A Probabilistic Notion of Correspondence and the Epipolar Constraint

Justin Domke and Yiannis Aloimonos
Computer Vision Laboratory, Dept. of Computer Science
University of Maryland, College Park, MD 20742 USA
{domke,yiannis}@cs.umd.edu

Abstract

We present a probabilistic framework for correspondence and egomotion. First, we suggest computing probability distributions of correspondence. This has the advantage of being robust to points subject to the aperture effect and repetitive structure, while giving up no information at feature points. Additionally, correspondence probability distributions can be computed for every point in the scene. Next, we generate a probability distribution over the motions, from these correspondence probability distributions, through a probabilistic notion of the epipolar constraint. Finding the maximum in this distribution is shown to be a generalization of least-squared epipolar minimization. We will show that because our technique allows so much correspondence information to be extracted, more accurate egomotion estimation is possible.

1. Introduction

It can easily be shown that, mathematically, a small number of corresponding points can be solved, by way of the epipolar constraint, to find the exact egomotion. Yet, after decades of research no completely satisfactory algorithm exists to do this. The current state-of-the-art algorithms first match points between the images, then essentially try to find the largest subset of these correspondences yielding a consistent motion. We propose that, instead of matching points, while acknowledging that many of the computed matches will be incorrect, it would be better to compute a structure which can be found reliably, for every point. We propose such a structure in *probability distributions of correspondence.* Given a point, to find such a distribution is a well-posed problem. We suggest that this generalization results in much less correspondence information being given up. First, a correspondence distribution may be computed for any point in the scene, regardless of if it is a 'feature point'. Second, the probability distribution can reflect exactly what low-level information gives about the correspondence- even if that is a single "peak" at the correct point.

As an example of the use of such correspondence distributions, we develop a generalized notion of the epipolar constraint, which explicitly takes into account the uncertainty in the correspondences. Instead of giving an exact mathematical constraint on the motions, the generalized version gives a soft constraint, reflecting the ambiguities in the correspondence. We will show that this constraint is a natural generalization of previous work. Namely, if correspondences are known (with some finite precision) the most likely motion for our constraint corresponds exactly to the motion giving the minimum least-squares epipolar error.

We will observe two major advantages to this approach. First, motion can be found despite scenes which are difficult to match, because of repetitive structure, etc. Second, because a very large number of correspondence distributions are available the motion can be estimated more accurately. The abundance of correspondence information reduces the translation-rotation ambiguity.

1.1. Related Work

In general, high-level information is necessary to estimate correspondences. As such, it is likely that correspondences cannot be reliably estimated from low-level measurements [12]. Commonly, many possible correspondences are computed, and a robust algorithm such as RANSAC [6] is used to search for a set of mutually coherent matches. A less common approach is to represent the ambiguity in correspondence more explicitly. Clocksin [2] estimates traditional flow vectors at each point, by first estimating flow probability distributions, and then combining this information through using spatiotemporal support regions. Simoncelli *et al*. [17] create probability distributions over the optical flow by assuming image gradients are corrupted by a Gaussian noise model. These distributions is then used to estimate traditional optical flow vectors. Rosenberg and Werman [16] use full, nonparametric probability distributions of correspondence for object tracking.

Our method to compute correspondence probability distributions uses the phase of tuned Gabor filters. We use the efficient Gabor filter implementation of Nestares et al. [13]. Stereo disparity [7] and optical flow [8] algorithms often use phase to guide the correspondence problem.

Egomotion and Structure from Motion are among the

most heavily researched areas of computer vision research, and rather than attempting to summarize all references, the reader is referred to a survey [14] and to more general recent books [9] [4]. Wexler et al. [1] present a method which aggregates information over multiple image pairs to learn the epipolar geometry. Dellaert et al. [3] present an algorithm which iteratively computes probabilities over both correspondence and motion through the Expectation-Maximization framework.

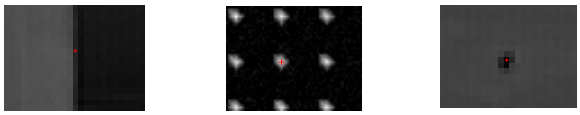
The approach most similar to the one here is by Makadia *et al.* [11]. There, the authors use traditional feature points, but rather than committing to an explicit matching, they search for a motion such that each feature point has a compatible point in the other image satisfying the epipolar constraint. Their approach can be phrased probabilistically. The principal difference with the current work is that we extract correspondence information for all points in the image, with out use of a feature detector. This means both that additional correspondence information is available, and that it is not necessary for the same point to be reliably detected as a feature. This drastically increased amount of correspondence information results in major increases in accuracy and robustness.

2. Correspondence

2.1. Ambiguities in Correspondence

There are 3 conditions affecting correspondence, resulting in it being hard to estimate: the aperture effect, repetitive structure, and the finite resolution of images. The aperture effect is illustrated in Fig. 1 (a). Suppose correspondence is sought for the marked point. Clearly, this is impossible- it can only be constrained to lie along the edge in the other image. Repetitive structure, meanwhile, is seen in Fig. 1 (b). Again, the correspondence cannot be found- instead the images constrain the correspondence to a set of disjoint points. Generally, feature based techniques will ultimately reject both of the above points, either because they will not be detected as feature points, or because the matching process will not have a single, 'outlying' corresponding point. In Fig. 1 (c), the correspondence would seem to be unambiguous. Even here, however, it is important to realize that the correspondence is only known with a finite precision, due to the resolution of the images.

Of course in realistic images, most points will not fall cleanly into one of the three categories above. For a given point, certain correspondences will be likely, others unlikely, and others virtually impossible. By computing a probability distribution, whatever ambiguity happens to be at hand can be represented.



(a) (b) (c)

Figure 1: In different situations, different information is available about correspondence.

2.2. Correspondence Distributions

How, then, can correspondence probability distributions be computed? Our focus in this paper is on the *usage* of these distributions, but we give the following simple approach used in our experiments. It is very likely that it can be improved with further research. The large literature on point matching is non-trivial to apply to the probabilistic case because the usual goal is only to find the *most likely* corresponding point. We need a quantitative measure of the relative likelihood of all points.

Our method uses Gabor filters tuned to different orientations and scales. It exploits the fact that for a given filter, matching points will have matching phase. (We do not use the amplitude of the filter response- we note tangentially that this gives the technique a high degree of contrast invariance.) Since we are computing correspondence probabilities over a quantized grid, the phase will not exactly match. For the filter with orientation γ and scale l, denote the phase by $\phi_{l,\gamma}$. Given this single filter, we take the probability that s matches most closely to $\hat{q}$ to be proportional to $\exp(-[\phi_{l,\gamma}(s) - \phi_{l,\gamma}(\hat{q})]_\pi^2) + 1$. Here $[\theta]_\pi$ denotes the principal angle of θ, given by adding multiples of 2π so that it is in the range $[-\pi, \pi)$. Thus, when all filters are combined, we have

$$\rho_s(\hat{q}) \propto \prod_{l,\gamma} (\exp(-[\phi_{l,\gamma}(s) - \phi_{l,\gamma}(\hat{q})]_\pi^2) + 1) \qquad (1)$$

The constant of 1 increases robustness to noise by limiting the influence of any single filter. Notice that because the Gabor filter responses are not fully orthogonal, the assumption of independence used to combine them here is only approximate. In our implementation, we found it convenient to use a threshold $\rho_{\min}$, where if the probability of a certain correspondence was below $\rho_{\min}$, it was set to zero, and removed from further consideration. Correspondence distributions are illustrated in Figs. 2 and 3 . Parts (c)-(j) show the probability that the point marked in (a) corresponds to each possible location in (b). Probabilities are encoded as color. It can be seen that the large scale Gabor filters (i.e. (c)) provide different distributions that the small scale filters

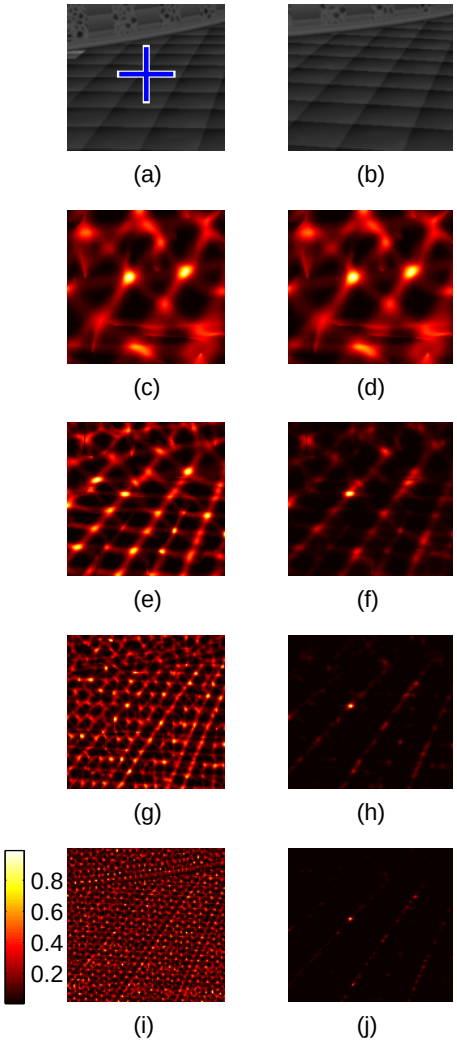


Figure 2: The computation of a correspondence distribution. (a) first image. (b) second image, over which correspondence is being considered. (c)-(i) distributions for specific, decreasing scales, each with all orientations. (d)- (j) distribution considering all previous scales. (j) final distribution.

(i.e. (i)). Nevertheless, it is shown in the right column that the combination of all filters leads to an excellent distribution.

A further problem exists in the finite resolution of images. Even in the best possible situation- correspondence for a 'corner', for example- the matching points will not be exactly known. We deal with this in a simple way. After finding the probability that a point s corresponds most closely to some *pixel* $\hat{q}$, we take the probability that it corresponds to some arbitrary point q (not necessarily having integer coordinates) by:

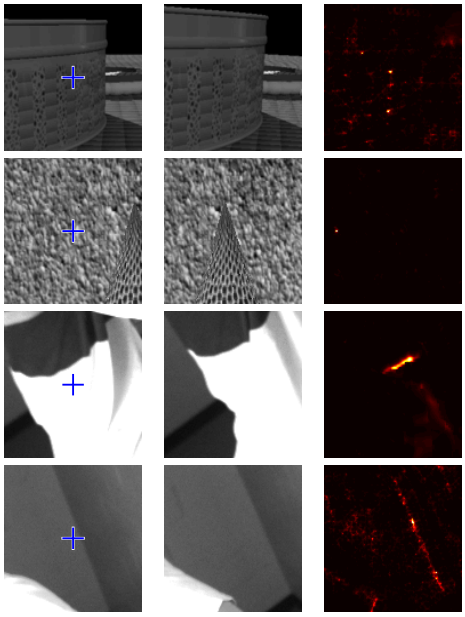


Figure 3: Several example correspondence probability distributions. The left column shows the first image, the middle shows the second image, and the right column shows the probability over each point in the second image.

$$\rho_s(q) \propto \max_{\hat{q}} \rho_s(\hat{q}) \exp(-|q-\hat{q}|^2) + \alpha \qquad (2)$$

Later in the paper, we will show that using this results in the most likely motion under our framework being exactly the motion with the least-squares epipolar error, in the case of known correspondences.

To reflect the possibility that the phase information is unreliable, we add a constant of α. This happens for example, if the point corresponding to s were independently moving, or were to become occluded in the second image. Notice that adding this constant is equivalent to taking a certain probability that the image information is unreliable, in which case a flat distribution is appropriate.

Since a filter with a given scale and orientation is assumed to match with the same filter in the second image, this method is only appropriate for situations with limited rotation or scale change.

3. The Probabilistic Epipolar Constraint

Supposing that the correspondences have been computed, what do they tell us about the motion? We would like to generalize the epipolar constraint to operate on probability distributions of correspondence. Our constraint on

the motion from one correspondence distribution is the following: *the probability of a motion is proportional to the maximum probability correspondence satisfying the epipolar constraint for that motion.*

So, if we are considering the correspondence of a point s, and E is the essential matrix we have

$$\rho_s(E) \propto \max_{q:q^T Es=0} \rho_s(q) \tag{3}$$

We can think of the traditional epipolar constraint as specifying $\rho_s(\hat{q})$ as 1 for the known correspondence, and 0 otherwise. We will show in the discussion that this, along with the Gaussian distribution about each point from the previous section is equivalent to least-squares epipolar minimization.

4. Egomotion

The probability of a motion is given by combining the information given by all points:

$$\rho(E) \propto \prod_s (\rho_s(E)) = \prod_s (\max_{q:q^T Es=0} \rho_s(q)) \tag{4}$$

Substituting the expression for $\rho_s(q)$ (Eqn. 2), we obtain:

$$\rho(E) \propto \prod_s (\max_{q:q^T Es=0} \max_{\hat{q}} \rho_s(\hat{q}) \exp(-|q-\hat{q}|^2|) + \alpha) \tag{5}$$

Hence,

$$\rho(E) \propto \prod_s (\max_{\hat{q}} \max_{q:q^T Es=0} \rho_s(\hat{q}) \exp(-|q-\hat{q}|^2|) + \alpha) \tag{6}$$

Notice that in this expression, we do not need to explicitly find the point q. We need only the minimum distance from $\hat{q}$ of any point from the line Es. We can therefore write the above expression in the final form, in which it is computed as:

$$\rho(E) \propto \prod_s (\max_{\hat{q}} \rho_s(\hat{q}) \exp(-(\hat{q}^T l_{(E,s)})^2) + \alpha) \tag{7}$$

where $l_{(E,s)}$ is the line Es, normalized such that $r^T l_{(E,s)}$ is the perpendicular distance from r to the line Es on the image plane:

$$l_{(E,s)} = \frac{Es}{\sqrt{(E_1 s)^2 + (E_2 s)^2}} \tag{8}$$

We denote the ith row of E by E_i. In all experiments using this probabilistic framework, $\alpha = 1/160$ was used. Our experiments suggest that the exact choice of α is unimportant.

4.1. Optimization

Given a single motion, E, its probability can be very quickly computed. Still, because there are 5 degrees of freedom, computing a full motion probability distribution is problematic- computational considerations demand such a coarse sampling of each dimension that the entire peak of the distribution may be missed. In our experiments, we will maximize the motion function through a simple heuristic optimization. First, random sampling (t on the sphere with $|t| = 1$, ω such that $|\omega| \leq .1$) is used at approximately 2500 points. The probabilities are then computed for each point, taking $E = [t]_\times R(\omega)$. Next, the Nelder-Mead simplex search method is used at the 100 highest scoring samples. The final maximum probability sample found is taken as the result. In practice, we found that several of the 100 searches resulted in very close answers, suggesting that missing the global maximum altogether is unlikely. This is consistent with results reported elsewhere for Egomotion techniques using nonlinear functions [15] [18] suggesting there will be several (but only several) local minima. Although this is in a sense a brute-force maximization, in practice the slowest part of our technique is often the computation of the correspondence distributions.

Our implementation for the full algorithm, taking images as input, and yielding egomotion as output, is available on the authors website.

5. Experiments

5.1. Synthetic Images

As a first test of the technique, we prepared a synthetic 3D model using the POV-ray software. This model was quite difficult, containing a great deal of repetitive structure. We rendered two sequences using this model- the first with a translation along the y-axis, while the second sequence represents a translation along the z-axis. Each sequence contained a small rotation. These motions were selected because the literature suggests that the rotation-translation ambiguity depends on the direction of translation but not the rotation [5] [18].

For each of these sequences we generated 10,000 correspondence probability distributions. For various sizes, we generated 50 random subsets of correspondences, and ran the algorithm was then run on each subset. Due to the computationally demanding nature of this experiment, only 25 searches were used in the second half of the motion optimization procedure. In Fig. 5, we plot the mean errors for each size, using the know ground truth motion, t_0, ω_0. The egomotion estimates continued to improve with more matches. For a rigorous comparison, we used the well known SIFT features [10] to provide matches. These matches were then manually filtered to provide a set of correct matches. One frame is shown in Fig. 4 with the inlying

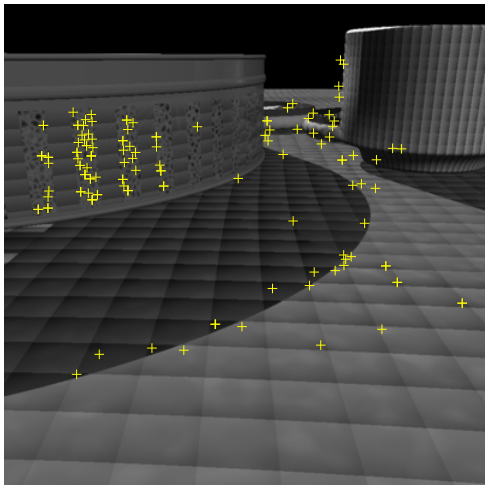


Figure 4: A frame from the synthetic model, with inlying feature points marked.

features marked. The algorithm was run on 50 random subsets of these matches for each size, with $\alpha = 0$. (Again, this is equivalent to least-squares epipolar minimization. See Section 6.1.) Though the SIFT features performed remarkably well, generating a large number of correct matches, the probabilistic approach is able to attain similar or lower errors. This is despite the fact that the probabilistic approach is fully automatic.

For another way of looking at these same results, Fig. 6 shows all 10 possible projections, for the algorithm run on the z-translation sequence, with 64 correspondence probability distributions. The projection of t_x vs. ω_y shows the rotation-translation ambiguity- a change in the translation can be compensated with a change in the rotation to yield a motion similarly consistent with the correspondence information. In Fig. 7 we see the same t_x vs. ω_y projection for several numbers of correspondences. We can observe that with a small amount of correspondence information, repetitive structure may result in a very wrong motion estimate. As the amount of correspondence information is increased this behavior disappears completely, but the egomotion estimate continues to improve, essentially because the translation-rotation ambiguity continues to decrease.

5.2. Real Images

Figure 8 shows epipolar lines and the projected solutions found for a real sequence, taken outdoors. This sequence was chosen for their difficulty- a great deal of repetitive structure is present. Notice that as the amount of correspondence information is increased the solutions all converge to a very small area.

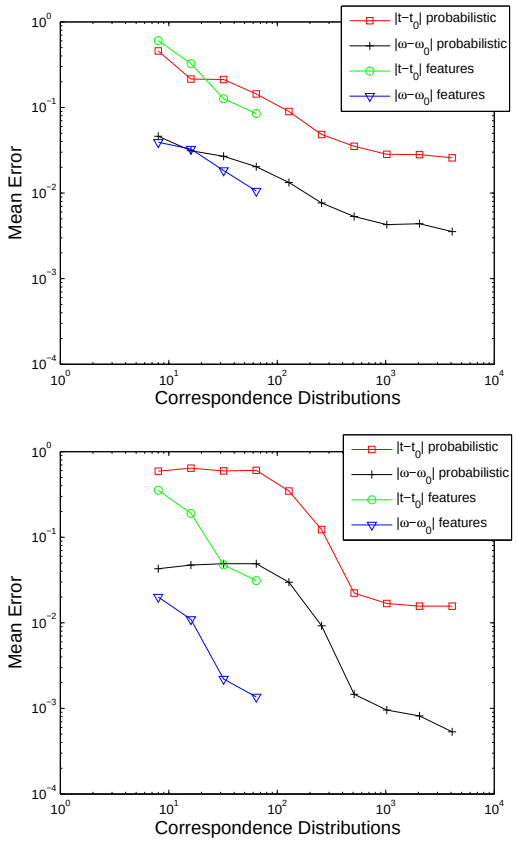


Figure 5: Errors for the synthetic sequences, with various number of correspondences or correspondence distributions. Top: translation along y-axis. Bottom: Translation along z-axis.

For a slightly different type of experiment, 9 shows epipolar lines and projected solutions for a wide-baseline indoor sequence. Whereas previous experiments used a local search region to create correspondence distributions, the wide-baseline nature of this sequence demands that search is done over the entire images. Note that the time complexity to create these distributions will then be $O(M \cdot N)$, where M is the number of distributions created, and N is the number of pixels in the image. As in other experiments, distributions are taken for points randomly distributed over the entire image.

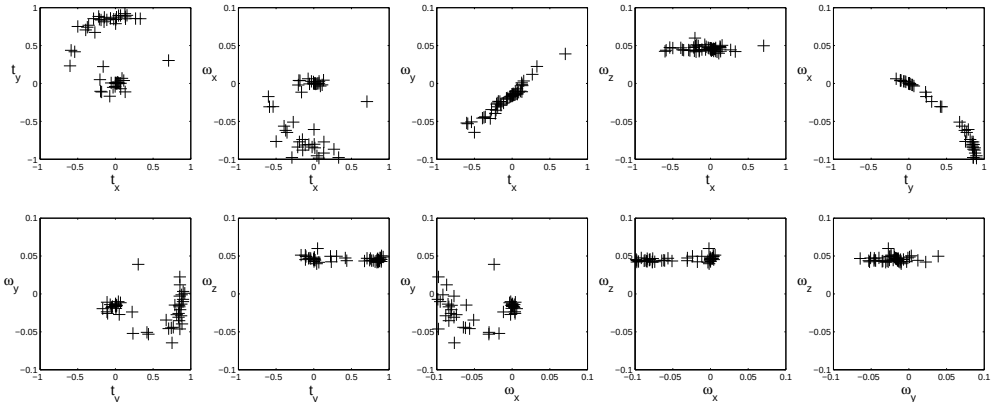


Figure 6: All 10 projections of the solutions found for the synthetic sequence with translation along the z-axis, with 64 correspondence distributions.

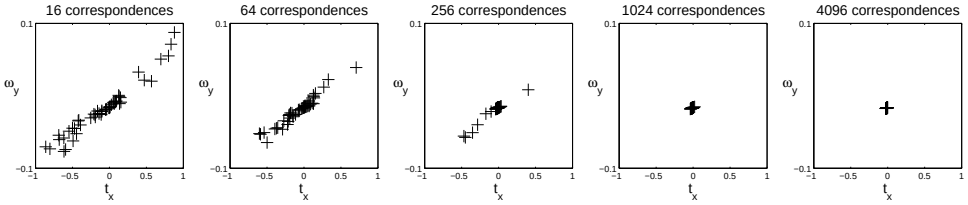


Figure 7: Projections of the solutions onto the t_x and ω_y plane, for the synthetic sequence with translation along the z-axis.

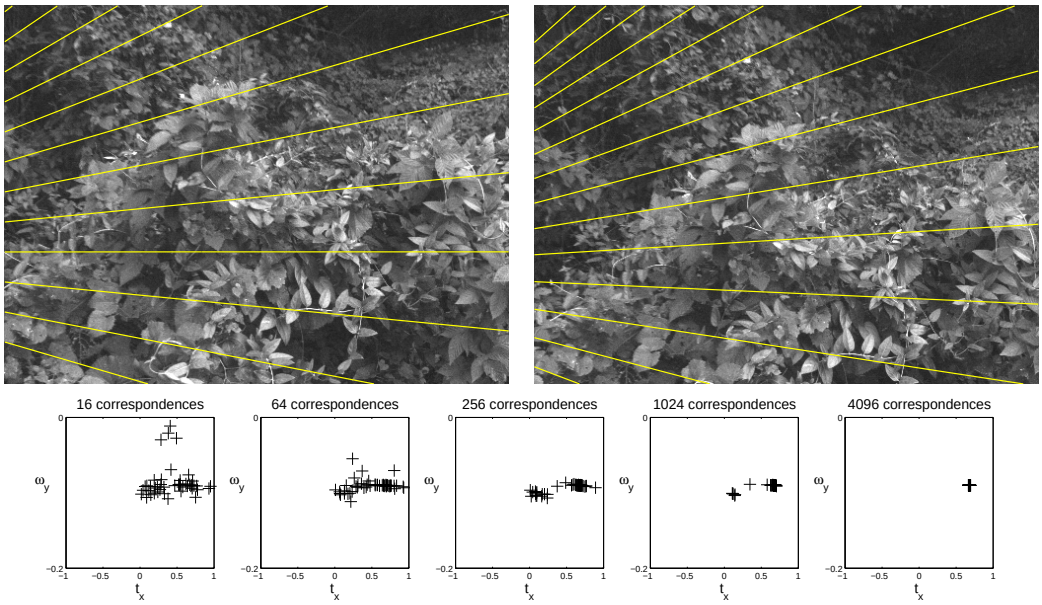


Figure 8: Results for a real sequence. Top: Epipolar lines found using all 10,000 correspondence distributions. Bottom: projected solutions using various numbers of distributions.

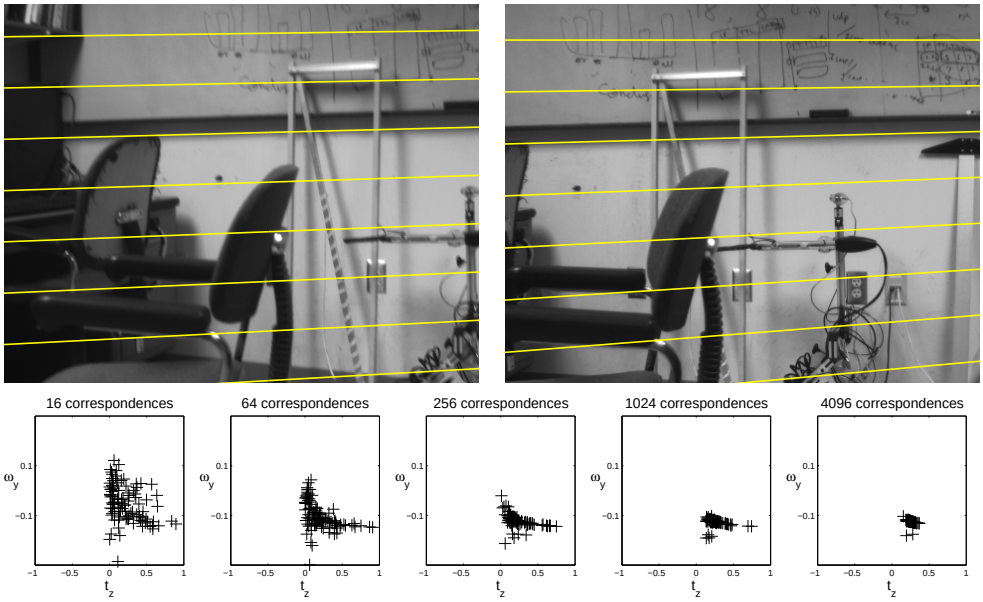


Figure 9: Results for a real sequence. Top: Epipolar lines found using all 10,000 correspondence distributions. Bottom: projected solutions using various numbers of distributions.

6. Discussion

6.1. Equivalence to Least-Squares Epipolar Minimization

We have mentioned several times that our approach reduces to least-squares epipolar minimization in the case of known correspondences. We will show this now. Suppose we have a set of manually selected matches $s_i \leftrightarrow \hat{q}_i$. Thus, for each s_i, we will have $\rho_{s_i}(\cdot) = 1$ for $\hat{q}_i$, and 0 otherwise. Furthermore, because of the known correspondence, α should be set to zero. If $\hat{E}$ is the most likely motion under the probabilistic framework, then:

$$\hat{E} = \arg\max_E \prod_{s_i} \max_{q:q^T E s_i = 0} \rho_{s_i}(q) \tag{9}$$

Now, since $\rho_{s_i}(q) \propto \max_{\hat{q}} \rho_{s_i}(\hat{q}) \exp(-|q-\hat{q}|^2) = \exp(-|q-\hat{q}_i|^2)$,

$$\hat{E} = \arg\max_E \prod_{s_i} \max_{q:q^T E s_i = 0} \exp(-|q-\hat{q}_i|^2) \tag{10}$$

Now using the same notation for $l_{(E,s)}$ introduced earlier, we have

$$\hat{E} = \arg\max_E \prod_{s_i} \exp(-(\hat{q}_i^T l_{(E,s_i)})^2) \tag{11}$$

Since the argument maximizing the quantity on the right will also maximize its logarithm, we have

$$\hat{E} = \arg\max_E \sum_{s_i} -(\hat{q}_i^T l_{(E,s)})^2 \tag{12}$$

Hence,

$$\hat{E} = \arg\min_E \sum_{s_i} (\hat{q}_i^T l_{(E,s)})^2 \tag{13}$$

This is the exact expression for the motion minimizing the least-squares epipolar error. In this way, our approach is a natural generalization of previous work.

6.2. Alternatives

The reader may object to several seemingly arbitrary choices made in our approach. When we take the Gaussian distribution about each point Eq. 2, why was a $\max$ chosen over all points $\hat{q}$, rather than a sum? Similarly, when we state our probabilistic notion of the epipolar constraint in Eq. 3, why is the probability of the motion E taken to be proportional to the maximum probability point q, rather than an integral over all points q? We are sympathetic to this view. Indeed, suppose we take these two alternatives, namely:

$$\rho_s(q) \propto \sum_{\hat{q}} \rho_s(\hat{q}) \exp(-|q-\hat{q}|^2) \tag{14}$$

$$\rho_s(E) \propto \oint_\Omega \rho_s(q) \tag{15}$$

where $\Omega = \{q : q^T Es = 0\}$. Briefly, we can then write:

$$\rho_s(E) \propto \oint_\Omega \sum_{\hat{q}} \rho_s(\hat{q}) \exp(-|q - \hat{q}|^2) \tag{16}$$

$$\rho_s(E) \propto \sum_{\hat{q}} \rho_s(\hat{q}) \oint_\Omega \exp(-|q - \hat{q}|^2) \tag{17}$$

Now, denote the point on Es closest to $\hat{q}$ by $q_{\min}$. We can see that the lines joining $\hat{q}$ to $q_{\min}$ and $q_{\min}$ to q must be perpendicular. This implies $|q-\hat{q}|^2 = |q-q_{\min}|^2 + |q_{\min} - \hat{q}|^2$, and therefore:

$$\rho_s(E) \propto \sum_{\hat{q}} \rho_s(\hat{q}) \oint_\Omega \exp(-|q-q_{\min}|^2) \exp(-|q_{\min}-\hat{q}|^2) \tag{18}$$

$$\rho_s(E) \propto \sum_{\hat{q}} \rho_s(\hat{q}) \exp(-(\hat{q}^T l_{(E,s)})^2) \oint_\Omega \exp(-|q-q_{\min}|^2) \tag{19}$$

However, since the integral on the right will be the same for any E, we can write:

$$\rho_s(E) \propto \sum_{\hat{q}} \rho_s(\hat{q}) \exp(-(\hat{q}^T l_{(E,s)})^2) \tag{20}$$

So, simply by changing the $\max$ in Eq. 7 to a sum, these alternatives can be tested. We have run experiments doing this- in practice, the results are very similar.

7. Conclusions

A cornucopia of applications in Computer Vision ranging from stabilization, registration, view generation, 3D motion estimation and the like have been addressed in an explicit first step where optical flow or correspondence is determined. This paper shows that this first step may not always be necessary. Instead, if one computes a probability distribution for the correspondence at each pixel, this is enough to obtain highly accurate results for, for example, egomotion, not obtained by existing state of the art algorithms. The promise of our robust solution suggests that this may represent a useful framework for low and intermediate level vision. Instead of attempting to estimate correspondence at all costs to begin with, we accept that we have uncertainty. Instead of computing flow or correspondence, we compute a probability distribution of it, thus carrying the uncertainty along and following the principle of "least commitment" [12]. As more information becomes available (e.g. 3D motion) the distributions become less uncertain and could be used for obtaining further distributions of depth, shape etc.

Future work should address both if efficient methods can be created to exploit correspondence distributions, and if more principled algorithms exist to explore probabilities in the space of motions.

References

[1] Y. F. A. Wexler and A. Zisserman. Learning epipolar geometry from image sequences. In *CVPR*, volume 2, pages 209–216, 2003.

[2] W. F. Clocksin. A new method for computing optical flow. In *BMVC*, 2000.

[3] C. T. F. Dellaert, S. Seitz and S. Thrun. Structure from motion without correspondence. In *CVPR*, 2000.

[4] O. Faugeras, Q.-T. Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001.

[5] C. Fermüller and Y. Aloimonos. Observability of 3d motion. *Int. J. Comput. Vision*, 37(1):43–63, 2000.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.

[7] D. Fleet. Disparity from local weighted phase-correlation. In *IEEE International Conference on SMC*, pages 48–46, 1994.

[8] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *Int. J. Comput. Vision*, 5(1):77–104, 1990.

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[11] A. Makadia, C. Geyer, and K. Daniilidis. Radon-based structure from motion without correspondences. In *CVPR*, 2005.

[12] D. Marr. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco, 1982.

[13] O. Nestares, R. Navarro, J. Portilla, and A. Tabernero. Efficient spatial-domain implementation of a multiscale image representation based on gabor functions. *Journal of Electronic Imaging*, 7:166–173, 1998.

[14] J. Oliensis. A critique of structure-from-motion algorithms. *Computer Vision and Image Understanding*, 80(2):172–214, 2000.

[15] J. Oliensis. The error surface for structure from motion. *NEC TR*, 2001.

[16] Y. Rosenberg and M. Werman. Representing local motion as a probability distribution matrix applied to object tracking. In *CVPR*, pages 654–659, 1997.

[17] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *CVPR*, pages 310–315, 1991.

[18] T. Tian, C. Tomasi, and D. Heeger. Comparison of approaches to egomotion computation. In *CVPR*, pages 315–320, 1996.