Models of Statistical Visual Motion Estimation

Minas E. Spetsakis

Dept. of Computer Science
York University
4700 Keele Street
North York, ONTARIO
CANADA, M3J 1P3

Abstract

Several models of Statistical Estimation of motion from visual input are derived and analyzed. We study a wide variety of models, ones that use least squares and ones that use Maximum Likelihood, with several different assumptions (dependent and independent noise, isotropic and non-isotropic noise), spherical and planar image surfaces and with different preprocessing (one based on correspondence and one based on disparity). We do all this analysis using only a few fundamental concepts from Statistical Estimation so the relative merits and shortcomings of all the methods become evident.

1. Introduction

In the recent years several papers appeared that deal with the issue of Visual Motion Analysis from a Statistical Estimation point of view. The increasing popularity of Statistical Estimation is caused by the problems posed to Motion Analysis research by noise. The situation if something like the following. There exist algorithms to solve a very wide variety of problems, but most of these algorithms do not work under even minimal noise. So researchers turned their attention to Statistics where a wealth of knowledge and experience has been accumulated. Several researchers have reported promising results [8, 9, 2]. But all these results come fragmented in different papers, each one derived in a different fashion and notation, using different assumptions etc. There has been some work on organizing and comparing them [7] but no simple way to derive and analyze them all with the same tools. In this paper we present a way to derive, analyze and organize most of the estimation methods, using very few fundamental ideas from Statistical Estimation. All the methods thus are easy to compare based on the assumptions they use, the difficulty of their computation and their mathematical elegance. The intuitive and heuristic component was kept at a minimum.

Not all of the methods analyzed here have appeared in the literature before. But since all of them seem to belong to a natural sequence they were included for completeness. In section 2 of the paper we derive and analyze several models, ones that use least squares and ones that use Maximum Likelihood, with several different assumptions (dependent and independent noise, isotropic and non-isotropic noise), spherical and planar image surfaces and different preprocessing models (one based on correspondence and one based on disparity).

2. Analysis of a few models

The objective in any estimation is to minimize the error in some sense. For a given application, after studying the particulars of the problem, the nature of the noise and the needs of the application, one can set up an estimation procedure. This procedure might be seen as minimizing the least square residual error or maximizing the probability that the computed solution is indeed the actual solution to the problem or computing a solution such that the error is orthogonal to the solution space, etc. All of the above lead to a different estimator, with different properties in general. With a notable exception. If the problem is linear and the noise has a Gaussian distribution all the above estimators are equivalent. In such a case the estimator has most of the desirable properties: it is consistent and unbiased and has minimum variance.

Unfortunately in structure from motion most of the problems are non-linear so we cannot expect many of the above nice properties. In general non-linear estimation problems are more difficult to

The support of the NSERC (App. No. OGP0046645) is gratefully acknowledged.

study and structure from motion is no exception.

In the next subsections we present and analyze a few estimators derived by methods like least squares minimization and MLM (Maximum Likelihood Method). While MLM is clearly the most interesting one, the others are included because they are instructive since they are based mostly on intuition. And it is interesting to see how much they differ.

2.1. Least Squares Models

Least Squares models can come in a large variety and several intuitive modifications can be added to them. The idea here is pretty simple (and old), but very powerful: Find an expression for the residual error for all the data, square it and sum it up for all data. Then try to find the values of the unknowns that minimize the sum of the square residuals. The values of these unknowns are the estimates we wanted.

2.1.1. Epipolar Constraint

The simplest thing we can do is use as residual error the discrepancy of the Epipolar Constraint. It is obvious that in the presence of noise there is no set of motion parameters R and T (or equivalently matrix E) that can satisfy the epipolar constraint for all points. We'll try to minimize the sum of the squares of the discrepancy of the epipolar constraint. The constraint has the form

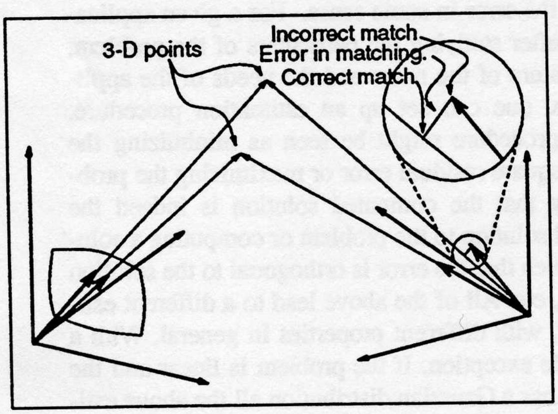

Figure 2.1.1: If we extend the lines from the two views of the same world point they should intersect. Since there is an error in the computation of the disparity (or flow or correspondence), they do not intersect. The distance of the actual and the computed match is the error (which in general is different from the residual).

$$p'^T_i E p_i = 0$$

So we have to minimize

$$\sum_i \left\{ p'^T_i E p_i \right\}^2 \qquad 2.1.1.1$$

Several methods have been suggested on how to minimize this expression [8, 6, 5]. But as pointed out in [8, 7] the result that comes out of it suffers from a very noticeable bias. To see why we rewrite the epipolar equation.

$$p'^T_i E p_i = p'^T_i \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t2 & t_1 & 0 \end{bmatrix} R p_i =$$

$$p'^T_i (T \times (R p_i)) \qquad 2.1.1.2$$

It is easily seen from 2.1.1.2 that, if the p_i's form a tight bundle then a value of T that is close to this bundle will make the cross product in 2.1.1.2 small. This, under moderate noise, will make a value of T close to the bundle preferable to a value of T close to the actual solution. In practice, this means that T will have the tendency to point towards the center of the scene with even small noise.

This high bias appears because there is this cross product in 2.1.1.2 which is proportional to the sine of the angle of the two vectors, T and $R p_i$. A simple way to deal with this bias is to divide each term of the expression by an amount equal to this sine, which is not a bad idea but appeals only to intuition.

2.1.2. Assume Isotropic Image Noise

Now we can be more careful about how we define the residual error. First of all let's assume the noise is isotropic (the same distribution in all directions). We relax this assumption completely in subsequent sections. Another assumption is that the input is the disparity field. Which means that for the points in the one frame we are given disparity vectors from which we can compute the corresponding points in the other frame. The error is in the disparity. An alternative is to assume that the error is in locating the pairs feature points. Since this second alternative leads to awkward formulae we shall use it in only two models where we can keep the math tractable. Which one of the two is best is clearly matter of taste and the application at hand.

The way we attack the problem is the following. Since we assume rigidity the epipolar constraint holds. Then we try to find a correction for the given disparities (or flow vectors or correspondences) to add to them to make them satisfy the epipolar constraint. And we try to minimize this correction. The intuition behind this is simple. If the error distribution is unimodal (has one maximum) and the mode is the same as the mean then small errors are more probable than large ones. But small errors require small corrections. So choose the motion parameters that require small corrections. For a rigorous analysis of this concept see [3].

If n_i is the noise vector of the disparity vector Δp_i then $\Delta p_i - n_i = \Delta p_{ci}$ is the "correct" disparity. Then

$$p'_i = p_i + \Delta p_i = p_1 + \Delta p_{ci} + n_i = p'_{ci} + n_i$$

or

$$p'_{ci} = p'_i - n_i$$

where p'_{ci} is the "correct" match of p_i. It is obvious that the epipolar constraint holds for the correct p'_{ci}.

$$p'^T_{ci} E p_i = (p'_i - n_i)^T E p_i = 0 \qquad 2.1.2.1$$

Eq. 2.1.2.1 is a constraint on the noise. We have then to minimize the sum of the square of the noise under this constraint.

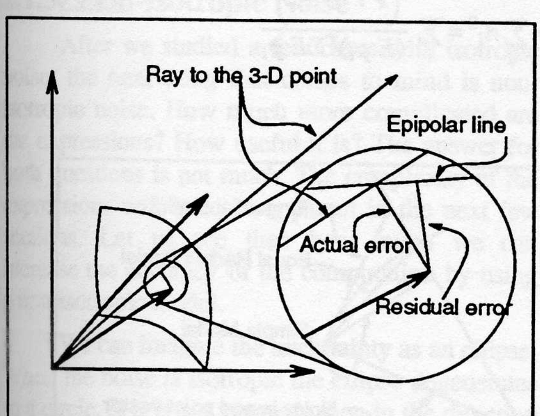

Figure 2.1.2: The epipolar constraint says that the correct match should be on the epipolar line, which is a function of the motion parameters. The residual error can end anywhere on the epipolar.

2.1.2.1. Simple Model with Spherical Image.

A simple model that gets rid of the high bias is:

$$\sum_i n_i^T n_i \to \min$$
$$\text{s.t } (p'_i - n_i)^T E p_i = 0 \qquad 2.1.2.2$$

which is easy to solve using Lagrangian multipliers [4]. Eq. 2.1.2.2 means we minimize the mean square of the correction n_i for what we think is the noise. The idea is simple. Form the expression

$$\mathbf{L} = n_i^T n_i + \lambda (p'_i - n_i)^T E p_i$$

and then get the equations by setting the derivatives of $\mathbf{L} = \sum_i \mathbf{L}_i$ equal to zero: $\frac{\partial \mathbf{L}}{\partial n_{ij}} = 0$, for $j = x, y, z$ and $\frac{\partial \mathbf{L}}{\partial \lambda_i} = 0$. This equation has $4 \times i + 5$ unknowns, 3 for the correction to the noise One for λ for each point correspondence and 5 for the motion parameters. Luckily the 4xi unknowns are local and can be eliminated. Then we have to solve for the 5 motion parameters only. So we eliminate all but the 5 unknowns by considering each $\mathbf{L}_i$ separately. Solving the equations we get

$$\lambda_i = 2\,\frac{p'^T_i E p_i}{(E p_i)^2}, \quad n_i = \frac{1}{2} \lambda_i E p_i \qquad 2.1.2.3$$

The noise vector n we computed above is a function of the motion parameters and satisfies the epipolar constraint written in the form of eq. 2.1.2.1. Then all we have to do is find the motion parameters that minimize $\sum_i \mathbf{L}_i$ which can be written as

$$\sum \mathbf{L}_i = \sum_i n_i^2 = \sum_i \frac{\left(p'^T_i E p_i\right)^2}{(E p_i)^2} \qquad 2.1.2.4$$

This expression was derived in [8]. The experimental results were very good compared to the biased approaches in the previous section. An iterative method was presented to find the solution that used an initial guess a biased solution. With this as a guess the convergence was quick.

It is easy to see that 2.1.2.4 can be transformed to

$$n_i^2 = p'^2_i \cos^2(angle(p'_i, E\, p_i))$$

where $angle(\cdot,\cdot)$ returns the angle between its two vector arguments.

2.1.2.2. Tangential Model with Spherical Image.

The above model is good up to error of 5-10 degrees (well... this is more than what we want anyways since on a wide angle camera this is 20-40 pixels). Another model, that would be preferable in some cases, is one that restricts the noise to be tangent to the image, which means normal to the p'_i. To do this we add one more constraint and we modify the Lagrangian $\mathbf{L}_i$ accordingly by using one more Lagrangian multiplier. Thus we have

$$\mathbf{L}_i = n_i^2 + \lambda_{1i}(p'_i - n_i)^T E\, p_i + \lambda_{2i} n_i^T p'_i$$

In the same way as before we get 5 equations that give us solutions

$$n_i = p'_i \times \left((E\, p_i) \times p'_i\right) \frac{p'^T_i E\, p_i}{\left((E\, p_i) \times p'_i\right)^2} \qquad 2.1.2.5$$

then as before the quantity we minimize is

$$\sum_i \mathbf{L}_i = \sum_i \frac{\left(p'^T_i E\, p_i\right)^2 p'^2_i}{\left((E\, p_i) \times p'_i\right)^2} \qquad 2.1.2.6$$

It is not difficult to see that 2.1.2.6 is equivalent to:

$$n_i^2 = p'^2_i \cot^2(angle(p'_i, E\, p_i))$$

2.1.2.3. Equal Radius Model

We can get yet another model by requiring the noise vector to be such that both p_{ci}' and p_i' are of the same length. So we replace the second constraint accordingly.

$$\mathbf{L}_i = n_i^2 + \lambda_{1i}(p'_i - n_i)^T E\, p_i + \lambda_{2i}(p_i'^2 - p_{ci}'^2)$$

And the solution is:

$$n_i^2 = 2\, p'^2_i\, (1 \pm \sin(angle(p'_i, E\, p_i))) \qquad 2.1.2.7$$

Of course from these two solutions we choose the one with the − when the sine is positive and + for the opposite.

2.1.2.4. Comparison of the Three Spherical Image Models

The three models that are described above do not differ much when the noise is low. If the noise is so high that these models differ significantly then most probably it is too high to do any motion and structure estimation. At least without using an unrealisticly high number of points.

2.1.2.5. Planar Image Model

We'll study two models. One with the same assumptions, namely noise in the disparity vector, isotropic noise and rigidity. And one that the noise is in the position of the correspondence pair. Again the story is the same. The extra constraint we have on the noise is that it rests on the image plane so that its $\hat{z}$ component is zero. We have

$$\mathbf{L}_i = n_i^2 + \lambda_{1i}(p'_i - n_i)^T E\, p_i + \lambda_{2i}\, n_i^T \hat{z}$$

Forming the equations in the usual manner we get

$$\lambda_{1i} = 2\frac{p'^T_i E\, p_i}{(E\, p_i)^T \Sigma E\, p_i}$$
$$\lambda_{2i} = \lambda_{1i}(E\, p_i)^T \hat{z} \qquad 2.1.2.8$$
$$n_i = \frac{\lambda_{1i}}{2}\Sigma E\, p_i$$

where $\Sigma = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix}$. And we minimize

$$\sum_i \mathbf{L}_i = \sum_i n_i^2 = \sum_i \frac{\left(p'^T_i E\, p_i\right)^2}{(E\, p_i)^T \Sigma E\, p_i} \qquad 2.1.2.9$$

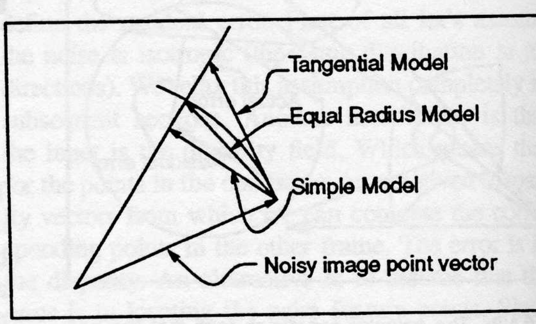

Figure 2.1.3: The three models with spherical image with a grossly exaggerated noise.

2.1.2.6. Planar Image Model with Noisy Correspondence Pair

We can now examine another model where the noise is assumed to be on both image points of a correspondence pair. The two points have the same noise distribution (this is not a strong assumption since the correspondence of the two points is established quite often using a measure of their similarity). We now have three terms to minimize on the Lagrangian

$$L_i = n_{1i}^2 + n_{2i}^2 + \lambda_{1i}(p'_i - n_{1i})^T E (p_i - n_{2i}) + \lambda_{2i} n_{1i}^T \hat{z} + \lambda_{3i} n_{2i}^T \hat{z}$$

Now we have three Lagrangian multipliers, but the problem is that we have some unpleasant non linearities associated with λ_{1i}. While the equations we get are still solvable using the program Maple, they are too complicated to do anything useful with them. So we get rid of all the second order terms of the noise and then take the derivatives. Then we have to minimize

$$\sum_i \frac{\left(p'^T_i E p_i\right)^2}{p'^T_i E \Sigma E^T p'_i + (E p_i)^T \Sigma E p_i} \qquad 2.1.2.10$$

where Σ is as before. Eq. 2.1.2.10 is the same as in [9, 6] and it makes more sense when the correspondence is established with discrete point matching. It should be noted though that this is an approximate expression.

2.1.3. Non-Isotropic Noise

After we studied a few cases with isotropic noise, the next thing that comes to mind is non-isotropic noise. How much more complicated are the expressions? How useful it is? The answer for both questions is not much. The complexity of the expressions will become apparent in the next few sections. Let us see then how much we can increase the accuracy of the computation by using a non isotropic model.

We can imagine the uncertainty as an ellipse. When the noise is isotropic the ellipse degenerates to a circle. If it is not isotropic then in the direction of the major axis of the ellipse the uncertainty is higher. The ellipse roughly corresponds to the area where the endpoint of noise is most likely to occur. It can also be seen as the ellipse defined from the inverse of the covariance matrix of the error. As shown in Fig. 2.1.4 when the uncertainty cloud is a

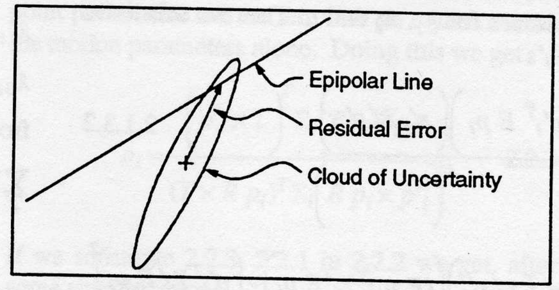

Figure 2.1.4: The cloud of uncertainty represents the area where the penalty for the residual error remains low. An elongated cloud of uncertainty will almost always provide us with a residual error, that terminates on the epipolar line and is within this cloud of uncertainty.

long and narrow ellipse a quite wide range of epipolar lines intersect it. So it does not constraint the position of the epipolar, and hence the values of the motion parameters, very much. And of course when due to the aperture problem the ellipse becomes really elongated there is no constraint at all.

This is not to say that the unisotropic model does not offer anything. It might not offer much to the computation of motion but it is very important to the computation of structure.

2.1.3.1. Spherical Image Model

If we assume image point vectors of constant length then the noise is a 3-D vector with an uncertainty that is described by a 3×3 symmetric non-negative definite matrix Σ. It is obvious that the uncertainty cloud, is not a 3-D ellipsoid but a 2-D ellipse projected on the sphere. So, matrix Σ has one eigenvalue equal to zero; the one that corresponds to the eigenvector normal to the 2-D ellipse. The plane of the ellipse is tangent to the image sphere at point p'_i. The Lagrangian then takes the form:

$$L_i = n_i^T \Sigma_i^{-1} n_i + \lambda_{1i}(p'_i - n_i)^T E p_i + \lambda_{2i} n_i^T p'_i \qquad 2.1.3.1$$

The last term above, forces the vector n to be normal to p'_i, in other words to belong to the tangent plane. But we run into a small problem with the inversion of Σ_i. The matrix Σ_i is singular, because it has an eigenvalue equal to zero; the one that corresponds to the eigenvector normal to the tangent plane (this eigenvector is parallel to vector p'_i). The easiest way to go around this is to use instead the matrix Σ_i' which is exactly the same as Σ_i except that the zero eigenvalue is now δ. Then we let δ go to zero. So we solve eq 2.1.3.1 as usual to get the values for λ_1, λ_2 and n. Then we substitute to the L_i's

$$L_i = \frac{\left(p'^T_i E p_i\right)\left(p'_i \Sigma_i' p'_i\right)}{F} \qquad 2.1.3.2$$

where

$$F = \left(p'^T_i \Sigma_i' p'_i\right)\left((E p_i)^T \Sigma_i' E p_i\right) - \left(p'^T_i \Sigma_i' E p_i\right)^2$$

Since p'_i is an eigenvector of Σ_i' with eigenvalue δ then we replace $\Sigma_i' p'_i$ with $\delta\, p'_i$ and take the limit $\delta \to 0$, Then we apply L'Hospital's rule and set $p'^2_i = 1$ and we get:

$$\sum_i L_i = \sum_i n_i^T \Sigma_i^{-1} n_i = \sum_i \frac{\left(p'^T_i E p_i\right)^2}{(E p_i)^T \Sigma_i E p_i} \qquad 2.1.3.3$$

which is the almost the same as eq. 2.1.2.9. The only difference is that we have a different meaning for matrix Σ_i. Now Σ_i is a non negative definite singular matrix. In eq. 2.1.2.9 it was just a diagonal matrix with two ones and a zero.

2.1.3.2. Planar Image Model

Indeed it is the same except that Σ_i is different

$$\Sigma_i = \begin{bmatrix} \sigma_1 & \sigma_2 & \\ \sigma_2 & \sigma_3 & \\ & & 0 \end{bmatrix}$$

All the rest is the same as above:

$$L_i = n_i^T \Sigma_i^{-1} n_i + \lambda_{1i} (p'_i - n_i)^T E p_i + \lambda_{2i} n_i^T \hat{z} \qquad 2.1.3.4$$

and after some math we get again

$$\sum_i L_i = \sum_i n_i^2 = \sum_i \frac{\left(p'^T_i E p_i\right)^2}{(E p_i)^T \Sigma_i E p_i} \qquad 2.1.3.5$$

2.1.3.3. Planar Image Model with Noisy Correspondence Pair

This case is very difficult, so we have to eliminate the second order terms from the Lagrangian.

$$L_i = n_{i1}^T \Sigma_{i1} n_{i1} + n_{i2}^T \Sigma_{i2} n_{i2} + \lambda_{1i} (p'_i - n_{1i})^T E (p_i - n_{2i}) + \lambda_{2i} n_{1i}^T \hat{z} + \lambda_{3i} n_{2i}^T \hat{z} \qquad 2.1.3.6$$

from which we get

$$\sum_i L_i = \sum_i n_{i1}^T \Sigma_{i1} n_{i1} + n_{i2}^T \Sigma_{i2} n_{i2} = \sum_i \frac{\left(p'^T_i E p_i\right)^2}{p'^T_i E \Sigma_{i1} E^T p'_i + (E p_i)^T \Sigma_{i2} E p_i} \qquad 2.1.3.7$$

This an approximate expression as before. The good thing is that it is not more complicated than the one with isotropic noise. This concludes our analysis of least squares models.

2.2. Maximum Likelihood

We have seen several least squares models. We are moving now to Maximum Likelihood. Instead of studying several special cases before we move on to more general ones, we go directly to the general case with planar image. It is evident from the previous sections that the general case is no more complicated than the special ones.

The MLM method tries to maximize the likelihood of a solution. To see what this means consider the following. Choose a solution at random for the motion parameters and structure. Project the resulting structure on the second image. If this solution was the ground truth then this means

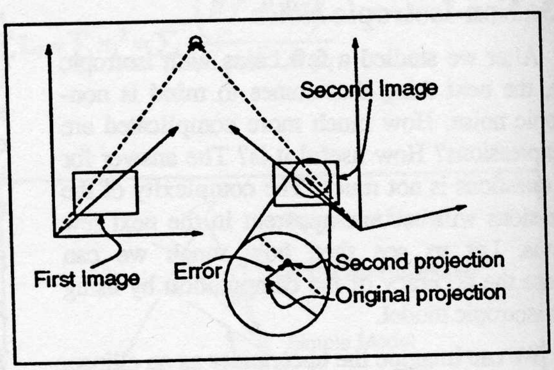

Figure 2.2.1: Let's say the point is ρ units away in the direction of p from the origin of the first (left in our figure) image. The point then projects to p''. The original projection is p'. The distance between these two projections is the error. To maximize the likelihood of ρ and the motion parameters we have to maximize the probability of this error.

that the distance between this second projection and the original projection is the error. If this error has very low probability to occur, and the same is true for all the points then it looks like a rare situation. If on the other hand the probability of such an error was high, it means that we are in a typical situation. Since it is natural to expect to encounter more common situations than uncommon, we should have preference for common ones. So a good solution should maximize the probability of the particular error pattern. We name this probability (or better the probability density) of the error pattern that comes from the choice of a particular solution *Likelihood.* And the method that finds solutions that maximize this probability *Maximum Likelihood* (Figure 2.2.1).

Let's say the 3-D point P_i is $\rho_i\, p_i$ in the system of reference of the first frame. We do not know ρ_i. The motion parameters are the rotation matrix R and the translation vector T. These motion parameters also describe the relative position of the systems of reference for frame one and two. The position of the same point in the system of reference of the second frame is

$$P'_i = \rho_i\, R\, p_i + T$$

and the projection of this point on the second image plane is

$$p''_i = \frac{P'_i}{\hat{z}^T P'_i} \qquad 2.2.1$$

In general this point p''_i is different than the point p'_i that is the given corresponding to p_i. The distance between them is the error $n = p''_i - p'_i$. Assuming Gaussian distribution the probability density of the error is

$$\frac{1}{2\pi\sigma}\, e^{-\frac{n^T \Sigma_i^+ n}{2}}$$

where Σ_i is defined as before, $\sigma = \sigma_1\sigma_3 - \sigma_2^2$ and the superscript + means pseudo-inverse. The likelihood for all the points is

$$\prod_i \frac{1}{2\pi\sigma}\, e^{-\frac{n^T \Sigma_i^+ n}{2}}$$

In order to maximize this we minimize the negative log of it or

$$\sum_i n^T \Sigma_i^+ n \rightarrow \min \qquad 2.2.2$$

Our unknowns are ρ_i (one for every image point pair) and the motion parameters. We can do the same as in previous sections and eliminate the per point parameters and be left with an equation with the motion parameters alone. Doing this we get

$$\rho_i = \frac{\left(p'_i \times T\right)^T \Sigma_i (T \times R\, p_i)}{(T \times R\, p_i)^T \Sigma_i \left(R\, p_i \times p'_i\right)} \qquad 2.2.3$$

If we substitute 2.2.3, 2.2.1 in 2.2.2 we get, after some simplifications

$$\sum_i n_i^T \Sigma_i^+ n_i = \sum_i \frac{\left(p'^T_i\, E\, p_i\right)^2}{(E\, p_i)^T \Sigma_i E\, p_i} \qquad 2.2.4$$

which is the same expression as 2.1.3.5.

The assumptions we used were Gaussian distribution, independence of noise (since we use one Σ_i matrix for each point) and rigidity. The Gaussian distribution is not a particularly strong assumption if we are able to obtain reasonable values for the Σ_i's and the actual distribution is unimodal (has one peak). This is not the case in a wide range of situations though (mismatches in a correspondence based algorithm or singularities in the computation of flow). But the alternatives we have are very limited and come with a heavy premium on the computational cost. The assumption of independence of the noise among various points is a quite strong assumption but we can relax it without major consequences.

2.3. Correlated Noise

When the noise cannot be considered independent (e.g. is correlated), we can still use MLM to estimate motion. The problem is that the expressions cannot be simplified. The same is true if we use least squares and Lagrangian multipliers as in section 2.1. Further, the representation of the covariance matrix is not easy because non-trivial correlations exist between pairs of flow vectors quite far apart (in fact we could not guarantee how far apart, without introducing more problems than we solve).

To proceed we need to state a few assumptions. One is that the method that provides the initial data for the disparity field is an overconstrained linear method or a Newton type method that involves a series of linear steps. So this excludes methods based on matching. We also assume that it makes sense to minimize the L_2 norm of the residual of the overconstrained linear system (e.g. minimize the sum of the squares of the residual vector). Then we use a trade off parameter λ to minimize

$$\mathbf{L} = (A\,x - b)^2 + \lambda \sum_i \frac{\left(p'^{\,T}_i \, E \, p_i\right)^2}{(E\,p_i)^T \Sigma_i E\,p_i} \qquad 2.3.1$$

where $A\,x-b$ is the residual of the overconstrained linear system of equations, x is the vector that contains all the disparity vectors (or flow vectors etc). Σ_i is either any reasonable approximation of the covariance matrix or

$$\Sigma_i = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix}$$

in the absence of anything better. The precise value of Σ_i is not important because $A\,x-b$ is playing exactly this role.

From a first look of eq. 2.3.1, one can tell that it is not an optimization using Lagrangian multipliers. It is closer to a regularization expression that uses rigidity instead of any kind of smoothness [1]. How can we use this then? If we have an initial guess on the motion parameters, then we can solve for x (or for p'_i which are directly related). This involves a minimization of $\mathbf{L}$ which is not a quadratic (one can drop the denominator of the expression in the summation and get a quadratic, or solve it with successive linearizations). Assume that we can do this minimization for a given set of motion parameters. Then $\mathbf{L}_{\min}$ is a function of the five motion parameters. The value of x, which is the vector of all the image disparities, that minimizes $\mathbf{L}$ is a set of disparities that almost fit the image data (since A depends on the image data) and at the same time satisfy the rigidity constraint if λ is large enough. If we minimize $\mathbf{L}_{\min}$ with respect to the motion parameters as well, the result is a set of disparities that fit the image data very closely while satisfying the rigidity constraint.

3. Conclusions

We presented an analysis of a series of models for Visual Motion Analysis. These include models that use planar or spherical image, isotropic noise or non-isotropic, use MLM or the simpler Least Squares, noisy correspondence or noise disparity vector and independent or correlated noise. Most of them turned out to be equivalent; some of them are expressed even by the same formula.

The question of whether some or any of these models is optimal remains almost open. The fact that some of them are derived from MLM is encouraging, but there are cases that MLM does not give an MVE (Minimum Variance Estimator), in which case no method can give one because such an estimator does nor exist. So the only thing we know is that if there exists an MVE, then we have found it because we used MLM.

Another question that remains open is the assumption of the noise distribution. All these methods work well with distributions that are similar to Gaussian [8] but we do not know what type of distributions to expect in realistic situations.

References

1. H. Ando, "Dynamic Reconstruction of 3-D Structure from Motion," *IEEE Workshop on Visual Motion: Representation and Analysis*, (1991).
2. T. J. Broida, R. Chellappa, and S. Chandrashekhar, "Recursive 3-D Motion Estimation from a Monocular Image Sequance," *IEEE Trans. on Aerospace and Electronic Systems* 26(4) pp. 639-656 (July 1990).
3. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton (1946).
4. B. K. P. Horn, *Robot Vision*, MIT Press (1986).
5. B. K. P. Horn, "Relative Orientation," MIT AI Lab Memo 994 (1988).
6. B. K. P. Horn, "Relative Orientation Revisited," *JOSA (accepted)*, (1990).
7. H. H. Nagel and K. Daniilidis, "Analytical Results on Error Sensitivity of Motion Estimation in Two Views," *ECCV*, (1990).
8. M. E. Spetsakis and J. Aloimonos, "Optimal Estimation of Structure from Motion from Point Correspondences in Two Frames," *Proc. ICCV*, (1988).
9. J. Weng, T. S. Huang, and N. Ahuja, "A Two Step Approach to Optimal Motion and Structure Estimation," *Proceedings of IEEE Computer Society Workshop on Computer Vision*, (1987).