Obstruction Removal Using Feature Extraction Through Time for Videoconferencing Processing

Elliott Coleshill, Dr. Deborah Stacey
University of Guelph, Guelph, Ontario N1G 2W1

Dr. Alex Ferworn
Ryerson University, 350 Victoria St. Toronto, Ontario, M5B 2K3

Abstract:

A major problem with front projection displays used for video conferencing is the potential for undesirable shadows to be cast onto the display screen by presenters. This paper provides a new processing approach for removing cast shadows within videoconference applications. By using sequential images through time shadows can be detected and removed and the presentation contents reconstructed.

Key words:

Video Conferencing, Shadow Removal, Image Processing

1. INTRODUCTION

The increase in affordability and portability of high quality projectors and the general availability of high-speed Internet access has generated a surge of interest in videoconferencing systems research [1][2][3]. One of the most common first steps in many computer vision applications--like virtual videoconferencing--is the detection and removal of unwanted artifacts. The removals are often based on frame differences or background subtraction [4][5].

A challenge that arises from videoconferencing applications is shadow removal. Shadows are caused by presenters standing in close proximity to the projector between the light source and the screen. Often, presenters are unaware that they are casting shadows the may interfere with the ability of the audience to see the projected image being blocked. There are many ways to detect and remove shadows. These include using multiple projectors and lights [6][7], using simple changes of intensity and saturation within the pixel colour values and other techniques. In this paper we propose a new method of removing shadows using sequential images through time to reconstruct presentation information obscured by shadows.

2. PROPOSED SOLUTION

Using a sequence of video frames taken through time, shadows can be detected and removed leaving the content of the presentation material on the display without distortion.

By using Mean Value Mapping, a tone/contrast map can be generated for each input video frame. A shadow caused by an obstruction usually creates a lower tone and contrast compared to the rest of the display. For each of the input frames all low tone and contrast areas are detected and removed. The final optimized image is then reconstructed using the remaining “good” information from each frame. We call this approach “Feature Extraction Through Time (FETT).

T. Sobh et al. (eds.), Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications, 131–133.

The use of FETT for removing obstruction shadows can be divided into the following steps:

1. Acquire Images

A series of frames are taken over time. Our approach relies on the motion of the presenter. During the time period the presenter's shadow will change due to this movement.

2. Perform Mean Value Tone Mapping

A grid is placed over each input frame where area [1,1] of frame-1 lines up with area [1,1] of frame-2 and so on. Mean mapping values are calculated for each image and are compared.

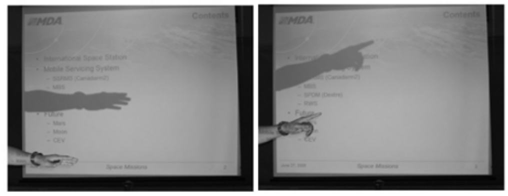
3. Generate New Image

The selected images and regions are then extracted from the original images and pieced together to generate a new image with removed.

3. RESULTS

A dataset of images has been generated with examples of how a presenter might cause occluding shadows on a display screen. Using the FETT method we have demonstrated how these shadows can be removed and replaced with the original content of the screen.

The first sets of tests were performed using the presenter's arm and hand as the obstruction causing the shadow. The test presenter moved their arm up and down over the screen as if they were pointing out details. Figure 1 (a) is a set of two frames taken with arm motion. It can be seen that every area of the screen is represented correctly within these two images. Using FETT Figure 1 (b) is generated with the arm/hand shadows removed.



(a)

(b)

Figure #1: Hand/Arm Shadows

A second set of tests was performed using more complex patterns displayed on the screen as well as the presenter being visible within the field of view. If has been found that if the presenter is wearing dark clothing--causing them to show up as low tone and contrast areas within the image--FETT can filter them out as well. Figure 2 (a) is an example of two frames taken where the presenter is causing a shadow in one, and is blocking the screen with their body in the second. Using FETT a new image, Figure 2 (b) is generated removing the shadow and presenter and reconstructing the display and its content.

(a)

(b)

Figure#2: Presenter/Body Shadow Example

4. APPLICABILITY

FETT as described in this paper is designed to work using a sequence of images from a single camera view with no changes to the scene other than the presenter and his/her shadow. This algorithm maps the presentation display to a specific area of the scene and reconstructs that area based on the same area of another input image containing no shadows. If the camera is moved the reconstruction of the display will not be reconstructed correctly.

Future enhancements to this approach include the use of multiple cameras. Generating a sequence of images from different cameras located around the classroom to reconstruct the video display on the screen.

5. CONCLUSION

This paper proposes a new method for using a sequence of frames through time to extract occlusion shadows created by presenters for video conferencing applications. With the use of Mean Value Mapping techniques, shadows can be detected and removed and the presentation reconstructed. Using the remaining "good" pixel information, an optimized image can be generated with no shadows blocking the viewable presentation information.

6. REFERENCES

[1] M. Draoli, G. Gambosi, and M. Lancia. "Videoconferencing on a LAN/MAN architecture: service evaluation and system dimensioning". IEEE Proceedings of ICCT'96, May 5-7, 1996. Volume 2 pp. 630-633.

[2] K. Liao, and J. Roberts. "Videoconference Traffic and Network Design". IEEE Transactions on Communications. March 1987. Volume 35, Issue 3, pp. 275-282.

[3] T. Turletti, and C. Huitema. "Videoconferencing on the Internet". IEEE/ACM Transactions on Networking (TON), June 1996. Volume 4, Issue 3. pp. 340-351.

[4] A. Limton, H. Fujiyoshi, and R. Patil. "Moving Target Classification and Tracking from Real-Time Video". Proceedings of WACV'98, pp. 8-14, 1998.

[5] A. Elgammal, D. Harwood, and L.S. Davis, "Non-parametric Model of Background Subtraction". Proceedings of ICCV'99 FRAME-RATE Workshop, 1999.

[6] T. Cham, J. Rehg, R. Sukthankar, and G. Sukthankar. "Shadow Elimination and Occluder Light Suppression for Multi-Projector Displays". Proceedings of the CVPR'03. March 2003.

[7] R. Sukthanker, T. Cham, and G. Sukthankar. "Dynamic Shadow Elimination for Multi-Projector Displays". Proceedings of CVPR'01. March 2001.

[8] O. Schreer, I. Feldmann, and P. Kauff. "Fast and Robust Shadow Detection in Videoconference Applications". VIPromCom-2002. IEEE Symposium on Video/Image Processing. June 16-19, Zadar, Croatia. Pp. 371-375.