ACTIVE SHAPE AND DEPTH EXTRACTION FROM SHADOW IMAGES

Lei Wang and James J. Clark

Division of Applied Sciences
Harvard University

ABSTRACT

We present a recursive estimation scheme for determining surface shape and depth. This technique relies on the control over the way in which shadows are cast in scenes, through variation of the position of illuminant. The estimation process is based on the iterated extended Kalman filter, with a near-optimal control of the movement of the light source to reduce the sensitivity of the state estimate to the sensory noise.

1. INTRODUCTION

Vision tasks often involve the recovery of object shape and depth in the scene. Traditional methods such as shape from shading and stereo are among the common approaches. Active determination of such parameters has received much attention as well. For example, Clark [2] presents shape from active photometric stereo; Aloimonos and Shulman [1] has treated active *shape from x* problem as a whole.

Shape from shadows is another area of surface information extraction. The advantage of using shadows is quite obvious: no surface reflectance map is needed. Instead of collecting shading information (such as in photometric stereo and shape from shading), a binarization of the image which divides the image into shadows and non-shadows is all that is required. This makes both generating and analyzing image information much easier. Shafer and Kanade [8] use identified shapes of shadow to generate constraints giving rise to surface orientation; Kender and Smith [6] extract surface shape information based on object self-shadowing under moving light sources; and Raviv *et. al.* [7] reconstruct visible and invisible surfaces by analyzing shadowgrams.

We have proposed a method for locally recovering both surface shape and depth information from a series of images of surface shadows [9], based on the controlled motion of a nearby point light source. Shadow regions are assumed to have been identified, see [5] for an way of doing this task.

This research was supported in part by the National Science Foundation, under grant number IRI-92-23676.

These techniques, however, employ sequences of previously taken images upon which analyses are based. In this paper we present a new method to recursively estimate shape and depth from moving shadows. How a new image is taken is dependent upon intermediate results, under a near-optimal control which minimizes the uncertainty in the output of the estimation process.

2. SHAPE AND DEPTH FROM SHADOW

In [9] we proposed an active vision technique for recovering object shape and absolute position information. We will summarize this technique below.

Under perspective projection for a pinhole camera, the movement of a nearby point light source generates a sequence of shadow images. Figure 1 depicts the geometric model for part of this sequence, where we assume a planar surface for the background patch and a quadratic surface for the foreground patch, and associate the superscript i $(i = 1, \ldots, k)$ with the i^{th} light source position.

In the camera coordinate system we denote $\boldsymbol{r}$ as a position vector, $\boldsymbol{i}$ a image vector, and $\boldsymbol{n}$ a normal vector of a surface. We let subscript p denote a point on the *cast shadow* boundary, q a point on the *self shadow boundary*, and l the light source. A symmetric, positive definite matrix, M, and a vector $\boldsymbol{t}_q$ describe shape of the quadratic patch. The focal length of the camera is f.

Each light source position gives rise to a set of constraint equations that describe the geometric relations among vectors of points on the surfaces and in the image plane. With measured and known system parameters $\boldsymbol{i}_p$, $\boldsymbol{i}_q$, and $\boldsymbol{r}_l$, we seek the least number of such equations from which we can solve for the unknown parameters: $\boldsymbol{r}_p$, $\boldsymbol{r}_q$, $\boldsymbol{n}_p$, $\boldsymbol{t}_q$, and $\mathbf{M}$. These constraints arise from the following different aspects of the geometric model:

- *surface patch*: $\boldsymbol{n}_p{}^T \boldsymbol{r}_p = -1$ & $\boldsymbol{r}_q{}^T \mathrm{M}\, \boldsymbol{r}_q + \boldsymbol{r}_q{}^T \boldsymbol{t}_q = 1$,
- *image formation*: $\begin{bmatrix} i_{px} \\ i_{py} \end{bmatrix} = -\frac{f}{r_{pz}} \begin{bmatrix} r_{px} \\ r_{py} \end{bmatrix}$ and for

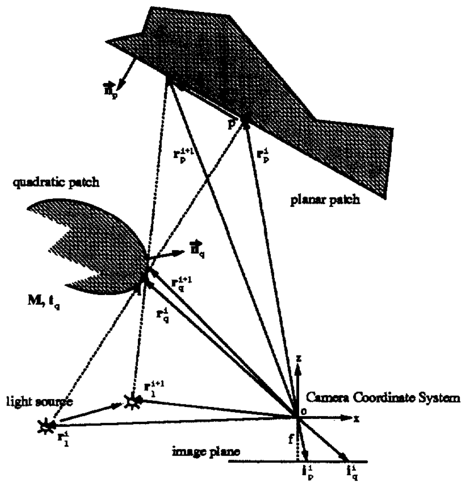

Figure 1: Geometry of the Shape Recovery Scheme

q,

- *illuminant collinear with shadow boundaries*: $\|\boldsymbol{r}_p - \boldsymbol{r}_q\| + \|\boldsymbol{r}_q - \boldsymbol{r}_l\| = \|\boldsymbol{r}_p - \boldsymbol{r}_l\|$, and
- *light ray tangent to quadratic surface*: $\boldsymbol{n}_q = 2\,\mathbf{M}\boldsymbol{r}_q + \boldsymbol{t}_q$.

It can be shown that the above constraints lead to a nonlinear system of great complexity even in the two dimensional space, and is therefore constructive to consider some special cases which are reasonable both in the sense of solving real-world problems and of reducing the degree of difficulty in finding a solution.

Such a special case, for example, considers some knowledge of the object shape *a priori*. In the world we live in objects often have edges. Hence it is reasonable to assume that the objects in our scheme have the same property, without greatly reducing the applicability of our scheme to real-world situations.

Under such a "sharp-edge" assumption, it can be shown that in two dimensions we need only solve the following linear system:

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{v}, \qquad (1)$$

where $\mathbf{A} = \begin{pmatrix} \eta^1/f & \omega^1 & 1 \\ \eta^2/f & \omega^2 & 1 \\ \eta^3/f & \omega^3 & 1 \end{pmatrix}$, $\boldsymbol{x} = (n_{px}, n_{pz}, 1/r_{qz})^T$, and $\boldsymbol{v} = (\xi^1, \xi^2, \xi^3)^T$. The quantities ω^i and ξ^i are in terms of the measurements and system parameters: $\omega^i = (fr^i_{lx} + i^i_{qx} r^i_{lz})/\gamma^i$ and $\xi^i = (i^i_{px} - i^i_{qx})/\gamma^i$, with $\gamma^i = fr^i_{lx} + i^i_{px} r^i_{lz}$. When $\gamma^i = 0$, depending on the terms at which γ^i vanishes, there will be either no solution or infinitely many solutions, and it is not possible to recover the object surface ($\mathbf{M}$).

When $\mathbf{A}$ is non-singular $\boldsymbol{x}$ has unique solution

$$\begin{aligned} n_{px} &= -\frac{1}{\Delta}\left[\left(\omega^1-\omega^2\right)\left(\xi^1-\xi^3\right) - \left(\omega^1-\omega^3\right)\left(\xi^1-\xi^2\right)\right] \\ n_{pz} &= \frac{1}{f\Delta}\left[\left(\eta^1-\eta^2\right)\left(\xi^1-\xi^3\right) - \left(\eta^1-\eta^3\right)\left(\xi^1-\xi^2\right)\right] \\ r_{qz} &= -f\Delta\left[\left(\omega^2\eta^3-\omega^3\eta^2\right)\xi^1 + \left(-\omega^1\eta^3+\omega^3\eta^1\right)\xi^2 \right. \\ &\quad \left. + \left(\omega^1\eta^2-\omega^2\eta^1\right)\xi^3\right]^{-1}, \end{aligned}$$

where $\eta^i = -\,i^i_{px}\omega^i$ and $\Delta \stackrel{\text{def}}{=} \det \mathbf{A}$. It can be easily seen that the unique solution exists so longs as any pair of the light source positions do not fall on the same radial line from the origin. The rest of the shape and depth quantities can be found using these values.

3. RECURSIVE ESTIMATION

The performance of the proposed active shadowing algorithm can significantly deteriorate due to measurement errors which arise in the measurement of positions of the shadow boundaries ($\boldsymbol{i}_p$ and $\boldsymbol{i}_q$) as well as the positions of the light source ($\boldsymbol{r}_l$). The sensitivity of the algorithm to the noise in the measurement drives the error in the output of the algorithm to an unacceptably high level. It has been shown in [9] that temporal integration of independent solutions by means of weighted least square or median filters alleviate the measurement errors only to a certain extent.

Better estimates for the shape and depth of a point over time can be achieved by the use of a recursive estimator, such as the Kalman filter. The Kalman filter is a common powerful tool for incremental estimation in dynamic systems. Let x_k be the state variables to be estimated at every time step k and $vecu_k$ be the control vector ($= \boldsymbol{r}_l$). Then given the measurements $z_k = h_k(x_k, vecu_k) + \hat{n}u_k$ where $h_k(x_k, vecu_k) = \boldsymbol{i}_{px}$ and $\hat{n}u_k \propto \mathcal{N}(0, R_k)$, we use an iterated extended Kalman filter [3] to produce estimates which have the least expected errors due to measurement noise:

$$\hat{x}_{k,i+1} = \hat{x}_{k-1} + K_{k,i}[z_k - h_k(\hat{x}_{k,i}) - H_k(\hat{x}_{k,i})(\hat{x}_{k-1} - \hat{x}_{k,i})],$$

where $K_{k,i} = P_k H_k^T(\hat{x}_{k,i})[H_k(\hat{x}_{k,i})P_k H_k^T(\hat{x}_{k,i}) + R_k]^{-1}$ and $P_{k+1} = [I - K_{k,i_{\max}} H_k(\hat{x}_{k,i_{\max}})]P_k$, for $i = 1, 2, \ldots, i_{\max}$.

In the above, $\hat{x}_{k,0} = \hat{x}_{k-1} = (k-1)^{st}$ estimate of the state, K_k is the *Kalman gain matrix*, P_k the *state estimate error covariance matrix*, and $H_k(\hat{x}_k)$ is obtained by linearizing h around the current estimate, $\hat{x}$: $h_k(x_k, vecu_k) = h_k(\hat{x}_k, vecu_k) + H_k(\hat{x}_k, vecu_k)(x_k - \hat{x}_k) + \cdots$, where $H_k(\hat{x}_k, vecu_k) = \frac{\partial h_k}{\partial x}(x, vecu_k)\big|_{x=\hat{x}_k}$.

We assume that the prior is such that $E[x_0] = \hat{x}_0$ and $\mathrm{Cov}[x_0] = P_0$.

4. NEAR-OPTIMAL CONTROL OF NEXT-LOOK

In *active vision*, one should strive to utilize the *active agent* to control the geometric parameters of the sensory apparatus and thence gather more information useful for the solution process. In other words, instead of having the agent move randomly or follow a predetermined path, one should move it to a point which is "optimal" in some sense.

In the Kalman filter the level of uncertainty in the state estimate is measured by the covariance matrix: $Tr(P_k) \stackrel{\text{def}}{=} Tr\{E[(x-\hat{x}_k)^T(x-\hat{x}_k)]\}$.[1] Then it is natural to find, for each k, $vecu_k$ such that this uncertainty is minimized. This is equivalent to

$$\max_{vecu_k} \delta_k = Tr(K_k H_k P_{k-1}). \tag{2}$$

A gradient descent on δ_k would usually solve a nonlinear optimization problem like this one [4], seeking the maximum of δ_k at each time step. However, practical limitations on the light source motion prevents the maximum being attained at each time step. For example, the light source cannot move behind the foreground object, or make the cast shadows move out of the image plane, or the robot manipulator which is placing the light source has its own limit on the extent of workspace. Hence we are solving a constrained optimization problem, which is usually very difficult. We are then forced to make approximations to the solutions, thereby settling for near-optimal solutions.

We use a variant of the gradient descent on δ_k: at each time step k, we move the light source in the gradient direction by a small amount. The local minimum of δ_k may never be attained this way since δ_k changes from one time step to the next due to the change in the Kalman gain matrix K as new observations are made. This again, will give us a near-optimal solution. If the constraint on the change in the control vector is such that the allowable change is small, and if the boundary of the region of allowable control vectors is convex, then it is likely that the optimal control vector will lie on the boundary formed by the constraints, and a gradient step to the boundary will therefore provide the optimal change.

[1] In practice, this measure is given by weighting the diagonal elements with the inverse squares of their corresponding current state values to offset the effect of different units and scales.

If we let $u_{k+1} = vecu_k + C\nabla_u \delta_k$, where C describes the magnitude of each move, it can be shown that

$$\frac{\partial \delta_k}{\partial u_i} = 2\; Tr\left(K_k \frac{\partial H_k}{\partial u_i} P_k\right) \tag{3}$$

for each component u_i of $vecu_k$.

5. SIMULATIONS

Based on the results and analyses above we perform a series of simulations on the algorithm for the two dimensional case in which the object has a sharp edge. In the following, all length measurements are in millimeters and all angle measurements are in degrees, and surface normals are unitless, unless otherwise stated. We set the scene as follows: the background has a normal of $(-10^{-3}, -10^{-3})$ and goes through the point $(0, 1000)$, the shadowing object has its sharp edge at $(42, 750)$ in front of the camera.[2]

Noise comes from mainly two sources: quantization noise in the image measurements and systematic noise in the light source placements. Since the resolution of the image cannot be better than one pixel we attach to all image measurements a uniform noise with a standard deviation (σ) of one pixel width for the worst case.

In the following we allow the noise level in the light source position to have a standard deviation of 0, 0.1, 0.3, and 1. Each plot below show the log of relative errors of estimate for each state. In Figures 2 and 3, the plots are arranged as follows: from left to right, top to bottom, $\sigma_{r_l} = 0, 0.1, 0.3, \text{and } 1$. Plotted are the diagonal entries of the error covariance matrices. These values are estimates for the variances in the estimates of the state variables.

In the iterated extended Kalman filter we let the light source follow a circular trajectory. This yields the plots shown in Figure 2.

The constrained optimization problem produces the trajectories for the light source along which the error covariance of the state estimate is nearly minimized at every step. In general, this leads to faster and better convergence than those of the previous two scenarios. The exception occurs when the uncertainty in the light source position is high; in that case the trajectory, based as it is on noisy information, will be quite different than the optimal one determined using exact information.

A worst case of about 10% error is achieved. More importantly, a much faster convergence (drop-off) rate

[2] The camera has the parameters of that in our lab: $f = 50$ and pixel size $= 17\mu$.

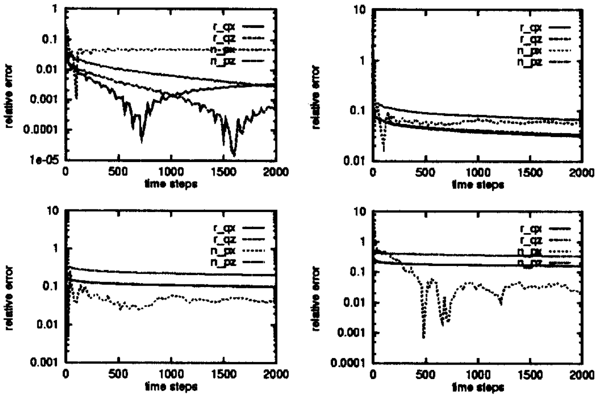

Figure 2: Log of Relative Estimation Error: IEKF (left to right, top to bottom): $\sigma_{r_l} = 0$ $\sigma_{r_l} = 0.1$, $\sigma_{r_l} = 0.3$, $\sigma_{r_l} = 1$

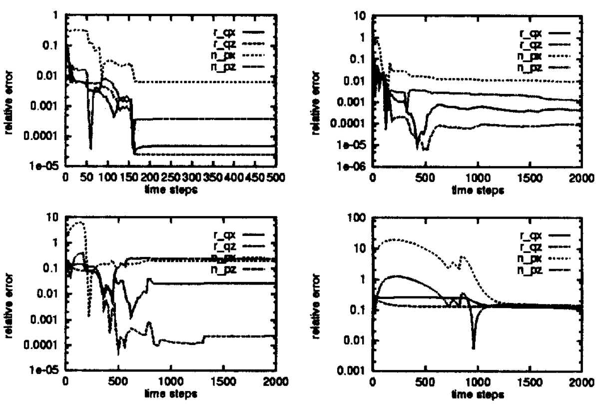

Figure 3: Log of Relative Error with Control (left to right, top to bottom): $\sigma_{r_l} = 0$, $\sigma_{r_l} = 0.1$, $\sigma_{r_l} = 0.3$, $\sigma_{r_l} = 1$

is realized for every case, making it a much more efficient recursive estimator. Since the Kalman filter is optimal when the measurement noise is Gaussian, the increasing error in the state estimate is to be expected here.

6. CONCLUSIONS

We present an algorithm to recover shape and depth from a sequence of shadow images generated by a near-optimal placement of an illuminant. Unlike traditional image sequence analysis which applies algorithms to a set of pre-taken images, each new image in our algorithm is dependent upon the intermediate results of the algorithm.

A constrained near-optimal optimization problem is solved recursively based on the iterated extended Kalman filter. This new technique uses vision as part of the control loop to minimize uncertainty level in the state estimate.

7. REFERENCES

[1] Aloimonos, J., and D. Shulman, **Integration of Visual Modules: An Extension of the Marr Paradigm,** Academic Press, Inc., San Diego, California, 1989

[2] Clark, J. J., "Active Photometric Stereo," in Proceedings of the 1992 IEEE Computer Vision and Pattern Recognition Conference, Champaign, IL, pp 29-35, June 1992

[3] Gelb, A. (ed.), **Applied Optimal Estimation,** MIT Press, Cambridge, MA, 1974

[4] Ho, Y.C. and A.E. Bryson, *Applied Optimal Control: Optimization, Estimation, and Control,* Hemisphere Publishing Corporation, N.Y.C., 1975

[5] Jian, C., and M.O. Ward, "Shadow Identification," in the Proceedings of 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 15-18, Champagne Illinois, pp. 606-612, 1992.

[6] Kender, J. R., and E. M. Smith, "Shape from Darkness: Deriving Surface Information from Dynamic Shadows," in the Proceedings of the First International Conference on Computer Vision, June 8-11, London, pp. 539-546, 1987.

[7] Raviv, D., Y. Pao, and K.A. Loparo, "Reconstruction of Three-Dimensional Surfaces from Two-Dimensional Binary Images," IEEE Transaction on Robotics and Automation, Vol. 5, No. 5, pp. 701-710, October 1989.

[8] Shafer, S. A., and T. Kanade, "Using Shadows in Finding Surace Orientations," Computer Vision, Graphics, and Image Processing, Vol. 22, pp. 145-176, 1983.

[9] Wang, L. and J. Clark, "Depth from active shadow motion", Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision XII, Boston, MA, September 1993, pp. 2-13