Dense Shape Reconstruction of a Moving Object under Arbitrary, Unknown Lighting

Denis Simakov, Darya Frolova and Ronen Basri*
Dept. of Computer Science and Applied Math
The Weizmann Institute of Science
Rehovot 76100, Israel

Abstract

We present a method for shape reconstruction from several images of a moving object. The reconstruction is dense (up to image resolution). The method assumes that the motion is known, e.g., by tracking a small number of feature points on the object. The object is assumed Lambertian (completely matte), light sources should not be very close to the object but otherwise arbitrary, and no knowledge of lighting conditions is required. An object changes its appearance significantly when it changes its orientation relative to light sources, causing violation of the common brightness constancy assumption. While a lot of effort is devoted to deal with this violation, we demonstrate how to exploit it to recover 3D structure from 2D images. We propose a new correspondence measure that enables point matching across views of a moving object. The method has been tested both on computer simulated examples and on a real object.

1. Introduction

This paper addresses the problem of shape reconstruction from multiple views of a moving object. We work directly with image intensities and produce a dense reconstruction. The illumination of the scene can be arbitrary but the same for all views, and our method does not rely on any information about light distribution. We do require prior knowledge of object motion (which is our primary source of information), along with either epipolar geometry (for matching) or camera calibration (for shape recovery). The object is assumed to be (mainly) matte and piecewise smooth, and the camera(s) to be orthographic.

*Research was supported in part by the Israeli Science Foundation, Grant No. 266/02 and by the European Commission Project IST-2000-26001 VIBES. The vision group at the Weizmann Inst. is supported in part by the Moross Foundation.

We concentrate here on developing a method for dense shape reconstruction with *known* motion and camera calibration. The last assumption is not very restrictive, since a few point correspondences are enough to estimate the required parameters. These small number of correspondences can be obtained either manually or by tracking a few prominent features that remain distinguishable even under changes of object orientation. We use this technique to perform an experiment with a real object (Section 6).

The main difficulty in our approach is to find a reliable way to match image intensities of the same object point in different views. Commonly, these intensity values are assumed constant. (This is referred to as the *brightness constancy* assumption). However, when the object changes its orientation with respect to the light sources, this assumption certainly does not hold, even in the simple case of completely diffuse (Lambertian) object. A major contribution of this work is the new *correspondence measure* we developed, which allows to find corresponding points when the brightness constancy assumption is violated. In fact, we show how to exploit this violation for our needs.

We propose to match points relying on a source of information about the object that is not exploited by the usual stereo matching techniques. Common stereo methods that exploit brightness constancy need different points on the object to look different, so that they can be distinguished. Therefore they need the surface of the object to be textured. Moreover, if different object points look the same, no matter how many views are available, they cannot be distinguished. We instead rely on the way the intensity of points changes as the object moves. Thus, our approach benefits from rich shape variations, not from texture variations. Every additional view provides extra information about the orientation of a surface element, thus making reconstruction more robust. It is interesting to compare this last observation with the previously noted complementarity of stereo and shading [3, 8]: the former performs well in textured regions, while the latter is better in smooth regions.

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

The paper is divided as follows. Section 2 describes some relevant previous work. Our shape reconstruction method is described in Sections 3, 4, and 5. Experiments demonstrating the applicability of our method are presented in Section 6.

2. Previous Work

Brightness constancy is a common underlying assumption in methods of finding correspondence between views, and much attention has been devoted to relaxing it. We mention several works that deal with this problem.

A common approach in *optical flow* estimation is to find a reasonable generalization of the brightness constancy constraint [16, 10, 2]. Black and Fleet [2] concentrate on building a robust statistical framework that can incorporate brightness variations. Negahdaripour [16] considers a general (linear) brightness variation model with additional constraints on its parameters (usually, smoothness). Haussecker and Fleet [10] consider specific physical models of changes in image appearance that occur due to, among others, object motion. All these methods consider infinitesimal changes and tend to end up with linearized models.

In the field of *tracking* several papers attempt to deal with brightness changes of tracked features, for example [9, 11]. Similar to the optical flow papers cited above, Jin et al. [11] model reflectance by invoking a physical model and then treating parameters as arbitrary changing, loosing their physical meaning. Hager and Belhumeur [9] cope with brightness changes by using a model of the object, which consists of a set of basis images computed offline from a set training images in which the object is stationary and illumination varies. Like them, we model the appearance of objects using low-dimensional linear representation, but work with the theoretically derived basis ([1, 17]) whose changes due to object motion we can predict.

[13, 8, 6] show important similarities with our approach. We shortly describe them below. In contrast to our work, these papers assume known lighting.

Lu and Little [13] recover shape and reflection properties of an object using its rotation. They have however much more constrained experimental setup than we do: they work with the light source in the direction of the camera, assume uniform reflection of the object and rely on the ability to control object motion.

Fua and Leclerc [8] present a surface reconstruction method that works in the common stereo setup (stationary object), but use a reflectance model to recover shape in non-textured regions (where usual stereo matching fails). Brightness is treated as a *shading* cue: spatial intensity variations impose constraints on shape variations along the object surface. We instead extract shape information from the way the object changes its appearance due to motion. The lighting model that they use (ambient light and single directional source) gives a good approximation for an arbitrary light ([1, 17]), although the authors were obviously unaware of this fact. This light model coincides with ours in a particular case.

Carceroni and Kutulakos [6] consider motion coupled with a reflectance model, similar to our approach. The entity they work with is dynamic *surfel* (*surf*ace *el*ement) which incorporates all relevant information about a small surface patch: shape, reflectance parameters and instantaneous motion. Surfel parameters are recovered by an optimization procedure that is quite involved and computationally heavy. They work with continuous motion, assumption that we do not require.

Finally, we want to mention a few papers that also exploit object motion to recover its shape under unknown lighting. Mukawa [15] models reflectance to explain brightness change due to motion, taking into account the influence of temporal variations on surface orientation (in common with our work), while ignoring the influence of spatial variations (surface orientation should vary slowly). Zhang et al. [19] recover the shape of a moving object explicitly using a reflectance model (by incorporating it into the optical flow subspace constraints). They utilize additional information present in object motion to recover shape even in untextured regions. The motion between views is assumed to be small.

The "geotensity" work of Maki, Watanabe and Wiles [14] is especially close to ours. They also address the problem of shape reconstruction of an object undergoing arbitrary motion under unknown lighting. Assuming that the rotating object is illuminated by a single point light source, they argue that the space of images is low dimensional, and use precomputed coefficients of expansion to constrain intensity values for every tentative 3D point.

3. Outline of Our Approach

We implement the following general approach to shape reconstruction. We work in a 3D space of possible points, and seek to extract a 2D surface that represents the solution. In the case of Euclidean shape reconstruction this 3D space can be the space of world points, and the extracted surface is the object surface. In the case when the calibration parameters are unknown (but the epipolar geometry is known) two dimensions represent the x, y coordinates of the point in the reference image plane, and the 3rd dimension is used to specify the corresponding point along the epipolar line, which is one-dimensional. The main requirement is that we should know how to project any point from this workspace onto all images.

We need to separate points in the workspace into two classes: points that belong to the object surface and points that do not. This is done by measuring consistency of points

with available input data. We define a *correspondence measure*, a function on the workspace that should be small for points on the object surface and large for points that are far from the surface. Under the brightness constancy assumption a natural correspondence measure is the variation of intensity values of the projections of the point. We seek to replace this measure by a measure that takes into account brightness changes due to motion.

To model reflectance we use results due to Basri and Jacobs [1] and Ramamoorthi and Hanrahan [17]. They justify previous empirical observations that the set of images of a Lambertian object under arbitrary, distant lighting is accurately approximated by a low-dimensional subspace. What is of crucial importance to us, is that they also provide an explicit expression for computing a basis for this subspace. We show that a correspondence measure based on this reflectance model indeed facilitates reliable reconstruction.

Once we define our correspondence measure we use the method developed by Boykov, Veksler and Zabih [5] to extract the sought surface. The problem is posed as an energy minimization problem: find a surface with minimal total energy. The energy functional incorporates constraints imposed by the input data, as well as prior information about the surface (usually some form of smoothness). Boykov, Veksler and Zabih achieve a good compromise between computational tractability and quality of solution: their algorithm is fast and, while it is not guaranteed to find the global optimum, successfully computes a generalized local optimum that is within a proved bound from the global optimum. They use a graph cut technique to obtain computationally efficient procedure. The algorithm accepts fairly general smoothness priors including ones that are robust and allow discontinuities in the surface. It also shows very good performance in practice, a fact that we also witness.

To summarize, this work uses a way of modelling object reflectance under general lighting conditions introduced in [1, 17] together with efficient energy minimization techniques [5] to solve a shape reconstruction problem using the motion of the object.

We provide details of our method below. Section 4 describes how we build a *correspondence measure* that allows matching of different projections of surface points. Section 5 gives a brief description of the algorithm that utilizes the information provided by the *correspondence measure* to find the most likely object surface given the input data. For simplicity, we consider the case of Euclidean shape reconstruction. All the derivations below can easily be adjusted to handle the case of an uncalibrated camera.

4. Correspondence measure

A *correspondence measure* is some function $\mu(\vec{P})$ defined on the (3D) workspace that for every point assigns a numerical value. This value is expected to be low for points on the true object surface and high for all other points. We consider a model that describes how image intensity changes with surface orientation. To build our measure we will put the available geometric information (how the object moves) and image data into this model. We will expect our data to fit the model quite accurately for real world points that are on the surface of the object.

4.1. Preliminaries

Reflectance model

We use here results due to Basri and Jacobs [1] and Ramamoorthi and Hanrahan [17], which found and justified a way of modeling the imaging of Lambertian objects under arbitrary distant lighting. They showed that one can obtain accurate approximations to the images by decomposing them into a small number of spherical harmonics, evaluated on the surface normals.

For approximation of order K we use harmonics of orders $k = 0, 1, \ldots, K$. There are $2k + 1$ harmonics for each k, so overall we have $(K + 1)^2$ basis vectors[1]. Let $\vec{l}$ be a $(K + 1)^2$-dimensional vector that represents light configuration ($\vec{l}$ consists of the spherical harmonic coefficients of the light function.) Then, the image intensity of a surface point $\vec{P}$ with normal $\vec{n}$ and albedo ρ is approximated by:

$$I \approx \rho \vec{l}^{\,T} \cdot \vec{Y}(\vec{n}). \tag{1}$$

Here the $(K + 1)^2$-dimensional vector $\vec{Y}$ is a known function on the unit sphere. Its components are proportional to the spherical harmonics of orders up to K, and are polynomials of degrees $0, 1, \ldots, K$ on the components of the normal $\vec{n}$.

This approximation appears to be sufficiently accurate already for small orders of approximation. More precisely [7], on average (assuming lighting in any direction is equally likely), the accuracy of this approximation is at least 87.5% for a first order approximation (involving four harmonics) and 99.22% for a second order approximation (nine harmonics). Moreover, empirical evidence suggests that this approximation remains valid even for fairly near lighting [7].

Equation (1) gives a simple low-dimensional model, that accurately describes the dependence of the image on the parameters of the object (albedo and normal) and the environment (lighting vector $\vec{l}$). It is also important to note, that in this model we do not assume any particular light configuration. In particular, this model permits arbitrary collections of point and extended light sources.

[1] Theoretically, under the distant light assumption harmonics of odd orders from order 3 and up do not contribute energy [1, 17] and can be omitted, but empirical study under near lighting implies that they are still worth including [7].

Camera geometry

We will need the following geometric information for our correspondence measure: (a) projection of 3D object points onto all images; (b) change of surface orientation (surface normal) from view to view. The first will be used to extract the intensities of points in all views. The latter will be used to model the change of intensities due to object motion.

Denote by $\vec{P}$ a point on the surface of our object, in some reference frame attached to the object (independent of the view). We assume that for every view j the projection of $\vec{P}$ onto this view is given by the relation:

$$\vec{p}_j = \text{proj}_j(\vec{P}), \tag{2}$$

where $\text{proj}_j(\cdot)$ is a known mapping from 3D space to 2D image coordinates.

Now if $\vec{n}$ is the normal vector at $\vec{P}$, then this normal in the j^{th} view is

$$\vec{n}_j = R_j \vec{n}, \tag{3}$$

where R_j is a 3×3 rotation matrix (we consider rigid transformations of the object). Surface orientations in all views are functions of a single vector $\vec{n}$, which has 2 degrees of freedom (we assume that the rotation matrices R_j are known).

4.2. Definition of the measure

We are now ready to turn to the correspondence measure itself. For a given 3D point $\vec{P}$ on the object surface[2] we can use (2) to find its projections to the view j. Assume that we are given images I_j taken from different views, then the point $\vec{P}$ has intensity $I_j(\vec{p}_j)$ in the view j.

If a Lambertian object does not move, these intensities should be the same (in the absence of noise). In this case several good correspondence measures can be devised, e.g., the variance of $\{I_j(\vec{p}_j)\}_j$, their range, etc.

In our, non-stationary scenario we have to use a more complicated approach. Instead of just brightness constancy we have the model (1). For a point $\vec{P}$ in view j we can write:

$$\rho \vec{l}^{\,T} \cdot \vec{Y}(R_j \vec{n}) \approx I_j(\vec{p}_j). \tag{4}$$

Here $\vec{p}_j$ is a known function of $\vec{P}$ and j (2). We are unable to evaluate this constraint (4) because ρ, $\vec{l}$ and $\vec{n}$ are unknown. We can, however, try to estimate these unknowns from additional views.

We consider equations (4) for some fixed point $\vec{P}$ observed in views $j = 1, \ldots, m$. As we deal with different views of *the same* point, we can simplify our notation: (a) we will write $\vec{l}$ instead of $\rho\vec{l}$, because in this case albedo ρ cannot be distinguished from the intensity of the lighting $\|\vec{l}\|$, (b) we will replace $I_j(\text{proj}_j(\vec{P}))$ (the intensity of the point projected onto view j) simply by I_j . Then we have the following system of equations:

$$\begin{cases} \vec{l}^{\,T}\vec{Y}(R_1\vec{n}) = I_1, \\ \quad\vdots \\ \vec{l}^{\,T}\vec{Y}(R_m\vec{n}) = I_m. \end{cases} \tag{5}$$

Here I_j, R_j and $Y(\cdot)$ are known, $\vec{l}$ and $\vec{n}$ are unknown.

Now, for given image intensities $I_1, \ldots, I_m$ of $\vec{P}$ in all views, we can compute a correspondence measure $\mu(\vec{P})$ in the following way: search for $\vec{l}$ and $\vec{n}$ that satisfy (5) the best, and take this best residual of (5) as our measure:

$$\mu(\vec{P}) = \min_{\vec{l},\vec{n}} (\textit{residual of (5)}). \tag{6}$$

If $\vec{P}$ lies on the object surface, the residual of (5) is guaranteed to be small for the correct values of the unknowns, so $\mu(\vec{P})$ will also be small. We expect that in general for points that are not on the surface, no values of unknowns will give a good fit to the system (5), and $\mu(\vec{P})$ will be large.

In this definition of the correspondence measure we consider every point $\vec{P}$ separately, so our measure will not ensure that the lighting is consistent for all points. Enforcing illumination consistency could potentially improve shape reconstruction, but we cannot do it directly because our workspace contains many more points outside the surface (which give rise to false correspondences) than points on the surface (true correspondences). On the other hand, we do not want to rely on just a few points to estimate the global illumination (as is done, for example, in [14]). If those points violate some of our assumptions (exhibit significant specular reflectance, fall in a cast shadow, etc.), this would deteriorate reconstruction for all points. A robust way of enforcing illumination consistency is a topic of our ongoing research.

4.3. Computation of the measure

In general, for reliable evaluation of the correspondence measure (6) we may need to perform global minimization in the space of $(\vec{l}, \vec{n})$. We want to avoid this because of very high computational cost. However, local optimization, while considerably faster, often gives solutions that are quite far from the global optimum. Below we show how we solve this problem in our case. We start by considering an important practical case.

First order approximation in the reflectance model (see section 4.1) gives us particularly simple relations while achieving reasonable accuracy. In this case spherical harmonics are linear functions and the system (5) becomes bi-

[2]Recall that $\vec{P}$ is in a coordinate system that is local to the object, and is not affected by its motion.

linear:

$$\begin{array}{l} \tilde{l}_0 + \vec{\tilde{l}}^{\,T} R_j \vec{n} = I_j \\ j = 1, \ldots, m. \end{array} \tag{7}$$

Here the scalar $\tilde{l}_0$ and the 3-vector $\vec{\tilde{l}}$ compose the 4-dimensional lighting vector $\vec{l}$ (scaled by the appropriate normalization constants of the spherical harmonics). This approximation in fact models light as a sum of ambient component $\tilde{l}_0$ and a single directional source $\vec{\tilde{l}}$, and, surprisingly, is quite accurate for an arbitrary light distribution.

We can eliminate $\tilde{l}_0$ by subtracting the first equation ($j = 1$) from all the rest. Then for $\vec{\tilde{l}}$ we obtain the following system:

$$\begin{array}{l} \vec{\tilde{l}}^{\,T} (R_j - R_1) \vec{n} = \Delta I_j \\ j = 2, \ldots, m, \end{array} \tag{8}$$

where $\Delta I_j = I_j - I_1$.

Let us consider the case in which an object rotates by small angles. Then, assuming without loss of generality that R_1 is the identity matrix, and exploiting the Rodrigues' formula for rotation matrices, we observe that (8) becomes:

$$-\theta_j \vec{\omega}_j^T \; (\vec{\tilde{l}} \times \vec{n}) = \Delta I_j j = 2, \ldots, m, \tag{9}$$

where $\vec{\omega}_j$ and θ_j are the rotation axis and angle of R_j respectively, and "$\times$" denotes the cross product of vectors.

This means that when we consider small angles the bilinear system becomes linear with respect to the new unknown $\vec{u} = \vec{\tilde{l}} \times \vec{n}$. We therefore have an inherent ambiguity of the solutions: there is a whole plane of normals $\vec{n}$ (and likewise of light $\vec{\tilde{l}}$) that solve our system. Importantly, all these solutions have the same residual.

We can directly use this results, for example, if we capture a moving object at a video rate, when rotations from frame to frame are very small. But it also appears, that this particular case analysis gives a good description of the general case. The ambiguity that arises with the first order harmonic approximation in the case of small angles (9) is observed also for the general system (5) both for fairly large angles (we had rotations by 5–30 degrees in our experiments) and for higher order approximations (we checked the case of second order approximation with 9 harmonics).

This has important implications for our problem of efficient and reliable computation of the correspondence measure (6). Starting from a random initial guess and minimizing the residual of the system (5) by a local iterative optimization technique, we will obtain different solutions that will generally be rather far from the true solution, but will lie on the ambiguity plane determined by (9). (This indeed is what we observe in our experiments). But all the solutions on the ambiguity plane have roughly the same residuals, so this ambiguity will not affect our computation of correspondence measure.

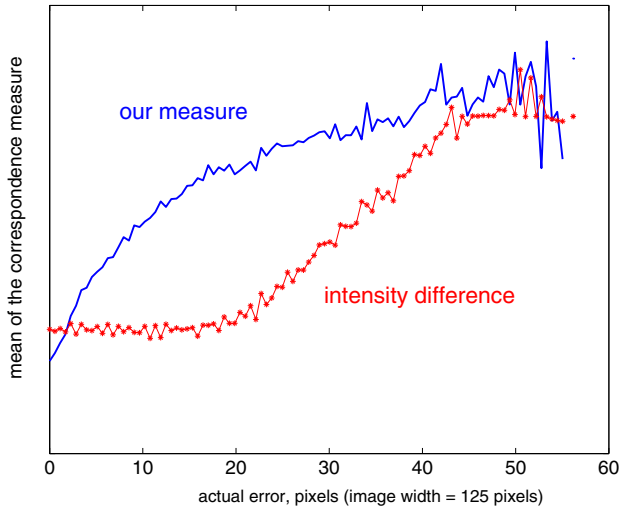


Figure 1. Comparison of correspondence measures. A good measure should increase as the actual error increases.

We can also gain two practical results for our problem of computation of our correspondence measure. First, we can use the residual of the linearized system (9) as the correspondence measure:

$$\mu_{\text{lin}}(\vec{P}) = \|(\Omega(\Omega^T\Omega)^{-1}\Omega^T - 1) \cdot \vec{\Delta I}\|, \tag{10}$$

where the $(m-1) \times 3$ matrix Ω consists of the rows $-\theta_j \vec{\omega}_j^T$ and $\vec{\Delta I} = [\Delta I_2, \ldots, \Delta I_m]^T$.

Alternatively, we can constrain the initial guess of the local optimization to lie on the ambiguity plane, that we can compute in advance. This improves the stability of the optimization.

Overall, experiments (see Section 6) show that our correspondence measure solves the task of separating "good" and "bad" points quite reasonably. Therefore, it can serve as an adequate replacement for measures that punish variation of image intensity values.

4.4. Comparison to the brightness constancy measure

In Section 6 we provide results of shape reconstruction using our correspondence measure (6), and compare these results to ones obtained using a measure based on intensity variation. However, here we want to take a look at the correspondence measures themselves, not yet involving the surface extraction algorithm, which introduces its own effects. We will show that the usual measure does not provide sufficient information to distinguish correct correspondences from incorrect ones.

Using the true shape, for every point $\vec{P}$ in a 3D volume containing the shape we can compute its *actual error*: $err(\vec{P}) = \|\vec{P} - \vec{P}^*\|$, where $\vec{P}^*$ denotes the point on the true object surface that is closest (in some sense) to $\vec{P}$. We can also evaluate our correspondence measure $\mu(\vec{P})$ at $\vec{P}$, or any alternative correspondence measure. Ideally, a correspondence measure should "mimic" the actual error function $err(\vec{P})$. To evaluate our correspondence measure we computed statistics of our measure $\mu(\vec{P})$ as a function of $err(\vec{P})$ for the dinosaur toy which we used in our real experiment (Section 6, images are shown in Fig. 3). We also computed statistics for a second measure $\tilde{\mu}(\vec{P})$ based on intensity difference. (We got similar behavior for all variants of $\tilde{\mu}$ that we checked: standard deviation of intensities and their range in different sets of images.)

Fig. 1 shows graph plots of the mean values of $\mu(\vec{P})$ and $\tilde{\mu}(\vec{P})$ for every value $err(\vec{P})$. The graphs show a significant difference between these two measures. While our measure behaves as expected (it has a steep minimum at the true solution and increases as the actual error increases), the intensity difference measure is nearly flat in a significant interval around the true solution. Obviously, no shape reconstruction algorithm can be expected to produce good results if it relies on a correspondence measure that does not differentiate correct points from incorrect ones. On the other hand, the measure we developed looks quite adequate for the task, and, as our experiments show, indeed provides accurate reconstructions.

5. Extraction of the optimal surface

We seek to find the most likely object surface given a sequence of images of an object and the transformation parameters, whose effect is captured by the *correspondence measure* (described in Section 4). As in standard stereo problems, using the notion of Markov Random Fields the problem of finding the most likely surface can be cast as an energy minimization problem. The energy function is a sum of two terms: a data term and a smoothness (prior) term. The data term is the sum of correspondence measure values over the surface. The smoothness term penalizes large jumps of the surface, that are often caused by random errors due to noise in the data and imperfections of the method. This term depends only on neighboring points on the surface. We perform energy minimization by adopting the *expansion* algorithm from [5]. This algorithm allows us to choose a robust (discontinuity preserving) smoothness term (we implemented the truncated linear functional).

6. Experiments

In this section we present experimental results obtained with our shape reconstruction algorithm. We performed both computer simulations and experiments with a real object. We compare our recovered shape with the ground truth one, which enables precise evaluation of the results.

Details of the implementation

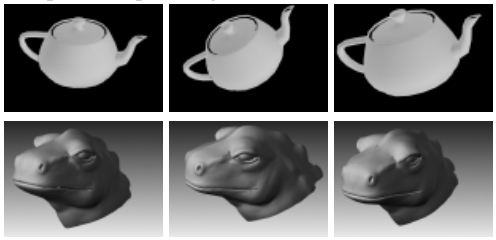
Our shape reconstruction implementation was based on a variant of the expansion algorithm from [5], implemented by Kolmogorov [12]. Shape was obtained by labeling the pixels in one of the images. Labels could be either depth or the location of a corresponding point in another image, with a special label meaning that no surface point exists (occlusion). We used an energy function that included a truncated linear smoothness term, which encourages piecewise smooth solutions. The correspondence measure (6) was computed with first order approximation (4 harmonics), and we used 6–10 input images.

In our implementation we parameterized the 3D space as it is commonly done in a non-calibrated setting. We took two reference images, and represented a 3D point by a point on the first image together with the position of its corresponding point along the epipolar line in the second image. (This is analogous to the notion of disparity in stereo). So the value that we recover is this position on the epipolar line, and the reconstruction error (or *correspondence error*) that we report is the error in this position. This error is very natural to evaluate results, and it is influenced by the performance of the algorithm rather than, say, camera configuration (as is the error in 3D position). The cameras are modeled as affine, and the projection from the object coordinate system to images (2) has the following form: $\text{proj}_j(\vec{P}) = A_j\vec{P} + \vec{t}_j$, where A_j is a 2×3 matrix and $\vec{t}_j$ is a 2D vector.

Time complexity

Time complexity of our shape reconstruction algorithm is in practice slightly more than $O(n^{3/2})$, where n denotes the image size. The main contribution is due to the minimum graph cut algorithm which is at the heart of the energy minimization technique. The min-cut algorithm, developed by Boykov and Kolmogorov [4], has n^3 worst case complexity (which is not as good as the state of the art), but is reported to run in effectively linear time on the special graphs that are used in these energy minimization methods (for example, [5]). Min-cut is computed for every label and we have $O(n^{1/2})$ labels, which results in $O(n^{7/2})$ worst case and about $O(n^{3/2})$ practical complexity. The energy minimization algorithm makes a constant number of iterations, and computation of our correspondence measure for one point is performed in constant time, so the asymptotic complexity is not changed. Note, that we do not need to

(a) Examples of input images (3 out of 10)



(b) Quality of reconstruction (the graphs show how many reconstructed points lie within given radius from their true positions).

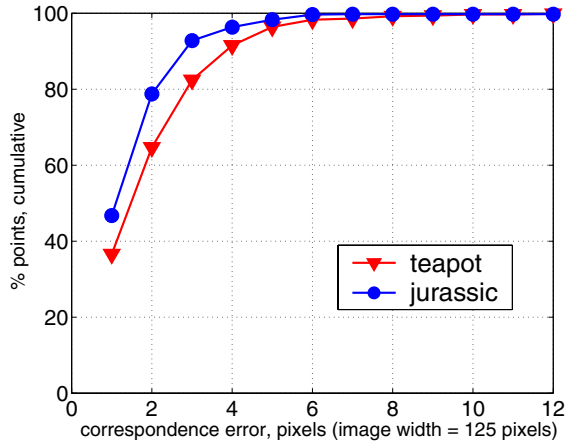


Figure 2. Simulation results.

compute the correspondence measure for more than $n^{3/2}$ points. To measure the actual complexity we ran the shape reconstruction algorithm for images of sizes from 10×10 to 300×300 pixels. We obtained a running time of $O(n^{3.2/2})$ with high accuracy, which supports the good practical complexity reported previously.

Simulations

We present here results of two computer simulations: a teapot (Fig. 2(a)) and a dinosaur head ("jurassic") (Fig. 2(b)). Both were performed by rendering images of 3D models using the 3D Studio Max software package. Along with the images we obtained exact camera calibration, object motion and ground truth shape. Both objects are perfectly Lambertian with uniform albedo. The largest motion between the most extreme views was 50 degrees. The scenes were lit by several (5-10) distant point light sources (placed manually at random positions). We compared the shape recovered by our algorithm to the ground truth shape. The results of these comparisons for both objects are presented in Fig. 2. The graphs show how many points of the reconstructed shape lie within a certain error radius from the correct points.

Experiment with a real object

In a real experiment, a dinosaur doll was rotated (rotation was within 40 degrees) and photographed with a camera (original 640×480 images were cropped and subsampled to obtain 125×50 input images for the algorithm). Ground truth shape was obtained by using a 3D laser scanner (which introduces its own errors). The camera was stationary (we do not need to move it, as in stereo, because the object itself moves), and calibration along with object motion were estimated from manually provided point correspondences in the images and the 3D scan.

We computed shape by running our algorithm and compared it against the shape obtained with the 3D scanner. We attempted also to reconstruct the shape using a correspondence measure that assumes brightness constancy (intensity difference). This was done to check how a standard measure copes with slight violations of the brightness constancy assumption. We chose the most similar pair of images (with the least difference of intensities of the corresponding points) and computed correspondence measure as intensity difference on these images. Note that this reconstruction problem is expected to be difficult for stereo algorithms that use the brightness constancy assumption because the dinosaur has roughly uniform albedo. The results of the experiment are presented in Fig. 3.

We see that our algorithm produces quite an accurate reconstruction and copes with brightness variations which considerably deteriorate the results obtained with the usual method. Our algorithm showed good performance in spite of inaccuracies in the estimation of parameters that are inevitable in practical applications.

In this experiment we used a few manually selected point correspondences to determine object motion and camera calibration (this enabled us to recover full calibration and compare results with ground truth). However a fully automatic implementation is possible, e.g., by tracking reliable point correspondences followed by motion estimation (e.g., using [18]).

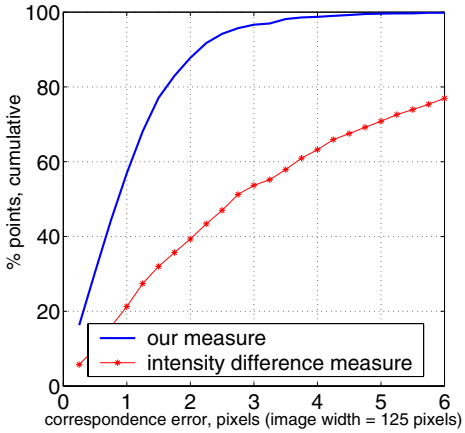
7. Conclusion

We presented a method for shape reconstruction of a non-stationary object under arbitrary lighting. We introduced a new *correspondence measure*, that indicates if points in different views should match, taking into account change in surface orientation and subsequent change in image intensity. We showed that our measure is suitable for solving the task. We described experiments with several objects, both synthetic and real, which illustrate the applicability of our method.

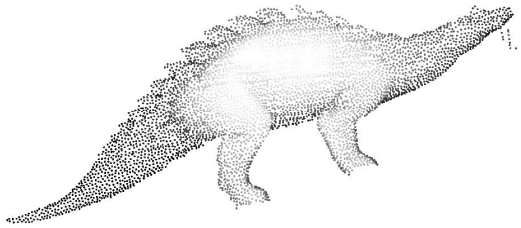
Our method relies on object motion, not the motion of the camera (as stereo methods). This allows us to use a fixed camera setup (which only simplifies the problem, remov-

(a) Examples of input images (3 out of 10)

(b) Quality of reconstruction for our measure and usual brightness constancy measure.



(c) Ground truth depth map (3D laser scan)



(d) Recovered depth map

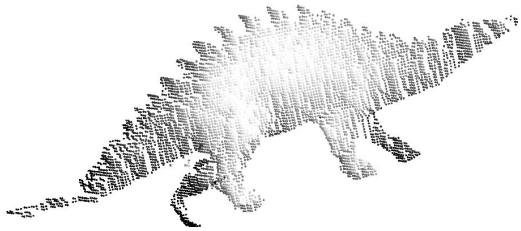


Figure 3. Dinosaur experiment.

ing ambiguity between camera and object motion). Such a setup is natural for video applications, and one can address shape reconstruction from video sequence of a moving object, with the benefit of a large number of views available to the algorithm.

References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, Feb. 2003.

[2] M. J. Black, D. J. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *CVIU*, 78(1):8–31, 2000.

[3] A. Blake, A. Zisserman, and G. Knowles. Surface description from stereo and shading. *Image and Vision Computing*, 3(4):183–191, 1985.

[4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *EMMCVPR*, pages 359–374, Sept. 2001.

[5] Y. Boykov, O. Vexler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, Nov. 2001.

[6] R. Carceroni and K. Kutulakos. Multi-View scene capture by surfel sampling: From video streams to Non-Rigid 3D motion, shape & reflectance. In *ICCV*, pages 60–67, June 2001.

[7] D. Frolova and R. Basri. Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. Forthcoming.

[8] P. Fua and Y. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *IJCV*, 16:35–56, Sept. 1995.

[9] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 27(10):1025–1039, Oct. 1998.

[10] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *PAMI*, 23(6):661–673, June 2001.

[11] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *ICCV*, pages 684–689, July 2001.

[12] V. Kolmogorov. Implementation of algorithms for energy minimization using graph cuts. http://www.cs.cornell.edu/People/vnk/software.html.

[13] J. Lu and J. Little. Surface reflectance and shape from images using a collinear light source. *IJCV*, 32(3):213–240, Aug. 1999.

[14] A. Maki, M. Watanabe, and C. Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. *IJCV*, 48(2):75–90, 2002.

[15] N. Mukawa. Estimation of shape, reflection coefficients and illumination direction from image sequences. In *ICCV*, pages 507–512, 1990.

[16] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *PAMI*, 20(9):961–979, Sept. 1998.

[17] R.Ramamoorthi and P.Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex lambertian object. *JOSA A*, pages 2448–2459, Oct. 2001.

[18] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography. *IJCV*, 9(2):137–154, 1992.

[19] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: unifying multiview stereo, photometric stereo, and structure from motion. In *ICCV*, Oct. 2003.