October 17, 2019

Measuring Stereo Camera Alignment Using Depth Map Smoothness

Edward T. Chang

Measuring Stereo Camera Alignment Using Depth Map Smoothness

Abstract:

This publication describes systems and techniques directed to measuring stereo camera alignment using depth map smoothness as a quality metric for the stereo calibration process. The quality of the depth map generated by the stereo cameras, as measured by smoothness, provides feedback indicative of a need (or not) for stereo calibration to account for camera misalignment. A smooth depth map indicates the stereo camera is aligned. Measuring depth map smoothness closes the feedback loop for gauging how well the stereo calibration process is working.

Keywords:

stereo cameras, camera alignment, camera calibration, depth map, depth map generation, depth measurement, depth sensing, depth map smoothness, depth value, face authentication

Background:

Dual cameras, also referred to as stereo cameras, enable the capture of stereo images of a scene because the two cameras are located at different locations and capture two slightly different views of the same scene. The two images are compared, and a disparity is calculated or generated, to obtain depth information about the scene. Many devices, including smartphones, use dual lenses or cameras (*e.g.*, a "stereo camera") for capturing stereo images and obtaining depth information about a scene.

Stereo cameras must be calibrated in order to be used for depth measurement. However, the calibration itself may be prone to error, which would lead to errors in depth measurement. A common problem in the stereo camera context is calculating a good depth map from the depth information of a scene. This difficulty exists partly because in order to generate accurate depth the two cameras need to be rectified, meaning their scan lines must correspond to each other. Namely, if the cameras are perfectly rectified or aligned, the scan lines of both cameras will lie along the same scan line (*e.g.*, overlap or be epipolar). However, if the scan lines are not epipolar, the cameras need to be recalibrated to rectify the scan lines to ensure proper depth calculations for the stereo camera. If the cameras are not aligned, then a depth map generated from the cameras will be unrealistic and irregular, degrading into visual noise or garbage.

Camera calibration includes measurement of the camera intrinsic properties such as lens distortion, focal length, and optical center, and extrinsic properties such as the relative positions of the cameras with respect to each other. If a camera is bumped or jolted while in use (e.g., a smartphone with a stereo camera is dropped), its calibration may need to be recomputed.

Camera calibration can be done either in a controlled or uncontrolled environment. A controlled environment uses specific charts and computer vision algorithms, for example, in a factory or a laboratory, and metrics such as reprojection error are used to measure the calibration quality. An uncontrolled environment uses real-world images, for example, when a user first turns on their camera, or each time they use their camera. However, this is an open-loop process with no means to detect the calibration quality.

Figure 1 illustrates the standard stereo camera calibration process. The calibration target is well-known in computer vision and tied closely to the calibration algorithm. Some example targets are checkerboard, Calibu, and ALVAR. Ideally, the stereo cameras should have perfect lenses (*e.g.*, no distortion) and be aligned such that the vector joining the physical camera centers is parallel to the scanlines of the camera (*e.g.*, the cameras lie on the same line, and point in directions parallel to each other). However, during manufacturing, it generally is not possible to achieve perfect alignment, which necessitates calibration. The calibration algorithm refers to the process of measuring each of the camera parameters (intrinsics) and the relative positions between the cameras (extrinsics). These constitute the calibration parameters as the result of Figure 1.

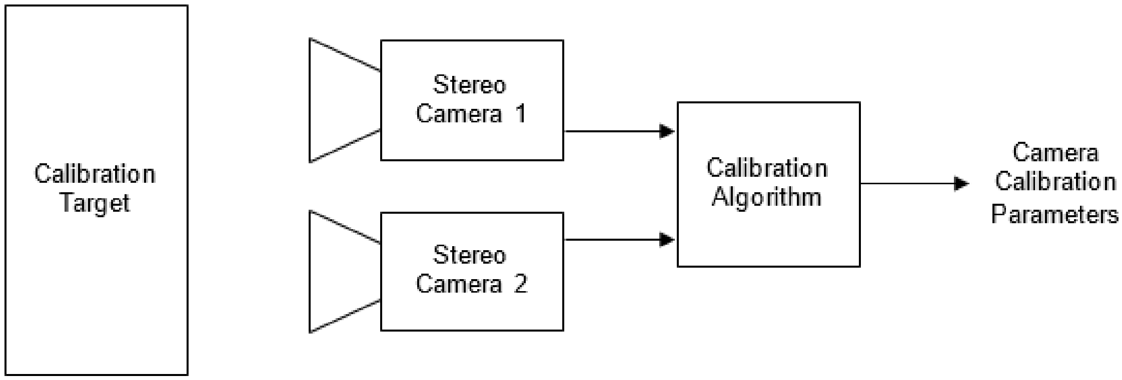

Figure 1

Description:

This publication describes systems and techniques directed to measuring stereo camera alignment (or camera calibration quality) using depth map smoothness as a quality metric for the stereo calibration process. The quality of the depth map generated by the stereo camera, as measured by smoothness, provides feedback indicative of a need (or not) for stereo calibration to account for camera misalignment. Generating the depth map includes the camera detecting an image target region that has low depth variation, such as a patterned plane facing the camera directly, or a human face. Depth map smoothness of an image is defined by neighboring pixel depth intensity values having similar disparities and minimal variations. Smoothness may be calculated in several ways including, for example, by using the standard deviation of the depth values, by using the number of invalid depth pixels returned by the depth map, or by measuring local variations in depth across a window and slide the window across a region of interest.

Measuring depth map smoothness helps close the feedback loop for gauging how well the stereo camera calibration process is working. A smooth depth map indicates the stereo camera is aligned. A non-smooth depth map indicates the stereo camera needs re-calibrating.

Figure 2 illustrates an example of the systems and techniques of this disclosure. Dual cameras (*e.g.*, a stereo camera), such as in a smartphone device, capture an image of a scene (a person's face in this instance), the image is detected and cropped to a smooth area, and a depth map is generated. The smoothness of the depth map is measured, and a camera calibration quality metric is determined based on the smoothness. If the camera calibration quality metric (smoothness) meets a pre-determined threshold of quality smoothness, then the stereo camera calibration is recognized as being sufficiently accurate for face authentication. However, if the depth map is not of satisfactory smoothness as indicated by the quality metric, then the camera needs calibration to better alignment. This entire process is beneficial for camera use in the field (*e.g.*, ordinary use) and where there is not a controlled environment such as where a laser range finder could sample every point of an image/scene to determine a real depth map.

The target scene is chosen such that it contains a smooth depth area (*e.g.*, a wall, a board, a human face, a mannequin face). The area should have texture to enable depth map generation, but with low depth variation to enable the generation of an accurate depth map that avoids noisy (not smooth) results. In this example, a person's face is used. Any object known to be smooth in depth space may be used as that provides confidence that the smoothness should be high for generating the depth map. If the image of a known smooth object is captured for generating a depth map, then a real or prior validated reference depth is unnecessary to determine how good a stereo depth map is generated.

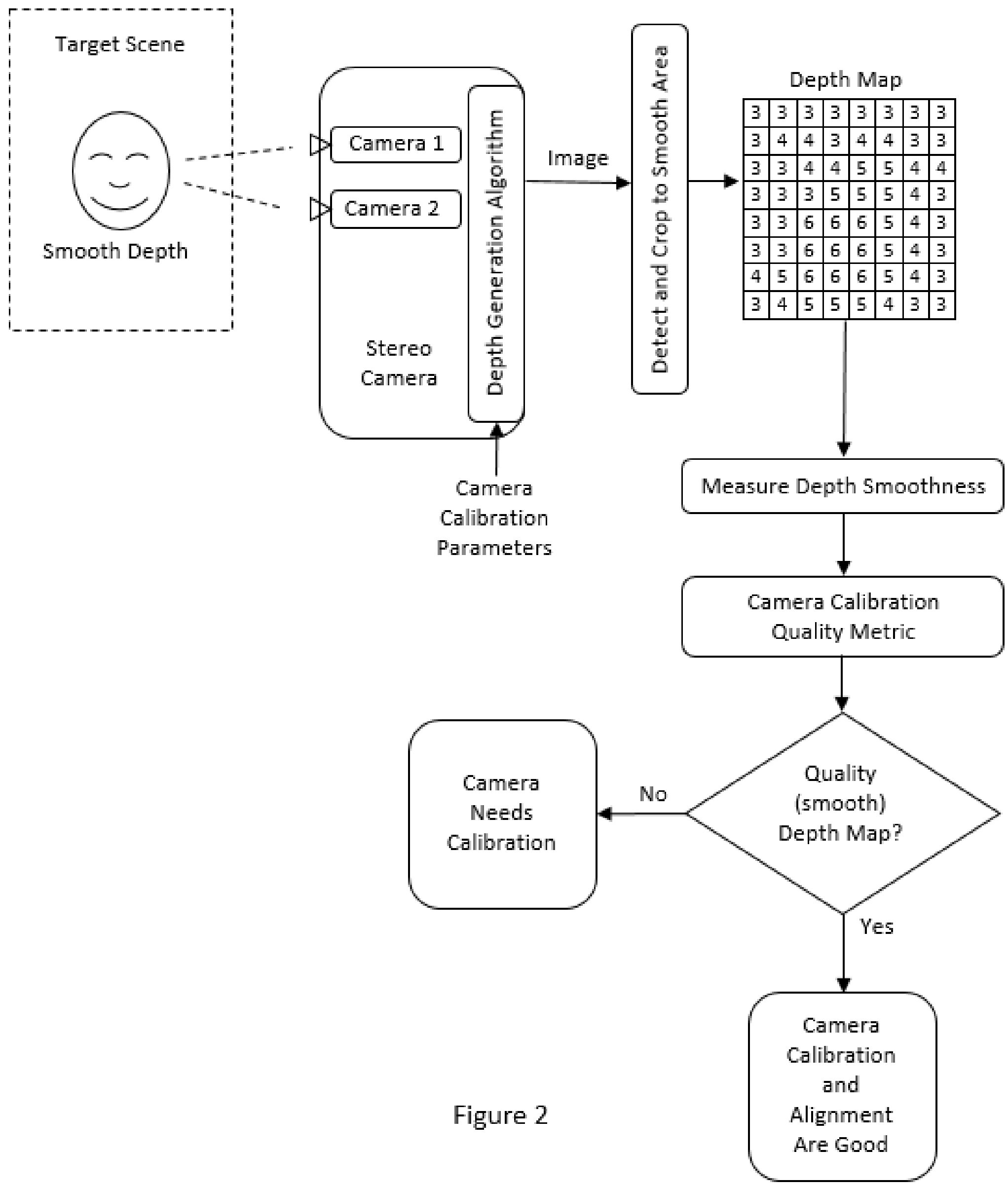

Figure 2

A standard detector for detecting the smooth depth area is used. For example, if the smooth depth area is a patterned plane, then the algorithm can be a pattern detection. As another example, if the smooth depth area is a face, then the algorithm can be a face detection.

The depth map generated by the stereo camera is represented in this example by a grid of varying pixel depth intensity values. Each small cell with a number in the grid represents a pixel of the face or target image, and differing numbers represent varying depth space measurements or pixel intensities. If the stereo camera is calibrated and aligned well, then a good quality, generally smooth depth map will be generated with minimal variations in pixel intensities. However, if the stereo camera is not aligned correctly, then typical characteristics of a poor-quality depth map include larger varying pixel depth intensity values across the depth map or discontinuous areas in the depth map.

Any well-known depth map generation algorithm may be used for generating the depth map. For example, a depth map generation algorithm may use active stereo that projects texture out onto the scene, or it may use passive stereo with no projection but uses both stereo cameras, or it may use machine learning-based algorithms that don't use traditional geometry but rather a neural net that trains off an extensive data set. One example algorithm is the Stereo Matching in O(1) with Slanted Support Windows (SOS) algorithm, and there are many other well-known algorithms in the computer vision literature. Note that for optimal performance, the camera calibration parameters (which are input to the stereo depth algorithm) must be very accurate. An example purpose of this disclosure is to measure the accuracy of those camera calibration parameters.

A smoothness metric is used to measure the quality of the depth map. High fluctuations in smoothness calculations of the depth map tend to indicate there is a problem with the camera alignment. For example, as local patches (areas) of depth values are calculated in areas across the depth map, and considered relative to accumulated depth value statistics taken across the entire depth map, a good determination of the smoothness of the depth map is achieved. A good quality, smooth map reflects only minimal changes in pixel intensity values. A bad depth map reflects high rates of change or discontinuous pixel intensity values.

Any well-known smoothness test, method, or algorithm may suffice for measuring the smoothness of the depth map. For example, the standard deviation of the depth values across the map could be measured. Alternatively, the number of invalid depth pixels returned by the map could be detected – the invalid depth pixels being regions where the depth map algorithm failed to produce depth value. Common areas for invalid pixels may be where an occlusion occurs between the two cameras (*e.g.*, one camera can see, and the other cannot (referred to as stereo shadow)). Another example would be to measure local variations in depth across a window and slide the window across a region of interest.

In considering the smoothness metrics of the depth map, generally, the higher the smoothness, the better the depth map because it more closely approximates a real-life scene. Real scenes typically include objects that are generally smooth in-depth and continuous in space – they are generally not objects that are close, and then far away, and then close, and then far, etc. Thus, under this disclosed technique, a real reference depth measurement is not needed to confirm how good the stereo depth map is. Rather, generating the stereo depth map from a smooth object suffices. Then, by measuring the smoothness of the depth map, the status of the camera alignment is detected.

The camera calibration quality metric is the final output of the system of Figure 2, which measures the quality of the calibration parameters generated by Figure 1. If the metric indicates a low quality as measured by depth smoothness, then the camera calibration parameters should be adjusted to restore the stereo depth algorithm to optimal performance. If the metric indicates a satisfactory quality as measured by depth smoothness, then the camera calibration parameters are sufficiently accurate.

In summary, proper camera calibration to account for stereo camera misalignment are necessary for the stereo camera algorithm to produce a good depth map. So to determine if the calibration is correct, we analyze the quality of the depth map. One quality metric to use is depth map smoothness, which is a qualitative feedback component indicative of whether the calibration is correct. If the depth map is smooth, then camera calibration is good. If the depth map is not smooth, then the camera needs recalibration.

References:

[1] Hahne, Uwe. May 3, 2012. Real-time depth imaging. 10.14279/depositonce-3244.

[2] Khoshelham, Kourosh & Oude Elberink, Sander. February 1, 2012. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. Sensors (Basel, Switzerland). 12. 1437-54. 10.3390/s120201437.

[3] Tankovich, Vladimir & Schoenberg, Michael & Fanello, Sean & Kowdle, Adarsh & Rhemann, Christoph & Dzitsiuk, Maksym & Schmidt, Mirko & Valentin, Julien & Izadi, Shahram. October 2018. SOS: Stereo Matching in O(1) with Slanted Support Windows. 10.1109/IROS.2018.8593800.