To appear in ICPR-94, Jerusalem, Israel, October 9-13, 1994.

Estimating 3D Motion and Shape of Multiple Objects Using Hough Transform*

Tina Yu Tian and Mubarak Shah
Computer Vision Lab, Computer Science Department
University of Central Florida, Orlando, FL 32816
email: {tian, shah}@cs.ucf.edu

Abstract

We present a robust method to determine 3D motion and structure of multiple objects. Rather than segmenting the scene containing multiple motions using 2D parametric model, we use the general 3D motion model and exploit Hough transform and robust estimation techniques to determine motion and segmentation simultaneously for an arbitrary scene. We divide the input image into patches, and for each sample of the translation space and each patch, we compute the rotation parameters using weighted least-squares fit. Each patch votes for a sample in the five-dimensional parameter space (translation and rotation). The multiple local maxima in the parameter space naturally correspond to the multiple moving objects. Our experimental results show that the proposed method is robust and relatively insensitive to noise.

1 Introduction

Determining three-dimensional motion and structure of *multiple* objects from two frames has been a challenging problem. The most common approach for motion analysis is based on two phases: computation of optical flow field and interpretation of this flow field. There are numerous methods for computing optical flow; Barron, Fleet and Beauchemin [3] present a comprehensive evaluation and comparison of existing optical flow methods. For interpretation of optical flow, a straightforward method is to segment the optical flow field first, and then apply ego-motion structure-from-motion (SFM) algorithm to each moving object. However, segmentation using optical flow field can not distinguish between real motion boundaries and depth discontinuities. Another approach for segmentation is based on the set of coherent motion parameters, independent of depth values. This approach (e.g. [1, 6, 7]) exploits 2D parametric motion approximations, ignoring the higher-order information of the displacement vector, and thus yields incorrect motion segmentation. Moreover, using a 2D motion model to segment a 3D scene can lead to ambiguities. Methods in [2] and [1] belong to Hough transform approach for SFM. The advantages of Hough transform are that it is relatively insensitive to noise and more robust as a global approach, and the multiple local maxima in the parameter space naturally correspond to the multiple moving objects. In [2], the candidate solutions over the entire five-dimensional parameter space have to be evaluated, and known depth is required, which makes the problem much easier. In [1], segmentation and motion computation is determined separately. Segmentation is performed based on 2D parametric model through a split-merge process.

In this paper, we attempt to solve the SFM problem for an arbitrary scene which may contain several moving objects with possible camera motion. We make no assumption about the scene (e.g. piecewise planar surface, known depth, etc) and use the exact general motion model, and exploit Hough transform and robust estimation technique to determine motion and segmentation *simultaneously*.

2 Motion Estimation Using Hough Transform

Let $\mathbf{T} = (\mathrm{T_x}, \mathrm{T_y}, \mathrm{T_z})^{\mathrm{t}}$ and $\mathbf{\Omega} = (\Omega_{\mathrm{x}}, \Omega_{\mathrm{y}}, \Omega_{\mathrm{z}})^{\mathrm{t}}$ respectively denote translation and rotation of a rigid object, $\mathbf{v} = [\mathrm{u}, \mathrm{v}]^{\mathrm{t}}$ denotes optical flow, and f denotes focal length. At each image point, (x, y), of the camera-centered coordinate system, the relationship between optical flow, $[u, v]^t$, motion parameters, and

*The research reported in this paper was supported by NSF grants CDA-9122006 and IRI-9220768.

corresponding depth is given by the following equations:

$$\mathbf{v}(\mathrm{x},\mathrm{y}) = \mathrm{p}(\mathrm{x},\mathrm{y})\mathbf{A}(\mathrm{x},\mathrm{y})\mathbf{T} + \mathbf{B}(\mathrm{x},\mathrm{y})\mathbf{\Omega}, \qquad (1)$$

where $p(x, y)$ is the inverse depth, and

$$\mathbf{A}(\mathbf{x},\mathbf{y}) = \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix},$$

$$\mathbf{B}(\mathbf{x},\mathbf{y}) = \begin{bmatrix} xy/f & -(f+x^2/f) & y \\ f+y^2/f & -xy/f & -x \end{bmatrix}.$$

Since inverse depth, $p(x, y)$, and translation, $\mathbf{T}$, can be determined only up to a scale factor, we only solve the translation direction and relative depth. Now $\mathbf{T}$ denotes a unit vector for translation direction and $p(x, y)$ denotes the relative inverse depth, $||\mathbf{T}||/Z$. Unit vector $\mathbf{T}$ can be represented by spherical coordinates in terms of slant, θ, and tilt, ϕ: $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^t$. Only half of the sphere has to be considered, since solutions on the front and back halves are the same, therefore, θ varies from 0^0 to 90^0, and ϕ varies from 0^0 to 360^0.

We can cancel depth $p(x, y)$ from (1) and obtain:

$$\mathbf{c}(\mathbf{T})\mathbf{\Omega} = \mathrm{q}(\mathbf{T}), \qquad (2)$$

where

$\mathbf{c}(\mathbf{T}) = [\ \mathrm{fT_z x} + \mathrm{T_y xy} - \mathrm{T_x}(\mathrm{f}^2 + \mathrm{y}^2),$
$\mathrm{fT_z y} + \mathrm{T_x xy} - \mathrm{T_y}(\mathrm{f}^2 + \mathrm{x}^2),\ \mathrm{fT_x x} + \mathrm{fT_y y} - (\mathrm{x}^2 + \mathrm{y}^2)\mathrm{T_z}\ ],$

$q(\mathbf{T}) = -\mathrm{fT_x v} + \mathrm{fT_y u} + \mathrm{T_z}(\mathrm{xv} - \mathrm{yu}).$

We collect N equations of (2) into the matrix form:

$$\mathbf{C}(\mathbf{T})\mathbf{\Omega} = \mathbf{q}(\mathbf{T}), \qquad (3)$$

where

$$\mathbf{C}(\mathbf{T}) = [\ \mathbf{c}_1(\mathbf{T}),\quad \mathbf{c}_2(\mathbf{T}),\quad \cdots,\quad \mathbf{c}_N(\mathbf{T})\ ],$$

$$\mathbf{q}(\mathbf{T}) = [\ q_1(\mathbf{T}),\quad q_2(\mathbf{T}),\quad \cdots,\quad q_N(\mathbf{T})\ ].$$

At least three image points have to be used to solve for $\mathbf{\Omega}$, rotation parameters. We compute least-squares estimate of rotation for a fixed choice of $\mathbf{T}$:

$$\mathbf{\Omega} = (\mathbf{C}(\mathbf{T})^{\mathrm{t}}\mathbf{C}(\mathbf{T}))^{-1}\mathbf{C}(\mathbf{T})^{\mathrm{t}}\mathbf{q}.$$

So, rotation is represented in terms of translation.

In order to deal with multiple moving objects, we divide the entire image into patches, and within each patch compute least-squares estimate of rotation for a given $\mathbf{T}$, and count the corresponding votes. The algorithm is given in Figure 1. Under the framework of standard Hough transform, instead of evaluating the entire five-dimensional parameter space, we only examine the two-dimensional translational parameter space, from which the corresponding optimal solution for three rotation parameters is computed.

Algorithm:

1. Quantize the parameter space of θ, ϕ, Ω_x, Ω_y, and Ω_z.
2. Form an accumulator array A(θ, ϕ, Ω_x, Ω_y, Ω_z) and initialize it to zero.
3. For each sample pair (θ, ϕ) do the following:
   For each patch in the image do
   Compute $\mathbf{\Omega} = (\mathbf{C}(\mathbf{T})^{\mathrm{t}}\mathbf{C}(\mathbf{T}))^{-1}\mathbf{C}(\mathbf{T})^{\mathrm{t}}\mathbf{q}$.
   Increment accumulator array A(θ, ϕ, Ω_x, Ω_y, Ω_z).
4. Find the local maxima in the accumulator array corresponding to multiple moving objects.

Figure 1: Hough algorithm for SFM

In the scene containing multiple moving objects, each image patch may contain multiple motions. The least-squares estimate is computationally efficient, but not robust, particularly to deal with multiple motions.

3 Robust Motion Estimation of Multiple Objects

In this section, we present a robust method for multiple motion estimation. Multiple motions within a patch can be treated as outliers with respect to the major motion. M-estimators can be expected to handle outliers and Gaussian noise in optical flow measurements simultaneously, so, we include redescending M-estimator in our scheme to obtain more robust rotation estimate of a major motion for a fixed $\mathbf{T}$ in a small patch, rejecting the other minor motions as outliers.

The M-estimators minimize the sum of a symmetric, positive-definite function $\rho(r_i)$ of the residuals r_i, with a unique minimum at $r_i = 0$. There are several possible choices for ρ function listed in [4]. Since it is relatively smooth, Beaton and Tukey's biweight function is used in our implementation:

$$\rho(r) = \begin{cases} (C_B^2/2)[1-[1-(r/C_B)^2]^3] & \text{if } |r| \le C_B \\ C_B^2/2 & \text{otherwise} \end{cases},$$

where r is residual, and C_B is a turning constant. It is recommended in [5] $C_B = 4.685$ to achieve superior performance for Gaussian noise. Since we are dealing with the patch which may contain the multiple motions, smaller turning constant should be used.

M-estimation problems are usually solved using an iterative weighted least-squares method [5], in which a weight is computed for each data point based on the residual error of the previous estimate. Initially, the weights are all set to 1, and the vector $\mathbf{\Omega}$ (denoted by $\hat{\mathbf{\Omega}}_0$) with the contribution of all data points in the patch is computed, then weights are updated according to the following:

$$w(r) = \begin{cases} [1-(r/C_B)^2]^2] & \text{if } |r| \le C_B \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The vector $\mathbf{\Omega}$ is refined through iterations:

$$\hat{\mathbf{\Omega}}_1 = \hat{\mathbf{\Omega}}_0 + (\mathbf{C}^{\mathrm{t}} < \mathrm{w}(\frac{\mathbf{q}-\mathbf{C}\hat{\mathbf{\Omega}}_0}{\sigma}) > \mathbf{C})^{-1}\ \mathbf{C}^{\mathrm{t}} < \mathrm{w}(\frac{\mathbf{q}-\mathbf{C}\hat{\mathbf{\Omega}}_0}{\sigma}) > (\mathbf{q}-\mathbf{C}\hat{\mathbf{\Omega}}_0), \quad (5)$$

where $< >$ denotes an $N \times N$ diagonal matrix, and σ is a scale parameter which can be estimated by

$$\sigma = 1.4826\ med\ |r_i - med\ r_i|, \quad (6)$$

where med denotes the median taken over the entire patch. We use the following measure to stop the iterations:

$$E^{(l)} = \sqrt{\frac{\sum_{i=1}^{n} w_i r_i^2}{\sum_{i=1}^{n} w_i}},$$

where l denotes the iteration number, and n denotes the number of nonzero weights corresponding to the number of inliers, which contribute to the robust estimate. If the difference of E at the current and the previous iteration is less than some predefined threshold or the maximum number of iterations is reached, the iteration is stopped.

4 Segmentation and Depth

A set of hypotheses on motion parameters can be obtained from local maxima in the parameter space. The image is then segmented based on this set of hypotheses. Once the motion and segmentation are known, the relative depth at each pixel, $(x,\ y)$, can be determined (see Equation (1)) by:

$$\frac{Z(x,y)}{|\mathbf{T}|} = \frac{(\mathbf{A}(\mathrm{x},\mathrm{y})\mathbf{T})^{\mathrm{t}}(\mathbf{A}(\mathrm{x},\mathrm{y})\mathbf{T})}{(\mathbf{A}(\mathrm{x},\mathrm{y})\mathbf{T})^{\mathrm{t}}(\mathbf{v}(\mathrm{x},\mathrm{y}) - \mathbf{B}(\mathrm{x},\mathrm{y})\mathbf{\Omega})}.$$

The motion parameters are relatively insensitive to noise in the optical flow measurements since the inputs are combined over the segment. The depth estimates are computed locally, and they thus are sensitive to the input noise. One possible way to improve depth estimates is to integrate information temporally through multiple frames.

5 Conclusion

In this paper, we present a robust method to determine 3D motion and structure of multiple objects. Rather than segmenting the scene containing multiple motions using 2D parametric model, we use the general 3D motion model and exploit Hough transform and robust estimation techniques to determine motion and segmentation simultaneously for an arbitrary scene. In our method, we do not have to evaluate the candidate solutions over the entire five-dimensional parameter space. We only examine the two-dimensional translation space. We divide the input image into patches, and for each sample of the translation space and each patch, we compute the rotation parameters using weighted least-squares fit, incorporating redescending M-estimator to reject outliers (either noise or minor motions in the patch). Each patch votes for a sample in the five-dimensional parameter space. Our experimental results show that the proposed method is robust and relatively insensitive to noise.

Acknowledgements: We would like to thank Prof. David Heeger for helpful discussions.

References

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Patt. Anal. Mach. Intell.*, 7:384–401, 1985.

[2] D. Ballard and O. Kimball. Rigid body motion from depth and optical flow. *Comput. Vision, Graph. Image Process.*, 22:95–115, 1983.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. In *CVPR'92*, pp. 236–242, 1992.

[4] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel. *Robust Statistics: An Approach Based on Influence Function.* Wiley, New York, 1986.

[5] P. Holland and R. Welsch. Robust regression using iteratively reweighted least squares. *Commun. Statist. – Theor. Meth.*, A6: 813–827, 1977.

[6] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV'92*, pp. 282–287, 1992.

[7] J. Wang and E. Adelson. Layer representation for motion analysis. In *CVPR'93*, pp. 361–366, 1993.