Robust Image Retargeting via Axis-Aligned Deformation

Daniele Panozzo[1,2]   Ofir Weber[3]   Olga Sorkine[2,3]

[1]University of Genova, Italy   [2]ETH Zurich, Switzerland   [3]New York University, USA

Abstract

We propose the space of axis-aligned deformations as the meaningful space for content-aware image retargeting. Such deformations exclude local rotations, avoiding harmful visual distortions, and they are parameterized in 1D. We show that standard warping energies for image retargeting can be minimized in the space of axis-aligned deformations while guaranteeing that bijectivity constraints are satisfied, leading to high-quality, smooth and robust retargeting results. Thanks to the 1D parameterization, our method only requires solving a small *quadratic program, which can be done within a few milliseconds on the CPU with no precomputation overhead. We demonstrate how the image size and the saliency map can be changed in real time with our approach, and present results on various input images, including the* RETARGETME *benchmark. We compare our results with six other algorithms in a user study to demonstrate that the space of axis-aligned deformations is suitable for the problem at hand.*

1. Introduction

Image retargeting resizes an input image to a given target resolution, where the aspect ratio changes. In order to avoid distorting the entire image by homogeneous scaling, or discarding important image parts by cropping, content-aware retargeting techniques were developed. These methods selectively deform the input image into the target dimensions according to a saliency map, preserving the shape of important image components while distorting unimportant background content. A few general methodologies for retargeting were proposed in the recent years, such as discrete carving/shifting, continuous warping and hybrid approaches [SS09]; some algorithms are even available in modern commercial image editing software [Ado10]. To help assess and further improve content-aware retargeting, a number of representative techniques were recently benchmarked and compared in a large-scale user study [RGSS10].

When analyzing the plethora of recent image retargeting approaches and their results, three prominent facts and challenges become apparent: (i) The quality of the resizing results depends immensely on the saliency map. Nearly every approach proposes its own variant of importance map computation, but it is disjoint from the content-aware resizing operator itself and can be easily exchanged. It is evident that a "silver-bullet" automatic saliency detection method does not exist, i.e., each saliency computation technique will fail on some input images, and user-guided importance specification is desired in such cases [SS09]. (ii) Fast or even re-
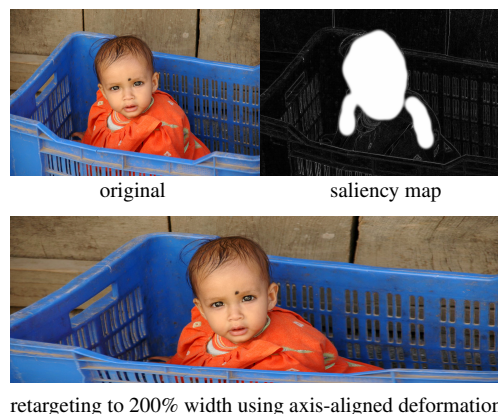

original   saliency map

retargeting to 200% width using axis-aligned deformations

Figure 1: *Retargeting an image to wide format. The importance map was generated using a gradient filter and refined by a few strokes. The computation of the retargeted image took 4 ms, and the overall process took 30 sec of user time.*

altime image resizing methods are extremely valuable, since they are more easily amendable to video retargeting and allow interactive control of the resizing process [KLHG09]. (iii) The robustness of the retargeting operator is another key factor that affects the quality of the results, namely the smoothness, predictability and avoidance of unwanted self-intersections (foldovers) in the resized image. Preventing foldovers appears to be a difficult task [CFK*10], requir-

ing complex, time consuming optimization that is not always guaranteed to be feasible and hence may not be robust.

This work focuses on continuous (warp-based) retargeting, which readily allows controlling the smoothness of the retargeting operator. Discrete approaches such as [RSA08, SCSI08, BSFG09, PKVP09] are excellent at resizing and general editing of images especially rich in texture content, but they are known to sometimes have smoothness artifacts, and are typically slower than warping approaches.

It may seem that a high-quality, foldover-free image retargeting method based on warping cannot run in realtime due to the computational costs involved. In particular, prevention of self-intersection poses a nonlinear constraint in the optimization. The contribution of our work is to show that this does not have to be the case, and even realtime interactive adjustment of the saliency map is possible, if an appropriate space of image deformations is taken for the retargeting operator. Observing the behavior of the state-of-the-art warping methods [KFG09, KLHG09, CFK*10], we notice that they hardly introduce any local rotations in the image deformation. This indeed makes sense, since if the resizing operator contains local, varying rotations, they manifest themselves as "swirls", which are highly noticeable distortion artifacts. Lack of local rotation leads the retargeting deformation to be *axis-aligned*, i.e., the isoparametric lines remain straight and parallel after deformation, only changing the spacing between themselves (see Fig. 2). Our key observation is therefore, that *the space of axis-aligned deformations is the appropriate space for content-aware image retargeting*.

This observation has important consequences. First, the deformation space can be parameterized in 1D, since an axis-aligned deformation is determined by the intervals between the vertical and horizontal isoparametric lines. Previous retargeting methods parameterize the deformations in 2D, leading to optimization problems in the order of M×N unknowns, where M, N are the vertical and horizontal input image resolution, whereas a 1D parameterization necessitates only O(M+N) unknowns. Further, preventing foldovers and controlling the stretching of axis-aligned deformations is simple and robust, since it merely poses linear inequality constraints on the isoparametric line spacing. These constraints can always be posed in a feasible way, contrary to the previously proposed 2D constraints on lack of foldovers.

In this work, we show how to build a complete image retargeting system based on the axis-aligned deformation space. We demonstrate that several deformation energies, meaningful for the content-aware retargeting application, such as the as-similar-as-possible and the as-rigid-as-possible energies, can be effectively optimized in this space while respecting foldover-free constraints. The resulting optimization problems are cast as small quadratic programming (QP) problems thanks to the 1D parameterization, which allows for extremely efficient CPU-based computation using off-the-shelf QP solvers. Our image retargeting
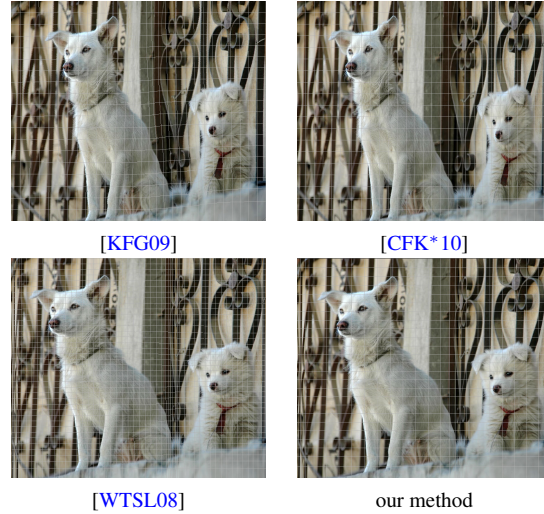

[KFG09]   [CFK*10]

[WTSL08]   our method

Figure 2: Excluding local rotations leads the retargeting deformation to be *axis-aligned*. The *Dogs* dataset is retargeted to 50% width using recent techniques and our method. Observe that although previous approaches do not explicitly enforce it, their deformations are nearly axis-aligned.

prototype runs in real time without requiring any precomputation and enables interactive realtime resizing and editing of the saliency map. Our solution is robust, since we minimize convex energies under feasible constraints, guaranteeing the convergence of the solver and quality of the results. We test our method on the RETARGETME benchmark [RGSS10], showing that results comparable or better than state-of-the-art can be obtained within few milliseconds.

2. Previous work

Warp-based content-aware retargeting methods define an energy functional, and then minimize this functional given the boundary constraints of the target image size. The energy typically measures local deviation of the warp from a shape-preserving deformation such as translation [GSCO06], rigid transformation [KFG09] or similarity [WTSL08, KFG09, ZCHM09, KLHG09], weighted by the importance map. Additional energy terms are introduced to mitigate artifacts such as bending of straight lines [WTSL08, KLHG09, CFK*10, LJW10] or edge blurring [KLHG09]. The energy is discretized over the 2D image domain, usually employing regular grids and finite differences (few works, e.g. [LJW10], use irregular triangle meshes). Earlier warping methods worked with quadratic energies which are highly efficient (only a sparse linear system needs to be solved, and its factorization can be precomputed and reused for arbitrary target image dimensions), but regrettably lead to artifacts, self-intersections and foldovers. Foldovers sometimes even result in "spills" of the retargeted grid outside of the image boundary; the spills are cropped, but the discontinuity in the result remains visible (see our Fig. 3 and Fig. 5

Figure 3: Top: the *Canalhouse* dataset retargeted using [ZCHM09]. Three foldovers occur in the top and bottom parts. Bottom: the foldover-free result of our algorithm.

in [WTSL08]). Similar techniques are used to retarget 3D models [KSSCO08].

Later techniques moved away from the linear least-squares formulations and proposed nonlinear energies and/or inequality constraints to improve the retargeting quality. They prevent self-intersections by iteratively penalizing grid edge flipping [WTSL08], constraining the size of grid cells [KLHG09] or explicitly posing positive scaling constraints on grid cells' transformations [CFK*10]. The method of Chen et al. [CFK*10] offers formal guarantees on lack of self-intersections and vanishing grid cells, but their quadratic program is not guaranteed to have a feasible solution. Nonlinear constrained optimization significantly increases computation time, nearly excluding realtime usage; one exception is the streaming method of [KLHG09] that uses custom tailored multigrid solver and performs all computations on the GPU, achieving impressive framerates at pixel-level discretization. Heavy reliance on graphics hardware may hinder applicability to computationally modest platforms such as mobile devices.

It is interesting to note that, although not obliged to it by the formulation, recent techniques exhibit nearly axis-aligned deformations in practice due to the inclusion of grid line bending penalties [WTSL08, KLHG09]; Chen et al. [CFK*10] even explicitly mention this behavior as an advantage. We directly encode axis-alignment in the deformation space and demonstrate that it not only dramatically reduces the size of the optimization problem but also provides high quality retargeting results with guaranteed robustness.

We refer to [RSA08, SCSI08, BSFG09, PKVP09] for a detailed discussion of discrete retargeting approaches. Discrete methods operate by removing or adding image pixels or patches to change the image's size and in general reshuffle its content (they can be seen as generalizations of cropping or

pasting). While these techniques do not guarantee smoothness and sometimes pose discontinuity artifacts, they are well-suited for images with repetitive or stochastic content that should be removed or duplicated rather than squeezed or stretched. Combining the advantages of discrete and continuous methods is a challenge; the method in [RSA09], albeit computationally expensive, was a successful attempt in terms of the achieved quality and user rankings [RGSS10].

3. Algorithm

We show how to cast content-aware image resizing as a small quadratic program in the space of axis-aligned deformations. The number of variables is linear in the size of the image boundary. The energy we minimize is convex, and finding the global minimum typically takes less than 4 ms.

Denote the input image width and height by W and H. Like most warp-based retargeting methods, we overlay a uniform grid over the image with N columns and M rows; the width of each column (and each cell) in the initial grid is then W/N and the height of each row is H/M. The task is to compute a deformed grid for the resized image, with the desired total width W' and height H'. In the continuous setting, an *axis-aligned deformation* can be fully described by the vertical and horizontal deformation derivatives along the boundary. In our discrete setting, we assume an axis-aligned deformation to be piecewise-linear (linear on each grid cell), such that it is fully determined by the widths of the deformed grid columns and the heights of the deformed grid rows.

Let $\mathbf{s}^{\text{rows}} = (s_1^{\text{rows}}, s_2^{\text{rows}}, \ldots, s_M^{\text{rows}})$ denote the unknown heights of the rows and $\mathbf{s}^{\text{cols}} = (s_1^{\text{cols}}, \ldots, s_N^{\text{cols}})$ the unknown widths of the columns. The axis-aligned deformation is therefore represented by the vector of unknowns $\mathbf{s} = (\mathbf{s}^{\text{rows}}, \mathbf{s}^{\text{cols}})^T \in \mathbb{R}^{M+N}$. The general form of the optimization that computes the retargeted image grid is:

$$\text{minimize } \mathbf{s}^T Q \mathbf{s} + \mathbf{s}^T \mathbf{b} \tag{1}$$

$$\text{subject to } s_i^{\text{rows}} \geq L^h, \quad i = 1, \ldots, M, \tag{2}$$

$$s_j^{\text{cols}} \geq L^w, \quad j = 1, \ldots, N, \tag{3}$$

$$s_1^{\text{rows}} + \ldots + s_M^{\text{rows}} = H', \tag{4}$$

$$s_1^{\text{cols}} + \ldots + s_N^{\text{cols}} = W'. \tag{5}$$

$Q \in \mathbb{R}^{(M+N) \times (M+N)}$ and $\mathbf{b} \in \mathbb{R}^{M+N}$ are determined based on the energy (see Sec. 3.1), and $L^h, L^w > 0$ are the minimum sizes allowed for rows and columns of the deformed grid. The inequalities (2) and (3) *guarantee* that our deformation is free from foldovers, since every cell on the grid cannot be smaller than the specified dimensions L^h-by-L^w and cannot invert (inside each cell the deformation is linear and therefore foldover-free). Eq. (4) and (5) fix the total dimensions of the deformed grid to the desired target size.

For the above quadratic program (QP) to be feasible, we just need $L^h \leq H'/M$ and $L^w \leq W'/N$; simple homogeneous scaling then provides a feasible solution. The feasible domain is bounded, since $\forall i, 0 \leq s_i \leq \max\{H', W'\}$, such that

Figure 4: The *Fatem* image resized with different energies.

the objective function in (1) is finite in the feasible region. The energy should be defined in such a way that Q is positive (semi)definite; our problem is then convex and can be solved with standard QP solvers.

3.1. Energy functions

Image retargeting methods rely on a saliency map $\omega(x,y)$ that assigns an importance value between 0 and 1 to every pixel of the image. Our goal is to compute a deformation that preserves the image in the salient zones as much as possible and concentrates the unavoidable distortion in less important areas. To integrate the saliency map ω in our formulation, we average its values inside every cell of the grid on the original image and we obtain the saliency matrix $\Omega \in \mathbb{R}^{M \times N}$. This per-cell integration of saliency is the proper FEM discretization in our piecewise-linear setting.

We consider two energies in our framework, often employed by prior successful retargeting methods: (1) the As-Similar-As-Possible (ASAP) energy [ZCHM09], which produces deformations that are locally close to similarities, and (2) the As-Rigid-As-Possible (ARAP) energy [KFG09], which penalizes all local deformations except translation and rotation. Fig. 4 shows an example result with these energies.

ASAP Energy. In the space of axis-aligned deformations, a similarity transformation is a combination of uniform scaling and translation, since rotations are eliminated. The ASAP energy thus minimizes non-uniform scaling:

$$E_{\text{ASAP}} = \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\Omega_{i,j} \left(\frac{M}{H} s_i^{\text{rows}} - \frac{N}{W} s_j^{\text{cols}} \right) \right)^2. \quad (6)$$

The two factors M/H and N/W compensate for the aspect ratio of the cells in the original grid.

To minimize this energy using our QP framework, we define the following matrix $K \in \mathbb{R}^{(MN) \times (M+N)}$:

$$K_{k,l} = \begin{cases} \Omega_{r(k),c(k)} \frac{M}{H} & \text{if } l = r(k), \\ -\Omega_{r(k),c(k)} \frac{N}{W} & \text{if } l = M + c(k), \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where $r(k) = \lceil k/N \rceil$ and $c(k) = ((k-1) \mod N) + 1$. From this equation, $K\mathbf{s}$ gives us the vector with energy terms per row, and $E_{\text{ASAP}} = \mathbf{s}^T K^T K \mathbf{s}$. Using the generic notation of Eq. (1), $Q = K^T K$ and $\mathbf{b} = 0$. Clearly, Q is a positive semidefinite matrix, such that the energy is convex.

ARAP Energy. In our axis-aligned deformation space, a rigid transformation is reduced to a translation, since rotations are not allowed by definition. The ARAP energy thus minimizes uniform and non-uniform scaling:

$$E_{\text{ARAP}} = \sum_{i=1}^{M} \sum_{j=1}^{N} \Omega_{i,j}^2 \left((\frac{M}{H} s_i^{\text{rows}} - 1)^2 + (\frac{N}{W} s_j^{\text{cols}} - 1)^2 \right).$$

To minimize this energy using our QP framework, we define the following two matrices $R^{\text{top}}, R^{\text{btm}} \in \mathbb{R}^{(MN) \times (M+N)}$:

$$R_{k,l}^{\text{top}} = \begin{cases} \Omega_{r(k),c(k)} \frac{M}{H} & \text{if } l = r(k) \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

$$R_{k,l}^{\text{btm}} = \begin{cases} \Omega_{r(k),c(k)} \frac{N}{W} & \text{if } l = M + c(k) \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where $r(k) = \lceil k/N \rceil$ and $c(k) = ((k-1) \mod N) + 1$. We also define the vector $\mathbf{v} \in \mathbb{R}^{MN}$, $v_k = \Omega_{r(k),c(k)}$. We can now rewrite the ARAP energy using matrix notation:

$$E_{\text{ARAP}} = \left(\begin{bmatrix} R^{\text{top}} \\ R^{\text{btm}} \end{bmatrix} \mathbf{s} - \begin{bmatrix} \mathbf{v} \\ \mathbf{v} \end{bmatrix} \right)^T \left(\begin{bmatrix} R^{\text{top}} \\ R^{\text{btm}} \end{bmatrix} \mathbf{s} - \begin{bmatrix} \mathbf{v} \\ \mathbf{v} \end{bmatrix} \right). \quad (10)$$

In the generic notation of Eq. (1):

$$Q = \begin{bmatrix} R^{\text{top}} \\ R^{\text{btm}} \end{bmatrix}^T \begin{bmatrix} R^{\text{top}} \\ R^{\text{btm}} \end{bmatrix}, \quad \mathbf{b} = -2 \begin{bmatrix} R^{\text{top}} \\ R^{\text{btm}} \end{bmatrix}^T \begin{bmatrix} \mathbf{v} \\ \mathbf{v} \end{bmatrix}. \quad (11)$$

Again, the form of the Q matrix clearly indicates that it is positive semidefinite, such that the ARAP energy is convex.

Note that even though the intermediate matrices K, R^{top}, R^{btm} have MN rows, they are extremely sparse and fast to construct procedurally. The resulting Q matrices of the QP are square with $M + N$ rows/columns, meaning they are dense but small. Note also that other energies can be similarly formulated in our space of axis-aligned deformations; we have chosen to concentrate on the above two since they are commonly used and typically provide good results. Furthermore, it is possible to linearly combine these two energies to obtain an optional degree of freedom.
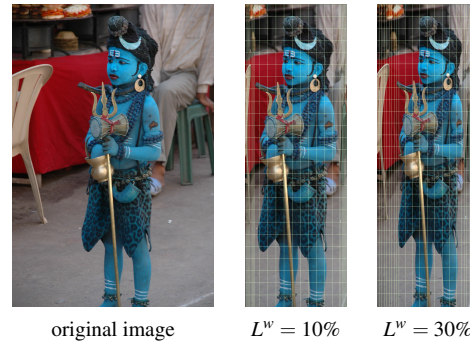

original image $L^w = 10\%$ $L^w = 30\%$

Figure 5: The minimal column width L^w and row height L^h can be prescribed. Middle: each column may not be compressed to more than 10% of the original width. Right: the minimal width is 30%, such that the deformation is less pronounced, since extreme squeezing is disallowed.
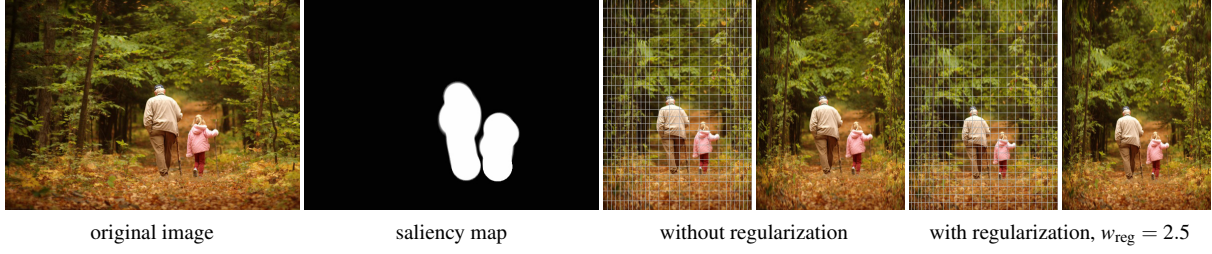
Figure 6: *The effect of the Laplacian regularization. The image on the left is retargeted to 50% width using a manually painted saliency map. As seen in the middle, retargeting using unregularized ASAP energy leads to strong variation in column width. This effect can be mitigated by adding a weighted Laplacian regularization term, as shown on the right.*

Figure 7: *Grid resolution does not have a dramatic effect on the energy minimization result (here, the ASAP energy was used).*

Cropping can be naturally incorporated in our method by setting the parameters L^w and L^h to zero. The optimization procedure is then able to collapse rows and columns, thus cropping the original image.

3.2. Laplacian regularization

We can enrich the energies shown above with a regularization energy that allows to increase the smoothness of the resulting deformation. Laplacian regularization allows to distribute the deformation more evenly across the image, and is particularly useful for manually painted saliency maps, since they tend to concentrate the saliency on distinct parts of the image and fall off abruptly to zero elsewhere (i.e., such saliency maps are highly non-smooth). See Fig. 6 for an example of the effect of the Laplacian regularization.

The Laplacian regularization term is defined as follows:

$$E_{\text{reg}} = \sum_{i=1}^{M-1} \left(\tfrac{M}{H} (s_{i+1}^{\text{rows}} - s_i^{\text{rows}}) \right)^2 + \sum_{j=1}^{N-1} \left(\tfrac{N}{W} (s_{j+1}^{\text{cols}} - s_j^{\text{cols}}) \right)^2.$$
(12)

The regularization penalizes two adjacent rows or columns that have large differences in size. Note that the deformation that minimizes the Laplacian is homogeneous scaling, such that this regularization term can be seen as a way to blend between homogeneous resizing and the ASAP or ARAP deformation, controlled by a weighting factor $w_{\text{reg}} \geq 0$.

To incorporate the regularization term into the QP (1), we simply add the term $\mathbf{s}^T (w_{\text{reg}} L) \mathbf{s}$ to the energy, where L is the standard Laplacian matrix corresponding to Eq. (12). In other words, we add the matrix $w_{\text{reg}} L$ to Q in Eq. (1). The Laplacian matrix is positive semidefinite, such that this energy term does not hurt the convexity of the problem.

Figure 8: *Images of high resolution may benefit from higher-order interpolation when using coarse grids for the retargeting optimization. The input image resolution is 2800×1800 and the retargeting uses a 25×25 grid. Bilinear interpolation leads to some smoothness artifacts (middle), while upsampling to a 100×100 grid using the spline technique described in Sec. 3.3 results in visually smooth interpolation.*

3.3. Cubic B-spline interpolation

The formulations of the ASAP and the ARAP energies, as well as the Laplacian regularization term, are proper linear FEM approximations of the continuous counterparts, such that convergence is expected under uniform grid refinement. We have observed that the results of the optimization are not greatly dependent on the grid resolution (see Fig. 7); this makes sense also because our constrained deformation space, parameterized in 1D, does not admit huge local variation. A coarse grid resolution of 25×25 (i.e, 50 optimization variables) suffices in most cases to faithfully describe the deformation map. However, such a coarse bilinear grid may be insufficient to provide high-quality results for images with very high resolution, because the bilinear interpolation is not smooth across grid lines.

To improve the interpolation results for high-resolution images, we can optionally employ B-spline interpolation to
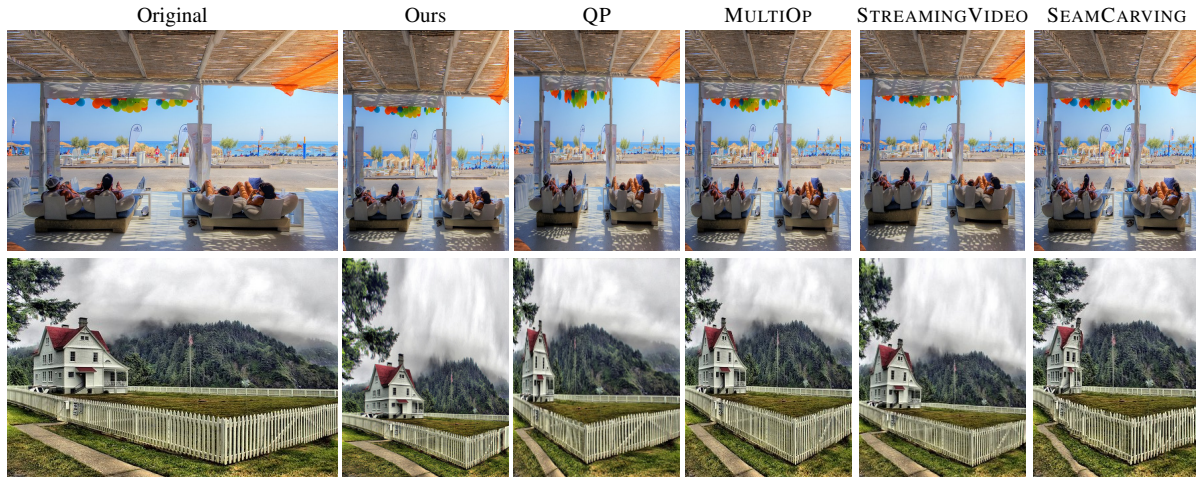
Original	Ours	QP	MULTIOP	STREAMINGVIDEO	SEAMCARVING

Figure 9: *Comparisons with recent image retargeting methods: QP [CFK*10], MULTIOP [RSA09], SV [KLHG09] and SC [RSA08]. The supplemental material provides full comparisons on the* RETARGETME *benchmark [RGSS10].*

upsample the retargeted grid. We define a uniform cubic B-spline using the deformed grid vertices as control points. This can be performed in 1D using two 1D cubic B-splines: one for rows and one for columns. We sample the splines for denser horizontal and vertical positions in order to produce a new grid of arbitrary resolution. The final retargeted image is created by bilinear interpolation on this finer grid.

The deformation described by our deformed grid is guaranteed to always be a bijection and, thanks to the variation diminishing property, the finer grid obtained with the spline is guaranteed to be foldover-free as well. An example of the improvement is given in Fig. 8.

4. Results

We implemented an interactive application for image retargeting that allows in real time to change the image size, adjust the saliency, and tweak the other two parameters of the optimization (the strength of Laplacian regularization and minimal grid cell size). We provide the full comparison with the RETARGETME benchmark [RGSS10] in the supplemental material. The accompanying video shows various recordings of interactive retargeting sessions.

We ran our experiments on a Core2Duo@2.4GHz, using a single core. OpenGL was used to compute the bilinear interpolation; our test system is equipped with a low-end laptop graphics card (the NVIDIA 320M). We employed an easy-to-use and fast QP solver CVXGen [MB10] to solve the QP in Eq. (1). Interactive frame rates (always above 60 fps) were attained in all our experiments. The average computation time to retarget an HD image using a 25×25 grid was 4 ms (250 fps). We observed that the ASAP energy usually produces slightly better results than ARAP, probably because of the additional flexibility of uniform scaling. Our method produces smooth, intersection-free images that pre-

serve the salient features well, and tend to respect the input image structure in general (Fig. 9).

Saliency maps. Our approach can be successfully applied in a fully-automatic mode. We experimented with the automatic saliency detection method of [IKN98] and [CZM*11], although of course any other method can be used. The main advantages of our method, the lack of local rotations and foldovers, are independent of the importance map choice. As discussed earlier, automatic saliency may fail for certain images due to the subjectiveness and ill-posedness of the problem. In such cases, a minimal amount of user intervention can improve the results; fine-tuning by the user is readily enabled by the realtime speed of our approach.

Fig. 10 shows the results computed using the ASAP energy with three different saliency maps. The first image uses an automatic saliency map; the person's head is not detected as salient and is distorted. Our application allows to interactively change the contrast of the saliency map to improve the result (Fig. 10, middle), and to simply paint over it. A single user stroke on the head is sufficient to greatly improve the outcome (Fig. 10, right). This example took 15 seconds of user time. On average, in our experiments we spent 30 seconds per image to paint the saliency map and adjust the parameters. The process is very intuitive, since the user can watch the result of any manipulation in real time.

Interactive, user-controllable parameters. Our method exposes a small numbers of parameter to the user to allow a high degree of control over the final result. Please refer to the video and the executable for demonstration of the realtime tuning effect. The image size can of course be changed interactively. The saliency map can be controlled by adjusting its brightness and contrast, and by manual painting (Fig. 10). The deformation can be controlled by changing the Laplacian regularization weight w_{reg} (Fig. 6). Finally, the minimal grid cell width and height can be selected (Fig. 5). We have

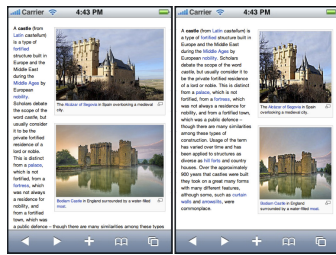| original | auto-saliency | result with auto-saliency | saliency contrast increase | increased contrast, manual marking (head) | result |

Figure 10: *The original image is retargeted to 150% width, first using an automatic saliency map of [IKN98]. The result is not perfect since the head has not been detected as a salient object. Increasing the saliency contrast improves the result (4th image), and a single stroke on the head produces even further improvement (right).*

found that saliency adjustment is the most useful control mechanism; the other parameters were usually left at their default values ($w_{reg} = 0.5$, $L^w = L^h = 20\%$).

Benchmarks. Our algorithm uses a single core of a CPU and does not rely on special graphics hardware or custom-tailored nonlinear solvers. Comparing to others methods, the complexity of the optimization is reduced by an order of magnitude, since the number of variables is proportional to the number of boundary samples rather than the number of quads. We allow real-time editing of the saliency map, which cannot be achieved by most other methods (including linear techniques that precompute the matrix factorization in advance). We process full HD (1920x1080) images in less than 4 ms using a single CPU core. [KLHG09] uses a GPU-based multigrid solver on NVIDIA GTX280 with 240 cores, and it is still 25 times slower than our CPU implementation. [KFG09] and [CFK*10] are about 500 times slower than our method. [RSA09] proposes a very expensive optimization that is 4 orders of magnitude slower than ours.

Integration in a web browser. Optimizing website layouts for different screen resolutions is a difficult task. To obtain good results in terms of usability and presentation, it is often necessary to design customized views for every aspect ratio.

Text can be easily rearranged to fit a window of any size, but images are only scaled homogeneously, limiting the layout optimization algorithm used in mobile web browsers and potentially leading to sub-optimal results. Image retargeting allows to change the aspect ratio of a picture, increasing the quality of the final layout and saving space.

To incorporate a retargeting system in a web browser, we should not only consider the retargeting quality but also its efficiency and space overhead, due to energy consumption and bandwidth limitations. Contrary to other methods, our algorithm can be easily integrated in a web browser with

negligible time and space overhead. We see two possible ways of extending any image format to store the information needed to retarget it (assuming a 25×25 grid, which suffices for high quality retargeting results up to full HD resolution):

(i) Storing the integrated saliency map. Our optimization procedure only requires the saliency matrix Ω that can be stored in 625 bytes if we quantize every matrix entry to one byte. We can then retarget to arbitrary aspect ratio using the mobile CPU.

(ii) Storing precomputed aspect-ratios. An axis-aligned grid is parametrized by 50 floats, since only the boundary has to be encoded. We can efficiently store multiple grids inside an image with a very small space overhead: encoding a grid requires 200 bytes, so that a set of 10 retargeted grids uses less than 2 KB. In this setting there is no computational overhead for the browser, as it only needs to select the desired grid for bilinear interpolation to map the image onto the screen. Tt is also possible to linearly interpolate two grids to obtain any intermediate aspect ratio; in our experiments this is very close to the exact retargeting result.

User study. We conducted a user study with 305 participants, following the protocol of [RGSS10]. Eight methods have been compared: manual crop (CR), nonhomogeneous warping (WARP) [WGCO07], Scale-and-Stretch (SNS) [WTSL08], MULTIOP [RSA09], shift-maps (SM) [PKVP09], streaming video (SV) [KLHG09], energy-based deformation (LG) [KFG09] and our algorithm (AA). All datasets in the study have been created by the authors of the respective methods, manually tweaking parameter values and sometimes the saliency to show the strengths of the retargeting algorithm and produce the best possible result. We note that the study participants had no reason to prefer a retargeted image over a (manually) cropped one since the study did not place the images in any semantic context. This biases the study in favor of manual cropping as it does not introduce any distortion. For this reason, cropping should be considered as a reference, not as a proper retargeting algorithm (for more details see the original paper [RGSS10]).

The study statistics are provided in the additional material. Fig. 11 provides a short summary that shows that our
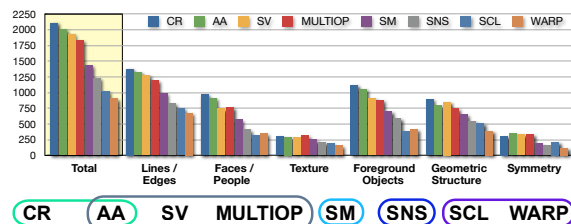
Figure 11: The number of votes for the 8 methods considered in our user-study for each image attribute. In the bottom, the operators within a group are statistically indistinguishable in terms of user preference. Our method ranks higher than others and it is statistically indistinguishable from CR.

deformation subspace is a good choice for content-aware retargeting. Our results have been considered superior with respect to six state-of-the-art methods and achieved a quality statistically indistinguishable to SV [KLHG09], while being simpler to implement, faster and not requiring a GPU implementation to obtain interactive frame rates. Our findings are in accordance with the original study [RGSS10], providing further validation of the consistency in the users' preferences.

5. Conclusion

We presented an image retargeting method that is based on axis-aligned deformations. This deformation space appears suitable for the problem at hand, and has multiple advantages, such as robustness and guaranteed lack of foldovers, smoothness, and realtime performance. The general approach of controlling a deformation energy by the domain boundary falls into the category of boundary element methods and allows for very efficient solutions in cases like ours.

Axis-aligned deformations have less freedom than general variational warps. We argue that in most cases, localized rotations are bad for image retargeting, because they lead to swirling or significant shearing. However, it is conceivable that in certain situations extreme shearing is preferable to axis-aligned scaling, for instance when the image background has completely uniform color, so that its shearing is not visible. As we exclude rotations from our warps, our method will not be capable of reproducing such effects. Note also that our method cannot guarantee the preservation of straight lines in the image if they are not strictly axis-aligned.

We are interested in extending our method to video retargeting in future work, as its speed and absence of precomputation overhead would enable online (streaming) execution. Video retargeting is very challenging due to the additional temporal coherence requirements. Our technique can be potentially generalized to video by making the warps "track" salient moving objects via deformations that are consistent with the optical flow.

Acknowledgements

References

[Ado10] ADOBE SYSTEMS INC.: Photoshop CS5, July 2010. http://www.adobe.com/photoshop/. 1

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D.: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph. 28*, 3 (2009). 2, 3

[CFK∗10] CHEN R., FREEDMAN D., KARNI Z., GOTSMAN C., LIU L.: Content-aware image resizing by quadratic programming. In *Proc. NORDIA* (2010). 1, 2, 3, 6, 7

[CZM∗11] CHENG M.-M., ZHANG G.-X., MITRA N. J., HUANG X., HU S.-M.: Global contrast based salient region detection. In *Proc. CVPR* (2011), pp. 409–416. 6

[GSCO06] GAL R., SORKINE O., COHEN-OR D.: Feature-aware texturing. In *Proc. EGSR* (2006), pp. 297–303. 2

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20* (1998), 1254–1259. 6, 7

[KFG09] KARNI Z., FREEDMAN D., GOTSMAN C.: Energy-based image deformation. In *Proc. SGP* (2009). 2, 4, 7

[KLHG09] KRÄHENBÜHL P., LANG M., HORNUNG A., GROSS M.: A system for retargeting of streaming video. *ACM Trans. Graph. 28*, 5 (2009). 1, 2, 3, 6, 7, 8

[KSSCO08] KRAEVOY V., SHEFFER A., SHAMIR A., COHEN-OR D.: Non-homogeneous resizing of complex models. *ACM Trans. Graph. 27* (2008), 111:1–111:9. 3

[LJW10] LIU L., JIN Y., WU Q.: Realtime aesthetic image retargeting. In *Proc. Eurographics Workshop on Computational Aesthetic in Graphics, Visualization, and Imaging* (2010). 2

[MB10] MATTINGLEY J., BOYD S.: CVXGEN: A code generator for embedded convex optimization, 2010. Manuscript. 6

[PKVP09] PRITCH Y., KAV-VENAKI E., PELEG S.: Shift-map image editing. In *Proc. ICCV* (2009). 2, 3, 7

[RGSS10] RUBINSTEIN M., GUTIERREZ D., SORKINE O., SHAMIR A.: A comparative study of image retargeting. *ACM Trans. Graph. 29*, 5 (2010). 1, 2, 3, 6, 7, 8

[RSA08] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Improved seam carving for video retargeting. *ACM Trans. Graph. 27*, 3 (2008). 2, 3, 6

[RSA09] RUBINSTEIN M., SHAMIR A., AVIDAN S.: Multi-operator media retargeting. *ACM Trans. Graph. 28*, 3 (2009), 23. 3, 6, 7

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Proc. CVPR* (2008). 2, 3

[SS09] SHAMIR A., SORKINE O.: Visual media retargeting. In *ACM SIGGRAPH Asia Courses* (2009). 1

[WGCO07] WOLF L., GUTTMANN M., COHEN-OR D.: Non-homogeneous content-driven video-retargeting. In *Proc. ICCV* (2007). 7

[WTSL08] WANG Y.-S., TAI C.-L., SORKINE O., LEE T.-Y.: Optimized scale-and-stretch for image resizing. *ACM Trans. Graph. 27*, 5 (2008), 118. 2, 3, 7

[ZCHM09] ZHANG G.-X., CHENG M.-M., HU S.-M., MARTIN R. R.: A shape-preserving approach to image resizing. *Comput. Graph. Forum 28*, 7 (2009), 1897–1906. 2, 3, 4