Spatial Data Acquisition from Motion Video

Mark Williams

Departments of Computer and Information Science

University of Otago, PO Box 56, Dunedin, New Zealand

Phone: +64 3 479 8488

mrwilliams@otago.ac.nz

ABSTRACT

Geographic information systems are an important tool for the field of geocomputing. A key component of every system is the data—spatial data has traditionally been labour-intensive to collect, and hence expensive. This paper establishes a new method of acquiring spatial data from motion video. The proposed method is based upon the principles of photogrammetry, but allows position to be calculated with feature tracking rather than point correspondence. By doing so, it avoids many constraints imposed by previous solutions. The new method is demonstrated with linear and rotational motion.

INTRODUCTION

The field of geocomputing utilises computer models of the real world. While in some cases there is no spatial correspondence between the computer model and reality, many cases require an accurate representation of the real world. For example, geographical information systems (GIS) commonly require accurate, scaled maps of land. A simple GIS may incorporate a flat two dimensional map from which the computer could, for example, measure distance or calculate areas. A more complex version may include a topographic map, which is essentially a layered set of two dimensional maps at fixed height intervals. From this information the GIS can calculate volumes of hills, or viewshed images that determine the visibility of, for instance, a large structure on a hilltop. More detailed GIS models require true three dimensional models of objects, such as tunnels, to measure volume. In all cases a GIS is only as good as the source data collection, hence the data collection technique needs to be both accurate and practical for the size of the object being modelled. In the case of traditional GIS, photogrammetry is well established as a data gathering method. This large-scale technique has also been proven successful for some small-scale tasks. For example Fraser (1988) demonstrated a semi-automated method for measuring ships' hulls and satellite dishes. On an even smaller scale, Rivett (1983) showed photogrammetry has been used for measuring human faces to evaluate post-operative swelling.

The quintessential problem of photogrammetry is to establish correspondence, which involves identifying the points in each image that represent a given point in space. This has proven to be an easy task for a trained human operator, but very difficult for a computer. Solutions often require that easily identifiable reference objects be placed within the scene. They may also require controlled lighting and manual identification of corresponding target points. There is, however, a class of problems where alteration of the scene is undesirable, making automated measurement very difficult.

A new method is proposed that aims at building a three dimensional model from easily obtained visual data. Currently the requirements of this method are normal lighting and controlled, functional camera motion.

This paper reviews methods of capturing three dimensional spatial data from two dimensional images. A new approach is proposed that is based upon the principles of photogrammetry, while circumventing the correspondence problem. The mathematical foundations of this method are described, then several worked examples are given to demonstrate this technique.

EXISTING METHODS OF SPATIAL DATA CAPTURE

The most common method for obtaining data for a GIS is photogrammetry. In the traditional case of obtaining land data, stereo pairs of aerial photographs are taken; smaller scale data capture may use terrestrial photogrammetry. Both cases process the stereo images to calculate depth of objects within the scene.

Stereo Geometry

Figure 1 shows the basic form of stereo geometry: consider this as a plan view of two cameras capturing a scene. The cameras have focal length f and are separated by the stereo baseline b. For simplification, the cameras are shown with parallel viewing directions, and subsequently parallel image planes. This is not essential, but if the cameras are not parallel a transformation is required to align the coordinate systems.

Two points A and B in the scene appear at A_l and B_l in the left image and A_r and B_r in the right image, and are at distances z_A and z_B from the cameras. Because of the different viewpoints of the cameras, the points appear in different places within each image—this difference is known as *disparity* or *parallax*. The parallax of each point is calculated from the x-coordinate of the image points:

$$p_A = A_l - A_r \qquad\qquad p_B = B_l - B_r$$

Using similar triangles, the depths of points A and B can be calculated:

$$z_A = \frac{fb}{p_A} \qquad\qquad z_B = \frac{fb}{p_B}$$

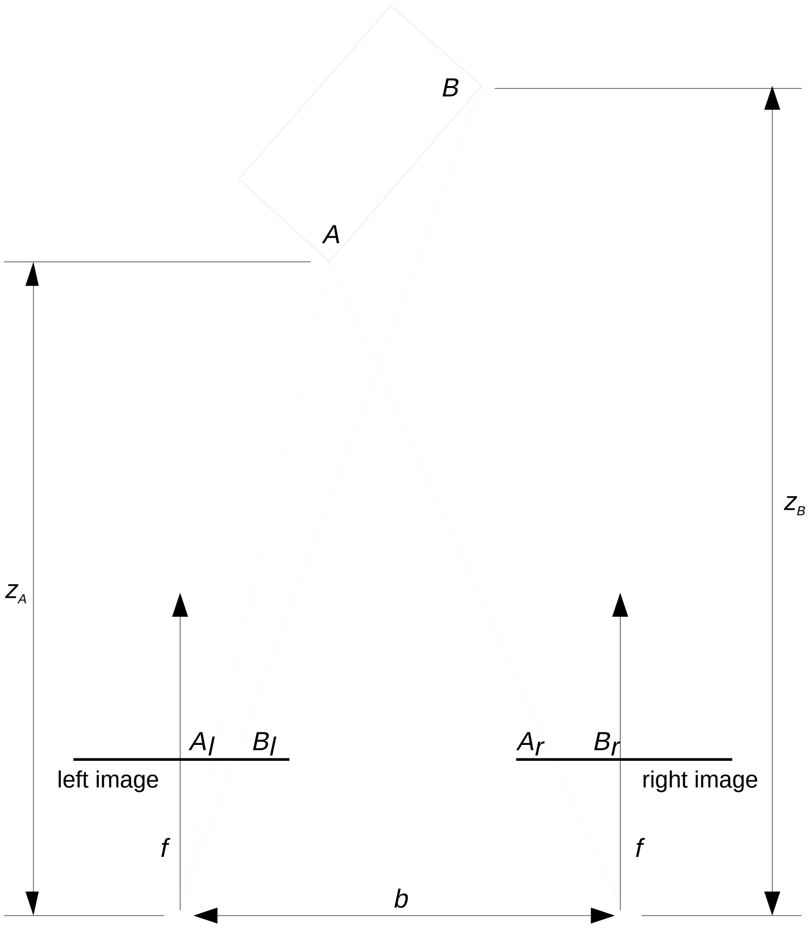


Figure 1: Geometry of stereo vision

These equations demonstrate the requirements of a depth-from-stereo algorithm: given a known baseline and focal length, align the images to reconstruct the relative orientation at the time they were captured, then measure the disparity of each matching pair of points to determine depth. While this process appears very simple, there is one fundamental problem: determining which point in the right image corresponds to a given point in the left image. Thus the key issue for calculating depth from stereo images is the *correspondence problem*. An important technique used by many researchers to assist solve the correspondence problem is the *epipolar constraint*. Figure 2 demonstrates the definition of the epipolar plane. For any point P in the scene, an epipolar plane is defined by the point and its representation in the left and right images. The epipolar plane intersects the images forming an epipolar line in each. When the two images are correctly aligned, these epipolar lines will be horizontal. Having located a point in one image, it is only necessary to search along the epipolar line to find the corresponding point in the second image. Hence the epipolar constraint reduces the correspondence problem to a one dimensional search of a scan line for the matching point.

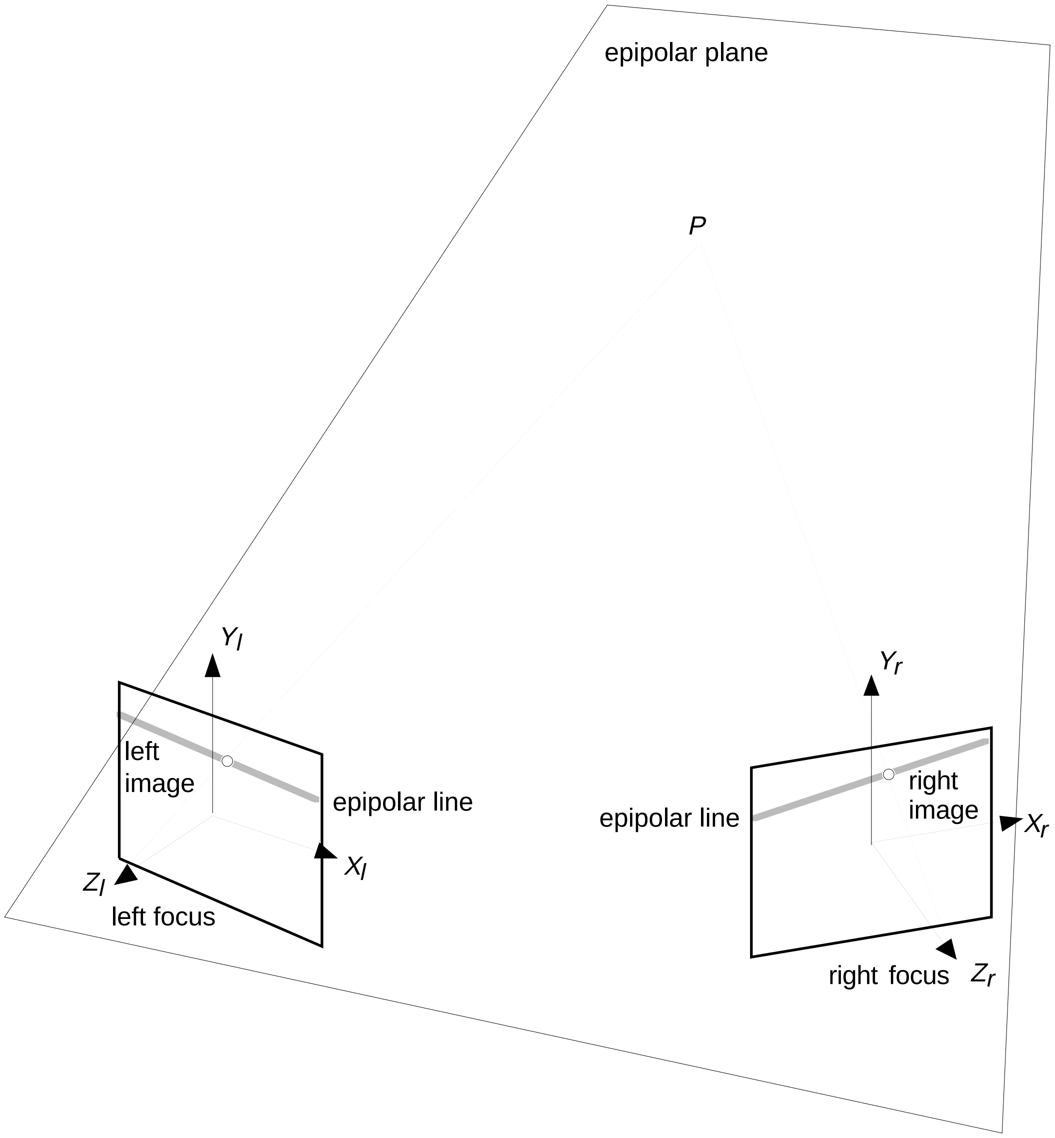


Figure 2: Epipolar Geometry for Stereoscopic Viewing

The epipolar plane is defined by a point *P* in space and the representation of that point in each image. The plane forms epipolar lines where it passes through each image. Having found a feature point in one image, it is only necessary to search for a matching point along the epipolar line in the other image.

Photogrammetry

Automatic photogrammetry systems have taken many different approaches to solving the correspondence problem. These approaches can be summarised in three categories: interactive methods which require a human operator to identify a match, target based systems that attach or project easily identifiable targets onto the object to simplify matching, and fully automatic systems that attempt to identify natural features within the scene to match.

Interactive systems, such as those demonstrated by Erez and Dorrer (1984) and Beyer (1992), are the most simple solution. An operator looks at a stereo pair of digital images and identifies matching feature points in each. The system then refines the location of each feature, for example by performing a least squares match with the intensity value of the pixels. The sub-pixel coordinates achieved by the refinement are then used to obtain an accurate measure of the parallax for that feature.

Target based systems aim to remove the need for operator assistance. This requires making the matching task as simple as possible. A common technique, as demonstrated by Fraser and Shortis (1994), Clarke *et al.* (1995) and Peipe and Schneider (1995), utilises retro-reflective targets and controlled lighting to ensure the targets are the only objects visible in the images. Using the epipolar constraint, matching targets in the stereo image pair is greatly simplified. As with the interactive case, the location of each target is refined by a least squares method using pixel intensities.

The most ambitious method of stereo matching attempts to identify the same point in each image without targets or controlled lighting. Rosenholm (1987a/b) attempted to match the intensity values of a patch or *window* of pixels in each image. Even using the epipolar constraint to reduce the search space to a single dimension, pixel based matching is extremely slow. Kölbl *et al.* (1991) addressed this issue by developing specialised hardware to perform the search in parallel. Koch (1992) reduced the search space by using a more complex evaluation criteria: instead of matching just the intensity of pixels, a gradient vector and match confidence is calculated for a pixel window, giving a more distinctive description to match.

While many photogrammetric systems have achieved satisfactory results for particular domains, the correspondence problem forces those systems to compromise generalisation for automation. Traditional stereo geometry requires correspondence to be solved, and that is the limiting factor in current techniques.

Computational Stereo

The field of computational stereo can be divided into three distinct areas: the human visual system, robot vision and image metrology. While all three areas share the fundamental framework of stereo vision, each has a unique goal, and the methods employed to reach those goals are very different. In the study of the human visual system, computer scientists attempt to model parts of the system, and use the computer models to help gain a better understanding of how human vision works. The emphasis is placed on biological and psychophysical accuracy, in order to ensure results remain relevant to the human model. Meanwhile robot vision is aimed at achieving the more practical goal of navigation and simple object identification. In the case of navigation, the important result is to avoid obstacles while moving in a “real world” environment. Industrial robots are also required to inspect objects, identify and sort them into appropriate groups. Both these tasks require some calculation of three dimensional position, but more emphasis is placed on recognising structure and shape. The final area of computational stereo, image metrology, closely parallels photogrammetry. The goal is to extract measurement of an object, by determining the three-dimensional coordinates of a series of points on its surface.

Barnard [Barn82] claims that the process of automatically matching features between a stereo pair of images is “the hardest and most significant problem in computational stereo”. This is certainly the case in image metrology research, and the variety of methods used fall into two broad categories: pixel- and feature-based matching. The former method attempts to match individual pixels or small clusters of pixels all over the images, creating a dense depth map, while the latter uses image processing techniques to identify sparse significant features to match between the images.

Pixel-based methods, such as those proposed by Cappellini *et al.* (1987), Trivedi (1985) and Zheng *et al.* (1990), rely on comparing the intensity value of individual pixels or small clusters of pixels. Arnold and Binford (1980) noted that pixel-based intensity matching was prone to variation from camera settings, lighting and so on. Instead, they proposed a simple feature-based method, identifying edges, which was more robust under the same conditions. Features are extracted by filter functions and are often edge pixels, line segments, curves or junctions. Marr and Poggio (1982) proposed the Laplacian of Gaussian (LoG) filter, which had a very similar effect to its biological equivalent in the human eye. The LoG filter detects edges by first smoothing in a circular, Gaussian fashion, then taking the second derivative of the smoothed intensity.

To reduce the extracted feature set and improve the uniqueness of each feature, the images can be filtered and features matched hierarchically. For example, Jin and Li (1988) proposed an edge-based hierarchical matching scheme. Edges are extracted by a selectable-resolution filter function. Matching starts with low resolution, gross features, and is continually refined to higher resolutions, using the previous resolution’s results to segment the image and provide depth estimates.

There are clearly more similarities than differences between the fields of photogrammetry and computational stereo, particularly when considering metric photogrammetry and image metrology. One notable distinction between the fields is the use of targets to identify feature points. This is common practise in industrial photogrammetry, and is often combined with controlled lighting to ensure the targets are easily identifiable in the images. Conversely computational stereo approaches aim for generalised solutions, extracting natural feature points or matching clusters of pixels. Accuracy is seldom quantified, while automation is a key goal.

Shape from Motion

Another popular method for obtaining three dimensional data from visual images is *shape from motion*. In such methods, two or more images are captured as the camera—or the scene—moves. The images are captured in rapid succession, so features within the scene will only move a few pixels from one image to the next. Basic principles of perspective dictate that features close to the camera will move differently from those further away. The essential problem is to determine optical flow—the differing motion of various regions within the image sequence. Once motion vectors have been calculated, they are used to determine the slope of surfaces in the scene.

Many techniques have been proposed to calculate optical flow, for instance differential methods, region matching, energy–based and phase–based techniques. A detailed description and comparison of examples of these methods can be found in a thorough review by Barron *et al.* (1992). All of these methods contain the same broad stages of processing. Generally the first step involves filtering or smoothing the images, followed by the extraction of basic measurements of motion. These measurements are then integrated into an optical flow field, which itself is smoothed before being used to determine the surface gradients within the scene.

Monocular Methods

While photogrammetry and computational stereo use stereo pairs of widely spaced images, and motion techniques use streams of closely spaced images, another class of methods for calculating depth use single images. *Shape from shading* analyses the intensity variations within an image to determine the shape of surfaces. The simplest form of shape from shading uses a single image, and makes assumptions about the nature and direction of the light source, as well as the albedo or reflectance properties of the object itself. The lighting graduations on the surface are then used to determine the curvature, and consequently the relative depth can be calculated. Bruckstein (1988) provides an excellent overview of the problem, and a simple recursive procedure to determine equal-height contours of a surface given a tightly constrained experiment.

Due to the number of assumptions required, single image shape from shading techniques are numerically unstable and their results may not be reliable. So the method was extended to *photometric stereo* where images are captured from the same viewing position under different lighting conditions. Lee and Kuo (1992) describe two methods for combining the multiple images: parallel and cascade schemes. The parallel scheme uses all images at the same time, formulating the intensities into a linear system of equations. The solution of these equations gives a height field for the scene. By contrast, the cascade scheme uses a single image shape from shading method for each image sequentially, where the result from one image is used as a starting condition to constrain the next.

Another approach uses photometric stereo with structured lighting. For example, instead of a conventional light source, Klette *et al.* (1995) use a single stripe of laser light projected into the scene. When the stripe of light hits an object it forms a contour line on the surface, which can be easily extracted from an image. By applying this technique in many places over the surface of the object, a very accurate model can be made.

Shape from contour uses a technique quite similar to the structured lighting method described above. Instead of using a laser stripe to illuminate a contour line, the silhouette of an object is extracted from many images captured as it rotates in front of a camera. The series of silhouettes can then be assembled and analysed to determine the object's shape. Unfortunately this algorithm fails with all but the simplest objects, because concave areas of a surface can be occluded in all of the silhouettes.

A more robust approach is described by Zheng (1994), which identifies an area of concavity as a discontinuity in the motion of a contour. The details of Zheng's method are beyond the scope of this paper, but one important aspect is the detection of occluded regions in an epipolar plane image, which is formed by extracting a single selected line from many images, and stacking them all together. Such a process is similar to that used, albeit for a different purpose, by the method proposed in this paper. Other shape from contour methods extract structure from a series of contours by recognising junctions or generalised cylinder, Ulupinar and Nevatia (1992/95).

Another type of single image method determines shape by recognising scene structure. These model based approaches, reviewed by Chin (1986), while particularly important in industrial applications, are not related to the method proposed in this paper, and are mentioned only for completeness.

All of the methods described above have proven to be successful in particular cases. All have constraints, however, including adding reference to the scene, controlling the lighting or making assumptions about the reflective properties of the surface.

SPATIAL DATA CAPTURE FROM FUNCTIONAL MOTION

An alternative approach for calculating depth from images is based upon the principles of photogrammetry, while circumventing the correspondence problem. In doing so many restrictions of other methods are avoided: no reference points are required, nor special markers on the object. Where photogrammetry requires a widely spaced pair of images, the new method uses a dense image set. The motion of the camera relative to the scene is determined by a simple function. Data from the camera is treated as a three dimensional block, consisting of X and Y spatial dimensions, and a temporal dimension. Figure 3 illustrates this *spatiotemporal* volume. Interesting features are tracked as they move within this space; the function of their motion determines the three dimensional location of each feature in space.

Bolles, Baker and Marimont [Boll86 and 87] first proposed slicing through the XT plane of a spatiotemporal volume to form an *epipolar plane image* (EPI). They noted the relationship between the path of a feature within the spatiotemporal volume, and the three dimensional coordinates of that feature in the scene. The aim of their research was a robot vision system, where a measurement of free-space within a scene was more important than measuring individual objects. Hence the purpose of their EPI analysis was to calculate depth and determine occlusions, which proved to be successful. However, they did not establish the relationship between the EPI analysis and the process of traditional photogrammetry.

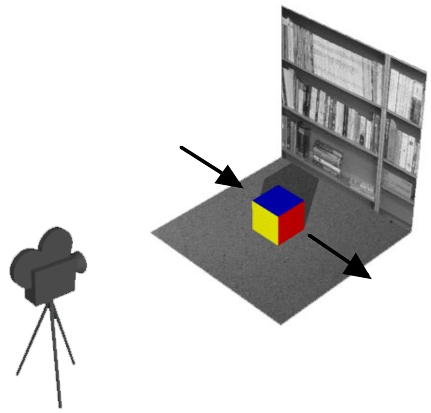
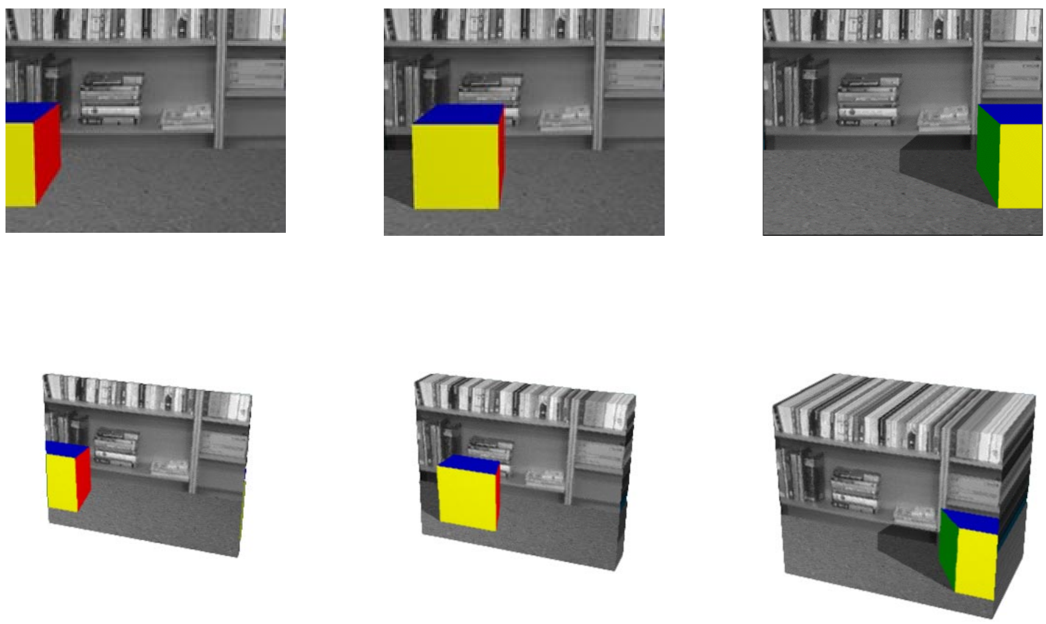


Figure 3: Spatiotemporal Volume

Left: Camera films the scene as block moves from left to right.

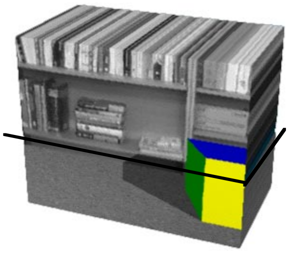
Below: Three frames from the image sequence.

Bottom: Frames are stacked to form XYT volume.

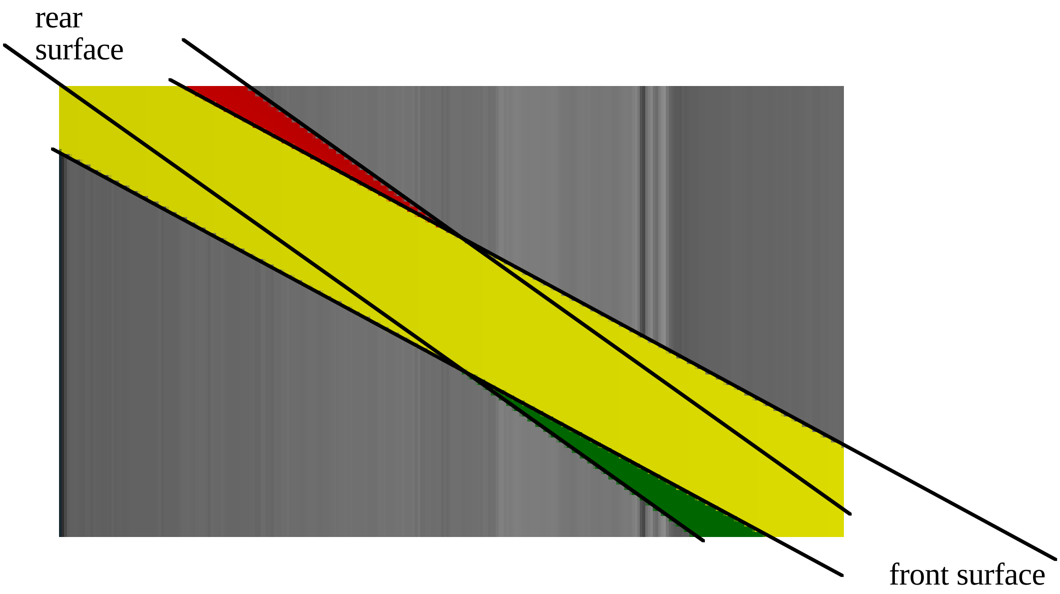


Since the relative motion between the camera and the scene is known, the general form of the motion of points within the spatiotemporal volume can be determined. Furthermore, the precise motion of every feature point is a parametrised form of the general case, where the parameters relate directly to the three dimensional position of each point in the scene. Hence the mapping from the given XYT coordinates to the desired XYZ coordinates becomes a task of fitting data to a function by adjusting these parameters.

Figure 4: Epipolar Plane Image



This is an epipolar plane image taken from the spatiotemporal volume in Figure 3. Time is shown on the vertical axis. The front and side surfaces of the cube have formed bands through the EPI, where the gradient of each edge is proportional to its position in XYZ space. The front edges are parallel to each other and at a different angle to the rear edges.



Linear Motion Example

The simplest case of the proposed method is linear motion. This will arise, for example, when the camera is stationary while an object moves past it. As the object moves, frames from the camera are captured to form a dense spatiotemporal volume. This volume can be segmented into XT slices: each image plots the paths of points on the object at one height, or Y value, in the XY frames. By considering the apparent motion of individual points over time, it can be observed that points at different distances from the camera move at different apparent velocities. All points move in straight lines, but those closer to the camera appear to move faster than those farther away. Imagine looking out the window of a moving train: trees close to the railway line appear to rush past in a flash, while mountains in the distance appear to move very slowly. When viewing this motion as an epipolar image, objects at different distances from the camera appear as lines of different angles. Measuring the angle between the lines allows a direct calculation of the difference in distance from the camera, and hence the size of the object.

For example, Figure 4 shows an EPI from the spatiotemporal volume shown in Figure 3. Note in this case the object moved through the scene while the camera stayed still. Hence the stationary background forms vertical stripes, while the object itself forms diagonal bands. From the epipolar plane image, it is easy to see the relationship between the path of a point in XYT and its location in XYZ. In this case, the function is a straight line, where the gradient is proportional to the depth of the point from the camera.

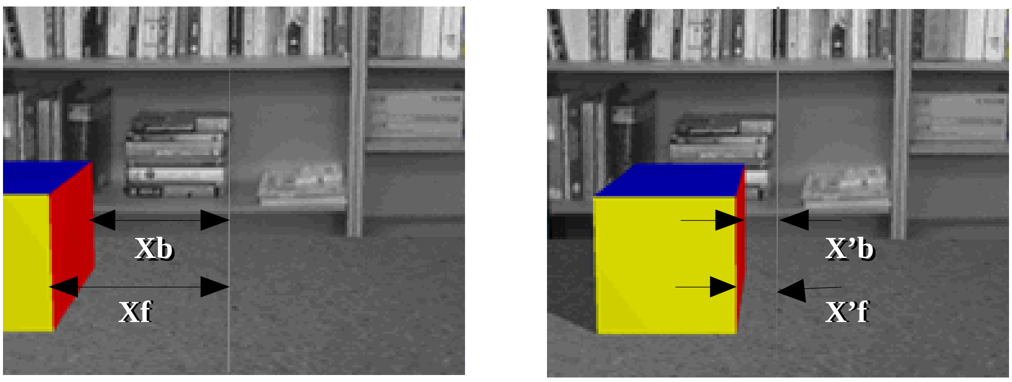


Figure 5: Parallax measurement in Photogrammetry
The distance is measured from the centre-line to the point of interest in each image. Parallax is then calculated as the measurement difference between one image and the other.

Now consider this case in traditional photogrammetric terms. Instead of a continual sequence of images, a stereo pair is selected (Figure 5). To calculate the depth of the cube, a parallax measurement is required for the front and rear surfaces.

$$\begin{aligned} Pf &= Xf - X'f \\ Pb &= Xb - X'b \end{aligned} \tag{1}$$

Given the camera baseline and focal length, depth can be calculated:

$$D = Bf\left(\frac{1}{Pf} - \frac{1}{Pb}\right) \tag{2}$$

Figure 6 demonstrates how the stereo pair images form two lines of the EPI, allowing the parallax measurement to be made directly. The ratio of spatial to temporal resolution in the XYT volume is 1:1, so the camera baseline is equivalent to the vertical distance between the stereo pair slice in the EPI. Hence equation (2) can be expressed as:

$$D = f\left(\tan\theta_f - \tan\theta_b\right) \tag{3}$$

For the linear motion case then, the proposed method can be expressed in terms of traditional photogrammetric approaches.

Rotational Motion Example

This method extends to rotational motion. As an object rotates before a camera, or the camera rotates around the object, individual points on its surface form sinusoidal paths in an epipolar image. Consider watching a glass of water placed at the edge of the turntable inside a microwave oven. The glass appears to move back and forth periodically: plotting this motion against time produces a sinusoidal curve. Now consider a second glass in the microwave, closer to the centre

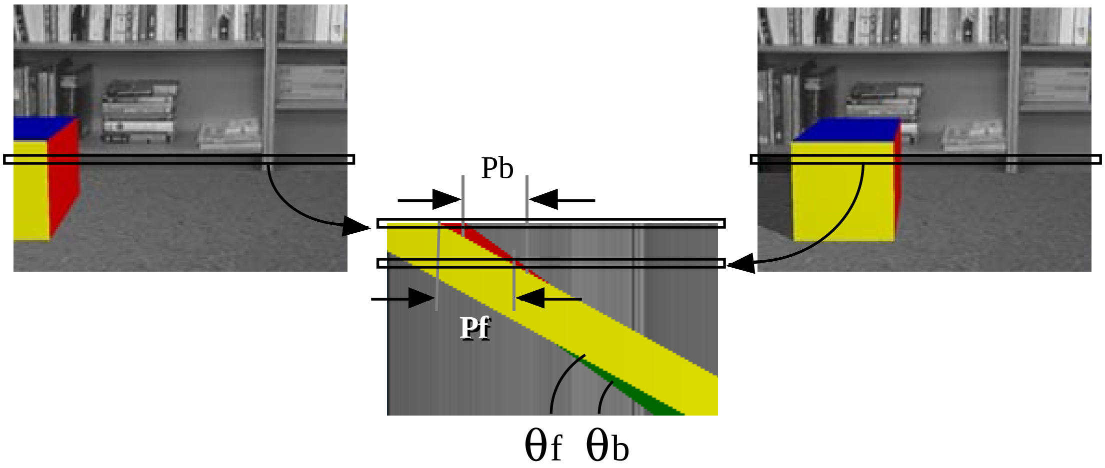


Figure 6: Stereo Pairs in an EPI

The stereo pair from Figure 5 can be considered as two lines within the EPI. Consequently the parallax can be measured directly from the EPI, and is proportional to the angle of the bands.

than the first. The epipolar image will contain two sinusoids, one with a smaller amplitude than the other. If the two glasses are in line with the centre, the sinusoids will be in phase. Shifting the first glass around the edge of the turntable will give the large amplitude sinusoid a different phase to the small amplitude one. By extracting the phase and amplitude of each sinusoidal curve in an epipolar image, the position of each tracked point relative to the camera can be calculated. Again this allows the size of the object to be determined.

In Figure 7 a cube is rotated around its centre and the XYT volume is formed and sliced through an XT plane. Figure 8 shows the resulting EPI, with one edge highlighted as it moves through the spatiotemporal volume. In this example, because the cube was rotated around its centre, all four edges form sinusoids with the same amplitudes, but different phases. The amplitudes and phases of the sinusoids are extracted and used to

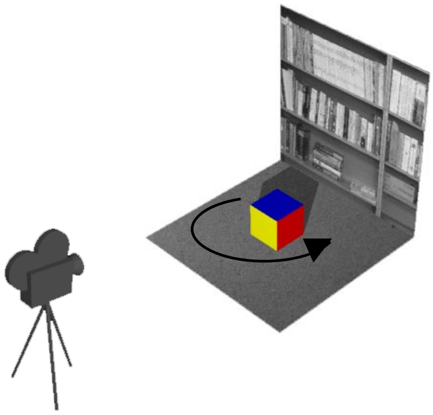
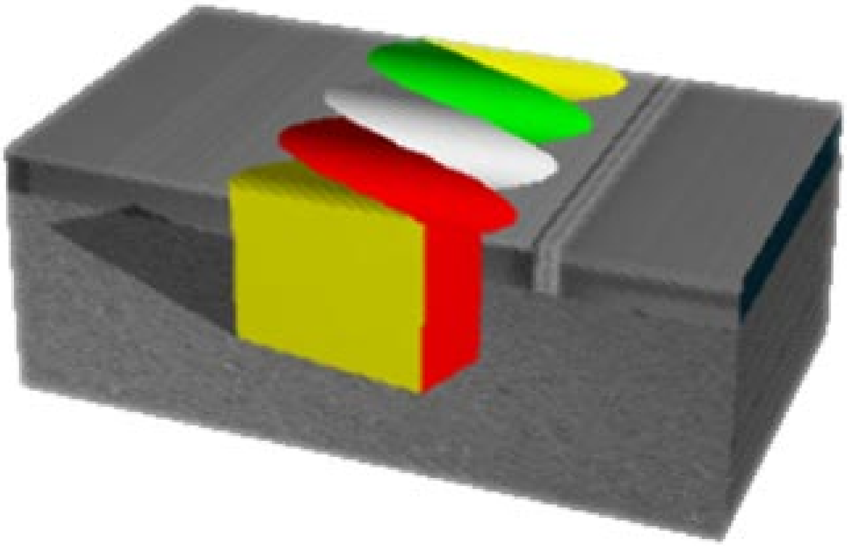


Figure 7: Spatiotemporal Volume

Left: Camera films the scene as block rotates about its centre.

Below: XYT volume sliced to show an EPI.



locate the edges in three dimensional space, in polar coordinates (Figure 9). Since all the sinusoids have the same amplitude, the edges are all the same distance from the centre. Similarly the phase of the sinusoids are at 90° intervals, as are the edges of the cube in polar coordinates.

Also note that the sinusoids are occluded for half of the EPI: the functional nature of each point's locus means once the parameters have been determined, the point can be "found" when it reappears after occlusion.

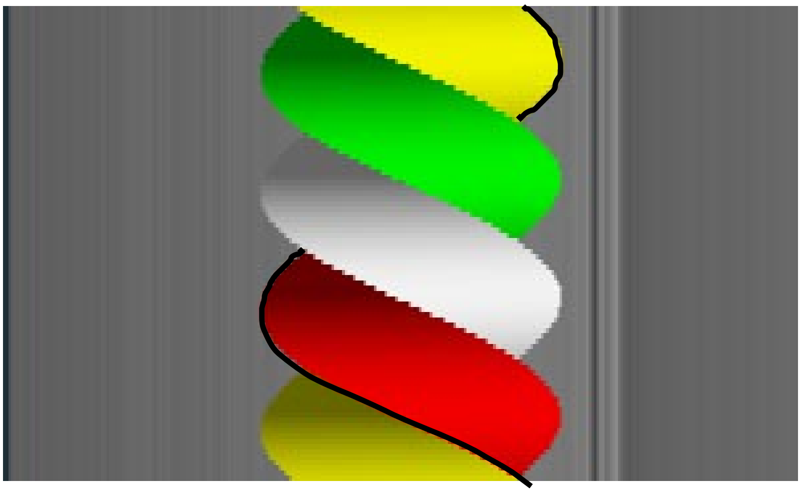


Figure 8: EPI of Rotational Motion

The motion path of each edge is sinusoidal. Time is shown on the vertical axis. Note the highlighted edge becomes obscured then reappears in an expected location.

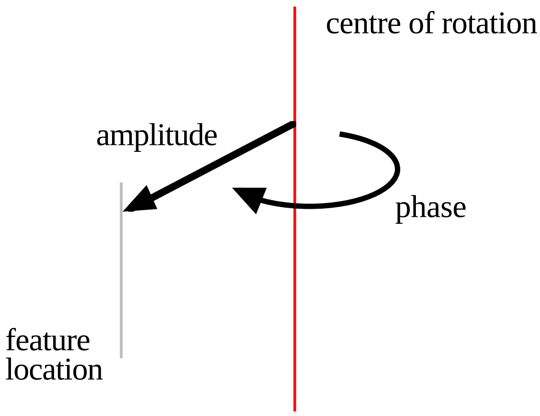


Figure 9: Location using Polar Coordinates

To locate the edge highlighted in Figure 8, the extracted phase determines the rotation and the amplitude determines the distance from the centre.

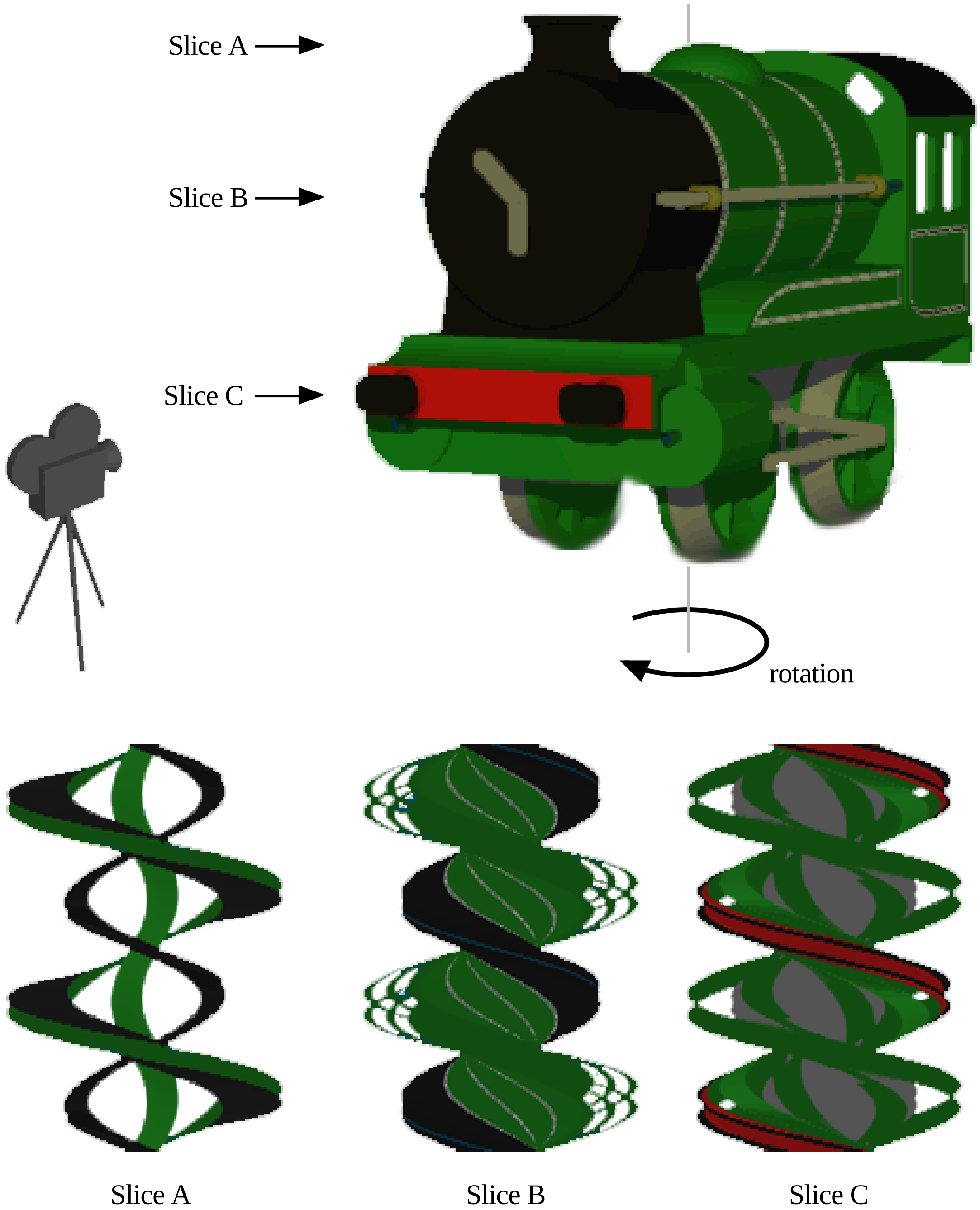


Figure 10: Complex Example of Rotational Motion

Top: The locomotive is rotated about its centre as the images are captured.

Bottom: Three EPIs from different positions within the spatiotemporal volume. Slice A shows the smoke-stack, bell top and cab roof, with clearly distinguishable phases and amplitudes. Slice B shows the boiler and cab. Slice C shows the front and rear buffers, including overhang at the rear.

Figure 10 shows a more complex example of rotational motion. The locomotive has been rotated about its centre, and the captured images formed into a spatiotemporal volume. Three EPI slices are shown from various heights within the volume. The first slice contains three sinusoidal bands, representing the smoke stack, bell top and cab roof. These bands clearly show the relationship between the phase and amplitude of each sinusoid, and the position of the feature in polar coordinates. Similarly the other slices show identifiable features as sinusoidal stripes.

Figure 11 shows an example of real video data, using the toy locomotive simulated in Figure 10. Klette *et al.* noted the transition from noiseless simulated data to noisy real data can be catastrophic for shape from motion techniques. Recall that such techniques, however, rely on calculating optical flow fields between consecutive frames, and use iterative methods to smooth the fields and calculate surface gradients. It is clear from the EPI slices in Figure 11 that tracking points through a dense spatiotemporal volume is more robust in noisy data than methods using only a few images.

Hierarchical Filtering Methods

To extract the parameters of the functional motion of each point, it is necessary to track points through the spatiotemporal volume. As Bolles *et al.* pointed out, the consistent image density of the spatial and temporal dimensions allows the data to be treated as a purely spatial image. Hence normal filtering and edge detection techniques work in all dimensions; for example a LoG operator can be applied to an EPI to find edges in the motion paths.

A *data crawler* can be used to step through the XYT volume, attempting to follow a feature point. Each step is determined by a set of constraints: firstly the neighbouring pixels are examined to find the candidates for a closest match given some standard image processing criterion. This search is constrained by the knowledge that all the points' motion is planar within the spatiotemporal block, and furthermore the motion will be functional. Hence the search space is reduced to a region fitting the general profile of the motion. As the motion path of a point is extracted, the error bounds are gradually reduced, so the search constraints can be tightened.

Given the density of the spatiotemporal volume, it would be impractical to attempt to follow every pixel individually. Hence a hierarchical filtering approach is used, similar to Rosenholm's multi-resolution filtering for feature matching. Initially a very broad filter extracts only the most distinct features, which are easily tracked. Once the motion paths of the top level features are determined, they can be used to segment the image for further processing. Multiple passes using progressively finer filters can then add detail to the model, using structure determined at higher levels to constrain the search space.

CONCLUSION

A new method for extracting spatial data from motion video has been proposed. The theory behind the linear and rotational motion has been established, and the methods for extracting parameters for the generalised functional motion of points have been determined. Examples of simple and complex cases with simulated and real data demonstrate that this method shifts the difficult task of solving correspondence to the simpler domain of tracking edges and fitting data points to functions.

The new method requires functional motion between the camera and scene, currently demonstrated for the first and second order cases. Ideally it would be possible to walk around any object with a hand held camera, and then have a computer system automatically produce a three dimensional model of that object. This research is an important step along the way.

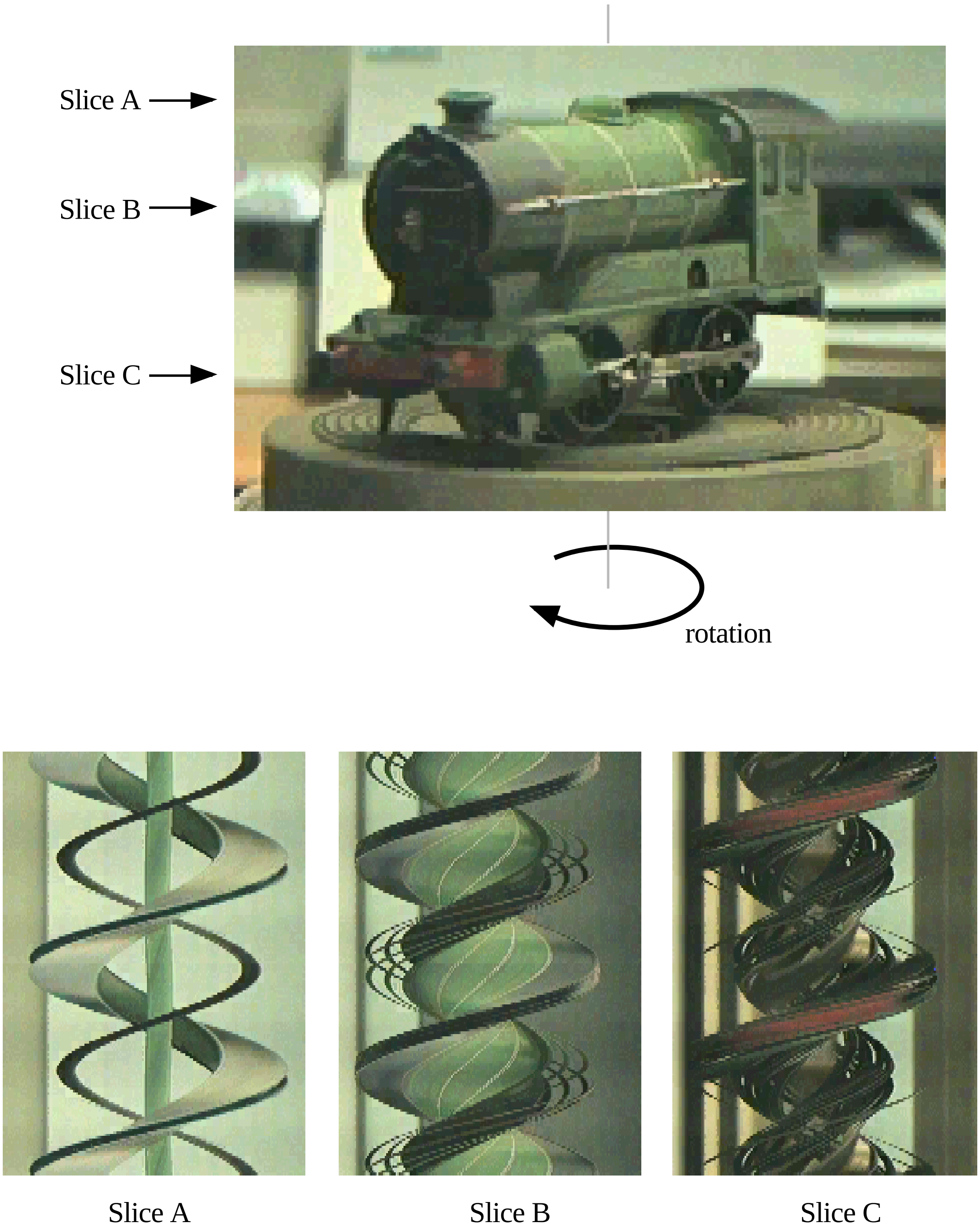


Figure 11: Real Video of Rotational Motion

Top: The locomotive is rotated on a turntable as the images are captured.

Bottom: Three EPIs similar to those in Figure 10. Note the sinusoidal bands are still clearly visible in the noisy video data.

REFERENCES

Arnold, R.D. and Binford, T.O., 1980. Geometric Constraints in Stereo Vision. SPIE Vol 238 Image Processing for Missile Guidance (1980), 281—292.

Baker, H.H., Bolles, R.C. and Marimont, D.H., 1986. A new technique for obtaining depth information from a moving sensor. Proceedings of the ISPRS Commission II Symposium on Photogrammetric and Remote Sensing Systems for Data Processing and Analysis, 1986.

Barnard, S. and Fischer, M., 1982. Computational Stereo. Computing Surveys, Vol 14, Number 4, 553—572.

Barron, J., Fleet, D., Beauchemin, S. and Burkitt, T., 1992. Performance of Optical Flow Techniques. IEEE CV&IP 3, 1992, 236—242.

Beyer, H., 1992. Automated Dimensional Inspection with Real-Time Photogrammetry. International Archive of Photogrammetry and Remote Sensing, 1992, 722—727.

Bolles, R., Baker, H. and Marimont, D., 1987. Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion. International Journal of Computer Vision, 1, 1987, 7—55.

Bruckstein, A., 1988. On Shape from Shading. Computer Vision, Graphics and Image Processing 44, 139—154.

Cappellini, V., Mecocci, A., Renault, S., 1987. Digital Processing of Stereo Images and Three Dimensional Reconstruction Techniques. Time-Varying Image Processing and Moving Object Recognition, 105—115.

Chin, R. and Dyer, C., 1986. Model-Based Recognition in Robot Vision. Computing Surveys, Vol 18 No 1, March 1986, 67—108.

Clarke, T.A., Cooper, M.A.R., Chen, J. and Robson, S., 1995. Automated Three Dimensional Measurement Using Multiple CCD Camera Views. Photogrammetric

Record, Vol 15 No 85, April 1995, 27—42.

Erez, M.T. and Dorrer, E., 1984. Photogrammetic Data Acquistion using an Interactive Computer Graphics Systems. Photogrammetric Engineering and Remote Sensing, Vol 50, February 1984, 183—188.

Fraser, C., 1988. State of the Art in Industrial Photogrammetry. International Archive of Photogrammetry and Remote Sensing, 27(B5), 166—181.

Fraser, C. and Shortis, M., 1994. Industrial Inspection Using a Still Video Camera. Proceedings Resource Technology '94, Melbourne, Australia, 362—375.

Klette, R., Mehren, D. and Rodehorst, V., 1995. An Application of Surface Reconstruction from Rotational Motion. Real-Time Imaging, 1 (1995) 127—138.

Koch, R., 1992. Model-Based 3D Scene Analysis from Stereoscopic Image Sequences. International Archive of Photogrammetry and Remote Sensing, 1992, 427—433.

Kölbl, O., Chardonnens, B., Gilliéron, P.-Y., Hersch, R.D. and Lutz, S., 1991. The DSR15T A System for Automatic Image Correlation. Proceedings of ASPRS/ACSM/ AUTO-CARTO 10, Baltimore 1991.

Lee, K. and Jay Kuo, C.-C., 1992. Shape Reconstruction from Photometric Stereo. IEEE CV&IP 3, 1992, 479—484.

Marr, D., 1982. Section 3.3: Stereopsis. Vision, 111—159. WH Freeman 1982.

Peipe, J. and Schneider, C.-T., 1995. High Resolution Still Video Camera for Industrial Photogrammetry. Photogrammetric Record, Vol 15 No 85, April 1995, 135—139.

Rivett, L.J., 1983. The Assessment of Post-Operative Swelling by Analytical Photogrammetry. Australiasian Physical & Engineering Science in Medicine (1983), Vol 6 No 3, 138—142.

Rosenholm, D., 1987a. Least Squares Matching Method: Some Experimental Results. Photogrammetric Record, 12(70): 493—512 October 1987.

Rosenholm, D., 1987b. Empirical Investigation of Optimal Window Size Using the Least Squares Image Matching Method. Photogrammetria 42 (1987), 113—125.

Sheng Jin and Yunming Li, 1988. Acquiring Range Data in the Binocular 3-D Computer Vision using Edge-Based Hierarchical Matching. IAPR Workshop on CV, October 12-14, 1988, Tokyo.

Trivedi, H., 1985. A Computational Theory of Stereo Vision. IEEE 1985 277—282.

Ulupinar, F. and Nevatia, R., 1995. Shape from Contour: Straight Homogeneous Generalized Cylinders and Constant Cross Section Generalized Cylinders. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 17 No 2, February 1995, 120—135.

Ulupinar, F. and Nevatia, R., 1992. Recovery of 3-D Objects with Multiple Curved Surfaces from 2-D Contours. IEEE CV&IP 3, 1992.

Zheng, J. Y., 1994. Acquiring 3-D Models from Sequences of Contours. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 16 No 2, February 1994.

Zheng, Y., Jones, D.G., Billings, S.A., Mayhew, J.E.W. and Frisby, J.P., 1990. SWITCHER: A Stereo Algorithm for Ground Plane Obstacle Detection. Image and Vision Computing, Vol 8 Num 1, February 1990.