Imperceptible Projection Blanking for Reliable Segmentation within Mixed Reality Applications

M. Kuechler and A. Kunz

Innovation Center Virtual Reality (ICVR)
Swiss Federal Institute of Technology ETH Zurich
Zurich, Switzerland
{kuechler, kunz}@inspire.ethz.ch

Abstract

Applications that run in a projection-based mixed reality environment frequently require simultaneous image acquisition of the user by a camera. For later processing, very often a segmentation of the user’s image is required, but along with the user, the camera captures the projected dynamic image in the background of the scene. This dynamic background complicates the later segmentation of the user in real-time. To ease segmentation, we introduce methods allowing to imperceptible blank the projected image for the camera. Four different methods are tested and compared. Acquired video streams are recorded and a successful segmentation of a user in front of a dynamic background is demonstrated. The paper concludes with an outlook on possible applications.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.5.1 [Multimedia Information Systems]: Virtual Realities

Additional Keywords: Imperceptible projection blanking, segmentation, mixed and augmented reality, gestural interfaces, networked collaboration

1 Introduction

In the last few years, large vertical interactive screens have become increasingly popular. Image display is mostly done by using a projector. Several techniques have been developed in order to make the screen interactive, including vision-based methods.

One vision-based interaction technique is to use a camera that acquires images of the hand gestures and upper body of the user standing in front of the vertical screen, while an image is projected onto the screen. A background subtraction algorithm segments the user from the dynamic background. Additional software then looks for known hand gestures and triggers events, e.g. drag and drop of displayed objects. For static or quasi-static projections like presentations, today’s background subtraction algorithms work fine. However, if dealing with dynamic backgrounds, like projected virtual realities in game worlds, background subtraction algorithms fail to correctly segment the user from the dynamic background. Thus, these algorithms should be supported by an additional projection blanking during image acquisition.

This paper describes and compares different shuttering methods that allow an imperceptible blanking the projected image to the camera acquiring the user who stands in front of the vertical screen. This leads to a static background in the acquired scene, independent of the projected content. In that way, segmentation of the user is successful even if the projected background is highly dynamic. The presented work is part of the project blue-c II, aiming to develop a modular hardware environment for local and remote collaboration.

2 Related work

A standard way to segment a dynamic foreground (e.g. a moving person) from a dynamic background is to acquire depth information about the different objects in the scene. This additional information then allows to correctly segment the image foreground from its background.

Conventional optical measurement systems capable of acquiring depth information are based on interferometry, triangulation, and structured light. Interferometry requires an additional device, e.g. a laser aperture for laser ranging. Triangulation needs at least two cameras and much computing power to achieve depth measurements at reasonable rates. Structured light is highly distracting and therefore not suitable for human-populated environments.

A new approach is depth measurement based on the time-of-flight principle with a camera, as presented by Lange and Seitz [1]. This technology is promising, but only...
monochrome imaging at 160x124 pixels has been done so far.

Cotting, Naef, et al. [2] presented front-projected invisible structured light to measure the depth of an acquired scene. This structured light is imperceptible to the human eye. However, for our envisioned application field, in which the user is standing in front of a screen, a rear-projection is more convenient. This means that in our setup a second projector would be required to illuminate the user in front of the screen with invisible structured light detectable by the camera.

3 Contribution

In this paper, we introduce and compare two well-known methods and two novel methods for blanking the projected image in order to ease segmentation within the scenario described in the above.

Figure 1 shows the principle setup of the system. It comprises a screen displaying a rear-projected image. Unless otherwise noted, the VGA signal from the computer graphics card is routed via a synchronization unit to the projector. A camera acquires the scene in front of the vertical screen, i.e. the hand gestures and upper body of the user standing in front of it. The camera’s image acquisition can be triggered via the synchronization unit.

Since the background’s dynamic part of the acquired scene consists only of the rear-projected image, it can be manipulated, namely blanked during the camera’s image acquisition. If the alternating image projection and blanking is done at a frequency above the threshold of perceiving and resolving flickering, the blanking is imperceptible to the user, and she or he is not distracted. For scotopic viewing, the frequency should be above 25 Hz, and for photopic viewing above 60 Hz. Thus, from the point of human perception, the blanking of the projection should occur only for a short time interval and at a high rate. However, from the technical point of view, this is limited by the sensitivity of the camera’s CCD, and the realizable speed of available shuttering technology.

We realized a system, which is capable to perform an imperceptible projection blanking for image acquisition. In the following, the individual components of the system are described.

4 Screen, synchronization unit, and camera

4.1 Screen

For the projection screen, we used a BlackScreen XRP3 from ScreenLab. It features a gain factor of 1.6 (DIN 67530), and horizontal and vertical half gain angles of 24 degrees. The screen preserves the polarization of the incident light. The screen surfaces are anti reflective (semi matte).

4.2 Synchronization unit

To synchronize the camera and the projector shutter for sequential image acquisition and display by time-multiplexing, we use a PIC16F877 microprocessor from Microchip Technology Inc. and DG2001 video switches from Vishay Intertechnology, Inc. The synchronization electronics is assembled on a standard PCIe card and mounted into an empty extension slot of the computer motherboard.

The synchronization unit generates various signals. Figure 2 shows a block diagram of the synchronization unit. Only the RGB signals are taken from the VGA input; the other signals, i.e. horizontal sync and vertical sync, are directly routed to the VGA output. Additionally, the vertical sync signal is also routed to the microprocessor.
The microprocessor generates the triggering signals for the camera, for the external shutter, and for the video switches. All timings of these signals are calculated with respect to the vertical sync signal. The synchronization electronic also generates a single-colored alternative video signal; the color can be set via three potentiometers. The triggering signal for the video switches is routed to the switches for the three video signals red (R), green (G), and blue (B). Depending on the triggering signal, the video switches route the original RGB signals or the alternative video signal to the VGA out jack. Switching of the RGB signals can be disabled by interrupting the triggering signal for the video switches. In this case, the original RGB signals are continually routed to the VGA out jack. The triggering signals for the camera and the external shutter are routed through another jack to the corresponding devices.

4.3 Camera

For our setup, we used an AVT Marlin F080C CCD firewire camera from ALLIED Vision Technologies. Images were taken with a resolution of 1024x768 pixels at rates varying from 15-18 Hz. Unless otherwise noted, the camera shutter was triggered by the synchronization unit. A trailing edge in the camera triggering signal releases the camera shutter. When sensing a trailing edge, our camera internally delays the acquisition by 8 μs, which can be neglected. According to the IIDC 1.31 standard, CCD integration time was set via the camera driver software (trigger mode 0).

5 Projectors and projector shuttering

We tested and compared four different methods for avoiding light from the projector onto the screen during image acquisition: (1) Periodic shuttering with an external shutter; (2) periodic shuttering by projecting black frames; (3) periodic shuttering by projecting black sub frames, and (4) projection blanking by using linear polarization filters. The different blanking methods make varying demands on the particular projector to be used for image display.

5.1 Periodic shuttering with an external shutter

Periodic shuttering of a projection by the use of an external shutter is a common technique used for active stereo projections, where two projectors alternately project an image for the left and the right eye, respectively. The user then wears actively triggered goggles, which also alternately shutter the left and the right eye. However, since we designed the system for a monoscopic projection only, no eyewear is required for the user. This is due to the fact that many industrial applications, e.g. in the early stages of product development, do not require stereoscopic viewing. However, our system can be easily adapted for stereoscopic viewing, if required.

For the external projector’s shutter, we used a custom-made ferroelectric (FE) liquid crystal shutter from CRL Opto. The FE shutter has a contrast ratio of approximately 750:1, and the switching time from 10 % to 90 % transparency and vice-versa is below 100 μs, as measured by the manufacturer. Opening and closing of the FE shutter was triggered by the synchronization unit.

For periodic shuttering with an external shutter, an LCD projector or a 3-chip DLP projector should be used. Since the three primary colors red, green and blue of single-chip DLP projectors are projected sequentially, shuttering with an external shutter would lead to color shifts of the projected image. For our setup, we used an XG-P25X three-panel LCD projector from Sharp. It has a native resolution of 1024x768 pixels, and a brightness of 4000 ANSI lumens.

Figure 3 shows the timings for periodic shuttering with an external shutter. Video switching of the synchronization unit was disabled. Signal R2 (bottom) represents the vertical sync signal of the computer graphics card. Since we use a vertical refresh rate of 60 Hz, the synchronizing rising edge of the signal occurs every 16.7 ms.

![Figure 3: Timings for periodic shuttering with an external shutter. R2: Vertical sync signal of the computer graphics card. R1: FE shutter triggering signal. 1: Relative brightness of the projected light. 2: Camera triggering signal.](image1)

![Figure 4: Timings for periodic shuttering by projecting black frames. R2: Vertical sync signal of the computer graphics card. 1: Relative brightness of the projected light. 2: Camera triggering signal.](image2)
Signal R1 is the FE shutter triggering signal. It is generated by the synchronization unit based on the vertical sync signal of the computer graphics card. If the triggering signal is low, the FE shutter is closed and therefore the projected light is hindered from striking the projection screen. In our setup, the projection is blocked for 5.6 ms at a rate of 60 Hz.

Signal 1 (top) represents the relative brightness of the projected light when projecting white (RGB 255/255/255). Since the FE shutter is closed for one-third of the time, the overall brightness of the projector is reduced by about one-third of its original intensity. However, the XG-P25X projector is bright enough (4000 ANSI lumens) that such a reduction in brightness can be tolerated. Note that the FE shutter does not fully block the projected image when closed. It reduces the amount of transmitted light only down to about 8% compared to the image brightness in its open state.

Signal 2 is the camera triggering signal. It is also generated by the synchronization unit. Since we acquire images with 15 Hz, only every fourth vertical sync signal is taken for generating the camera triggering signal. The synchronization unit is programmed in a way that the camera starts to acquire the images 0.4 ms after every fourth time the FE shutter starts to close, when there is no light being projected anymore. The CCD integration time for the camera is set to 5.2 ms.

5.2 Periodic shuttering by projecting black frames

For periodic shuttering by projecting black frames, we exchanged every second frame in the VGA signal with a black frame, before sending the signal to the projector. Video signal manipulation was done by enabling the synchronization unit’s video switches.

Note that with this method, both the optical refresh rate and the overall brightness of the projected image are halved. To avoid a flickering or dark projection, a bright projector with a high optical refresh rate must be used. For our setup we used a Mirage 5000 3-chip DLP projector from Christie. It has a native resolution of 1280x1024 pixels and a brightness of 5000 ANSI lumens. The device features a stereo mode, in which it runs frame synchronized on input signals with vertical refresh rates up to 108 Hz. In the stereo mode, images are projected with one frame delay, without any frame dropping. We used a 108 Hz input signal, which gave us a black frame lasting 9.3 ms at a rate of 54 Hz.

Figure 4 shows the timings for the projector shuttering by projecting black frames. Signal R2 (bottom) represents the vertical sync signal of the computer graphics card. The synchronizing rising edge of the signal occurs every 9.3 ms.

Signal 1 (top) represents the relative brightness of the projected light when projecting gray (RGB 220/220/220). The changing level of the measured light shows the generation of gray by mixing black and white over time. The information of every second frame is replaced by white segment for boosting additional light. If this white boost is disabled, a short black sub frame is generated. Its duration and rate depend on the speed of the color wheel. Within this phase, the camera acquires the image of the user. This would not be possible during a ‘normal’ white boost, since the camera’s dynamic range would be exceeded and thus the user would appear dark.

Figure 5 shows the sequential color projection of a single-chip DLP projector. Signal R2 (bottom) is the vertical sync signal of the computer graphics card.

Signal 1 (top) represents the relative brightness of the projected light when projecting white (RGB 255/255/255). By sequentially projecting the three primary colors red (R), green (G), and blue (B). The color is switched by a rotating color wheel, which is in the light path of the projector. Beside the RGB segments, most of today’s projectors’ color wheels also comprise a white segment for boosting additional light. If this white boost is disabled, a short black sub frame is generated. Its duration and rate depend on the speed of the color wheel. Within this phase, the camera acquires the image of the user. This would not be possible during a ‘normal’ white boost, since the camera’s dynamic range would be exceeded and thus the user would appear dark.

Periodic shuttering by projecting black sub frames requires a single-chip DLP projector, which provides a white boost feature that can be disabled. A bright projector is preferred, since the overall brightness drastically decreases by completely turning off the white boost. In order to acquire an image during the black sub frame, the projector’s optical output must be synchronized to the VGA input signal.

For our tests we used an F1+ SX+ single-chip DLP projector from Toshiba. It has a native resolution of 1400x1050 pixels and a brightness of 2500 ANSI lumens. The device runs frame synchronized on input signals with vertical refresh rates from 48-62 Hz. In that case, the images are displayed with a latency of one frame, and the projector’s color wheel has a speed of twice the vertical refresh rate. The color wheel comprises an 80 degree wide white segment for boosting additional light. The white
boost feature can be disabled via the menu settings of the projector.

Figure 6 shows the timings for projector shuttering by projecting black sub frames. Signal R2 (bottom) again represents the vertical sync signal of the computer graphics card. Since we use a vertical refresh rate of 50 Hz, the synchronizing rising edge of the signal occurs every 20 ms.

Signal 1 (top) represents the relative brightness of the projected light when projecting white (RGB 255/255/255). The speed of the projector’s color wheel is twice the vertical refresh rate, therefore every color is projected twice per frame. No light is projected between green (G) and blue (B) because we disabled the white boost feature. This results in a black sub frame (K), lasting for approximately 2.5 ms. Note that this period is longer than expected. We calculated a period of $20 \text{ ms} \times \frac{1}{2} \times \frac{80^\circ}{360^\circ} = 2.2 \text{ ms}$ for the black sub frame.

The additional 0.3 ms results from the fact that no light is projected either during the transitions from the green to the white segment and from the white to the blue segment. Also note that image frames start while blue (B) is being projected.

Signal 2 (middle) is the camera triggering signal. Since we acquire images with 16.7 Hz, only every third vertical sync signal is taken for generating the camera triggering signal. Here, the synchronization unit is programmed in a way that the camera starts to acquire the images 5.8 ms after the vertical sync, right at the beginning of the first sub frame of every third image frame. The CCD integration time for the camera is set to 2.5 ms.

5.4 Projection blanking by using linear polarization filters

Polarization filters are widely used for passive stereo projections; the working principle is explained e.g. in [4]. Here, two projectors simultaneously (or one projector alternately) project polarized images for the left and the right eye. The user wears goggles with polarization filters as well. They ensure that only the right image can be seen by the right eye, and the left image can only be seen by the left eye. For passive stereo projections, mostly circular polarization filters are used.

Figure 7 shows the setup for blanking the projection by using linear polarization filters. For polarizing the light, we used two HN38 linear polarization glass filters from ScreenLab. They both provide a transmittance of 38 % in the visible spectrum. If the two filters are placed in series and twisted by 90 degrees with respect to each other, the transmittance goes down to 0.05 %. The two linear polarization filters were oriented with 45 degrees and 135 degrees to the projection, respectively. One filter in front of the projector only allows 45 degrees polarized light to pass through and strike the screen. The user sees the projected monoscopic image without any auxiliary means. Another filter in front of the camera lens only allows 135 degrees polarized light to pass through and hit upon the imaging sensor. Since the screen does not depolarize the light, the projected image is not visible to the camera. Ghosting is negligible.

For the projection, a DLP projector should be used. Its light can be polarized with an external polarization filter in any desired orientation. This is not the case with most of the LCD projectors, which already emit linearly polarized light that has not the desired orientation. Their polarization could only be changed by larger modifications of the projector. A circular polarization can not be used because of the employed projection screen, which only allows linear polarization. In addition, the linear polarization allows a better channel separation than with a circular one. A bright projector is preferred, since the polarization filter in front of the projector only transmits 38 % of the light. Therefore, the brightness of the projector is reduced by 62 %. For our tests we used an F1+ SX+ single-chip DLP projector from...
Figure 7: Setup for projection blanking by using linear polarization filters. The filters are mounted in front of the projector and camera lenses in a way that they polarize light with 45 degrees and 135 degrees, respectively. The screen must not depolarize the light projected onto it.

Projection blanking by using polarization filters allows simultaneous image acquisition and display, making any synchronization needless. The CCD integration time for the camera is not limited by the system. We operated the camera internally triggered at a frame rate of 15.0 Hz, and set the integration time to 17 ms.

5.5 Discussion of the different shuttering methods

Periodic shuttering, as done with the external shutter, by projecting black frames, and black sub frames, leads to sequential image acquisition and display by time-multiplexing. In order to have a flicker-free projection, the frame rate of the displayed image must be at least about 60 Hz, which gives a maximum duration of about 16.7 ms for one full cycle of image acquisition and image display. However, both the image acquisition time and the image display time should be as long as possible. When using a standard camera, a too short image acquisition time results in a dark picture. Only expensive (> 5'000 $) high-speed cameras (> 100 Hz) provide a good image quality even with short exposure time and office light conditions. On the other hand, shortening the image display time results in a reduced brightness of the projected image, since the duty cycle of the projector is reduced. One could compensate this effect by using a brighter projector, but the bright peaks are not convenient and possibly irritate the user. Therefore, time-multiplexed image acquisition and image display is always a tradeoff between a high quality of either the acquired or the displayed image.

The advantage of a periodic shuttering is that the screen must not meet special requirements, i.e. preserving the polarization of the incident light. Periodic shuttering by projecting black frames with a stereo projector gives the best results. Today’s stereo projectors even support optical refresh rates up to 120 Hz, which leads to an image projection time of 8.3 ms at a rate of 60 Hz, and a maximum CCD exposure time of 8.3 ms. However, stereo projectors are still expensive non-consumer products.

Periodic shuttering with an external shutter gives full flexibility concerning the timing for the image projection and image acquisition. The drawbacks of this method are the reduced image quality due to the FE shutter in the light path of the projector, and residual projected light during the image acquisition.

Projection blanking by using polarization filters is the only method that allows simultaneous image acquisition and display. The theoretically unlimited CCD exposure time is an important advantage to the other shuttering methods. In practice, the CCD exposure time should not exceed about 17 ms, otherwise the acquired image is not sharp anymore if the user moves around. On the other hand, the polarization filter in front of the camera lens only lets 38 % of the incident light pass through and hit the CCD sensor, leading to a longer CCD exposure time. The drawback of projection blanking by using polarization filters is that it does not work with screens that depolarize light.

Table 1 summarizes the achievable shuttering and optical update rates for the four presented shuttering methods, as well as the corresponding maximum CCD exposure times. The choice of which shuttering method to use also depends on the application. Due to the limited CCD exposure times, not all shuttering methods allow the acquisition of high quality images, as they might be required by certain applications.

6 Segmentation and possible applications

In the following we show the segmentation that can be achieved when blanking the projection by using linear polarization filters in front of the projector and camera lens. Figure 8 shows a user who stands in front of a rear-projected dynamic content from an interactive virtual reality application. As shown in figure 9, the dynamic rear-projected content is almost completely removed.

After acquisition, all images undergo a segmentation of the user from the background. The algorithm we use is based on the illumination invariant method proposed by Mester et al. [5]. Assuming static backgrounds, previously

<table>
<thead>
<tr>
<th>Shuttering method</th>
<th>Shuttering and optical update rate</th>
<th>Maximal CCD exposure time</th>
</tr>
</thead>
<tbody>
<tr>
<td>External shutter</td>
<td>60 Hz</td>
<td>16.7 ms</td>
</tr>
<tr>
<td>Black frames</td>
<td>60 Hz</td>
<td>8.3 ms</td>
</tr>
<tr>
<td>Black sub frames</td>
<td>100 Hz</td>
<td>3.5 ms</td>
</tr>
<tr>
<td>Polarization filter</td>
<td>continuous</td>
<td>not limited</td>
</tr>
</tbody>
</table>

Table 1: Maximal CCD exposure times and shuttering rates for the different shuttering methods.
Figure 8: A user stands in front of a screen that displays rear-projected content of an interactive virtual reality application. The monoscopic projected image is linearly polarized, but can be seen by the user without any auxiliary means. The image is taken without the linear polarization filter in front of the camera lens.

Figure 9: This image shows exactly the same scene as depicted in figure 8, but with a linear polarization filter in front of the camera lens. By properly aligning the two polarization filters in front of the projector and camera lens, the projected image can be blanked for the camera.

Figure 10: Result of the segmentation with input images as shown in figure 9. Dark regions are hard to segment from the also dark background, e.g. black hair.

captured sequences of the screen without any user in front of it are compared to the actual image. This is done by using a statistic criterion that measures the colinearity of the actual color and the expected background in the color space. The result from the segmentation is shown in figure 10.

After successful segmentation, the movements of the acquired user could be analyzed for hand and upper body gestures. This opens the way to novel direct interaction with projected virtual objects, which could be interesting e.g. for video games. To acquire an image of the user, today’s systems use camera positions that ensure static or quasi-static backgrounds. However, such camera positions do not support direct virtual object interaction but only the control of a projected pointer [6]. All other gestures require full body acquisition, which can not be achieved by cameras parallel to the screen.

Another possibility is to use the output of the segmentation as an alpha mask for the acquired video. In that way, a video of only the foreground is available. This video can then be overlaid over the other digital content, enabling interesting applications.

One application could be for instance an advanced networked collaboration. If two electronic whiteboards are interconnected over a network, and a video of the scenery in front of each electronic whiteboard is acquired, the displayed shared electronic content can be enriched with a live video from the remote station. This allows transferring meta information in addition to the regular shared content.

7 Conclusion

A new system was presented that allows segmenting a user standing in front of a screen that displays a rear-projected dynamic image. Four different methods to blank the projected image in the background to the camera acquiring the scene were described and compared. One method was further analyzed concerning the segmentation, and first results were shown. Finally, some possible application scenarios were introduced.

8 Acknowledgments

This work is carried out within the context of the blue-c II project, funded by ETH grant No. 0-21020-04 as an internal poly-project. We would like to thank Konrad Wegener from inspire AG for supporting this work. We would also like to thank Roland Kehl for the segmentation,
Daniel Cotting, Christian Bacs, and Marcus Kroon for the fruitful discussions and help, and all the people whose comments contributed to this paper.

References


