Vertical Parallax from Moving Shadows

Yaron Caspi
Faculty of Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot, 76100, Israel

Michael Werman
School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem 91904, Israel

Abstract

This paper presents a method for capturing and computing 3D parallax. 3D parallax, as used here, refers to vertical offset from the ground plane, height. The method is based on analyzing shadows of vertical poles (e.g., a tall building's contour) that sweep the object. Unlike existing beam-scanning approaches, such as shadow or structured light, that recover the distance of a point from the camera, our approach measures the height from the ground plane directly. Previous methods compute the distance from the camera using triangulation between rays outgoing from the light-source and the camera. Such a triangulation is difficult when the objects are far from the camera, and requires accurate knowledge of the light source position. In contrast, our approach intersects two (unknown) planes generated separately by two casting objects. This omits the need to precompute the location of the light source. Furthermore, it allows a moving light source to be used. The proposed setup is particularly useful when the camera cannot directly face the scene or when the object is far away from the camera. A good example is an urban scene captured by a single webcam.

1 Introduction

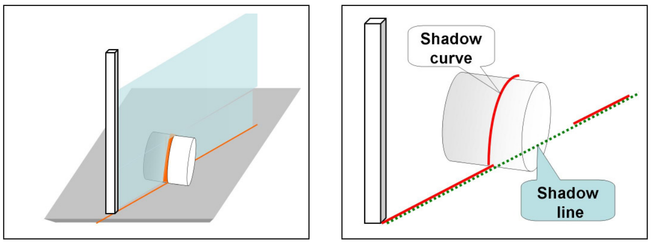
This paper deals with the recovery of a three-dimensional structure from images. The recovered structure is generally derived from point correspondences where the output is represented as a depth map, distance from the camera, or a height map, distance to some reference plane. An alternative to point correspondences is the use of structured light or cast shadows ([1, 3, 16] to list but a few). By shadow or structured light methods, we refer to methods that combine a single camera with a source of light, e.g., a laser beam, light strip (or stripes) from a projector or a cast shadow edge. However, in contrast to shadow-based methods that produce depth maps, in this paper we describe a method that produces height maps. To the best of our knowledge, 3D reconstruction from cast shadow edges or light strips were limited to recovery of depth maps as their intermediate representation.

To derive height maps, we use the Plane + Parallax (P+P) representation. Plane + Parallax (P+P) has been touted as an excellent representation for 3D reconstruction (e.g., [9, 11, 13, 14, 15, 17]). The proposed method works as follows. Every image pixel, and its unknown 3D location in the scene, are linked to the 3D point vertically below it on the reference plane, or more precisely to the 2D image location of this corresponding point. These 2D image offsets are converted to a dense height map using known techniques from P+P studies.

The input for the proposed method consists of two image sequences of a scene, static except for moving shadows cast by vertical obstructions, where either the light source, the obstructor(s), or both move. To illustrate these requirements, our scenario is an object in an outdoor scene which is ”swept” by shadow edges twice, each time by a different vertical obstruction (e.g., two buildings). The ”sweep” results from the relative motion of the sun (see top rows of Fig. 5). The cast shadow of a building edge on the reference plane is a line. The 3D object on the reference plane deforms this line. We present a simple constraint that captures each point's parallax by combining the information of two such deformations.

Once our setup and the accompanying reconstruction is complete, we can assign real-world Euclidean units (up to a global scale factor) to every visible point above the reference plane. The theoretical results on P+P representation can be applied for this task. In particular, the formulation in this paper is a direct derivation from the methods of [6] and [17]. The advantage of our approach over [6] is that the vertical correspondence is generated automatically for every scene point. This replaces the need to manually mark the point correspondences as in ([6]).

The main benefit of the proposed approach over previously proposed shadow-based or other structured light systems is when the objects are far away from the camera (e.g., outdoor scenes). In these cases, depth variations are fairly



(a) (b)

Figure 1. Shadow curves and shadow lines. *The left-hand side (a) displays a pole or tall building casting a shadow on an object (the cylinder) placed on the ground plane. The right-hand side (b) illustrates two notions that are used in this paper to describe a shadow edge: (i) "shadow-curve" and (ii) "shadow line". The shadow curve is the visible trace of the casted shadow (displayed in red). The shadow line is the line generated by the intersection of the shadow vertical plane with the ground plane (marked by a dotted green line).*

small with respect to the distance to the camera. Therefore, meaningful depth maps are difficult to compute. In contrast, the recovered parallax map (height from the reference plane) using a P+P representation is not affected by the distance to the camera, and is therefore more accurate in these cases.

The computation of depth maps in classic shadow-based approaches (e.g., [1]) is based on triangulation, i.e., on intersecting rays from the light source with rays from the camera. This constrains the light source to be in a fixed and known position, typically computed in a preprocessing calibration step. In contrast, the method proposed in this paper does not place constraints on the location of the light source and/or the obstruction. It can, therefore, use the sun (and its motion) as the source of the cast shadow. Alternative technologies for remote capturing of structure such as laser range finder may also be used in our scenario, see for example [8]. Such active methods are beyond the scope of this paper.

The rest of paper is organized as follows: Section 2 describes the setup. Section 3 describes the recovery process. Section 4 shows an outdoor example. Before, concluding the paper we describe implementation (Section 5) issues, and discuss factors effecting the accuracy of our method (Section 6).

2 Capturing Parallax

Before describing the setup and its associated geometry we present some notations used in this paper. Assume a tall pole (or a building) casts a shadow on a plane with an object on it as illustrated in Fig. 1.(a). The shape of the shadow edge (transition from dark to bright) results from the intersection of a vertical plane with the object and the ground plane. This plane is shown in transparent blue in Fig. 1.(a). This intersection defines two concepts that are used in this paper: *shadow curve* and *shadow line*, both are illustrated in Fig. 1.(b). The shadow curve is the trace of the vertical plane shadow on the object and on the ground. The shadow line is the line generated by the intersection of the shadow's vertical plane with the ground plane. Note that while the shadow line might not be seen at some points on the ground plane it is still well defined.

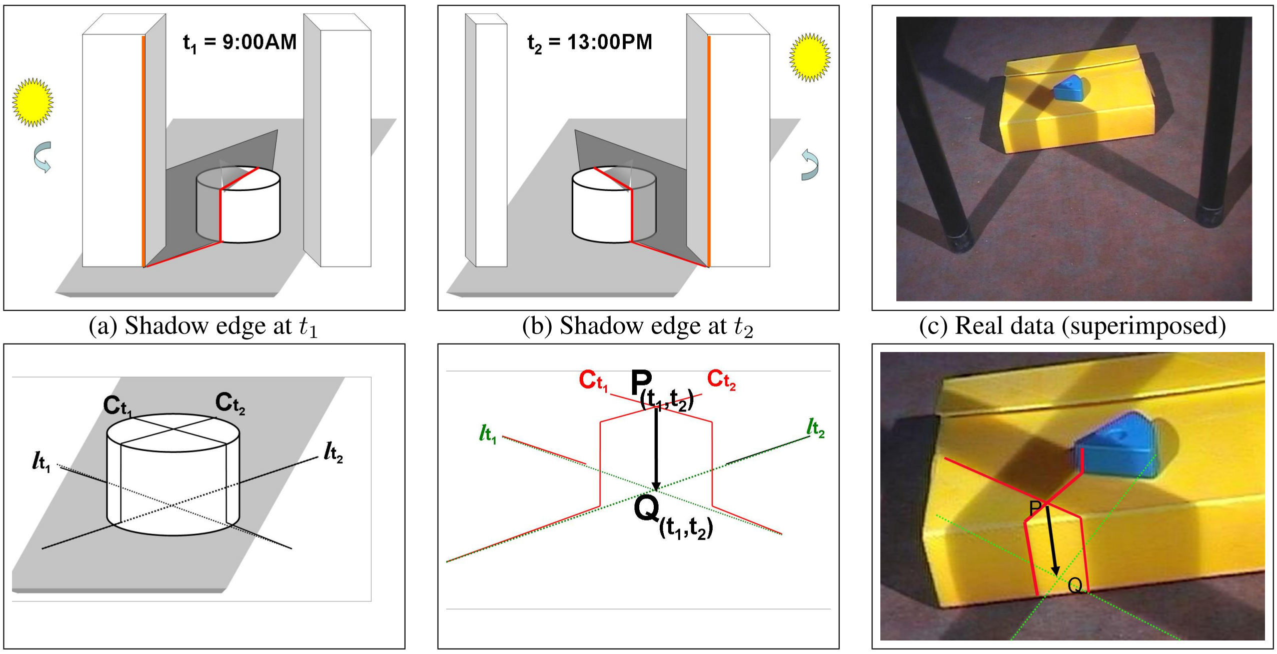
2.1 The Setup

The capturing "scenario" is illustrated in Figure 2. It illustrates two tall buildings casting shadows on a small object lying on the ground plane. The goal is to model the small object (the cylinder). This figure illustrates a fairly common scenario in urban environments, where two buildings or two corners of the same building cast shadows on the same objects (at different times during the day).

In general, it suffices to assume that the poles (building contours) that cast the shadows are parallel and not necessarily orthogonal to the ground plane. For example, the ground plane need not be horizontal, it can be slanted. In this case the buildings are vertical, but not orthogonal to the plane. The recovered height will be along the direction of the parallel poles (vertical), and the height is the distance to the corresponding point (below it) on the slanted plane. Since in most cases the poles are vertical, we refer to them as vertical poles[1].

Given 2 images captured at different times, t_1 and t_2, using the same static camera, we have two shadow lines and two shadow curves. These are denoted by l_{t_1} and l_{t_2} and by C_{t_1} and C_{t_2}, respectively, and are displayed in Fig 2 (d). Given these 2 lines and 2 curves, we define a real intersection point (curve intersection) and a virtual intersection point (line intersection), both are displayed in Fig 2 (e). The real (curves) intersection point is denoted by P and belongs to both shadow curves. The virtual (lines) intersection point is denoted by Q and is defined by the intersection of the two shadow lines (we ignore pathological cases where two different poles cast the same shadow line from different directions, i.e., exactly 12 hours apart). Once P and Q are defined, we define the *vertical mapping*. The vertical mapping is the mapping from P to Q (see Fig 2. (e)). The vertical mapping is a mapping from real intersection points (intersection of curves, possibly not on the plane) to the intersection points of the virtual lines on the reference

[1] If the poles are only parallel but not vertical, the "height" reconstruction will be along their joint direction.



(a) Shadow edge at t_1 (b) Shadow edge at t_2 (c) Real data (superimposed)

(d) Shadow curves and lines (e) The parallax mapping (f) Real data

Figure 2. The vertical mapping. *The top row illustrates the setup. It displays shadows cast at two different times, (t_1, t_2). Graphically in (a and b), and superimposed real data in (c). The second row (d) -(f) illustrates the geometrical constraint imposed by the shadow curves and lines. (d) Displays two pairs of curve shadows and line shadows superimposed on the same frame (see Fig. 1 for notations). (e) Displays the real and virtual intersection points: $P_{t_1,t_2} = C_{t_1} \bigcap C_{t_2}$ and $Q_{t_1,t_2} = l_{t_1} \bigcap l_{t_2}$. These define the vertical mapping $P_{t_1,t_2} \rightarrow Q_{t_1,t_2}$ indicated by an arrow. The same construction is illustrated in (f) on real data.*

plane. The P's (curve intersections) are always visible, but the Q's (line intersections) may be occluded by the object. For example, all visible points on the plane are mapped by the vertical mapping onto themselves.

A further illustration and explanation of vertical mapping and its properties is illustrated in Fig. 3. It explains why such a setup (two vertical poles) provides the desired output, i.e., a vertical mapping in 3D. The crucial observation is that the mapping is parallel to the vertical poles. This is because it is induced by an intersection of two vertical planes. Again, this result holds even if the poles are just parallel and not vertical. Then the parallax mapping vectors will be parallel to both poles (but not vertical).

The above explanation focuses on a single scene point. To generate a dense mapping (i.e., a vertical mapping for every scene point), we use two sequences of shadow images, each sweeping the scene/object once. The temporal length of such sequences depends on the time each shadow edge sweeps the object. In our outdoor experiment, each sweep was 15-minutes long, and they were taken about 2 hours apart.

3 Three-Dimensional Reconstruction

Thus far, we have not considered the camera. Given a sequence of images of such a scene with cast shadows, we only have access to projections of P and Q. In this section, we discuss what can be computed about real 3D values (height), once the projection of the vertical mapping pairs of points are known.

Previous theoretical results on plane + parallax may be used to analyze the geometry of the vertical mapping presented in this study (see for example [9, 10, 11, 13, 14], or [15, 12] for recent results). They all show that the 3D height of a point from the reference plane can be determined up to a global scale factor.

To derive height equations for every point, we take a stratification approach, as proposed in [17], where the reconstruction problem is decomposed into two steps: (i) affine calibration/rectification of the reference plane and (ii) reconstruction using affine equations.

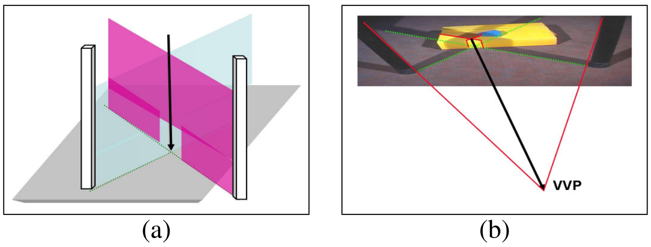


Figure 3. The vertical mapping properties. *The left hand side (a) illustrates the cause of verticality. It shows two poles. The shadows cast by the left pole must be on the left vertical plane (blue). The shadows cast by the right pole must be on the right vertical plane (red). Thus their intersection is a vertical line denoted by a dark arrow. The right hand side (b) illustrates "verticality" by looking at the vertical vanishing point (VVP) - the 2D intersection point of projections of 3D vertical lines. In this example the vertical vanishing point can be recovered by intersecting the lines of the table legs. Finally, (b) also shows that if the vertical vanishing point is known, a single shadow sweep suffices (i.e., we need only one pole). Simply because the intersection of the line connecting a shadow pixel p and vpp with the single shadow line defines the point q.*

3.1 Shadow-based Affine Calibration

To affinely calibrate the reference plane, we used parallel shadow lines. If in some frames two (or more) shadows induced by different vertical poles are visible simultaneously, their corresponding shadow lines on the reference plane are parallel. The intersection of parallel lines provides a point on the line at infinity l_∞ (see Fig. 4). Given 2 pairs (or more) of parallel lines, it is possible to compute l_∞. Once $l_\infty = [l_1, l_2, l_3]^T$ is known, we can calibrate the reference plane using the following homography $H_{affine\ calibration} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix}$ (see [7]).

3.2 Relative Height Reconstruction

To simplify the derivation of relative height equations, we induce the rectified image affine coordinate system (X and Y) to the ground plane. The Z axis is in the direction of the poles (i.e., vertical direction), with the ground plane being at $Z = 0$. As a result, we get a simple form for the camera projection matrix.

Denote by $p = (u, v, 1)$ and $q = (u', v', 1)$ the images of a pair of scene points $P = (X, Y, Z, 1)$ and $Q = (X, Y, 0, 1)$, i.e., ($p \cong MP$ and $q \cong MQ$). Our special coordinate system implies that $u = X$, and $v = Y$, thus M the 3×4 camera matrix has the following form:

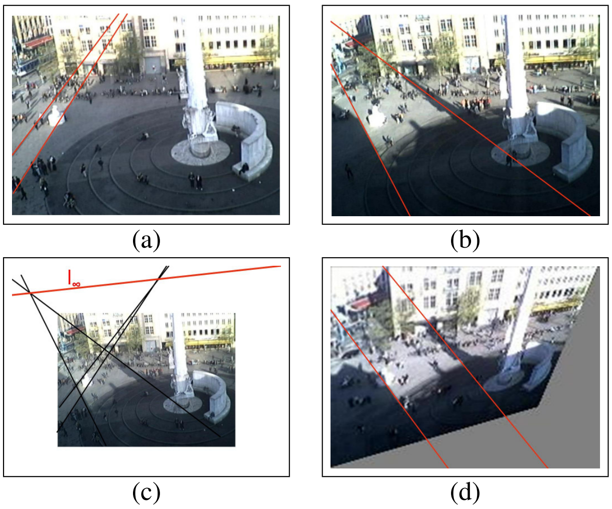


Figure 4. Shadow-based affine calibration. *Figures (**a**) and (**b**) display pairs of parallel shadow lines (red lines). Their intersection points are on the line at infinity (l_∞). Thus it can be recovered (**c**). The resulting affinely calibrated image is shown in (d). The shadow lines are now parallel.*

$$M = \begin{pmatrix} 1 & 0 & \alpha & 0 \\ 0 & 1 & \beta & 0 \\ 0 & 0 & \gamma & 1 \end{pmatrix} \tag{1}$$

This simple form results from the fact that using this coordinate system, M preserves the X and Y values of points on the reference plane (with $Z = 0$). Applying M on P and Q yields:

$$u = \frac{X + \alpha Z}{1 + \gamma Z}\ , v = \frac{Y + \beta Z}{1 + \gamma Z} \tag{2}$$

and

$$u' = \frac{X + \alpha \cdot 0}{1 + \gamma \cdot 0} = X\ , v' = \frac{Y + \beta \cdot 0}{1 + \gamma \cdot 0} = Y. \tag{3}$$

Eliminating Z from 2 and substituting X and Y gives:

$$u\beta + u' v\gamma - u' \beta - v\alpha - v' u\gamma + v' \alpha = 0, \tag{4}$$

a linear homogeneous equation in (α, β, γ), so that 2 pairs of points are sufficient to compute (α, β, γ) up to scale. Once α, β, γ are found, we can compute Z:

$$Z = \frac{v' - v}{v\gamma - \beta} = \frac{u' - u}{u\gamma - \alpha}. \tag{5}$$

Again, the results are up to one global scale that has to be assumed, known or guessed.

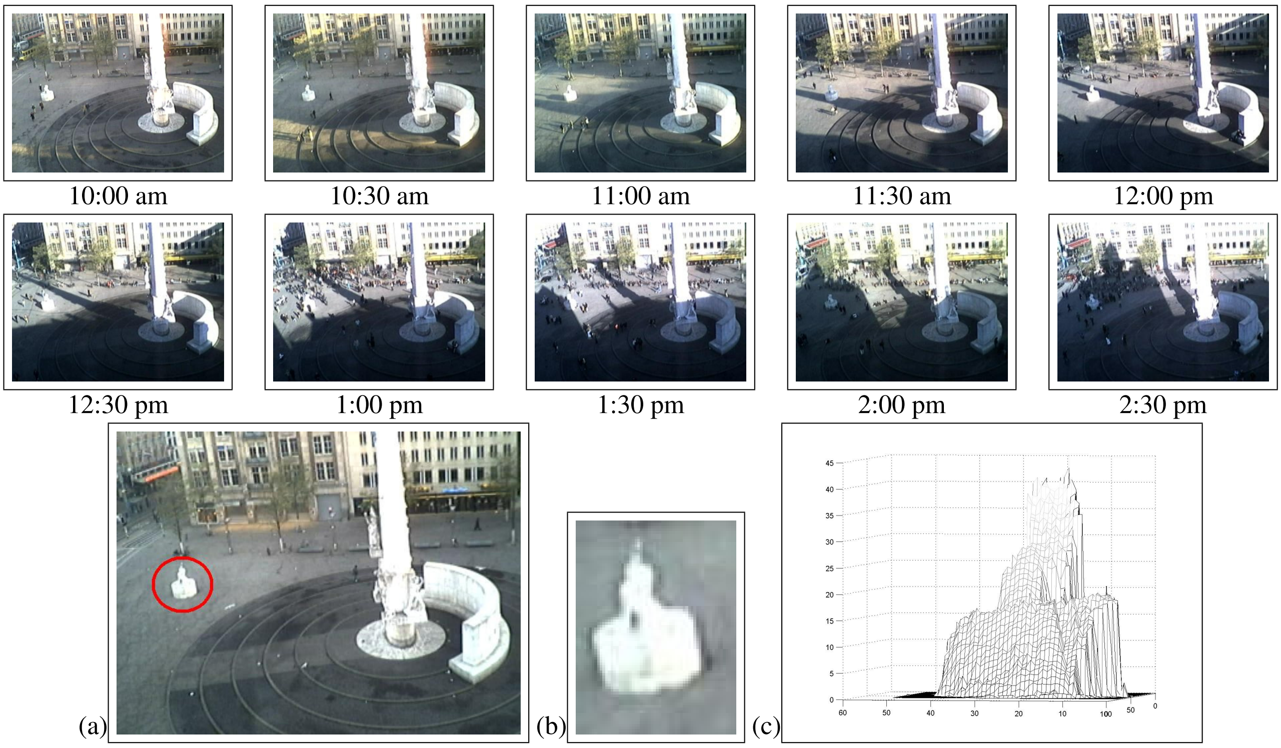


Figure 5. Outdoor example. *The first two rows display representative frames downloaded from a webcam in Amsterdam (www.damstaete.nl/damsite). The shadows used in this example are from the pole in the center of the square (around 11:30 am), and from a tall building invisible in the frame (around 1:30 pm). The task is to reconstruct the 3D structure of the statue of a lion (shown from its back) and its circular base marked by a red circle in (a). A magnified view is displayed in (***b***). The reconstructed height map is shown in (***c***).*

An equivalent approach has been proposed in ”single view metrology” [6, 5]. The geometry is similar but, there pairs of points above each other are selected manually. Height of such a manually selected point from its corresponding point on the reference plane is computed using the cross ratio. It is constructed from the following 4 image points: The 2 manually marked points, the vertical vanishing point, and the intersection of the line defined by these points with the line at infinity.

It is interesting to note that there are applications that do not require any calibration. Chuang et al. [4] proposed a similar setup to the one proposed here. They also sweep the scene/object twice by a shadow, generated by translating the obstructing object (a pole) in front of the camera. Their algorithm generates a mapping from the reference plane to scene points, that is later used for insertion of a novel shadow. Since they focus on shadow rendering in the specific light+camera setup, their method does not require calibration. This implies that the results are limited to the particular constellation of the camera and sun at the time of taking the input sequences.

4 An Example

This section illustrates the applicability of our method for scenarios that are highly challenging for other methods. We downloaded data from a remote webcam[2], at intervals of 5 seconds, on a sunny day. Representative images of the input data are shown in Fig. 5. The task is to reconstruct the 3D structure of a statue of a lion. It is marked by a red circle in Fig. 5.(a) (the lion is facing the other direction). Fig. 5.(b) shows a magnified view. The algorithm was applied to two 15-minute sequences (around 11:30AM and 1:30PM), when shadow edges were ”sweeping” the statue. Please note the challenges of this scene: 1) Only a single webcam is used. 2) The shallow viewing angle. 3) The object is textureless. 4) The distance from the camera (about 100m). 5) The poor input resolution, object dimensions are 60 by 80 pixels. The reconstructed height map is shown in Fig. 5.(c).

[2] Data was taken from a camera at http://www.damstaete.nl/damsite. To the best of our knowledge the camera was temporarily removed.

5 Extracting Shadow Lines and Curves

Our shadow-extraction algorithm is based on temporal analysis as proposed in Bouguet and Perona [2]. Namely: (i) for each pixel, find its minimal and maximal value along time, (ii) the time a shadow edge passes a particular pixel is when its intensity value equals the average between minimum and maximum values. However, some modifications have been introduced to address temporal illumination changes.

First we apply a temporal median filter to prune out non static objects such as walking people. We manually marked planar regions around the object, and identified shadow lines at the beginning and end of each sequence. We now transform the input images (only the region of interest) to a polar coordinate system with origin at the pole. In this coordinate system (angle, radius, time) shadows move horizontally along image rows. This allows us to enforce temporal smoothness within each row. The resulting 3D semi-polar volume (an xt cut of this volume) was analyzed using dynamic programming. The dynamic programming cost function was the distance from average of minimum and maximum temporal values. The results where warped back to the image coordinate system.

6 Error Analysis

This section explores possible sources of errors and analyzes their effect on potential accuracy. Such errors might result from violation of our assumptions and/or noise in measurements. This section analyzes whether our processing magnifies these errors or not. One possible assumption that may be violated is the planarity of the reference ground plane. Namely, that the region where the planar shadow lines are approximated is on a different plane than the object (see Fig. 6.(a)). In this case, the height will be approximated with respect to Q', the point resulting from extrapolation of the shadow line in the left side of the plane. The actual height should be measured with respect to Q - a point on a different shadow line (on a different plane). To minimize this extrapolation error, it is recommended that the shadow line fitting be done from both sides of the object (i.e., replace extrapolation by interpolation).

Similarly, the casting poles may not be vertical (non parallel to each other). The resulting ”vertical mapping” will be slanted (see Fig. 6.(b)). Such a pole slant may be decomposed orthogonally into two components: one in a plane parallel to the light’s direction, and the other perpendicular to it. Note that only the later biases the results. Along this direction it shifts the matching point on the reference plane as a sin of the drift angle of the pole ($sin(\alpha)$) times the height from the reference plane. This is further increased by a factor of $\frac{1}{sin(\beta)}$, when searching for the intersection point with the second shadow line, where β is the angle between these lines (see Fig. 6.(b)). Therefore, the height output is more robust when the angle between shadow lines is close to 90^o.

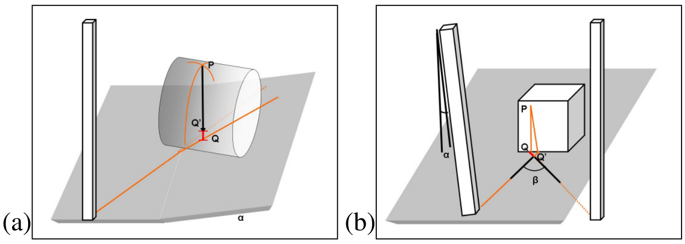


Figure 6. Error analysis. *This figure presents the effects of errors in our assumptions about the scene. (a) illustrates the error resulting from non planarity of reference plane. (b) illustrates the contribution of non vertical pole. The input error angle is marked by α, while β marks the angle between shadow lines. Output offsets are marked by a red bar. See text for further details.*

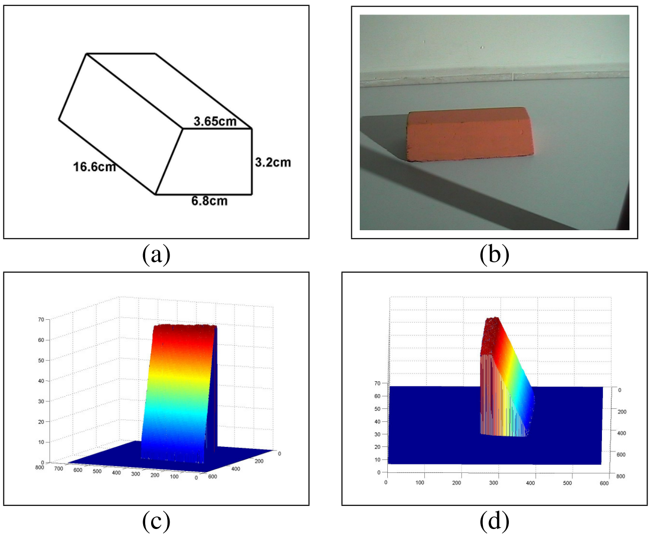
6.1 Testing the Accuracy

It should be appreciated that only the ”verticality” of the poles is required for capturing parallax. Therefore, both the pole and/or the light source may move or change. Moving the pole and not the light source simplifies the shadow edge extraction and results in a better reconstruction. In this experiment the user sweeps the object with a shadow of a narrow moving object and maintains the verticality of the obstructing object. The object was a piece of wood with a trapezoid profile. The test object and its dimensions are illustrated in Fig. 7(a). A representative frame from the input sequences, is shown in Fig. 7(b).

In (c) and (d) we can see the reconstructed height map, from front and side views. To measure the accuracy of the reconstruction results, we compute RMS error with respect to each dominant plane. The error was less than 1%, which translates to less than 0.16 mm.

Conclusions

We presented an approach for capturing parallax (height from the ground plane) from moving shadows cast by vertical obstructors. The main benefits of the presented method are for outdoor scenes, when the object is far from the camera, e.g., the webcam example in Fig. 5. Previously proposed shadow-based (or structured light) approaches are impractical in these cases. As our method (i) recovers height and not depth, and (ii) does does not constrain the



(a) (b)

(c) (d)

Figure 7. Accuracy test. *This figure presents results of measuring height in a controlled setup. (a) illustrates the objects dimensions graphically. (b) displays a representative input image. (c) and (d) represent the reconstructed parallax. See text for further details.*

location of the light source, it can use the sun and is less affected by the distance to the camera. It is interesting to note that methods based on inputs from multiple cameras (e.g., stereo) will also face difficulties in such outdoor examples. The ratio between depth variations and distance from the camera/s will require an impractical base line between the cameras. Therefore, for such a scenario, the proposed method subsumes previously published approaches.

References

[1] J.-Y. Bouguet and P. Perona. 3d photography on your desk. In *International Conference on Computer Vision (ICCV)*, Bombay, India, January 1998.

[2] J.-Y. Bouguet and P. Perona. 3d photography on your desk. Technical Report TR-136-93, CA. Institute of Technology, 1998.

[3] N. L. Chang. Efficient dense correspondences using temporally encoded light patterns. In *PROCAMS*, Nice, France, October 2003.

[4] Y.-Y. Chuang, D. B. Goldman, B. Curless, D. H. Salesin, and R. Szeliski. Shadow matting and compositing. In *ACM Transactions on Graphics, (SIGGRAPH)*, San Diego, CA, July 2003.

[5] A. Criminisi. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. Springer-Verlag London Ltd, 2001.

[6] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision (IJCV)*, 40(2):123–148, Nov 2000.

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[8] O. Incorporated. Ilris 3d. http://www.optech.ca.

[9] M. Irani and P. Anandan. Parallax geometry of pairs of points for 3D scene analysis. In *European Conference on Computer Vision (ECCV)*, Cambridge, UK, April 1996.

[10] M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multi-view parallax geometry and applications. In *European Conference on Computer Vision (ECCV)*, pages 829–845, Freiburg, June 1998.

[11] R. Kumar, P. Anandan, and K. Hanna. Direct recovery of shape from multiple views: a parallax based approach. In *International Conference on Pattern Recognition (ICPR)*, pages 685–688, Jerusalem, 1994.

[12] C. Rother. Linear multi-view reconstruction of points, lines, planes and cameras, using a reference plane. In *International Conference on Computer Vision (ICCV)*, pages 1210–1217, Nice, France, October 2003.

[13] H. S. Sawhney. 3D geometry from planar parallax. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 929–934, Seattle, WA., June 1994.

[14] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 483–489, Seattle, WA., June 1994.

[15] B. Triggs. Plane + parallax, tensors and factorization. In *European Conference on Computer Vision (ECCV)*, pages 522–538, Dublin, Irlande, June 2000.

[16] G. Wang, Z. Hu, F. Wu, and H.-T. Tsui. Projector-camera based system for fast object modeling. In *PROCAMS*, Nice, France, October 2003.

[17] D. Weinshall, P. Anandan, and M. Irani. From ordinal to euclidean reconstruction with partial scene calibration. In *SMILE*, pages 208–223, 1998.