Pervasive Cameras: Making Sense of Many Angles using Radial Basis Function Interpolation and Salience Analysis

Yotam Mann, Adrian Freed•

CNMAT
Center for New Music and Audio Technologies
UC Berkeley Dept. of Music
1750 Arch Street, Berkeley, California

yotammann@berkeley.edu, adrian@adrianfreed.com

Abstract. This paper describes situations which lend themselves to the use of numerous cameras and techniques for displaying many camera streams simultaneously. The focus here is on the role of the "director" in sorting out the onscreen content in an artistic and understandable manner. We describe techniques which we have employed in a system designed for capturing and displaying many video feeds incorporating automated and high level control of the composition on the screen using radial basis function interpolation.

1 A Hypothetical Cooking Show

A typical television cooking show might include a number of cameras capturing the action and techniques on stage from a variety of angles. If the TV chef were making salsa for example, cameras and crew might be setup by the sink to get a vegetable washing scene, over the cutting board to capture the cutting technique, next to a mixing bowl to show the finished salsa, and a few varying wider angle shots which can focus on the chef's face while he or she explains the process or on the kitchen scene as a whole. The director of such a show would then choose the series of angles that best visually describe the process and technique of making salsa. Now imagine trying to make a live cooking show presentation of the same recipe by oneself. Instead of a team of camera people, the prospective chef could simply setup a bunch of high quality and inexpensive webcams in these key positions in one's own kitchen to capture your process and banter from many angles. Along with the role of star, one must also act as director and compose these many angles into a cohesive and understandable narrative for the live audience.

Here we describe techniques for displaying many simultaneous video streams using automated and high level control of the cameras' positions and frame sizes on the screen allowing a solo director to compose many angles quickly, easily, and artistically.

• Adrian Freed: Concordia University, SIP/GRAND Researcher

Pervasive 2011, June 12 – 15, San Francisco, CA, USA

2 Collage of Viewpoints

Cutting between multiple view points would possibly obscure one important aspect in favor of another. In the cooking show example, a picture in picture approach would allow the display of both a closeup of a tomato being chopped and the face of the chef describing his or her cutting technique. The multi angle approach is a powerful technique for understanding all of the action being captured in a scene as well as artistically compelling and greatly simplifies the work of the director of such a scene. The multi-camera setup lends itself to a collage of framed video streams on a single screen (fig. 1 shows an example of this multi-angle approach). We would like to avoid a security camera style display that would leave much of the screen filled with unimportant content, like a view of a sink not being used by anyone.

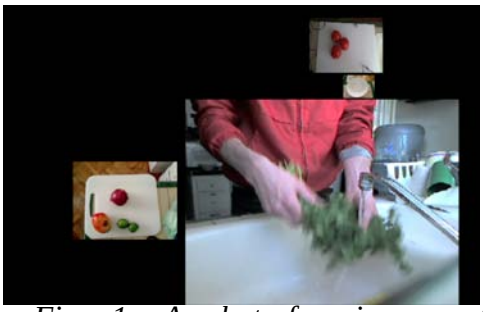

Fig. 1: A shot focusing on the cilantro being washed

We focus attention on a single frame by making it larger than the surrounding frames or moving it closer to the center. The system is designed to take in a number of scaling factors (user-defined and automated) which multiply together to determine the final frame size and position. The user-defined scaling factor is set with a fader on each of the video channels. This is useful for setting initial level and biasing one frame over another in the overall mix. Setting and adjusting separate channels can be time consuming for the user, so we give a higher level of control to the user that facilitates scaling many video streams at once smoothly and quickly. CNMAT's (Center for New Music and Audio Technologies) *rbfi* (radial basis function interpolation) object for Max/MSP/Jitter allows the user to easily switch between preset arrangements, and also explore the infinite gradients in between these defined presets using interpolation[1]. *rbfi* lets the user place data points (presets in this case) anywhere in a 2 dimensional space and explore that space with a cursor. The weight of each preset is a power function of the distance from the cursor. For example, one preset point might enlarge all of the video frames on particular positions in the rooms, say by the kitchen island, so as the chef walks over the kitchen island, all of the cameras near the island enlarge and all of the other cameras shrink. Another preset might bias cameras focused on fruits. With *rbfi,* the user can slider the cursor to whichever preset best fits the current situation. With this simple, yet powerful high level of control, a user is able to compose the scene quickly and artistically, even while chopping onions.

Fig. 2: A screenshot focusing on the mango being chopped and the mixing bowl

Fig. 3: A screenshot of the finished product and some cleanup

3 Automating Control

Another technique we employ is to automate the direction of the scene by analyzing the content of the individual frame and then resizing to maximize the area of the frame containing the most salient features. This approach makes for fluid and dynamic screen content which focuses on the action in the scene without any person needing to operate the controls. One such analysis measures the amount of motion which is quantified by taking a running average of the number of changed pixels between successive frames in one video stream. The most dynamic video stream would have the largest motion scaling factor while the others shrink from their relative lack of motion. Another analysis is detecting faces using the openCV library and then promoting video streams with faces in them. The multiplied combination of user-defined and automated weight adjustment determines each frames final size on the screen. Using these two automations with the cooking show example, if the chef looks up at the camera and starts chopping a tomato, the video streams that contain the chef's face and the tomato being chopped would be promoted to the largest frames in the scene while the other less important frames shrink to accommodate the two.

Positioning on the screen is also automated in this system. The frames are able to float anywhere on and off the screen while edge detection ensures that no frames overlap. The user sets the amount of a few different forces that are applied to the positions of the frames on the screen. One force propels all of the frame towards the

center of the screen. Another force pushes the frames to rearrang their relative positions on the screen. No single influence dictates the exact positioning or size of any video frame; this is only determined by the complex interaction of all of these scaling factors and forces.

4 Telematic Example

Aside from a hypothetical self-made cooking show, a tested application of these techniques is in a telematic concert situation. The extensive use of webcams on the stage works well in a colocation concert where the audience might be in a remote location from the performers. Many angles on one scene gives the audience more of a tele-immersive experience. Audiences can also experience fine details like a performer's playing subtly inside the piano or a bassist's intricate fretboard work without having to be at the location or seated far from the stage. The potential issue is sorting out all of these video streams without overwhelming the viewer with content. This can be achieved without a large crew of videographers at each site, but with a single director dynamically resizing and rearranging the frames based on feel or cues as well as analysis of the video stream's content.

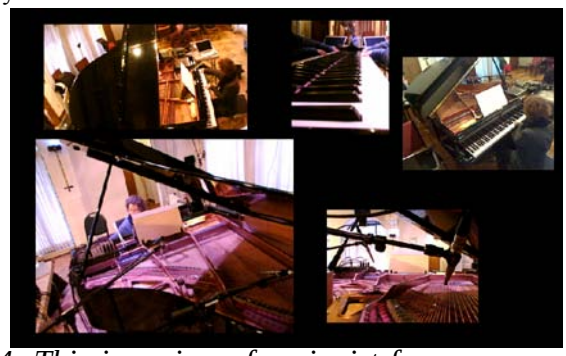

Fig. 4: This is a view of a pianist from many angles which would giving an audience a good understanding of the room and all of the player's techniques inside the piano and on the keyboard.

References

1. Freed, A., MacCallum, J., Schmeder, A., Wessel, D.: Visualizations and Interaction Strategies for Hybridization Interfaces. New Instruments for Musical Expression (2010)