A SIMPLE METHOD FOR 3D SHAPE RECONSTRUCTION
Tomislav Pribanić, Mario Cifrek, Stanko Tonković
Faculty of Electrical Engineering and Computing
Unska 3, HR-10000 Zagreb
Croatia

Abstract
To infer 3D descriptors of the scene one can use image(s) captured from camera(s). With respect to the illumination of the scene there are passive and active approaches. The later ones are particularly suitable for objects with no texture. The paper represents 3D reconstruction system with commercially available pair of cameras and video projector, as source of active illumination. Furthermore, algorithms involved are simple and easily implemented. Practical reconstruction results along with systems strengths and weaknesses are presented as well.

Key Words: 3D Reconstruction, Lasers, Epipolar Geometry, Medical Imaging

1. Introduction
A 3D reconstruction of point in space is a task present in variety of areas and applications. Different 3D reconstruction system/methods are characterized by accuracy, speed, hardware/software implementations, cost etc. List of priorities that particular method has to meet is determined by the specific application. If we look closely at sport and medicine, then a condition of method being non-invasive is at the top of the list. A very convenient way to fulfill this condition is to extract three-dimensional information from two-dimensional image coordinates acquired by (video) cameras. Systems based on that principles have its origins in photogrammetry [1] and are usually referred as 3D kinematic systems [2]. Furthermore such a system may relay on retroreflective markers (accompanied with infrared cameras) attached to a subject. Immediate system outputs are kinematic parameters of attached markers: position, speed and acceleration. One of the main advantages of 3D kinematic systems for the purpose of biomechanical analysis is that acquired kinematic parameters combined with anthropometrical subject data can yield additional kinetic parameters by the means of inverse dynamic calculation [3]. On the other hand, the drawback of these systems is that we are actually reconstructing only the rather limited number of subject points in space, i.e. attached markers, and not the entire subject shape which is within the cameras field of view. Still there are many applications, even outside sport and medicine area, which demand full 3D shape reconstruction [4]. Reconstruction of (almost) every pixel in the cameras image has always been quite challenging compared to reconstruction of ‘only’ small number of retro reflective marker points. Nevertheless, nowadays the so-called 3D laser scanners routinely carry out 3D reconstruction of partial/full body shape. Besides being rather accurate their major obstacle to wider use is large cost [5]. However, laser technology is not the only way to obtain dense points reconstruction. Alternative approaches may employ different lightening/shadow technique [4]. Although such non-laser scanners may be inferior to laser ones in terms of accuracy they are still very much satisfactory in many cases. The purpose of this paper was to construct cheap and simple, yet effective 3D body shape reconstruction system. In the following text principles of the system are given and experimental results are shown along with major weaknesses and possible future improvements.

2. Materials and Methods

Figure 1. Reconstruction system setup
The idea is to place the subject in dimmed (preferably dark) environment and to use projector in order to scan a white line across the subject (Figure 1). On images captured by video cameras light-dark edge of projected white line is detected within each frame. Finally, 3D positions of detected edge pixels are calculated. The presented system for a 3D shape reconstruction consists of several straightforward steps which can be summarized as follows:

- **Video camera calibration**: calibration was performed in traditional manner using set of calibration points from calibration cage. 50 calibration points were equally spaced throughout calibration volume of size 30 cm \( \times \) 40 cm \( \times \) 50 cm (Width \( \times \) Height \( \times \) Depth).
- **Scanning the subject**: subject of interest (i.e. vase) was placed inside the calibration volume and scanned via white vertical line in dimmed ambient light (Figure 2).

![Figure 2. A projected line on the vase in dimmed ambient light](image)

- **Recording the scanned subject**: simultaneous recording by pair of video cameras of projected line took place (Figure 1).
- **Saving the recordings on PC**: video frame grabber captured the VHS recordings from cameras into avi file and individual images were extracted from saved avi files.
- **Extracting the image portion of interest**: prior to further image processing user is given a chance to crop the image part where the subject was actually positioned. This step can be performed in matter of seconds and although it is not absolutely necessary it is still highly recommended since it can speed up the image processing time. Apart from recording procedure that is in fact the only user intervention involvement. The rest of the processing is fully automatic.
- **Edge detection**: on every image the dark-light edge is detected (Figure 3)
- **Corresponding image points search**: one video camera is taken as referent and with respect to it fundamental matrix [6] is calculated. Within each frame, for each detected edge point on the referent camera and fundamental matrix, epipolar line [7] on the second camera image is found. Intersection of the epipolar line with the detected edge on the second camera image yield corresponding image point of second camera (Figure 3, Figure 4).

- **3D shape reconstruction**: calibrated cameras and found pairs of corresponding image points put together gave 3D positions of image points. In fact, acquired 3D position of corresponding points served as input for mesh drawing procedure which resulted in final output as 3D subject shape.

![Figure 3. Examples of epipolar lines intersecting detected dark-light edge (white stars)](image)

![Figure 4. Corresponding image points search: note intersection of epipolar line and interpolated edge](image)

3. Results and Discussion

Beside vase shape reconstruction (Figure 7) shape of the footprint left into sponge was acquired in similar manner (Figure 8). The camera model used here for
calibration and reconstruction is effectively the same as those used in photogrammetry for a long time [8]. Throughout the time many variations of it have been proposed [9], however all of them are based on the same pinhole model augmented with additional parameters to compensate linear and non-linear distortion [10]. Following the principle 'as simple as possible' chosen camera model was linear in its nature meaning that parameters for non-linear distortion (mostly due to lens imperfection [11]) were left out. Encouragement to use linear camera model came largely from two reasons. Firstly, as it will be seen, other algorithms and methods in the rest of the reconstruction and process were simple as well. Secondly, reconstruction accuracy analyses of set of 'unknown points' [12] showed very good results. Beside calibration points additional 150 points served as 'unknown' points which 3D positions were found and root means square values (RMS) was calculated between its true and calculated positions [13]. Obtained RMS values for X (width), Y (height) and Z (depth) coordinate direction were 0.407 mm, 0.372 mm and 1.044 mm respectively.

Figure 5. Footprint in the sponge: Regions excluded from further processing due to light occlusion

For a successful automatic image processing during which edge detection takes place ambient light conditions are critical. Ideal situation would be scanning the subject in completely dark room. Even than it is possible that some areas may appear darker (occluded) than one would expect when the white line (light) scans across them (Figure 5). That is due to inherent physical characteristic of the subject being scanned and its position with respect to projector (source of light) These portions of images (sets of pixels) are considered as impossible to reconstruct since edge detection of scanned line over such areas is considerably harder at least in the sense of full automatic detection. Therefore such pixels ‘immune’ to scanning are excluded from further processing after they are identified as follows. For a certain pixel on the image gray level values were considered throughout the frames. Typical changes of gray level values vs. frames for a pixel included and excluded from edge detection are given in Figure 6. It is evident that for pixel being excluded change in gray level intensity (+ dotted line in Figure 6) is negligible in comparison to pixel being included in processing (solid line in Figure 6). For a pixel to be considered during edge detection difference between minimum and maximum gray level intensity values has to be above some threshold.

Figure 6. Gray level values vs. frames: line with + markers case of pixel being excluded from processing

VHS cameras used for recording have had a low end commercial characteristics. Involvement of any quality digital video cameras would ultimately give better images and speed up the frame grabbing procedure on PC.

Edge is by the definition place where abrupt change in gray level intensity occurs [14]. There are many edge detectors, which differ mostly in how many pixels in the neighborhood they consider in order to calculate gradient of gray level change [15]. Here, an even simpler method was applied to detect the so-called smooth step edge [16]. Typical transition of gray level intensity for certain set of pixels in horizontal direction is shown in Figure 6. As it can be seen small changes of gray level values are coming from noise and in accordance with normal practice in image processing smoothing was applied [17] via moving average filter. After smoothing, the minimum, maximum and average values were calculated for that set of pixels. Finally pixel having the closest value to the average one is labeled as the edge pixel. Same procedure is repeated along the vertical image direction to find all edge pixels on the image of particular frame (Figure 3).
Fundamental matrix is a major algebraic representative of epipolar geometry [18]. As such it is a very powerful tool in algorithms for image points correspondence search. The basic expression involving fundamental matrix states that when some point from 3D space projects in the image planes of two cameras then knowing the image coordinates in the first camera and fundamental matrix we are able to calculate epipolar line in the second camera image. Coordinates of projected point on the second camera lays somewhere on the epipolar line [6]. Since we also recognize that our image points lay on the detected edge, for an edge point in the first image its correspondent is found as intersection of calculated epipolar line and edge detected on the second image. To improve accuracy of correspondence, detected edge on the second image was additionally linearly interpolated (Figure 4).

Having the set of corresponding image points and pair of calibrated cameras it was a simple matter to calculate 3D positions of it. As a first experiment a vase was scanned (Figure 2) and its shape reconstructed. A vase is quite attractive for reconstruction since it does not have abrupt surface changes and as such is good detector for a smooth reconstruction. There were more than 60000 detected edge points on each of the camera images through 420 frames recorded. At the end the correspondence algorithm passed about 40000 points mainly because points (edges) were not visible on both cameras. Such a dense cloud of points was to a certain extent subject to many sources of errors and noise during the different phases of reconstruction. Therefore a surface mesh involving smoothing and averaging with closest neighborhood points gave final output (Figure 7).

When one is about to have his shoe insole made common practice is to step into sponge and leave his footprint, which serves (along with some other information) as guide in making shoe insole. Therefore, reconstructing the footprint is effectively very much reconstruction of the foot itself. This footprint reconstruction in sponge was somewhat harder primarily because of two things. First, the shape of the footprint is such that some portions (i.e. pixels on images) of it will remain in occluded regions where scanning light cannot fully reach (Figure 5). Secondly, with only two cameras the large common filed of view is harder to achieve. In this case were also a large number of reconstructed 3D points smoothed out by surface mesh reconstruction (Figure 8).

4. Conclusion

A rather simple reasoning and involvement of very basic hardware can produce quite well 3D shape reconstruction system. Particularly if we consider performance/price ratio than presented system has to come forward. In every step of calculation were intentionally used only the minimum requirements and basic algorithms/methods, i.e. two low end VHS video cameras, linear camera model, 8-point algorithm for calculation of fundamental matrix [19], basic edge detection technique etc. This leaves us with considerable room for improvement and then studying its effect. It can be expected that any changes to more sophisticated calculation/hardware would primarily come into effect when going to larger volumes, such as full human body shape. This is exactly where future work might be directed after the testing phase in every day clinical practice of the present system satisfactory ends.
Figure 7. Different views of vase reconstruction

Figure 8. Example of insole (footprint) shape reconstruction
References