Automatic real-time zooming and panning on salient objects from a panoramic video

Vamsidhar Reddy Gaddam, Ragnar Langseth, Håkon Kvale Stensland, Carsten Griwodz, Pål Halvorsen

Simula Research Laboratory & University of Oslo, Norway

{vamsidhg, ragnarla, haakonks, griff, paalh}@ifi.uio.no

ABSTRACT

The proposed demo shows how our system automatically zooms and pans into tracked objects in panorama videos. At the conference cite, we will set up a two-camera version of the system, generating live panorama videos, where the system zooms and pans tracking people using colored hats. Additionally, using a stored soccer game video from a five 2K camera setup at Alfheim stadium in Tromsø from the European league game between Tromsø IL and Tottenham Hotspurs, the system automatically follows the ball.

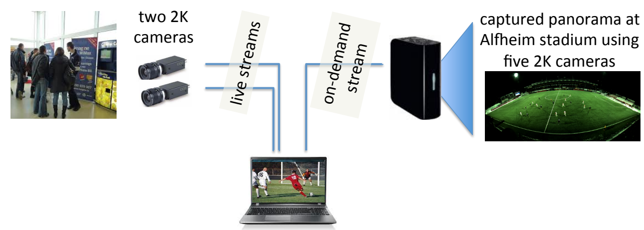

Figure 1: Demonstration setup

1. INTRODUCTION

In several fields, especially in surveillance and sports, Pan-Tilt-Zoom (PTZ) cameras have gained popularity with their ability to use the camera sensor efficiently. Yet, one problem that a physically moving system inherently has is that it does not capture and store the data that is not currently in it's field of view. Here, we present a system that generates panorama video in real-time, and we introduce a *virtual* PTZ camera system extracting data from the panorama that is capable of following targets in real-time.

Such a system can be useful when multiple PTZ camera views are demanded for multiple targets. In order to keep the costs low, it is useful to use the same camera setup to create multiple camera views. Several panoramic capture systems have been demonstrated before with different shortcomings like ease of mobility [3], cost [7] and manual input [2].

In this demonstration, we present two versions of our system. The first demonstration is a live, scaled down version of our sport stadium system using two 2K cameras zooming into live recorded video at the conference venue. A few funny colored hats are provided to users that act like targets. Then, the system will be able to track and zomm into the desired target automatically. The second is a demonstration that uses a recorded five 2K-camera panorama video from Alfheim stadium where our current prototype is installed. The demo will allow users to interact with the virtual camera in both stored and live scenes.

ACMMM'14, November 3-7, 2014, Orlando, Florida, USA.
ACM xxx-x-xxxx-xxxx-x/xx/xx.

2. SYSTEM OVERVIEW

The system presented in this demo is part of a larger system where the current prototype is running at Alfheim stadium in Tromsø. This system has been presented before [5, 4], but without the possibility to automatically zoom and pan into a high-resolution panorama tracking selected objects.

The relevant part of the system in this context is divided into a panorama generation part, the object-tracking part and a video delivery part. The delivery side of the system supports both user-controlled interactive and automatically tracking virtual cameras. A sketch of our system is given in figure 1, and in the subsections below, we give more details of the different components of our system.

2.1 Real-time Panorama Video Capture

To generate a live panorama video, we use two 2K Basler industry vision cameras each of resolution 2046×1086 pixels along with 8 mm lenses. The cameras are shutter synchronized using our custom trigger solution. A recording machine grabs the frames, aligns them and stitches into a cylindrical panorama. The system automatically avoids effects due to parallax by finding dynamic seams that do not pass through moving objects.

Furthermore, to capture the entire soccer field, we use the same Basler cameras. To maximize the panorama resolution, the cameras are rotated by 90°, giving a panorama video of 4450x2000. Moreover, the cameras are mounted in a circular pattern. This means that they are pitched, yawed and rolled to look directly through a point 5 cm in front of the lenses reducing the parallax effects, and the capture system also dynamically determines the required exposure depending on changing light conditions. After the generation of the panorama video, it is encoded and compressed using x264. At the moment, with a focus on video quality, the system still demands a large network bandwidth to the full resolution panorama video. However, there are large potentials for trading of quality for lower bandwidth requirements.

2.2 Live object tracking

In the current setup, the object tracking module is a stand-alone program to support later improvements like distributed processing and multi-sensor fusion. The current tracker is a relatively simple tracking by detection. The major requirement is to be able to track objects introduced or disappeared during the process.

First, an adaptive background subtraction is performed, where the background is updated at regular intervals. Then, color-thresholding followed by object detection leads to a position of objects of different color. Once the positions for a video segment are succesfuly found, the position data is made available to the viewer which is described next. Even though this process currently shares the same resources as the stitching process, tracking is acheived with an average execution time of 7 ms per frame.

2.3 Live Automatic/Manual virtual viewer

Once the panoramic videos are encoded, they are made available to the client program running the virtual camera via HTTP segment streaming along with the position data if available. The client has two modes, an automatic mode and a manual mode. The manual mode allows the user to interact with the virtual camera, where one can pan/tilt/zoom manually into the live stream.

More importantly in this demo, the automatic mode follows a few heuristics to keep the demanded target in the virtual view yet provide a smooth video [1]. The zoom for the on-site demo is set to be fixed because of the unknown 3D structure of the area, whereas in the soccer field the zoom variable also changes depending on the ball position in the field.

In addition to the virtual view, a preview window is presented where the portion of panorama that is being fetched, corrected for a perspective view and displayed is highlighted. This preview window proves to be a rather useful feature when developing the servoing algorithms. In addition, it also demonstrates that the virtual view is not a simple crop from the panorama video.

All the processing except for downloading and decoding the video frame happens on a GPU to utilize the parallel processing power. The client has been tested on both Mac and PC with different capabilities. Nevertheless the average processing time is approximately 12 ms per frame on commodity graphics hardware.

3. DEMONSTRATION

In this demo, we present a system for real-time interactive zooming and panning of panorama video. The general setup is shown in figure 1. We use two types of panorama input. We use a stored panorama video captured in the Alfheim soccer stadium, i.e., the European League game between Tromsø IL and Tottenham Hotspurs [6]. The video was recorded using five 2K industrial cameras, and processed and stitched in a distributed system. We also use a live system using two 2K cameras on site to generate the panorama video. Both the stored-video and the live-video demos work in real-time where the system zooms and pans into tracked objects, i.e., the ball in the soccer game and people wearing colored hats in the live panorame video[1].

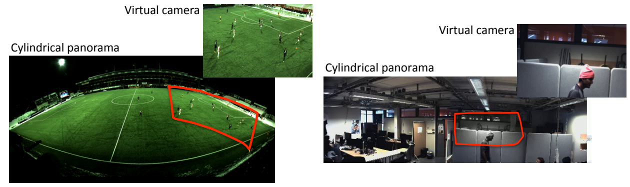

Figure 2: The virtual camera is generated from the region of interest marked in the panorama video. Note that it is not a simple crop from the bigger video.

Acknowledgments

This work has been performed in the context of the *iAD* center for Research-based Innovation (project number 174867) funded by the Norwegian Research Council.

4. REFERENCES

[1] Algorithms and user-studies submitted to acm mm 2014 as full paper, 2014.

[2] P. Carr and R. Hartley. Portable multi-megapixel camera with real-time recording and playback. In *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, pages 74–80, 2009.

[3] P. Carr, M. Mistry, and I. Matthews. Hybrid robotic/virtual pan-tilt-zom cameras for autonomous event recording. In *Proc. of ACM MM*, pages 193–202, 2013.

[4] V. R. Gaddam, R. Langseth, S. Ljødal, P. Gurdjos, V. Charvillat, C. Griwodz, and P. Halvorsen. Interactive zoom and panning from live panoramic video. In *Proc. of ACM NOSSDAV*, Mar. 2014.

[5] P. Halvorsen, S. Sægrov, A. Mortensen, D. K. Kristensen, A. Eichhorn, M. Stenhaug, S. Dahl, H. K. Stensland, V. R. Gaddam, C. Griwodz, and D. Johansen. Bagadus: An integrated system for arena sports analytics – a soccer case study. In *Proc. of ACM MMSys*, pages 48–59, Mar. 2013.

[6] S. A. Pettersen, D. Johansen, H. Johansen, V. Berg-Johansen, V. R. Gaddam, A. Mortensen, R. Langseth, C. Griwodz, H. K. Stensland, and P. Halvorsen. Soccer video and player position dataset. In *Proc. of ACM MMSYS*, Mar. 2014.

[7] O. Schreer, I. Feldmann, C. Weissig, P. Kauff, and R. Schafer. Ultrahigh-resolution panoramic imaging for format-agnostic video production. *Proceedings of the IEEE*, 101(1):99–114, Jan 2013.

[1] http://heim.ifi.uio.no/ vamsidhg/acmdemo.mp4