ROBUST BAYESIAN CAMERAS MOTION ESTIMATION USING RANDOM SAMPLING

Gang Qian[†]*, Rama Chellappa*[‡]*, and Qinfen Zheng*[‡]

[†]Department of Electrical Engineering and
Arts, Media and Engineering Program
Arizona State University
Tempe, AZ 85287-8706
gang.qian@asu.edu

[‡]Center for Automation Research
Institute for Advanced Computer Studies
University of Maryland
College Park, MD, 20742-3275
{rama,qinfen}@cfar.umd.edu

ABSTRACT

In this paper, we propose an algorithm for robust 3D motion estimation of wide baseline cameras from noisy feature correspondences. The posterior probability density function of the camera motion parameters is represented by weighted samples. The algorithm employs a hierarchy coarse-to-fine strategy. First, a coarse prior distribution of camera motion parameters is estimated using the random sample consensus scheme (RANSAC). Based on this estimate, a refined posterior distribution of camera motion parameters can then be obtained through importance sampling. Experimental results using both synthetic and real image sequences indicate the efficacy of the proposed algorithm.

1. INTRODUCTION

Recently, multiple cameras and wide baseline stereo cameras have attracted much attention to tackle computer vision problems that are difficult to solve using a monocular camera, such as human motion analysis (robust human body part tracking), wide area surveillance (distributed tracking), etc. To observe the events of interests from distinct view points, cameras are usually widely separated, with significant orientation changes. Relative camera positions and orientations need to be solved so that video streams or pre-processing results from these cameras can be effectively integrated. This essentially is the traditional structure-from-motion problem in computer vision. Two issues need to be addressed here: feature correspondence across wide baseline stereo, and robust camera motion estimation. Between the two issues, the first one has been considered a much more difficult problem than the second one. There is a common understanding that once the feature correspondences can be found accurately, the camera motion estimation is straightforward. Based on this belief, in the past few years, considerable research effort has been taken to address the wide baseline stereo feature matching problem and a multitude of methods has been proposed to solve the problem, based on PDE [1], affine viewpoint invariant features transformation [2,3] and moving objects trajectory matching [4], to name a few. Although these wide baseline feature matching algorithms usually work reasonably well, significant errors sometimes still exist in the found feature correspondences. In the presence of feature matching errors, posterior probability density function (pdf) is a better representation of the camera motion parameters space than a single optimal solution in some sense. Since the posterior pdf can well describe the uncertainty as well as any ambiguities of the motion parameters. In this paper, we propose a two step coarse-to-fine computational framework to obtain the posterior pdf of the camera motion parameters. First, a coarse prior distribution of camera motion parameters is estimated using the random sample consensus scheme (RANSAC) [5]. Based on the initial estimates, a refined posterior distribution can then be obtained via importance sampling [6].

1.1. Related work

The sequential importance sampling has been used to recursively estimate the motion of a continuously moving camera and scene structure, using a video taken by the moving camera [7]. However, possible large camera orientation changes make it impossible to apply the algorithm proposed in [7] to compute camera motion, which assumes smooth camera rotation changes in continuous camera motion.

The RANSAC technique has also been widely applied to camera motion estimation, [8, 9, 10], especially in the presence of mismatch features. When using RANSAC, feature sets containing the minimal number of feature correspondences are selected randomly. Each set produces a hypothesis for the relative camera motion. These hypotheses are then measured by robust statistical criteria using all feature correspondences. The best solution is further refined iteratively using global optimization techniques such as bundle-adjustment [11]. Although based the scores of these hypotheses, a pdf of the camera motion parameters can be found, it is not

accurate enough to reflect the true solution space, which is because that the sampling is done in the observation data set (feature pairs), but not directly in the parameter space. In the proposed algorithm, we apply RANSAC to obtain an initial prior distribution and based on that, a refined posterior pdf of the motion parameters is computed through the importance sampling.

2. BACKGROUND

2.1. Epipolar Geometry

Consider two separate cameras. Let C and C' be the two related camera centered coordinate systems. Let the rotation matrix and translation vector from C to C' be (R, T) such that for a 3D point P in C, its new coordinate in C' is given by $P'=R(P-T)$. Five parameters are used to describe the relative motion between C and C', $\boldsymbol{x}=(\psi_x, \psi_y, \psi_z, \alpha, \beta)^T$, where ψ's are rotation angles and α and β are the elevation and azimuth angles of translation direction. Let γ be the translation magnitude. The translation T is then given by $\gamma(\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)^T$. Translation magnitudes of multiple cameras can only be recovered up to a scale. In the case of two cameras, it can be used as the unit for length variables such as the feature point depths.

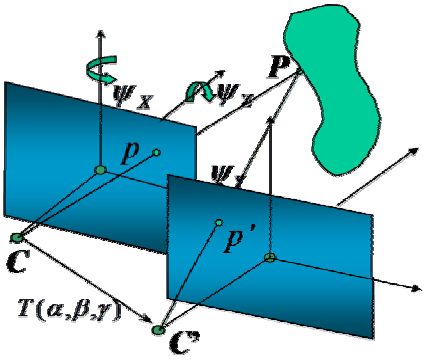

Figure 1. Projective geometry of two cameras

Figure 1 illustrates the parameterization of relative motion between two cameras. Let $p=(u,v,1)$ and $p'=(u',v',1)$ be the projections of $\boldsymbol{P}$ in two related image planes in homogeneous coordinates. The well-known epipolar constraint states that the epipolar distance $d(p',l)$ from p' to its related epipolar line $l=(a,b,c)$ is zero. $d(p',l)$ is given by

$$d(p',l)=\frac{p'l^T}{\sqrt{a^2+b^2}}=\frac{au'+bv'+c}{\sqrt{a^2+b^2}} \qquad \textbf{(1)}$$

The epiploar line $\boldsymbol{l}$ related to $\boldsymbol{p}$ is determined by the camera motion.

$$[a,b,c]^T = A'^{-T}[-RT]_\times RA^{-1}[u,v,1]^T$$

A and A' are the known calibration matrices of the two cameras. $[K]_\times$ denoates a skew symmetric matrix, such that for any 3D vector y, $[K]_\times y=K\times y$.

2.2. Importance Sampling

Drawing samples from a target distribution sometimes is not an easy task. Importance sampling has been proposed to use weighted samples from a trial distribution to represent samples from the target distribution, in terms of properly weighted samples. A random variable X drawn from a distribution g is said to be **properly weighted** by a weighting function $w(X)$ with respect to the distribution π if for any integrable function h,

$$E_g h(X)\, w(X) = E_\pi h(X)$$

A set of random draws and weights $\{x^{(j)};w^{(j)}\}$; $j = 1; 2..,n$, is said to be properly weighted with respect to π if

$$\lim_{n\to\infty}\sum_{j=1}^{n}\frac{w^{(j)}h(x^{(j)})}{\sum_{j=1}^{n}w^{(j)}}=\int h(x)\pi(x)dx$$

for any integrable function h. The weight function $w(X)$ is the ratio of the target and trial distributions.

$$w(X)=\frac{\pi(X)}{g(X)}$$

Given nosiy feature matches Y, we would like to obtain samples that are properly weighted with respect to the posterior camera motion parameters, namely, $p(X|Y)$. In the proposed approach, we first use RANSAC to get an initial distribution for the motion parameters. Using this initial resutls as the prior knowledge for camera motion, we then apply importance sampling to obtain refined posterior camera motion distributions.

3. INITIALIZATION

To improve the numerical condition of the observation matrix [12], the coordinates of the correspondences are normalized to the range of [-1,1]. The following procedure using the RANSAC based on the eight-point algorithm [12] is then applied to the normalized correspondences to obtain an initial distribution of the camera motion parameters.

a) Randomly select N sets of eight matched feature pairs. N depends on the percentage of good matches and desired probability that at least one set contains all correctly matched features.

b) Compute the essential ($\boldsymbol{E}$) matrix using one of the feature set by the eight-point algorithm.

c) If the major portion of the feature correspondences (e.g. 90%) can be interpreted by the fundamental matrix such that the epipolar distances computed by (1) are less than a pre-chosen threshold (e.g. 4 pixels), this $\boldsymbol{E}$ matrix is accepted as a good sample.

d) The camera motion parameters $\boldsymbol{x}$ is then recovered from the $\boldsymbol{E}$ matrix [10].

e) Assume the standard deviations of feature matching errors along two axes in image plane are both σ. A weight $\boldsymbol{w}$ is computed based on the epipolar distance Δ, with respect to this sample. Since the true projection of the point in the image plane is unknown, we assume that it's uniformly distributed along the image plane. Hence the sample weight, which is the likelihood of the motion estimation is computed by

$$\begin{aligned} \boldsymbol{w} &= \boldsymbol{p}((\boldsymbol{u}',\boldsymbol{v}') \mid \boldsymbol{x}) \\ &= \frac{1}{2\pi(\boldsymbol{r}_1+\boldsymbol{r}_2)\sigma^2}\int_{-\boldsymbol{r}_1}^{\boldsymbol{r}_2} \exp-\frac{\boldsymbol{r}^2+\Delta^2}{2\sigma^2}\boldsymbol{dr} \\ &= \frac{\exp\{-\frac{\Delta^2}{2\sigma^2}\}}{2(\boldsymbol{r}_1+\boldsymbol{r}_2)\sqrt{2\pi}\sigma}\left[\boldsymbol{erf}(\frac{\boldsymbol{r}_2}{\sqrt{2}\sigma})+\boldsymbol{erf}(\frac{\boldsymbol{r}_1}{\sqrt{2}\sigma})\right] \end{aligned} \quad \textbf{(2)}$$

where r_1 and r_2 are the lengths of the line segments from the p_l (the projection of p' on l) to the two valid terminals of l on the second image plane (as illustrated by Figure 2). The two valid terminals can be determined using camera motion parameter $\boldsymbol{x}$, previous project p and the cheirality (positive-depth) constraint.

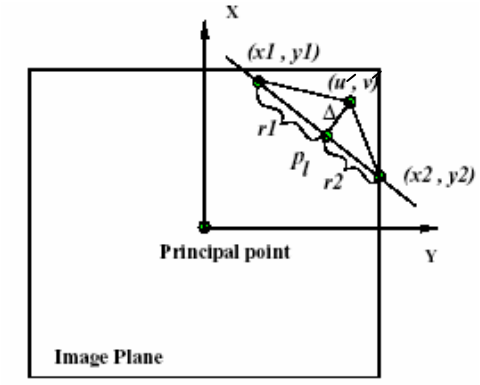

Figure 2. Epipolar line segments

f) Go to step b) and continue, till all N sample sets have been tested. Keep all the accepted fundamental matrix samples.

Let $\{\boldsymbol{x}^{(j)},w^{(j)}\}_{j=1}^{n}$ be the valid weight sample set resulting from this initialization procedure. The motion samples can be viewed as samples drawn from a priori distribution $p_0(\boldsymbol{x})$.

4. BAYESIAN MOTION ESTIMATION

According to Bayes' rule

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p_0(\boldsymbol{x})$$

Given samples and weights representing prior motion distribution $p_0(\boldsymbol{x})$, we would like to drawn more samples from the prior distribution $p_0(\boldsymbol{x})$ and evaluate the weights, which is $p(\boldsymbol{y}|\boldsymbol{x})$, the likelihood of the samples according to importance sampling. The resulting refined weighted samples are properly weighted with respect to the posterior distribution. To increase the search space of the motion parameters, a mixed Gaussian distribution is used in place of the original approximate prior distribution $p_0(\boldsymbol{x})$. Using each sample of $\{\boldsymbol{x}^{(j)}\}$ as seed samples, new samples are drawn according from a seed-sample-centered normal distribution and the weights of the new samples are evaluated properly. The detailed sampling procedure is as follows.

a) *Re-sample the initial distribution.* According to the initial distribution $\{w^{(j)}\}_{j=1}^{n}$, draw seed sample $\boldsymbol{x}_s$.

b) *Local importance sampling.* Around $\boldsymbol{x}_s$, draw M samples $\boldsymbol{x}_r^{(k)}$ from a trial distribution, $g(\boldsymbol{x})$. Here we select $g(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}_r^{(k)},\Sigma)$, where Σ is a diagonal matrix with small positive elements. Compute epipolar distances related to these new motion samples using **(1)** and evaluate weights $\boldsymbol{w}_r^{(k)}$ using **(3),** according to the weight evaluation rule in importance sampling.

$$\boldsymbol{w}_r^{(k)} = \frac{\boldsymbol{p}((\boldsymbol{u}',\boldsymbol{v}') \mid \boldsymbol{x}_r^{(k)})}{\boldsymbol{g}(\boldsymbol{x}_r^{(k)})} \quad \textbf{(3)}$$

where the likelihood is computed using **(2).**

c) Go to step a) and continue, till a sufficient number of seed samples have been drawn from the prior distribution.

The new samples and weights $\{\boldsymbol{x}_r^{(j)},w_r^{(j)}\}_{j=1}^{n}$ are then used to represent the posterior distribution of the camera motion parameters.

5. EXPERIMENTAL RESULTS

The proposed algorithm has been tested using both synthetic and real image sequences. One example is included here. In this example, two cameras with perpendicular looking directions were used to watch a parking lot. Each camera has tracked a walking person. The centroids of this walking person in two views at corresponding frames are used as matched feature points between two cameras. The two centroid trajectories are show in Figure 3. The motion estimation results are shown in Figure 4. The left column lists the initial camera motion distributions obtained using RANSAC and the right column shows the refined motion distributions after importance sampling. The refined camera motion distribution is more much accurate in terms of low uncertainties. Although the ground-truth of the camera motion is not available, the motion estimation results are

consistent to the knowledge of the rough relative position of the two cameras when the sequences were captured.

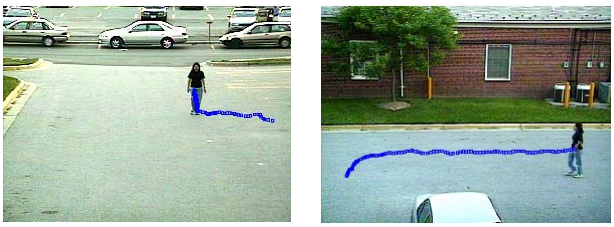

Figure 3. Human tracking results from two cameras

4. CONCLUSIONS

In this paper, we have proposed a two-step coarse-to-fine approach to robust Bayesian motion estimation of wide baseline cameras. RANSAC is deployed first to obtain an initial distribution of the motion parameters. Importance sampling is then applied to refine the distribution, using the initial distribution as prior knowledge of the camera motion. Weighted samples are obtained to represent the posterior distribution of the camera motion parameters. The proposed algorithm is robust to feature matching errors across wide baseline cameras.

5. REFERENCES

[1] Strecha, C., Tuytelaars, T., and Gool, L. V., Dense Matching of Multiple Wide-baseline Views, The Ninth IEEE International Conference on Computer Vision, Nice, France, 2003.

[2] Baumberg. A., (2000) "Reliable feature matching across widely separated views". *CVPR*, 2000.

[3] Xiao, J. and Shah, M., (2003) Two-Frame Wide Baseline Matching, The Ninth IEEE International Conference on Computer Vision, Nice, France, 2003.

[4] Lee, L., Romano, R. and Gideon S. Monitoring Activities from Multiple Video Streams: Establishing a Common Coordinate Frame. *IEEE Trans. PAMI*, vol. 22, pp. 758-767, 2000

[5] Fischler, M. A. and Bolles, R. C.,Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, Vol 24, pp 381-395, 1981.

[6] Doucet, A., Freitas, N. and Gordon, N, *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, Series Statistics for Engineering and Information Science, 2001

[7] Qian, G. and Chellappa, R., Structure from motion using sequential Monte Carlo methods, in *Proceedings of International Conference on Computer Vision*, vol II, pp. 614-621, Vancouver, BC, Canada, July 9-12 2001

[8] P. Torr and D. Murray, The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix, *International Journal of Computer Vision*, 24(3):271-300, 1997

[9] Z. Zhang, Determining the Epipolar Geometry and its Uncertainty: a Review, *International Journal of Computer Vision*, 27(2):161-195, 1998.

[10] Nister, D., An Efficient Solution to the Five-Point Relative Pose Problem, CVPR 2003

[11] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon, Bundle Adjustment - a Modern Synthesis, *Springer Lecture Notes on Computer Science*, Springer Verlag, 1883:298-375, 2000.

[12] Hartley, R. I., In Defense of the Eight-Point Algorithm, *IEEE Trans. PAMI*, vol. 19, pp. 580-593, 1997

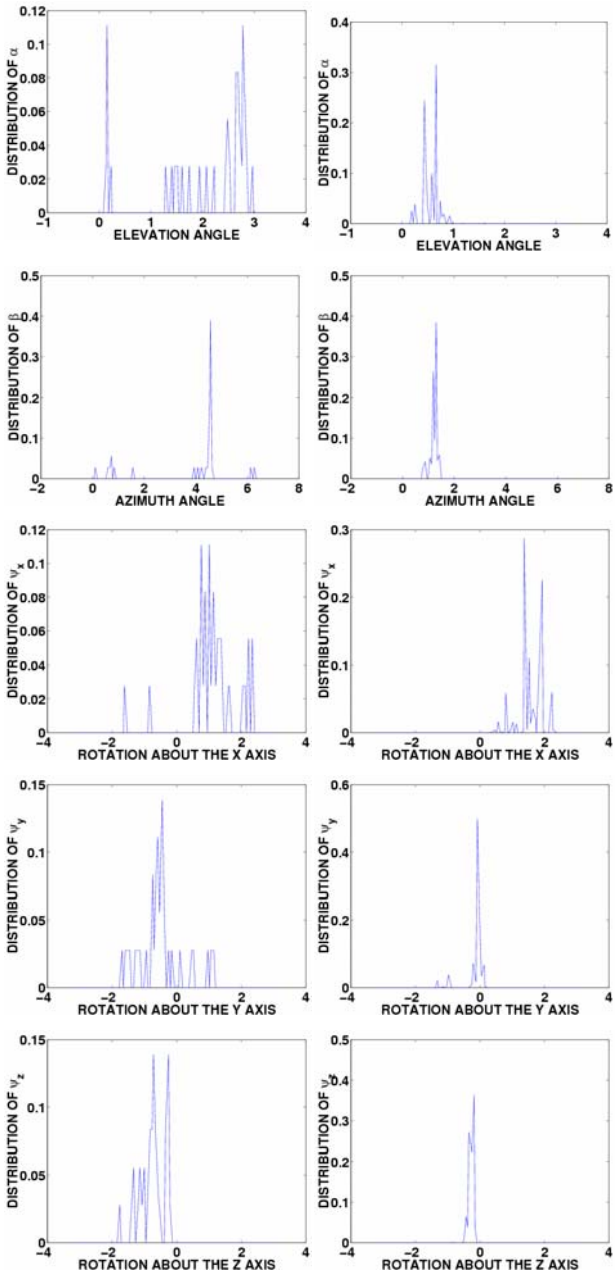

Figure 4. Camera motion estimation results. The left column presents the initial distributions obtained using RANSAC and the right column shows the refined distribution after importance sampling.