FOVEAL DISPLAYS: ON INTEGRATING MULTI-RESOLUTION INPUT AND OUTPUT DEVICES

[1]Li-Wei Chan(詹力韋), [2]Wei-Shian Ye(葉韋賢), [2]Shou-Chun Liao(廖守鈞), [3]Yu-Pao Tsai(蔡玉寶), [1,2]Chieh-Chih Wang(王傑智), [1,2] Hao-hua Chu(朱浩華), [1,2]Jane Hsu(許永真), and [1,2]Yi-Ping Hung(洪一平)

[1] Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei, Taiwan
[2] Dept. of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
[3] Dept. of Computer and Information Science,
National Chiao Tung University, Hsinchu, Taiwan

ABSTRACT

Previous research on human visual perception system had conducted studies on the spatial variation in human visual resolution. Their studies had shown that only the fovea region of human eyes can afford sharp vision with acute visual details. In this paper, we propose an *interactive multi-resolution monitor* system composed of a multi-resolution input channel and output display. The multi-resolution input channel consists of a wide angle camera and a Pan/Tilt/Zoom (PTZ) camera. The multi-resolution output display consists of a fixed projector and a steerable projector. The fixed projector takes inputs from the wide-angle camera to provide an overview of a scene, while the steerable projector takes inputs from the PTZ camera to provide a high-resolution sub-view of the same scene. Our system combines multi-resolution input devices and output devices to create a *cost-effective*, *attentive* user interface that can help reduce human operator's cognitive loads during their monitoring tasks. The contributions are two folds: (1) enabling interactive capture and display of high-resolution region-of-interest for center of attention while providing a low-resolution background for peripheral awareness, and (2) balance the requirement between capture and display resolution and system cost.

1. INTRODUCTION

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them. Previous research [7] on human visual perception system had conducted studies on the spatial variation in human visual resolution. Their studies had shown that only the fovea region of human eyes can afford sharp vision with acute visual details. In comparison, peripheral region of human eyes perceives rough percipience of the world with coarse visual details. This spatial variation in human perception has been widely exploited for many eye-tracker based applications, such as level-of-detail rendering [14] and video compression [13], in reducing display resolution outside the fovea region. Therefore, these applications can lower the amount of computation and economize system cost. Research in building large, low-cost displays can also take advantage of this human perception characteristic. For example, the concept behind a multi-resolution wall-sized display is to (1) provide peripheral awareness over a large display region at a low resolution matching human peripheral visual acuity, and (2) put high resolution only at a small display region of interest matching his/her fovea visual acuity. [17]

In this paper, we propose an interactive multi-resolution monitor system composed of a multi-resolution input channel and output display. An overview of our system design is shown in Figure 2. The multi-resolution input channel consists of a wide angle camera and a Pan/Tilt/Zoom (PTZ) camera. The wide-angle camera provides a wide field of view at a lower resolution, whereas the PTZ camera provides a limited sub-view of the same scene at a high resolution. These two cameras are calibrated to simultaneously provide both a rough overview and a detailed sub-view from a scene.

The multi-resolution output display consists of a fixed projector and a steerable projector. These two projectors cooperatively project a multi-resolution display wall. The fixed projector takes inputs from the wide-angle camera to provide an overview of a scene, while the steerable projector takes inputs from the PTZ camera to provide a high-resolution sub-view of the same scene onto the display wall.

Our multi-resolution monitor system is interactive to accommodate movements of human eyes to different regions of interest. Interaction is provided in the following steps. (1) A user indicates his/her target region of interest by directly pointing on a display wall.

(2) This pointing action and its pointed spatial region are detected by a wide-angle camera. (3) A PTZ camera rotates and zooms to capture a high-resolution view of the target region of interest. (4) A steerable projector projects the high resolution view onto the display wall.

Our system combines multi-resolution input devices and output devices to create a cost-effective, attentive user interface that can help reduce human operator's cognitive loads during their monitoring tasks. In a typical monitoring task such as video surveillance, a human operator often needs to split his/her cognitive load between center of attentions (e.g., a person/object of interest) and peripheral awareness (e.g., a large background). Our system is designed to ease this cognitive load: high resolution foreground sufficient for center of attentions and low resolution background adequate for peripheral awareness.

Our contributions are two folds: (1) enabling interactive capture and display of high-resolution region-of-interest for center of attention while providing a low-resolution background for peripheral awareness, and (2) balance the requirement between capture and display resolution and system cost.

The remainder of this paper is organized as follows. Section 2 describes related work. Section 3 presents design and implementation of our multi-resolution monitor system. Section 4 discusses evaluation of our system through user studies. Section 5 draws conclusion and future work.

2. RELATED WORK

In the following, we first give a review on previous monitor systems, which are the target application of the proposed system, and then followed by related researches on the multi-resolution input and output systems.

2.1.Monitor System

Monitoring-type tasks usually use static or pan-tile-zoom camera to watch a scene. These camera views are tiling up or are presented in a split screen, which is hard for the viewers to preserve and organize the spatial relationships between the camera views. For example, it is common to see more than ten or hundred of camera views tiled in a surveillance center. Although several guarders concentrating on monitoring task are probably missing track of a moving target, especially when the target is crossing from one camera view into another.

Another surveillance system that integrates a wide angle camera with a PTZ camera is more flexible to use, because the guarder can control the PTZ camera by simply pointing on the wide angle camera view. The system is cost effective from which only two cameras are required to cover a large scene. However, the two camera views are usually presented separately on two display screens or on two disjoint regions of single screen. Figure 1 shows the type of surveillance system set in our laboratory. The system visualizes input streams in an overview+detail concept proposed in [8, 14]. Once a target shows up, the guarder has to switch between two screens, tracking in one screen and seeing detail in another screen. If the target keeps moving, the guarder is tend to lose the target during transition from the overall view to the detailed view and vice versa. This is due to when the users point on the target on the overall view; they assume the target will present on the center of the detailed view. Once lost, the users start a full search on the detailed view. Otherwise when transferring back from the detailed view to the overall view, they assume that the target remains appeared nearby last showup, or do full search.

Figure 1. A commercial surveillance system uses the overview+detail interface. (the left is a contextual overview of a scene, and the right is a detailed view)

L-3 GSI [1] presented an immersive video surveillance system that combines the 3D model of the monitored environment and a number of cameras settled in the environment. The camera poses and camera coordinates are calibrated to the 3D model. They demonstrate monitoring the environment by navigating in the virtual world. Camera views are presented as a textured boards standing in the virtual world. Their approach gives strong connections among the cameras and the world. However, building such a surveillance environment requires tedious cost and effort.

2.2.Multi-Resolution Output Display

Utilizing multi-resolution concept on displaying technique, S. Feiner and A. Shamash [4] tried to improve 2D window management by introducing hybrid user interface, which combines high/low resolution display and interaction technologies that permit the user to see and interact with both. Their implementation includes a head mounted display to present the low resolution surround and a flat-panel display as the high resolution area fixed in front of the user. The user's head is tracked to determine what part of the world to display, depending on where the user is looking. This makes the head mounted display to present the full window desktop, while the flat plane display acts as a window on a large surround within which the user and the high-resolution display reside.

P. Baudisch et al. [10, 12] proposed Focus Plus Context Screens, a multi-resolution display to simultaneously display focus and context information. Their work helps the user accesses overview information while working on details. Their multi-resolution display is built by seamlessly integrating a LCD in the center with a customized projection surface surrounding it.

The two woks provide prototypes of multi-resolution display by integrating different units as high/low resolution displays. However, these works make the high resolution area fixed in front of the user. In other words, the user is limited not to move his/her eyes and still can perceive high resolution display area.

Other researches [5, 8, 11, 15] related to multi-resolution display include software-based approaches that design zooming interface for navigation in multiscale documents /maps under a resolution limitation of the display screen. These approaches try to gain resolution on portions of a large document by introducing different zoom strategies for presenting levels of details of the document/map.

2.3.Multi-Resolution Input Channel

The multi-resolution input channel is commonly used in a surveillance system. As mentioned on the study of monitor system, the surveillance systems that integrate a wide angle camera and a PTZ camera have benefits both on its flexibility and the cost. However, these systems present two levels of inputs on separate output display screens or by one display screen divided by one boundary, so that the users have to switch between display regions. Since the two regions are for different meanings of inputs from cameras, the users switch not only waving their heads, but changing their metals.

Our multi-resolution system simultaneously presents both the overall and detailed views which are seamless form a large videoconference on the wall. Our system naturally connects meanings of two camera views, as well as strengthens the spatial relationship between them.

3. SYSTEM DESIGN AND IMPLEMENTATION

Figure 2 illustrates the design of the proposed interactive multi-resolution monitoring system, which consists of the two main components: one is a multi-resolution input channel composing of a wide-angle camera and a PTZ camera, and the other is a multi-resolution output display composing of a fixed projector and a steerable projector.

To provide a high quality integration of two components, the system addresses two problems. First, the spatial relationship between the wide-angle camera and the PTZ camera is required to stitch two video sources into a multi-resolution input. Second, the spatial relationship among the fixed projector, the steerable projector, and the display wall is required to provide a seamlessly large-scale, multi-resolution output display. These critical issues are described in more details in the following sections.

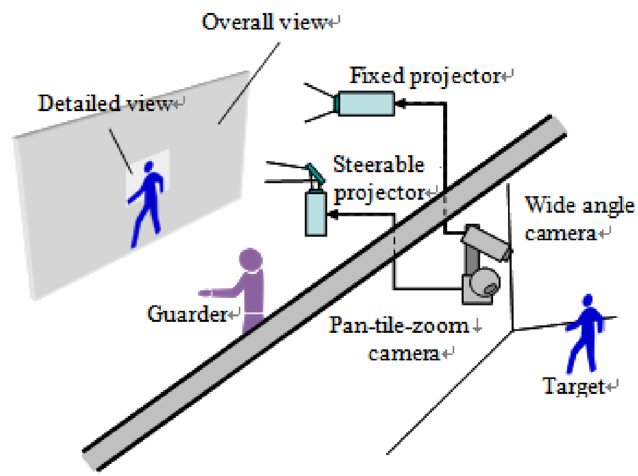

Figure 2. System design.

3.1.Transformation Calibration of Multi-Resolution Input

The multi-resolution input devices are shown in Figure 3. To seamlessly stitch two camera views, the transformation between them is needed. In practice, 2D homography between two cameras can be computed to accomplish this task if images are taken from the same center of projection. Unfortunately, the closely mounted wide-angle and PTZ cameras do not share single camera origin in which it is impossible to find a perfect homography. Note that the pose and zoom changes of the PTZ camera make this task much more difficult.

In this section, we proposed a novel and feasible approach. In the calibration phase, the cameras are placed to face a large planar wall for satisfying the homography assumption. Without losing the accuracy of the calibration, all possible pose and zoom changes of the PTZ camera are randomly sampled. Homographies between the sampled PTZ camera poses and the wide-angle camera are determined. To further improve the accuracy, the lens distortion of the wide-angle camera is removed with the use of the calibration tool based on OpenCV but with better corner detection released by V. Vladimir [2] for wide angle camera image. Regarding the camera calibration techniques, readers can refer to the paper [6]. All homographies between the sampled PTZ camera poses and the wide-angle camera are stored in a lookup table.

In the monitoring phase, once a dedicated point in the wide-angle camera is specified, the best pose of the PTZ camera is generated automatically and a homography, H_{wide}^{ptz}, that best covers the observed scene is fetched from the table in real time.

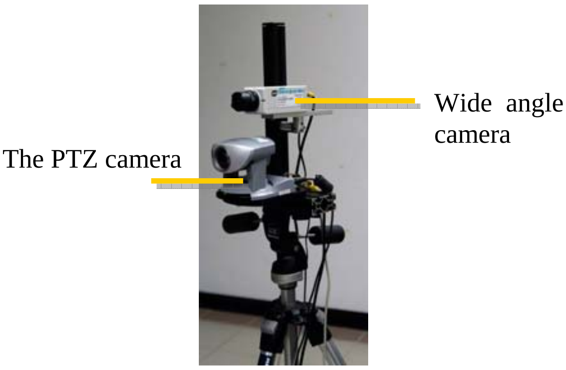

Figure 3. The multi-resolution input components.

3.2.Transformation Calibration of Multi-Resolution Output

The multi-resolution output devices are shown in Figure 4. Our output system currently consists of three stationary projectors which are to cooperatively provide a large display on a wall, and a steerable projector which is to provide high resolution display of a sub-region on the wall. As the three fixed projectors are properly calibrated with respect to the wall surface, these fixed projects can be treated as a fixed projector without violating the assumption of a fixed projector. In the rest of the paper, the fixed projector represents both a fixed projector and the multiple calibrated fixed projectors. More details about calibration of multiple projectors can be found in our previous work [16].

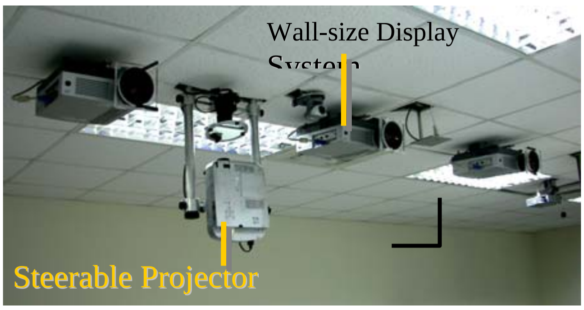

Figure 4. The multi-resolution output devices.

By determining the spatial transformation between the fixed projector and the wall surface, we can have large correct display coverage on the wall surface [16]. In order to overlay the projection of the steerable projector onto the display wall, the projected images from both the fixed projector and the steerable projector must be matched perfectly on the display wall in which the spatial relationship between the steerable projector and the fixed projector is critical. In other words, homographies between the steerable projector and the fixed projector must be estimated correctly. To automate this calibration process, another PTZ camera is used during the calibration phase. First, the fixed projector projects a grid pattern as shown in Figure 6 and then the camera captures this grid pattern for computing the homography between the PTZ camera and the fixed projector. Next, the steerable projector displays another grid pattern. The same process is carried out to find the homography between the PTZ camera and the steerable projector. Finally, the homography between the steerable projector and the fixed projector is determined by integrating these two homographies. For techniques about the projector calibration, reader can refer to [9].

3.3.Integration of Multi-Resolution Input and Output

Upon completion of the calibration of the input and output systems, we now describe the procedures to integrate these two components. First, two images, I_{wide}, I_{ptz}, from the wide-angle camera and the PTZ camera are captured. The I_{wide} is further calibrated to remove the lens distortion, notated as $I_{wide\text{-}dlens}$. Next, the homography H_{wide}^{ptz}, which is determined in calibration phase, is applied on I_{ptz}, transforming I_{ptz} into the coordinate system of the wide-angle camera, notated as $I_{ptz\text{-}wide}$. The two images, $I_{wide\text{-}dlens}$ and $I_{ptz\text{-}wide}$, are sent to the fixed projector and the steerable projector, respectively. Lastly, the fixed projector applied the homography $H_{surf}^{fixed^{-1}}$ on $I_{wide\text{-}dlens}$ to display an undistorted image on the wall surface. The steerable projector first multiplies $I_{ptz\text{-}wide}$ with H_{steer}^{fixed}, transforming the image to the coordinate system of the steerable projector's frame buffer. The transformed image is then multiplied with $H_{surf}^{steer^{-1}}$, which removes the distortion of the steerable projector with respect to the wall surface. Figure 5 illustrates the integration process and Figure 6 presents the result after applying the integration process.

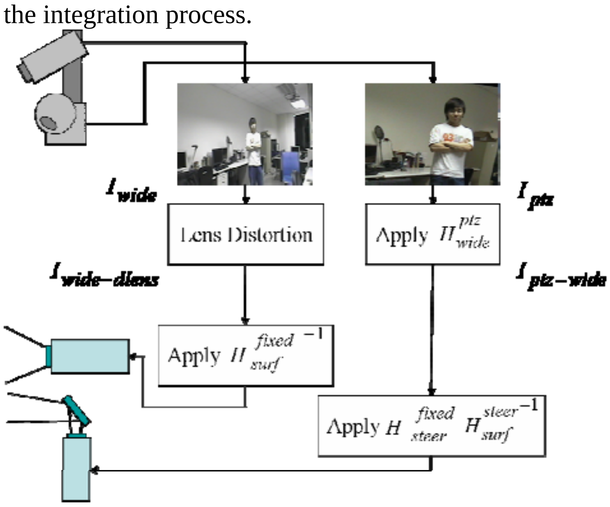

Figure 5. This flow chart illustrates the integration process of the multi-resolution input and output devices.

Figure 6. A multi-resolution presentation of multi-resolution input projected on the display wall.

To reduce the computational power while minimizing the loss of image quality during transforming, we concatenate homographies into one transformation for two transforming flow. Therefore, only one re-sampling operation is applied to the image sources before projected.

Note that the users are allowed to indicate an interesting target on the display wall by using a mouse cursor in our current implementation. A more nature way to interact with the system, a direct pointing using laser pointer, is in progress.

4. APPLICATION DEMONSTRATION – CARE WALL

We have applied our multi-resolution monitor system to a healthcare application called Care-wall. The motivation for the Care-wall is to enable remote family members to closely monitor and attend to physical and mental well-beings of their elders. The goal of the Care-wall is to support face-to-face like care between caregivers and their remote care-receivers. Care-wall hopes to create a digital illusion that two physically-disjoint spaces are intercon-nected together. In the following, we demonstrate a sample scenario of utilizing the Care-wall.

A grandfather and a grandson are living in two physically-disjoint spaces. One day is the grandson's birthday. A birthday party is held in the living room of the grandson's home. Unfortunately, grandpa cannot physically attend this birthday party. When the party starts, the grandpa joins this birthday party in front of the Care-wall, where the wall is virtually connected to the grandson's remote living room. Through the Care-wall, the grandpa can enjoy the birthday party in which children are happily playing around. The grandpa sequentially points to each child, and the Care-wall interactively brings an enhancement in video resolution on a specified target. The grandpa soon finds the grandson and sees a big smile on his face. Figure 7 shows an illusion of this sample scenario.

Figure 7. This figure illustrates the scenario of Care Wall.

5. USER STUDY (DISCUSSION)

In the performance study, we compared our system with a commercial surveillance system having multi-resolution input from a wide angle camera and a PTZ camera as well. The reference system follows the overview+detail interface, henceforth called "o+d display", to present two input steams onto two separate LCD screens placed side by side. The left LCD screen displays the overview, and the right screen displays the detailed view. Each LCD screen is 15-inch in size and has a resolution of 1,024 x 768. The subjects click on any specific region on the overview in the left display to obtain the corresponding detailed view of the selected region in the right display. In both systems, the subjects use mouse to issue pointing operation.

5.1.Test Design

We designed two tests that simulate respectively two common events during a monitoring task. The first required each subject to read the word carried by a moving target at various speeds. The second test required each subject to perform a similar task. However, multiple targets are allowed to appear during a test. The two tests are aimed at a quantitative evaluation of the two systems in performing monitoring tasks.

5.1.1.Evaluation on single moving target

One of the tasks that commonly happened in real world monitoring is tracking a moving target. For example, in a transportation station, a security guard monitors the station in the surveillance center. Whenever a robbery takes place, the guard must keep track of the robber and try to take a clear shot of the robber before his/her escaping the station.

To simulate the event, the evaluation program generates a moving circle as a target per task. Each circle carries a word inside. The subject is instructed to catch up with the moving circle and read the word out aloud. Due to the limitation of the displays' resolution, the word is only legible in high resolution. In other words, in our system, the word is only readable within the high resolution region of the multi-resolution display wall, while in the o+d display, it is only readable on the detailed view screen.

In this test, we specify three levels of speed for the moving target, which is defined by the number of pixels a moving circle could traverse within a second. Speed levels are set at 300, 400 and 500 pixels respectively. For each speed level, we required the subject to finish ten tasks. The subjects were instructed to finish the tasks as quickly as possible. Subjects were given written instructions describing the test procedure and the basic operations of the two interfaces. Before starting a test, a subject is given a practice round of 3 tasks. The practice time is not included in the evaluation result.

5.1.2.Evaluation on concurrent moving targets

The second test is to handle multiple targets entering the monitoring scene concurrently. This test is intended to simulate the following situation. When the security guard focuses on a current event, he might miss the other events that happen around the same time. The ability to concentrate locally while perceiving the global scene is important. For example, in a healthcare scenario, the caregiver may provide care to multiple remote healthcare receivers. Any lost of incoming information could result in a serious outcome.

In this test, we control the maximum number of tasks to exist concurrently as an evaluation parameter in a test round. As is different from the previous test, each target will disappear in a pre-defined time (e.g. 5 seconds in the current experiment). Subjects are instructed to read the words before the targets expire. At the start of any test, one new target enters the screen per second till the maximum number of tasks has reached. Each test is done at the maximum level of 2, 3 and 5 moving targets. For each level, a total of 20 targets will enter the screen. The performance is measured by calculating the *miss rate*, which is the percentage a subject failed to track the word. Before a real test, subjects were given written instructions describing the test procedure and went through a simple practice.

5.2.Subjects

The subjects are mainly recruited from members of various laboratories in the computer science department building. A total of five subjects participated in the user study. All are computer science majors, specializing in different research topics. None is in the field of user interface.

5.3.Results

Figure 9 plots the average time (in seconds) used with respect to different speed levels of a single moving target. The results show that the Fovea Display outperforms the o+d Display. On average, the Fovea Display saves 45.9% of time in the first speed level, 26.14% in the second level, and 17.4% in the fastest level. This is likely due to the fact that the o+d display shows the overall and detailed views in two separate screens. Subjects had to click in the overall view to capture a moving circle and then to decipher the text inside the corresponding circle in the detailed view. During the test process, a subject often had to turn his/her head in order to complete the tasks; the physical movement might result in a loss of target in the detailed view as the speed goes up. Besides, switching between two screens with different resolution might increase the cognition overhead of subjects. In comparison, the proposed Fovea Display integrates multi-resolution views within a single display, which should be more nature for subjects to catch an interesting target.

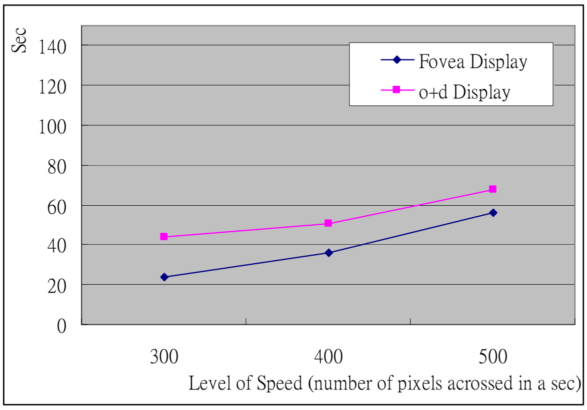

Figure 9. Average Time in the Single Moving Target Task

Figure 10 plots the target miss rate (in percentage) with respect to max number of tasks to appear on screen at the same time. Results show that the miss rate in the Fovea Display test rises slower than the o+d dispaly.

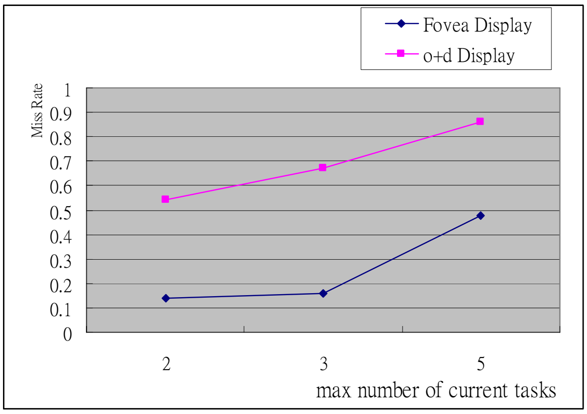

Figure 10. Averaged miss rate of tasks in the multiple moving target task.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed an interactive multi-resolution monitor system, which is composed of a multi-resolution input channel and output display, to provide an intuitive user interface for monitoring task. With the moving capability both on the input and output system, the proposed integration attends two achievements. First, the system provides the users to interactively catch an interesting target without loss of global information. Second, the system balances the requirement between resolution and cost of building a wall size, high resolution monitor system. Based on the system, we developed a care wall to demonstrate a sample use of the system in the healthcare domain. Other applications related to monitoring task are with high potential to profit from the interactive multi-resolution monitor system.

In the current implementation, we approximate the transformation between the wide angle camera and the PTZ camera by estimating homographies between them, which might produce artifacts. This is due to the two cameras not sharing the same camera origin and also the monitoring task of an environment is not simply a planar case. A nonlinear transformation might compensate the misalignment or a special mechanism that places two camera origins closer could provide possible solutions. Also, we will implement a laser pointer detector to provide direct pointing, so that the user can guide the high resolution region (involves moving the steerable projector and the PTZ camera) more intuitively.

Besides, the interactive multi-resolution monitor system is with potential for face-to-face like care between caregivers and their remote care-receivers. Introducing advanced interactive techniques which support direct or indirect interactions to the two physically disjoint spaces would be interesting.

7. ACKNOWLEDGMENTS

This work was supported in part by the grants of NSC 95-2752-E-002-007-PAE and by the Ministry of Economic Affairs, Taiwan, under Grant 95-EC-17-A-02-S1-032.

8. REFERENCES

[1] L-3 Communications Government Services, Inc. (L-3 GSI) – Immersive Video Surveillance: http://www.l-3gsi.com/index.cfm

[2] Graphics & Media Lab, Moscow State University: http://graphics.cs.msu.su/en/research/calibration/index.html

[3] Questionnaire for User Interaction Satisfaction (QUIS) : http://lap.umd.edu/quis/

[4] Feiner, S. and A. Shamash, "Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers" *In Proceedings of the ACM Symposium on User Interface Software and Technology* (U1ST '91), pp. 9-17, Hilton Head, SC, 1991.

[5] B. B. Bederson, and J. D. Hollan, "Pad++: A zooming graphical interface for exploring alternate interface physics," *In Proceedings of the ACM Symposium on User Interface Software and Technology* (U1ST '94), 17–26. ACM press.

[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334, 2000.

[7] Loschky, L. C., and McConkie, G. W., "User Performance With Gaze Contingent Multiresolutional Displays," *In Eye Tracking Research & Applications Symposium* (Palm Beach Gardens, FL, 2000), ACM, pp.97--103.

[8] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for Using Multiple Views in Information Visualization," *In Proceedings of AVI 2000*, pp. 110–119.

[9] R. Sulcthankar, R. Stockton, and M. Mullin, "Smarter presentations: Exploiting homography in camera-projector systems," *In Proceedings of International Conference on Computer Vision*, 2001.

[10] P. Baudisch, N. Good, and P. Stewart, "Focus Plus Context Screens: Combining Display Technology with Visualization Techniques," *In Proceedings of the A CM Symposium on User Interface Software and Technology* (U1ST '01), pp. 31-40, 2001.

[11] K. Hornbaek, and E. Frokjaer, "Reading of electronic documents: the usability of linear, fisheye, and overview+detail interfaces," *In Proceedings of CHI 2001*, 293–300.

[12] P. Baudisch, N. Good, V. Bellotti, and P. Schraedley, "Keeping Things in Context: A Comparative Evaluation of Focus Plus Context Screens, Overviews, and Zooming," *In Proceedings of CHI 2002*, Minneapolis, MN, April 2002, pp. 259-266.

[13] L. Itti, "Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention," *IEEE Transactions on Image Processing*, Vol. 13, No. 10. (Oct 2004), pp. 1304-1318.

[14] Javier E. Martínez, Ali Erol, George Bebis, Richard Boyle and Xander Twombly , "Rendering Optimizations Guided by Head-Pose Estimates and Their Uncertainty," *Advances in Visual Computing: First International Symposium, ISVC 2005*, Vol. 3804, Lake Tahoe, NV, USA, December 5-7, 2005.

[15] Forlines, C.; Shen, C., "DTLens: Multi-user Tabletop Spatial Data Exploration", *In Proceedings of the A CM Symposium on User Interface Software and Technology* (U1ST '05), October 2005.

[16] Yu-Pao Tsai, Yen-Nien Wu, Shou-Chun Liao, Yi-Ping Hung and Zen-Chung Shih, "Automatic Geometric and Photometric Calibration for Tiling Multiple Projectors with a Pan-Tilt-Zoom Camera," *IEEE International Conference on Multimedia & Expo* (ICME '06), July 9, 2006.

[17] Yu-Pao Tsai, Yian-Nien Wu, Yi-Ping Hung, "Generating a Multiresolution Display by Integrating Multiple Projectors," *IEEE International Workshop on Projector-Camera Systems*, Nice, France, October 2003.