Proc. of the 14th ACM international conference on Multimedia (MM) (Video Demo), pp. 957-958, Santa Barbara, CA, USA, October 2006
by Stephan Kopf, Fleming Lampi, Thomas King, Wolfgang Effelsberg

Automatic Scaling and Cropping of Videos for Devices with Limited Screen Resolution

Stephan Kopf, Fleming Lampi, Thomas King, Wolfgang Effelsberg
Dept. of Computer Science IV, University of Mannheim
Mannheim, Germany
{kopf,lampi,king,effelsberg}@informatik.uni-mannheim.de

ABSTRACT

A large number of previously recorded videos cannot be directly visualized on mobile devices like PDAs or mobile phones due to an inappropriate screen resolution of their displays. Transcoding can be used to change the resolution, however, the usual transcoding algorithms have problems preserving the semantic content. For instance, superimposed text is unreadable if the character size drops below a certain value. In this paper, we present a novel adaptation algorithm to scale and crop videos while preserving their semantic content. Semantic features in a shot are combined to select a suitable region to be presented in the adapted video.

Categories and Subject Descriptors

I.4.3 [**Image processing and computer vision**]: Enhancement; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding

General Terms

Algorithms

Keywords

video adaptation, content repurposing, transcoding

1. INTRODUCTION

Videos are no longer limited to television or personal computers due to the technological progress in the last years. Nowadays, many different devices such as Tablet-PCs, Handheld-PCs, PDAs, notebooks or mobile phones support the playback of videos. These devices are build for different purposes and hence vary in shape and size. This leads to a great number of different display resolutions.

The specific display resolution of a particular mobile device must be considered to archive a reasonable playback of a video. *Automatic video adaptation techniques* facilitate the adaptation and playback of videos. For such techniques, the most important goal is the preservation of the semantic information in the adapted videos. Although much work was done on the transcoding of videos, only few approaches focus on the *semantic adaptation* of videos [1, 3, 4].

MM'06, October 23–27, 2006, Santa Barbara, California, USA.
ACM 1-59593-447-2/06/0010.

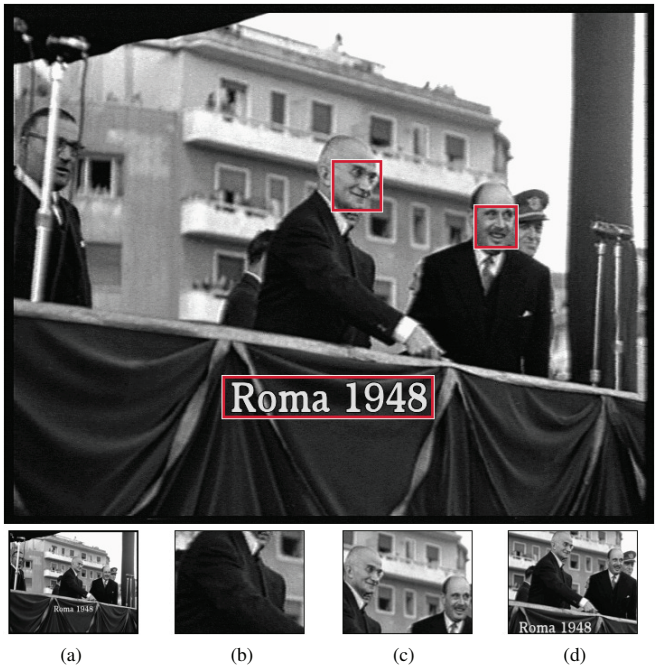
2. ADAPTATION OF THE RESOLUTION

In the following, we propose our novel adaptation approach. The change of the resolution is done by *scaling* or *cropping* (selecting a rectangular region); we use a combination of scaling and cropping. Additionally, we apply artificial camera zooms or camera motions to highlight different semantic content. Our algorithm utilizes the following four heuristics:

- Regions with relevant semantic content should be clearly visible in the adapted video. A region should not be part of the adapted video if the semantic content is no longer recognizable due to its limited resolution. For instance, text in an image is worthless if the character size is too small to be readable, or if characters are truncated. We identify *text regions* by analyzing projection profiles, *faces* by applying a neural network, as well as *persons* and *moving objects* by constructing a background image for each shot and analyzing shape features of the (segmented) foreground object [2].
- Regions without expressive content should not be part of the adapted video. Dark borders (e.g., black stripes in wide screen movies) or large monochrome regions (e.g., sky) adjoining an image border are typical candidates of irrelevant regions.
- The identified region is scaled to the appropriate screen resolution. The aspect ratio of the selected region should be identical with the aspect ratio of the adapted video.
- The goal is to maximize the visible information in each frame. An artificial camera motion is inserted if several important regions are detected. A continuous modification generates an artificial camera motion; a sudden shift of the camera produces a hard cut.

A rating of semantic features is necessary to identify a region of interest. The size of a feature in the adapted image is relevant for the rating. A pure scaling or cropping does not produce acceptable results (see Figure 1 (a) and (b)). The example in Figure 1 (c) and (d) illustrates that a combination of scaling and cropping leads to better results.

The following approach guarantees a well-balanced trade-off between scaling and cropping: Each semantic feature is characterized as a rectangle. We assume a proportional coherence between the size of a semantic feature in the adapted video and the visual information in the image. A *minimum perceptible size* is defined for each feature. The content is no



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

Figure 1: Adaptation of the resolution with scaling (a) and cropping of image borders (b). The quality of the adapted image is much better if two (c) or three (d) semantic features are considered.

longer recognizable if the size of a feature drops below this value. Additionally, we can define an upper size for each feature (*maximum reasonable size*). For instance, if a text is large enough to be readable, an additional enlargement of the characters does not provide new information.

It is our goal to specify the position and size of the rectangular region which maximizes the visible information in the adapted video. The total amount of information $V_{sum}(R)$ based on the semantic features i in the selected region R is defined as:

$$V_{sum}(R) = \sum_i S_i(R) \cdot V_i(R) \quad \text{with} \tag{1}$$

$$V_i(R) = \begin{cases} \frac{H_{max}}{H_i(R)} & H_i(R) > H_{max}, \\ \frac{H_i(R)}{H_{max}} & H_{min} \leq H_i \leq H_{max}, \\ 0 & H_i(R) < H_{min}, \end{cases} \tag{2}$$

$$S_i(R) = \begin{cases} 1 & \text{if } i \text{ is part of } R, \\ 0 & \text{else.} \end{cases} \tag{3}$$

$V_i(R)$ estimates the information of the semantic feature i and the selected region R. The binary variable $S_i(R)$ specifies whether the semantic feature i is totally included in the region R. Usually, the aspect ratios of the region and the adapted video do not match. To fix this problem, the height or width of the selected region is enlarged until both ratios fit. The entire frame is selected if no semantic feature is detected at all.

So far, the selected region maximizes the visible information for each frame. This approach is not suitable for videos because small variations of the size or position of a region between adjacent frames have a very disturbing jitter effect. On the other hand, a continuous change of the position and size of the visible image is a typical element in the video production process and should be possible within a shot.

The size of the region is adjusted in a first step. A linear function calculates an approximation of the height of the regions for all frames in a shot. The width depends on the aspect ratio of the adapted video. In a second step, the position of the region is smoothed in each frame. The horizontal and vertical position values are approximated independently with a linear function. For instance, the camera will track a moving object so that it is always centered in the adapted video.

We apply the following procedure if only two relevant regions are detected: If the size and position of these regions do not significantly change within a shot, one region is selected for the first frame and the second region is used for the last frame in this shot. A linear transition between these two regions is calculated if the distance between both regions is under a certain threshold. Otherwise, an artificial hart cut is inserted in the middle of the shot.

3. CONCLUSIONS

Video clips of our examples and a video demo are available on our website [5]. Additionally, the source code of our novel algorithm is available.

The semantic adaptation of videos for mobile devices is an important issue due to the large number of different devices and the huge amount of digital videos. We focused on the *semantic adaptation of the image resolution* in videos. Another field of application is the publication of videos on the Web where a limited video resolution is often desired. Our new algorithm for the adaptation of the screen size is also suitable to improve the quality of these videos.

4. ACKNOWLEDGMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG).

5. REFERENCES

[1] L.-Q. Chen, X. Xie, X. Fan, W.-Y. Ma, H.-J. Zhang, and H.-Q. Zhou. A visual attention model for adapting images on small displays. In *ACM Multimedia Systems Journal*, volume 9(4), pages 353–364. ACM Press, 2003.

[2] 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*, volume 8, pages 561–591. CRC Press, Boca Raton, FL, USA, 2003.

[3] X.-S. HUA, L. LU, and H.-J. ZHANG. Photo2video. In *ACM International Conference on Multimedia*, pages 592–593. ACM Press, 2003.

[4] J.-G. Kim, Y. Wang, and S.-F. Chang. Content-adaptive utility-based video adaptation. In *IEEE International Conference on Multimedia and Expo*, pages 281–284, 2003.

[5] Movie Content Analysis Homepage (MoCA), Department of Computer Science IV, University of Mannheim, http://www.informatik.uni-mannheim.de/pi4/projects/moca/Project-ResolutionAdaptation.html, last checked: July 2006.