Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

Self-calibration of depth sensing systems based on structured-light 3D

Vikas Ramachandra, James Nash, Kalin Atanassov, Sergio Goma
Qualcomm Inc., 5775 Morehouse Drive, San Diego CA 92121, USA

ABSTRACT

A structured-light system for depth estimation is a type of 3D active sensor that consists of a **structured-light projector**, that projects a light pattern on the scene (e.g. mask with vertical stripes), and a camera which captures the illuminated scene. Based on the received patterns, depths of different regions in the scene can be inferred. For this setup to work optimally, the camera and projector must be aligned such that the projection image plane and the image capture plane are parallel, i.e. free of any relative rotations (yaw, pitch and roll).

In reality, due to mechanical placement inaccuracy, the projector-camera pair will not be aligned. In this paper we present a **calibration process** which measures the misalignment. We also estimate a scale factor to account for differences in the focal lengths of the projector and the camera.

The three angles of rotation can be found by introducing a plane in the field of view of the camera and illuminating it with the projected light patterns. An image of this plane is captured and processed to obtain the relative pitch, yaw and roll angles, as well as the scale through an iterative process. This algorithm leverages the effects of the misalignment/ rotation angles on the depth map of the plane image.

Keywords: Structured light, 3D, depth map, yaw, pitch, roll, scale, projective transforms, and calibration.

1. INTRODUCTION

1.1 Background

A structured-light system (Fig. 1) for depth estimation is a type of 3D active sensor that consists of:

- A **light projector**, that projects an illumination pattern (e.g. mask with vertical stripes) on the scene
- One or more **camera sensors**, which capture the illuminated scene.
- **Algorithms** that process the captured scene and estimate the depth of the objects based on information extracted from the projected pattern

Based on the received patterns, depths of different regions in the scene can be inferred. For this setup to work optimally, the camera and projector must be aligned such that the projection image plane and the image capture plane are parallel, i.e. free of any relative rotations (yaw, pitch and roll) and translations.

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

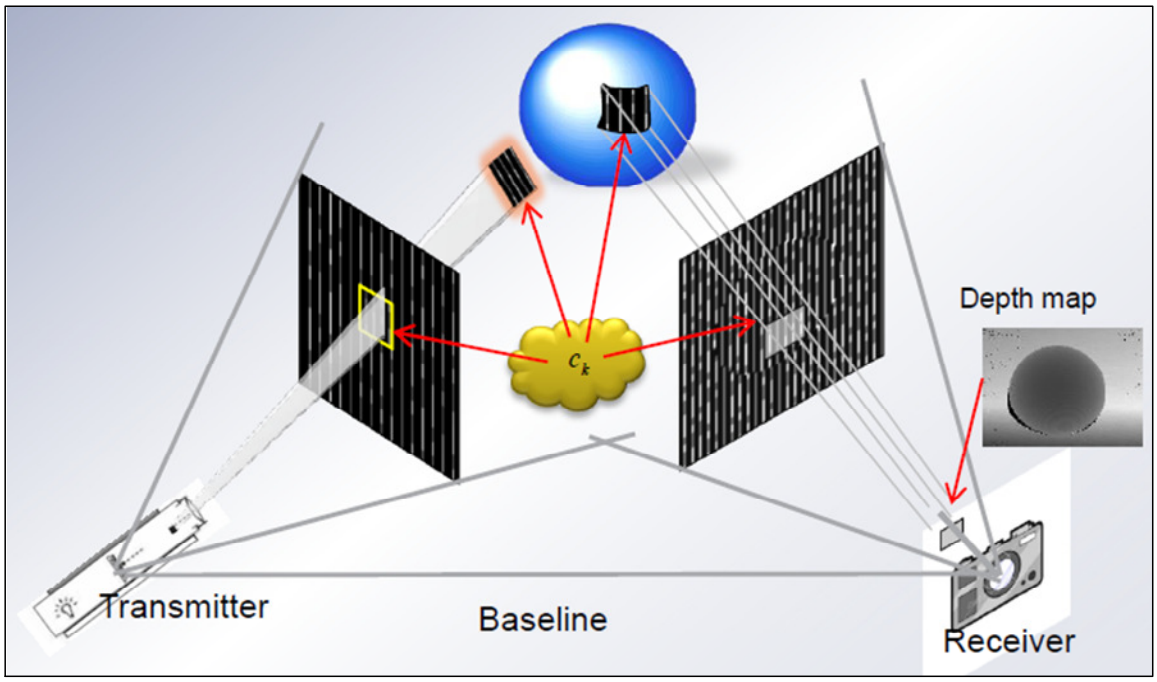


Figure 1: Structured-Light System

1.2 Motivation

When the images are parallel, the pixels in the projected image differ from those in the received image by horizontal shifts and there is no vertical disparity between the two images. However, in a real world scenario, the camera/receiver and the projection system/ transmitter cannot be mounted perfectly parallel. In reality, due to mechanical placement inaccuracy, the projector-camera pair will not be aligned (Fig. 2). To overcome this, we have implemented a calibration process which measures the relative misalignment between the projector and the camera. We also estimate a scale factor to account for differences in the focal lengths of the projector and the camera.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

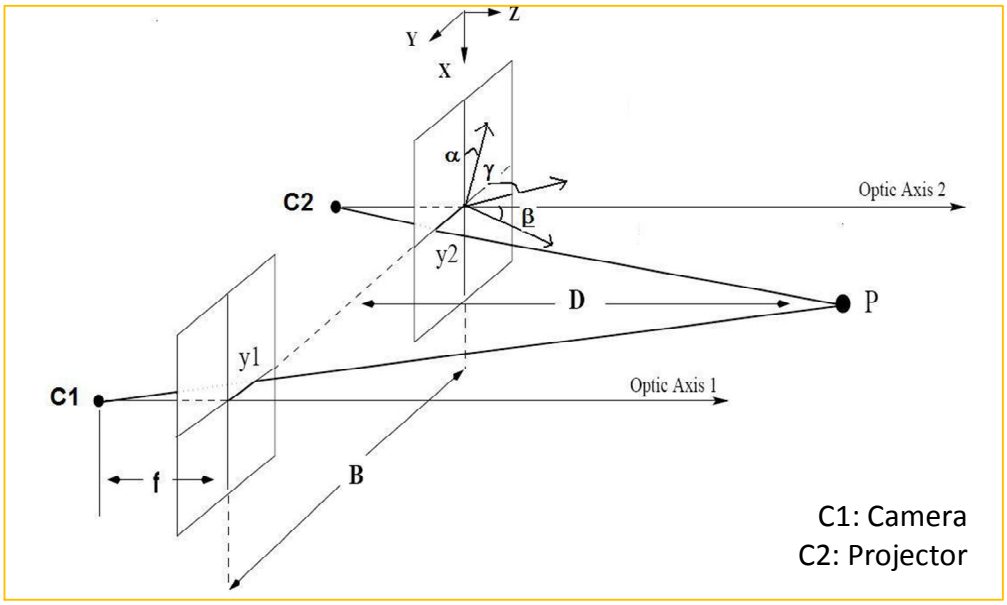


Figure 2: Model of projector/ camera

1.3 Previous work

Previous efforts in this direction broadly fall into two categories: Full calibration and partial calibration.

Techniques based on full calibration, i.e. estimating all intrinsic and extrinsic parameters of the projector and the camera [1-7], are not useful in a practical scenario that involves a consumer device. This is because these techniques require images of special charts (such as the checkerboard) at multiple locations, making the calibration process cumbersome and not user friendly. Moreover, these techniques require the knowledge of 2-dimensional coordinates of several points in the projected calibration pattern and its captured image, which cannot be obtained from a system which provides a depth map (which has only horizontal disparities).

Another set of techniques [8-11] assume that only a limited set of parameters need to be estimated. Thus, they do not need special charts and setup. However, in the existing literature, those techniques concentrate only on estimating the yaw angle and the variation of the baseline with time. We believe that this is too restrictive for our application.

In our case, based on the system specifications, we need to estimate all three rotation angles. However, we do not need to estimate the intrinsic parameters and the relative shifts between the projector and the camera. Moreover, we do not have the luxury of capturing multiple images nor do we assume the use of special charts or the availability of 2D coordinate positions of points on such charts. We assume that the only input we have to estimate the calibration parameters is the estimated depth map obtained from the projection of the light pattern using the system. These limitations and requirements led us to design a specific calibration algorithm to suit our needs.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

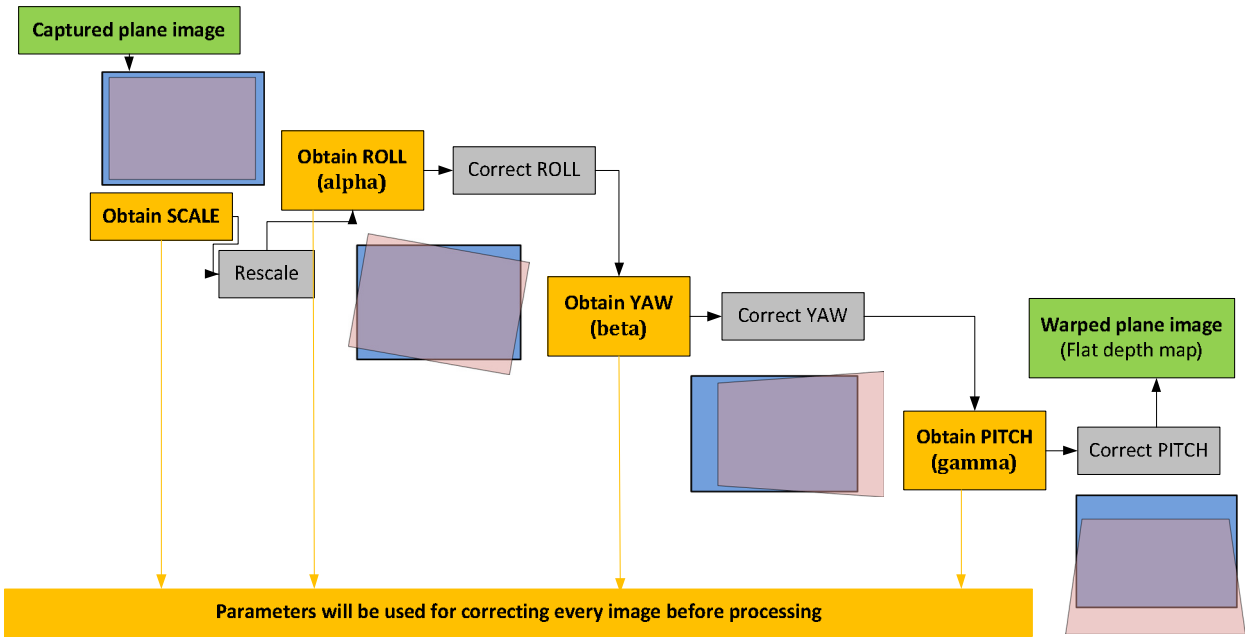


Figure 3: Calibration Algorithm

2. PROPOSED METHOD

2.1 Overview

Based on manufacturing tolerances, it was found that the shifts along the X and Z axes are negligible (translation along Y gives the baseline distance, see Fig. 2). Therefore, we only need to remove the relative rotations between the camera and the projector. The three angles of rotation can be found by introducing a plane in the field of view of the camera and illuminating it with the projected light pattern. An image of this plane is captured and processed to obtain the pitch, yaw and roll angles, as well as any scaling difference between the projected and received patterns. Yaw, pitch and roll angles affect the received image in particular ways, as shown. Their effect can be teased out by analyzing the captured image. Thus, the calibration process involves the estimation of the three rotation angles and a rotation matrix is built from these angles.

During the rectification stage, all received images are warped. This involves the application of a projective transform on the received images, according to the estimated rotation matrix in the above step. This will make the projected and received image planes parallel as well as maintain the scale of the pattern, as required by further processing stages to estimate the depth map.

To summarize, given a general setting for the structured-light system, the calibration process will provide the parameters needed to transform this general system to the particular configuration of our model, that is, with both the projection and the capture plane located in the same plane of the 3D space (XY plane in the model) with no vertical disparities, separated by a given baseline B, and scaled such that the codebit is defined by 4 pixels along both the horizontal and vertical direction (a codebit is a ‘0’ or a ‘1’ and it occupies a certain number of pixels in the projected pattern). The proposed calibration method consists of the steps described in the following sections and shown in Fig. 3.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

2.2 Scale estimation

This module processes the captured image to measure the number of pixels per codebit (Fig. 4). The input image is first filtered using a vertical blurring filter (Fig. 5). This removes the short vertical code stripes. Then, the Fourier transform of the image is calculated. The transformed image is center shifted and normalized. Since the projected pattern has periodic structure, this is reflected as peak frequencies in the Fourier domain. Peaks are detected in the normalized Fourier transform (Fig. 6). The scale is then detected using the extreme peak locations found. This is given by the following relation:

Scale = desired pixels per bit * (peak1 location-peak2 location)/(image width).

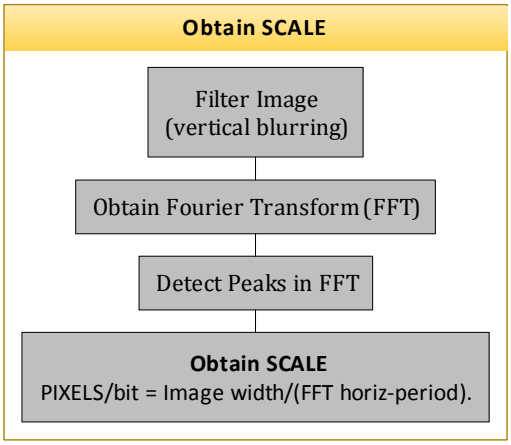


Figure 4: Scale estimation

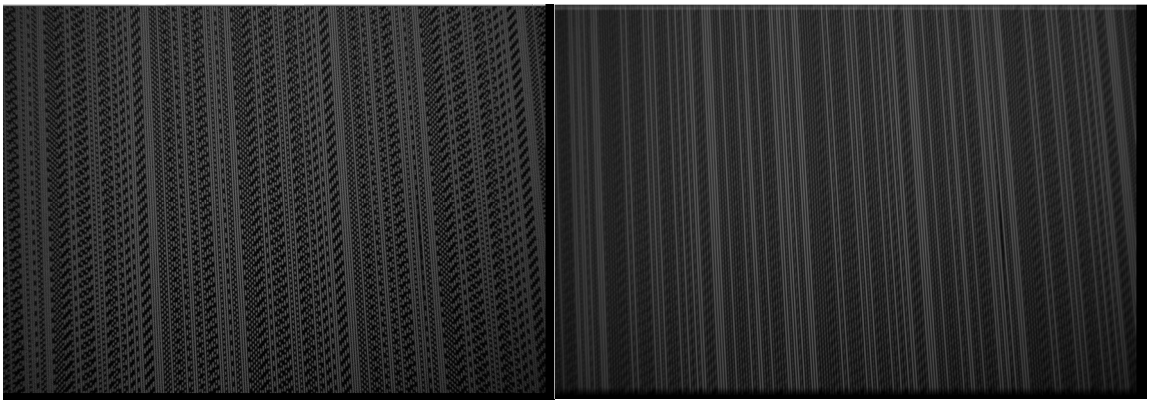


Figure 5: Captured image of plane (left) and filtered image (right)

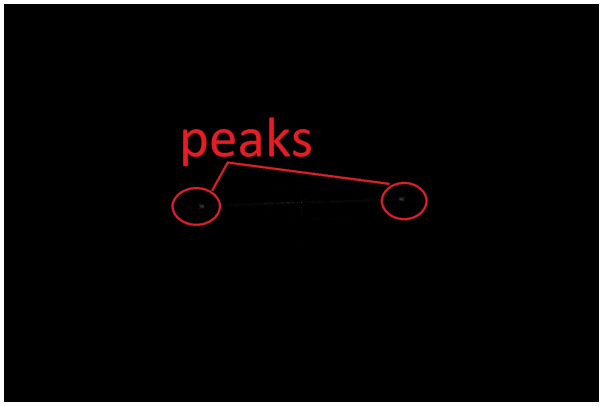


Figure 6: Shifted/scaled Fourier transform of filtered image: detected peaks excluding the center peak.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

2.3 Roll estimation

To estimate roll (Fig. 7), first, the edge map of the image is found using the Canny edge detector (Fig. 8 left). The edge map is converted to the Hough space. The Hough transform parameterizes lines based on their slope and distance from the Y axis. In the Hough space, peaks are detected and peaks which have similar parameters are joined to form lines. Only long lines are considered. Finally, orientations for all the long lines are found (Fig. 8 right) and their median is selected as the roll angle.

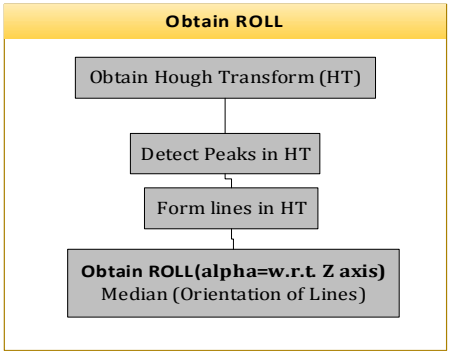


Figure 7: Roll estimation

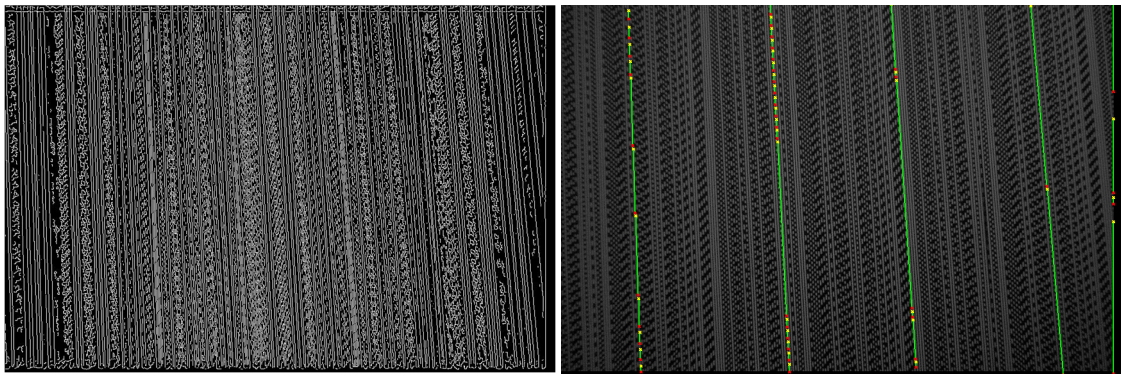


Figure 8: Edge detected image(left) and Long Hough lines detected in green (right)

2.4 Yaw estimation

To estimate the yaw angle (Fig. 9), first, the depth map of the roll/scale corrected image of the plane is found (Fig. 10 right). The middle row of the depth map is extracted (Fig. 10 right). For the middle row, the slope of the depths across the image width is considered. We look for valid depth values (depth values which are detected and flagged as correct, i.e. excluding missing depths and non-decoded depths) in that row, and estimate the depth slope from the extrema values and their column locations. The reason for selecting the central row to estimate the yaw is that for the rows about the centre of the received image, the effect of the pitch angle is minimal, and they are mostly affected by the yaw rotation misalignment.

There is a direct relation between this slope and the yaw angle, based on projective geometry. The relation between depth slope and yaw is given by:

$$\delta d = f\frac{y_1 + z_1 a}{-y_1 a + z_1} - f.y_1 / z_1$$

Where x,y,z: 3D coordinates, f: focal length, a=yaw.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

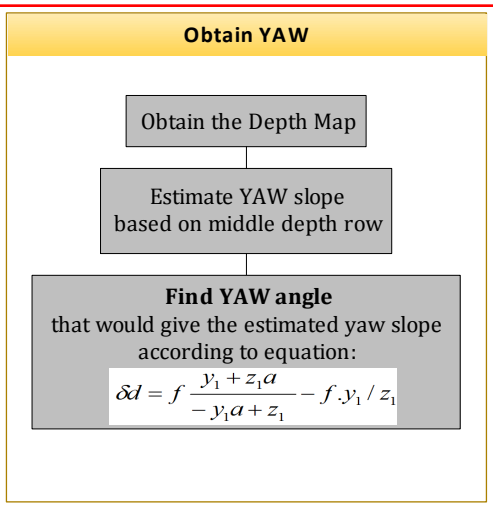


Figure 9: Yaw estimation

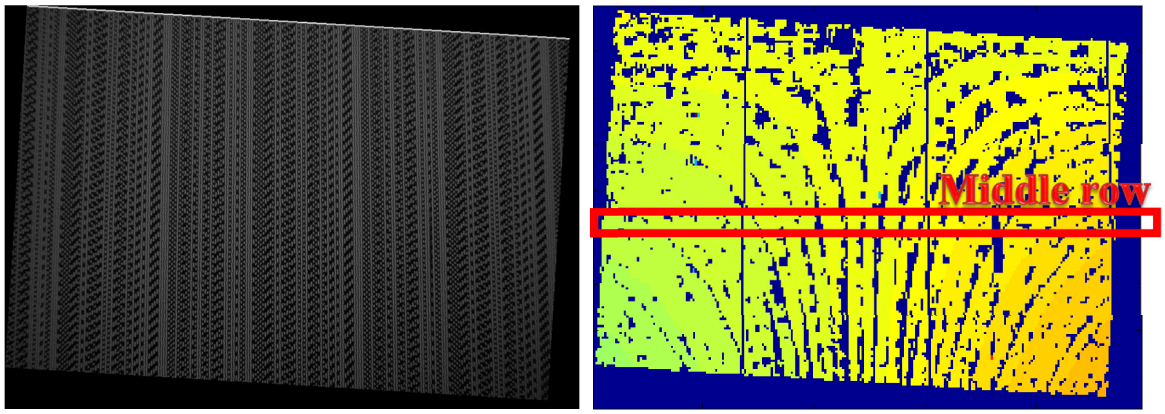


Figure 10: Roll/scale corrected image (left) and Depth map to get yaw angle (right).

2.5 Pitch estimation

To estimate the pitch angle (Fig. 11), first, the depth map of the roll/scale/yaw corrected image of the plane is found (Fig. 12 right). For the depth map column in the right (or left) corner, the slope of the depths is considered (from top to bottom). The reason for selecting the corner column to estimate the pitch is that for the extreme corner columns of the received image, the effect of the yaw angle is minimal, and they are mostly affected by the pitch rotation misalignment.

There is a direct relation between this slope and the pitch angle based on projective geometry, as given by the equation below. This is used to fit the best possible pitch angle. The relation between depth slope and pitch is given by:

$$\delta d = f\frac{y_1}{x_1 b + z_1} - f.y_1 / z_1$$

x,y,z: 3D coordinates, f: focal length, b=pitch.

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

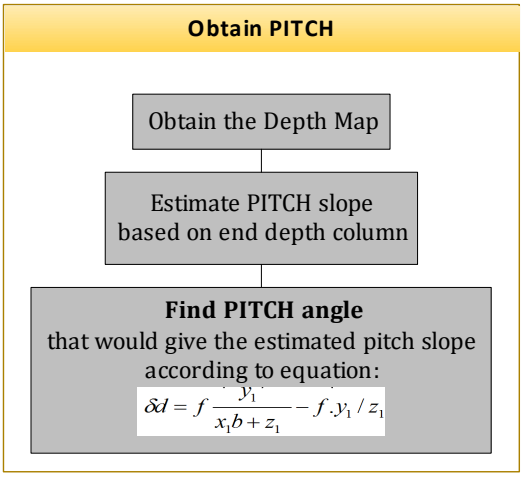


Figure 11: Pitch estimation

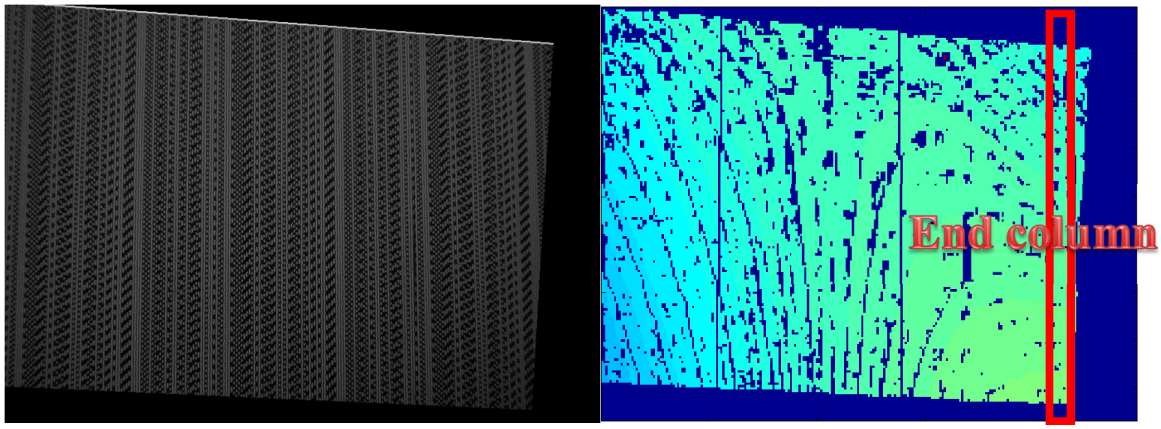


Figure 12: Roll/scale/yaw corrected: input to pitch estimation (left) and depth map for pitch estimation (right)

2.6 Rectification

Once the rotation angles have been found (as explained above) all future captured frames can be warped (i.e. rectified) to make them parallel to the projected image. The warping process will use the projection matrix obtained from the estimated rotation angles (yaw, pitch, roll).

The rectified images are obtained by projecting all the images captured by the structured light system using the projection matrix (Fig. 13). Fig. 14 left shows the input and Fig. 14 right shows the image warped to make the plane parallel to the camera-projector rig.

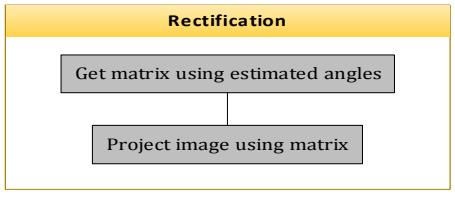


Figure 13. Rectification

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

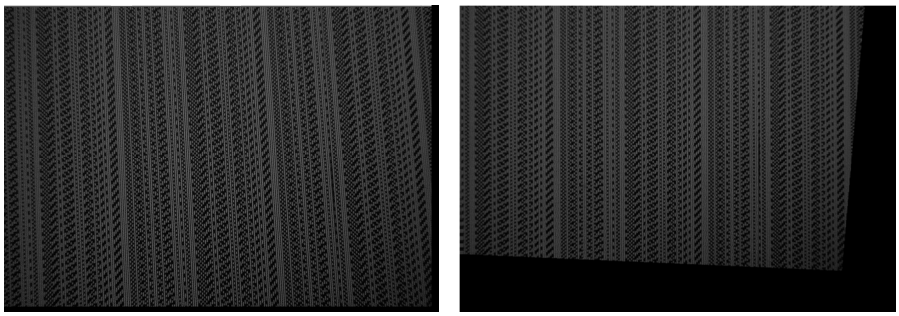


Figure 14: Input image and warped image

3. RESULTS

The proposed calibration algorithm was tested with both synthetic and real world images. For synthetic images, the input image was warped with known rotation angles and fed as input to the calibration algorithm. The estimated rotation angles were compared with their expected values. It was found that the proposed algorithm was accurate to within 0.1 degree for all the rotation angles tested.

For real world images, depth maps before and after calibration were compared visually. As seen below (in Fig. 16) it is clear that the calibration process greatly reduces the depth artifacts due to the misalignment of the camera and projector. The artifacts can be seen as warped depth planes, as indicated by bands of different shades on the depth map of a flat plane (these arifacts also lead to banding in depth maps of 3D scenes with objects at different distances from the rig). In an ideal situation, the depth map of a flat plane parallel to the camera-projector rig should be of a single shade (i.e. flat). This is what the proposed method tries to achieve, as seen in the post-calibration depth map (Fig. 16). Fig. 17 shows a real world scene image before and after warping (based on the calibration parameters), and Fig. 18 shows their corresponding depth maps.

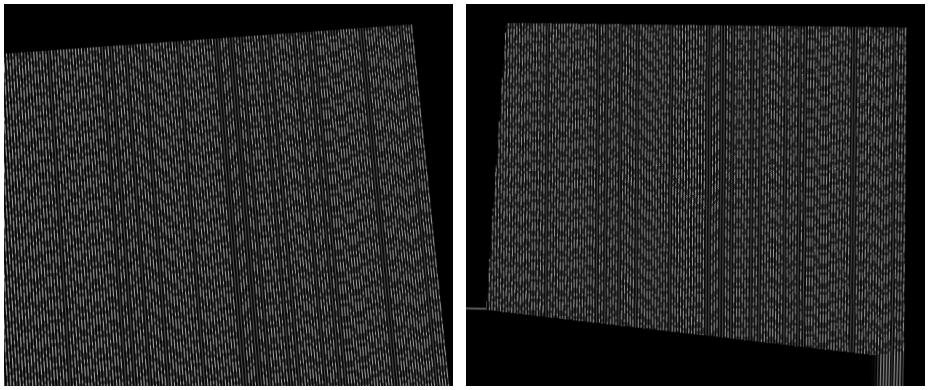


Figure 15: Plane image before (left) and plane image after calibration(right)

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

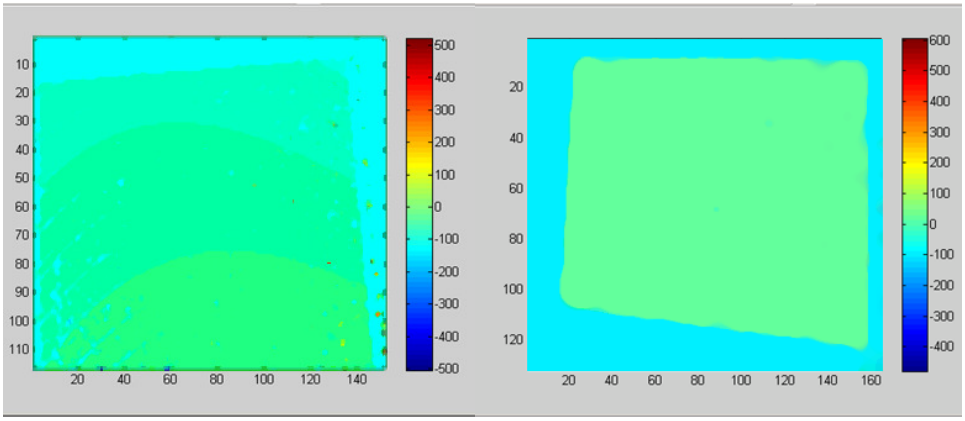


Figure 16: Depth map before (left) and depth map after calibration (right): flat

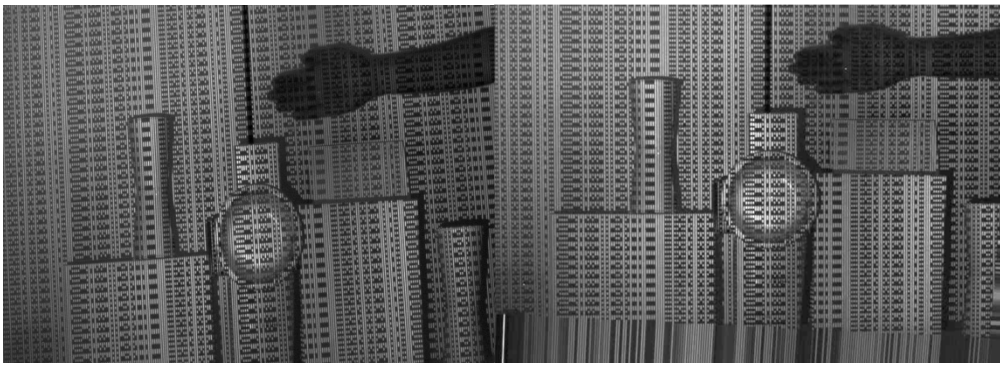


Figure 17: Scene before (left) and scene after calibration (right)

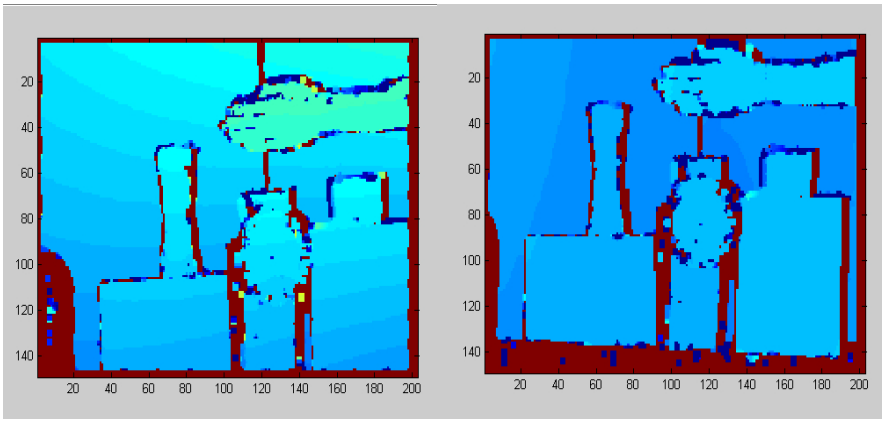


Figure 18: Depth map: misalignment artifact Artifacts removed due to calibration

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

4. CONCLUSION

In this paper, a calibration procedure is proposed which aligns the camera and the projector of a structured light system for depth sensing. Due to mechanical placement inaccuracy, the projector-camera pair will not be aligned. We have proposed a **calibration process** which measures the misalignment in terms of the yaw, pitch and roll angles. We also estimate a scale factor to account for differences in the focal lengths of the projector and the camera.

The three angles of rotation can be found by introducing a plane in the field of view of the camera and illuminating it with a projected light pattern. The image of this plane is captured and processed to obtain the relative pitch, yaw and roll angles, as well as the scale through a sequential process. Once the angles have been estimated, these angles define a projection matrix. The calibration process only needs to be run during the setup of the system. After that, future captured frames will be rectified based on the calibration parameters.

The proposed method was found to be accurate in estimation, upto 0.1 degrees of the actual rotation angles. Also, this method was found to work well for a real world camera-projector setup, based on a visual inspection of the depth map post calibration, which is free of banding artifacts introduced due to the projector-camera misalignment.

REFERENCES

[1] Shuntaro, Y., Mochimaru, M., and Kanade, T. "Simultaneous self-calibration of a projector and a camera using structured light," IEEE Computer Vision and Pattern Recognition Workshop. (2011)

[2] Schmalz, C., "Robust Single-Shot Structured Light 3D Scanning," University of Erlangan, Thesis, (2011)

[3] Fernandez, S, and Joaquim, S. "Planar-based camera-projector calibration," 7th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE. (2011)

[4] Zhang, Y., and Fang, J.. "Calibration for structured-light vision system based on homography." 9th International Conference on Electronic Measurement & Instruments, IEEE. (2009)

[5] Liao, J, and Cai L. "A calibration method for uncoupling projector and camera of a structured light system," IEEE/ASME International Conference on Advanced Intelligent Mechatronics. IEEE. (2008)

[6] Zhang L., and Nayar S. "Projection defocus analysis for scene capture and image display," Transactions on Graphics (TOG). Vol. 25. No. 3, ACM. (2006)

[7] Huynh, Q. "Calibration of a structured light system: a projective approach," Conference on Computer Vision and Pattern Recognition, IEEE. (1997)

[8] Peng, T."Algorithms and Models for 3-D Shape Measurements Using Digital Fringe Projections," University of Maryland, Thesis. (2006)

[9] Li, Y. F., and Chen, S. Y. "Automatic recalibration of an active structured light vision system," Transactions on Robotics and Automation, IEEE. (2003)

[10] Chen, S. Y., and Y. R. Li. "Self recalibration of a structured light vision system from a single view." International Conference on Robotics and Automation, IEEE. (2002)

[11] Sansoni, G., Carocci, M., & Rodella, R. "Calibration and performance evaluation of a 3-D imaging sensor based on the projection of structured light," Transactions on Instrumentation and Measurement, IEEE. (2000)

SPIE USE: ____ DB Check, ____ Prod Check, Notes: