A Multi-Operator Retargeting Scheme for Compressed Videos

Dai-Yan WEI, Yung-Chieh CHOU and Po-Chyi SU, *Member, IEEE*

Abstract**--This research presents a multi-operator retargeting scheme, in which content-based cropping, seam carving/insertion and scaling are applied sequentially to adjust the video frames to the target resolution. Assuming that the video to be processed is encoded with H.264/AVC, compressed-domain data in the bitstream are utilized to classify video shots into different types for further processing. SLIC superpixels are formed to identify boundaries of objects and a saliency map will determine the visual significance of frame areas for appropriate cropping. A motion feature map is constructed to locate moving objects or contents so that possible distortions on them can be avoided. For static scenes, a local-significance-aware seam carving scheme based on one-dimensional gradients is applied. A SSIM-based blur detection method is also developed to extract sharp foreground objects. The experimental results show that the proposed method performs well in various kinds of video shots.**

I. INTRODUCTION

Content retargeting that modifies image/video resolutions in an adaptive way for better demonstrating these imagery data has drawn tremendous research interests in recent years. The methodologies of retargeting can be roughly classified into four categories [1], including content-based cropping, warping, seam carving and the usage of multi-operator. Among them, multi-operator approaches employing various operations to cope with different situations should have better performance. Most retargeting designs focus on still images given the facts that videos are composed of images/frames and that offline transcoding to modify the resolution is allowed in considered applications. Since computational load is always an important factor in video processing, we aim at developing a multi-operator retargeting scheme for videos stored in a compressed format, e.g. H.264/AVC. Compressed-domain data are used to decide suitable operations to facilitate content retargeting.

II. PROPOSED SCHEME

A. Framework

Fig.1 shows the framework of the proposed scheme. The input data is an H.264/AVC compressed video. The scene-change detection mainly utilizing coding modes helps to divide the video into shots. The motion information is further used to classify shots into two basic types, i.e., static and non-static shots, which are processed in different manners. Static shots contain video frames with a steady background and possible moving objects within frames. The proposed scheme will employ a so-called local-significance-aware seam carving to modify the width and height of frames so that the resolution can be closer to the target one. To facilitate effective boundary cropping, three operations are executed in parallel. A visual saliency map based on luminance difference of multi-scale neighboring pixels is constructed to highlight important areas of frames. SLIC superpixels [2] are formed to appropriately protect the boundaries of content in the subsequent cropping process. We observed that many static scenes in videos have a blurry background so a SSIM-based blur detection [3] is used to identify the areas of sharp foreground objects and blurry background. It should be noted that these two processes are applied to the first (scene-change) frame only, which is usually encoded as an I frame and can be extracted efficiently. We will further take the motion information from other frames, which are assumed to be encoded as P frames. These extracted motion data will construct a motion feature map. The visual saliency or possible extracted sharp objects will be combined with the motion feature map to determine suitable cropping limits. For non-static shots, the motion features are rather unstable and the seam carving may not be a good option either. Therefore, we use the visual saliency only to decide suitable cropping locations. Details of each step are described below.

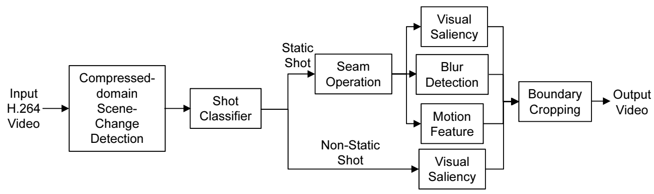

Fig. 1 The framework of the proposed scheme

B. Compressed-domain scene change detection

The scene-change detection procedures basically follow our previous work [4]. It is assumed that the encoding adopts a fixed GOP, which contains I and P frames only. The basic idea is to compare the intra coding modes of adjacent I frames. A large difference of intra coding modes may indicate that a scene change happens between them so these two frames are expanded from the compressed stream to further examine their color histograms. If a scene change may exist, the numbers of intra coding modes in the P frames between the two I frames will be checked to determine the exact scene-change frame.

C. Shot classifier

The frames following a scene-change will be considered in the same video shot. Shot classification is based on the motion features in inter-coding frames around the frame boundaries. Static shots are those having a static background with certain moving objects in the middle while non-static shots are those

This research is supported by the Ministry of Science and Technology in Taiwan, R.O.C., under Grants MOST 105-2221-E-008-086 and MOST 106-2221-E-008-086.

with large camera motions. In static shots, the leading scene-change frame will be processed by seam carving/insertion first. The same seams will be cropped/added in the entire video shot. The visual saliency and blur detection are then applied to the scene-change frames only for determining suitable cropping limits. In non-static shots, the I frames in the same shot are extracted to compute the saliency for deciding cropping limits.

D. Visual saliency and blur detection

We modify [5] to calculate the visual saliency map. Following our previous scheme, SCAN [6], superpixels are formed to determine more accurate boundaries of possible objects. If reduction of frame width is required, horizontal scans from both sides of a frame will be applied and adjacent blocks/areas transforming from background to foreground will indicate the cropping limits. The blur detection process [3] is mainly used in the frames with a blurry background. The idea is to apply blurring operations on a frame first and then calculate SSIM before and after such processing. If a frame is ruled as one containing a blurry background, the extracted foreground objects will serve as the cropping limit. Otherwise, the visual saliency will be the major reference for cropping.

E. Local-significance-aware seam carving

The seam carving operation in the proposed scheme helps to reduce the degree of cropping, which may cause content loss near the frame boundaries. Unlike traditional seam carving methods, our scheme [6] relies on one-dimensional gradients. Some saliency points are identified, either by large gradients or the masks formed via object recognition or motion features, to prevent selecting the seams passing through sensitive areas. A slanted-line detection is used to avoid distortions introduced by the seam operations on these slanted straight lines, which have weak gradients. The design principle here is to avoid any distortion by considering the local significance.

F. Motion feature map

The motion feature map is formed by the extracted motion vectors in the compressed stream. The objective is to avoid fully expanding the video frames for processing. Since some frames may contain slight camera motions, global motion vectors will be checked and subtracted from the extracted motion vectors. Besides, if consistent global motions exist, the seam carving process will not be applied to prevent possible "background content drift.". Blocks with large motions indicate important objects so more suitable cropping limits can be identified than those determined by using still frames only.

III. EXPERIMENTAL RESULTS

The experiments are applied on various kinds of videos to show the feasibility of the proposed scheme.

A. Experiment results of static video retargeting

For the cases of static shot video retargeting, we classify videos using the motion information of the macroblocks near the boundaries of each frame, and apply seam carving, blur detection, content-based cropping and scaling to reach the target aspect ratio. Fig. 2 shows the experimental results of static shot video "Silent". Twelve continuous frames are shown and the lady shown in the middle of these frames is preserved well. Using the visual saliency and the motion features, we won't crop off the body of the lady.

(a) (b)

Fig. 2. The experiment results of static shot "Silent". (Ratio = 0.5) (a) Original frames (frame 150-161) and (b) seam carving + cropping

B. Experiment results of non-static video retargeting

For the cases of non-static shot video retargeting, we apply content-based cropping and scaling only. As shown in Fig. 3, the relative positions of the objects are kept well so the objects will not be stretched severely by scaling. The entire face of the person can be fully contained in this zooming-in case.

(a) (b)

Fig. 3. The experiment results of non-static shot "Coraline". (Ratio = 0. 5) (a) Original frames (frame 11-22) and (b) the results of cropping

REFERENCES

[1] D. Vaquero, M. Turka, K. Pullib, M. Ticob, and N. Gelfandb, "A survey of image retargeting techniques," Proceedings of SPIE the International Society for Optical Engineering, vol. 7798, p. 779814, 2010.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "SLIC superpixels," EPFL Technical Report, no. 149300, June 2010.

[3] Yung-Chieh Chou, Chih-Yun Fang, Po-Chyi Su and Yu-Chien Chien, "Content-based cropping using visual saliency and blur detection," IEEE Ubi-Media, Pattaya, Thailand, 2017.

[4] Po-Chyi Su, Chin-Song Wu, "Efficient copy detection for compressed digital videos by spatial and temporal feature extraction," Multimedia Tools and Applications, Jan. 2017, Vol. 76, Issue 1, pp 1331-1353.

[5] S. Montabone, and A. Soto., "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," Image and Vision Computing, vol. 28, no. 3, pp. 391-402, 2010.

[6] Po-Chyi Su, Zi-Hao Xiang and Hao-Wei Wu, "SCAN: A multi-operator image retargeting scheme," APSIPA Annual Summit and Conference 2014, Siem Reap, City of Angkor Wat, Cambodia, Dec. 9-12, 2014.