Stabilizing Video while Keeping Resolution and Capturing Intention

Bing-Yu Chen* Jong-Shan Lin* Wei-Ting Huang*
National Taiwan University

1 Introduction

Annoying shaky motion is one of the significant problems in home videos, since hand shake is an unavoidable effect when capturing by using a hand-held camcorder. Video stabilization is an important technique to solve this problem. However, the stabilized videos resulted by current methods usually have decreased resolution and are still not so stable. In this sketch, we propose a novel, robust, and practical method of video stabilization while considering users' capturing intention. Our method can produce full-frame stabilized videos, and not only the high frequency shaky motions but also the low frequency unexpected movements are removed. To guess the user's capturing intention, we first consider the regions of interest (ROI) in the video to estimate which regions or objects the user wants to capture, and then use a polyline to estimate a new stable camcorder motion path while avoiding the user's interested regions being cut out. Then, we fill the dynamic and static missing areas caused by frame alignment from other frames to keep the same resolution and quality as the original video. Furthermore, we smooth the discontinuous regions by using a 3D Poisson-based method. After the above automatic operations, a full-frame stabilized video can be achieved and the important regions can also be preserved.

2 Camcorder Path Estimation

To estimate the global camcorder motion path, we first extract the feature points of each frame by SIFT (Scale Invariant Feature Transform), which is invariant to scaling and rotation of the image. The feature points on every consecutive frames are matched if the distances between the feature descriptions are small enough and RANSAC (RANdom SAmple Consensus) is used to select the inliers of the matched feature pairs. Then, a 3×3 affine transformation matrix between the two consecutive frames can be achieved, which contains six parameters. Once the transformation matrices between the consecutive frames are obtained, all of the transformations can be combined to derive a global transformation chain.

To extract the video ROI from the input video, we take the temporal and spatial attention models into consideration to produce the spatiotemporal video saliency maps. The spatial attention model is based on image ROI and the temporal attention model is extracted by considering the moving objects in the video. To obtain the spatiotemporal attention model by combining the temporal and spatial attention models, the spatiotemporal saliency map $Sal(i)$ of frame i is defined as $Sal(i) = kt_i \times SalT(i) + ks_i \times SalS(i)$ [Zhai and Shah 2006], where $SalT(i)$ and $SalS(i)$ are the temporal and spatial saliency maps of frame i, and the weighting parameters kt_i and ks_i are defined as $kt_i = \alpha_i/(\alpha_i + \beta)$ and $ks_i = \beta/(\alpha_i + \beta)$, where $\beta \in (0, 1)$ is a constant value and $\alpha_i = SalT(i)/(\max(SalT(i)) - \min(SalT(i)))$.

To obtain a stabilized camcorder motion path without not only the high frequency shaky motions but also the low frequency unexpected movements, we use a polyline to fit the estimated global camcorder motion path while considering video ROI. Once the camcorder motion path is fitted by a polyline, the video frames are aligned along the polyline fitted camcorder motion path.

3 Video Completion

After aligning the video frames along the stabilized camcorder motion path, there are several missing areas in the new stabilized video. To complete the video, we first detect the moving objects to segment the video to a static background region and some dynamic moving object regions. Then, we complete the missing areas by filling dynamic regions and static regions respectively. In order to detect moving objects, we evaluate the optical flow of them by using an efficient and less noisy optical flow approach to obtain the motion vector of each pixel, and the length of the motion vector shows the motion value. The motion values in the moving object regions are considered to be relatively larger than those in the static background region. If the missing area falls in the regions where the neighboring pixels have been masked as the dynamic region, this area is treated as the dynamic region and motion inpainting [Matsushita et al. 2006] is used to complete the area, otherwise we recover the area by mosaicing.

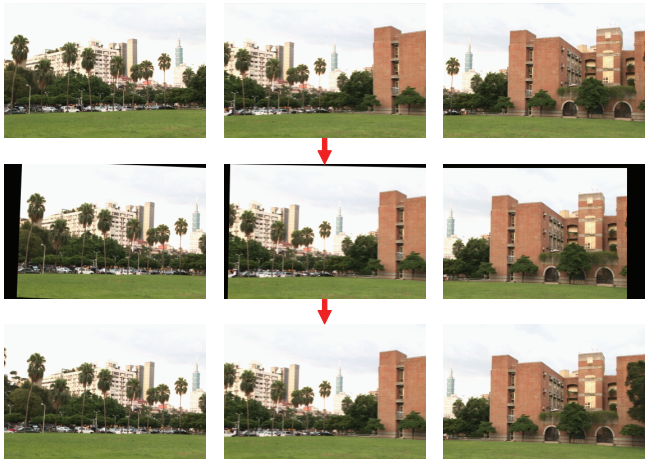

Figure 1: *Top: The original video with annoying shaky motions. Middle: Stabilized frames with missing areas. Bottom: Our result.*

Although the missing areas caused by the stabilized camcorder motion path are completed, there may be a discontinuous boundary between the recovered pixels and the original frame, since the missing areas may be large and needed to be filled from the frame far from the current one. In order to keep the spatial and temporal continuity, we provide a 3D Poisson-based smoothing method before filling in a pixel from other frames, the Poisson equation is applied to obtain a smoothed pixel by considering its neighboring pixels in the same frame and neighboring frames.

4 Result

In Figure 1, the user wants to use the hand-held camcorder to capture a panorama view. Without a tripod, the captured video are shaky due to the hand shakes. Although the camcorder motion path can be stabilized by a polyline-based motion path, without taking video ROI into consideration, the stabilized motion path may cause the building to be cut out.The bottom row of Figure 1 shows our result which is stabilized as captured by using a tripod and the building could be preserved in the stabilized video.

References

MATSUSHITA, Y., OFEK, E., GE, W., TANG, X., AND SHUM, H.-Y. 2006. Full-frame video stabilization with motion inpainting. *IEEE TPAMI 28*, 7, 1150–1163.

ZHAI, Y., AND SHAH, M. 2006. Visual attention detection in video sequences using spatiotemporal cues. In *Prof. ACM MM 06'*, 815–824.

*e-mail: {robin,maruko,weiting}@cmlab.csie.ntu.edu.tw