A Handheld Texel Camera for Acquiring Near-Instantaneous 3D Images

B.M. Boldt

Scott E. Budge
Utah State University

R.T. Pack

P.D. Israelsen

Follow this and additional works at: https://digitalcommons.usu.edu/ece_facpub

Part of the Electrical and Computer Engineering Commons

Recommended Citation
A Handheld Texel Camera for Acquiring Near-Instantaneous 3D Images

Brandon M. Boldt, Scott E. Budge, Robert T. Pack, and Paul D. Israelsen

Abstract—A Texel Camera is a device which synchronously captures depth information via a ladar and digital imagery of the same scene. The ladar and digital camera are co-boresighted to eliminate parallax. This configuration fuses the ladar data to the digital image at the pixel level, eliminating complex post-processing to register the datasets.

This paper describes a handheld version of a Texel Camera which can be used to create near-instantaneous 3D imagery. The hardware configuration of the Texel Camera, issues and method associated with ladar/camera calibration, and representative imagery are presented.

I. Introduction

The ability to model three-dimensional surfaces has been an active area of research for many years. Several methods for capturing real-world objects and representing them digitally exist for various applications.

This paper describes the development of a handheld Texel Camera. A Texel Camera is a device which synchronously captures ladar depth information and digital imagery of the same scene[1]. Since the two datasets are captured synchronously, they can be “fused” together at the pixel level directly from the hardware. Previously built Texel Cameras have relied on tripod-mounted systems, but the development of a smaller device will aid in the flexibility of the system for many applications.

A textured model can be developed from a handheld Texel Camera, such as from the prototype shown in Fig. 1. Note the size of the camera compared to a US Quarter. The digital camera, evident on the right in the Figure, captures visual information while the ladar system measures three-dimensional data. The two cameras are co-boresighted, as well, to minimize any parallax present between the depth information and the digital imagery. Because the data is gathered in a structured geometry, a wire frame mesh can be constructed. After the mesh is known, the visual image can then be appropriately placed on the resulting surface which becomes a textured model.

A flash ladar system is used in the handheld prototype. It consists of a modulated light source, such as a laser or LED, an array of sensors capable of detecting the phase of the incoming light, and an optical system which focuses the light onto the sensor array. This allows for an entire scene of ladar points to be captured simultaneously. Because the Texel Camera contains no moving parts, it can be reduced to a handheld size.

II. Integration of Mechanical Systems

In order to test the feasibility of a handheld Texel Camera, the various components were integrated together at Utah State University. This integration consisted of mechanically placing the two prebuilt sensors so that their optical pathways coincide as shown in Fig. 1. This way the visual camera and the ladar system can be triggered together and their information can be spatially and temporally aligned.

The major components used to build the Texel Camera are a Micron 1280 x 1024 CMOS imaging sensor, a Canesta 64 x 64 CMOS flash ladar sensor, a cold mirror (which separates visible and infrared light by reflecting wavelengths below about 760 nm), and a mechanical mount to hold the cold mirror and allow the sensors to be co-boresighted. Because both of the sensors capture information one frame at a time, instead of one data point at a time, this setup allows the capture of both visual and ladar information to be performed simultaneously for an entire scene.

Note that the Canesta ladar sensor takes the measured range to the object at each pixel and converts it to a depth perpendicular to the focal plane array.
III. Minimizing Effects of Parallax Between Cameras

It is important in this configuration to consider spatial parallax between the two cameras. If the two cameras have parallax, there will be a difference in the angle in which the two cameras view an object. This angular change is not a constant shift with respect to distance and angle. Thus, two cameras which have a significant amount of parallax cannot have their scenes matched together by a simple angular shift. In order to eliminate the effects of parallax, the two cameras need to optically view the same scene from the same angle.

The point about which a camera can rotate in any direction without suffering a parallax effect observed between a frame taken from one pose and a frame taken from another pose is called the center of perspective (COP) [2]. In order for the two sensor systems which make up the Texel Camera not to suffer the effects of parallax, it is important that the centers of perspective of each camera coincide optically. The center of perspective of each camera is measured and each camera is then mechanically placed in a position for the respective centers of perspective to coincide.

There are four steps required to perform the co-boresighting of the two cameras after the cameras are assembled on a common mechanical mount. These are:

1. Adjust the tilt of the cold mirror so that the visual image is in focus evenly throughout the image, and the center of the image is located at a point in front of the ladar sensor lens and perpendicular to the Texel Camera.
2. Locate the COP of each of the cameras. To locate the position of the COP of a camera, a device called a panoramic mount is used. This device is typically used to allow cameras to capture scenes in such a way that the scenes can be stitched together into a panorama [3]. Once the assembly is mounted to the panoramic mount, the cameras are rotated vertically and horizontally to determine the COP.
3. Translate the mirror and/or the visual camera until the COPs are coincident.
4. Iterate steps 1–3 if necessary.

IV. Calibration of the Cameras

Once the mechanical alignment of the cameras has been accomplished, it is necessary to calibrate the images so that the effects of imperfect lens distortion is removed and the visual camera image is correctly mapped to the ladar data. This is done in three steps:

1. Calibrate the ladar intensity image to remove lens distortion by finding the intrinsic parameters of the ladar sensor.
2. Find the nonlinear polynomial matching parameters that map the visual image to the ladar intensity image.
3. Correct the ladar sensor measurements to a pinhole model to get improved ladar 3D measurements.

A. Background on Intrinsic Parameters

The intrinsic parameters of a camera system are the parameters that affect how the image is created on the focal plane. These parameters include variables such as the horizontal and vertical focal distances, the sensor offset from the center of the lens, and the radial distortion coefficients. These parameters are used to correct the ladar intensity image before matching the ladar image to the visual image, as well as to transform ladar’s range image into 3D space. A brief background in intrinsic parameters is given here.

The principal point of a camera is the intersection of the principal axis with the sensor array. For a rectangular sensor array, the camera frame of reference is defined as the frame of reference which has an origin that is located at the principal point, the z-axis located on the principal axis going away from the sensor array, the x-axis located parallel to the rows of imaging sensors, and the y-axis located parallel to the columns of the imaging sensors.

With this definition in place it is possible to locate where a point in space, defined as

\[
P_r = [X_r, Y_r, Z_r],
\]

would be projected onto the camera sensor array [4].

Given a pinhole model of a camera, the 2D projection of this point onto the sensor array \( (p_n) \) would be

\[
p_n = \begin{bmatrix} \frac{x_n}{y_n} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix}.
\]

This model assumes no distortion occurs in the image, even though this is rarely the case. The point after including the distortion caused by the optical system \( (p_d) \) can be calculated using the following model [4]:

\[
p_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = d_r \begin{bmatrix} x_n \\ y_n \end{bmatrix} + d_t,
\]

where \( d_r \) is the radial distortion multiplier and \( d_t \) is the tangential distortion vector which are defined by

\[
d_r = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6,
\]

where \( r = \sqrt{x_n^2 + y_n^2} \).
\[ d_t = \left[ \frac{2k_3x_ny_n + k_4(r^2 + 2x_n^2)}{k_3(r^2 + 2y_n^2) + 2k_4x_ny_n} \right], \]  
and  
\[ r^2 = x_n^2 + y_n^2. \]

The values \( k_1, k_2, \) and \( k_3 \) are the radial coefficients (dimensionless) and \( k_3 \) and \( k_4 \) are the tangential distortion coefficients (dimensionless).

The pixel position is related to the location on sensor array by the following:

\[
\begin{pmatrix}
  x_p \\
  y_p
\end{pmatrix}
= K
\begin{pmatrix}
  x_d \\
  y_d
\end{pmatrix},
\]

where \( p \) is the pixel position (in pixels), and \( K \) is the camera matrix. \( K \) is defined as:

\[
K = \begin{bmatrix}
  f_1 & 0 & c_{c1} \\
  0 & f_2 & c_{c2} \\
  0 & 0 & 1
\end{bmatrix},
\]

where \( f_1 \) and \( f_2 \) are, respectively, the horizontal and vertical focal distances. The parameters \( c_{c1} \) and \( c_{c2} \) are the horizontal and vertical position of the principal point with reference to the origin of the pixel matrix (such that \((0,0)\) is located at the upper left corner) \[4\].

The horizontal and vertical focal distances, \( f_1 \) and \( f_2 \), are scaling factors to scale the normalized coordinates, \( p_n \) or \( p_d \), to pixel coordinates, \( p_p \), in the horizontal and vertical directions, respectively \[5\]. When \( f_1 = f_2 \), then each pixel is square.

The camera parameters given in \((4)\)–\((8)\) must be found for the ladar system so that the effects of the lens can be removed. Methods such as those proposed by Zhang\[6\], Heikkila and Silven\[7\], or Tsai\[8\] can be used to do this.

### B. Polynomial Mapping

Each of the cameras has a different lens which focuses the light onto its sensors. This allows the sensor arrays to have a desired field of view, and it also leads to the two camera lenses having different optical systems. It follows that the two cameras will have a different field of view as well as a different amount of optical distortion.

When selecting the components chosen to build the Texel Camera, the optics of the visual camera should have as large or larger field of view as the ladar camera. This can lead to the visual camera having a larger amount of optical distortion than the optics used for the ladar sensor array. A method for matching the two systems to each other is a necessary step towards developing a Texel Camera.

We will now map the visual camera coordinates directly to the ladar camera coordinates. This renders the direct calibration of the visual camera unnecessary. Matching these optical systems together can be thought of as mapping one system to the other system. This process is described by

\[
x = f(u, v),
\]
\[
y = g(u, v),
\]

where \((x, y)\) are the coordinates of the visual camera system and \((u, v)\) are the coordinates of the corrected ladar intensity system. A polynomial mapping was chosen to transform the two cameras into the same projective space. This is given by Mayampurath \[9\] as

\[
x = a_x + b_x u + c_x u^2 + d_x u^3 + e_x v + f_x u v + i_x v^2 + j_x u v^2,
\]
\[
y = a_y + b_y u + c_y u^2 + e_y v + f_y u v + g_y v^2 + i_y v^2 + j_y u v^2. \]

The constant terms of each equation allow for an adjustment between the centers of distortion in the two cameras, the single value terms compensate for scaling between the two systems, the cross terms account for image tilting, and the higher terms account for the difference in curvature of the two systems.

In order to determine the 16 coefficients given in \((10)\), the two sensors were assembled and aligned, and several frames of a calibration image (a checkerboard pattern) were taken by both cameras. The images were preprocessed to extract all of the \((x, y)\) and \((u, v)\) camera coordinates matched to positions of the corners of the checkerboard squares in the calibration images. Then a least-squares fit method employed by Mayampurath was used to determine the coefficients needed for the polynomial mapping.

### C. Transformations of the Ladar Depth Data into 3D Space

Equations \((1)\)–\((8)\) describe how to locate where a point in space projects on the pixel array of a camera. However, it is desirable to calculate the location of the point in space based from the measurements of the depth data obtained from the array of the ladar sensor. The depth \((Z_r)\) of the real point in 3D space is found from ladar sensor measurements, but the other two \((X_r, Y_r)\) coordinates are unknown due to the distortion caused by the non-ideal ladar lens.

To find these coordinates, the undistorted normalized coordinates \((x_n, y_n)\) need to be found using the information given. Notice that there is no analytical solution to the inverse projection with distortion problem using \((3)\)–\((8)\). Given the equations for the distorted normalized coordinates \((x_d, y_d)\), and given the undistorted normalized coordinates \((x_n, y_n)\), these equations reduce to two seventh order polynomial equations:

\[
x_d(x_n, y_n) = k_5 x_n^7 + 3k_5 x_n^5 y_n^2 + k_2 x_n^5 + 3k_5 x_n^3 y_n^4 + 2k_2 x_n^3 y_n^2 + k_1 x_n^3 + 3k_4 x_n^2 y_n^3 + k_5 x_n y_n^6 + k_2 x_n y_n^4 + k_1 x_n y_n^3 + 2k_2 x_n y_n^2 + k_1 x_n y_n^2 + k_1 y_n^3 + 3k_3 y_n^2 + y_n.
\]

where \((x, y)\) are the coordinates of the visual camera system and \((u, v)\) are the coordinates of the corrected ladar intensity system. A polynomial mapping was chosen to transform the two cameras into the same projective space. This is given by Mayampurath \[9\] as

\[
x = a_x + b_x u + c_x u^2 + d_x u^3 + e_x v + f_x u v + i_x v^2 + j_x u v^2,
\]
\[
y = a_y + b_y u + c_y u^2 + e_y v + f_y u v + g_y v^2 + i_y v^2 + j_y u v^2. \]

The constant terms of each equation allow for an adjustment between the centers of distortion in the two cameras, the single value terms compensate for scaling between the two systems, the cross terms account for image tilting, and the higher terms account for the difference in curvature of the two systems.
No analytical solution can be used to solve these equations for \((x_n, y_n)\), though several methods have been developed to solve these equations in the literature [7]. An iterative method for solving an approximation to the undistorted normalized coordinates \((p_n)\) from the distorted normalized coordinates \((p_d)\) has been proposed by Melen [10]. Algorithm 1 summarizes this approach, where \(\delta(p_n) = d_r p_n + d_t\), and \(d_r\) and \(d_t\) are given in (4) and (5), respectively.

Algorithm 1 Iterative approach to transforming the image coordinates into 3D space.

Input:
- Distorted Normalized Coordinates, \(p_d = (x_d, y_d)\).
- Radial and Tangential Distortion Coefficients, \(k_1 - k_5\).
- Number of Iterations to Perform, \(n\).

Output:
- Approximate Undistorted Normalized Coordinates, \(\tilde{p}_n = (\tilde{x}_n, \tilde{y}_n)\).

Begin
For Each Pixel
- \(p_n = p_d\)
Loop For \(n\) Iterations
- \(p_n = p_d - \delta(p_n)\)
Next Iteration
Next Pixel
End

Since the ladar system gives the depth to the point in space, \(Z_r\), the other point coordinates can easily be solved for using (2) which inverted gives the equation for calculating a point in space given the normalized image locations, \(p_n\), and the depth of the point, \(Z_r\):

\[ P_r = [X_r, Y_r, Z_r] = [x_n Z_r, y_n Z_r, Z_r]. \quad (12) \]

Note that the values of \(x_n\) and \(y_n\) for each pixel in the ladar focal plane array must be computed only once during the calibration process and can then be stored in a lookup table.

V. INTEGRATION OF LADAR AND VISUAL INFORMATION

Because the ladar information and the visual data are captured together, the data can be viewed in real time with a 3D viewer providing a real-time presentation and eliminating the need for laborious post processing. Since the ladar information and the visual data are co-boresighted, the images can be accurately fused onto the generated point clouds from the ladar system. Because the handheld Texel Camera uses flash ladar, one visual image and one ladar image are used. This is a shift from earlier Texel Cameras [1], which take one small image for every ladar shot and the small images are then tiled together to create a full frame. With the use of flash ladar, this need to stitch patch images together to provide one complete image for the scene is eliminated.

With a corrected image and the points in space calculated, the data can be rendered into a textured model. This involves generating a texture map which relates the points in the visual image (“texture”) to the points in 3D space. The map creation is trivial since it is generated through the calibration process described above. The resulting texture map and point cloud can then be viewed in a 3D viewer.

An example of the steps in the process of creating a 3D image is given in Fig. 2. The corrected image from the visual camera is given in (a), the corrected ladar point cloud from the ladar sensor is given in (b), the structured mesh from the ladar is presented in (c), and the final 3D texel image generated directly from the Texel Camera is given in (d).

VI. CONCLUSION

This paper focused on the development of a Texel Camera which is small enough that it can be handled by one person without the need of a tripod. It was noted that there are non-ideal optical components associated with the ladar and low-cost digital camera which introduce nonlinear distortions in the captured data. These distortions were reduced through a calibration procedure that corrects data from both the visual camera and the ladar sensor system and fuses the data. Once the calibration was completed, no further post-processing is needed to obtain 3D images.

REFERENCES

Fig. 2. Steps to form a textured model from a Handheld Texel Camera: (a) Image taken from visual camera. (b) Point cloud taken from the lidar system and viewed from a different point of view. (c) Mesh constructed from known geometry. (d) Texel image viewed from same point of view as in image (b).