July 01, 2019

Methods of Detecting and Mitigating Light Contamination in Active Stereo-Cameras

Michael Schoenberg

Charlie Wang

Methods of Detecting and Mitigating Light Contamination in Active Stereo-Cameras

Abstract:

This publication describes methods for detecting and mitigating image light-contamination when using active stereo cameras. Positioning an illuminator (or light projector) close to a camera underneath a cover glass of a user equipment (UE), such as a smartphone, can result in a direct optical-path reflection from the illuminator to the camera. If light "leaks" from the illuminator into one or both cameras, the emitted light contaminates the image(s) with extra light, which makes stereo matching difficult. To mitigate image light-contamination the UE leverages the symmetry of the cameras. First, the UE manufacturer calibrates both cameras to have consistent exposure with respect to each other. At the time of use, the UE captures one image from each camera with the illuminator enabled. The UE may perform this step when it detects a significant amount of energy (brightness) difference between the images. After the UE captures one image from each camera with the illuminator enabled, the UE performs a two-dimensional (2D) image (signal) processing. It is possible and, in some cases, prudent, the UE may directly perform statistical measurements in 2D signals. Nevertheless, performing statistical measurements of the 2D signals is computationally expensive. To simplify the computations, the UE converts the 2D signals into one-dimensional (1D) signals. After the UE converts the 2D signals into two 1D signals, the UE computes a statistical measure of the difference between the two 1D signals. After the UE has calculated the statistical measurements to determine whether image contamination has occurred, the UE may ask for user input when the distance is larger than a predefined threshold distance. Lastly, the UE performs stereo-matching using the images.

Keywords:

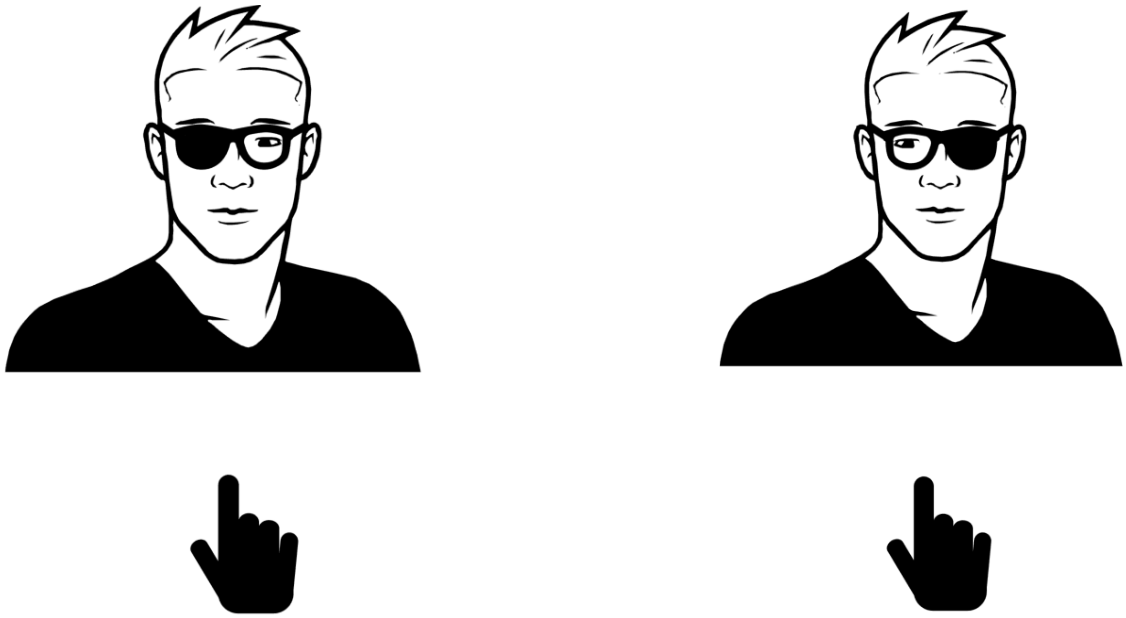
Active stereo-cameras, active stereo-system, assisted stereo-cameras, assisted stereo-system, image processing, signal processing, detecting optical-path reflections, image contamination, light contamination, two-dimensional signal, 2D signal, one-dimensional signal, 1D signal, 2D-to-1D signal conversion, parallax.

Background:

In an active stereo system, positioning an illuminator (or light projector) close to a camera underneath a cover glass of a user equipment (UE), such as a smartphone, can result in a direct optical-path reflection from the illuminator to the camera. The optical-path reflection contaminates an image captured by the camera and damages stereo matching. To better understand the problem, it is worthwhile to briefly discuss stereo-matching methods and optical-path reflections.

Stereo-Matching Methods

Depth from stereo is a computational technique for estimating the depth of points in a scene. To do this, correspondence needs to be found between a point in one image that is captured by a first camera and the same point in another image that is captured by the second camera. Given prior knowledge of the location and optical characteristics of the cameras, the parallax between the point in two cameras can be used to estimate how far away the point is from the two cameras. To illustrate the parallax between the point in two cameras, consider the example in Figure 1.



1A) Covered Right Eye **1B) Covered Left Eye**

Figure 1

Assume Tom goes for a routine eye check-up at his optometrist. The optometrist begins the examination by asking Tom to cover his right eye and look at the optometrist's index finger, as illustrated in Figure 1A. Then, the optometrist proceeds by asking Tom to cover his left eye and look at the optometrist's index finger, as illustrated in Figure 1B. The optometrist has not moved or shifted the position of their index finger. Tom, however, perceives a shift in the positions of the optometrist's index finger depending which eye has an unobstructed view. This visual effect describes the parallax between the same point that is viewed by two different cameras.

Dense stereo-matching systems try to estimate this effect for every pixel in an image. To find which points in each image match, it may be necessary to examine not just the pixel itself, but the "neighborhood" of the pixel—in other words, a patch. Like many engineering solutions, there is a tradeoff for such an approach because bigger patches are more computationally costly. In addition, by examining the patch (instead of the pixel) it may lead to added error because each pixel in the patch typically has a different shift, and these shifts are blended together when using the patch for searching. On the other hand, when trying to search for a single pixel or an exceedingly-small patch in the image captured by the first camera, there is a risk of not finding the corresponding pixel or the corresponding small patch in the image captured by the second camera. Thus, patch size is an important parameter of a stereo matcher that is tuned to the data.

Furthermore, using patches to disambiguate points in the image may be challenging because portions of the image may lack texture. As described herein, image texture refers to a set of metrics calculated in image processing that are designed to quantify the perceived texture of the image. Image texture provides information about the spatial arrangement of color or color intensities in the image or a selected region of the image. When the image lacks texture, it is difficult to determine what part of the image captured by the first camera corresponds to the image captured by the second camera. For example, if a patch in the image captured by the first camera is grey, it will correspond with every grey patch in the image captured by the second camera. In this instance, all-but-one grey patch is the correct-corresponding grey patch. To mitigate this problem, some stereo systems project light into the scene to provide texture to the image. These types of stereo systems are often referred to as active (or assisted) stereo systems.

Optical-Path Reflections

Although often avoidable with proper optical simulation and design, manufacturing active stereo systems poses challenges when a light projector is placed near a camera, such as in the case of the smartphone. If light "leaks" from the projector into one or both cameras, the emitted light contaminates the image(s) with extra light—light that is not normally present in the scene. Causes of such "leakage" can include reflection off the smartphone's cover glass, other optical elements placed in the light path, thin material layers placed in the optical path, or other reflecting sources. For example, placing a mirror in the optical path can reflect light back into a camera. Similarly, some thin material layers, such as skin oil from a user's hands, can return a sufficient percentage of the emitted light into the camera(s) to significantly contaminate the image(s). The magnitude of this effect depends on the exact optical design of the system and, in the case of thin films, can typically be avoided using optical simulation. In instances when an optical-path reflection is not avoidable, the optical-path reflection results in image contamination. Image contamination is a problem for stereo-matching methods because it causes a scene to look different between the cameras. Consequently, image stereo-matching may be even more challenging when image contamination occurs in only one of the cameras.

Therefore, for proper image stereo-matching, it is desirable to have a technological solution to handle optical-path reflections that contaminate images captured by one or both cameras.

Description:

This publication describes methods for detecting and mitigating image light-contamination when using active stereo cameras. Positioning an illuminator (or light projector) close to a camera underneath a cover glass of a user equipment (UE), such as a smartphone, can result in a direct optical-path reflection from the illuminator to one or both cameras. If light "leaks" from the projector into the camera(s), the emitted light contaminates the image(s) with extra light. Causes of such "leakage" can include reflection off the smartphone's cover glass, other optical elements placed in the light path, thin material layers placed in the optical path, or other reflecting sources. The optical-path reflection contaminates the image(s) captured by the camera(s) and makes stereo matching more difficult.

Using the symmetry of the cameras, this publication describes statistical measurements to determine whether image contamination has occurred. If there is greater image contamination in one camera compared to the other camera of the active stereo system, the image captured by the first camera does not look the same as the image captured by the second camera. Untangling this difference may be challenging. For one, the cameras are not in the same position in space, and they view a different scene; recall the parallax in Tom's example in Figure 1. In addition, the cameras capture a different view of the scene, and the captured images have different content. To determine whether an image is contaminated, the UE needs to decipher the context between both camera views. Figure 2 helps explain how the UE detects and mitigates image smudges due to image light-contamination.

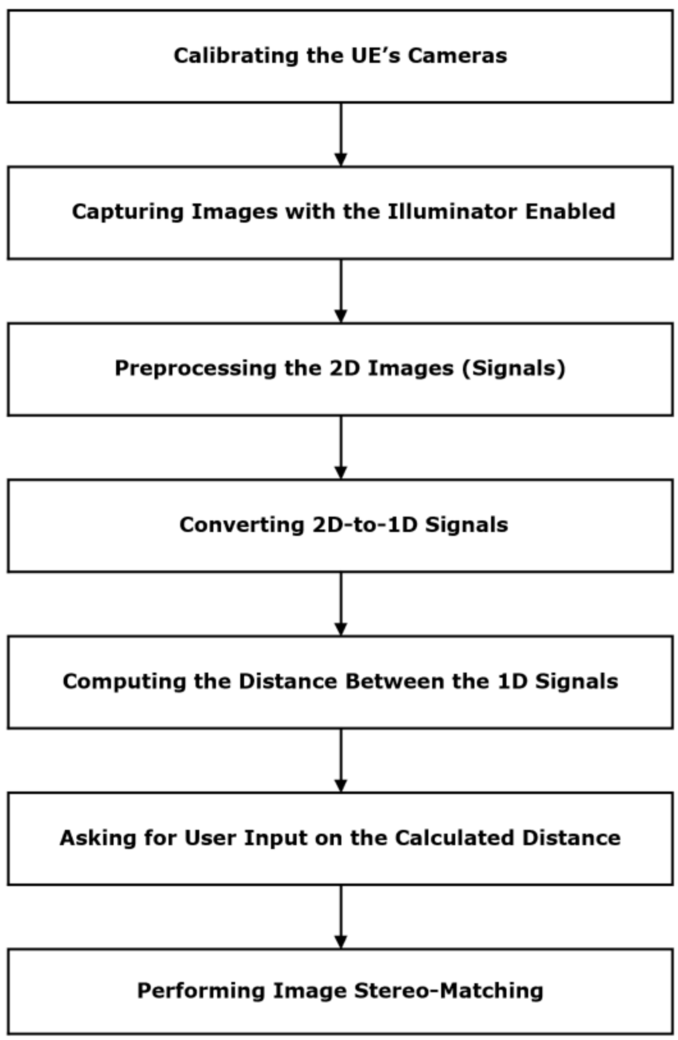


Figure 2

As illustrated in Figure 2, the UE manufacturer calibrates both cameras to have consistent exposure with respect to each other. Calibrating both cameras with respect to the exposures is important because when both cameras view the same scene, it is expected that the images captured by both cameras will have a similar amount of energy—a similar amount of brightness. If there is a significant amount of extra brightness in one of the images, that is a good indication that light contamination has occurred.

At the time of use, the UE captures one image from each camera with the illuminator enabled; refer to Figure 2. The UE may perform this step when it detects a significant amount of energy (brightness) difference between the cameras. As described herein, brightness refers to the perceived brightness of an object by a user or the UE. A low brightness may refer to an energy level that is less than a predefined threshold level, such as approximately 50%, 40%, 25%, 15%, and so on. This predefined threshold may be set by the UE manufacturer or defined by a setting selected by the user. A high brightness may refer to a brightness level that is greater than a predefined threshold level, such as approximately 50%, 60%, 75%, 85%, 95%, or 100%.

After the UE captures one image from each camera with the illuminator enabled, the UE performs a two-dimensional (2D) image (signal) preprocessing; refer to Figure 2. In some embodiments, the UE may crop non-overlapping portions of the images, due to different camera views of the scene, using the UE manufacturer's camera-calibration information. In some embodiments, the UE may blur the images to remove high frequency information. In some embodiments, from the images captured with the illuminator enabled, the UE subtracts another pair of images captured by the cameras without enabling the illuminator.

It is possible and, in some cases, prudent, the UE may directly perform statistical measurements of the 2D signals. Nevertheless, performing statistical measurements of the 2D signals is computationally expensive. To simplify the computations, the UE may convert 2D signals into one-dimensional (1D) signals using various techniques; refer to Figure 2. In some embodiments, the UE may compute a 1D signal by summing rows or columns of the 2D signals. In some embodiments, the UE may compute a 1D signal by finding the mean of rows or columns of the 2D signals. In some embodiments, the UE may compute a 1D signal by finding the median of rows or columns of the 2D signals. In some embodiments, the UE may compute a 1D signal by finding the variance of rows or columns of the 2D signals. In general, the UE generates a 1D signal across one dimension of the 2D signals by collapsing the other dimension into some type of summary statistic.

After the UE converts the 2D signals into two 1D signals, the UE computes a statistical measure of the difference between the two 1D signals using various techniques; refer to Figure 2. In some embodiments, the UE may perform a one-tailed or a two-tailed Kologomorov-Smirnov test of the 1D signals. In some embodiments, the UE may compute the Cramer-von Mises criterion of the 1D signals. In some embodiments, the UE may compute the Wasserstein distance of the 1D signals. It is worth noting that computing the Wasserstein distance is particularly advantageous, because it achieves a high degree of accuracy using a relatively low computational power. In some embodiments, however, the UE may skip the 2D-to-1D signal conversion, and the UE may directly compute the statistical distance of the 2D signals.

After the UE has calculated the statistical measurements to determine whether image contamination has occurred, the UE may ask for user input when the distance is larger than a predefined threshold distance; refer to Figure 2. The UE may ask the user for input on the calculated difference using a user interface, such as a toast or a small pop-up on the bottom of the screen (*e.g.*, a smartphone's screen), an audible question (*e.g.*, a UE's voice-assistant feature), a pop-up graphical user interface (*e.g.*, a notebook's screen), a preview of the images, and so forth.

Lastly, the UE performs stereo matching using the images; refer to Figure 2. When performing stereo matching using the images, the UE uses the calculated distance to mitigate the differences in the stereo matcher. In some embodiments, the UE may down-weight the image direct current (DC) (or direct brightness) term using the calculated distance. In some embodiments, the UE may linearly blend the gradient of the images with the images themselves, using the distance as a weight. In some embodiments, the UE may perform local contrast normalization. In some embodiments, the UE may blend local contrast normalization with the original images using the calculated distance as a weight.

In conclusion, using statistical measurements to determine whether image contamination has occurred and mitigating the contamination using signal processing improves the quality of the images and enhances user experience.