J. Vis. Commun. Image R. 24 (2013) 81–92

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Dynamic distortion maps for image retargeting

Xuejie Zhang [a], Yiqun Hu [b], Deepu Rajan [a,*]

[a] School of Computer Engineering, Nanyang Technological University, Singapore
[b] School of Computer Science & Software Engineering, University of Western Australia, Australia

ARTICLE INFO

Article history:
Received 25 April 2011
Accepted 13 November 2012
Available online 1 December 2012

Keywords:
Image retargeting
Image warping
Visual distortion
Geometrical distortion
Dynamic distortion map
Mesh optimization
Sparse linear system
Multi-level optimization

ABSTRACT

We present an image retargeting method that incorporates image content distortion into a mesh optimization process through the generation of dynamic distortion maps. The warping process is driven by the distortion produced by the warping process itself. We retarget the image through an iterative mesh optimization process to minimize the visual distortion. An adaptive distortion map is iteratively constructed to describe the visual distortion between the original image and the retargeted image. The mesh mapping from the source image to the retargeted image is optimized through an energy minimization process. The objective of the optimization is to allow the distortion produced by the retargeting process to be distributed to smooth and highly textured regions which cause less visual distortion while preserving the geometrical structure of the mesh at regions that may cause distortion after retargeting. Experimental evaluation of the algorithm is done both subjectively and objectively.

1. Introduction

Image retargeting is the process of adaptively resizing an image for devices whose displays may vary in size, aspect ratio and resolution. During this process, the ‘important’ contents in the image should be preserved and minimal distortion should be introduced. With the availability of a wide range of devices with varying display characteristics, image retargeting has become an important practical problem as reflected by recent research efforts. Intuitively, cropping an image (or video frame) based on characteristics such as gradient, saliency and motion (in case of videos) to fit the smaller display size seems to be a straightforward way for retargeting [1–3]. While this might be sufficient in some cases, a more effective content-aware method would be necessary for instances where important regions are not entirely contained within the cropping window. One of the approaches for image retargeting belongs to the class of warping based methods, of which the technique presented in this paper is an example. These methods try to distribute the deformation in a mesh due to resizing, to the unimportant regions of the image through a mesh optimization process [4,5]. The important regions are identified by the analysis of image content, like visual attention detection, object/face detection, feature analysis etc. Warping methods that use motion information have also been developed for resizing video clips [6,7].

Most existing methods that use warping for image retargeting rely on an importance map that determines which contents should be preserved [5,8–11]. In these methods, the amount of warping is solely determined by a *fixed* importance map calculated from the original image. As a result, the warping mechanism does not take into account the distortion in image content produced during the warping process. Such a distortion-aware technique is inherent in the forward-energy based seam carving method which considers the gradient produced by the seam removal or insertion [12].

In this paper, we propose a new framework in which the retargeting process is guided by the distortion produced in the image due to deformation in the mesh. In contrast to most warping based methods that use a fixed importance map, a dynamic distortion map is iteratively updated by comparing the intermediate retargeted result with the input image and by deriving the distortion at each vertex of the rectangular mesh. Another major difference between the proposed framework and other warping-based approaches is the inclusion of distortion in the image content in addition to the geometric distortion of the mesh. It is evident that there is change in image content during the retargeting process and it is important to include it in the energy function for mesh optimization. We also develop a multi-level mesh optimization method in which the constraints for each vertex is defined over its neighbors at different levels to accommodate image distortions at multiple scales. Finally, we include an objective evaluation of the experimental results using the Earth Mover’s Distance (EMD) and the SIFTflow, which have been shown to be mostly correlated with human perception among a set of objective measures [13].

* Corresponding author. Address: School of Computer Engineering, Nanyang Technological University, Block N4, Nanyang Avenue, Singapore 639798, Singapore. Fax: +65 67926559.

E-mail address: asdrajan@ntu.edu.sg (D. Rajan).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2012.11.002

The rest of the paper is organized as follows. Section 2 is an introduction to the related work in the literature. Section 3 gives an overview of our method. Section 4 describes the dynamic visual distortion and geometric distortion energies we used. Section 5 explains the multi-level mesh optimization for image warping. Section 6 shows subjective and objective evaluation results of the proposed method. Section 7 concludes the paper.

2. Related work

Image and video retargeting techniques can be generally divided into two groups – discrete and continuous [14]. Discrete methods remove (insert) content from (to) the original image. These include cropping based techniques that remove contents from boundaries of the image to reduce image size [1–3]. Seam carving removes or inserts connected seams with lowest possible gradient energy [15]. This is extended to video resizing by incorporating forward energy and a framework for removing 2D seam manifolds from 3D space–time volumes [12]. Discontinuous seam carving in both time and space domains has also been developed for video retargeting [16]. The shift-map determines an optimal mapping for each pixel in the retargeted image with respect to the input image [17]. Some part of the input image is not present in the output image due to the shift of pixels. The patch transform divides the input image to patches and the output image is reconstructed by adding/removing/reorganizing the patches subject to resizing constraints [18]. Based on user studies that showed a preference towards combinations of multiple operators such as cropping, scaling and carving, multi-operator approaches have also been developed to mitigate disadvantages of a single operator [19].

Continuous methods optimize a mapping or warping from the input image to the retargeted image. This mapping or warping is a continuous function such that important contents are preserved and unimportant contents are permitted to be distorted. The optimized scale-and-stretch (SNS) method [5] belongs to this class. The SNS method divides the image into quads using a mesh and lets quads with high importance undergo a similarity transformation. The SNS method is extended by combining a novel cropping operator and the warping mechanism [7]. The non-homogeneous content-driven video retargeting method (WARP) utilizes an importance map produced from a combination of gradient, object (face) and motion information to guide the retargeting process [20]. Retargeting is achieved by solving a sparse linear system which is constructed by defining importance based constraints for each pixel in the input image. In [21], the authors analyze the input video or image using saliency, edges and motion to produce an importance map for warping. In [22], an input image is partitioned into vertical strips using gradient information and the retargeted image is obtained by adaptively resizing the vertical strips.

Warping-based methods have the advantage of producing less distortion compared to the discrete methods which removes contents from the image. The gradual distribution of vertices in the mesh preserves the continuity of image features. However, most existing warping or mesh parametrization methods rely on a fixed importance map obtained from analyzing the input image. As the mesh optimization is fully controlled by this importance map, this may cause visual distortion of image content as it is not able to predict the distortion produced by the warping process, especially for some cases of extreme warping. Even when the initial importance map is well defined, the retargeting process itself may introduce new visual distortion since the retargeting process is highly affected by the density and distribution of the intensities in the importance map and the desired output image size.

3. Overview

The proposed method is an iterative strategy in which the deformation of the mesh is guided by the visual distortion caused by the mesh deformation in the previous iteration. Experimental studies have shown that humans are highly sensitive to distortions of specific types of features such as faces, structures and symmetry [13]. Regions that are smooth or highly textured would be visually pleasing even after dramatic warping. On the other hand, structured regions will cause visual distortion since the underlying structure will be easily damaged by the warping process. Hence, the central question to be addressed is how to define visual distortion and to detect such distortions at a certain location in the retargeted image. Note that previous warping-based approaches for retargeting have not considered this phenomenon; instead, they focus only on the geometric distortion of the mesh. Our approach to include visual distortion involves comparison of the original image with the retargeted image. A patch dissimilarity measure together with a context suppression mask is used to derive an adaptive distortion map that guides the retargeting process. The measured distortion is used in a multi-level mesh optimization process to warp the original image to the new dimensions. The optimization of the mesh lets the vertices in the retargeted mesh compete with each other and achieves a stable and optimal distribution of visual distortion.

Given the original image of size $H \times W$ with a uniform rectangular mesh V defined on it, the objective is to estimate an optimal mesh V' on a resized version of the original image so that visual distortion of the image content is minimal. Initially, the original image is scaled down directly to the desired output size. Let V'_0 be the corresponding warped mesh. The visual distortion at each vertex is calculated for corresponding positions in V and V'_0. The total distortion energy that contains both the visual distortion as well as the geometric distortion of the mesh is minimized by solving a sparse linear system for the deformed mesh V'_1, following which the input image is warped according to V'_1. The output image is again compared to the input image and the process repeated till a termination condition – either the vertices move by less than 0.5 pixel or 30 iterations are completed – is attained. Vertices with large distortion tend to maintain its distance to its neighbors while those with low distortion tend to be blended together. Through the iterative process, regions which may cause large distortion after deformation tend to be preserved and the change in distortion map reflects the decrease of distortion in the structured parts of the image.

Fig. 1(a) shows an input image overlaid with a uniform rectangular mesh. This uniform mesh is initially scaled directly to the target size of half its original width, as shown in Fig. 1(b). This scaled mesh needs to be deformed such that the content around its vertices that could possibly cause large visual distortion be kept unchanged. Fig. 1(c) shows the retargeted image in which the foreground giraffe is preserved and the background with smooth or textured contents is deformed. Fig. 1(d) and (e) show the initial and final distortion maps, respectively, in which the initial high distortion in the structured parts is reduced in the final map.

4. Distortion energies

One of the key characteristics of the proposed method is that the mesh deformation is guided by comparison between the original image and the retargeted image. The distortion measure for each vertex is defined as a combination of two energy terms. The first term is the visual distortion based on change in the image content around a vertex before and after retargeting. The second term is the geometric distortion of the mesh structure that also includes