All Points Considered: A Maximum Likelihood Method for Motion Recovery

Daniel Keren[1], Ilan Shimshoni[2], Liran Goshen[2], and Michael Werman[3]

[1] Department of Computer Science, University of Haifa
Haifa 31905, Israel
dkeren@cs.haifa.ac.il
[2] Faculty of Industrial Engineering, Technion
Technion City 32000, Israel
{ilans,lirang}@ie.technion.ac.il
[3] School of Computer Science and Engineering, The Hebrew University of Jerusalem
Jerusalem 91904, Israel
werman@cs.huji.ac.il

Abstract. This paper addresses the problem of motion parameter recovery. A novel paradigm is offered to this problem, which computes a maximum likelihood (ML) estimate. The main novelty is that *all* domain-range point combinations are considered, as opposed to a single "optimal" combination. While this involves the optimization of non-trivial cost functions, the results are superior to those of the so-called algebraic and geometric methods, especially under the presence of strong noise, or when the measurement points approach a degenerate configuration.

1 Introduction

A key problem in motion analysis is the recovery of motion between two successive frames (the first of which will be referred to as the *domain*, and the second as the *range* – these names are meant to indicate that the sought transformation operates on points in the first image, and transforms them into the second image), given a set of point correspondences. Two facets of this problem are studied in this paper:

1. If the measured points in the domain and range frames are denoted $\{P_i\}_{i=1}^m$ and $\{Q_i\}_{i=1}^m$, find a transformation which "maps P_i close to Q_i"; there are two common methods for defining this notion of "closeness". The first, which is usually referred to as the *algebraic method*, seeks a transformation T, restricted to be of a certain class (Euclidean, linear, affine, projective etc.) which minimizes $\sum_{i=1}^{m} \|T(P_i) - Q_i\|^2$. The *geometric method* searches for a set of points $\{\hat{P}_i\}_{i=1}^m$ and a transformation T, such that

$$\sum_{i=1}^{m} \left[\|\hat{P}_i - P_i\|^2 + \|T(\hat{P}_i) - Q_i\|^2 \right]$$

T. Asano et al. (Eds): Geometry, Morphology, ... 2002, LNCS 2616, pp. 72–85, 2003.

is minimal. One may think of a domain-range combination $\{\hat{P}_i, \hat{Q}_i\}$ such that $T(\hat{P}_i) = \hat{Q}_i$ as "legal"; the goal is then to find a "legal" combination which is closest to $\{P_i, Q_i\}$.

2. The FOE (focus of expansion) problem: given are two images of a set of n 3D points, I_1 and I_2, taken from a translating calibrated camera. Let P_1 and P_2 be the projections of the points in I_1 and I_2. Every pair of corresponding points is collinear with an epipole F. Assuming that the camera is moving forward, the points in the second image will be further away from the epipole than their corresponding points in the first image. The points that have been measured in the images are not collinear with the epipole due to noise, and therefore the lines connecting all pairs of corresponding points will not intersect at a single point. For this problem one can also define the algebraic and geometric distances.
   The algebraic approach to determine the epipole F is to use a linear least squares algorithm which finds the point closest to all the lines passing through the pairs of points. The analogue of the geometric method is as follows. Given a candidate epipole F, for each pair of points $p_{1i} \in P_1$ and $p_{2i} \in P_2$, compute the line through F that is closest to the two points. Now, measure the distance of the points from this line, and add the square of each of these two distances to the error function. As opposed to the algebraic approach for which there is a closed-form solution, the geometric requires non-linear optimization techniques. For a comprehensive survey of these methods as well as other motion recovery problems, see [4, 7, 12, 3, 1, 9].

In this paper a new method is introduced, which is compared to these two methods. We first study the linear transformation model, and then extend it to FOE recovery. The method is based on computing the maximum likelihood (ML) estimate for the unknown parameters [5, 6]. The major difference compared with the algebraic and geometric methods is that the ML approach seeks not only a single "good" combination of domain-range points, but looks for a combination which has a "wide support" in the vicinity of the noisy measurement points. In all experiments, it performed considerably better than both other methods, at the price of optimizing more complicated cost functions.

Our approach resembles the one in [11], in which the compatibility between a model and a data point is obtained by integrating over the entire model. The differences between our method and the Bayesian approaches in [10, 2] are in that a) we integrate out the "real" domain points, and b) we do not integrate over all possible models, but find the mode of the distribution of the model. Since the full probability distribution is computed, we could also integrate over all models, but this is beyond the scope of this paper.

2 Linear Transformation

In this section the ML estimate to the parameters of a linear transformation is derived, and its performance compared to that of the algebraic and geometric

methods. We chose to start with the 1D case as it is technically simpler, yet it captures the basic idea of the ML method, and also lends itself to a simple, intuitive explanation.

2.1 The 1D Case

Let us first analyze the ML estimate in a very simple case: a linear transformation from $\mathcal{R}$ to itself. As will be seen, even this case (transformation with one parameter) involves non-trivial computations and admits no explicit solution.

We seek to estimate the parameter a of a mapping from $\mathcal{R}$ to itself, $g(t) = at$, given two noisy measurement points: Y (range) and X (domain). If additional pairs are given, the probability is the product over all pairs, assuming independence. The algebraic and geometric estimates for the single pair case are both equal to $\frac{Y}{X}$.

The ML estimate proceeds to compute the density function of every a conditioned on the measurements. Denote this density $f(\cdot)$ (hereafter $f()$ will denote probability density). Denote by x, y the "true" values of the domain and range points; hence, if the noise model is n, then $X = x + n$, $Y = y + n$, and $g(x) = y$. Hereafter we will assume a Gaussian noise with variance σ^2, so $f(X|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x-X)^2}{2\sigma^2})$ and $f(Y|y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(y-Y)^2}{2\sigma^2})$.

Since x is not known, it has to be marginalized, or "integrated out". The first step is to write down the expression for the joint probability density of a and x:

$$f(a, x/Y, X) = \left|\frac{\partial(ax, x)}{\partial(a, x)}\right| f(ax, x/Y, X) = |x|\frac{f(Y, X/ax, x)f(ax, x)}{f(Y, X)} \propto$$
$$|x|f(Y, X/ax, x) \propto |x| \exp(-\frac{(x - X)^2 + (ax - Y)^2}{2\sigma^2}) \quad (1)$$

We used the fact that $f(Y, X)$ is constant given the measurement, and assumed a uniform prior on ax, x. While the latter is subject to criticism, it is not pertinent to the theme of this paper; given a better prior on ax, x, it can easily be incorporated into the ML computation. The choice of the "correct" prior lies outside the scope of this paper. Next, the expression in Eq. 1 must be integrated over x, to obtain the residual density for a. A closed-form expression can be obtained for this integral, but it is rather cumbersome and we therefore omit it. We have not been able to obtain a closed-form expression for the ML value of a, and it was found by numerical optimization (which is easy in this case, the search space being one-dimensional). Let us demonstrate how the ML estimate differs from the algebraic and geometric estimates, by studying some special cases. Fig. 1 (left) depicts a plot of $f(a)$ as a function of a, when $X = Y = \sigma = 1$. It is readily seen that the ML estimate is rather different from 1, which is the algebraic and geometric estimate; the optimal ML value for a is about 0.467. For low levels of noise (as demonstrated in Fig. 1, right) the ML estimate converges to 1.

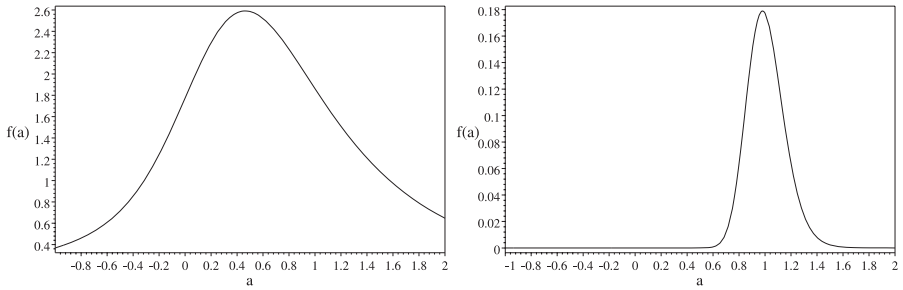

Fig. 1. Probability of the 1D transformation given by $g(x) = ax$, when the domain and range measurements are 1. Left: the noise is Gaussian with unit standard deviation; right: the noise is Gaussian with standard deviation 0.1. In both graphs the horizontal axis stands for a and the vertical axis for $f(a)$

2.2 Interpretation of the Results: Cloud-to-Cloud Match

The ML estimate for a defines a line which does not touch the measurement point (Y, X) – as opposed to the algebraic estimate, which goes through the point. This result, which may appear counter-intuitive, can be explained as follows. The ML estimate seeks to find a slope a which has the largest *support*, that is, such that there is a large "volume" of pairs (ax, x) that are close to the measurement point (Y, X). In other words, it does not seek an optimal points-to-points match (like the algebraic and geometric estimates do), but an optimal *cloud-to-cloud* match: one cloud is the neighborhood of (Y, X), the other is the set $\{(ax, x)/x = -\infty...\infty\}$. That is, the ML estimate a seeks not only that aX be close to Y, but that for a large volume of points x which are close to X, ax will also be close to Y. An explanation is presented in Fig. 2. Note that correspondence still has to be asusmed: the "clouds" here do not refer to the aggregate of points, but to "probability clouds" which surround each range and domain point.

We note that least square analysis is inappropriate for the "cloud-to- cloud" matching problem, as demonstrated in [11].

2.3 Stability

One of the characteristics of the suggested ML estimate is its stability. Consider, for example, the case $\sigma = 1$. Since the geometric and algebraic estimates for a are both equal to $\frac{Y}{X}$, they are very unstable: since X and Y in this case are equal to $x + N(0,1), y + N(0,1)$ ($N(0,1) =$ normal distribution with zero mean and unit variance), the distribution of a is equal to that of the quotient $\frac{y+N_1(0,1)}{x+N_2(0,1)}$. The expectation of this estimate is equal to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y+N_1}{x+N_2} \exp(-\frac{N_1^2+N_2^2}{2}) dN_1 dN_2$$

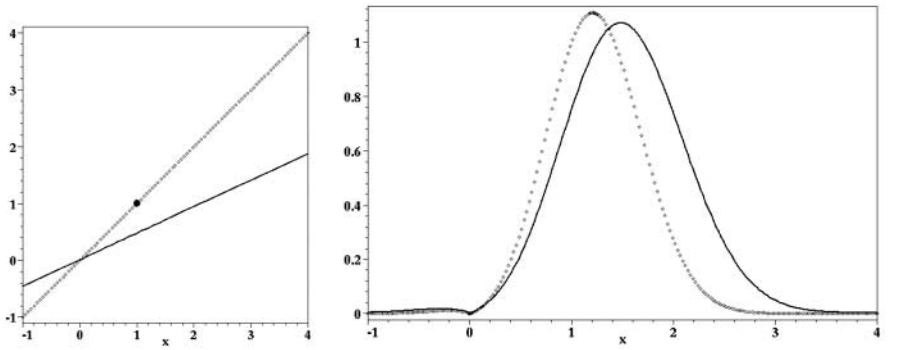

Fig. 2. Left: two lines, corresponding to the geometric choice of slope, $a = 1$ (dotted line), and the ML choice $a = 0.467$ (continuous line). Both are for the same data as in Fig. 1 (left), $X = Y = \sigma = 1$. On the right, the plots for $|x| \exp(-\frac{(x-X)^2+(ax-Y)^2}{2})$ for the two cases. While the solution for $a = 1$ passes through (X, Y) – and therefore the corresponding plot on the right has a higher peak – the *total* probability for $a = 0.467$ is larger. Plots are for the range $x = -1..4$

which is undefined due to the singularity when the denominator is zero (at $x = -N_2$). On the other hand, the ML estimate is always well-behaved in the sense that its expectation is bounded. The proof is presented in the Appendix.

2.4 What Does the Result Mean?

As noted, the transformation a recovered by the suggested ML method may appear counter-intuitive, because the recovered transformation does not map the domain data point onto the range data point. The explanation we offer is that the quality of the result should not be measured by the proximity of the transformed measurements of the domain points to the measurements of the range points, neither by the proximity of a configuration of "legal" domain and range points (see the Introduction), to the measurements. We suggest that the quality of the results should be measured by studying the *entire* probability distribution over all possible sources of the measurement points.

To illustrate this for the one-dimensional case, consider again the case $X = Y = \sigma = 1$. We know that there exist "real" points x, y such that $x + N_1(0, 1) = X, y + N_2(0, 1) = Y$, and that the "real" a satisfies $ax = y$. Thus, the distribution for a is given by $\frac{1+N_1(0,1)}{1+N_2(0,1)}$ (note that $N(0, 1)$ is symmetric around the origin). One can, in fact, compute the distribution of this expression, and prove that it is identical to the distribution derived in Section 2.1.

The same considerations hold for the problems of 2D motion recovery discussed in the following sections.

2.5 The 2D Case

The case of a 2D linear transformation proceeds in a manner very similar to the 1D case. Given noisy measurements of two domain points, P_1 and P_2, and of the corresponding range points, Q_1 and Q_2, the probability of the transformation $T(x, y) = (ax + by, cx + dy)$ can be computed in a manner similar to the 1D case, albeit more complicated, resulting in:

$$f(a, b, c, d/P_1, P_2, Q_1, Q_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1y_2 - x_2y_1)^2 \times \exp\left(\frac{\sum_{i=1}^{2}\left[\|(x_i, y_i) - P_i\|^2 + \|T(x_i, y_i) - Q_i\|^2\right]}{2\sigma^2}\right)dx_1dy_1dx_2dy_2 \tag{2}$$

The explicit form of the integral is too long to include here. However, there is an interesting similarity between this integral and Eq. 1. Note, especially, the weighing factor $(x_1y_2 - x_2y_1)^2$. As in the 1D case, a higher weight is assigned to domain measurement points which are farther from the origin; however, the weight factor also penalizes two domain points which are in similar directions. This is not surprising, since when two points are in the same direction, they are linearly dependent, hence they yield less information on the transformation T.

2.6 Optimization

Given a pair of domain and range measurements, the ML linear transformation is obtained by maximizing Eq. 2. We were not able to find an explicit solution, so the maximum was recovered using a general-purpose optimization method, the Nelder-Mead simplex method with simulated annealing, as presented in [8]. When more measurement points are provided, the expression for the probability factors into a few expressions as Eq. 2, one for each combination of two domain-range pairs. The time it took to recover the motion between two frames was on the average 0.03 seconds on a Digital workstation, for 100 point pairs.

2.7 Some Results

We tested the suggested method for a few cases, and it consistently performed better than both the algebraic and geometric methods. Results for two cases are presented: a nearly degenerate configuration and a stable one.

Nearly Degenerate Configuration In this experiment, four domain and four range points were chosen according to the following simple rule: for $k = 1..4$, the k-th range point as well as the k-th domain point is equal to $(k + n, k + n)$,

where n is zero mean Gaussian with 0.01 standard deviation (with different instances of n for every coordinate). We study this case in order to illustrate the stability of the method.

For example, in one of the tested cases, the noisy measurements were:

$$T(0.999, 1.009) = (1.023, 1.015), \quad T(2.021, 1.989) = (2.015, 1.985)$$
$$T(2.983, 3.009) = (3.012, 3.002), \quad T(4.007, 4.006) = (3.996, 4.002)$$

The geometric method results for this type of data were very unstable. Denoting as before $T(x, y) = (ax + by, cx + dy)$, we have that $a(k + n) + b(k + n) = k + n$ for very small n, so $a + b \approx 1$, and similarly $c + d \approx 1$. In order to display the results, it therefore suffices to show the values of a and c. For ten such tests, the results are shown in the scatter diagram below (Fig. 3). The results are very unstable, although the point configurations are nearly identical. The same phenomena occurs when the algebraic method is applied to these point sets.

However, the ML estimate studied here gave very stable results; for all cases, the values of the transformation coefficients were between 0.49 and 0.51, that is, the recovered transformation matrix was very close to

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

The explanation for this result is that the transformation tries to map points which are close to (k, k) (for $k = 1..4$), as close as possible to (k, k) (see previous discussion about the "cloud to cloud" matching in Section 2.2). Of all linear combinations of noisy measurements of (k, k) whose expectation is k, the one with weights $(\frac{1}{2}, \frac{1}{2})$ has the smallest variance, hence it achieves the best concentration around the range point (k, k). In this case and others, the ML estimate has the property of choosing a "simple" and stable solution when the motion recovery becomes ill-posed, and this is achieved without imposing an external simplicity constraint, as in regularization or minimal description length based methods.

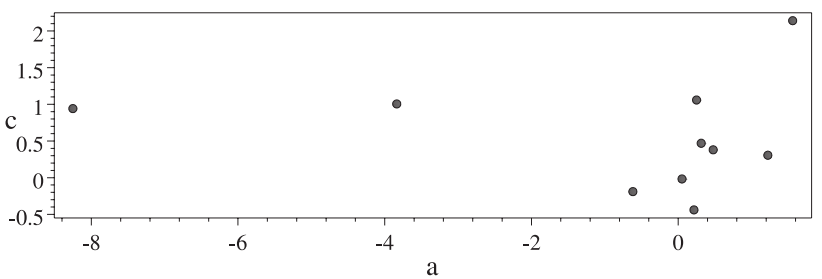

Fig. 3. Scatter diagram for the coefficients (a, c) of the linear transformation recovered by the geometric method in the nearly degenerate case

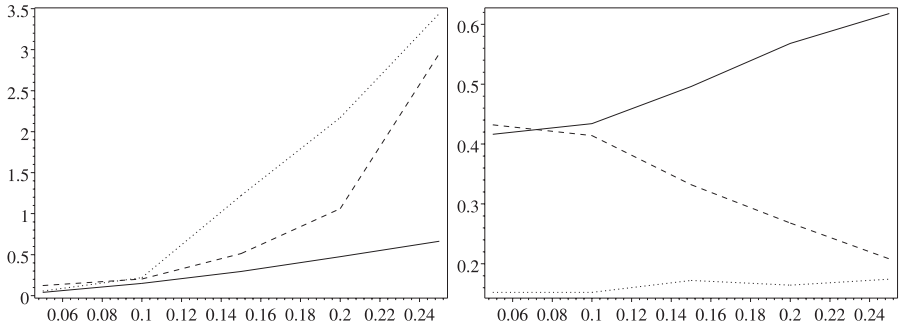

Fig. 4. Performance of the three methods for motion recovery. Left: horizontal axis stands for noise standard deviation, vertical for average l^2 error in the recovery of the motion parameters. The ML estimate is depicted as a solid line, dashed line stands for the algebraic method, and dotted line for the geometric method. Right: horizontal axis stands for noise standard deviation, vertical for the ratio of experiments in which the respective method performed best

Non-degenerate Configuration In this set of experiments, the accuracy with which the three methods recover the transformation

$$\begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}$$

was studied. The domain points formed a stable configuration: $\{(1,1),(2,1),(1,2),(2,2)\}$, and various levels of noise were tested. An algorithm's error was defined as the l^2 difference between the original transformation and the recovered one. We have also charted the percentage of cases in which each method performed best. The tests were run 500 times for each noise variance. Results are presented in Fig. 4.

The suggested ML estimate was clearly superior to the algebraic and geometric methods. Its average error was much smaller, and it also performed best in most cases except for the smallest noise variance, in which it was very slightly (1.6%) surpassed by the geometric method (however, in this case its average error was 29% smaller than the geometric method's error). On a side note, our experiments indicated that, although the geometric method performs better than the algebraic method over a wide range of noise variance, its average error is larger when the noise increases.

3 Focus of Expansion

An important problem in computer vision is the recovery of the *focus of expansion* (FOE). In the simplest instance of the problem, four points are given in the plane: $P_1 = (P_1^x, P_1^y), P_2 = (P_2^x, P_2^y), Q_1 = (Q_1^x, Q_1^y), Q_2 = (Q_2^x, Q_2^y)$, which are

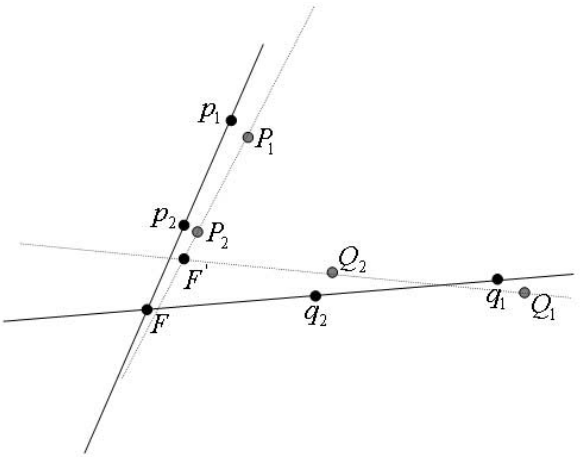

Fig. 5. Layout of the FOE problem. The point to be recovered is F, however only $\{P_1, P_2, Q_1, Q_2\}$, which are the noised versions of $\{p_1, p_2, q_1, q_2\}$, are given. For comparison, the FOE corresponding to $\{P_1, P_2, Q_1, Q_2\}$ is the point F'

the noisy versions of $p_1 = (p_1^x, p_1^y), p_2 = (p_2^x, p_2^y), q_1 = (q_1^x, q_1^y), q_2 = (q_2^x, q_2^y)$. The point $F = (F^x, F^y)$ is sought that such $\{F, p_1, p_2\}$ and $\{F, q_1, q_2\}$ are collinear. It is common to assume that p_2 respectively q_2 is between F and p_1 respectively q_1 (Fig. 5). The problem is to find a good estimate for F, given $\{P_1, P_2, Q_1, Q_2\}$ (if more than four points are given then the ML estimate maximizes the product of the densities for the subsets of size four).

In order to extend the ML paradigm to the FOE recovery problem, we integrate over all the possible choices of $\{p_1, p_2, p_1, p_2\}$. Specifically, given a candidate F, we parameterize p_2, q_2 by F, p_1, q_1: $p_2 = \lambda_1 p_1 + (1 - \lambda_1)F, q_2 = \lambda_2 p_2 + (1 - \lambda_2)F$ (where $0 \leq \lambda_1, \lambda_2 \leq 1$). Then, to compute the density of F given $\{P_1, P_2, Q_1, Q_2\}$, we proceed as before, using Bayes' law and an appropriate change of variables. Alas, the computation is more cumbersome. For simplicity we assume that the noise variance satisfies $2\sigma^2 = 1$, but obviously this does not sacrifice any generality.

$$
\begin{aligned}
&f(F, p_1, q_1, \lambda_1, \lambda_2 / P_1, P_2, Q_1, Q_2) = \\
&f(F^x, F^y, p_1^x, p_1^y, q_1^x, q_1^y, \lambda_1, \lambda_2 / P_1^x, P_1^y, P_2^x, P_2^y, Q_1^x, Q_1^y, Q_2^x, Q_2^y) = \\
&f(p_1^x, p_1^y, q_1^x, q_1^y, \lambda_1 p_1^x + (1-\lambda_1)F^x, \lambda_1 p_1^y + (1-\lambda_1)F^y, \lambda_2 q_1^x + (1-\lambda_2)F^x, \lambda_2 q_1^y + (1-\lambda_1)F^y / \\
&P_1^x, P_1^y, P_2^x, P_2^y, Q_1^x, Q_1^y, Q_2^x, Q_2^y) \cdot \\
&\left| \frac{\partial(p_1^x, p_1^y, q_1^x, q_1^y, \lambda_1 p_1^x + (1-\lambda_1)F^x, \lambda_1 p_1^y + (1-\lambda_1)F^y, \lambda_2 q_1^x + (1-\lambda_2)F^x, \lambda_2 p_1^y + (1-\lambda_1)F^y)}{\partial(F^x, F^y, p_1^x, p_1^y, q_1^x, q_1^y, \lambda_1, \lambda_2)} \right| = \\
&(1-\lambda_1)(1-\lambda_2) \left| p_1^y F^x - F^x q_1^y - p_1^y q_1^x + p_1^x q_1^y - p_1^x F^y + F^y q_1^x \right| \cdot \\
&\exp(-(p_1^x - P_1^x)^2 - (p_1^y - P_1^y)^2 - (q_1^x - Q_1^x)^2 - (q_1^y - Q_1^y)^2 - \\
&(\lambda_1 p_1^x + (1-\lambda_1) fx - P_2^x)^2 - (\lambda_1 p_1^y + (1-\lambda_1) fy - P_2^y)^2 - \\
&(\lambda_2 q_1^x + (1-\lambda_2) fx - Q_2^x)^2 - (\lambda_2 q_1^y + (1-\lambda_2) fy - Q_2^y)^2)
\end{aligned}
\tag{3}
$$

Next, the "nuisance parameters" $p_1^x, p_1^y, q_1^x, q_1^y, \lambda_1, \lambda_2$ have to be integrated out. The resulting six-dimensional integral is difficult to explicitly compute because of the absolute value expression (Eq. 3). We have therefore applied numerical integration, adopted from the *qgauss* method [8].

3.1 Results

As before, the results may appear counter-intuitive. For example, consider the case
$P_1 = (10, 20), P_2 = (7.5, 15), Q_1 = (24, 12), Q_2 = (20, 10)$, with noise satisfying $2\sigma^2 = 1$. The ML result for the FOE is $(3.24, 4.03)$, quite different from $(0, 0)$ which is the result of the geometric and algebraic methods. Again, the intuitive explanation is that the entire probability space for p_1, p_2, q_1, q_2 is sampled by adding noise to P_1, P_2, Q_1, Q_2, and each time the FOE is computed by intersecting the lines between p_1, p_2 and q_1, q_2. This results in a cloud of points whose largest density is around the ML estimate. To make this empirical explanation

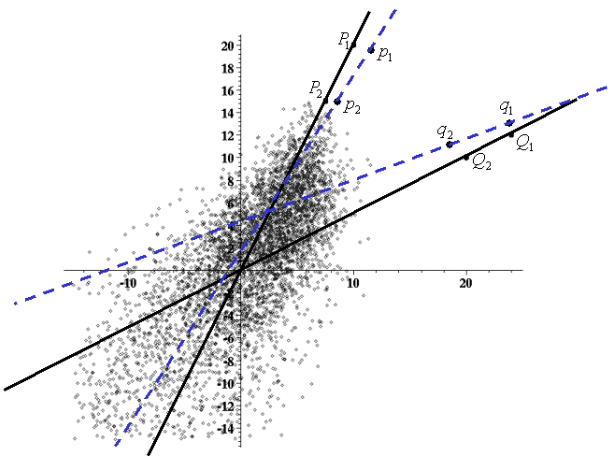

Fig. 6. Empirical demonstration of the ML FOE estimate. The measured points are $P_1 = (10, 20), P_2 = (7.5, 15), Q_1 = (24, 12), Q_2 = (20, 10)$. The FOE corresponding to the measurements is $(0, 0)$ (represented by the intersection of the two dark lines). The ML estimate, on the other hand, is the FOE with the highest probability density; this density can be sampled by randomly adding noise to $\{P_1, P_2, Q_1, Q_2\}$, thus obtaining $\{p_1, p_2, q_1, q_2\}$, and computing the FOE for $\{p_1, p_2, q_1, q_2\}$. A scatter diagram for 4,000 such FOE's is shown, together with the measurement points and a sample $\{p_1, p_2, q_1, q_2\}$, with its corresponding FOE (represented by the intersection of the blue dashed lines). Only points in the square $[-15, 15] \times [-15, 15]$ are shown

of the results clear, a sample of this point cloud is presented (Figs. 6,7). It can be observed that it clusters not around the origin, but around the ML estimate.

4 Conclusion and Future Research

We have presented a Bayes based ML estimate for motion recovery. The basic idea is to consider not only the noisy measurement points, but to integrate over all possible combinations of "real" domain and range points. The method yields results which are correct in the sense of obtaining a motion estimate which has the highest support among all these combinations. Due to this global property, the method is very stable even when the point configurations are nearly degenerate, or when the noise is large relative to the size of the point set. The price is the computational cost of evaluating the probability of the candidate motion parameters, which involves integrating over all possible domain configurations.

Future work will address the following questions: is a ML estimate good enough, or should the "fully Bayesian" estimate be computed by averaging the

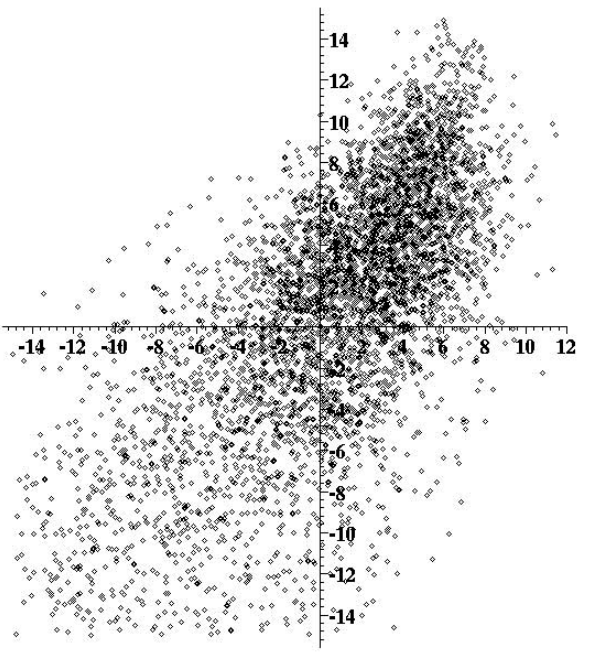

Fig. 7. Close-up of scatter diagram of Fig. 6: denser (darker) areas correspond to higher values of the FOE probability density function. It can be observed that the highest density is not around the origin. The ML estimate is $(3.24, 4.03)$. Note that the density is nearly flat in a large area, indicating that it is not possible to produce a single highly reliable estimate for the FOE

motion models weighed by their probabilities? How reliable are the ML estimates (equivalently, is the probability distribution uni-modal and strongly peaked)? Also, the method will hopefully be extended to cover more difficult problems such as the recovery of the fundamental matrix. In addition, applications of the theoretical results to real problems will be studied.

References

[1] A. Adam, E. Rivlin, and I. Shimshoni. Computing the sensory uncertainty field of a vision-based localization sensor. *IEEE Trans. on Robotics and Automation*, 17(3):258–267, June 2001. 73

[2] D. A. Forsyth, S. Ioffe, and J. Haddon. Bayesian structure from motion. In *ICCV99*, pages 660–665, 1999. 73

[3] R. Hartley and A. Zisserman. *Multiple Views Geoemtry in Computer Vision.* Cambridge University Press, 2000. 73

[4] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(6):580–593, June 1997. 73

[5] K. Kanatani. Geometric computation for machine vision. In *Oxford University Press*, 1993. 73

[6] K. Kanatani. Statistical-analysis of geometric computation. *CVGIP*, 59(3):286–306, May 1994. 73

[7] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981. 73

[8] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988. 77, 81

[9] G. Speyer and M. Werman. Parameter estimates for a pencil of lines: Bounds and estimators. In *ECCV*, 2002. 73

[10] P. H. S. Torr and A. Zisserman. Concerning Bayesian motion segmentation, model averaging, matching and the trifocal tensor. In H. Burkharddt and B. Neumann, editors, *ECCV98 Vol 1*, pages 511–528. Springer, 1998. 73

[11] M. Werman and D. Keren. A Bayesian method for fitting parametric and nonparametric models to noisy data. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23(5):528–534, May 2001. 73, 75

[12] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of unknown epipolar geometry. *Artificial Inteligence*, 78(1-2):88–119, 1995. 73

Appendix

In this appendix we will prove that the expectation of the maximum likelihood estimator for the recovery of the one-dimensional scale parameter, discussed in Section 2.1, is well-defined and finite. Although this is a simple case, it serves to demonstrate the stability and convergence properties of the suggested method.

Recall (Eq. 1) that if the mapping is given by $T(x) = ax$, and the domain respectively range measurements are X and Y, then the probability density of a is given by

$$f(a) = \int_{-\infty}^{\infty} |x| \exp(-[(x - X)^2 + (ax - Y)^2])dx \tag{4}$$

(where we have assumed for simplicity $2\sigma^2 = 1$). The proof will proceed by demonstrating that the integral of Eq. 4 is smaller than $f(0)$, for a large enough a. Hence, the maximum likelihood estimate (which is obtained at the global maximum of $f(a)$) will be bounded, and then its expectation can be estimated by integrating over all possible Y, X.

The integral of Eq. 4 equals

$$\exp(-X^2 - Y^2) \int_{-\infty}^{\infty} |x| \exp(-[(a^2+1)x^2 - 2(X+aY)x])dx$$

if $A > 0$,

$$\int_{-\infty}^{\infty} |x| \exp(-[Ax^2 - Bx])dx = \frac{2A^{3/2} + \sqrt{\pi}AB \exp(\frac{B^2}{4A}) erf(\frac{B}{2\sqrt{A}})}{2A^{5/2}}$$

then, noting that $|erf(x)| < 1$, and substituting $A = a^2 + 1, B = 2(X + aY)$, $f(a)$ can be bounded by

$$\exp(-X^2 - Y^2) \left[\frac{1}{a^2} + \frac{X+aY}{a^3} \exp(\frac{(X+aY)^2}{a^2}) \right] \tag{5}$$

here we gave assumed that $a, Y, X \geq 0$, but this only means that in the remaining part of the proof we should replace these three variables by their absolute values; to make notations simpler, we will omit the absolute values.

Note that

$$\begin{aligned} f(0) &= \int_{-\infty}^{\infty} |x| \exp(-[(x-X)^2 + Y^2])dx \\ &= \exp(-X^2 - Y^2)[1 + X \exp(X^2) erf(X)] \geq \exp(-X^2 - Y^2). \end{aligned}$$

So, in order to bound $f(a)$ by $f(0)$, we can cancel out $\exp(-X^2 - Y^2)$. After discarding $\frac{1}{a^2}$ in Eq. 5, as it is asymptotically negligible compared to the other expressions, it is then sufficient to bound $\frac{X+aY}{a^3} \exp(\frac{(X+aY)^2}{a^2})$ by 1.

Lemma 1. *There exists a polynomially bounded expression $p(Y, X)$ such that if $a > p(Y, X)$ then*

$$\frac{X+aY}{a^3} \exp(\frac{(X+aY)^2}{a^2}) < \frac{1}{a^{3/2}} \exp(Y^2 + 1)$$

Proof: It is straightforward to verify that this inequality holds for a polynomial larger than $\max\{2, 2X, 2Y, 2X^2, 4XY, 2Y^2\}$. Denote $p(Y, X) = \max\{2, 2X, 2Y, 2X^2, 4XY, 2Y^2\}$.

Lemma 2. *If $a > \max\{\exp(\frac{4}{5}(Y^2 + 1)), p(Y, X)\}$, then $f(a) < f(0)$.*

Proof:

$$f(a) < f(0)\left[\frac{X+aY}{a^3}\exp(\frac{(X+aY)^2}{a^2})\right] < \frac{f(0)}{a^{3/2}}\exp(Y^2+1) <$$
$$\frac{f(0)}{\exp(\frac{6}{5}(Y^2+1))}\exp(Y^2+1) = \frac{f(0)}{\exp(\frac{1}{5}(Y^2+1)} < f(0).$$

Since the ML estimate is obtained at the global maximum of $f(a)$, it is bounded by $\max\{\exp(\frac{4}{5}(Y^2+1)), p(Y,X)\}$, and its expectation can be bounded by integrating this bound over all Y, X pairs:

Theorem 1. *The expectation of the maximum likelihood estimate for a is finite.*

Proof: Suppose that the "real" domain and range points are x, y respectively. Then, the expectation of a is bounded above by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-(X-x)^2-(Y-y)^2)\max\{p(Y,X),\exp(\frac{4}{5}(Y^2+1))\}dXdY$$

and this integral is convergent, because the left exponent is dominated by $\exp(-X^2-Y^2)$ as $Y, X \to \infty$. Hence asymptotically the integrand is bounded by $\max\{\exp(-X^2-Y^2)p(Y,X), \exp(-X^2-\frac{1}{5}Y^2)\}$, and the integral of this expression clearly converges.