Activity-Aware Video Stabilization for BallCam

Ryohei Funakoshi
Tokyo Institute of Technology
funakoshi.r.aa@m.titech.ac.jp

Vishnu Naresh Boddeti
Carnegie Mellon University
naresh@cmu.edu

Kris Kitani
Carnegie Mellon University
kkitani@cs.cmu.edu

Hideki Koike
Tokyo Institute of Technology
koike@cs.titech.ac.jp

ABSTRACT
We present a video stabilization algorithm for ball camera systems that undergo extreme egomotion during sports play. In particular, we focus on the BallCam system which is an American football embedded with an action camera at the tip of the ball. We propose an activity-aware video stabilization algorithm which is able to understand the current activity of the BallCam, which uses estimated activity labels to inform a robust video stabilization algorithm. Activity recognition is performed with a deep convolutional neural network, which uses optical flow.

Author Keywords
BallCam, Activity Segmentation, Video Stabilization

Introduction
In the ball sports,ball plays a central role of sports, and players moving around the ball. If we see the sports from the ball's viewpoint,that movie gives us more sense of realism,and football players can check their formation by view from above. But ball moves extremely while playing,so we need some stabilization for recorded movies. In this paper,we make BallCam, similar to [1], which is an action camera embedded in an American Football. Since the motion of BallCam induces extreme levels of camera motion, there is a need to innovate new algorithms for stabilization. The purpose of BallCam is to capture and generate a stabilized video from the ball's point-of-view to enable an new viewing paradigm for American football.

One of the primary reasons why it is difficult to stabilized BallCam video is because the motion of the camera transitions between very different types of motion based on the action of the athlete. This suggests, that it is important that the video stabilization algorithm is able to understand ball's actions and the impact it will have on camera motion.

UIST'16 Adjunct October 16-19, 2016, Tokyo, Japan

ACM ISBN 978-1-4503-4531-6/16/10.

DOI: http://dx.doi.org/10.1145/2984751.2984767

Based on this observation, we propose an *activity-aware* algorithm for stabilizing videos. To enable accurate action segmentation, we utilize a neural network trained to recognize various camera actions.

Related Work
Throwable Cameras: With the current advances in action camera, there has been several ball mounted camera systems. Kitani et.al. [1] introduced BallCam!, with a single camera embedded at the side of an American football, providing a unique POV enhancing the viewing experience of such sports. The camera in our BallCam system is embedded in the tip of the American football.

Video Stabilization: Video stabilization is typically done by estimating the camera path, smoothing the camera path and synthesizing the stabilized video using the smoothed camera path. Grundmann et. al.[2] approximate the smoothed camera path as segments of no,linear and parabolic motion. The entire camera path is optimized subject to constraints on the possible affine transformations that the crop window can undergo. We propose a video stabilization algorithm that is aware of the activity of the camera and generates specialized optimization constraints for different motion types.

Approach
Stabilizing videos from ball camera systems that undergo extreme egomotion and varied motion types is a challenging task. This challenge is exacerbated by radial distortion in the lens. To address these challenges, we propose an activity-aware video stabilization system.

In our activity-aware video stabilization,We first account for radial distortions in the image by calibrating the lens. After that,we perform video stabilization based on the linear programming framework proposed by Grundmann et.al. [2].

And we customize the stabilization across different motion types by adopting motion specific constraints in the linear programming framework. We propose a deep convolutional network to segment the video into different action. We categorize the different motion types that our BallCam undergoes: hold, throw, fly and run. Figure 1 shows our network architecture for recognizing the different action types. Our network takes dense optical flow [3] features stacked over multiple frames as input.

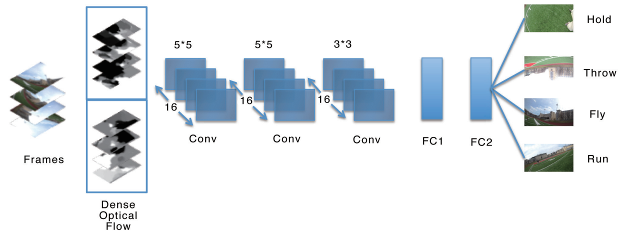

Figure 1. BallCamNet. Convolutional neural network architecture for recognizing the actions for ball camera systems.

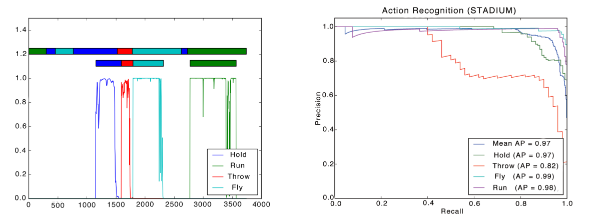

Figure 2. Left:Action Recognition Results for the test Sequence and Right:Precision-Recall curves for the test Sequence

Video Stabilization

We build upon the video stabilization algorithm introduced by Grundmann et.al. [2] where the camera path is approximated by segments of constant, linear and parabolic motion. The stabilized camera path P_t can be computed from the camera path C_t (computed from feature tracks) by an update transform B_t, $P_t = C_t B_t$. Then we crop images according to crop window matrix B_t The stabilized camera path can be optimized subject to constraints on the update transform $p_t = \begin{pmatrix} a_t & b_t & d_t^x \\ c_t & d_t & d_t^y \\ 0 & 0 & 1 \end{pmatrix}$, and the four corners $c_t i = (c^x_{ti}, c^y_{ti}), i = 1, \dots, 4$ of the crop window in frame t,

$$\begin{aligned} \min_{a,b,c,d,d^x,d^y} \quad & w_1|D(P)|_1 + w_1|D^2(P)|_1 + w_1|D^3(P)|_1 \\ s.t. \quad & 0.9 \le a,d \le 1.1, -0.1 \le b,c \le 0.1 \\ & -0.05 \le b+c \le 0.05, -0.1 \le a-d \le 0.1 \\ & 0 \le d^x + ac^x_{0i} + bc^y_{0i} \le w \\ & 0 \le d^y + cc^x_{0i} + dc^y_{0i} \le h \end{aligned}$$

Different motion types needs different constraints on the transformation that the crop window can undergo. We use the motion segmentation outputs from our deep learning system to guide our video stabilization.

Results

We evaluate the various components of our activity aware video stabilizing method for action cameras, namely, learning based motion segmentation, and activity aware video stabilization.

Action Recognition Results

Our deep convolutional network for activity segmentation was trained to predict the motion type on a per-frame basis. The network parameters were optimized using stochastic gradient descent with momentum. Figure 2 shows a plot of the classification output for each action category and a plot of the precision-recall curve for BallCam sequence. In the classification output,the top row and the bottom row shows the predicted and ground truth activity segmentation. Our deep learning based system for motion segmentation is highly effective resulting with an Average Precision (AP) of 0.97 for the test sequence.

Conclusion

We presented a robust video stabilization algorithm for ball-camera systems. Our activity-aware stabilization approach is composed of two main modules: deep learning based activity classification, and a customized motion-aware video stabilization algorithm. The motion labels from our activity recognition network serve to guide the motion-aware video stabilization algorithm. We showed that our deep learning based activity segmentation results is highly-performant and can effectively guide our video-stabilization algorithm. Our experiments showed that being aware of the motion type and designing custom algorithms for different motion types results in visually pleasing stabilized videos in comparison to many state-of-the-art baseline approaches.

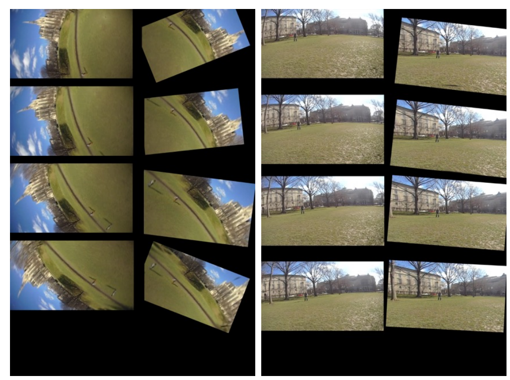

Figure 3. Raw video and Qualitative results of our proposed video stabilization method for flying(left) and running(right) sequences

REFERENCES

1. Kitani, K., Horita, K., Koike, H.: Ballcam!: dynamic view synthesis from spinning cameras. In: Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology, ACM (2012)
2. Grundmann, M., Kwatra, V., Essa, I.: Auto-directed video stabilization with robust l1 optimal camera paths. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2011)
3. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Image analysis. Springer (2003) 363–370