CONSTRAINED SAMPLING FOR IMAGE RETARGETING

Tongwei Ren, Yanwen Guo, Gangshan Wu, Fuyan Zhang*

State Key Lab for Novel Software Technology, Nanjing University, Nanjing 210093, P. R. China

ABSTRACT

In this paper, we present a new approach for retargeting large images to mobile devices with small screens. As the core of image retargeting, information fidelity is adequately considered in terms of reservations of salient regions, edge integrity, and image layout. By taking these aspects as constraints, image retargeting is formulated as a constrained sampling task. Each pixel in image is first represented with a vector encoding the constraints. Then, pixels with the same vector values combine to form blocks, and the original image is thus converted into a graph representation. Thereafter, the sampling ratio of each block is determined with a balanced minimum cost flow algorithm. Final result is generated by an interpolated sampling scheme and direct scaling. Experiments demonstrate the effectiveness of the proposed approach.

1. INTRODUCTION

With the prevalence of digital cameras and scanners, digital images with high resolution can be easily acquired nowadays. It is usually necessary to display the images on mobile devices, such as PDAs and smartphones, for transmission, share, and exchange of information. For the limited screen sizes of such devices, how to quickly browse the original images and grasp important information is crucial to understanding image content. This problem is commonly referred to as *image retargeting*.

Existing methods typically seek solutions to retargeting based on user's attention, and can be roughly classified into two kinds. The first kind [1] [2] [3] realizes navigation among different regions with respect to attention priority. The region with high attention value, i.e., salient region, is first displayed, while the rest parts are displayed serially by interaction or with an optimal path browsing scheme. Apparently, these methods are unsuitable to the rapid browse of large image collections. To overcome this limitation, the second kind tries to exploit the redundancy in image content, and displays the whole image with one-shot operation such as cropping. Nevertheless, direct cropping is helpless to redundancy removal among multiple objects. Other schemes emphasize salient objects with non-uniform deformation. This can be fulfilled by a fisheye like non-linear warping function [4], or using image segmentation and re-composition [5]. Although background can be completely retained, these methods possibly alter relative scales/positions of objects, easily misleading the user. Avian [6] proposed to iteratively carve unnoticeable seams to reduce image size. This method has shown remarkable results, but image composition is not addressed. With the evolvement of carving, strong edges and relative positions of objects may be destroyed.

In this paper, we propose a new approach of image retargeting. Our main observation is that image retargeting essentially can be viewed as a constrained sampling problem. Besides emphasizing salient regions as most previous methods have done, our approach also explicitly considers information fidelity in terms of reservation of edge integrity as well as image layout. We interpret these aspects as constraints of retargeting. Based on the constraints, each pixel in the original image is represented with a vector. The pixels with the same vector value constitute blocks. The original image is then transformed into graph by treating blocks as nodes. We compute the sampling ratio of each block over the graph with a balanced minimum cost flow algorithm. Final result is generated by interpolated sampling, coupled with direct scaling.

Corresponding to: Yanwen Guo. E-mail: ywguo@nju.edu.cn. This project is supported by the National Natural Science Foundation of China (Grant 60533080, 60723003, 60703084) and the Natural Science Foundation of JiangSu Province of China (Grant BG2007037, BK2007571).

2. CONSTRAINTS ON INFORMATION FIDELITY

The core of image retargeting is to reserve information fidelity. A key factor of information fidelity is the reservation of salient regions. Besides, since edges are important visual features, their integrity should be maintained as rigid as possible. Furthermore, image layout, which reflects image composition and spatial relations of objects, cannot be changed. We interpret these aspects as constraints of retargeting.

2.1. Salient region

Retargeting should maintain salient image regions, which can be identified by low-level visual attention model or high-level detector. We first use the method in [7] to compute a normalized attention value for each pixel. The region with high attention value generally corresponds to salient region. As special objects like human faces tend to attract most of user's attention, we also detect faces with face detection algorithm

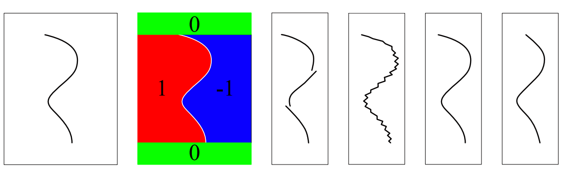
[8], and their attention values are assigned 1. Additionally, user interaction is allowed to designate specific salient objects that cannot be automatically identified.

In our approach, different image regions are dealt with differently, whereas within each region sampling is exerted uniformly. To achieve this, we segment the original image with mean-shift method [9]. The attention value of each region is set to the mean attention of all its pixels.

2.2. Edge integrity

Edges are crucial visual features. Their integrity provides vital clue for evaluation of information fidelity. Such edges come from two parts. One part is the region boundary of mean-shift segmentation. The other is local edges detected by Canny operator. They should be preserved as rigid as possible.

Fig. 1 shows several retargeting effects of an edge (Fig. 1(a)). Random sampling easily causes broken or zigzag effects (Fig. 1(c) and Fig. 1(d)). Hence, we need more restrictive sampling. We construct a trimap (Fig. 1(b)) for the edge, in which the red/blue region is the left/right of the edge, and green is irrelevant to the edge. It is obvious that if the same number of pixels in the red/blue region are removed in each row, the edge will be intactly kept as Fig. 1(e). However, with such strategy, the sampling ratio is restricted when edges increase. To resolve this issue, we adopt an approximate strategy. That is, when sampling horizontally, we ignore horizontal components of edges and only retain vertical components. The vertical sample is performed similarly. In practice, the result (Fig. 1(f)) is still acceptable. With such strategy, we can remove more redundant pixels, and retain edges as rigid as possible.



(a) (b) (c) (d) (e) (f)

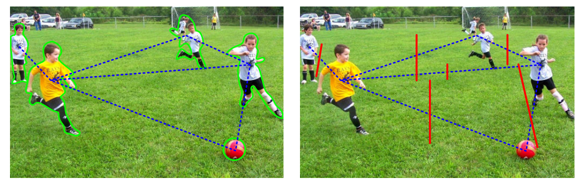
Fig. 1. Retargeting effects using different sampling strategies.

2.3. Image layout

In general, image layout mainly comprises image composition and relative positions of objects.

Image composition means the main components of the image, for example the salient object like people, and less salient sky, water, and land. In our view, completely discarding the less salient regions is infeasible. For instance, when retargeting an image containing the person and scenery, it is judicious to preserve background scenery. In our approach, the sampling ratio of each region is proportional to its attention value. In this way, only those regions with low attention value and extremely small sizes are fully discarded.

Image often contains multiple salient objects, e.g., the photograph of football game. One key point for retargeting such image is to keep the relative positions of objects. Once the attention value is computed for each region, our approach automatically appends some structure edges (Fig. 2) to indicate the relative positions. The structure edges exert the same effect as edges described in the above subsection. During sampling, relative positions will not be changed.



(a) (b)

Fig. 2. Structure edges for indicating relative positions. (a) Relative positions of salient regions; (b) Structure edges (red segments) among objects.

3. CONSTRAINED SAMPLING

In our approach, retargeting is viewed as a constrained sampling process, and can be divided into two independent parts, i.e., horizontal and vertical sampling. We only describe here horizontal sampling. Vertical sampling performs in a similar way.

Basically, horizontal sampling has the following stages. First, represent each pixel with a vector encoding edge constraints. Second, convert the image into graph representation by treating pixels with the same vector value as node, and compute the sampling ratio of each block over graph. Third, sample the image with respect to attention value by interpolated sampling.

3.1. Pixel encoding under constraint

Preserving edge integrity forms explicit sampling constraint. As aforementioned, edges used in our approach include three parts: region boundary resulting from mean-shift segmentation, local edges detected by Canny, and structure edges appended. To reduce computational complexity, only those salient local edges are remained. To achieve this, we calculate an importance value I_e of edge e with:

$$I_e = \sum_{p \in e} (\omega_s \cdot S_p + \omega_g * G_p), \tag{1}$$

where p is the pixel on e. S_p and G_p represent the attention value and gradient at p. ω_s and ω_g denote corresponding weights balancing the actions of attention and gradient (ω_s=ω_g=0.5 in our experiments).

To exploit constraint on edge integrity, we need to decompose the edges into connected components, each of which contains at most one pixel in each row.

For one edge e, a map matrix $\mathcal{M}_e$ with the same size of the image exists by assigning its left pixels 1 (including itself), right -1, and the rest 0 (Fig. 1(b)). Considering all edges, each pixel p can be encoded by a N-dimensional vector $\{V_p(n)|n \in 1, ..., N\}$ satisfying that every $V_p(n)$ is in $\{1,\text{-}1,0\}$, in which N is the number of edges. Obviously, this vector encodes p's position relative to edges. Based on such representation, we define two pixels p_1 and p_2 are *consistent*, if for $\forall n \in N$, $V_{p_1}(n) * V_{p_2}(n) \geq 0$.

To preserve the rectangular shape of retargeted image, a valid sampling should reserve the same number of pixels in each row. Actually, it can be decomposed into a series of basic removing operations, each of which removes one pixel in each row every time. For each edge e, the $\mathcal{M}_e$ values of removed pixels in a basic removing operation that keeps e's integrity can be expressed by a sequence $\{0, ..., 0, 1, ..., 1, 0, ..., 0\}$ or $\{0, ..., 0,\text{-}1, ...,\text{-}1, 0, ..., 0\}$. Considering all edges, it is evident that if each pixel is consistent to its previous one, the basic removing operation will keep integrity of all edges.

Therefore, if the sampling is composed of a series of basic removing operations in which each pixel is consistent to its previous pixel, the sampling will intactly keep edge integrity.

3.2. Graph representation

The pixels with the same vector value cluster into a block (Fig. 3(a)). Analogically, we define that two blocks are *consistent* if their pixels are *consistent*. Apparently, each block has at least one consistent block above it and one below it except the top and bottom blocks. We can therefore construct a graph (Fig. 3(b)) by treating blocks as nodes and connecting adjacent nodes if they are consistent.

It is evident that there exists one path from a top node to a bottom node in graph. Based on the analysis in the above subsection, we thus have the following sampling strategy ensuring that all edges are completely retained. That is, the same number of pixels are removed in each row within every passed blocks over path. In this way, the sampling procedure can be converted into a problem of distributing the number of pixels to be removed on the graph. Regarding removed pixels as flow, we use minimum cost flow algorithm [10] to compute the sampling ratio of each block.

For each block B_k, its valid number of removed pixels in each row is determined by two factors. First, it is limited to the minimum width of all rows $Wid(B_k)_{min}$. Second, we sample B_k according to the attention value S_{B_k}, the bigger S_{B_k} is, the more pixels should be reserved. Hence we calculate a valid removing number related to S_{B_k} using $Num(B_k)_{sal} = (Area_{B_k} - S_{B_k} * Area_{B_k})/h_{B_k}$. Here $Area_{B_k}$ and h_{B_k} represent B_k's area and height separately. The valid number of removed pixels in each block of B_k, named as B_k's capacity, is then computed with,

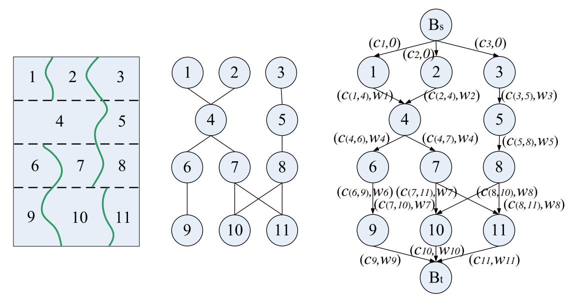
$$c_k = min(Wid(B_k)_{min}, Num(B_k)_{sal}). \tag{2}$$

In addition, B_k's weight is calculated by,

$$w_k = S_{B_k} * h_{B_k}. \tag{3}$$

We then add a virtual source block and a sink one to the graph, and define a unified direction that points downwards as the directions of all arcs (Fig. 3(c)). For each arc in graph, its capacity is the minimum of its two supporting nodes' capacities, and weight is set to its starting node's weight.

Till now, we can use the minimum cost flow algorithm [10] to compute the sampling ratio of each block. For minimum cost flow algorithm is greedy in procedure, it may cause bias in sampling (Fig. 6(b)). To balance flow distribution, we select several minimum cost paths each time and distribute the current flow to each path.



(a) (b) (c)

Fig. 3. Graph representation. (a) Original image; (b) Undirected graph representation; (c) Weighted directed graph for balanced minimum cost flow algorithm, here $c_{(i,j)} = min(c_i, c_j)$.

3.3. Interpolated sampling

Direct uniform sampling will cause broken effect when adjacent blocks have different sampling ratios. We employ here interpolated sampling instead. We perform uniform sampling in the middle row of each block and compute the sampling positions of rest rows by interpolation.

The edges spreading widely may terminate sampling before reaching the assigned width. In this situation, we uniformly sample the horizontal parts that are not fully sampled. Finally, retargeting result is generated by incorporating direct scaling.

4. EXPERIMENTAL RESULTS AND ANALYSIS

We have implemented our approach, and tested the algorithm using several typical images. Fig. 4 shows the results for retargeting an image on screens with various aspect ratios. We also compare our result with previous methods (Fig. 5). The

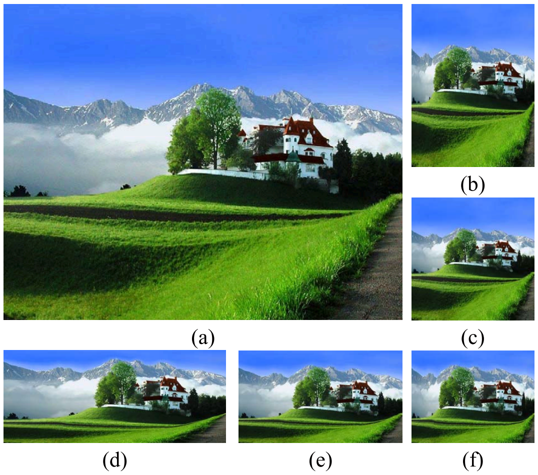


Fig. 4. Retargeting results with different aspect ratios. (a) Original image; (b)-(f) Results with aspect ratios of 3:4, 1:1, 2.4:1, 16:9, 4:3.

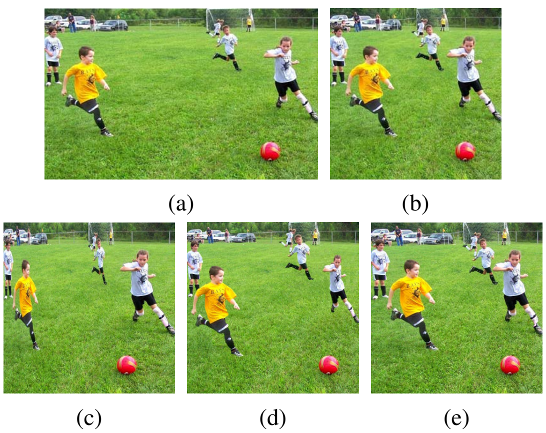


Fig. 5. Comparison of our approach with previous methods. (a) Original image; (b) Our result; (c) Result of fish-eye warping [4]; (d) Result of [5]; (e) Result of seaming carving [6].

methods in [4] [5] alter the relative scales/positions of objects in the undesirable manners (Fig. 5(c) and Fig. 5(d)). Seam carving produces a good result (Fig. 5(e)), but some important edges are broken such as the girl's left leg. Overall our approach attains high information fidelity and shows comparable result (Fig. 5(b)). Fig. 6 verifies that our approach is immune to the horizontal and vertical sampling orders.

5. CONCLUSIONS AND FUTURE WORK

We have presented a new image retargeting approach, which takes aspects of information fidelity as constraints and formulates image retargeting as a constrained sampling problem. Experiments demonstrate the effectiveness of our approach.

Retargeting is intrinsically difficult. There exists limitation to be addressed. Currently, user interaction is sometimes required to specify salient regions. In future, we intend to further automate this procedure with high-level semantic detectors, and extend our approach to video retargeting.

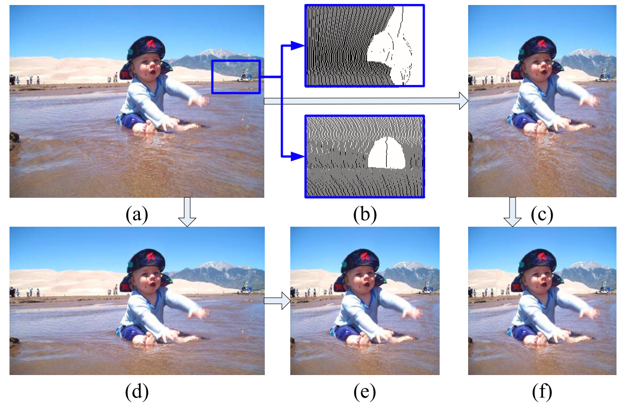


Fig. 6. Retargeting results with different sampling orders. (a) Original image; (b) Sampling map with minimum cost flow (top) and after balance (bottom); (c) and (f) Result of horizontal-vertical sampling; (d) and (e) Result of vertical-horizontal sampling.

6. REFERENCES

[1] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, and H. Q. Zhou, "A visual attention model for adapting images on small displays," *Multimedia Systems*, vol. 9, no. 4, pp. 353–364, 2003.

[2] H. Liu, X. Xie, W. Y. Ma, and H. J. Zhang, "Automatic browsing of large pictures on mobile devices," *ACM Multimedia*, pp. 148–155, 2003.

[3] S. Jiang, H. Liu, Z. Zhao, Q. Huang, and W. Gao, "Generating video sequence from photo image for mobile screens by content analysis," *IEEE ICME*, pp. 1475–1478, 2007.

[4] F. Liu and M. Gleicher, "Automatic image retargeting with fisheye-view warping," *ACM UIST*, pp. 153–162, 2005.

[5] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch, "Retargeting images and video for preserving information saliency," *IEEE CG&A*, vol. 27, no. 5, pp. 80–88, 2007.

[6] S. Avian and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. on Graphics*, vol. 26, no. 3, pp. 267–276, 2007.

[7] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Conf. CVPR*, pp. 511–518, 2001.

[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. PAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[10] M. T. Goodrich and R. Tamassia, *Algorithm Design: Foundations, Analysis, and Internet Examples*, John Wiley & Sons Inc, 2001.