

A Flexible Laser Range Sensor Based on Spatial-temporal Analysis

Jiang Yu ZHENG

Kyushu Institute of Technology
Iizuka, Fukuoka 820-8500, Japan
Zheng@mse.kyutech.ac.jp

Abstract

This paper introduces a smart laser range sensor with many advantages over the current products. The device is portable for its lightness and possessing simplest components. The image data is recorded in digital videotape for late processing, which allows it to be used in wide areas and places just as shooting pictures by a camera. The measuring scope can be freely adjusted to cover various distances, which greatly saves manpower in measuring immovable objects or a site. The system employs a post-calibration so that it can be used easily onsite. Further, we obtain color of measured points, which has been realized so far only by complicated systems in constrained environments. All these functions are realized by spatial temporal analysis of the video images. This sensor system has been used in measuring unearthed relics at a world heritage site.

1. Introduction

The recent progresses in multimedia require use of 3D range data. Although various laser range sensors have been put on market, they have many inconveniences in measuring large scale, multi-type objects in limited spaces and general environments. Such devices are usually precisely mounted and have fixed sensing scope. This prevents them to be used in an unstable out-lab environment such as excavation site, museum, photo shop, etc. A device is very desired not to involve any heavy calibration so that a non-specialist can handle it. The sensing environment should not be limited to a dark room, but anywhere with reasonable ambient light. The device should be light enough to be moved frequently from position to position for sensing a large number of objects.

In order to facilitate the use of laser range finder in different areas, a revolutionary upgrade is demanded. We build a laser range sensor that is probably the simplest one for its simple structure and components, as well as its easy utilization. Nevertheless, it has many advanced properties:

- (1) Camera direction and covering scope are adjustable.
- (2) Neither pre-calibration nor onsite calibration of the system is needed. A post-calibration is carried out after images have been taken.
- (3) It is portable for its lightness because it only consists of a finger-type laser projector and a small digital video camera that are driven on a rail [1-4].

- (4) It catches not only depth but also color of surface points [3,4].

We can start with a casual camera directing towards objects of interest and then do laser scanning and taking video. Since image processing will be done later, measuring a large site is just recording video sequences continually from position to position.

The principle used in the image processing is the spatial temporal analysis. Based on the vanishing point principle of linear motion, a simple post-calibration method is designed. Colors of measured 3D points are obtained by looking at later images in the sequence according to the image velocities of the points. Every scanning yields an image sequence, which further results several intermediate images, i.e., two dynamic projection images [5], a depth map and a color map that contain all the 3D and color information. This compact and standard data set is effective for multimedia communication.

2. A Smart Laser Range Sensor

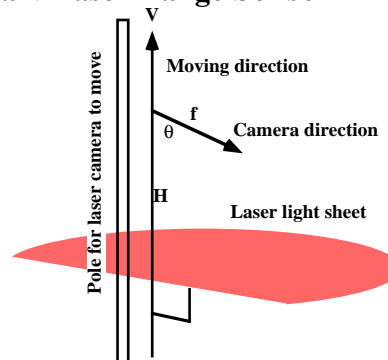


Fig. 1 Basic configuration of our laser range finder.

A laser-camera set is moved linearly in a selected constant speed (V in Fig. 1). A sheet of laser light is set orthogonal to the moving direction. The camera takes 30 images per second. The internal parameters of the camera such as focal length f , image center $o(x_0, y_0)$ and frame distortion are known through a calibration.

After properly setting the camera towards a distance where targets may locate, the pitch θ is fixed. The image coordinate system may have a roll α with respect to the plane containing the motion direction and the camera axis (Fig. 2). The pan of camera is not important since the laser sheet is wide enough to keep the same system geometry under a slight change of pan. The camera exposure is set onsite at a level where the laser stripe on

the surface will not be over-exposed and surface color is also taken in.

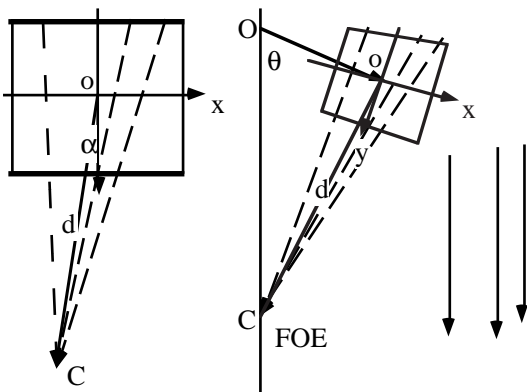


Fig. 2 The vanishing point in a linear motion.

If laser is on, there is a red stripe appearing on the object surface. It is taken into the camera. As the laser-camera set moves, the red stripe scans object surfaces. For each image coming in, our algorithm searches the laser stripe vertically, extracting edges and marking belts with a proper width between two edges. The hue and brightness are examined to verify the laser color. If several candidate belts are found, the best is always kept. After obtaining segments of laser stripes in an image, the algorithm extends segments from long ones to insure continuity and remove noise. The image position of the laser stripe in image t is registered as $L(x, y(x, t), t)$. For different t and x , we obtain a distribution $y(x, t)$ called *depth map*.

3. Spatial Temporal Analysis

As a camera moves along a line, all space points have a relative move opposite to the camera. It is well known that the projections of these parallel moves in the images approach to a *vanishing point* $C(x_c, y_c)$ (Fig. 2) or *FOE*. In most of our cases, the FOE is out of the camera frame. The approaching velocity of a point is constant for the fixed V and is dependent on the distance from the camera.

Further, we have a conclusion that the FOE on the image plane is also on the trace of camera focal point in the 3D space. In other word, FOE is also the projection of the motion vector of camera focal point on the image plane. This result can be easily proved.

If we pile the video images along the time axis, a spatial temporal volume is formed (Fig. 3). All points in the volume have their linear traces and finally reach a line that is parallel to time axis at the position of FOE. Most of the points will pass the top and bottom lines in the image (Fig. 2) if the camera moves over a long distance. In the spatial temporal volume, these points start from top plane and end at bottom plane.

In the spatial temporal volume, those red stripes form piece-wise smooth surface (Fig. 4) due to discontinuity in depth, occlusion by other surfaces and far background where the laser stripe exceeds the camera sight.

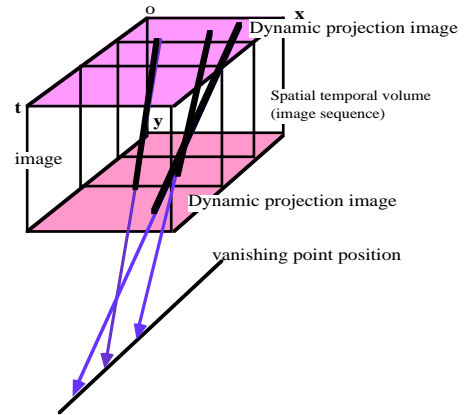


Fig3 Motion traces of points towards FOE.

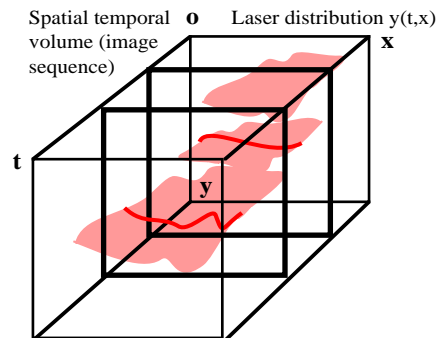


Fig. 4 Laser stripe distribution in the spatial temporal volume.

4. Post-calibration of the System

After a lot of video clips have been taken, we select those clips that the camera pose was first changed to compute θ and α , as well as distance H between the focus and laser plane. These system parameters will be used until the next clip where camera pose changes (Fig. 5).

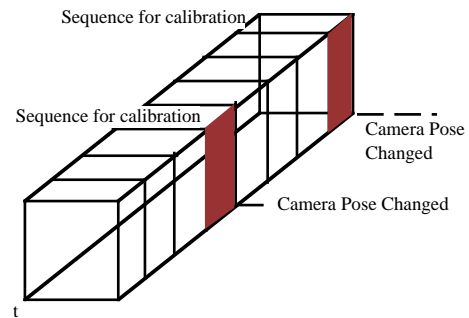


Fig. 5 Calibrating camera poses using video sequences.

For a video clip, we first compute the FOE from a set of motion vectors by locating distinguishable points. As described above, many image points in the top plane of the spatial temporal volume move to the bottom plane. If we collect both top and bottom plane images, we are able to find correspondences of those feature points, and further compute their image velocities. Suppose y_t and y_b are y

coordinates of the top and bottom pixel lines, the top and bottom planes are two *Dynamic Projection Images* obtained from these two lines [5]. In these two dynamic projection images, we manually find a number of correspondences (x_i, t_i) and (x'_i, t'_i) , $i=1, \dots, n$. The FOE is the intersection of all the vectors $(x'_i - x_i, y_b - y_t)$, where $y_b - y_t$ is the image size. Therefore, $C(x_c, y_c)$ should be on all these vectors and satisfy their linear equations.

$$\frac{size_y}{x_i - x_i} = \frac{y_c - y_i}{x_c - x_i}$$

We use the least squared error method to estimate (x_c, y_c) . After obtaining FOE, we find the distance and roll as

$$d = \sqrt{x_c^2 + y_c^2} \quad \alpha = \tan^{-1} \frac{y_c}{x_c}$$

We therefore obtain each point (x, y) in a rectified coordinate system as (x_r, y_r) by applying a 2D rotation using obtained angle α . The angle θ is then

$$\theta = \tan^{-1} \frac{d}{f}$$

In order to obtain H , we need to take the laser stripe into consideration. If we select a pair of corresponding points (x_i, y_i, t_i) and (x'_i, y_b, t'_i) in dynamic projection images, which are the projections of the same 3D point, we can fully determine the linear trace of that point in the spatial temporal volume. If that point is at very front of the surface, it should be scanned by laser at a known time. Since the pixel trace of that point in the volume will become red somewhere. We therefore search along the point trace and find its intersecting position with a red laser surface. We can compute H from it by using image velocity equation that will be given in the next section.

5. 3D Measure of Surface Points

As we obtain a laser stripe in the image, we register the positions (x, y, t) in another image called *depth map* ($y=y(x, t)$). The intensity of the image is the y coordinates of laser points.

When we compute 3D positions of laser lit points, we first calculate their position in the rectified coordinate system without roll. Suppose the obtained coordinate is (x_r, y_r, t) , the 3D position of the point becomes

$$X = \frac{Hx_r}{f \cos \theta + y_r \sin \theta} \quad Y = t \frac{\partial Y}{\partial t} \quad Z = H \frac{f \tan \theta - y_r}{f + y_r \tan \theta}$$

where $\partial Y / \partial t = |v|$.

6. Color Acquisition of 3D Points

Getting colors of surface points is necessary in many applications. Some laser range finder uses two cameras: one with a filter for laser detection and the other for color acquisition. It complicates the system. Some devices take images when laser is off, which requires synchronizing the laser and camera. There is another consideration to obtained color from projected white light, which may require a dark environment.

In our system, we get the exact colors of measured

points after those points move out of the laser stripe. As laser points are extracted, their 3D positions become known. Since the camera move is controlled at a constant speed, the image velocities can be determined as

$$\frac{\partial x_r}{\partial t} = - \frac{x_r (f \cos \theta + y_r \sin \theta)}{Hf} \cos \theta \frac{\partial Y}{\partial t}$$

$$\frac{\partial y_r}{\partial t} = \frac{f}{H} \frac{\partial Y}{\partial t} \left(\cos \theta + \frac{y_r}{f} \sin \theta \right) \left(\sin \theta - \frac{y_r}{f} \cos \theta \right)$$

Because the image velocity of a surface point in y direction only depends on its y coordinate when it is lit by laser, a lookup table is enough to get its value from laser position. For each point on the laser stripe, we compute its image velocity and predict where it moves in the later images. Waiting for a time delay Δt , we pick up color at the predicted position where the laser has passed away. If such a position is still lit by laser because of a possible rapid change in depth, we can extend the delay further. The color of every point is registered in another image called *color map* $T(x, t)$, at the same position as in *depth map*. These data will be used in building 3D surface model. For a 3D model, points neighboring in *depth map* are connected to compose triangular patches. The surface model is covered with all these small patches.

7. Experiments

We have measured twenty statues at an archaeology excavation site --- the Museum of Terra Cotta Warriors and Horses. The measure is done from different directions. Figure 6 is the dynamic projection images for post-calibration after the video sequence has been taken. Post-calibration is carried out for video sequences. Figure 7 gives a set of depth maps and Fig 8 is the measured 3D surface models.

The finest resolutions of the measured data are 0.3mm in Y direction, 0.5mm in X direction and 1mm in Z direction, covering a 3D space of 0.5m, 1m, 0.8m in X, Y, Z directions respectively. It can cover larger 3D region with a coarser resolution. Depending on a required resolution in Y direction (3mm~0.3mm), each scan takes 10~100 seconds.

In our analysis, we suppose two dynamic projection images are taken from top and bottom planes for the simplicity. In real case, there is no necessity to obey it. Directing the camera center at the average depth of objects to measure, laser stripes will frequently appear at the center part of images. Selecting two pixel lines away from center not only avoids red laser appearing in the dynamic projection images, but also increases the accuracy in the estimation of FOE. We can also select two pixel lines closer if the camera has a wide angle and objects are far away (object may not appear in both dynamic projection images due to limited length of linear move).

Our system use neither auxiliary equipment nor standard model in the calibration; the accuracy of a standard model itself is questionable sometimes. The only parameter our system relies on is the linear translation speed V guaranteed as accurate as 0.1mm at

the speed of 10mm/sec. The final position error is from θ and H , caused by picking up corresponding points at dynamic projection images, which might have several pixels uncertainty. With the robust estimation, these errors can be reduced to minimum.

8. Conclusion

We built a flexible laser range finder that is probably the simplest one so far, not only because of its handy components but also because of its convenience in setting, measuring and carrying. Spatial temporal analysis is extensively used in image processing to simplify calibration and obtain color, which were two problems of using laser range finder in wide areas and for different applications.



Fig. 6 Two dynamic projection images for post calibration

9. References

[1] J. Y. Zheng, Z. L. Zhang, Virtual Recovery of Excavated Relics, IEEE Computer Graphics and Application, May-June, pp. 6-11, 1999.
 [2] J. Y. Zheng, Z. L. Zhang, Digitizing and virtual recovering excavated relics in a world heritage site, International Conference on Virtual Systems and Multimedia 98, Vol. 2, pp. 654-660, 1998.
 [3] J. Y. Zheng, Z. L. Zhang, Digital Archiving of an Archaeological Excavation Site for Multimedia Display, 14th International Conference on Pattern Recognition, Vol. 2, 1492-1496, 1998.
 [4] J. Y. Zheng, Z. L. Zhang, N. Abe, Virtual Recovery of Excavated Archaeological Finds, IEEE Multimedia Systems Conference 98, pp. 348-357, 1998.
 [5] J. Y. Zheng, S. Tsuji, Generating Dynamic Projection Images for Scene Representation and Understanding, Computer Vision and Image Understanding, Academic Press,

Vol. 72, No. 3, December, pp. 237-256, 1998.
 [6] P. Besl, Active, optical imaging sensors, Machine Vision and Application, 1988, pp. 127-152.
 [7] J. Clark, E. Trucco, H-F. Cheung, Improving laser triangulation sensors using polarization. ICCV95, pp. 981-985, 1995.



Fig. 7 Depth maps of a measured statue from different directions.



Fig. 8 Measured 3D surface model of the statue in Fig. 7.