Monte-Carlo Image Retargeting

Roberto Gallea, Edoardo Ardizzone and Roberto Pirrone

DICGIM - Universita’ degli Studi di Palermo, Viale delle Scienze, Ed.6, III Piano, 90128, Palermo, Italy

Keywords: Image Resizing, Image Retargeting, Monte-Carlo, Visual Saliency.

Abstract: In this paper an efficient method for image retargeting is proposed. It relies on a monte-carlo model that makes use of image saliency. Each random sample is extracted from deformation probability mass function defined properly, and shrinks or enlarges the image by a fixed size. The shape of the function, determining which regions of the image are affected by the deformations, depends on the image saliency. High informative regions are less likely to be chosen, while low saliency regions are more probable. Such a model does not require any optimization, since its solution is obtained by extracting repeatedly random samples, and allows real-time application even for large images. Computation time can be additionally improved using a parallel implementation.

The approach is fully automatic, though it can be improved by providing interactively cues such as geometric constraints and/or automatic or manual labeling of relevant objects.

The results prove that the presented method achieves results comparable or superior to reference methods, while improving efficiency.

1 INTRODUCTION

The diffusion of display devices coming with different aspect ratios and resolutions, entails using content-aware resizing techniques. Simple cropping is not sufficient due to severe information loss. On the other hand, homogeneous scaling with aspect ratio variation, introduces unwanted distortions in the images. A proper non-homogeneous resizing operator is required in order to preserve image content, introducing deformations just in the low-importance regions of the image.

In this paper, we present a novel image retargeting technique, which is both efficient and effective. Differently from many literature approaches, such a method does not require neither energy minimization nor functional optimization, and relies just on Monte Carlo sampling. Our model estimates the deformation likelihood of each image region, according to the image saliency. Then, by extracting random samples over this probability distribution, less important regions get more deformed, while high-saliency ones are preserved. Another advantage of using a sample-based approach is that it can be implemented easily using a parallel scheme, thus improving efficiency.

Salient regions can be extracted using several content relevance estimators, such as visual saliency maps (Itti et al., 1998; Hou et al., 2012), corner detectors (Harris and Stephens, 1988), eye-gaze measurement (Santella et al., 2006), etc. Additionally, both automatic or interactive cues can be given to improve the results: people detectors (Dalal and Triggs, 2005) or face detectors (Viola and Jones, 2001) can help in preserving people and faces in the images. Finally, other geometric constraints can be provided by the user to preserve structures explicitly.

2 RELATED WORK

In general, the resizing operators used by image processing applications work by resizing images to a target size by means of homogeneous shrinking or enlarging operators. After early works based on cropping, like (Suh et al., 2003), more recent approaches use adaptive image resizing. The idea is to preserve important image features by applying a non-linear content driven resize operator. Remarkable works were done using seam carving, (Avidan and Shamir, 2007; Rubinstein et al., 2008) where 1D seams are removed/added to reduce/increase the image size. Such seams are chosen from low energy regions of the image. However, due to the discrete nature of this method, notches in the objects may appear. In ad-
3 IMAGE RESIZING APPROACH

In our model an input image \( I \) is considered as a set of \( n \) lines (the columns or the rows) \( I = \{l_0, l_1, ..., l_{s-1}\} \), where \( l_i \) are the initial lines positions and \( s \) is the initial image size along the considered dimension. Thus, \( l_i = i \quad \forall \quad i \in \{0, s-1\} \). To resize the image to the new dimension \( s' \) we look for the new set \( I' = \{l'_0, l'_1, ..., l'_{s-1}\} \) where distances between two consequent lines should be preserved in most informative image regions in order not to introduce distortions as in Equation (1),

\[
(l_i - l_{i-1}) = (l'_i - l'_{i-1}).
\]  

(1)

Obviously, some distances have to be necessarily changed due to resizing, and some deformation must be introduced. The model is built in order to spread the required deformations across the whole image in a non uniform way that obeys to a probability distribution. This is done applying multiple atomic resize operators that are sampled from a proper probability mass function. Such a function is built according to lines significance. The idea is to apply less atomic deformations in salient regions of the image, while the most deformation affects the unimportant zones.

The whole system is realized by means of a chain, which is schematized in the block diagram in Figure 1. The input image \( I \) is given as input to the system. The saliency estimator generates a saliency map, which is used in turn to build a deformation probability mass function \( d\text{pmf} \). Such a function is sampled to move the lines of the image non-homogeneously. Finally, the image is reconstructed and the final retargeted image is produced as output.

3.1 Model Formulation

The proposed method is based on two concepts:

- a resizing operator
- a deformation strategy

**Resizing operator.** The resizing operator we introduce is considered as the multiple application of atomic resizing operations. Each atomic resizing operates on a single line \( l_i \), moving it by a given quantity \( k_{lod} \), which is expressed in (fractions of) pixels, and defines the level of detail of the transformation. Relative movement between two consequent lines has the effect of deforming the underlying image. Changing the quantity \( k_{lod} \) affects:

- the size of each atomic resizing, which is equal to \( k_{lod} \), and defines the movement of the the single line \( l_i \)
- the number of atomic resizing \( n_r \) (see Equation 2) to produce the required image final size.

These two quantities define the resolution of the global resizing operator. Of course, as the level of detail gets finer, the computational burden gets heavier, due to an increasing number of atomic resizing operations. In the following paragraphs, the selection of the \( k_{lod} \) parameter will be discussed in detail.

\[
n_r = s' / k_{lod}.
\]  

(2)

**Deformation strategy.** In order to apply the described resizing operator, a line selection strategy needs to be designed to determine which line should be moved at each resizing step. The required procedure has to exhibit a dual behavior: firstly, it should select less important lines from a visual content importance perspective, as the candidates for resizing since distortions should be preferentially introduced in low-importance or homogeneous regions. On the other hand, deformations should be distributed across
the whole image in order not to remove whole image regions, thus introducing severe artifacts. In order to attain such behaviors, we define a deformation probability mass function \( p_d(x) \) over the image \( I \). Such a \( p_d \) indicates the likelihood that a single resizing operation would affect the image line \( l_i \).

Intuitively, such a probability should be related to the image visual content importance. In particular, line relevance \( R(i) \) is extracted from the dual form of visual saliency, i.e. visual inconspicuousness. Visual saliency \( S(i, j) \) of each pixel is extracted using either Itti’s saliency detector (Itti et al., 1998) or signature saliency (Hou et al., 2012). Such values are then projected along the considered resizing axis using the maximum operator, and are complemented as in Equation (3).

\[
R(i) = 1 - \max_j S(i, j) \quad (3)
\]

Finally, the values are normalized w.r.t. their summation, as in Equation (4) to recover \( p_d(i) \); an example is reported in Figure 2.

\[
p_d(i) = R(i)/\sum_j R(j) \quad (4)
\]

Here high-value points correspond to image regions with high-probability of being deformed, while low-value points correspond to image regions that should be preferentially preserved. Note that visual saliency can also be improved either interactively by adding constraints, or automatically using people (Dalal and Triggs, 2005) or face detectors (Viola and Jones, 2001), and modifying \( R(i) \) to have low values in presence of constraints or people/faces. Note that this operation must be performed prior to normalization reported in Equation (4) to preserve the \( p_d(i) \) integral to sum to 1, thus being a valid \( p_d \).

![Figure 2: Plot of the deformation probability mass function \( p_d(i) \) related to the example image. The function represents the probability that a given line \( l_i \) will be subject to an atomic resizing step during the retargeting operation.](image)

3.2 \( k_{lod} \) Parameter Selection

The whole procedure is automatic, just the parameter \( k_{lod} \) requires to be tuned. Since it influences the quality of the result, it should be as smaller as possible. However, the computation time is in inverse proportion to the parameter value, so it should be determined as the best trade-off between quality and efficiency. For a visual evaluation purpose, we report the results using different values for \( k_{lod} \) (1, 0.5, 0.1, 0.05 and 0.01), see Figure 3. The results show that using a coarse level of detail causes artifacts. However, using too fine level of detail is not useful, since the resulting image quality does not get remarkable improvements. However, visual inspection is not sufficient to determine how to choose \( k_{lod} \). More objective cues are derived by measuring the variations of two image difference indexes: Root Mean Squared Error - \( RMSE \) and Structural Similarity - \( SSIM \) (Wang et al., 2004). Measures have been computed between the image produced using the highest level of detail (approximated using a very low value of \( k_{lod} = 0.001 \)) and the one resulting using a \( k_{lod} \) value varying in the interval \([0.001, 1]\). The results of this experimentation are shown in Figure 5. In Figure 5(a) the results are better for lower values of \( RMSE \). In Figure 5(b) the results are better for higher values of \( SSIM \). The plots exhibit jumps which are due to interpolation artifacts that arise during image reconstruction.
Figure 3: Detail of an output image using different values for $k_{lod}$ and scaling ratio $s_w = 0.5$. From left to right: (a) original image, (b) $k_{lod} = 1$, (c) $k_{lod} = 0.5$, (d) $k_{lod} = 0.1$, (e) $k_{lod} = 0.05$, (f) $k_{lod} = 0.01$. Even though the effect is more noticeable when the images are larger, as it can be seen in this image portion, the finest results are provided using a high level of detail. However, values of $k_{lod}$ smaller than 0.1 do not provide remarkable improvements.

Figure 4: Plot of the sampling of the $d_{pmf}$ for the picture shown in Figure 2. Each column corresponds to the number of times each column was drawn from the $d_{pmf}$. Such artifacts get smaller as $k_{lod}$ decreases. When $k_{lod}$ approaches values around 0.1, jumps disappear, the trend becomes asymptotic and the quality of the result has very low variations. As a consequence $k_{lod} = 0.1$ is assumed to be used for the referred results in the next sections.

4 EXPERIMENTAL RESULTS AND DISCUSSION

The described method was implemented on a PC with Quad CPU 2.30 GHz. The system can benefit from parallel computation by leveraging gGPU capabilities being implemented using Nvidia CUDA API extensions (CUDA, 2007).

Comparison. To evaluate the results of our system, we compared it with other literature retargeting systems. For space reasons, this paper references four methods: Multi-operator (Rubinstein et al., 2009), non-homogeneous warping (Wolf et al., 2007), seam carving (Rubinstein et al., 2008) and scale-and-stretch (Yu-Shuen Wang and Lee, 2008). However, several other methods were compared and the reader is referred to the supplemental material provided with this paper. The evaluation was assessed using the datasets and measures provided by the RetargetMe comparative study (Rubinstein et al., 2010). Examples of comparisons are shown in Figure 6.

Additionally to qualitative images inspection, an objective evaluation was also taken. Two comparative measures were used for this purpose: Earth Mover’s Distance (EMD) (Pele and Werman, 2009) and SIFT-flow (Liu et al., 2008). These are two commonly used similarity metrics, which do not require the two datasets to be the same size, a binding property for image retargeting. Both measures use a dense SIFT descriptor (Lowe, 2004), which captures structural properties of the image robustly, while EMD also uses a state of the art color descriptor (ciede2000). The two measures both endorse their solutions to small and smooth local displacements, reflecting the way human vision system tolerates deformations and the operations applied by retargeting operators.

Results, summarized in Table 1, show that the images produced with the proposed method provide measures comparable to literature methods, or even better. Most of the existing literature methods, tend to warp the whole image and make them fit it into the new frame size. However, often the periphery of the
image is not required to be kept. Our method allows intrinsically to discard the whole periphery data, if it is not salient, by strongly compressing it, thus achieving a certain extent of cropping. This allows to keep more space for important image regions, which can be better preserved without introducing heavy deformations.


Table 1: EMD and SIFTflow measures for images of RetargetMe framework.

<table>
<thead>
<tr>
<th>Measure</th>
<th>EMD</th>
<th>SIFTflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monte-carlo</td>
<td>8.01 ± 3.25 · 10^3</td>
<td>3.98 ± 2.02 · 10^5</td>
</tr>
<tr>
<td>Multi-operator</td>
<td>8.30 ± 3.58 · 10^3</td>
<td>3.94 ± 1.99 · 10^5</td>
</tr>
<tr>
<td>Non-homogeneous</td>
<td>8.68 ± 3.73 · 10^3</td>
<td>4.12 ± 2.15 · 10^5</td>
</tr>
<tr>
<td>Scale and stretch</td>
<td>8.69 ± 3.60 · 10^3</td>
<td>4.09 ± 2.38 · 10^5</td>
</tr>
</tbody>
</table>

4.1 Complexity Considerations

Looking at the proposed retargeting operator from a complexity perspective, is possible to take both memory and computational considerations.

The memory amount required to store all the data needed to retarget an image composed of s into one composed of s’ lines is the following:

- s real values to store the positions of the lines l,
- s real values to store the dpmf.
- s’ · klad real values to store the samples extracted from the dpmf.

As a consequence, the proposed method needs a total of 2s · s’ · klad real values, keeping the memory complexity polynomial.

From a computational point of view, the main burden is related to the saliency extraction which is common in all of the retargeting methods, so it is not considered. For the same reason, image reconstruction is not taken into account. The rest of the process is accomplished by the following operations:

- Design of the dpmf. Each value pd(i) is designed starting from the saliency S(i, j) using the max(·) operator → polynomial.
- Sampling pd(i). This operation is repeated s’ · klad times → polynomial.
- Updating of the lines position l according to the extracted samples → polynomial.

Being all of the subprocess polynomial, the whole procedure is polynomial too. In addition, all of the previous operations can be easily implemented in parallel, since little or no dependencies exists both between data and processes. This allow very fast one-shot retargeting of images, opposed to many of the reference literature methods relying onto iterative optimization.

5 CONCLUSIONS AND FUTURE WORKS

A novel efficient method for image retargeting was presented. It is based on Monte Carlo sampling of the deformation probability mass function of the image, which is defined using the image saliency map. This allows its use for real-time applications. Experimental results show that its performance are comparable or even superior tested against more complex existing systems. The method keeps its complexity very low both from a memory and computational perspective, also leveraging the parallelization of its processes.

Further work will involve overall system improvements and its extension to video resizing. This issue requires the introduction of a time-coherent saliency map and further constraints. Additionally, the model will be embedded in systems making use of retargeting for real-time applications, such as personalized media content distribution on mobile devices or the web.

REFERENCES


