SeamCrop: Changing the Size and Aspect Ratio of Videos

Johannes Kiess, Benjamin Guthier, Stephan Kopf, Wolfgang Effelsberg
Department of Computer Science IV
University of Mannheim, Germany
{kiess|guthier|kopf|effelsberg}@informatik.uni-mannheim.de

ABSTRACT

The resizing of videos is currently in high demand due to the popularity of mobile devices with very small screen sizes. As the screen resolution of these devices increases, the resizing is mostly needed to change the aspect ratio of the videos to fit the display size. In this paper we present a novel and fast approach for the retargeting of videos which combines cropping and seam carving. The idea is to find an optimal cropping window and to get more useful content into it by additionally removing seams. We formulate the search for a cropping window in terms of a 2D rectangle representing the possible positions of the window over all frames. Finding the optimal position can be solved efficiently with dynamic programming. A user study was conducted that compares our approach to a similar state-of-the-art video retargeting technique, showing that our new algorithm has fast computation times while also having a comparable quality of results.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

Keywords

video retargeting, video resizing, seam carving, seam crop

1. INTRODUCTION

With the increasing popularity and distribution of smartphones and similar mobile devices, the demand for media to consume on the go rises. As most images and videos today are captured with HD or higher resolutions, there is a need to adapt them in a content-aware fashion before they can be watched comfortably on screens with small sizes and varying aspect ratios. This process is called retargeting. The change of aspect ratio often causes the most distortions when resizing. Thus retargeting mainly focuses on adapting the aspect ratio of a video while the rest can be scaled uniformly. An example would be the viewing a widescreen movie on a device with a 4:3 display. The movie can be scaled down to fit the height without loosing important details but black borders have to be introduced or the borders have to be cropped in order to fit the width.

A recently published study about the *resizing of images* indicates that people generally prefer the loss of content over the insertion of artifacts [8]. While many researchers concentrate on complex techniques like warping for retargeting, it is also suggested that the search for an optimal cropping window is still a viable research topic. Cropping describes a simple operation were the borders of an image are removed to fit it to the target size. As it can be assumed that the suggestion is also true for the retargeting of video, we present a novel combination of cropping and seam carving [1]. In the latter, connected paths of pixels called *seams* are removed from within the image in order to reduce its size.

The intention of SeamCrop is to have fast computation times while providing improved or at least comparable results to other state-of-the-art algorithms. Many of these algorithms use an optimization over the entire video sequence which is very time consuming. Therefore, we only use the whole sequence for the computing of a cropping window while the rest is done on each frame separately. This makes the optimizations that have to be solved less complex and consequently reduces the processing time.

The main idea of the algorithm is to find an optimal path for the cropping window that reduces the width or height and then use seam carving to get more useful content into it without creating visible artifacts. An optimal path is a sequence of cropping windows of a fixed size over all frames that captures the most important content. The borders of the cropping window are extended on both sides by a small amount. This extra space is intended to prevent objects from being cut off when they are near a border. After the content outside of the extended borders has been cropped, seam carving is used on each frame separately to reduce them back to the target size. To ensure temporal coherence between the frames, we use a simplified and much faster version of temporal coherence costs [2]. With these costs, the energy map of a frame is temporarily modified for each seam to have lower values in the positions of the corresponding seam from the preceding frame.

In contrast to multi-operator retargeting suggested by Rubinstein et al. [9], we do not use non-homogeneous scaling in our approach as it may introduce squeezing artifacts to faces and objects. A preliminary user study with 16 participants was conducted in which multi-operator retargeting,

MoVid'12, February 24, 2012, Chapel Hill, North Carolina, USA

cropping and a combination of cropping and seam carving were compared for the retargeting of images. The evaluation shows that all subjects find squeezing artifacts in faces or persons disturbing while some still find them acceptable for images showing only buildings or a landscape. As our algorithm is intended to resize different kinds of videos that may also depict persons, this supports our decision not to include asymmetric scaling.

Our main contributions are as follows:

- Define a novel and fast method for finding the optimal path of a cropping window in a video.
- Present a fast combination of cropping and seam carving for changing the aspect ratio of videos.

The outline of this paper is as follows: Chapter 2 presents the current state of the art of video retargeting. The SeamCrop algorithm is described in detail in Chapter 3. A comparison of our novel technique with the multi-operator approach is shown in the evaluation in Chapter 4. Chapter 5 concludes the paper.

2. RELATED WORK

Warping techniques resize a video by attaching a grid to each frame and then combining them into a 3D construct which consists of all frames. This grid is deformed in a content-aware fashion in order to keep highly salient regions as unchanged as possible.

The warping algorithm proposed by Krähenbühl et al. [6] performs a warp on the pixel level and for each frame separately. In order to preserve temporal coherence, the per-frame importance maps are combined with a lookahead of a few frames.

Wang et al. [10] present a technique which separates the retargeting process into a temporal and a spatial component. Each frame is resized independently of the others first to analyze the resulting motion pathlines. These lines describe the motion trajectories of a set of points over the course of the video. Incoherent deformations in them cause temporal artifacts which can be prevented by minimizing their divergence. The resizing is then repeated with additional information gained from the motion path lines.

Seam carving was introduced by Avidan and Shamir [1] and retargets an image by removing or inserting paths of monotonic and connected pixels called *seams*. These seams go either from top to bottom when changing the width or from left to right when changing the height. Each removed or inserted seam alters the width or height of the image by one pixel (see Figure 1). First, each pixel is assigned an energy value. The cost of a seam is defined as the summed up energy values of the pixels contained in it. A seam has to satisfy two conditions: the pixels must be connected, and only one pixel is assigned to it in each row. Dynamic programming is used to determine the optimal seam with the lowest overall costs under these conditions. This seam is then removed, and the process is repeated until the target size is reached.

Seam carving is extended by Rubinstein et al. [7] for the resizing of videos. Graph cuts are used instead of dynamic programming to find a 2D manifold in the 3D video cube. This cube depicts the width, height and the time of a video. Also, forward energy is introduced which considers the energy brought into a frame by the newly formed edges instead of the energy that is removed.

Figure 1: Left: Original image. Center: Seams found in the image. Right: Image after the seams have been removed.

Seam carving is also the basis of other current video retargeting techniques. Kopf et al. [5] align the frames of a shot into a background image using a median filter. Seams are then calculated on the background image and re-traced to the individual frames by applying an inverse camera model. The seam carving operator used in this approach is enhanced by Kiess et al. [3] to preserve the shape of straight lines and symmetric structures.

In the approach introduced by Grundmann et al. [2], the seams are allowed to be temporally and spatially discontinuous. Temporal coherence is achieved by using a measure that takes the seam from the previous frame as an initial seam and determines a cost for the change that would be introduced into the frame by removing other points. Similarly, there is a cost that allows seams to be discontinuous in the spatial domain.

A combination of bi-cubic scaling, cropping and seam carving is proposed by Rubinstein et al. [8]. To determine in which order the operators should be employed, an image similarity measure called bi-directional warping is introduced. Dynamic programming is used to find the order in a multi-dimensional space that combines the operators. The method distinguishes between regular paths, where the operators are used only once one after another, and mixed paths, where the operators have no distinct order and are used as often as necessary. When used to retarget videos, the algorithm searches for the best regular paths for the key frames and interpolates how often each operator is used in the intermediate frames.

A different approach has been presented by Kopf et al. [4]. Instead of combining two or more video retargeting operators, the authors analyze each shot and based on its visual content select the most suitable retargeting technique. Either seam carving or cropping based on regions of interest is used as the video retargeting operator.

The current techniques try to squeeze as much content as possible in the retargeted result. Our novel approach is based on the motivation to prefer the loss of content over the insertion of artifacts.

3. THE SEAMCROP ALGORITHM

The algorithm can be divided into two parts: the first part calculates the positions of a cropping window with the target size over the course of the video, the second extends the cropping window and then removes seams to reach the target size again. As the movement of the cropping window depends on the content shown on the screen, the effect of a virtual pan may be introduced.

In the following, the reduction of width is used to illustrate our algorithm. The reduction of height is achieved in a similar manner.

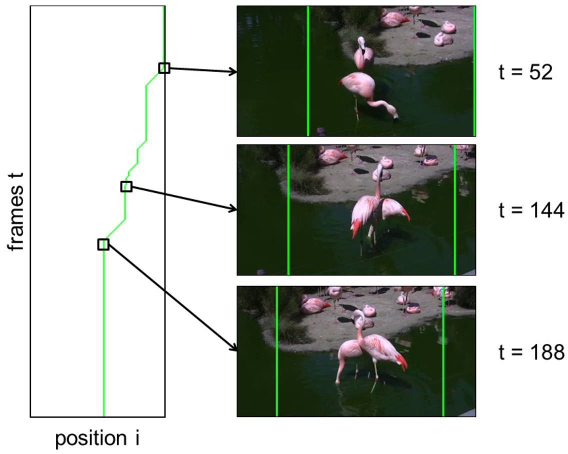

Figure 2: Path of the cropping window over time. Each point in the path stands for the position of the cropping window in the corresponding frame.

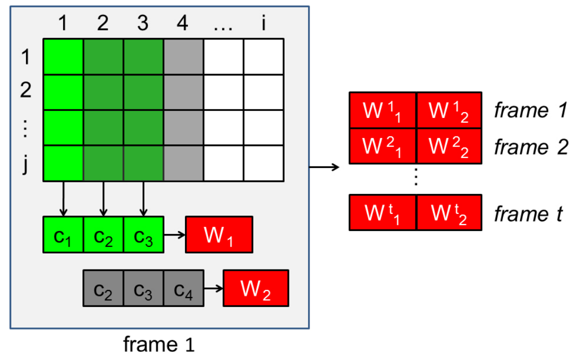

Figure 3: The energy values in each column are summed up first. Next, this is also done with the summed up values included in a cropping window position (e.g., the green values are summed up to W_1). Lastly, the values for the positions of all frames are combined to a 2D array.

3.1 Finding a cropping window

The video sequence to be resized consists of a sequence of T frames F^t. Unless required for clarity, we omit the time index t. Each frame has the original width m and height n and is resized to a target width of $m' < m$. At any given point in time, a cropping window can thus take on $m - m' + 1$ possible horizontal positions. Over the duration of the entire video, the path of a cropping window is a path through a two-dimensional space of size $(m - m' + 1) \times T$. It is calculated via dynamic programming, similar to the computation of a seam in the seam carving algorithm [1]. Figure 2 shows the path of the cropping window over time.

We begin by computing an energy map E for each frame in the video. Each pixel in E is a value representing the importance of the corresponding pixel in the frame. The energy map is composed of the rate of temporal change of a pixel (i.e., its motion) E_M and the measure E_G depicting the L_1 length of its gradient normalized to $[0..1]$. We combine them in the following way using weights that worked well in our experiments:

$$E = \frac{3}{4} E_M + \frac{1}{4} E_G. \tag{1}$$

The motion of a pixel is estimated by the difference of its values between the preceding and the following frame. As pixel values may differ slightly between frames due to lighting or small camera movements instead of object motion, the values in E_M are thresholded to either 0 or 1 using a threshold T_M. It is computed for each frame individually. First, the maximum difference value for each column i of E_M is determined as

$$e_i = \max_{j=1,..,n} E_M(i, j). \tag{2}$$

The threshold T_M is then set to 25% of the average of these maxima:

$$T_M = 0.25 \sum_{i=1}^{m} \frac{e_i}{m}. \tag{3}$$

Since thresholded values of 1 mainly appear on the edges of moving objects, a Gaussian smoothing filter is used to distribute the non-zero energy values over the total space of the object. Empirical tests have indicated that this simple measure works sufficiently well for detecting object motion in a video.

The values of the columns of the total energy map E computed above are now summed up. The result is a list of column costs

$$c_i = \sum_{j=1}^{n} E(i, j) \tag{4}$$

for each column $i = 1, .., m$. These costs are then used to determine the total energy W_i contained within each possible cropping window position $i = 1, .., (m - m' + 1)$

$$W_i = \sum_{k=0}^{m'-1} c_{i+k}. \tag{5}$$

Calculating this total energy for every frame in the video yields a 2D array W_i^t where each value stands for the total energy of one cropping window position i in one frame t (see Figure 3).

On this 2D array, dynamic programming is used with similar restrictions as in the seam carving algorithm in order to find the path with the maximum energy [1]. The energy of a path is the sum of the energy values of all path positions. It is determined by traversing the W_i^t space along the time axis and calculating the cumulative maximum energy $\widetilde{W}_i^t$ for each position as

$$\widetilde{W}_i^t = W_i^t + \max(\widetilde{W}_{i-1}^{t-1}, \widetilde{W}_i^{t-1}, \widetilde{W}_{i+1}^{t-1}). \tag{6}$$

The maximum value of $\widetilde{W}_i^t$ in the last row ($t = T$) indicates the total cost of the path with the highest energy. By backtracking from this maximum, we find the optimal path for a cropping window. As the paths found by the algorithm are connected, the resulting cropping window path is temporally coherent but it may contain jitter. In order to lessen the effect of jitter of the window, the computed positions are smoothed with a Gaussian filter.

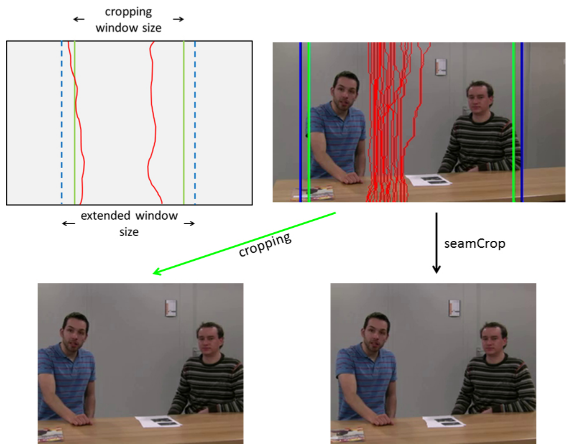

Figure 4: Top row: The estimated cropping window (green lines) is extended (blue dotted lines) and then reduced to the target size again by removing a small number of seams (red). Bottom row: results of the frames being reduced by the cropping operator alone (left) and by SeamCrop (right).

3.2 Finding seams

In order to include more important content in the cropping window without introducing artifacts, the borders of the cropping window are slightly extended, and then seam carving is used frame by frame to reach the target size again. To ensure temporal coherence, the energy map E^t of each frame is modified by the seams found in the previous frame. We use a simplification of the temporal coherence costs introduced by Grundmann et al. [2]. They base their measure on the new edges that are introduced by moving the seam to a different pixel position. We on the other hand use costs increasing linearly from the position of the previous seam because it is sufficient for our technique and the computation time is much lower.

As the cropping window already has the target size, it has to be extended before seams can be removed. It is extended equally on both sides (see Figure 4). The amount of enlargement is a parameter that controls the tradeoff between seam carving and cropping. We chose to enlarge the window by 20% in our experiments. If the cropping window moves towards a border of the frame and the added space would lie outside of the border, the space on the other side is extended by an equal amount. Similarly, if the window leaves the border, the space is again equally distributed.

The k-th vertical seam in frame t is defined by the list of pixels it includes. It includes exactly one pixel per row of the frame. The seam can thus be fully described by a horizontal pixel position $s_k^t(j)$ for each row $j = 1, .., n$. We omit the row index j when referring to the entire seam. In contrast to the computation of the cropping window, which is optimized over all frames, the seams are calculated for each frame separately. To prevent temporal discontinuities and jitter that may occur by a frame-wise search for seams, temporal coherence costs are used to modify E before computing a seam.

We assume that a number of seams s_k^{t-1} have been calculated in the previous frame. Now, the same number of

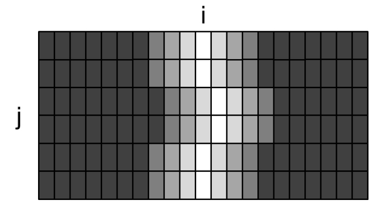

Figure 5: Visualization of temporal coherence costs. The positions of the corresponding seam s_k^{t-1} from the preceding frame are white, and the energy increases the darker the positions get.

seams must be calculated for the current frame. We want to make it more likely for a new seam to be close to the seam with the same index in the previous frame. We thus add temporal coherence costs C_k^t to the energy map E before calculating seam s_k^t. The costs C_k^t are zero at the location of the corresponding seam s_k^{t-1} in the previous frame. They then increase linearly with increasing horizontal distance to the seam up to an upper bound β beyond a pixel distance of α (see Figure 5). The temporal coherence costs are thus defined as

$$C_k^t(i,j) = \begin{cases} \frac{\beta}{\alpha}\,|i - s_k^{t-1}(j)| & for\ |i - s_k^{t-1}(j)| < \alpha \\ \beta & otherwise. \end{cases}$$

$(E + C_k^t)$ is then used as the energy function for calculating seam s_k^t. The parameter β adjusts how strongly the position of the seam of the previous frame is imposed on the energy map. We chose an upper bound of $\beta = 0.3$ for a normalized energy map. The choice of the parameter α is dependent on the image size. We set it to 3% of the image width in our experiments.

Seams are calculated in a manner similar to finding the optimal path of the cropping window, as described earlier. To calculate the k-th vertical seam in frame t, the modified energy map $(E + C_k^t)$ is traversed from top to bottom (in the direction of index j). The *cumulative minimum energy* for each pixel position is calculated by adding the energy of the current pixel to the minimum of the cumulative energy values of the three adjacent pixels above (similar to Equation 6).

The positions of all seams s_k^t are stored and are all removed in one step after an additional condition is checked. If a position of a previous seam $s_k^{t-1}(j)$ lies outside the extended window, all costs in row j are set to the upper bound: $C_k^t(i,j) = \beta, \forall i$. When more than a certain percentage of the seam lie outside the extended window, no temporal costs are added to E so that a new seam can be found. We found that setting this value to 20% gives good results. With this restriction, the seams may first be deferred a few frames so that more points of the seam lie at the border before they disappear, which is visually less disturbing in the result.

4. EVALUATION

We conducted an evaluation in order to compare our algorithm to the similar state-of-the-art technique of *multi-operator retargeting* [9]. The evaluation was a no-reference comparison where the original sequence was not shown to the participants. This is comparable to the real-world situation in which the user only gets to see the resized result

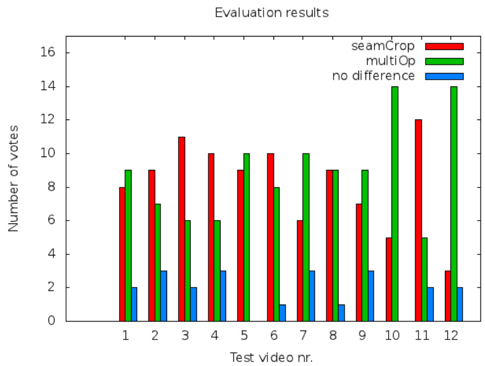

Figure 6: Votes of the evaluation.

on his device. As test sequences, twelve videos belonging to different categories like animations, movies or sports videos were used. These videos are available on the Web[1].

The evaluation was conducted online on a web site for scientific surveys[2] and consisted of twelve comparisons and five questions at the end. In each comparison, a video was shown which presented the retargeted results side by side. The side on which the results appeared on was chosen randomly for each video. For each comparison the participants were asked which result they prefer or if they could not find any visual differences. The last page was meant to get feedback on the decisions the subjects made while rating the videos. It was asked if their decisions were influenced by squeezing artifacts, visible deformations in general, cut off objects or abnormal camera motion. Lastly, there was an optional open question about the reasons for their rating.

A total of 19 participants took part in the evaluation. 13 were volunteers, 6 were colleagues from our department. A total of 228 video comparisons were thus evaluated, and only fully executed surveys were utilized.

Analysis and Discussion

The survey shows that none of the techniques is superior in all of the test sequences, each has its strong and weak points (see Figure 6). SeamCrop is preferred where persons or objects that appear large on the screen are squeezed, e.g. in a video where a train with visible passengers crosses the screen or a sequence with football players. Multi-operator retargeting is favored in sequences with many moving objects or with structured backgrounds, e.g. a small person walking between buildings with a structured facade.

This coincides with the answers the participants gave at the end of the survey as responses to the posed questions. The most disturbing artifacts are deformations of relevant objects, followed by cut of objects. Also, the squeezing of persons is explicitly stated in the optional open question about the reasons of their voting. Some participants additionally noticed unstable backgrounds and do not like camera pans from one side to another and then vice versa in one scene. Figure 7 shows some examples of the videos used in the survey. The original frame and a symmetrically scaled down version are added for better comparison. It can be clearly seen that some objects like persons or cars get squeezed by the scaling operators.

Many current state-of-the-art video resizing techniques solve optimization problems on the whole video cube, which takes a lot of processing time. For example, the resizing of a video sequence (400×300 pixels, 400 frames) to 50% of the original width takes about 10 to 20 minutes with seam carving based on graph cuts [7]. Also, multi-operator retargeting has average optimization times of 10 minutes for one key frame with resolutions between 600×400 and 400×300 pixels. In contrast, our new approach is very fast as it only uses information of the entire video to calculate the optimal cropping path as a 2D problem, and searches for seams in each frame individually. The performance values of the following Table 1 were created on a Intel Core 2 Quad desktop with 2.4 GHz and 4 GB memory. Please note that our algorithm is not implemented to benefit from CPU-based parallel processing.

	400×300 400 Frames	720×432 261 Frames	1920×1080 72 Frames
Processing time	1 min 6 sec	3 min 21 sec	18 min 27 sec

Table 1: Processing time for three test videos with different screen resolutions. All sequences were retargeted to 50% of the original width.

Limitations

SeamCrop produces clearly visible distortions in some of the shots. Like all resizing techniques, the algorithm depends on the functions that measure the importance. If the cropping window is positioned in the wrong spot, relevant persons or objects might get cut. Also, even though only a small percentage of the resizing is done with seams, fast moving objects may cross their path before they can jump to a new position. Additionally, artifacts may be introduced if the important content takes the whole screen.

5. CONCLUSIONS

In this paper, we presented a combination of cropping and seam carving called SeamCrop for changing the size and aspect ratio of videos. An optimal path of a cropping window is computed first, and then seams are used to get additional content into the window without creating visible artifacts. A major advantage of our technique is its fast computation time, with comparable results to a similar state-of-the-art algorithm, which is indicated by a user evaluation.

SeamCrop may produce visible errors if it positions the cropping window in the wrong spot or if an object moves too fast for the seams to evade it. In future work, we would like to improve our approach by making the size and distribution of the extended window for the removal of seams more flexible by depending it on the content. Further, we want to extend the algorithm to be able to compute streaming videos and work in real-time.

[1] http://ls.wim.uni-mannheim.de/de/pi4/research/projekte/retargeting/
[2] http://www.soscisurvey.de

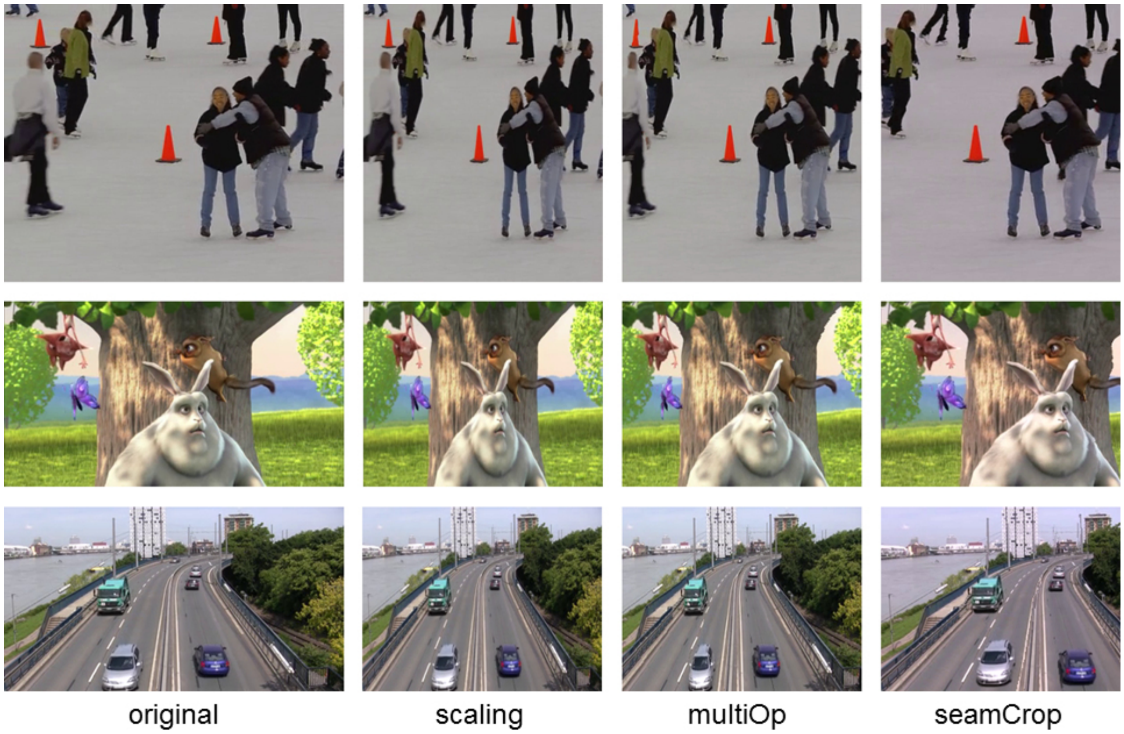

Figure 7: Example frames from the evaluation. The original frame and an asymmetrically scaled version are added for better comparison of the results.

Acknowledgements

We would like to thank Michael Rubinstein for providing the sequences resized by multi-operator retargeting. The image and video examples were provided by the following users and websites under the creative commons license[3]: adrielsocrates (beach, via flickr.com), Eugenia Loli-Queru (flamingo, via vimeo.com), media.xiph.org [4] (ice) and Blender Foundation[5] (Big Buck Bunny). Furthermore, the authors acknowledge the financial support of this project granted by the *Deutsche Forschungsgemeinschaft* (DFG).

6. REFERENCES

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics, SIGGRAPH 2007*, 26(3), 2007.

[2] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Discontinuous seam-carving for video retargeting. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 569 –576, june 2010.

[3] J. Kiess, S. Kopf, B. Guthier, and W. Effelsberg. Seam carving with improved edge preservation. In *Proceedings of IS&T/SPIE Electronic Imaging (EI) on Multimedia on Mobile Devices*, volume 7542(1), pages 75420G:01 – 75420G:11, January 2010.

[4] S. Kopf, T. Haenselmann, J. Kiess, B. Guthier, and W. Effelsberg. Algorithms for video retargeting. *Multimedia Tools and Applications (MTAP), Special Issue: Hot Research Topics in Multimedia, Springer Netherlands*, 51:819–861, January 2011.

[5] S. Kopf, J. Kiess, H. Lemelson, and W. Effelsberg. FSCAV: Fast seam carving for size adaptation of videos. In *Proceedings of the 17th ACM International Conference on Multimedia (MM)*, pages 321–330, October 2009.

[6] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross. A system for retargeting of streaming video. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–10, New York, NY, USA, 2009. ACM.

[7] M. Rubinstein, S. Avidan, and A. Shamir. Improved seam carving for video retargeting. *ACM Transactions on Graphics, SIGGRAPH 2008*, 27(3), 2008.

[8] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. In *ACM SIGGRAPH Asia 2010 papers*, pages 160:1–160:10, New York, NY, USA, 2010. ACM.

[9] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *ACM Transactions on Graphics, SIGGRAPH 2009*, 28(3):1–11, 2009.

[10] Y.-S. Wang, J.-H. Hsiao, O. Sorkine, and T.-Y. Lee. Scalable and coherent video resizing with per-frame optimization. *ACM Transactions on Graphics, SIGGRAPH 2011*, 30(4), 2011.

[3] `http://creativecommons.org`
[4] `http://media.xiph.org/video/derf/`
[5] `http://www.bigbuckbunny.org`