This is a preliminary version of an article published in
Proc. of the 17th ACM international conference on Multimedia (MM), pp. 321-330, Beijing, China, October 2009
by Stephan Kopf, Johannes Kiess, Hendrik Lemelson, Wolfgang Effelsberg

FSCAV - Fast Seam Carving for Size Adaptation of Videos

Stephan Kopf, Johannes Kiess, Hendrik Lemelson, Wolfgang Effelsberg
Department of Computer Science IV
University of Mannheim
Mannheim, Germany
{kopf|kiess|lemelson|effelsberg}@pi4.informatik.uni-mannheim.de

ABSTRACT

The presentation of multimedia data and especially of high resolution videos on small mobile devices is still a great challenge today. Both cropping of borders and scaling of frames may result in the removal of essential content of videos or lost details due to the reduced size of the visual content. Another major problem emerges if the aspect ratio of the original video and the display of the mobile device differ. User evaluations indicate that changing the aspect ratio may reduce the visual quality of videos significantly. In this paper, we present the new *FSCAV* algorithm (*F*ast *S*eam *C*arving for Size *A*daptation of *V*ideos) to adapt the size of videos to the limited display resolution and different aspect ratios of handheld mobile devices. The general idea of the seam carving algorithm for still images is to remove seams in images so that the essential content is preserved. We extended this technique which works very well for images to create videos without jitter or visible artifacts. A major feature of our *FSCAV* algorithm is the low computational complexity which enables an efficient adaptation of videos to small screens. Nevertheless, severe distortions are clearly visible in some shots of the adapted videos. We present a new heuristic to identify shots with such a low visual quality. If the quality drops below a threshold, a different adaptation technique is used for this shot (e.g., scaling or cropping). User evaluations confirm a very high visual quality of our approach.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

Algorithms

Keywords

video adaptation, video retargeting, seam carving

1. INTRODUCTION

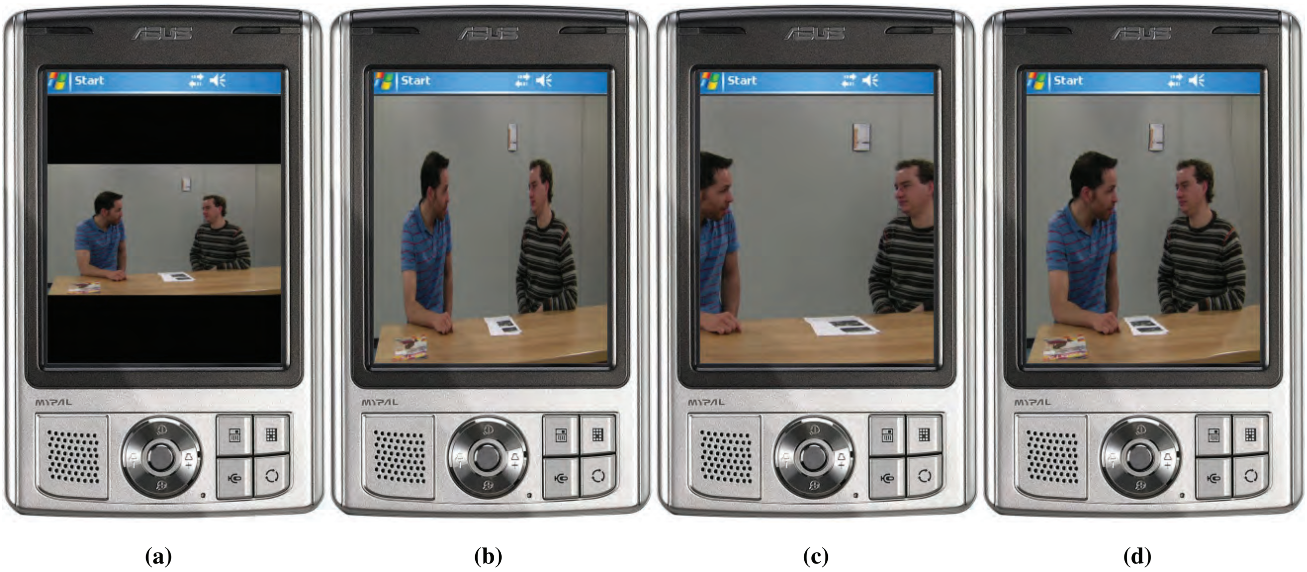
The use of multimedia content on handheld mobile devices has increased significantly in the last few years due to technological improvements of these devices and the increased capacities of wireless networks. Mobile devices also play a major role as a new medium for presentation or simply for entertainment. Today, almost all handheld mobile devices support playback of high resolution videos, but presentation of multimedia content and especially the visualization of such videos is still a great challenge due to the limited screen resolution of mobile devices.

In 2007, a very interesting technique called seam carving was presented by Avidan and Shamir [1] for content-aware resizing of images. This algorithm surpasses traditional scaling techniques because the size can be changed independently of the aspect ratio without distorting the image in most cases. This makes seam carving very interesting for video adaptation. Using this technique for videos is not straightforward and causes a large amount of instability because different seams in adjacent frames cause visible jitter. It is especially important to avoid flickering noise caused by temporal changes in the video. Therefore, we have to consider camera and object motion explicitly when applying seam carving.

In this paper, we present a novel approach for the automatic adaptation of videos and analyze the advantages and disadvantages of the seam carving technique. The distinct features of our approach are:

1. *FSCAV* is a novel and very fast technique to adapt the resolution of videos.
2. The new concept of *robust seams* preserves the visual stability of the video. In addition, we have developed a verification mechanism which classifies the quality of robust seams. This is especially relevant in case of camera motion or zoom operations.
3. The novel heuristic for quality measurement identifies adapted videos in low quality and enables the automatic switch-over to another adaptation technique when seam carving does not work well.

The outline of the paper is as follows: The following section gives an overview of previous work in the context of image and video adaptation. Section 3 illustrates our new *FSCAV* algorithm (*F*ast *S*eam *C*arving for Size *A*daptation of *V*ideos) and the new heuristic for quality measurement. In Section 4, we present experimental results and user evaluations. We conclude the paper in Section 5.



(a) (b) (c) (d)

Figure 1: Adapted videos based on letterboxing (a), scaling (b), cropping (c), and our FSCAV algorithm (d). Every technique causes some errors, e.g., lost details (a), unnatural aspect ratio (b), lost content (c), and modification of the paper on the table and the arms of the left person (d).

2. RELATED WORK

Retargeting of videos is much more difficult than image retargeting due to the increased complexity caused by the temporal dimension. For example, in order to reduce the screen size of a video sequence each frame has to consider its successor and predecessor. Otherwise, visual artifacts may occur if the frames are reduced independently. This area of research is not as well explored as the resizing of images.

Scaling the frame or *cropping* frame borders are the most frequently used methods to adapt the resolution of video. Another challenge of video adaptation are different aspect ratios of the source and target sequences. *Letterboxing* preserves the aspect ratio of the original video: E.g., if widescreen cinema movies (aspect ratio: 1.85:1) are adapted for television (aspect ratio: 4:3), black bars are added to the top and bottom of the image to get the desired aspect ratio. The full TV screen is no longer used which makes it more difficult to recognize all details.

If *cropping* is used to resize cinema movies for television, the left and the right parts of the picture are removed, which causes a different kind of information loss like truncated or missing objects. This is especially disturbing if a person is no longer completely visible or if text is cropped. Figure 1 visualizes typical problems which are caused by letterboxing, cropping and scaling. The adapted video based on our FSCAV algorithm is presented on the right of Figure 1.

Liu and Gleicher [13] use a mix of scaling and cropping to adapt cinema films to various display sizes. The reduction is formulated as a minimization task where the main concern is to keep the loss of information as small as possible. Several variations of the destination sequence are calculated first. The quality of a sequence is judged based on the loss of important content information.

El-Alfy et al. [3] developed a method to process surveillance videos in real-time. First, the video is divided into overlapping segments which are processed independently. For each segment, a trajectory for a variable-size window is processed which maximizes the captured saliency. The content of the window is then magnified and depicted in a corner of the original video.

Wang et al. [20] present a method to visualize videos on different mobile devices. The idea is to identify a region of interest in each shot and zoom into it while displaying the video. If no region of interest is detected, the entire screen is shown. It is also possible to reduce the length of the video in case of low available bandwidth. If this is the case, shots with a high user attention value are transmitted without any adaptation, while the frame rate of other shots is reduced.

The problem of visualizing videos on small displays is also handled by Fan et al. [4]. They propose a technique to present a section of the original video. In contrast to automatic cropping, the user may choose a region of the frame while watching the sequence. A fully automatic selection of such regions is also supported.

Tao et al. [19] propose an algorithm for video retargeting which uses an active window to capture the most interesting content. They assume that the actions and movements of foreground objects are always more interesting for users. Therefore, the active window is automatically set on these objects. The size of the target display specifies the area covered by the active window.

Another method of video retargeting is presented by Wolf et al. [21]. The idea is to identify the relevance of a region and shrink it depending on its relevance. Each frame considers the predecessor for the mapping. Finally, each frame is rendered by a forward mapping technique.

In previous work, we have presented a video adaptation technique where we put the focus on the preservation of foreground objects [9, 11]. Regions of interest are identified based on objects, faces and superimposed text. The goal is to maximize the visible information of these regions in the

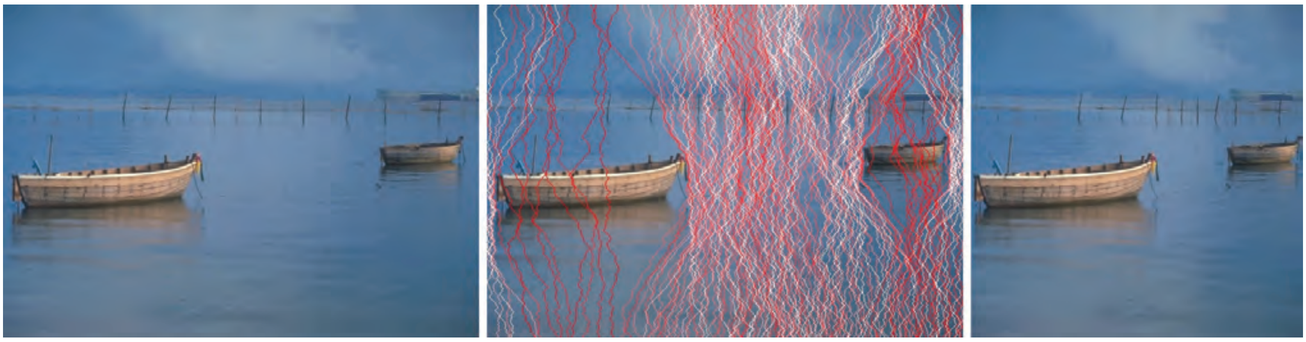


Figure 2: Left: Original image. Center: The color indicates the order of the seams (white seams with low energy are removed first, red seams are preserved as long as possible). Right: Image after the seams have been removed.

adapted video subject to the constraints to preserve a minimum perceptible size for each region but also to limit each region to a maximum reasonable size. The solution of an optimization problem (maximize the visible content) specifies the parameters for scaling and cropping [12]. An additional application was presented, that merges spatial and temporal adaptation techniques based on foreground objects to generate video summaries with limited screen resolutions [10].

All these existing methods concentrate on scaling, cropping or zooming into regions of interest. This leads to the problem that some details are lost due to the small size in the adapted video (scaling) or that relevant content is removed (cropping / zooming). The seam carving technique removes regions of low relevance within an image. The major advantage of seam carving is that these may be located in inner parts of an image.

2.1 Image adaptation based on seam carving

The seam carving technique was proposed by Avidan and Shamir [1] to enable a content-aware adaptation of the resolution of images. The idea of image adaptation based on seam carving is to identify and remove a path (*seam*) of pixels with low relevance for the content of an image. Each removed seam causes a reduction of the image size by one, where vertical seams are used to change the width of an image and horizontal seams to change its height.

A *vertical seam* $\mathbf{s} = \{(x(i), i)\}_{i=1}^{H}$ of an image I with height H is defined as a path of pixels from top to bottom with the following constraints:

1. One and only one seam pixel is selected in each row.

2. The horizontal distance between two adjacent seam pixels $|x(i) - x(i-1)|$ does not exceed a threshold T. If $T = 1$, seam pixels of vertical seams are vertically or diagonally connected (8-connected).

These conditions define a vertical seam:

$$\mathbf{s} = \{s_i\}_{i=1}^{H} = \{(x(i), i)\}_{i=1}^{H}, s.t. \forall i : |x(i) - x(i-1)| \leq T \quad (1)$$

An *energy function* defines which pixels should be selected for a seam. The optimal seam $\mathbf{s}^*$ is defined as seam with minimum cost based on an energy function e:

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} E(\mathbf{s}) = \arg\min_{\mathbf{s}} \sum_{i=1}^{H} \{s_i\} \qquad (2)$$

Avidan and Shamir [1] have evaluated the performance of several energy functions and recommended using the gradient. The cost of a seam $E(\mathbf{s})$ is defined as the sum of all absolute gradient values of path pixels. Dynamic programming is used to identify the optimal seam. A *horizontal seam* is an 8-connected path from left to right which enables the change of the height of an image. It is defined in a similar way by switching rows and columns of an image.

Depending on the desired size of the adapted image, an optimal seam $\mathbf{s}^*$ is identified in each iteration. The path pixels of the optimal seam are removed, and all pixels are shifted horizontally or vertically to fill the gap. Figure 2 shows an example of vertical seams that are detected in each iteration. The seams with the lowest costs are marked in white.

Compared to scaling, a major advantage of seam carving is the fact that changes of the aspect ratio are handled very well. The method also focuses on preserving the important content of the image. However, large objects like buildings lead to clearly visible artifacts.

2.2 Video adaptation based on seam carving

As Rubinstein et al. presented in [16], the quality of adapted videos is very low if seam carving is used on each frame separately. This leads to visible artifacts, and the video becomes blurred and shaky. The errors are caused by small differences in consecutive frames, like lighting changes, object or camera motion, noise and compression errors. The seam carving algorithm is very sensitive to these changes. Even with a static camera, small differences in pixel values lead to different seams.

The absolute difference of pixel values in the top right image of Figure 3 visualizes minor differences between two adjacent frames of the original video. The differences increase significantly after seam carving is applied on individual frames (Figure 3, lower right). In this example, background objects like buildings or pillars are constantly changing their proportions and positions. These observations underline the importance of processing the sequence as a whole and not each frame separately.

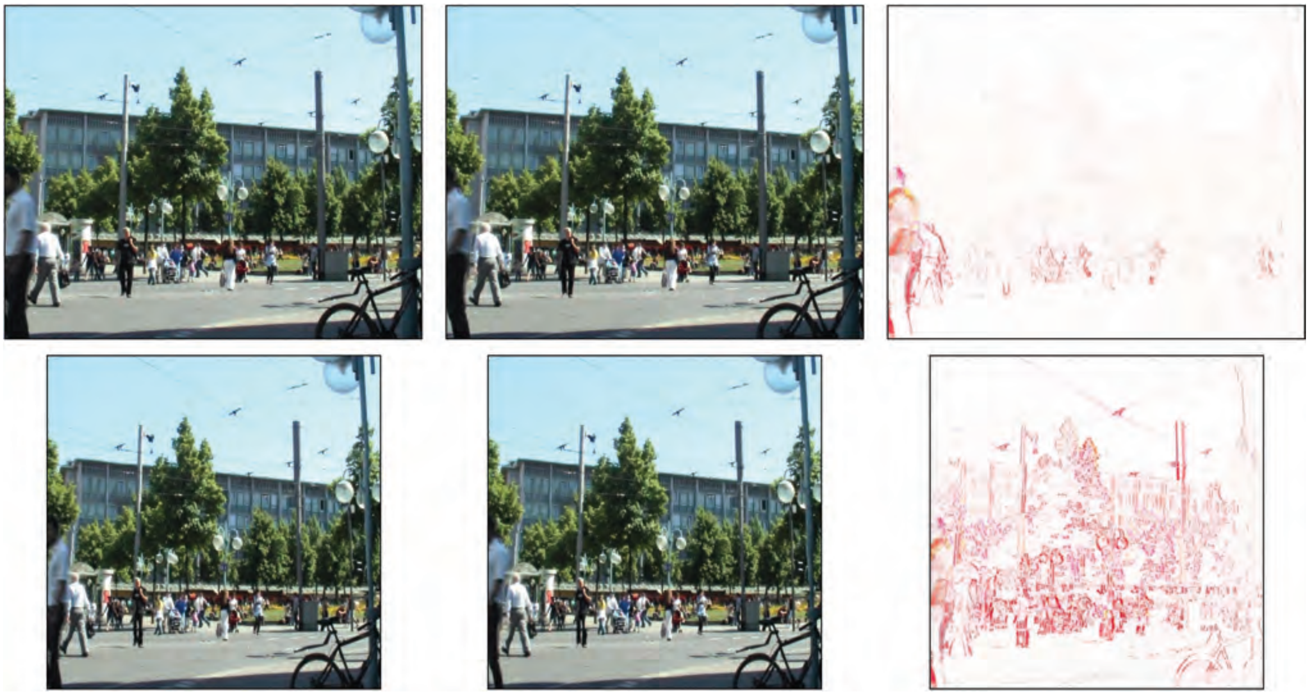


Figure 3: Top: Two original video frames and difference image. Bottom: Seam carving applied on individual frames and difference image. The image difference of the adapted video frames is significantly higher and causes a shaky video.

Rubinstein et al. [16, 17] proposed an extension of the seam carving technique for videos which removes 2D seam manifolds from 3D space-time volumes. They use graph cuts instead of dynamic programming. Whereas both are used to find the seams with low overall energy values, the latter one does not work for videos. A seam is given by a minimal cut in the graph and has to be monotonic and connected in order to be valid. They also present a new energy criterion which looks ahead in time and considers the energy brought into the destination image by removing seams.

The major disadvantage of this extension for videos is the high computational complexity of the approach. The graph cut algorithm takes a lot of processing time and is not applicable to HD videos if all pixels of the video are considered as nodes in the graph. Even for short sequences, the total amount of data required for the graph cut approach cannot be handled in memory [16]. Therefore, the graph is approximated temporally and spatially, and refined in several iterations. In spite of this optimization, the time to adapt a video is still significantly higher compared to scaling, cropping or our *FSCAV* algorithm. A precalculation of all seams for longer video sequences would generate a huge amount of data and is not applicable, too.

3. THE FSCAV ALGORITHM

The FSCAV algorithm operates on the level of shots (continuous camera recordings). Even if we adapt full-length videos, the algorithm considers all shots separately. We put the focus on the following requirements for the development of the FSCAV algorithm:

1. An optimal seam should be *robust*, that is to say the removal of this seam should not cause a blurred or shaky video.
2. To limit the computational effort and memory consumption, 1D seams should be calculated in images instead of 2D seams manifolds in 3D space-time volumes.

We assume in the first part of this section, that we want to adapt a video without object motion, and changes are only caused by camera motion. The first requirement is valid if a pixel of the optimal seam represents the same visual content in all frames (we call it *robust seam*). If a robust seam is deleted, the same object regions are removed in all frames, and the video does not become shaky. The idea of FSCAV is to identify robust seams by estimating and compensating the camera motion between all frames.

The analysis of the camera motion also makes it also possible to hold the second requirement: Instead of searching 2D seam manifolds in a 3D cube, we analyze and compensate the camera motion between consecutive frames. This enables the aggregation of pixel values respectively energy values into a single image. The seams of the aggregated image are robust seams, because they can be mapped back into all frames of the video by applying the inverse camera motion. This guarantees that optimal seams describe the same image content in all frames.

We present details of the implementation of the FSCAV algorithm in the next sections and consider the following challenges:

1. How should the algorithm avoid that an optimal seam removes too many pixels from moving foreground objects?

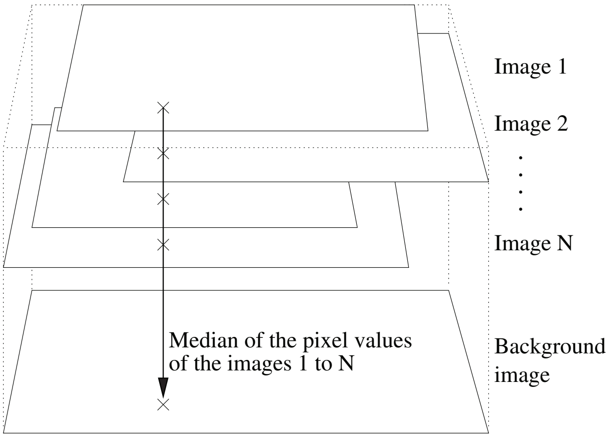


Figure 4: The background image is constructed by applying the median filter on the aligned frames.

2. How to handle robust seams that are not visible in all frames of a sequence (e.g., in case of a camera pan)?
3. How many seams can be removed from a video without reducing the visual quality too much?

3.1 Camera motion compensation

Image registration techniques can be used to track the image content in consecutive frames. We use the projective camera model [5] which uses eight parameters to describe the motion of the camera in consecutive frames:

$$x' = \frac{a_{11}x + a_{12}y + t_x}{p_x x + p_y y + 1}, \qquad y' = \frac{a_{21}x + a_{22}y + t_y}{p_x x + p_y y + 1}. \tag{3}$$

Six parameters (a_{ij}, t_x, t_y) specify affine motion and two parameters (p_x, p_y) a change of the perspective. To identify the parameters of the model, point correspondences between two frames have to be identified first. A large number of techniques have been proposed to identify characteristic feature points in images like the Moravec detector [15], corner detectors like Harris [8] or SUSAN [18], or detectors that are invariant to image transformations like SIFT [14]. Frame rates of at least 25 fps are typical in videos that were produced for cinema or television, and we do not require invariant features to track the correspondences because the camera motion between consecutive frames is relatively small. We have selected the Harris detector with sub-pixel refinement due to its better repeatability and accuracy [5].

In a second step, correspondences between features have to be identified. The features of two frames are considered only if the spatial distance between two feature points is below a predefined threshold ($T_S = 30$). The visual distance between two arbitrary feature points is defined as the sum of absolute differences in a small local window of 8×8 pixels. A greedy-based approach selects corresponding feature points by choosing the two features with the smallest visual distance. The algorithm terminates if the distance exceeds a threshold ($T_D = 4000$).

We use a robust algorithm for motion estimation due to the wrong assignment of some feature points (outliers). One of the most popular techniques is the RANSAC algorithm [7]. A subset of four corresponding features is randomly drawn and the parameters of the camera model are calculated based on these features. We classify the number of inliers (features that fit to the model) and outliers and keep the parameters with the largest number of inliers. By repeating this step (we use 300 iterations), the probability that those features chosen describe the camera motion correctly is very high.

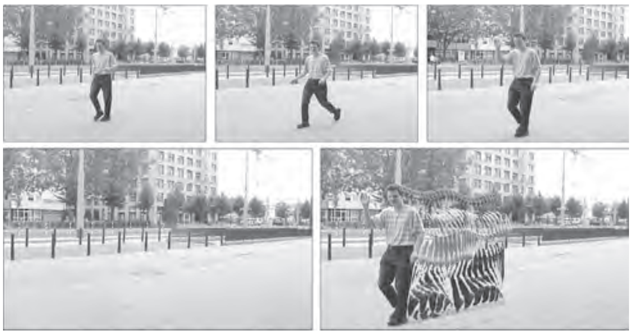


Figure 5: Top: Three sample frames of a video sequence with horizontal pan. Bottom: Background image based on the median (left) and foreground objects (right).

3.2 Aggregation of frames

In the next step an *energy map* based on all motion compensated frames is calculated. The idea is to aggregate the frames into one *background image* [6]. The center frame of a shot is selected as a reference frame and all other frames are aligned to this frame. After alignment, the background objects are always located at the same absolute pixel position. A background image without moving foreground objects is constructed by applying a median filter to all pixels at one position. Figure 4 visualizes the construction of the background image. The background in videos is preserved well, if an optimal seam is detected in the background image and transformed back into all frames of the shot.

On the other hand, this approach does not consider foreground objects at all. To reduce deformations in moving objects, we identify foreground objects by comparing each aligned frame with the background image. A pixel is characterized as object pixel if the absolute difference of aligned pixels exceeds a threshold ($T_A = 30$). All object pixels are copied into the background image which is used for the identification of optimal seams. An energy map based on the gradient magnitude of each pixel is calculated for this background image. Figure 5 shows an example of a background image and an image with foreground objects.

3.3 Identification of robust seams

To identify *robust seams*, we apply the seam carving algorithm to the energy map of the background image and get a list of suitable seams. Iteratively, the seams are mapped to the individual frames of a shot by applying the camera motion model with inverse parameters. However, the direct utilization of the mapped seams is problematic and causes visible artifacts in the adapted video.

The characteristics *completeness* and *visibility* of seams are introduced in the following, to validate whether a mapped seam is suitable or not. A seam is *visible* if it is included in all frames of the shot. For example, in case of a camera pan, the seams at the borders are not visible in most cases. A seam is *complete* if a pixel is assigned in each row (vertical seam) or in each column (horizontal seam). This corresponds with the first constraint of the definition of a seam (see Section 2.1). We classify a seam as *robust seam* if it is *visible* and *complete* in all frames. In the following, we analyze how to process seams that are not robust:

- *Missing seams:* In case of camera motion like a pan or tilt, it may occur that the image content covered by a seam is not included in all frames (these seams are called missing seams). Figure 6 (left) shows an example of missing seams, where two seams are only mapped to frame A and another to frame B. In Figure 6 (right), it is not possible to map the seam of the background image into frame A or B because it does not pass through the frame from top to bottom. Missing seams are ignored to avoid shaky videos.

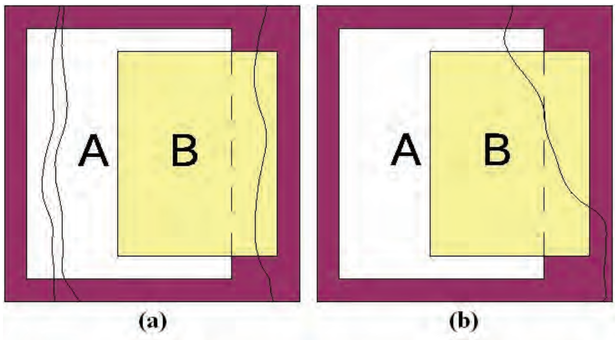


Figure 6: Example of missing seams: The seams are located in frame A or B, but not in both frames (a). The vertical seam does not pass through frame A or B from top to bottom (b).

- *Gaps in seams:* In case of a camera zoom, it is possible that horizontal or vertical gaps appear in seams due to the mapping from the background image. These gaps appear periodically and cause frayed edges in the adapted frames. The gaps are filled by interpolating adjacent seam pixels.

- *Occupied pixels:* Another typical problem is caused by rounding errors or inexact parameters of the camera model. In this case, two different seam pixels of the background image may be mapped to the same pixel in a frame. Replacing the previous pixel or removing the new pixel would lead to gaps in seams and should be avoided. Figure 7 visualizes our approach to detect the next unoccupied pixel position. Because pixel A is already occupied, the adjacent pixel B and C are checked next. The search is implemented by using a counter which changes its sign and increases its absolute value by 1 in each step.

- *Fast camera motion:* Some camera operations like long camera pans are a great challenge for this algorithm. No robust seams are detected if the first and the last frame of the shot do not share any visual content. To avoid this problem, the sequence is split in the middle if the total number of robust seams is insufficient (this depends on the desired resolution of the video). Both video segments are then processed separately. We split the sequence so that the video segments overlap by 0.5 seconds. The overlap is necessary to fade from the first to the second segment and to reduce the visible error at the transition of the segments. Each segment is split recursively until a sufficient number of robust seams are detected.

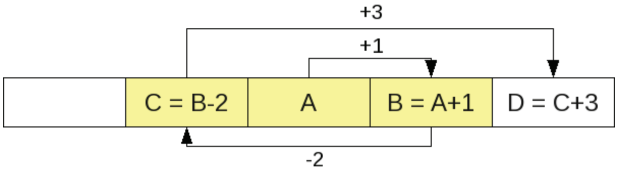


Figure 7: Example of the search for unoccupied pixels. If the position of a robust seam pixel is already occupied (yellow pixels) by another seam, a new and unused position for this pixel is found.

3.4 Quality measurements for video adaptation

In the following, we present a heuristic to estimate the degradation of the quality of an adapted video. We analyze how much the seam to be removed next reduces the image quality. Although shaky videos were not observed, other noticable image errors may occur by deformations of foreground objects. The seam carving algorithm stops if the heuristic indicates that the quality drops below a certain level. In this case, no more seams are removed from the shot and another adaptation technique is chosen to adapt the video to its final size (we use scaling).

In a first step, the parameters of the camera model are analyzed to guarantee a correct background image. Errors occur in case of large foreground objects or an insufficient number of characteristic features in the image background. An error is detected if a parameter is not within a characteristic interval or if a parameter considerably changes between adjacent frames. The binary variable $C_{i,i+1}$ defines whether the parameters of the camera model $(t_x, t_y, a_{i,j}, p_x, p_y)$ between frames i and $i+1$ are correct or not:

$$C_{i,i+1} = \begin{cases} 1 \text{ if } & |\frac{t_x}{W}|, |\frac{t_y}{H}| \leq 0.05 \;\wedge \\ & \text{rotation angle} \leq 5 \text{ degrees} \;\wedge \\ & \text{scaling factor} \leq 4 \text{ percent} \;\wedge \\ & |p_x|, |p_y| \leq 10^{-5} \\ 0 & \text{else.} \end{cases} \quad (4)$$

The thresholds were estimated empirically, the parameters W and H define the width and height of a frame. We switch to the alternative adaptation technique if the parameters of at least one pair of adjacent frames are invalid.

In a second step, we analyze the cost of the next seam to be deleted. We assume that pixels from relevant objects are deleted and errors become obvious in case of seams with high cost values. No more seams are deleted if the cost $E(s_i^*)$ of the optimal seam in iteration i exceed a threshold T_S. After the removal of the first i seams ($E(s_i^*) \leq T_S < E(s_{i+1}^*)$), the image is scaled to the final resolution. The threshold T_S is defined based on the costs of the seams in the original image:

$$T_S = \alpha \cdot E(s_1^*) + (1 - \alpha) \cdot E(s_1^{MAX}) \qquad (5)$$

$E(s_1^*)$ specifies the cost of the optimal seam in iteration 1, $E(s_1^{MAX})$ the maximum cost of a seam in this iteration. The costs are weighted by a parameter $\alpha = 0.3$.

4. EVALUATION

4.1 Fast adaptation techniques

In a first step, we compare the quality of the adapted videos based on scaling, cropping and *FSCAV*. 45 video sequences (shots) have been selected from television, the Internet or recorded with a HD camcorder. We consider individual shots because seam carving operates on the level of shots. The resolution of the sequences varies between PAL resolution (720 × 576 pixel, 25 fps) and HD resolution (1920 × 1080 pixel, 25 fps).

Due to the fact that *FSCAV* is especially sensitive to object motion, the video sequences have been grouped into 5 categories: *static* (no camera motion, average number of object pixels is less than 1 percent of all pixels), *camera motion only*, *small object motion* (1-10 percent object pixels), *high object motion* (>10 percent), and *large objects* (occupy at least 50 percent of the height or width of an image). Camera motion is usually visible in all videos except static sequences. The following table gives an overview of the sequences in each category.

Type of Sequence	Number of Videos Sequences	Length [frames]
Static	5	40 – 120
Camera motion only	12	60 – 250
Small object motion	15	50 – 500
High object motion	11	90 – 260
Very large objects	2	100 – 250

One video of each category was selected for the evaluation. Ten students between 21-24 years evaluated the test series by watching the original video sequence first and the three adapted versions in the following. The order of the adapted sequences was unknown to the users and changed with each test series. The width of each video was reduced by 45 percent (400×568 pixels in case of PAL resolution). In case of cropping, we had to set the borders manually to get acceptable results. The quality of cropping is completely unacceptable without this manual setting. The subjects filled out a questionnaire and answered the following questions for the test series: How well are details preserved? What kind of disturbing effects did you recognize? Which visual errors did you recognize? What is your overall impression of the adapted video? Answers were given on a scale from 1 (excellent) to 5 (insufficient) and additional comments were collected. Another task was to sort the adapted sequences of each test series by visual quality.

Cropping always leads to a loss of relevant content in the adapted video and achieved the worst results in the evaluation (4, *sufficient*). Ranking the quality of *scaling* and *FSCAV* is not so easy: although the average quality of *FSCAV* is better than scaling (between 3, good and 2, very good) the evaluations of the different sequences differ a lot. *FSCAV* is significantly better when object and camera motion are relatively low (*static*, *camera motion only* and *small object motion* sequences) and ranges from 1 (excellent) to 2 (very good).

The visual quality of sequences with high camera or object motion depends on the direction of the motion: The quality of *FSCAV* is very good if the camera or the objects move in parallel to the seams (e.g., vertical motion does not degrade the image quality so much when the width of a video is reduced). On the other hand, the visual quality drops significantly if objects move orthogonally to the direction of a seam. The best case occurs when the seams are uniformly distributed causing an effect similar to scaling (see Figure 8 (d)). If many seams are located in a small area, a moving object is significantly distorted in this region. This is very annoying and reduces the average visual quality to 4 (sufficient). Most problematic are sequences with large objects where the object covers a major part of a frame. The quality is insufficient in these cases and the test persons ranked the sequence even below the cropped videos.

Figure 8 shows sample video frames from each category and their adapted versions. In case of the static sequence (a), *FSCAV* preserves the building and trees much better compared to the scaled frame. In (b), the visual quality of the adapted videos are very similar. The quality of sequence (c) based on *FSCAV* is rated significantly higher, because parts of both persons are missing in the cropped video and the faces are heavily deformed in case of scaling. Many fast moving objects (cars) cross the image in sequence (d). The differences of scaling and seam carving are very low especially if we compare the cars in the foreground. Major differences can only be recognized in the building. The quality of the scaled video is much higher in sequence (e). The tram crosses the full image and is heavily deformed by the removal of seams. Another general problem of seam carving is also recognizable at the overhead contact lines of the tram: Straight lines in the original image become curved. This is very disturbing compared to scaling.

4.2 Video adaptation based on seam carving

The second part of the evaluation compares the two seam carving based adaptation techniques (*graph cut* and *FSCAV*). The visual differences of adapted videos based on these approaches are very low. In most cases, it is not possible to recognize differences if only one frame is observed (this is also true for all video sequences of Figure 8). In videos without object motion (*static* or *camera motion*) the quality of *FSCAV* adapted videos is slightly better. The reason is that the seams calculated by graph cut change from frame to frame and introduce a small amount of shakiness, which is especially obvious in background objects.

The graph cut technique generates videos of higher quality if small objects move in parallel to the direction of the seams (seams usually avoid the foreground objects). In case of *fast moving objects* or *very large objects*, the visual quality of the adapted videos based on both seam carving techniques is much lower. Due to the fact that 2D seam manifolds are connected in the temporal direction, the position of a seam pixel may only change one pixel position in adjacent frames. Even slow moving objects that move orthogonally

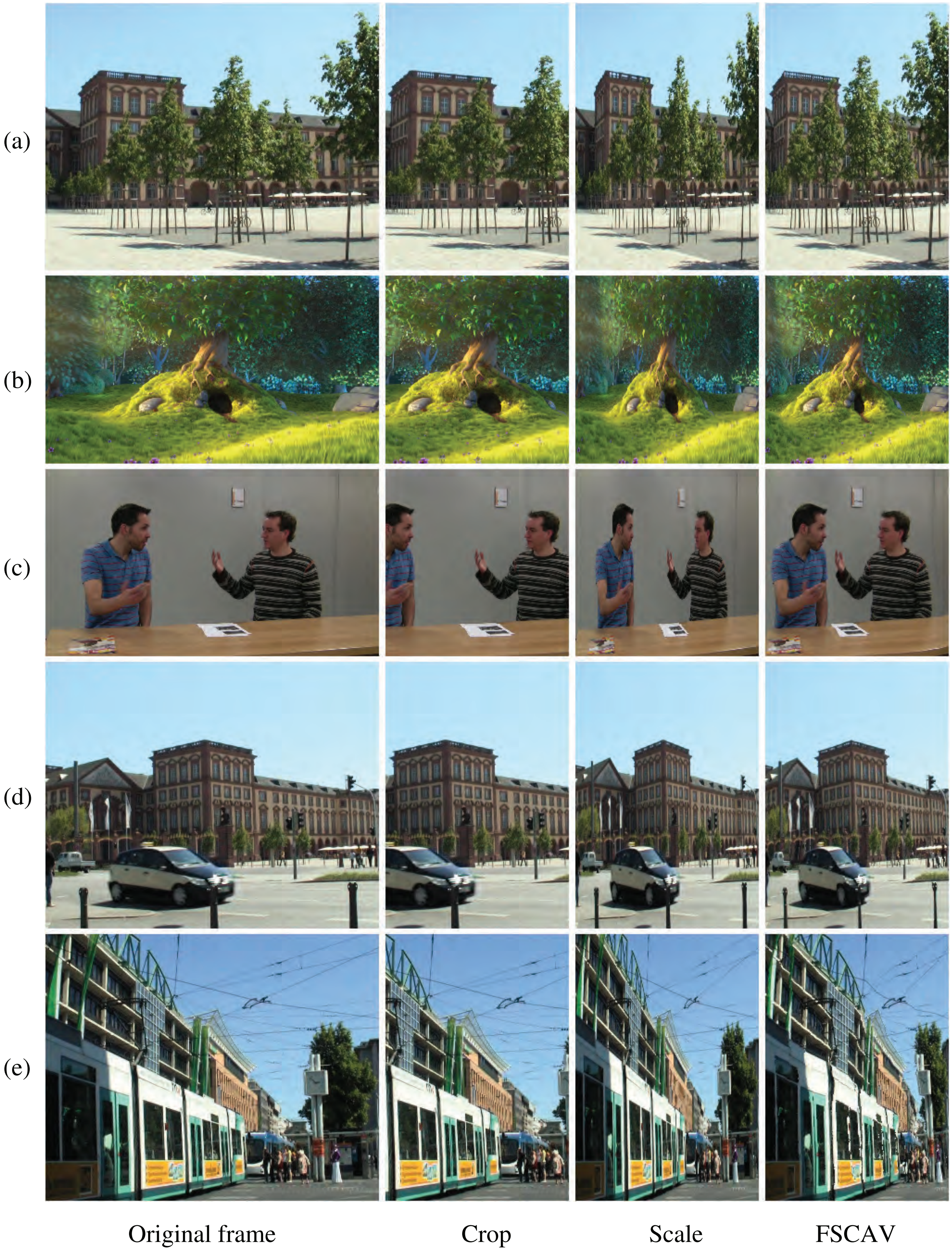


Figure 8: Top to bottom: Sample sequences from the 5 categories: static (a), camera motion only (b), small object motion (c), high object motion (d), and large objects (e). The width of each video is reduced by 45 percent.

to the seams will cross them and cause visual errors that are comparable to *FSCAV*.

Another aspect is the *computational effort* and *working memory* required by both algorithms. *FSCAV* is separated into an *analysis* and an *adaptation* phase. As result of the analysis phase, an image with global seams and parameters of the camera model is stored. The global seams are mapped to each frame in the adaptation phase. The adaptation is very efficient and can be handled nearly in real-time on a standard personal computer if the decoding and re-encoding of the adapted video stream is not considered.

The memory requirements of *FSCAV* are defined by the maximum number of frames which are loaded into memory at one time. Sequences in PAL resolution are processed entirely in memory. If sequences in HD resolution are processed, repeated decoding of a video sequence might be necessary to limit the total amount of memory. In this case, the memory requirements are less than 200 MB, but runtime increases due to hard disk access.

The memory requirements of the *graph cut* algorithm are very high making this technique only applicable to low resolution videos. Each pixel of the video sequence is represented as a node in a 3D spatio-temporal cube and several edges are connected to each node.

We analyze the memory requirements and the computational effort of the different algorithms for three test sequences (low resolution, PAL resolution and HD resolution). In our implementation, we used the max-flow algorithm presented by Boykov and Kolmogoro [2]. The following table lists the measurements of the runtime and the memory requirements on a standard personal computer (Athlon 64 Dual Core, 2.4 GHz, 2 GB RAM). In case of the PAL and HD sequences, it was not possible to run the original graph cut algorithm on our system due to the required amount of memory. The numbers in brackets show the theoretical requirements of the max-flow algorithm for these sequences. Both memory requirements an computational effort of the *FSCAV* algorithm are much lower.

	Low res.	*PAL*	*HD*
	120 × 68	720 × 576	1920 × 1080
	50 frames	150 frames	200 frames
Crop	<1 s	5 s	32 s
Scale	<1 s	6 s	36 s
FSCAV			
- Analysis	14 s	8 min	51 min
- Adaptation	1 s	11 s	83 s
Graph Cuts	17 min	N/A	N/A
	290 MB	(44 GB)	(292 GB)
Graph Cuts	N/A	49 min	123 min
(hierarchical)	N/A	530 MB	820 MB

A hierarchical approximation was proposed by Rubinstein et al. [16] to reduce the memory requirements. We implemented a hierarchical graph cut algorithm which reduces the resolution of a video to the low resolution video defined above. In case of HD videos, the spatial and temporal resolutions are reduced by a factor of 16 and 4 respectively. Seam manifolds in the smallest video cube are detected and mapped to the next level. A 16 pixel wide video slice is selected in the next hierarchy level so that the cut is located at its center. The width of the slice is set to 8 and 4 pixels in the last 2 iterations. The major disadvantage of using a hierarchical approximation is the fact that a seam may be trapped in a local minimum and detection of the optimal cut is no longer guaranteed.

4.3 Limitations

Calculation of the camera parameters fails in some shots, especially in case of fast camera motion, large foreground objects, missing feature points in background objects, or dolly shots. For instance, it was not possible to create a background image in several shots of a soccer game, and the sequence could not be processed with the *FSCAV* algorithm. The quality of scaled videos is much better if objects move orthogonally to the direction of the seams. Just like seam carving for images, *FSCAV* has problems with objects covering a large part of the screen, and some content may increase the perceptible errors significantly like straight lines which become curved. Another disadvantage in comparison to scaling or cropping is the computational effort. Despite these limitations, the visual quality of adapted videos based on *FSCAV* is significantly higher in most cases.

5. CONCLUSIONS

We presented our new *FSCAV* algorithm to adapt videos to the limited screen resolution of handheld mobile devices. A major advantage of the *FSCAV* algorithm is the possibility to change the aspect ratio of a sequence without distorting important objects. Objects appear unchanged because the algorithm removes the less relevant pixels only. The analysis is done only once for each shot. The major advantage of our approach compared to other techniques is the fact that the adaptation of a video is very efficient. User evaluations indicated the high quality of the adapted videos. The visual quality of adapted videos based on *FSCAV* and graph cut is very similar. The image background of video sequences is more stable in case of *FSCAV* whereas small foreground objects or objects in slow motion are better preserved if graph cut is used.

The *FSCAV* algorithm produces clearly visible errors in some shots. We have presented new measurements to describe the visual quality of an adapted video. *FSCAV* is replaced by scaling if the quality is insufficient. Typically, a shot which contains only few important regions is processed with *FSCAV*, while sequences with a high percentage of important screen content are scaled. The combination of both techniques is used if the number of robust seams is insufficient to perform the adaptation.

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] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *IEEE Transactions on PAMI*, volume 26(9), pages 1124–1137, September 2004.

[3] H. El-Alfy, D. Jacobs, and L. Davis. Multi-scale video cropping. *MM '07: ACM international conference on Multimedia*, pages 97–106, 2007.

[4] X. Fan, X. Xie, H.-Q. Zhou, and W.-Y. Ma. Looking into video frames on small displays. *MM '03: ACM international conference on Multimedia*, pages 247–250, 2003.

[5] D. Farin. *Automatic Video Segmentation Employing Object/Camera Modeling*. PhD thesis, Technische Universiteit Eindhoven, Einhoven, The Netherlands, 2005.

[6] D. Farin, T. Haenselmann, S. Kopf, G. Kühne, and W. Effelsberg. Segmentation and classification of moving video objects. In B. Furht and O. Marques, editors, *Handbook of Video Databases: Design and Applications*, volume 8 of *Internet and Communications Series*, pages 561–591. CRC Press, Boca Raton, FL, USA, September 2003.

[7] M. Fischler and R. Bolles. Random sample concensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Communications ACM*, volume 24(6), pages 381–395. ACM Press, 1981.

[8] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of Alvey Vision Conference*, pages 147–151, 1988.

[9] S. Kopf and W. Effelsberg. Mobile cinema: Canonical processes for video adaptation. In *Multimedia Systems*, volume 14(6), pages 369–375. Springer, December 2008.

[10] S. Kopf, T. Haenselmann, D. Farin, and W. Effelsberg. Automatic generation of summaries for the web. In *Proceedings of IS&T/SPIE conference on Storage and Retrieval for Media Databases*, volume 5307, pages 417–428, Januar 2004.

[11] S. Kopf, F. Lampi, T. King, and W. Effelsberg. Automatic scaling and cropping of videos for devices with limited screen resolution. In *ACM Multimedia (video program session)*, pages 957–958, October 2006.

[12] S. Kopf, F. Lampi, T. King, and W. Effelsberg. Automatic scaling and cropping of videos for devices with limited screen resolution. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 957–958. ACM Press, Oktober 2006.

[13] F. Liu and M. Gleicher. Video retargeting: Automating pan and scan. *MM '06: ACM international conference on Multimedia*, pages 241–250, 2006.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 60(2), pages 91–110. Kluwer Academic Publishers, November 2004.

[15] H. Moravec. Visual mapping by a robot rover. In *Proceedings of the 6th International Joint Conference on Artificial Intelligence*, pages 599 – 601, August 1979.

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

[17] A. Shamir and S. Avidan. Seam carving for media retargeting. *Commun. ACM*, 52(1):77–85, 2009.

[18] S. M. Smith and J. M. Brady. Susan – new approach to low level image processing. In *International Journal of Computer Vision (IJCV)*, volume 23(1), pages 45 – 78, May 1997.

[19] C. Tao, J. Jia, and H. Sun. Active window oriented dynamic video retargeting. *Proceedings of the Workshop on Dynamical Vision, ICCV 2007*, 2007.

[20] J. Wang, M. Reinders, R. Lagendijk, J. Lindenberg, and M. Kankanhalli. Video content presentation on tiny devices. *ICME '04: IEEE International Conference on Multimedia and Expo*, pages 1711–1714, 2004.

[21] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV-07)*, 2007.