

MANAGING A CAMERA SYSTEM TO SERVE DIFFERENT VIDEO REQUESTS

Qiong Liu, Don Kimber, Lynn Wilcox, Matthew Cooper, Jonathan Foote, John Boreczky

{liu, kimber, wilcox, cooper, foote, johnb }@fxpal.com

FX Palo Alto Laboratory, 3400 Hillview Ave., Palo Alto, CA94304

ABSTRACT

This paper presents a camera system called FlySPEC. In contrast to a traditional camera system that provides the same video stream to every user, FlySPEC can simultaneously serve different video-viewing requests. This flexibility allows users to conveniently participate in a seminar or meeting at their own pace. Meanwhile, the FlySPEC system provides a seamless blend of manual control and automation. With this control mix, users can easily make tradeoffs between video capture effort and video quality. The FlySPEC camera is constructed by installing a set of Pan/Tilt/Zoom (PTZ) cameras near a high-resolution panoramic camera. While the panoramic camera provides the basic functionality of serving different viewing requests, the PTZ camera is managed by our algorithm to improve the overall video quality that may affect users watching details. The video resolution improvements from using different camera management strategies are compared in the experimental section.

1. INTRODUCTION

Given the sharp drop in camera prices and easy access to the Internet, people increasingly prefer watching video remotely under various circumstances. Videoconferences and distance learning are good examples of remote video access. During a video broadcast of a class or meeting, people sometimes prefer a view of their choice over one selected by someone else.

Because of physical limitations, a traditional pan/tilt/zoom (PTZ) camera system cannot allow multiple remote users to point the camera to different positions at the same time. To serve different viewing requests, a straightforward approach is to capture the event with a panoramic camera that covers every possible view and serve different viewing requests through electronic pan/tilt/zoom. However, a panoramic camera generally lacks the required resolution. FlySPEC is designed to balance the multiple-view requirement and the resolution requirement with a hybrid camera construction.

The rest of this paper is organized in the following way. Section 2 of this paper provides a brief review of related research on camera system management. Section

3 illustrates the design of the FlySPEC system including the camera hardware, the graphical user interface for remote users, and the algorithm for camera resource management. In section 4, we move on to the experimental method and results. The concluding remarks and future work are given in Section 5.

2. RELATED RESEARCH WORK ON CAMERA MANAGEMENT

Existing camera systems can be separated into two categories: systems that always provide the same video stream to all users, and systems that can provide different video streams to different users. In the first category, we note the traditional single-operator-controlled camera system and the work from [1][3][4][5]. These systems give full control to the operator who produces the same video stream for all remote users without considering users' individual needs.

The panoramic video system described in [2] is a system in the second category in which systems can provide different video streams to different users. This video system allows remote viewers to acquire their own views by cropping and scaling regions in the panoramic view. Furthermore, it does not require additional computation and bandwidth if remote users are interested in the same view. However, the work presented in [2] is limited by the resolution when some users are interested in small regions in the panoramic view. The FlySPEC system is designed to tackle this problem.

3. THE FLYSPEC SYSTEM



Figure 1. The FlySPEC Camera

FlySPEC is the name of our SPot Enhanced Camera system. It is constructed with a set of PTZ cameras and a panoramic camera. This system allows a user to watch a resolution-reduced context (panoramic) video and a

customized close-up video at the same time. The panoramic video is the same for all users. The close-up video for a user is selected by marking a region in the user's panoramic view using a simple gesture.

3.1. The FlySPEC Camera

Figure 1 shows a picture of a FlySPEC camera. In this picture, we see that the FlySPEC camera is a hybrid camera constructed with a PTZ camera and a panoramic camera. The panoramic camera provides an overview to remote users, and can also be used for electronic pan/tilt/zoom. That solves the problem that a PTZ camera may encounter when multiple users want to watch video shots in different directions. On the other hand, the PTZ camera can compensate for the limited panoramic camera resolution. The close proximity of the panoramic and PTZ cameras makes it easy to determine where to point the PTZ camera for a close-up view of any given region in the panoramic video. In summary, this system is designed to balance operational flexibility and video frame resolution.

3.2. The FlySPEC System Structure

Figure 2 shows the system structure of our FlySPEC system. The FlySPEC camera captures videos with a PTZ camera and a panoramic camera. The NTSC video from the PTZ camera is connected to an off-the-shelf video server that can translate analog video shots into digital video streams for the Internet. It can serve a remote user through the Internet at a maximum speed of 30 frames per second. When multiple users access the server, the serving speed may slow down gracefully. To serve a large number of video users, a high performance network server should be used. Since the FlySPEC prototype system does not deal with a large number of users at its early stage, connecting the video server directly to the Internet provides us reasonable system performance.

Besides connecting to the video server through a NTSC connection, the PTZ camera is also connected to a networked workstation through a RS-232 serial link. The workstation runs a camera control server. The camera control server collects users' requests through the Internet and sends pan/tilt/zoom commands to the PTZ camera. When no user requests customized video, the control server operates the PTZ camera based on commands from automatic control units.

The panoramic camera is connected to a workstation that can capture video from fixed cameras and stitch these video inputs into panoramic video in real-time. A video server on this workstation sends reduced-resolution panoramic video to remote users, and sends close-up videos according to users' requests. On the right side of Figure 2, remote automatic control systems and human clients are connected to the FlySPEC system through the

Internet. When no human user controls the FlySPEC system, the automatic control units will control the system as default users. Remote users can watch FlySPEC videos controlled by automatic units, or can control the PTZ camera from their own computers.

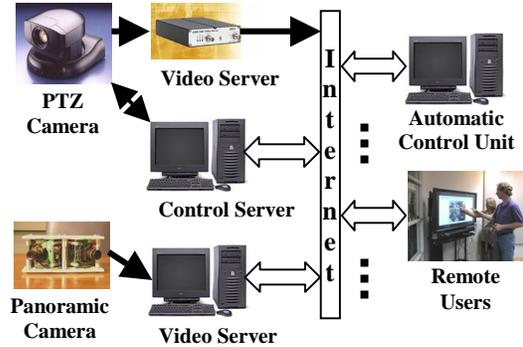


Figure 2. The FlySPEC System

3.3. The Graphical User Interface (GUI)

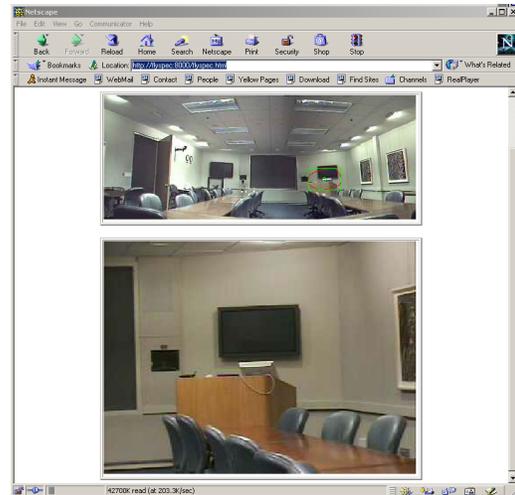


Figure 3. Web-based Graphical User Interface for Remote Users

Figure 3 shows the web-based graphical user interface (GUI) for remote users. In the web browser window, the upper window shows a resolution-reduced video from the panoramic camera, and the lower window shows the customized close-up video. By selecting a region in the panoramic window with a simple mouse-based gesture (i.e. drag the mouse through the region in the upper window when the left mouse button is pressed), a PTZ camera or a virtual camera will shoot a close-up view of that region, and show that close-up view in the lower window. If a touch screen is available, the region selection can be performed with a pen or a finger.

3.4. Managing the FlySPEC Camera for Simultaneous Multiple Accesses

A FlySPEC system has only a limited number of PTZ cameras. If multiple users compete for the PTZ camera resources and a user cannot get the control of a PTZ camera, the user will not be able to see his/her desired view with good resolution. This is a problem for remote applications like classes, seminars, or sports games. To tackle the problem caused by multiple users, an algorithm can be designed for efficient FlySPEC resource management based on optimizing a management cost function. Many choices exist for the design of the cost function. Our FlySPEC system currently uses the overall electronic zoom-in factor as the cost function. More specifically, our management software tries to move PTZ cameras to positions that can minimize the overall electronic zoom-in factor, as defined below. Through reducing the overall electronic zoom-in factor, the overall resolution of video streams can be improved.

In the FlySPEC system, users' video selections are enclosed in bounding boxes. Since the aspect ratio of a selection is generally different from the aspect ratio of the user's close-up view window, to ensure that the user can see the entire selection, it may be necessary to adjust the bounding box of each selection to the proper aspect ratio. Let i be the index of a selection, w_i and h_i be the width and height of selection i respectively. Assume 640 by 480 is the size of the close-up view window. Then the adjusted width, W_i , of bounding box i can be described with eq. (1).

$$W_i = \max(w_i, \frac{640}{480} \cdot h_i), \quad (1)$$

In the FlySPEC prototype, we only have one PTZ camera in the system. Let A be the set of all requests served by the PTZ camera, P be the set of all requests served by the panoramic camera, n be the total number of requests, W_A be the width of the bounding box of all requests in A (measured by pixels in the panoramic image), W_P be the width of the panoramic view, R_{wi} be the resolution (window width in pixels) of user- i 's viewing window on his/her computer screen, R_A be the resolution of the PTZ camera, R_P be the resolution of the panoramic camera. Then we can get the electronic zoom-in factor $Z_{A,i}$ of request i with eq. (2) provided that request i is served by the PTZ camera.

$$Z_{A,i} = \frac{R_{wi}}{(W_i/W_A)R_A} = \frac{W_A \cdot R_{wi}}{W_i \cdot R_A} \quad (2)$$

Similarly, if request i is served by the panoramic camera, the electronic zoom-in factor $Z_{P,i}$ of request i can be calculated with eq. (3).

$$Z_{P,i} = \frac{R_{wi}}{(W_i/W_P) \cdot R_P} = \frac{W_P \cdot R_{wi}}{W_i \cdot R_P} \quad (3)$$

Eq. (2) and (3) are used in different situations depending on whether the PTZ camera or the panoramic camera is used to serve the request. If C_i is used to represent the cost of request i , our camera management idea can be defined with eq. (4),

$$\begin{aligned} C_i &= \max(Z_{A,i}^2, 1) \Leftrightarrow (i \in A) \\ C_i &= \max(Z_{P,i}^2, 1) \Leftrightarrow (i \in P) \cdot \\ A &= \arg \min_A \sum_{i=1}^n C_i \end{aligned} \quad (4)$$

The definition of C_i in eq. (4) helps the system to select the right formula for zoom-in factor calculation; it also helps the system to ignore the zoom-in factors when the FlySPEC can provide sufficient pixels to remote users. In general, remote users' screens have high enough resolutions to show all available pixels clearly. Assuming R_{wi} is large enough, $Z_{A,i}$ and $Z_{P,i}$ will always be larger than 1. Also considering $W_P=R_P$, we can get a simplified equation – eq.(5) – for camera management.

$$\begin{aligned} C_i &= \frac{W_A^2}{R_A^2 \cdot W_i^2} \Leftrightarrow i \in A \\ C_i &= \frac{1}{W_i^2} \Leftrightarrow i \in P \\ A &= \arg \min_A \sum_{i=1}^n C_i \end{aligned} \quad (5)$$

In eq. (5), W_A is the width of the bounding box of all requests in A . It is measured by pixels in the panoramic image. When no high zoom-in request exists (i.e. every request has a bounding box larger than R_A), eq. (5) suggests the system to serve all requests with the panoramic camera. When there are high zoom-in requests, the system will suggest a request combination that has a W_A smaller than R_A . If W_A is smaller than R_A , this equation will suggest the system to serve high zoom-in requests (i.e. small W_i requests) with the PTZ camera, and serve all other requests with the panoramic camera. It aligns well with our intuition. Eq. (5) gives the optimization cost function for a FlySPEC that has one PTZ camera. Following similar procedures, it is not difficult to derive the cost function for a FlySPEC that has multiple PTZ cameras.

With this optimization cost function, we can perform exhaustive search for the best set A . The computation of the exhaustive search is expensive when the number of requests is large. To reduce the computational complexity, we can upper-bound the number of requests used in the optimization process with heuristics. For example, if the number of requests passes 10 during a

specified interval, the system can pick the most recent 10 requests for the optimization process, and serve all requests based on this optimization result. If the number of requests is smaller than 10, the system can use all requests in the optimization process. The upper-bound of 10 is used the following experiment, where a 1-GHz PC can finish the exhaustive search within 1 second.

4. EXPERIMENTAL METHOD AND RESULTS

To test the system’s effectiveness on electronic zoom-in factor reduction, we deployed the system in a corporate conference room and collected image shots, corresponding user requests, and corresponding PTZ camera positions during 20 presentations. Figure 4 shows a typical shot from the experiment. In this figure, users’ selections are represented by bounding boxes of their gestures. These bounding boxes are painted with green color. The PTZ camera position is marked with a thick red box. Assume the panoramic camera has 1500x492 resolution and 100° field-of-view, the PTZ camera has 640x480 resolution and 48.8° maximum field-of-view, and every user can view their requests in a 640x480 window. If we calculate the average electronic zoom-in factor with eq. (6),

$$Z_{avg} = \frac{1}{n} \sum_{i=1}^n C_i, \quad (6)$$

the average electronic zoom-in factor of this shot is 28.038 without the PTZ camera. This zoom-in factor indicates that the system must generate 28 times as many pixels for the user image as are actually available from the camera. If the system uses the PTZ camera to serve the most demanding zoom-in request, the average electronic zoom-in factor is 21.748. Through managing the PTZ camera with our algorithm, Z_{avg} of this shot reduces to 4.697.

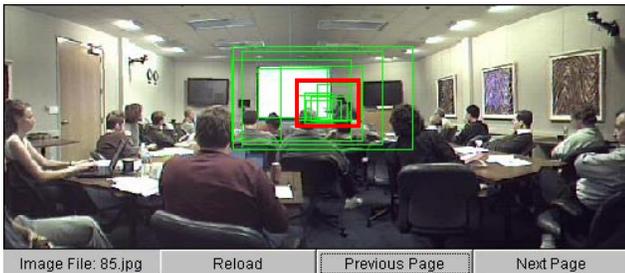


Figure 4. Users’ Video Requests and the PTZ Camera Position

We processed all collected data with the same procedure. The results are listed in Table 1. Results in Table 1 show great electronic-zoom-in-factor reduction achieved by our camera-management algorithm. If we install two or more PTZ cameras in the system, and modify our algorithm for

multiple PTZ-camera management, the overall video frame resolution can be further improved.

Table 1. Statistical Results of Electronic Zoom-in Factors Under Different Situations

Statistical Results of Zoom-in Factors	Without Using PTZ Camera	Using PTZ Camera for Highest-Zoom-in Request	Managing the PTZ Camera with Our Algorithm
Mean	14.786	11.342	3.873
Std.	4.492	3.378	1.513
Median	14.246	10.068	3.747

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a hybrid camera system, called FlySPEC, that can customize views for remote users. To ensure good display resolution for each video stream, a camera management algorithm is designed to optimize the use of the camera system by minimizing the overall electronic zoom-in factor. The system and the algorithm were tested in our organization’s conference room. Experimental results strongly support our camera management idea. This project can be extended in many aspects. For example, the overall electronic zoom-in factors can be further reduced through caching high-resolution images of fixed objects, and generating some requested videos through “image mosaics” based on cached images. Cost functions other than the overall electronic zoom-in factor could also be used in the optimization. Another interesting topic in this project is to teach an automatic control system to improve itself gradually based on manual operations.

6. REFERENCES

- [1] M. Bianchi, “AutoAuditorium: a fully automatic, multi-camera system to televise auditorium presentations,” *Proc. of Joint DARPA/NIST Smart Spaces Technology Workshop*, July 1998.
- [2] J. Foote and D. Kimber, “FlyCam: Practical Panoramic Video,” *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. III, pp. 1419-1422, 2000.
- [3] Q. Liu, Y. Rui, A. Gupta, J. Cadiz. "Automating Camera Management in a Lecture Room", *Proceedings of ACM CHI2001*, vol. 3, pp. 442 – 449, Seattle, Washington, USA, March 31 - April 5, 2001.
- [4] S. Mukhopadhyay, and B. Smith, “Passive Capture and Structuring of Lectures,” *Proc. of ACM Multimedia’99*, Orlando.
- [5] C. Wang, M. Brandstein, “A hybrid real-time face tracking system,” *Proc. of ICASSP98*, pp. 3737-3740, Seattle, May 1998.