ratio is defined as follows:

$$\Re(T) = \left(\sum_{i=1}^{n} A(S_i(T))\right) \Big/ A(I), \tag{22}$$

where $A(S_i(T))$ denotes the areas of $S_i(T)$ and $A(I)$ denotes the area of the original image.

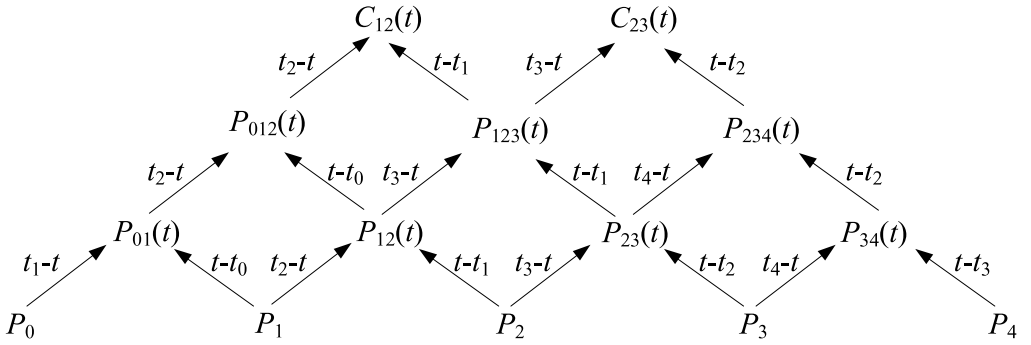

Fig. 2. The cubic Catmull–Rom spline curve algorithm. The first two levels from the bottom were obtained using Neville's algorithm and the topmost level was obtained using the de Boor algorithm.

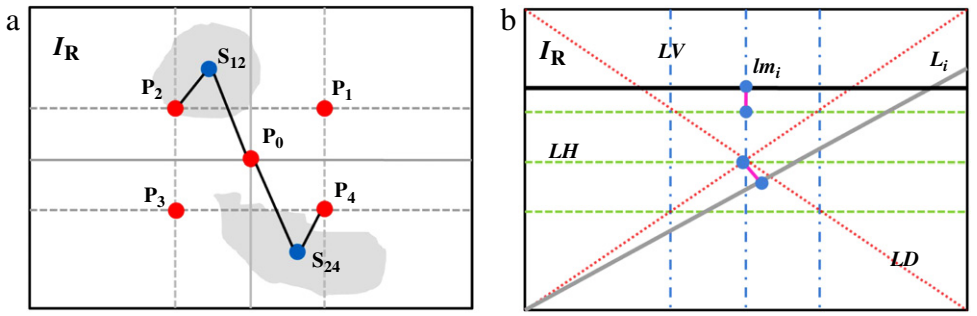

Fig. 3. Aesthetic rules of image composition. (a) The rule of dominant points. (b) The rule of dominant lines.

Next, we find a suitable ratio $\Re_b$ on the basis of Eq. (22). To this end, we first obtain some samples $(\Re_i, T_i),\ 0 \le i \le N$. We use a Catmull–Rom spline [20–22] to interpolate these samples to obtain a continuous curve. The Catmull–Rom spline $C_n(t)$ is defined by [23]

$$C_n(t) = \sum_k N_{k,n-1}(t) P_{k\cdots k+n}(t), \tag{23}$$

where $N_{k,n-1}(t)$ is a B-spline curve obtained by the de Boor algorithm, and we describe it as follows:

$$N_{k,n-1}(t) = \frac{t - t_k}{t_{k+n-1} - t_k} N_{k,n-2}(t) + \frac{t_{k+n} - t}{t_{k+n} - t_{k+1}} N_{k+1,n-2}(t). \tag{24}$$

Also $P_{k\cdots k+n}(t)$ can be obtained by Neville's algorithm for a Lagrange polynomial:

$$P_{k\cdots k+n}(t) = \frac{t_{k+n} - t}{t_{k+n} - t_k} P_{k\cdots k+n-1}(t) + \frac{t - t_k}{t_{k+n} - t_k} P_{k+1\cdots k+n}(t). \tag{25}$$

A cubic Catmull–Rom spline can be calculated by using the pyramid algorithm (see Fig. 2).
After the interpolation, we calculate the most suitable ratio $\Im$ by using the following formula:

$$\Im = \max_{T \in T_{\text{in}}} p(T), \tag{26}$$

where T_{in} is the set of inflection points of p. In this way, we find the optimal threshold at the same time. It is obvious that we can obtain these inflection points by computing the second derivative for the cubic Catmull–Rom spline curve.

3. Optimal image composition rules

Image composition rules usually reflect the positions of objects in an image such that the distribution of these objects follows aesthetic theory. Image composition rules are commonly used in photography and painting. Such rules are usually taught in professional books or courses of photography and guide people as regards how to take excellent photographs. Here below we introduce these rules for image composition and take advantage of them in computing the aesthetic score for image retargeting.

3.1. Aesthetic rules of image composition

Recently, Liu et al. [6] presented an excellent approach for setting these rules. In their approach, they developed a computable model for evaluating the composition aesthetics of a given image based on quantizing several well-grounded composition rules, including rules of thirds, diagonal dominance and visual balance. Their approach summed up the aesthetic rules from a photographic point of view; however, their approach is more empirical and lacks a uniform relationship among these rules. Therefore, some conflicts may exist. In fact, image composition is basically based on the composition of feature points and lines. In contrast, we present some novel composition rules based on image features. Although based on the empirical rules, they abstract the geometric elements from these empirical rules, and formulate the rule of dominant points and lines. The two rules can form a uniform relationship on the basis of the geometric composition. Therefore, they can be used to evaluate the aesthetic scores for the extracted sub-images.

3.1.1. The rule of dominant points

The well-known rule of thirds is that where one divides the image into nine equal parts by means of two equally spaced horizontal lines and two vertical lines; see the gray dotted line in Fig. 3(a). Then, the intersections of these four lines and the center of the original image form a five-point structure, called the dominant points. In the rule of dominant points, we calculate the minimum distance between the centroid of each salient region and these dominant points. This distance is called the dominant point distance of this saliency and denoted by $d(S)$. See Fig. 3(a); the dominant points are marked in red and the centroids of the salient regions (gray region) are marked in blue. At the same time, the whole image is divided into four quadrants by a horizontal solid line and a vertical solid line. A better dominant point image composition means a smaller sum of all the distances of salient regions. In Section 3.2, we will extract a sub-image with the best image composition among a set of sub-images.

3.1.2. The rule of dominant lines

Lines are another important feature of an image, especially straight lines. Actually, the straight lines laid on the image can be divided into three types: horizontal, vertical and diagonal, denoted by LH, LV and LD, respectively. See Fig. 3(b); these three types of line are marked as light green, light blue and light red dotted lines respectively. In our rule of dominant lines, we establish the dominant lines from the thirds lines, the median lines and two diagonal lines of an image. According to the rule of dominant lines, we will calculate the minimum distance between the median point of each salient line L and the dominant lines. This distance is called the dominant line distance of the salient line and denoted by $d(L)$. A better dominant line image composition means a smaller sum of all the distances of salient lines.

3.2. Measurement computation

In order to measure the aesthetics of an image, we use an aesthetic score to identify the aesthetics of the image. In terms of the above mentioned rule of dominant points and rule of dominant lines, the aesthetic scores are determined by the distances of dominant points and dominant lines, respectively. $L = \{LH, LV, LD\}$ denoting the corresponding line is one of three dominant line types (see Fig. 3(b)).

3.2.1. The dominant point score

The aesthetic score V_p of the dominant points is given by the following equation:

$$V_p = \left(\sum_{i=1}^{n} \frac{1}{1 + d(S_i)} F(S_i)\right) \Big/ \left(\sum_{i=1}^{n} F(S_i)\right), \tag{27}$$

where S_i, $1 \le i \le n$, are all the saliencies and $F(S_i)$ is defined by Eq. (7). As shown in Fig. 3(a), after dividing an image into four quadrants, to obtain the dominant point distance of a saliency S, one only needs to calculate the Euclidean distances $d(P_S, P_C)$ and $d(P_S, P_I)$, where P_S is the centroid of S, P_I is the centroid of the original image I, and P_C is the dominant point located in the same quadrant as P_S. In this way, we can reduce the calculation.

3.2.2. The dominant line score

The aesthetic score V_l of dominant lines is given by

$$V_l = \left(\sum_{i=1}^{m} \frac{1}{1 + d(L_i)} G(L_i)\right) \Big/ \left(\sum_{i=1}^{m} G(L_i)\right), \tag{28}$$

where L_i, $1 \le i \le m$, are all the saliencies and $G(L_i)$ is defined by Eq. (20).

Table 2
Time costs of different approaches with the test images in Fig. 4.

	Scaling (s)	Optimized scale-and-stretch (s)	Forward seam carving (s)	Ours (s)
Roadster	0.078	2.759	4.135	1.803
Elephant	0.077	2.859	4.087	1.588
Skiing	0.077	2.795	4.328	1.711
Rider	0.078	2.783	4.572	1.297

3.2.3. The total aesthetic score

The total aesthetic score is defined by a combination of the above aesthetic scores as follows:

$$V = \frac{\omega_p V_p + \omega_l V_l}{\omega_p + \omega_l}, \quad (\omega_p, \omega_l \in [0, 1]), \tag{29}$$

where ω_p and ω_l are the empirical weights, and their values are chosen within 0 to 1. ω_p and ω_l can be set as a constant (e.g. 0.5) or depend on the value of V_p and V_l. Generally, we assure that the sub-images have one salient object at least by means of the region extraction operator. That is, the ω_p value must be larger than 0. However, $\omega_l = 0$ if there is no straight salient line in the image; otherwise $\omega_l \neq 0$. According to Eq. (29), we can choose the best sub-image corresponding to the maximal total aesthetic score.

4. Implementation and experimental results

We implement our approach on a computer with a 2.8 GHz Pentium Duo CPU, 4 GB RAM and MATLAB 2009a. The steps of our approach are as follows. Firstly, we input an arbitrary aspect ratio of an extracting window. Then, we get some samples by using multiple windows. Later, we evaluate the aesthetic scores of dominant points and dominant lines. Finally, we choose the sub-image which has the highest aesthetic score.

4.1. Sub-image extraction

The quality of cropping is decided by the position and the size of the extracted regions. We use two cropping approaches to produce suitable sub-image samples. One is the 8-DoF approach and the other the R-DoF approach. In the 8-DoF approach, the extracting window is separately translated, starting from an initial sub-image with the same center as the input image and at the same step, to the boundary of the image in eight directions. I_w and I_h denote the width and height of the input image, and R_w, R_h and (x_0, y_0) denote the width, height and left upper corner of the initial window, respectively. Generally speaking, the initial window should satisfy $R_w(0) > \frac{I_w}{2}$ and $R_h(0) > \frac{I_h}{2}$. R_w/R_h is the aspect ratio. R-DoF means random DoF, i.e., the left upper corner of the cropping window is determined by a random function.

There are remarkable differences among the extracted sub-images. The salient object is extracted in different ways, producing different image compositions and aesthetics.

4.2. Experiments

Recently, Liu et al. [24] presented an assessment based on a top-down approach that extracted the global geometric structures of input and output images, and established the pixel correspondence. Rubinstein et al. [25] gathered a set of images as benchmarks and conducted a user study to compare some state-of-the-art retargeting approaches. In our experiments, we use some commonly used approaches and results gathered from RetargetMe [25], such as the scaling approach, forward seam carving approach (seam based) and optimized scale-and-stretch approach (mesh based).

The comparisons are given in Table 2, Figs. 4 and 5. Obviously, we can see that Fig. 4(a) and (b) have some serious distortion of the salient objects while Fig. 4(c) has some discontinuities in the image. Our results are given in Fig. 4(d). Table 2 shows the time costs of different approaches. Apparently, the simple scaling costs the least time, but both the optimized scale-and-stretch and the forward seam carving approaches take much more time than our approach. In Fig. 5, we compare our results with those gathered from RetargetMe[2] benchmarks.

5. Conclusions

We present an approach of multiple-cue region extraction and optimized image composition to produce a high-definition retargeting image. On the one hand, our approach can reduce the computing time significantly by using an extracting-based image retargeting approach. On the other hand, on the basis of our image composition rules (the rules of dominant points and dominant lines), our approach can also significantly improve the aesthetic effect of the image. We sum up six properties for evaluating the visual salient objects and present the region extraction. Because both points and lines are essential elements

[2] RetargetMe website: http://people.csail.mit.edu/mrub/retargetme/.

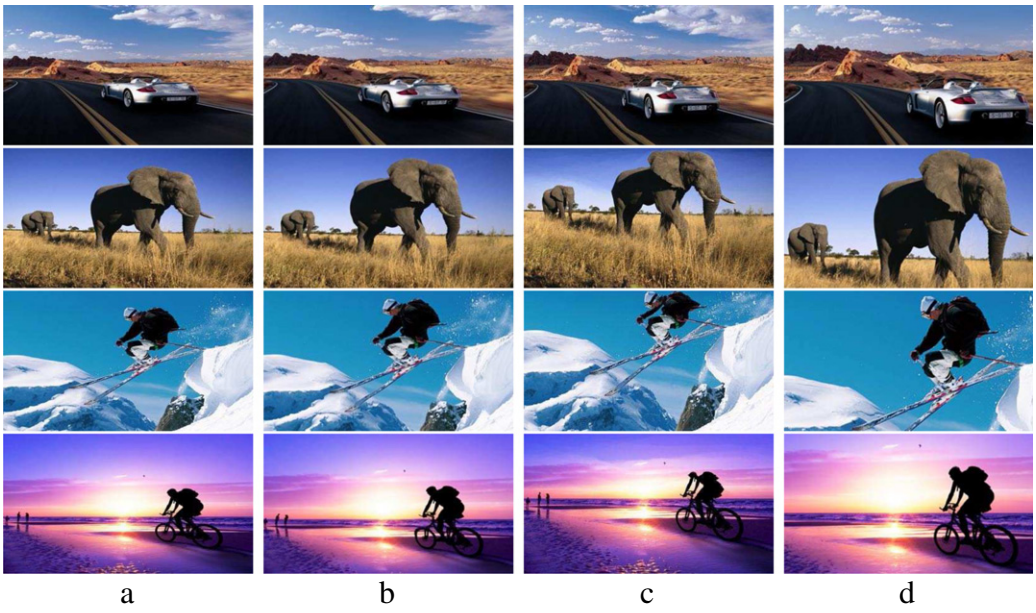

Fig. 4. Comparison of different retargeting approaches. (a) Scaling. (b) Optimized scale-and-stretch [12]. (c) Forward seam carving [1]. (d) Ours.

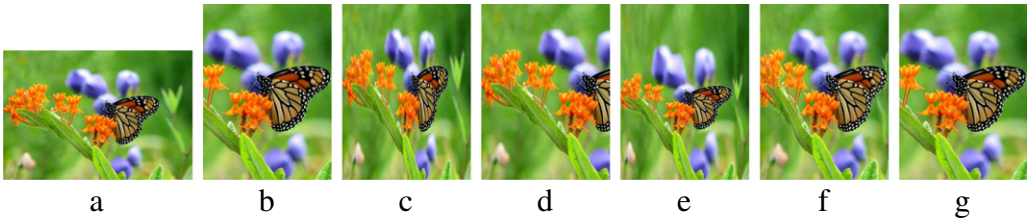

Fig. 5. Comparison of different retargeting approaches using RetargetMe. (a) Input image. (b) Cropping. (c) Scaling. (d) Warping [11]. (e) Optimized scale-and-stretch [12]. (f) Seam carving [1]. (g) Ours.

of an image, their distributions basically reflect the aesthetics of the image. Our image composition rules provide some useful guidelines for retargeting; however, some limitations are also obvious. For example, some critical image information could be lost. In addition, the semantics of the original image could be changed or the global semantics destroyed.

In the future, we will devote attention to improving the efficiency of the algorithm so that it can be used on a video stream in real time and present more image composition rules based on a deeper study of the geometric and color characteristics of an image.

References

[1] A. Shamir, O. Sorkine, Visual media retargeting, in: ACM SIGGRAPH Asia 2009 Courses, 2009, pp. 1–13.
[2] G.X. Zhang, M.M. Cheng, S.M. Hu, R.R. Martin, A shape-preserving approach to image resizing, Computer Graphics Forum 28 (7) (2009) 1897–1906.
[3] Y.W. Guo, F. Liu, J. Shi, Z.H. Zhou, Image retargeting using mesh parameterization, IEEE Transactions on Multimedia 11 (5) (2009) 856–867.
[4] W.M. Dong, N. Zhou, J.C. Paul, X.P. Zhang, Optimized image resizing using seam carving and scaling, in: ACM SIGGRAPH Asia, 2009, pp. 1–10.
[5] M. Rubinstein, A. Shamir, S. Avidan, Multi-operator media retargeting, in: ACM SIGGRAPH, 2009, pp. 1–11.
[6] L.G. Liu, R.J. Chen, L. Wolf, D. Cohen-Or, Optimizing photo composition, Computer Graphics Forum 29 (2) (2010) 469–478.
[7] A. Santella, M. Agrawala, D. Decarlo, D. Salesin, M. Cohen, Gaze-based interaction for semi-automatic photo cropping, in: ACM SIGCHI Conference on Human Factors in Computing Systems, 2006, pp. 771–780.
[8] G. Ciocca, C. Cusano, F. Gasparini, R. Schettini, Self-adaptive image cropping for small displays, IEEE Transactions on Consumer Electronics 53 (4) (2007) 1622–1627.
[9] Y. Pritch, E.K. Venaki, S. Peleg, Shift-map image editing, in: IEEE International Conference on Computer Vision, 2009, pp. 151–158.
[10] C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman, PatchMatch: a randomized correspondence algorithm for structural image editing, in: ACM SIGGRAPH, 2009, pp. 1–11.
[11] L. Wolf, M. Guttmann, D. Cohen-Or, Non-homogeneous content-driven video-retargeting, in: Proceeding of the 11th IEEE International Conference on Computer Vision, 2007, pp. 1–6.
[12] Y.S. Wang, H.B. Fu, O. Sorkine, T.Y. Lee, H.P. Seidel, Motion-aware temporal coherence for video resizing, in: ACM SIGGRAPH Asia, 2009, pp. 1–10.
[13] Y.F. Zhang, S.M. Hu, R.R. Martin, Shrinkability maps for content-aware video resizing, Computer Graphics Forum 27 (2008) 1797–1804.
[14] B.C. Yin, Q. Shi, B. Lei, C.Z. Wang, Improved adaptive principal component extraction algorithm and its application in image feature extraction, Journal of Computational Information Systems 1 (2) (2005) 253–257.
[15] S.Q. Su, B.C. Yin, A robust face detection method, in: Processing of the Third International Conference on Image and Graphics, ICIG'04, 2004, pp. 18–20.

[16] Y.F. Ma, H.J. Zhang, Contrast-based image attention analysis by using fuzzy growing, in: ACM International Conference on Multimedia, 2003, pp. 374–381.
[17] L. Itti, C. Koth, Computational modeling of visual attention, Nature Reviews Neuroscience 2 (2001) 194–204.
[18] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Advances in Neural Information Processing Systems 19 (2007) 545–552.
[19] R.G. von Gioi, J. Jakubowicz, J.M. Morel, G. Randall, On straight line segment detection, Journal of Mathematical Imaging and Vision 32 (3) (2008) 313–347.
[20] G. Wolberg, I. Alfy, An energy-minimization framework for monotonic cubic spline interpolation, Journal of Computational and Applied Mathematics 143 (2002) 145–188.
[21] M. Hu, J.Q. Tan, Adaptive osculatory rational interpolation for image processing, Journal of Computational and Applied Mathematics 195 (2006) 46–53.
[22] D. Qi, K. Djidjeli, W.G. Price, E.H. Twizell, Weighted rational cubic spline interpolation and its application, Journal of Computational and Applied Mathematics 117 (2) (2000) 121–135.
[23] P. Costantini, M. Rossana, An algorithm for computing shape-preserving cubic spline interpolation to data, Calcolo 21 (1984) 295–305.
[24] Y.J. Liu, X. Luo, Y.M. Xuan, W.F. Chen, X.L. Fu, Image retargeting quality assessment, eurographics 2011, Computer Graphics Forum 30 (2) (2011).
[25] M. Rubinstein, D. Gutierrez, O. Sorkine, A. Shamir, A comparative study of image retargeting, in: Proceeding of ACM SIGGRAPH Asia 2010, ACM Transactions on Graphics 29 (6) (2010).