Laser Plane Range Finder
The Implementation at the CVL[1]

Tomáš Pajdla
Computer Vision Laboratory
Czech Technical University
CZ-121 35 Praha 2, Karlovo nám. 13
pajdla@vision.felk.cvut.cz

Research report No. K335-1995-98

October 27, 1995

Czech Technical University
Faculty of Electrical Engineering
Department of Control
Computer Vision Laboratory

[1]This research was supported by the Grant Agency of the Czech Republic, grants 102/93/0954, 102/95/1378 and European Union grant Copernicus No. 1068 RECCAD.

Abstract

This report describes the implementation of a 3-D scanner - Laser Plane Range Finder. The sensor builds on the triangulation and active illumination using laser light plane. It delivers 3-D coordinates of points in the volume of size 20 cm $\times$ 15 cm $\times$ 15 cm with the precision about 0.5 mm. Visual surface of planar topology can easily be constructed due to sensor's construction.

1 Introduction

Range finders utilizing active structured illumination attract attention since 3-D scene analysis has become widely popular. Besides passive techniques relying on stereo algorithms various kinds of active 3-D measurement devices were developed. See [1, 2] for the overview of the whole gamut of methods.

This text documents the implementaion of the Laser Plane Range Finder (LPRF) at the Computer Vision Laboratory. The text is organized as follows. In Section 2, the principles of the sensor are explained. Concise description of the system is given in Section 3.

2 The principle of the sensor

Basic principle of a structured light 3-D sensor builds on the concept of stereo vision. Two perspective cameras placed at different locations in space look at the same 3-D point. Each camera therefore provides a light ray in the space that projects the 3-D point onto the retinas by connecting the point, the centers of projection, and the point in the retina. The intersection of two rays then readily gives the desired 3-D cordinates of the point in space. Computing the coordinates 3-D points by intersecting the rays passing through the point's image projections is usually called triangulation. Bottleneck of this approach consists in automatic extraction of corresponding projections from images which turns out to be rather difficult problem as it is not easy to predict how the projections are related without knowing where the point is.

The use of an active controlled illumination can remove or at least alleviate "chicken and egg" correspondence problem. Usually, one camera is replaced by a source of light rays and a coding scheme allowing unique identification of the rays from the image of the illuminated scene. For instance, the most trivial setup would use just one ray at a time so that the image had contained maximally one bright point. Such a setup leads to very trivial solution to the matching of projections. However, it turns out to be very impractical since the whole image has to be captured and processed to measure one 3-D point.[1]

More advanced but still easy and efficient ray coding scheme projects planar pencil of rays at a time, so that an intersection curve, light trace, can be seen in each image. Ray identification is still unique as shows the following argument.

The figure 1 illustrates a camera with the center of projection at F_C and the retina plane π_C and a light projector providing a planar pencil of rays with the center F_P lying in the plane π_P. It is easy to see that there is a one-to-one linear relation, a planar homology with the fixed point F_C, relating points in π_P to points in π_C if $F_C \neq F_P$. This homology also forms a one-to-one relation between the pencil of lines in π_P passing through F_P and the pencil of epipolar lines in π_C passing through $\mathbf{e}$. The point $\mathbf{e}$, the epipole, is the image of the projector center F_P as seen by the camera. The image of each projected ray is therefore uniquelly identified by its corresonding epipolar line. There is

[1] Only to capture the images for measuring 512^2 points in 3-D would take almost 3 hours at CCIR video rate.

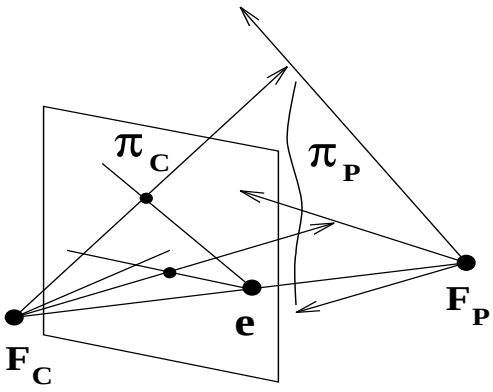

Figure 1: *Each ray emanating from F_P projects into its line passing through* **e**. *Each ray is uniquely identified by its corresponding epipolar line.*

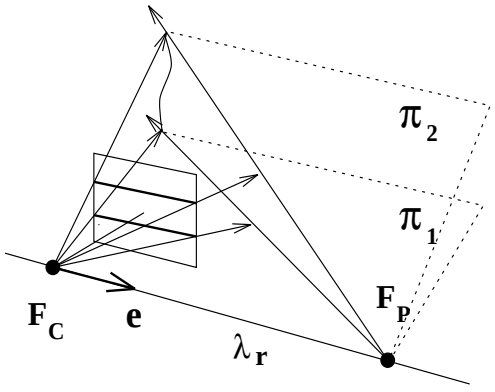

Figure 2: *By placing the center of projection, F_P, at the line λ_r, the search can be restricted to dertecting the brightest point in each image row.*

yet another fortunate consequence of the above relation. There can be only one intersection of an active light ray with any surface and the image of this intersection has to be found on the epipolar line in π_C. This means that the projection of a light trace into the retina intersects each epipolar line passing through **e** maximally in one point. Simultaneous search for the light trace along a pencil of lines in the image can be performed.

Pixels in CCD cameras are usually arranged in a rectagular array. Easy access to rows or columns suggests to use a "natural" pencil of lines in π_C corresponding, for instance, to the set of lines passing through the rows in the image. Such lines are parallel, hence forcing **e** to be at infinity in the direction parallel to the direction of rows, see Figure 2. The position of **e** depends on the position of F_P with respect to the camera. Placing F_P on the line of common intersection of all planes passing through the rows and F_C, λ_r, forces **e** to lie at infinity in desired direction.

Sending **e** to infinity not only allows easy access to the pencil of lines but also leads to optimal use of discrete sensor since the resolution does not change along the epipolar lines as it does in the general case.

Notice also, that F_P can be placed anywhere on λ_r and that the angle between the light plane and the camera can be set almost freely[2] what gives enough freedom to adapt the sensor setup to various conditions of measurement.

3 Sensor Description

The setup of the sensor utilizes the idea of triangulation in its simplest form. The light plane is generated by a special light projector that delivers planar pencil of rays emanating from one focus point. In our setup, the camera and

[2]The light plane should have reasonably large angle with all the row planes to provide numerically stable intersections.

the light plane projector are mounted on the fixed holder, forming the camera-projector rig, so that the projector center, F_P, lies on the line of intersection of all the planes passing through rows in the image and the camera projection center F_C. The rig itself is then mounted on the platform allowing a translation along the x axis of the world coordinate system, see Figure 4. Since the camera-

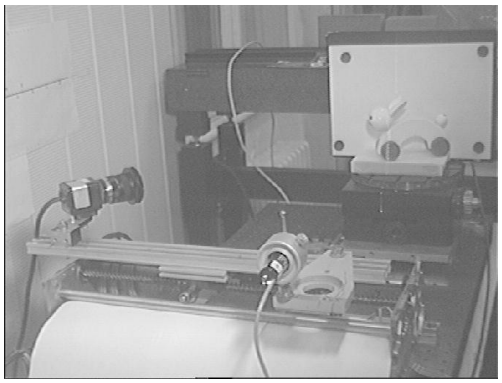

Figure 3: *The picture of the LPRF range finder.*

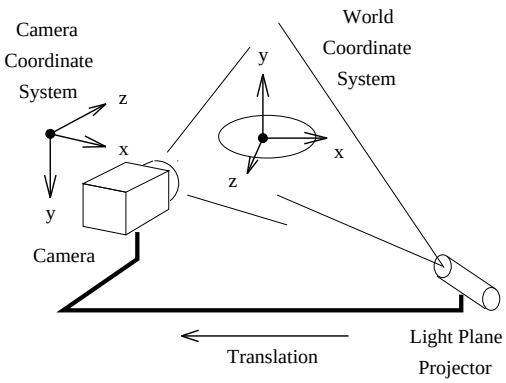

Figure 4: *The chart of the LPRF setup.*

projector rig is fixed, its calibration remains same for the whole sequence of light traces taken during the measurement of a 3-D surface. Moreover, the points reconstructed in particular scanlines lie in a sequence of parallel planes with even resolution in x direction.

The intersection curve of the light plane with a surface, the light trace, is captured by the camera looking at the surface, Figure 5. Due to the special arrangement of camera-projector rig, maximally one intersection per row is detected as the brightest pixel in the row. Image coordinates, (u, v), of each such pixel altogether with the position of the light plane in space provide data for the finding spatial coordinates of the point by the triangulation. 3-D data come from the sensor in the world coordinate system attached to the rotary platform. Figure 4 shows the coordinate systems.

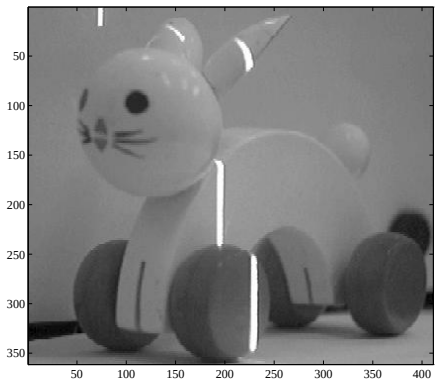

Figure 5: *The image of the light trace.*

The lens of the camera is equiped with a linearly polarized filter to suppress

unvanted mirror reflections and inter-reflections of the laser rays from shiny objects. Rays directly reflected from dielectrical surfaces come to the camera with linear polarization and can be elliminated by a perpendicularly polarized filter. On the other hand, desired image of the light trace comes to the camera as diffuse light with eliptic polarization.

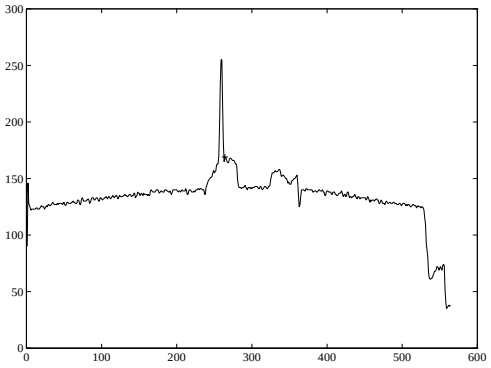

Figure 6: *The plot of the brightness in one "ideal" row*

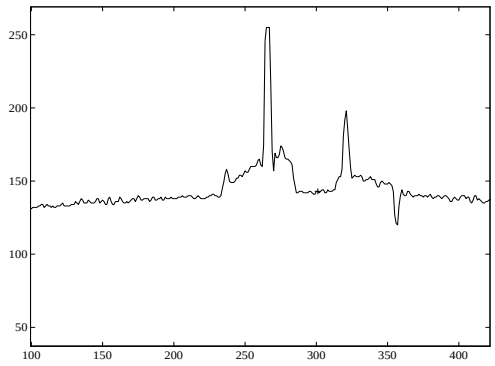

Figure 7: *The plot of brightness in one row with the saturation at pixel 265 and an inter-reflection at the pixel 325.*

3.1 Light trace analysis

Each row in the image is processed independently. It is assumed that the intersection of the light plane with a surface projects into the row as the brightest pixel. So, the pixel with maximal brightness in the row provides the most probable light trace pixel, see Figure 6. This strategy is very simple and it is obvious that a number of improvements can be thought. First, subpixel detection of the maximum would certainly improve resolution in z direction from current apx. 300 levels at least by the factor of 5. This is particularly desirable in the case when the maximum brightness reaches saturation level of the digitizer, i.e., in our case brightness = 255. Then, the precision drops significantly as it is apparent from Figure 7. Furthermore, if interreflections are present, as it is usual for metal objects, correct image of light trace can be masked by the light reflected from the surface, see peak at the pixel 325 in Figure 7 that is caused by inter-reflected light.

Currently, the value of brightness at the maximum is also used to express sort of certainty that detected maximum is acctually a part of the light trace. This should be soon replaced by more meaningful measure of belief.

4 Sensor calibration

Geometrical relation between the camera, the projector, and the world coordinate system, see Figure 8, is established through the explicit calibration of the camera-projector rig. The calibration uses known scene, i.e., the set of points

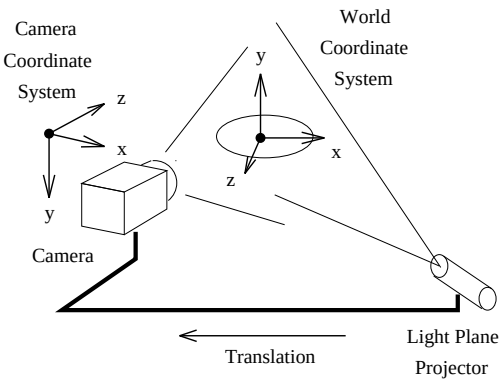

Figure 8: *Camera, projector, and a world coordinate system.*

coordinates of which are a priory known, to obtain necessary parameters for building the mapping $(u,\ v,\ t) \rightarrow (x,\ y,\ z)$, where u and v stand for image coordinates of light trace pixels and t parametrizes the position of the rig during the translation. The coordinates x, y, and z are then 3-D coordinates of the reconstructed point in the world Euclidean coordinate system. The calibration of the sensor consists of three steps. First, the camera is calibrated with respect to the world coordinate system. Then, position of the light plane with respect to the world coordinate system is fixed and measured. Finally, the parameters of rig translations are defined. The translation of the rig is parallel to the x axis, hence it is fully defined by two parameters, its origin and the length of a standard step along the translation vector.[3]

4.1 Camera calibration

The mathematical model of the camera used in this sensor utilizes a linear perspective camera with compensated radial distortions

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \tilde{u} + k_1(\tilde{u}^2 + \tilde{v}^2)\tilde{u} \\ \tilde{v} + k_1(\tilde{u}^2 + \tilde{v}^2)\tilde{v} \end{pmatrix}, \tag{1}$$

$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \begin{pmatrix} \frac{p}{r} \\ \frac{q}{r} \end{pmatrix}, \tag{2}$$

$$\begin{pmatrix} p \\ q \\ r \end{pmatrix} = M \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{3}$$

where u, v stand for nonlinearly distorted affine image pixel coordinates of detected point, i.e., coordinates that are accualy measured in the image, $\tilde{u}$, $\tilde{v}$ are undistorted linear affine coordinates, and p, q, r are linear homogenous coordinates of the pixel. The x, y, z represent Euclidean coordinates of the point in space that projects into the pixel $(u,\ v)$. Linear projection matrix M

[3]The mobile platform translating the rig is driven by a step motor.

is decomposed through QR decomposition

$$M \;=\; K\,R\,\left(\; I \mid -T \;\right) \tag{4}$$

into the product of orthonormal matrix R, representing the rotation of the camera with respect to the world coordinate system and an upper triangular matrix K containing the internal camera calibration parameters. T stands for the position of camera projection center F_C in the world coordinate system. Nonlinear distortion is assumed to be symmetric around the principal point of the camera and proportional to the squared distance from the principal point. The influence of nonlinear distortion is weighted by the constant k_1.

Notice, that the above nonlinear model is a good aproximation of real symmetrical nonlinear distortions only if coordinate axes, i.e. rows and columns of a CCD element, in image plane are almost orthogonal and the aspect ratio is close to one. More elaborate camera models are widely used in long-range photogrammetry as well as in precise vision instrumentation [3, 4]. Nevertheless, our eperience shows that the above model suffice for close-range fotogrammetry.

Camera calibration procedure is a variation of Tsai's bundle adjustment method [3] and can be summarized in two steps. First, nonlinear distortion is neglected and a linear system of equations is formed by combining (2) with (3), where a set of known points provide both 3-D and image coordinates. The estimate of the projection matrix M obtained by the linear algorithm provides the starting point for nonlinear optimization. The sum of squared distances between computed, equation (1), and real image coordinates is minimized using a general Loevenberg-Marquart minimization to provide an estimate of linear internal parameters as well as the measure of radial distortion, k_1.

4.2 Projector calibration

The projector is modeled by a plane which, in our setup, is forced to be perpendicular to the xz plane of the world coordinate syetem

$$n_x\; x + n_z\; z + d \;=\; 0. \tag{5}$$

The normal vector of the light plane is constant whereas d changes as the rig is translated, Figure 8. The nomal vector is calibrated by explicitly measuring its angle with x axis.

5 Triangulation

3-D coordinates of the points are computed by intersecting the camera ray corresponding to the vector $(u, v, 1)$, u, v being its pixel coordinates, with the light plane given by (5). The intersection is computed in the coordinate system of the camera where the formulas gain a particularly simple form.

5.1 Constructing the topology

In general, our ultimate goal is to measure the surface, not only points on it. The difference between a set of points and the surface passing through a set

of points consists in defining the topology, in this case the neighbourhoodship of the points. Our sensor is discrete by the construction[4], hence implicitly providing only points.

However, a surface can be constructed from points by assuming its local continuity. Since the part of the surface, visible from the projector, is homeomorphic to some rectangle $\mathcal{D}$ we can construct the neighbourhoodship of points in 3-D space by constructing the 4-neighbourhoodship along suitably chosen orthogonal parametrization of $\mathcal{D}$. As the camera-projector rig remains fixed during the translation, once selected rows in the retina uniquely determine a pencil of rays emanating from the projector, Figure 9. This pencil is translated thus forming its sweep plane, Figure 10. Then, very convenient parametrization of the surface can be forced through parametrizing fixed pencil of rays by their corresponding row coordinate u. The second surface parameter, t naturally arises from parametrizing the translation, see Figure 10. This parametrization

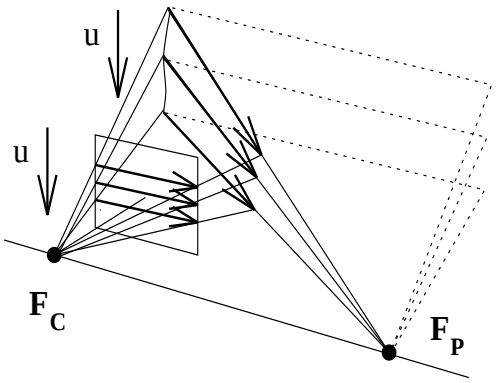

Figure 9: *The parametrization of the rays emanating from the projector.*

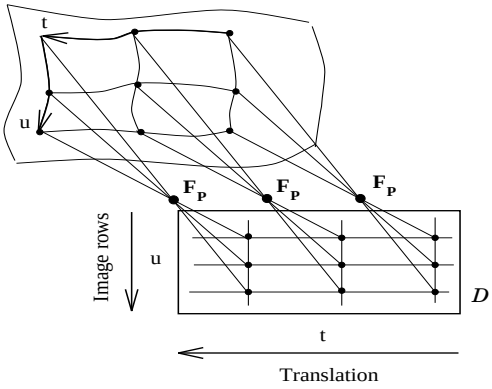

Figure 10: *Surface parametrization is composed from ray parametrization and the translation parametrization.*

allows easy construction of 4-connected mesh following parametric curves just by connecting points found in neghbouring rows of the image and points in neghbouring scans with the same image row coordinate.

The assumption of surface continuity is implemented as the restriction on the distance of neighbouring points. Only the neighbours closer than a predefined threshold are considered to lie on one surface one next to the other. The points with no close neighbours are assumed to be outliers and are removed from measured data.

[4]First, a CCD camera consists of discrete elements and second, the scanlines are taken in discrete steps.

6 Technical Data

Camera	BW CCD Panasonic CCIR, 500×582 pixels
Lens	Computar, effective f = 30 mm
Laser	Lasiris, 680 nm, compensated plane generator
Resolution x/y/z	0.5 mm / 0.29 mm / 0.5 mm
Spatial uncertainty	less than 0.5 mm

6.1 Format of 3-D data

3-D data are stored to an ASCII file arranged to lines delimited by CR-LF characters. Each line of the file contains data describing one point. Data are stored columnwise, i.e. one scanline follows the other. The format of each line, i.e. point, is the following:

x y z north nbr. west nbr. south nbr. east nbr. certainty

The first three items hold the coordinates, next four hold links to the four neighbours as the indices of the rows in the file, 0 stands for missing neighbour. The last item provides the measure of the certainty of point detection.

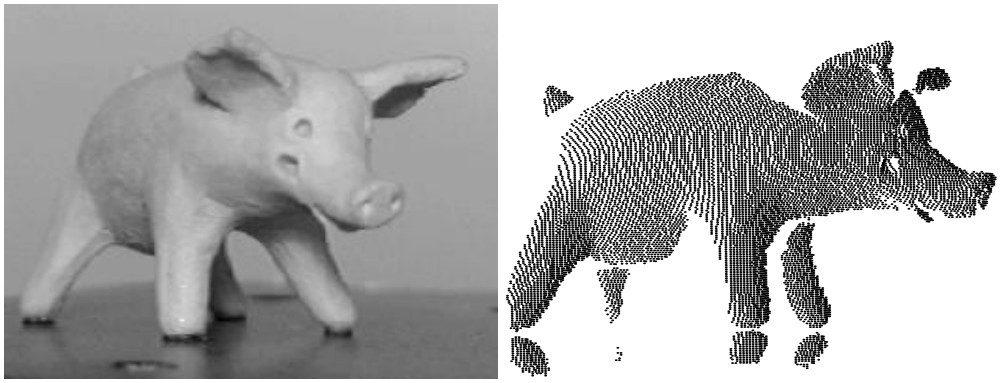

Figure 11: *Gray level image of a pig and its 3-D reconstructuion.*

7 Conclusion

The prototype of Laser Plane Range Finder provides data at medium resolution and precision when compared to best comercially available scanners for less than one tenths of their price. Howewer, a number of improvements are still thought and currently developed. Subtle technical details as well as the manual showing how to calibrate and use the system are described in [5] and [6].

References

[1] R. Jarvis. Range sensing for computer vision. In A.K. Jain and P.J. Flyn, editors, *Three Dimensional Object Recognition Systems*, volume 1 of *Ad-*

vances in Image Communication, pages 17–56. Elsvier Science Publishers B.V., 1993.

[2] P. Vuylsteke and A. Oosterlinck. Range image acquisition with single binary-encoded light pattern. *IEEE Pattern Analysis and Machine Intelligence*, 12(2):143–164, 1990.

[3] R.Y. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[4] Trond Melen. Geometrical modelling and calibration of video cameras for underwater navigation, trond.melen@itk.unit.no. Master's thesis, Norges Tekniske Høugskole Trodheim, November 1994.

[5] Tomáš Pajdla. Laser plane range finder. Computer Vision Laboratory, Prague, Memo CVL-LPRF-94-2, July 1994. pajdla@vision.felk.cvut.cz.

[6] Tomáš Pajdla. Camera calibration from know scene. Computer Vision Laboratory, Prague, Memo CVL-CAMCAL-94, November 1994. pajdla@vison.felk.cvut.cz.