MAP ESTIMATION OF EPIPOLAR GEOMETRY BY EM ALGORITHM AND LOCAL DIFFUSION

Wenfeng Li *Baoxin Li*
Department of Computer Science & Engineering
Arizona State University, Tempe, AZ 85282 USA
E-mail: {wenfeng.li, baoxin.li}@asu.edu

ABSTRACT

Finding epipolar geometry for two images is a fundamental problem in computer vision. While this typically relies on feature point correspondence, the epipolar constraint can also be used for improving the accuracy of correspondence. We propose a probabilistic framework for estimating the epiploar geometry, in which the geometry and the feature correspondence are estimated iteratively at the same time. Using the EM algorithm to maximize a posteriori, our approach updates feature correspondence with estimated epipolar geometry. The correspondence is further improved with local diffusion on a prior Markov Random Field model. In turn, more accurate epipolar geometry is recovered. Experiments show this approach produces more accurate fundamental matrix compared with typical methods and can handle some challenging situations such as view rotation and scale changes.

Index Terms— Epipolar Geometry, EM Algorithm, MAP, local diffusion

1. INTRODUCTION

Two perspective images of a common scene are constrained by the so-called epipolar geometry [1]. It can be described as, given any point $\boldsymbol{x}$ in the first image, if it is the projection from a 3D point $\boldsymbol{X}$ in the scene, the projection $\boldsymbol{x'}$ in the second image must be on a line determined by $\boldsymbol{x}$ which is called epipolar line. The epipolar geometry can be written as

$$\boldsymbol{xFx'} = 0$$

where $\boldsymbol{F}$ is a 3×3 matrix called the fundamental matrix.

A typical way to find the epipolar geometry from two images includes two main stages. In the first stage, two sets of feature points are detected in the two images separately, and then inter-image correspondence is established for the features. Among others, commonly-used feature detection and correspondence methods include those based on Harris corner detector [2] and those using Shift Invariant Feature Transform (SIFT) [3]. In the second stage, the fundamental matrix is estimated using the corresponded features. This usually starts with a linear solution, followed by nonlinear optimization (e.g., LMedS [4]). Most methods for this stage can be viewed as maximum likelihood estimation (MLE), and the quality of the estimate mainly relies on the accuracy of the feature correspondence, which remains a challenge despite many years of research. On the other hand, if the epipolar constraint is known, it makes the correspondence problem much easier since possible matches for a given feature are constrained to points on an epipolar line.

In this paper, we propose a probabilistic framework, in which the Expectation-Maximization algorithm (EM) is used to estimate the epipolar constraint and feature correspondence iteratively at the same time. Explicit correspondence is avoided by a probabilistic representation. For more reliable probabilistic description of feature matching between two images, we encode the smoothness on adjacent pixels in two ways. Pixels around a corner within a small patch are treated as a unit by enforcing planar constraint. A prior is approximated by a Markov Random Field model to aggregate support from neighbor regions. We use local diffusion to solve this model. Therefore this problem becomes *Maximum A Posteriori* (MAP) estimation.

2. RELATED WORK

MAP has been proposed to replace simple RANSAC and MLE methods for epipolar geometry estimation [5][6]. Cham and Cipolla [5] present a multiscale method for feature matching in uncalibrated image mosaicing. They use parameters propagated from a coarse level as prior for the fine level estimation. This is pointed out as being problematic in [6] since the fine level data is not independent of that of the coarse level. Torr and Davidson [6] also use a multiscale scheme. The posterior distribution is passed from coarse level to fine level by the technique of sampling-importance-resampling and MCMC. While such multiscale methods prove to be an efficient way in solving the matching problem, texture details may be lost in the coarse level, causing some feature points useless. Not limiting to feature points, Domke and Aloimonos [7] present a probabilistic framework on frequency space, where all points could be used for matching after applying Gabor filter. A practical difficulty is that the method can be computationally very costly.

3. PROPOSED METHOD

3.1. EM Algorithm-based Formulation

We first describe a general formulation of the problem. Given two perspective images $\mathbf{I} = \{\mathbf{I}_0, \mathbf{I}_1\}$ from a common scene, a feature point detector generates two sets of feature

points $\mathbf{U}_0=\{\boldsymbol{u}_{0j}\}$ and $\mathbf{U}_1=\{\boldsymbol{u}_{1k}\}$, where $j=1,\ldots,m$ and $k=1,\ldots,n$, with m and n being the numbers of features in the two images respectively. These points are projection of a set of 3D points $\mathbf{X}=\{\boldsymbol{x}_i\}$, $i = 1,\ldots q$. In general m is not equal to n, and thus q is also undetermined. For some point in $\mathbf{I}_0$, there may be no corresponding point in $\mathbf{I}_1$ and vice versa. But for each point in $\mathbf{X}$, there must be a pair of projections in $\mathbf{U}_0$ and $\mathbf{U}_1$. Since the correspondence is unknown, a hidden value $\mathbf{J}$ is introduced to model the projection from $\mathbf{X}$ to $\mathbf{U}_i$, $i=0,1$, which also can be viewed as a geometry transformation from $\mathbf{X}$ to $\mathbf{U}_i$. Since the viewpoints of $\mathbf{I}_0$ and $\mathbf{I}_1$ are in general different, a transformation $\mathbf{R}$ is further introduced. The goal is to maximize the posteriori probability of parameters $\mathbf{\Theta}=\mathbf{G}$ given the data measurement $\mathbf{U}$ (the union of $\mathbf{U}_0$ and $\mathbf{U}_1$) and the hidden value $\mathbf{T}=\{\mathbf{J}, \mathbf{R}\}$, where $\mathbf{G}$ is the epipolar geometry expressed by a fundamental matrix. This is equal to maximizing the logarithm of the joint distribution, which is proportional to the posteriori by the Bayes rule

$$\mathbf{\Theta}^* = \arg\max_{\Theta} \log\{P(\mathbf{U},\mathbf{\Theta})\} = \arg\max_{\Theta} \log\sum_{\mathbf{T}} P(\mathbf{U},\mathbf{T},\mathbf{\Theta}) \quad (1)$$

This computation needs to be done with all possible $\mathbf{T}$, which results in combinatorial explosion and should be avoided. This is exactly the prime motivation of using EM algorithm here. By Jensen's inequality, a lower bound is obtained by transforming a log of sums to a sum of logs

$$\begin{aligned}\log\sum_{\mathbf{T}} P(\mathbf{U},\mathbf{T},\mathbf{\Theta}) &= \log\sum_{\mathbf{T}} f^t(\mathbf{T})\frac{P(\mathbf{U},\mathbf{T},\mathbf{\Theta})}{f^t(\mathbf{T})} \\ &\geq \sum_{\mathbf{T}} f^t(\mathbf{T})\log\frac{P(\mathbf{U},\mathbf{T},\mathbf{\Theta})}{f^t(\mathbf{T})}\end{aligned} \quad (2)$$

Where $f^t(\mathbf{T})$ is defined as a probability distribution given an arbitrary transformation $\mathbf{T}$ from the hidden value space and it is constrained by

$$\sum_{\mathbf{T}} f^t(\mathbf{T}) = 1$$

A Lagrangian function can be written as

$$L(f^t) = \sum_{\mathbf{T}} f^t(\mathbf{T})\log\frac{P(\mathbf{U},\mathbf{T},\mathbf{\Theta})}{f^t(\mathbf{T})} + \lambda\left(\sum_{\mathbf{T}} f^t(\mathbf{T}) - 1\right)$$

Taking the derivative and solving the Lagrange optimizer, we obtain

$$f^t(\mathbf{T}) = P(\mathbf{T}\mid\mathbf{U},\mathbf{\Theta}^t)$$

From (2), given current guess of parameters $\mathbf{\Theta}^t$, an optimal lower bound of the objective function is

$$\begin{aligned}&\sum_{\mathbf{T}} f^t(\mathbf{T})\log\frac{P(\mathbf{U},\mathbf{T},\mathbf{\Theta}^t)}{f^t(\mathbf{T})} \\ &= \sum_{\mathbf{T}} P(\mathbf{T}\mid\mathbf{U},\mathbf{\Theta}^t)\log\frac{P(\mathbf{U},\mathbf{T},\mathbf{\Theta}^t)}{P(\mathbf{T}\mid\mathbf{U},\mathbf{\Theta}^t)} = \log P(\mathbf{U},\mathbf{\Theta}^t)\end{aligned}$$

Therefore the two step EM algorithm can be written as:

E-step: Given current guess $\mathbf{\Theta}^t$, compute

$$f^t(\mathbf{T}) \equiv P(\mathbf{T}\mid\mathbf{U},\mathbf{\Theta}^t) \quad (3)$$

M-step: Update $\mathbf{\Theta}$ by maximizing the lower bound of objective function as

$$\mathbf{\Theta}^{t+1} = \arg\max_{\Theta}[\log P(\mathbf{U},\mathbf{\Theta}^t)] \quad (4)$$

3.2. The Likelihood Model

To obtain the likelihood of data $\mathbf{U}$ given parameters $\mathbf{\Theta}$, a simple approach is to consider the probability of each point in $\mathbf{U}$ as i.i.d. (we also assume that $\mathbf{U}_0$ and $\mathbf{U}_1$ play symmetric roles in the computation). For a 3D point $\boldsymbol{x}_j$, assuming it projects to $\boldsymbol{u}_{ia}$ on $\mathbf{I}_i$. The probability that it also projects to $\boldsymbol{u}_{(1-i)b}$ on $\mathbf{I}_{1-i}$ is determined by both the geometry constraint $\mathbf{G}$ and the distance between $\boldsymbol{u}_{ia}$ and $\boldsymbol{u}_{(1-i)b}$ in terms of feature descriptor measurement, noted as p_G and p_M. The objective function becomes

$$\begin{aligned}\log P(\mathbf{U},\mathbf{\Theta}) &= \sum_i \log\prod_j p(u_{ij}\mid\Theta)p(\Theta) \\ &= \sum_i\sum_j \log\{p_G(u_{ij}\mid\Theta)p_M(u_{ij}\mid\Theta)p(\Theta)\}\end{aligned} \quad (5)$$

For the geometry constraint, we adopt the commonly used reprojection error which is

$$d_G = d(u_{ia}, Fu') + d(u', Fu_{ia})$$

where d is a Euclidian distance function. Note that not every point has a corresponding point in another image due to occlusion and failure of feature detector. We use a contaminated Gaussian model as a robust penalty function

$$p_G(u\mid\Theta) = (1-\varepsilon_G)\exp(-d_G^2/2\sigma_G^2) + \varepsilon_G \quad (6)$$

The matching probability p_M is defined as

$$p_M(u\mid\Theta) = (1-\varepsilon_M)\exp(-d_M^2/2\sigma_M^2) + \varepsilon_M \quad (7)$$

where d_M is Euclidian distance defined on pixel intensity or color values, which depends on image format.

3.3. Feature Matching

We use Harris corner as feature point. For each feature point, we assume that it is at the center of a small planar surface comprised of 5×5 pixels. When a planar patch matches to another view, they undergo certain geometry transformation which can be modeled by a homography, written as

$$s\begin{bmatrix}x\\y\\1\end{bmatrix} = \begin{bmatrix}A & t\\ v & 1\end{bmatrix}\begin{bmatrix}x'\\y'\\1\end{bmatrix}$$

where A is a 2×2 matrix, t is a translation vector and s is a scale factor. Since the translation is already considered in the projection hidden value $\mathbf{J}$ and if the patch is centered well with the feature point, t can be set to zero. A can be decomposed as

$$A = R(\theta)R(-\varphi)DR(\varphi),\ D = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda 2\end{bmatrix}$$

where $R(\theta)$ and $R(\varphi)$ is a rotation matrix with rotation angle θ and φ, θ is the patch rotation angle, and D and φ determines the patch deformation ratio and direction [8]. In

practice, patch matching is only sensitive to rotation while relatively insensitive to deformation given the fact that most patches in the images are not drastically deformed. Therefore we can safely approximate A with a rotation matrix parameterized by θ. 16 evenly spaced rotation angles are tested at the first matching iteration, sum of absolute difference (SAD) is used as matching score and the rotation angle with the smallest SAD is picked as an initial value for the EM algorithm.

When the fundamental matrix is available, let l and l' be the epipolar lines that pass the two patches to be matched, it is easy to prove that two continuous pixels a and b along the epipolar lines will not change their relative position. As shown in Figure 1, the patch rotation angle θ is obtained by

$$\theta = \cos^{-1}(l \cdot l')$$

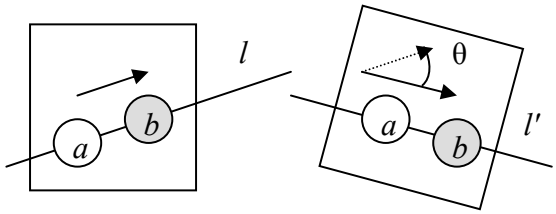


Figure 1: Estimate patch rotation from epipolar lines.

The scale factor s can be estimated as follows: Draw an epipolar line l with the center point, draw two epipolar lines l', l'' with the top center point a and the bottom center point b. The line segment that is perpendicular to l and through the center point in the transformed image intersects l' and l'' at a' and b'. Then s is approximately estimated by

$$s = \frac{\overline{a'b'}}{\overline{ab}}$$

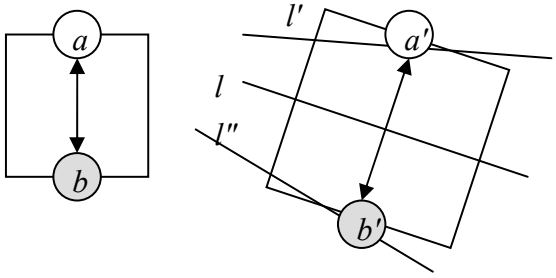


Figure 2: Estimate patch scale from epipolar lines.

When a detected feature lies on object contours, the patch model is not accurate if pixels of the patch come from both the background and the foreground. Such features give unreliable information for matching but their inconsistent votes will typically be dominated by other correctly matched corners, assuming that more features do not lie on the object contours.

3.4. The Prior Model

In the likelihood model we consider the probability of each feature as i.i.d. for simplicity. In real images, features are highly correlated. Adjacent pixels are still adjacent after being transformed by function **T** to an image from another view, which is also called smoothness constraint. We use a Markov Random Field (MRF) [9] to encode preference of surface smoothness into a prior model p_P

$$p_P(\mathbf{T}) = \frac{1}{Z_P}\exp(-E_P(\mathbf{T}))$$

where Z_P is a normalizing factor and the potential function E_P is the sum of clique potentials. The idea of using MRF is to aggregate support from neighbors and to use this support as a prior. There are two potential issues with using MRF here. First, we do not have cliques for sparse feature points; secondly, it requires a large memory and expensive computation for probabilities of all possible **T.** We propose a local diffusion strategy to construct cliques for each feature point by expanding to its neighbor regions. The diffusion is undertaken only in local minimums obtained from the likelihood computation. We use nonlinear diffusion from [10] for its capability to handle multiple ambiguities in clique. The implementation can be described as follows. For a point $\boldsymbol{u}_{0j}$, if it finds a local minimum of match score at $\boldsymbol{u}_{1k}$ with **T',** we also compute matching scores for 8 neighbor patches of $\boldsymbol{u}_{0j}$ with **T'** and a shift on disparity d=-3,…,+3. We rewrite the log of probabilities in Eqn. (5) with E to represent energy. The diffusion process updates E as

$$E \leftarrow E_0 + \sum_{N_8} E_S$$

where E_0 is the initial energy from $\log(p_M p_G)$. E_S is the log of the smoothed probability distribution p_S

$$p_S = \sum_d \exp(-d^2)\, p_M\, p_G$$

The updating is iterated only twice since the diffusion region is very small.

3.5. The Complete Algorithm Workflow

With all components introduced above, we present the complete algorithm as follows:

Step 1. For each image, a set of Harris corners are detected. Feature matching is performed with all combination of corner pairs from two images. The matching score is stored to compute p_M.

Step 2. For a point $\boldsymbol{u}_{0j}$, if its best match is $\boldsymbol{u}_{1k}$, and the best match for $\boldsymbol{u}_{1k}$ is also $\boldsymbol{u}_{0j}$, such consistent matched pairs are selected as seeds. We run random sampling and standard 7-point algorithm to compute fundamental matrix from these seeds. The difference from normal RANSAC is that the criterion is the probability in Eqn. (5) and all points are computed, not only those matched pairs.

Step 3. The fundamental matrix that produces the best result will be used as the initial value for EM algorithm. Matching score is re-computed with the estimated fundamental matrix.

Step 4. Use random sampling or standard nonlinear optimization to find a better fundamental matrix based on

re-computed matching score. If no better result can be found, exit, otherwise go to step 3.

Note that when we use Eqn. (5) to maximize the posteriori, we do not enforce the uniqueness constraint. It means that one point may be matched to two or more points. The reason is that explicitly enforcing uniqueness is a deterministic method and to find exclusively matched pairs to maximize the posteriori requires a combinatorial explosion of enumeration of all possible pairs. Since local diffusion is used to enforce the smoothness constraint, it in some way remedies the loss of the uniqueness constraint and our experiments show that this works reasonably well with object occlusions (as in Figure 4).

4. EXPERIMENTAL RESULTS

Sample experimental results are listed in Figure 3-5. We compare our results with the implementation of typical RANSAC and LMedS from OpenCV [11]. While OpenCV works fine in normal situations and gives the same result as our method, there are difficult cases where it gives obviously wrong result, as in Figure 3. Although there are many parameters used in both the EM algorithm and local diffusion in our approach, we use the same set of parameters for all experiments, which shows the performance is robust to parameter settings.

5. CONCLUSION

We present an approach to estimating the epipolar geometry and feature point correspondence more accurately and efficiently with EM algorithm. Experiments show the effectiveness of the approach, in comparison with typical existing techniques. By applying local diffusion and adaptive matching, our feature point matching is found to be relatively invariant to image rotation and scale changes. We plan to extend the system to be invariant to illumination change and able to segment pixels on different surfaces. This would lead to a surface reconstruction method.

6. REFERENCES

[1] Q-T. Luong and O.D. Faugeras, "The fundamental matrix: Theory, algorithms, and stability analysis," *International Journal of Computer Vision*, 17(1), pp. 43-75, 1996.

[2] C. Harris and M. Stephens, "A combined corner and edge detector," Proceedings of the 4th Alvey Vision Conference, pp. 147-151, 1988.

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60(2), pp. 91–110, 2004.

[4] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," Tech. Rep. 2927, Institut National de Recherche en Informatique et en Automatique, July 1996.

[5] T.J. Cham and R. Cipolla, "A statistical framework for long-range feature matching inuncalibrated image mosaicing," In *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp.442-447, 1998.

[6] P.H.S. Torr and C. Davidson, "IMPSAC: synthesis of importance sampling and random sample consensus," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(3), pp.354-364, 2003.

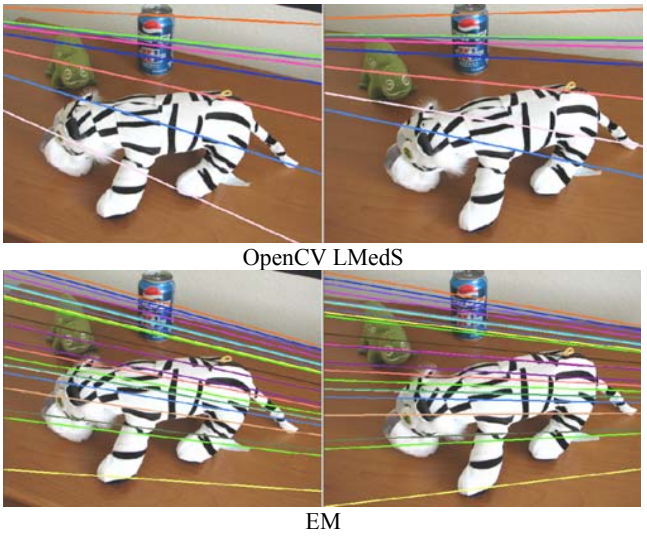
[7] J. Domke and Y. Aloimonos, "A probabilistic notion of correspondence and the epipolar constraint," *Proc.* 3DPVT (International Symposium on 3D Data Processing Visualization and Transmission), 2006.

[8] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge, UK, 2000.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6), pp.721-741, 1984.

[10] D. Scharstein and R. Szeliski, "Stereo matching with nonlinear diffusion," *International Journal of Computer Vision*, 28(2), pp.155-174, 1998.

[11] OpenCV, http://sourceforge.net/projects/opencvlibrary



OpenCV LMedS

EM

Figure 3: Top: Incorrect epipolar geometry due to mismatches on zebra patterns. Bottom: Correct epipolar geometry with our method.

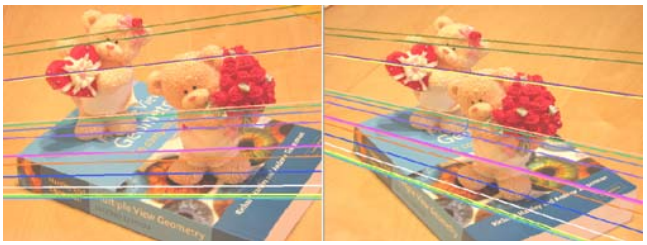


Figure 4: Epipolar geometry computed on two images with view rotation and object occlusions.

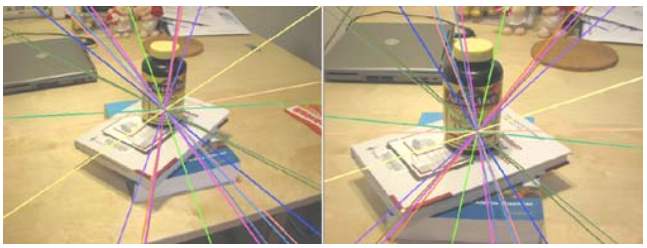


Figure 5: Epipolar geometry computed on two images with view rotation and scale change.