What do planar shadows tell about scene geometry?

Jean-Yves Bouguet†, Markus Weber‡ and Pietro Perona††

† California Institute of Technology, 136-33, Pasadena, CA 91125, USA
‡ Università di Padova, Italy
{bouguet,rmw,perona}@vision.caltech.edu

Abstract

A method for reconstructing 3D scene geometry from a set of projected shadows is presented. It is composed of two stages. First, the scene geometry is retrieved up to three scalar unknowns using only the information contained in the observed shadow edges on the image plane. Then, the three remaining unknowns are computed making use of the known depths at three points. This technique improves upon previous results [2] in that it does not require the presence of a reference plane in the background. A mathematical analysis is presented using dual-space geometry, a formalism that provides adequate tools to carry out all the derivations in a compact and intuitive manner. A linear algorithm based on singular value decomposition (SVD) is presented leading to a closed form solution for reconstruction.

1 Introduction and Motivation

The problem of acquiring three dimensional models of real objects has recently received a lot of interest, due to numerous applications, such as animation and industrial design. Three dimensional scanners are among the most successful solutions for acquiring shape accurately and densely [1, 7, 9]. These systems use active devices (Laser, LCD projector) to project patterns on the object, and commonly need motorized parts to achieve full scene coverage (translation and/or rotation stages). This makes them very accurate, but they are often expensive and/or bulky.

We have recently proposed a method for capturing 3D surfaces based on projecting shadows onto the scene using a pencil (or another straight edge) and a simple desk lamp [2]. Aside from being cheap and accurate, this approach has the advantage of achieving full Euclidean reconstruction, however it requires the presence of a background plane used as a reference. One question remained: Can we do without a reference surface, or what information can be extracted about the scene geometry from a set of projected shadows alone? In this paper, we demonstrate that, under the assumption that the light source position is known, planar shadows provide sufficient information for Euclidean 3D reconstruction (up to three global scalars) and propose a simple algorithm for achieving such a reconstruction.

We start with the description of the method in Sec. 2, followed by experimental results in Sec. 3. We end with conclusions in Sec. 4.

Figure 1: Scanning method using a reference plane Πd: Assume that the positions of the light source S and the plane Πd (reference plane) in the camera reference frame are known from calibration. During scanning, the user casts a shadow on the scene observed as a curved edge E on the image. The goal is to estimate the 3D location of the point P in space corresponding to any point p on E. Denote by Π the corresponding shadow plane. Assume that two portions of the shadow projected on the desk plane are visible on two given rows of the image (top and bottom rows on the figure). Consider the two points a and b lying at the intersection of E and the two reference rows. Their corresponding points A and B in the scene may be found by intersecting Πd with the optical rays (Oc,a) and (Oc,b) respectively. The shadow plane Π is then inferred from the three points in space S, A and B. Finally, the point P is retrieved by intersecting Π with the optical ray (Oc,p) (triangulation stage).

2 Description of the method

Let us consider the scanning scenario as previously introduced in [2]. Figure 1 recalls the complete geometry corresponding to this technique.

The central observation is that for a given stick position, once the shadow plane Π is identified, so is the 3D position of the entire shadow edge E (by geometrical triangulation). The reference plane Πd constitutes then a direct means for locating the shadow plane in space (through the top and bottom reference rows) and achieve Euclidean reconstruction. This paper answers the question: Is there a way of recovering Π without Πd?
We will first describe the scanning scenario. The light source, $S$, is located at a known position with respect to the camera and the camera is calibrated (for the description of the calibration procedure, refer to [2] and [8]). During scanning, the user projects a series of $N$ shadows onto the scene (with a straight edge), generating shadow edges, $\mathcal{E}_1, \ldots, \mathcal{E}_N$, on the image plane. The problem of reconstructing scene geometry then leads to the problem of estimating the locations of the associated $N$ shadow planes $\Pi_p$ in space. As one shadow edge is added to the list of edges, in principle three unknowns are added to the problem, corresponding to the three degrees of freedom of the associated shadow plane in space. Are there enough constraints imposed by the images that allow to estimate this total of $3N$ variables?

2.1 A constructive approach

Let us first build some intuition about the geometry of the problem. For that purpose, we state the following two properties.

**Property 1 - Depth propagation along an edge:** If the depths $Z_A$ and $Z_B$ of two distinct points $a$ and $b$ along a given shadow edge $\mathcal{E}$ are known, then so is the depth $Z_P$ of any point $p$ along that edge. In other words, depth information at two distinct points on an edge propagates along the entire edge.

**Proof:** Let $\mathbf{x}_A$ and $\mathbf{x}_B$ be the homogeneous coordinate vectors of the two points $a$ and $b$ on the image plane: $\mathbf{x}_A = [x_A, y_A, 1]^T$ and $\mathbf{x}_B = [x_B, y_B, 1]^T$. Then, if the two depths $Z_A$ and $Z_B$ are known, so are the full coordinate vectors of the associated points $A$ and $B$ in the 3D scene: $\mathbf{X}_A = Z_A \mathbf{x}_A$ and $\mathbf{X}_B = Z_B \mathbf{x}_B$. Therefore, the associated shadow plane $\Pi$ is the unique plane passing through the three points $A$, $B$, and $S$. Once $\Pi$ is recovered, any point $p$ along the edge $\mathcal{E}$ may be triangulated leading to $Z_P$.

This approach was implicitly used in [2], in connection with a reference plane. Consider Fig. 1, where the depths of the two points $a$ and $b$ are known due to the fact that they lie on the known reference plane $\Pi_b$.

The following property is central to the new technique and carries most of the intuition.

**Property 2 - Depth propagation from edge to edge:** Let $a$, $b$, and $c$ be three distinct points on the image (in the scanable area). Assume their respective depths $Z_A$, $Z_B$, and $Z_C$ known. Then, by "appropriate" shadow scanning, one may retrieve the depth at any point $p$ on the image (in scanable areas).

**Proof:** The proof of this property is constructive. First project a shadow edge $\mathcal{E}_a$ that goes through the points $a$ and $b$, and another one ($\mathcal{E}_b$) through $a$ and $c$ (see Figure 2). Given that two points ($a$ and $b$) on $\mathcal{E}_a$ are of known depths, according to Property 1, one may compute the depth of every point along that edge. The same holds for $\mathcal{E}_b$. For any point $p$ in the image, project an edge $\mathcal{E}_p$ that passes through $p$ and intersect $\mathcal{E}_a$ and $\mathcal{E}_b$ at any two distinct points $p_1$ and $p_2$ (different from $a$). Since $p_1$ and $p_2$ lie on the two known edges $\mathcal{E}_1$ and $\mathcal{E}_2$, their depths are known. Therefore, following the depth propagation principle (Property 1), the depth of every point along $\mathcal{E}_p$ may be computed, in particular that of the point $p$.

A direct consequence of that property is that the knowledge of the depth at three distinct points in the image is enough to recover the entire scene depth map. Therefore, fixing three scalar parameters (the three depths) is sufficient to retrieve a complete Euclidean reconstruction of the scene (of course restricted to the scanable areas). Notice that we have not yet shown that this is a necessary condition. In other words, there could exist an alternative scanning strategy that requires only 2 or less scalar identifications in order to achieve the same Euclidean reconstruction. Below we show that the condition is also necessary.

Notice that the basic constraint that we used in order to propagate depth information from $\{a, b, c\}$ to $p$ made direct use of the intersecting points $p_1$ and $p_2$ of the shadow edges. As the scanning progresses, more edges are projected on the scene, generating more and more intersections. In fact, while approaching the end of the scanning procedure, it is very likely to find a lot more than two intersecting points per edge. Therefore, in presence of noise in the measurements, this direct constructive method may not be the most robust technique to propagate depth information across the image through the edge-web (defined as the entire set of edges). A better algorithm exists in order to make appropriate use of all the edge intersections at once. It will be presented in Section 2.3.

An edge, $\mathcal{E}$, is an isolated edge if and only if it does not intersect with at least two other edges on the image. Notice that depth information cannot possibly be propagated to any isolated edge from the rest of the edge-web. Therefore, every isolated edge should be rejected prior to depth computation.

2.2 Dual-space formalism

This section gives a short description of the dual-space geometry, the mathematical framework that will be used for deriving the main reconstruction algorithm (in Sec. 2.3). Let $(E) = \mathbb{R}^3$ be the 3D Euclidean space. A plane $\Pi$ in $(E)$ is uniquely represented by the 3-vector $\mathbf{w} = [\omega_x, \omega_y, \omega_z]^T$ such that any point $P$ of coordinate vector $\mathbf{X}_C = [X_C, Y_C, Z_C]^T$ (expressed
in the camera reference frame) lies on $\Pi$ if and only if $\langle \omega, X_c \rangle = 1$ ($\langle \cdot, \cdot \rangle$ is the standard scalar product operator). Notice that $\omega = \pi/d$ where $\pi$ is the unitary normal vector of the plane and $d \neq 0$ its distance to the origin. Let $\Theta = \mathbb{R}^3$. Since every point $\omega \in \Theta$ corresponds to a unique plane $\Pi$ in (E), we refer to $\Theta$ as the dual-space$^*$. 

Similarly, a line $\lambda$ on the image plane is represented by the 3-vector $\lambda$ (up to scale) such that any point $p$ of coordinates $x_c = [x_c \ y_c \ 1]^T$ lies on this line if and only if $\langle \lambda, x \rangle = 0$. See [5, 6]. This $\omega$-parameterization differs from conventional dual space parameterizations (see [3]) in that it does not allow to represent planes crossing the origin. However, that does not constitute a limitation in our application. Furthermore, this new representation exhibits useful properties allowing to naturally relate objects in 3D (planes, lines and points) to their perspective projections on the image plane (lines and points). The following property is one of them.

**Property 3:** Consider two planes $\Pi_a$ and $\Pi_b$ in space, with respective coordinate vectors $\omega_a$ and $\omega_b$ ($\omega_a \neq \omega_b$), and let $\Lambda = \Pi_a \cap \Pi_b$ be the line of intersection between them. Let $\lambda$ be the perspective projection of $\Lambda$ on the image plane, and $\lambda$ its coordinate vector. Then $\lambda$ is parallel to $\omega_a - \omega_b$. In other words, $\omega_a - \omega_b$ is a valid coordinate vector of the line $\lambda$.

Consequently, the coordinate vector $\omega$ of any plane $\Pi$ containing the line $\Lambda$ will lie on the line connecting $\omega_a$ and $\omega_b$ in dual-space $\Theta$. We denote that line by $\Lambda$ and call it the dual image of $\Lambda$. This concept of dual image may be generalized to other geometrical objects [3].

### 2.3 3D reconstruction algorithm

In this section, we derive the complete algorithm for 3D reconstruction from planar shadows. One may find a summary of it at the end of the section.

Let $\Pi_i$ be the $i$th shadow plane generated by the stick ($i = 1, \ldots, N$), with corresponding plane vector $\omega_i = [\omega_{i_1} \ \omega_{i_2} \ \omega_{i_3}]^T$ (in dual space). For all vectors $\omega_j$ to be well defined, it is required that none of the planes $\Pi_i$ contain the camera center $O_c$ (see Section 2.2). Denote by $E_i$ the associated shadow edge observed on the image plane, then the $N$ vectors $\omega_i$ constitute the main unknowns in the reconstruction problem (3N unknown variables). Given the scanning scenario, every shadow plane $\Pi_i$ must contain the light source point $S$. Therefore, denoting by $X_S = [X_{S_1} \ Y_{S_1} \ Z_{S_1}]^T$ the known light source coordinate vector in the camera reference frame, we have for all $i = 1, \ldots, N$, $\omega_i(X_S) = 1$ (see Sec. 2.2). One may then explicitly use that constraint, and parameterize the vectors $\omega_i$ using a two-coordinate vector $\mathbf{v}_i = [u_{i_1} \ u_{i_2}]^T$ such that:

$$\omega_i = \mathbf{W} \mathbf{v}_i + \omega_0 = \begin{bmatrix} \omega_{a1} & \omega_{a2} \end{bmatrix} \mathbf{v}_i + \omega_0 \quad (1)$$

where $\omega_0$, $\omega_{a1}$, and $\omega_{a2}$ are three vectors defining the parameterization. For example, if $X_S \neq 0$, one may keep the last two coordinates of $\omega_i$ as parameterization: $u_i = [u_{i_1} \ u_{i_2}]^T$, picking $\omega_{a1} = [-Y_S/X_S \ 1 \ 0]^T$, $\omega_{a2} = [-Z_S/X_S \ 0 \ 1]^T$ and $\omega_0 = [1/X_S \ 0 \ 0]^T$. Any other choice of linear parameterization is acceptable (there will always exist one given that $S \neq O_c$).

In order to define a valid coordinate change, the three non-zero vectors $\omega_{a1}$, $\omega_{a2}$, and $\omega_0$ must only satisfy the three conditions (a) $\langle \omega_{a1}, X_S \rangle = 1$, (b) $\langle \omega_{a2}, X_S \rangle = \langle \omega_{a2}, X_S \rangle = 0$, (c) $\omega_{a1} \neq \omega_{a2}$. After that parameter reduction, the total number of unknown variables reduces to $2N$: two coordinates $u_{i_1}$ and $u_{i_2}$ per shadow plane $\Pi_i$. Given that reduced plane vector parameterization (called $\mathbf{v}_i$-parameterization), let us derive the analytical basis of the global reconstruction algorithm.

As described in the previous section, the only elements that let depth information propagate from edge to edge are the intersecting points between the edges themselves. These points provide the only geometrical constraints that may be extracted from the images. Therefore, the first step consists of studying the type of constraint provided by an elementary edge intersection.

Assume that the two edges, $E_{n}$ and $E_{m}$, intersect at a point $p_k$ on the image ($n \neq m$), and let $\Pi_n$ and $\Pi_m$ be the two associated shadow planes with coordinate vectors $\omega_n$ and $\omega_m$ (see Figure 3). Let $\pi_k$ be the homogeneous coordinate vector of $p_k$ on the image plane, and $Z_k$ the depth of the corresponding point $P_k$ in the scene. Then, the two edges $E_n$ and $E_m$ intersect in space at $P_k$ if and only if the planes $\Pi_n$ and $\Pi_m$ and the scene surface intersect at a unique point in space $(P_k)$. Equivalently, the depth $Z_k$ may be computed by triangulation using either plane $\Pi_n$ or $\Pi_m$. This condition translates into the two constraint equations $Z_k = 1/\langle \omega_n, \pi_k \rangle = 1/\langle \omega_m, \pi_k \rangle$ (standard triangulation.

![Figure 3: Elementary edge intersection](image-url) The point $p_k$ lies at the intersection of the two edges $E_n$ and $E_m$ on the image plane. What does $p_k$ tell us about the corresponding shadow planes $\Pi_n$ and $\Pi_m$?
equation in dual-space) that may be written in the form of a single equation not involving the depth $Z_k$:
\[
\langle \bar{x}_k, \bar{w}_n - \bar{w}_m \rangle = 0
\] (2)

This unique equation captures all the information that is contained in an elementary edge intersection. There is a very intuitive geometrical interpretation of that equation: Let $\lambda_k$ be the line of intersection between the two planes $\Pi_n$ and $\Pi_m$ in space, and let $\lambda_k$ be the perspective projection of that line onto the image plane. Then, the vector $\bar{\lambda}_k = \bar{w}_n - \bar{w}_m$ is one coordinate vector of the line $\lambda_k$ (see Property 3 in Sec. 2.2). Therefore, (2) is merely $\langle \bar{x}_k, \bar{\lambda}_k \rangle = 0$, which is equivalent to enforcing the point $p_k$ to lie on $\lambda_k$ (Figure 3).

Equation 2 has both advantages of (a) not explicitly involving the depth $Z_k$ (which may be computed afterwards from the shadow plane vectors) and (b) being linear in the plane vector unknowns $\bar{w}_n$ and $\bar{w}_m$. The same constraint may also be written as a function of $\bar{u}_n$ and $\bar{u}_m$, the two $u$-parameterization vectors of the shadow planes $\Pi_n$ and $\Pi_m$:
\[
\langle \bar{y}_k, \bar{u}_n - \bar{u}_m \rangle = 0
\] (3)

where $\bar{y}_k = \mathbf{W}^T \bar{x}_k$ (a 2-vector). Notice that this new equation remains linear and homogeneous in the reduced parameter space.

Let $N_p$ be the total number of intersection points, $p_k, (k = 1, \ldots, N_p)$, existing in the edge-web (the entire set of edges). Assume that a generic $p_k$ lies at the intersection of the two edges $E_{n(k)}$ and $E_{m(k)}$ ($n(k)$ and $m(k)$ are the two different edge indices). The total set of constraints associated to the $N_p$ intersections may then be collected in the form of $N_p$ linear equations:
\[
\forall k = 1, \ldots, N_p, \quad \langle \bar{y}_k, \bar{u}_{n(k)} - \bar{u}_{m(k)} \rangle = 0
\] (4)

which may also be written in a matrix form:
\[
\mathbf{A} \bar{U} = \mathbf{0}_{N_p}
\] (5)

where $\mathbf{0}_{N_p}$ is a vector of $N_p$ zeros, $\mathbf{A}$ is an $N_p \times 2N$ matrix (function of the $\bar{y}_k$ coordinate vectors only) and $\bar{U}$ is the vector of reduced plane coordinates (of length $2N$): $\bar{U} = [\bar{u}_{11}^T \quad \bar{u}_{12}^T \quad \cdots \quad \bar{u}_{N_p1}^T \quad \bar{u}_{N_p2}^T \quad \cdots \bar{u}_{N_pN}^T]^T$. The vector $\bar{U}$ will be sometimes denoted $\bar{U} = [\bar{u}_i]_{i=1,\ldots,N}$. According to Eq. 5, the solution for the shadow plane vector lies in the null space of the matrix $\mathbf{A}$. It is therefore essential to identify the rank of that matrix or equivalently the dimension of its null space. Notice that the number of intersection points grows faster than the number of edges in the image. Essentially, the condition $N_p \geq 2N$ is not too demanding.

**Definition 2 - Fully connected edge-web:** The edge-web is fully connected if and only if it cannot be partitioned into two groups of edges which have less than two points in common. In particular, a fully connected edge-web does not contain any isolated edge. Notice that under that condition only, depth information can freely propagate through the entire web following the constructive approach described in Section 2.1. A normal scanning scenario is then defined as a scenario where the edge-web is fully connected and the total number of intersections is larger than $2N$ (this last condition will be relaxed later).

**Theorem 1:** In a normal scanning scenario, the rank of the matrix $\mathbf{A}$ is exactly $2N - 3$ (or alternatively, the null space of $\mathbf{A}$ is of dimension 3).

**Proof:** We presented in Section 2.1 a constructive method that allows to compute the entire geometry of the scene based on the depth at three distinct points (from propagation of depth information from edge to edge). Therefore, in the case of a normal scanning scenario, the reconstruction problem has at most three free parameters. Consequently the dimension of the null space of $\mathbf{A}$ is at most 3, or equivalently, $\mathbf{A}$ is of rank at least $2N - 3$.

Consider now the two-dimensional linear subspace $S$ of vectors $\bar{U}$ that have the following form: $\bar{U} = \bar{U}_{\alpha,\beta} = [\alpha \beta \alpha \beta \cdots]^T ((\alpha, \beta) \in \mathbb{R}^2)$. It is straightforward to show that for any value of $\alpha$ and $\beta$, the vector $\bar{U}_{\alpha,\beta}$ lies in the null space of $\mathbf{A}$: $\mathbf{A} \bar{U}_{\alpha,\beta} = \mathbf{0}$. Therefore, $S$ is included in the null space of $\mathbf{A}$ which is therefore of dimension at least 2. Equivalently, the rank of $\mathbf{A}$ is less or equal than $2N - 2$.

Finally, the null space of $\mathbf{A}$ cannot be restricted to $S$, otherwise, the only solutions to the problem would reduce to sets of identical shadow planes in space ($\bar{u}_i = [\alpha \beta]^T, \forall i = 1, \ldots, N$), which is impossible in practice. Therefore, the dimension of the null-space of $\mathbf{A}$ must be greater than two (in order to allow for distinct shadow planes) leading to the rank of the matrix being at most $2N - 3$. Therefore the rank of $\mathbf{A}$ is exactly $2N - 3$.

A direct consequence of Theorem 1 is that, no matter which strategy one adopts in solving for the set of constraints, there will always be three free parameters to set in order to achieve Euclidean reconstruction. In addition, since the linear system is rank $2N - 3$, only a minimum of $2N - 3$ intersection points in the edge-web is needed (and not $2N$). It is straightforward to show that this condition is always automatically satisfied if the edge-web is fully connected. Therefore, a normal scanning scenario may be re-defined as a scenario in which the edge-web is fully connected. Notice that Figure 2 shows the minimum configuration for scanning: $N = 3$ and $N_p = 2N - 3 = 3$. The following corollary is another important consequence of Theorem 1.

**Corollary 1:** Let $\bar{U}' = \{\bar{u}_i\}$, $(i = 1, \ldots, N)$ be a non-trivial solution for the linear system (5) (by non-trivial, we mean a solution vector such that at least two vectors $\bar{u}_i$ and $\bar{u}_j$ are distinct for some $i \neq j$). Then, for every solution vector $\bar{U} = \{\bar{u}_i\}$ to Equation 5, there
exists three scalars $\alpha$, $\beta$ and $\gamma$ such that:
\[
\forall i = 1, \ldots, N, \quad \mathbf{u}_i = \gamma \mathbf{u}_i^0 + \mathbf{u}_0
\]
with $\mathbf{u}_0 = [\alpha \ \beta]^T$. Conversely, for any scalars $\alpha$, $\beta$ and $\gamma$, the vector $\mathbf{U} = \{\mathbf{u}_i\}$, given by Equation 6, is a solution of the linear system (5). The vector $\mathbf{U}^0$ is called a `seed' solution from which all solutions of the linear system may be identified. Notice that not all these solutions lead to the exact Euclidean reconstruction of the scene, only one of them does (corresponding to one particular set of coefficients $\alpha$, $\beta$ and $\gamma$). However, one seed vector enables to identify explicitly the whole set of possible solutions, each one of them having a valid 3D scene interpretation agreeing with the set of observed shadows. The following part of the paper will focus on that seed vector identification.

First, any non-trivial solution vector $\mathbf{U}^0 = \{\mathbf{u}_i^0\}$ may be used as seed (as long as it is one solution of (5)), let us pick the one (called unitary seed vector) that satisfies the two extra normalizing conditions: (a) $\sum_{i=1}^{N} \mathbf{u}_i^0 = 0$ (zero mean), (b) $\sum_{i=1}^{N} ||\mathbf{u}_i^0||^2 = 1$ (unit norm). Those conditions assure a non-trivial solution (all $\mathbf{u}_i$ cannot be identical). The resulting unitary seed vector $\mathbf{U}^0$ satisfies the linear equation $\mathbf{B} \mathbf{U}^0 = 0_{N,2}$, where $\mathbf{B}$ is the following augmented $(N_p + 2) \times 2N$ matrix:

\[
\mathbf{B} = \begin{bmatrix}
A & 0 & 1 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix}
\]

The two last rows of $\mathbf{B}$ enforce the zero-mean constraint, bringing the rank of $\mathbf{B}$ to $2N - 1 (= 2(N - 3) + 2)$. Therefore, the dimension of the null space of $\mathbf{B}$ is one, leading to $\mathbf{U}^0$ being the unitary eigenvector associated with the unique zero eigenvalue of $\mathbf{B}$. Consequently, the singular value decomposition of $\mathbf{B}$ may be used to retrieve $\mathbf{U}^0$. Such a decomposition leads to the following relation:

\[
\mathbf{B} = \mathbf{U} \mathbf{S} \mathbf{V}^T
\]

where $\mathbf{U}$ and $\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_{2N} \end{bmatrix}$ are two unitary matrices of respective sizes $(N_p + 2) \times (N_p + 2)$ and $2N \times 2N$, and $\mathbf{S}$ is the $(N_p + 2) \times 2N$ matrix of singular values. The unitary seed vector $\mathbf{U}^0$ is then the column vector of $\mathbf{V}$ associated to the zero singular value. Without loss of generality, assume it is the last column vector: $\mathbf{U}^0 = \{\mathbf{u}_i^0\} = \mathbf{V}_{2N}$. Alternatively, one may retrieve the same $\mathbf{V}$ matrix by applying the same decomposition to the smaller $2N \times 2N$ symmetric matrix $\mathbf{C} = \mathbf{B}^T \mathbf{B}$. Such a matrix substitution is advantageous because the so-defined matrix, $\mathbf{C}$, has a simple block structure: $\mathbf{C} = [\mathbf{C}_{i,j}]_{i,j \leq N}$, where each matrix element $\mathbf{C}_{i,j}$ is of size $2 \times 2$. Let us derive a closed form expression for $\mathbf{C}_{i,j}$. Since two shadow planes can only intersect once (two shadow planes intersect along a line that can only intersect the scene at a single point), for a given pair of shadow indices $(i,j)$ there exists at most one index $k$ such that $p_k = E_i \cap E_j$. This mapping will be denoted $k = k(i,j)$ (notice, for example, that $k(i, j) = k(j, i)$ and $k(n(k'), m(k')) = k'$). If the two edges do not intersect (such as an edge with itself), or if that intersection is not visible on the image plane, then the point $p_{k(i,j)}$ does not exist. However, we will still refer to it as a phantom point of coordinate vector $\mathbf{x}_{k(i,j)} = [0 \ 0 \ 0]^T$ (leading to $\mathbf{y}_{k(i,j)} = \mathbf{W}^T \mathbf{x}_{k(i,j)} = [0 \ 0]^T$). Adopting that formalism, all the following algebraic equations remain valid even in the case of missing intersections. Given this notation, it may be shown that all matrices $\mathbf{C}_{i,j}$ have the following expressions:

\[
\mathbf{C}_{i,j} = I_2 - \mathbf{y}_{k(i,j)} \mathbf{y}_{k(i,j)}^T \quad \text{if } i \neq j
\]

\[
\mathbf{C}_{i,i} = I_2 - \sum_{n=1}^{N} \mathbf{y}_{k(n,i)} \mathbf{y}_{k(n,i)}^T
\]

where $I_2$ is the $2 \times 2$ identity matrix. Observe that every off-diagonal matrix element $\mathbf{C}_{i,j}$ ($i \neq j$) depends only on the intersection point $p_{k(i,j)}$ between edges $E_i$ and $E_j$ (Equation 9). Every diagonal block $\mathbf{C}_{i,i}$ however is function of all the intersection points of $E_i$ with the rest of the edge-web (the sum is over all points $p_{k(n,i)}$, for $n = 1, \ldots, N$).

Once the matrix $\mathbf{C}$ is built, $\mathbf{V}$ is retrieved by singular value decomposition. This technique allows then for a direct identification of the unitary seed solution $\mathbf{U}^0 = \{\mathbf{u}_i^0\}$ leading to the set of all possible solutions of (5) (following Equation 6 in Corollary 1). Euclidean reconstruction is thus achieved up to the three parameters $\alpha$, $\beta$ and $\gamma$.

From this analysis, one may question the optimality of such an estimation scheme in presence of noise in the measurements (intersection coordinates $\mathbf{x}_i$). Essentially, in a noisy situation, the matrix $\mathbf{B}$ becomes full rank (the smallest singular value is not exactly zero), and then SVD really finds the (non-trivial) unit length vector $\mathbf{U}^0$ that minimizes the square norm of $\mathbf{A} \mathbf{U}^0$. The question is really about the geometrical meaning of such a cost minimization. Let $\mathbf{U} = \{\mathbf{u}_i\}$ be the real set of shadow plane coordinates corresponding to Euclidean reconstruction. Then, according to Corollary 1, there exists a unique set of coefficients $\alpha$, $\beta$ and $\gamma$ such that: $\mathbf{u}_i = \gamma \mathbf{u}_i^0 + \mathbf{u}_0$ with $\mathbf{u}_0 = [\alpha \ \beta]^T$, for all $i = 1, \ldots, N$. Those three global scalar parameters upgrade the reconstruction to Euclidean. Then, the full plane coordinate vectors $\mathbf{v}_i$ have the expression $\mathbf{v}_i = \mathbf{W} \mathbf{u}_i + \mathbf{w}_0$ (from Equation 1). The depth $Z_k$ of a given intersection point $p_k$ may now be computed from either planes $P(k)$ or $P_m(k)$ leading, in an ideal noiseless scenario, to the same estimates. In the presence of noise, however, those two quantities may
be different. Let us denote them by $Z_k^{(1)}$ and $Z_k^{(2)}$:

$$Z_k^{(1)} = \frac{1}{\gamma (\langle \tilde{w}_{m(k)} \tilde{y}_k \rangle + \langle \tilde{w}_o, \tilde{y}_k \rangle) + \langle \tilde{w}_o, \tilde{w}_k \rangle}$$

$$Z_k^{(2)} = \frac{1}{\gamma (\langle \tilde{p}_{m(k)} \tilde{y}_k \rangle + \langle \tilde{w}_o, \tilde{y}_k \rangle + \langle \tilde{w}_o, \tilde{w}_k \rangle)}$$

The inverse depth difference is a direct function of the seed vector coordinates: $c_k = 1/Z_k^{(1)} - 1/Z_k^{(2)} = \gamma (\tilde{y}_k, \tilde{p}_{m(k)} - \tilde{p}_{o})$. This may also be written in a matrix form: $\tilde{c} = \gamma A \tilde{U}^o$, where $\tilde{c} = [c_1 c_2 \ldots c_N]^{T}$. Therefore, independently from the global parameters $\alpha$, $\beta$ and $\gamma$, the least squares solution given by SVD minimizes the 2-norm of $\tilde{c}$. Consequently, this estimation scheme is optimal in the sense of minimizing the sum, over all intersection points, of the square of the inverse depth errors. This cost function may not exactly achieve the goal of reconstructing the “best” possible three-dimensional model of the scene. For example, a more natural cost function would directly involve the depth differences $Z_k^{(2)} - Z_k^{(1)}$ rather than the inverse depth differences. However, the main difficulty about using such a cost function comes from the fact that the quantity $Z_k^{(2)} - Z_k^{(1)}$ remains a function of the three scalars $\alpha$, $\beta$ and $\gamma$ that are not known until three additional geometrical constraints are enforced, such as the depth at three points.

Once the seed solution $\tilde{U}^o = \{\tilde{p}_i\}$ is found (by SVD), one may identify the final “Euclidean” solution $\tilde{U} = \{\tilde{p}_i\}$ if the depth of (at least) three points in the scene are known. Without loss of generality, assume that these points are $p_k$ for $k = 1, 2, 3$ (with depths $Z_k$). Those points provide then three linear equations in the unknown coefficient vector $\tilde{c} = [\alpha \ \beta \ \gamma]^{T}$:

$$\begin{bmatrix}
\tilde{y}_k^{T} \\
\langle \tilde{w}_{m(k)} \tilde{y}_k \rangle
\end{bmatrix} \ \tilde{c} = 1/Z_k - \langle \tilde{w}_o, \tilde{w}_k \rangle$$

(11)

for $k = 1, 2, 3$ resulting in a linear system of three equations and three unknowns. This system may then be solved, yielding the three coefficients $\alpha$, $\beta$ and $\gamma$, and therefore the final solution vector $\tilde{U}$ (through Eq. 6). Complete Euclidean shape reconstruction is thus achieved. If more points are used as initial depths, the system may be solved in the least squares sense (once again optimal in the inverse depth error sense). Notice that the reference depth points do not have to be intersection points as Eq. 11 seems to infer. Any three points in the edge-web may be used.

Finally, the proposed method for 3D reconstruction may be summarized into five successive steps:

**Step 1**: Acquire a set of shadow images, extract the shadow edges and compute their intersections.

**Step 2**: Reject all isolated edges (and isolated groups of edges) so that the entire edge-web is fully connected (Def. 2). Obtain a set of $N$ edges $E$, and $N_p$ intersection points $p_k$.

**Step 3**: Build the $2N \times 2N$ matrix $C$ (Eqs 9 and 10), and compute the unitary seed vector $U^o$ by SVD. Euclidean reconstruction is then achieved up to the three scalars $\alpha$, $\beta$ and $\gamma$.

**Step 4**: Select (at least) three points in the scene with known depths, and solve linearly for the remaining scalars $\alpha$, $\beta$ and $\gamma$ (Eq. 11).

**Step 5**: Compute the list of shadow plane vectors $\Pi_i$ (Eq. 6 and 1) and triangulate all the points in the edge-web. The resulting set of 3D points may then be triangulated into a surface mesh for visualization purposes (Figs 4 and 5).

### 3 Experimental Results

Figures 4 and 5 show experimental results obtained from two real scenes. The first one consists of two parallel planes 5cm apart, and the second one a small object (a moon) on a plane. In both experiments, first the seed solution vector $U^o$ was computed by SVD (following steps 1 to 3 of the method) and then $\alpha$, $\beta$ and $\gamma$ were recovered using the three known depths at the three circled points on the figures. For that purpose, the background plane was calibrated in both cases, but only to recover the depth at the three reference points (the entire background planes were not used for scanning).

There are different ways to assess reconstruction accuracies. The first one consists in looking at the depth errors $Z_k^{(1)} - Z_k^{(2)}$ at the intersection points $p_k$ relative to their absolute depths in the camera reference frame. In both experiments, this error is approximately 3mm (in standard deviation) over an average scene depth of 25cm. This leads to a relative depth error of 1.2%. However, in modeling applications, a more relevant quantity to look at is the reconstructed surface roughness relative to the size of the object of interest. In that case, this error is approximately 3mm over object sizes of 5 to 10cm. In addition to surface roughness, it is essential to check for possible global distortions in the final reconstructed scene. For that purpose, one may quantify how well a number of intrinsic geometrical properties of the scene are preserved after reconstruction. Plurality of planes, angles between planes, or size of objects are typical samples of such properties. For example, in both experiments, the reconstructed planes deviate from planarity by approximately 4mm (in maximum). This concerns both planes of Scene 1 and the background plane of Scene 2. The height of the top plane of scene 1 is estimated after reconstruction to 46cm±3mm (the error accounts for the surface roughness), its real value being 5cm. Finally, in the first scene reconstruction, parallelism of the two planes is recovered within approximately 3 degrees of error.

One may notice that the reconstruction accuracies achieved on those two scenes are not as good as the ones achieved when using the original shadow scan-
Figure 4: Experiment 1 - Two planes scene: Top row: The initial scene with a shadow projected on it and the total set of $N = 26$ shadow edges generating $N_p = 173$ intersection points. Bottom row: Two views of the final 3D reconstruction (in the form of a mesh). The processed images were $320 \times 240$ pixels.

Figure 5: Experiment 2 - Luna scene: A total number of $N = 122$ shadow edges are intersecting at $N_p = 3156$ points.

The scanning technique that we presented in [2] (the previous method achieves surface errors of the order of 0.1 mm in standard deviation on similar scenes). The main reason for this is that in this present method, the shadow edge was extracted on the image through spatial processing (based on image gradient) instead of temporal processing as in the method presented in [2]. This illustrates the fact that temporal processing is more reliable than spatial processing because it is a lot less sensitive to changes to surface albedo and occlusions. This was originally demonstrated by Curless and Levoy in [4]. Nevertheless, it is not strictly possible to compare reconstruction accuracies of the two shadow scanning methods, given that we used here very few shadow edges for shape estimation (an order of $N = 100$ edges) while in the previous shadow scanning technique [2], an order of 700 to 1000 images are often necessary to achieve good quality reconstructions.

4 Conclusion and Future Work

We have presented a linear closed form solution for 3D scene geometry recovery from planar shadows. The method is composed of two fundamental stages. The first one consists in retrieving the Euclidean scene geometry up to three scalar unknowns using only the information contained in the observed shadow edges on the image plane. The solution to this problem reduces to a singular value decomposition of a matrix that is only function of the edge intersection coordinates. In the second stage, the three remaining unknowns are computed making use of the known depths at only three points in the scene. Dual-space geometry provides an appropriate framework for carrying the complete mathematical analysis elegantly and intuitively. As part of future work, we intend to carry out a sensitivity analysis of the method, and study alternate geometrical clues for achieving Euclidean reconstruction (other than direct depth measurements at three points). In addition, we intend to merge this reconstruction technique with the one presented in [2] for achieving best local surface reconstruction qualities (taking advantage of temporal processing) while dealing with scenes that do not contain any reference surface.

References


