Feature Grouping in a Hierarchical Probabilistic Network

Will Dickson*

Robotics Research Group
Department of Engineering Science
Oxford University
Oxford OX1 3PJ

The goal of feature grouping is to provide efficient codings of the necessary information for much of the scene interpretation and object recognition applications in machine vision. This paper offers a sound theoretical background for feature grouping processes, using a Bayesian approach which makes explicit the world knowledge which is applied at any stage. We describe a framework which can integrate many different forms of grouping and different levels of information. We also report on a preliminary implementation within this framework to group parallel lines in a perspective image. In support of this we develop a mapping from the image uncertainty to the orientation uncertainty for the hypothesised groups.

Figure 1: *A simple line image*

The grouping of features to form perceptual structure is basic in many machine vision applications. It has usually been divided into two problems: *segmentation* which is expected to yield a single, correct interpretation of the data and *grouping* where it is accepted that a single solution is not achievable from the available data. We address ourselves to the latter problem, with the hope of creating a system which is able to adapt to the level of ambiguity in the observed scene and thus exhibit qualities of *opportunism* and *graceful degradation*.

INTRODUCTION

The output of a feature grouping algorithm is usually a set of hypothesised groups, ranked according to a measure of their quality or likelihood. In general these algorithms are intended to feed into another system which brings in higher-level data in order to disambiguate the multiple solutions. A particular area of concern for us, indeed our motivation for research into this problem, is the use of a model-matcher as this higher-level system. The desired end-product is a system which could autonomously identify and localise instances of known objects from a *single* video image. To restrict the problem further, we only concern ourselves with a polyhedral world and use as our features line segments formed from the output of an edge detector. A simple example of such an image is shown in figure 1. This is the result of applying the Canny edge detector to an image, followed by a simple algorithm which finds linear segments in the Canny edge map. Some of the problems in the interpretation of these images can be seen here: duplicate edges due to shadows etc; blending of edges between objects;
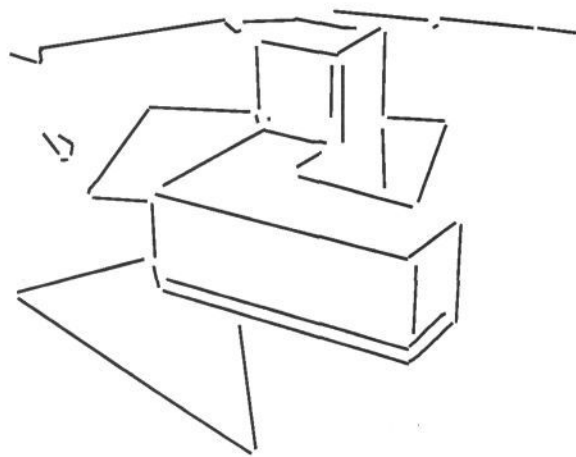
the occasional non-general viewpoint and the edge detector's endearing habit of breaking edges at occlusion boundaries. Feature grouping is even more critical in interpreting more complex images – for example images of occluded pallets. Here the grouping is critical both in extracting the salient features (parallel sets of lines in this case) and in reducing the data complexity.

After a brief mention of some of the other work in this field, this paper goes on to outline the principles of the application of Bayes' Rule to feature grouping. We then concentrate on applying these principles to a particular grouping cue: parallelism in perspective images. An algorithm is described which applies some of the ideas of the Bayesian analysis, and results are shown of the application of this algorithm to a simple scene.

PREVIOUS WORK

Our work was inspired by Lowe's SCERPO model-based vision system [5]. He develops a simple statistical analysis of grouping processes which he applies to finding particular perceptual structures in an image. He then matches these structures to corresponding structures in the model in order to generate a first estimate of viewpoint, which is then refined iteratively as further matches are made. Some problems with Lowe's system were described by McIvor in [6], one of which was an oversimplified statistical treatment. Horaud [3] has developed a similar system which attempts to generate more complex perceptual structures. Jacobs [4] has recently developed a system which finds perceptual structure based on convexity and uses that structure to recognise model subparts.

In terms of our approach, Pearl's description of Bayesian networks in [7] has been very influential.

STATISTICAL APPROACH

Our approach centres around Bayes' Rule [1]:

$$p(h|e) = \frac{p(e|h)p(h)}{p(e)} \qquad (1)$$

where

$$p(e) = \sum_{h \in H} p(e|h)p(h) \qquad (2)$$

In our case h is a grouping hypothesis and e is the available evidence. The prior expectations $p(h)$ for our grouping hypotheses are determined either empirically or heuristically.

Probabilistic Networks

Our initial concept was of a network where the nodes were the observed features, and the arcs corresponded to hypothesised relationships between them; the hope was to use a propagation mechanism similar to Pearl's in order to evaluate the likelihoods of the different relationships. We begin with the simplest possible case: two features X_1, X_2 and a hypothesised group H made up from them. Associated with each feature is a likelihood function $f_i(x_i)$ defined over the space of their parameters x_i. The false and true states h_f and h_t of the hypothesis have respective likelihoods $(1 - p_t)$ and p_t. Furthermore, we split the h_t state into a state space X for the parameters of the group and we can associate with this a distribution $f_h(x)$; h can now take values from the set $\{h_f\} \cup X$. If we consider the total evidence e_{tot}:

$$
\begin{aligned}
f(e_{tot}|h_f) &= (1 - p_t) \int_{X_1} f_1 dx_1 \int_{X_2} f_2 dx_2 \\
f(e_{tot}|h_t) &= \int_X f(e_{tot}|x) dx \\
&= p_t \int_X f_h f_1' f_2' dx
\end{aligned}
$$

where

$$f_i'(x) = \int_{X_i} f_i(x_i) f(x|x_i) dx_i \qquad (3)$$

We can multiply the likelihoods together in this way only if they are independent; here we have chosen the state-space for H such that the two feature likelihoods are *conditionally* independent so long as we know h.

Following the ideas presented by Pearl in [7] we could extend this to an arbitrary number of features as long as we preserve the above conditional independence; for this we would need to know that the hypotheses relating to a feature X_i were independent once x_i was known, ie that X_i is the only link between these hypotheses. This condition proves to be unattainable in a real system: we cannot generate a singly-connected network from the possible relationships between the features in a real image.

The solution to this lies in explicitly representing the groups within the image. Unfortunately this solution in its naive form has a completely unmanageable level of complexity. It lies in fact somewhere between e^n and n^n where n is the number of features in the image[1]. For even simple images such as figure 1 n is around 60, so there is little hope of success from a frontal attack. Fortunately we do not need to evaluate the entire hypothesis space, as much of it will have a negligible likelihood. Furthermore the data will tend to separate into cliques separated by clear 'boundaries'. The groups involving data from more than one clique will have a very low likelihood. Thus we can evaluate the groups independently within the cliques rather than over the whole space, with the exponential gain which results.

The Vanishing Direction

In this paper we will concern ourselves only with grouping line segments into *parallel* sets in perspective images. A complete feature grouping system needs to concern itself with many grouping processes; we have not addressed the additional problems posed by such a system here. We chose to concern ourselves with parallelism because it is exemplary of the class of grouping problems where a single feature leaves one (or more) of the grouping parameters unconstrained: in this case a line segment constrains the orientation to lie within the *tangent* plane (defined by the line segment in the image plane and the camera focal centre) but leaves the orientation otherwise unconstrained (see figure 2(a)).

The traditional parameter space for orientations in space is the Gaussian sphere. Traditionally also this has been parametrised by two angles in a global coordinate system. We represent our orientations by a unit vector $\vec{x}$ in space which we call the *vanishing direction* (ie: a point on the Gaussian sphere) with which we associate a normal plane (the plane tangent to the Gaussian sphere) within which we represent covariances etc... in local coordinates without the singularities associated with the global coordinate systems. We project the likelihoods from the image onto the Gaussian sphere as described in the appendix, and then back-project them onto the local tangent plane; the equations developed do not depend on the orientation vector $\vec{x}$ being normalised which considerably eases this back-projection.

The vanishing direction corresponds to the unit vector along the direction of the line from the camera focal centre to the vanishing point in the image plane. We can construct the vanishing direction vector from the projections l_1 and l_2 of two parallel lines L_1 and L_2 in the following manner (see figure 2(b)):

1. construct the two normals $\vec{n_1}$ and $\vec{n_2}$ to the tangent planes P_1 and P_2 corresponding to each of the line segments respectively;

2. construct $\vec{x}$ as the direction of the intersection of the two planes: $\vec{x} = \vec{n_1} \wedge \vec{n_2}$.

To show that this direction is also that of the parallel lines in space (up to camera calibration errors):

- lines L_1 and L_2 lie within planes P_1 and P_2 respectively;

---

[1] We have a recurrence formula which shows that the complexity lies close to but remains less than n^n.
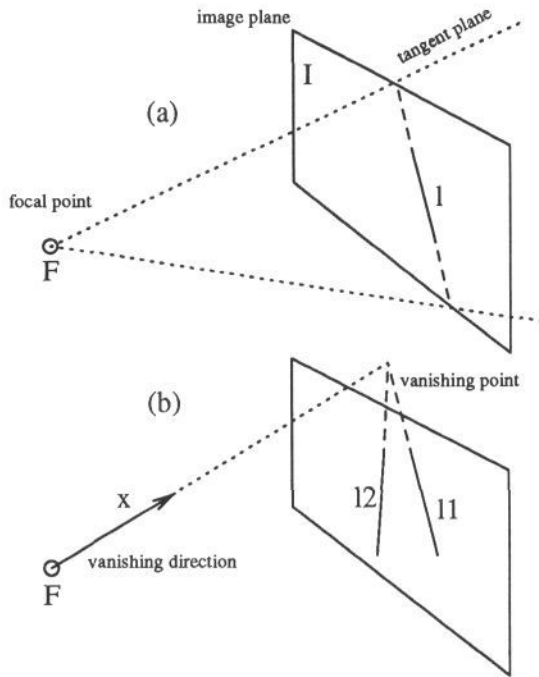
Figure 2: *(a): constructing the tangent plane; (b) inter-secting two planes to find the vanishing direction.*

- L_1 and L_2 never meet, so must both be parallel to the intersection of P_1 and P_2;

- $\vec{x}$ is parallel to $P_1 \cap P_2$, so must be parallel to both L_1 and L_2.

The vanishing direction has therefore two distinct advantages over the vanishing point as a grouping parameter:

1. the lack of singularities when the observed parallel set is parallel to the image plane;

2. the correspondence between the vanishing direction and the true direction of the line segments in space.

Computing the hypothesis likelihoods

The expression in equation (3) corresponds to a summation over the subspace of the feature's parameter space which is consistent with the particular value x of the group's parameters – in this case the vanishing direction. In the Appendix we develop an analytical expression for the likelihood of a line segment given a vanishing direction:

$$p(e|\vec{x}) = \left[\frac{\vec{x}^T A' \vec{x}}{\vec{x}^T B \vec{x}} \right]^{\frac{1}{2}} \exp\left(-\frac{\vec{x}^T A \vec{x}}{\vec{x}^T B \vec{x}} \right) \qquad (4)$$

where the matrices A, A' and B are derived from the observed line segment. The above expression[2] is linearised about the current estimate of $\vec{x}$ in order to approximate it by a Gaussian function. This approximation greatly simplifies the calculations.

Combining likelihoods

The expression for a normalised Gaussian distribution

---

[2] For what are euphemistically known as 'historical' reasons, we take $\frac{\vec{x}^T A' \vec{x}}{\vec{x}^T B \vec{x}}$ to be constant in our implementation. The full effects of this have not been analysed.

is:

$$N(\overline{p}, M^{-1}) = \frac{\sqrt{|M|}}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}(p - \overline{p})^T M(p - \overline{p}) \right) \quad (5)$$

where n is the dimension of the parameter vector p. When we multiply two of these together, the result is a non-normalised Gaussian:

$$N(\overline{p}_1, M_1^{-1}) \times N(\overline{p}_2, M_2^{-1}) = \lambda e^{-\kappa} \times N(\overline{p}', M'^{-1}) \quad (6)$$

where

$$M' = M_1 + M_2 \qquad (7)$$

$$\overline{p}' = M'^{-1}(M_1\overline{p}_1 + M_2\overline{p}_2) \qquad (8)$$

$$\kappa = \frac{1}{2}(\overline{p}_1 - \overline{p}_2)^T M_1 M'^{-1} M_2(\overline{p}_1 - \overline{p}_2) \qquad (9)$$

$$\lambda = \left(\frac{|M_1||M_2|}{|M'|(2\pi)^n} \right)^{\frac{1}{2}} \qquad (10)$$

We use κ, the statistical distance (or normalised distance) and λ in our estimator for the likelihood of a group. The expression for $\overline{p}'$ is used to find an initial estimate of the group's parameters; we then re-linearise about this point to provide better estimates for all four parameters. The likelihood of the data once grouped is then $\lambda e^{-\kappa}$ times the separate likelihoods. This could be viewed as a *likelihood gain*.

IMPLEMENTATION

We use a split-and-merge approach in our experimental implementation. We start from an initial global interpretation where each feature forms a distinct group which we "know" is indivisible. From this we create new hypotheses by merging those hypotheses when the likelihood of the resultant hypothesis is greater than the combined likelihood of the two separate hypotheses. The search is carried out depth-first in parallel: each hypothesis can operate as a distinct process. Our implementation is currently on a serial machine and thus the parallelism is only simulated.

The basic algorithm is as follows:

```
Initialise NETWORK
repeat
    for each active NODE in NETWORK
        if Ok-to-merge(NODE)
            Merge-with-best(NODE)
until Termination-condition
write out all active nodes
```

The Initialise function allocates a hypothesis node for each feature, and sets up an initial neighbourhood of possible matching features according to the value of the statistical distance between them. In order to restrict the combinatorial complexity the statistical distance is augmented by a measure of the actual distance between a pair of line segments, scaled inversely according to their lengths. This distance was not subsequently used in the evaluation of the hypotheses, but only in the statistical distance measurement.

The main loop attempts to create new hypothesis nodes by merging the existing ones. The condition tested

in `Ok-to-merge` is that the best merge for `NODE` has `NODE` as its best merge. This interlocking ensures that the most likely hypotheses are generated before the less likely ones, but the assessment is of course only based on the local evidence. Then the `Merge-with-best` function carries out the merging, creating a new neighbourhood from the old neighbourhoods, based on the same statistical distance function as `Initialise`; it also sets a flag in the neighbourhood pointers of the two nodes being merged to prevent them from merging again, and rejects any node whose likelihood is very low (as a result of the non-linearities in the likelihood function). We have not determined any sophisticated termination conditions; at present we use an iteration count or an execution-time bound.

The simple approach outlined above is not satisfactory as it stands and has two serious problems. The first is that many duplicate hypotheses are formed. While we prevent any two nodes from merging more than once, we cannot prevent the same node being created via several different paths. Thus the functionality of `Merge-with-best` needs to be extended to determine whether or not the new node being created already exists. The second problem is that the actual likelihood gain is not taken into account. This is where the notion of an *active* node comes in. We define as inactive any node whose likelihood is less than the combined likelihood of its 'component parts' or descendants (those nodes which merged to create it). Thus the algorithm does not proceed along any 'downhill' paths.

In order that the most complex groups are created as early as possible, any hypothesis which has an active ascendant waits until its ascendants have exhausted all the possible moves before becoming active again itself. Alone, this leads to deadlocks; these can be detected by the ascendant which then temporarily returns control to its descendents until the deadlock has been resolved.

For the statistical distance to be valid, we need to ensure that the two evidence-sets being compared are independent. When we create a new hypothesis, its neighbourhood is generated from the neighbourhoods of its two descendents. In doing this we discard any duplicate neighbours; in order to ensure independence we also discard any neighbours whose intersection with the new group is non-null.

As the algorithm proceeds we maintain a network of ascendants and descendents of each hypothesis node. The ascendants are those hypotheses which are created from the current node, while the descendents are those hypotheses which merged to form the current node, or merged at a later time and were added to the network. This network is (at present) used for two things:

1. it allows an efficient, local search for duplicate nodes;

2. it provides a means by which likelihoods can be compared in a statistically sound manner, since all pairs of descendents (and all their pairs of descendents) are independent.
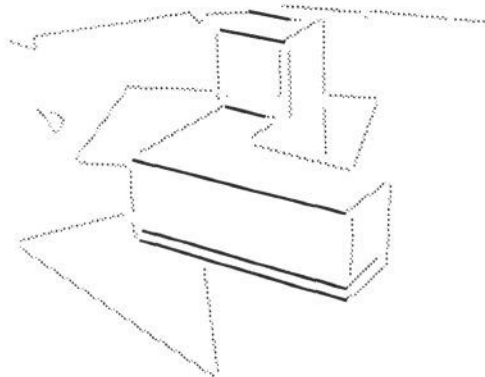

Figure 3: *best group from widget image*

The ability to compare likelihoods should allow us to implement a statistically sound hypothesis-pruning strategy. Furthermore the ascendant-descendent structure should provide an efficient way to extend matches between model and image features in a model-based vision system.

PRELIMINARY RESULTS

The algorithm was run on the image of figure 1. The results are shown in figures 3 and 4. Figure 3 is the best group obtained; the second and third best (not shown here) are subsets of this group. The group in figure 4 is the fifth-best. The fourth best group was formed from two segments in the background behind the object. The remaining groups are either not particularly significant or are subsets of one of these two.

The likelihoods are expressed as the negative logarithm of the actual likelihoods – effectively a squared error term. The reference for these (likelihood = 1 or error = 0) is the initial state, where all the features are considered as separate. The error for the first group shown above is -5.6 and for the second it is -2.0, giving likelihoods of 270 and 7 respectively. The parameters for these likelihoods were chosen in a fairly *ad hoc* manner such that they gave reasonable performance on a range of similar images. The algorithm is written in C and ran unoptimised in 8 seconds on a Sun SPARCstation.

DISCUSSION

The results achieved are encouraging: the statistical approach is apparently competent, and the computational burden of our implementation is reasonable for simple images. We do not however claim that perceptual structures based on parallelism alone will ever be sufficiently rich to allow robust model-matching. We will also argue here that a critical concept is lacking from our statistical knowledge at this time.

In order to extract useful perceptual structure from an image, we need to apply more than one grouping process; at the very least we would wish to extract vertices and make use of collinearity to repair some of the defects due to the edge detection. The full Bayesian statistics provide for competition between these different processes where they conflict and cooperation where they agree;
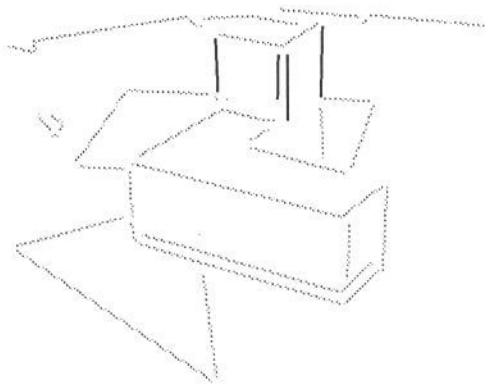
Figure 4: *another group from widget image*

Figure 5: *Line segment parametrisation*

our computational approach needs to capture both of these.

The results shown here correspond to parameters which were chosen because they gave reasonable results rather than because they were theoretically sound. There is good reason for this. We wish here to group segments according to perspective parallelism; however we can construct a vanishing direction for *any pair* of line segments with zero error but a relatively small prior likelihood. In order that our algorithm should progress we need to estimate this likelihood as being higher than the more theoretically sound one, so our parameters are biased towards this. What would be far more satisfactory from our point of view is a method of estimating the *uncertainty* in a likelihood estimate, or of estimating the utility of an action according to a loss function related to the expected likelihood of the ascendants of a possible hypothesis. This is the critical concept lacking here in our statistical treatment.

CONCLUSION

We have outlined a statistically rigorous approach to feature grouping in its most general sense, and successfully applied that approach to a particular case of feature grouping where there exists a substantial amount of ambiguity. In doing this we have also developed a likelihood function for spatial orientation of a line segment given an observed segment. We believe that the formal nature of these developments is the major contribution of this paper.

Our future aims are to extend the implementation to cover a greater range of grouping processes in order to extract complete perceptual structures. We hope to be able to integrate the instantiation and localisation of a known object model within the same framework as we use to evaluate the other hypotheses. We need to generate a theoretical backing for a fine-grained utility function to direct the search, and for a pruning strategy which will allow us to tackle more complex images. Finally we need a method of estimating the prior likelihoods for hypotheses to replace our current heuristic methods. See [2] for full details.
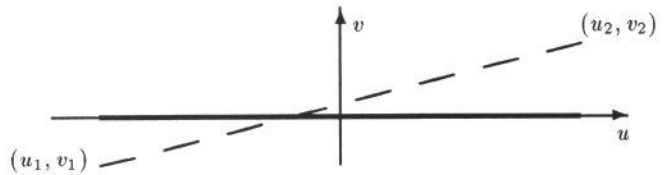
APPENDIX
Constrained likelihood for a line segment

We describe the statistics of a line segment in 2D in terms of its two endpoints $\vec{p_1}$ and $\vec{p_2}$. The longitudinal error is assumed normally distributed and independent for each endpoint as well as independent of the lateral errors. With a line segment we associate a local coordinate system u, v as in figure 5 with the origin at the midpoint of the line. In this figure, the bold line is the observed segment and the dashed line is a hypothesised true segment. We concern ourselves here purely with the lateral errors v_1 and v_2. These are described by a joint normal distribution based on the squared normal distance of the observed segment from the supporting line of the hypothesised segment. This gives an error expression of the form:

$$l_0 \epsilon^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(v_1 \frac{\frac{1}{2} - u}{l} + v_2 \frac{\frac{1}{2} + u}{l} \right) du$$
$$= \frac{l}{3} \left(v_1^2 + v_1 v_2 + v_2^2 \right)$$

The resulting information matrix is:

$$M = \frac{l}{l_0 \sigma_n^2} \begin{bmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{bmatrix} \qquad (11)$$

where $l_0 \sigma_n$ is the expected value of the squared distance. The resulting probability density function is:

$$f(\vec{v}) = \frac{\sqrt{|M|}}{2\pi} \exp \left(-\frac{1}{2} \vec{v}^T M \vec{v} \right) \qquad (12)$$

The eigenvector and eigenvalue matrices for M are:

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = W^{-1}$$

$$\Lambda = \frac{l}{l_0 \sigma_n^2} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1}$$

where σ_1^2 and σ_2^2 are the decoupled variances. Let us now assume that v_1 and v_2 are linearly related:

$$v_2 = a v_1 + b \qquad (13)$$

We can decompose this along the principal axes of the distribution, giving:

$$e_1 = \frac{(1+a)v_1 + b}{\sqrt{2}}$$

$$e_2 = \frac{(1-a)v_1 - b}{\sqrt{2}}$$

Which then gives us:

$$e_2 = \frac{(1-a)e_1 - \sqrt{2}b}{1+a} \tag{14}$$

Thus our expression for the total error becomes:

$$\frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} = \left(\frac{1}{\sigma_1^2} + \frac{(1-a)^2}{(1+a)^2\sigma_2^2}\right)e_1^2$$
$$- \frac{2\sqrt{2}(1-a)b}{(1+a)^2\sigma_2^2}e_1 + \frac{2b^2}{(1+a)^2\sigma_2^2}$$

We can write this in the form:

$$\frac{(e_1-\mu)^2}{\sigma^2} + \rho \tag{15}$$

which gives us:

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{(1-a)^2}{(1+a)^2\sigma_2^2}$$
$$\mu = \frac{\sqrt{2}(1-a)b\sigma_1^2}{(1-a)^2\sigma_1^2 + (1+a)^2\sigma_2^2}$$
$$\rho = \frac{2b^2}{(1-a)^2\sigma_1^2 + (1+a)^2\sigma_2^2}$$

There remains to determine the direction along which the likelihood should be integrated:

$$|d\vec{s}| = \sqrt{de_1^2 + de_2^2}$$
$$= de_1\sqrt{1 + \frac{(1-a)^2}{(1+a)^2}}$$
$$= \beta de_1$$

So finally, the likelihood of the line segment data given a and b is:

$$p(e|a,b) = \beta\sigma\sqrt{2\pi}\exp(-\rho) \tag{16}$$

Mapping to Orientation Space

We now reformulate the results obtained in the previous section in terms of the vectors in orientation-space. The values of a and b which constrain the line in equation 13 correspond to the choice of a vanishing point in the image through which the line segment must pass. If we refer all our vectors to the focal centre of the image, let $\vec{p_1}$, $\vec{p_2}$ refer to the two observed endpoints in the image, $\vec{x}$ refer to the vanishing direction and $\vec{n}$ be the normal to the line in the image plane. The condition that the line segment should pass through the vanishing point is equivalent to asking that the two *actual* endpoint vectors and the vanishing direction be coplanar, which we can express via the vector triple product:

$$[\vec{p_1} + v_1\vec{n}] \wedge [\vec{p_2} + (av_1+b)\vec{n}].\vec{x} = 0 \tag{17}$$

This gives us the two equations:

$$(a\vec{p_1} \wedge \vec{n} - \vec{p_2} \wedge \vec{n}).\vec{x} = 0$$
$$(\vec{p_1} \wedge \vec{p_2} + b\vec{p_1} \wedge \vec{n}).\vec{x} = 0$$

which give us:

$$a = \frac{\vec{p_2} \wedge \vec{n}.\vec{x}}{\vec{p_1} \wedge \vec{n}.\vec{x}}$$
$$b = -\frac{\vec{p_1} \wedge \vec{p_2}.\vec{x}}{\vec{p_1} \wedge \vec{n}.\vec{x}}$$

Some further manipulation gives us:

$$\rho = \frac{2(\vec{p_1} \wedge \vec{p_2}.\vec{x})^2}{[(\vec{p_1}-\vec{p_2}) \wedge \vec{n}.\vec{x}]^2\sigma_1^2 + [(\vec{p_1}+\vec{p_2}) \wedge \vec{n}.\vec{x}]^2\sigma_2^2}$$
$$(\beta\sigma)^2 = \frac{2\sigma_1^2\sigma_2^2\left[(\vec{p_1} \wedge \vec{n}.\vec{x})^2 + (\vec{p_2} \wedge \vec{n}.\vec{x})^2\right]}{[(\vec{p_1}-\vec{p_2}) \wedge \vec{n}.\vec{x}]^2\sigma_1^2 + [(\vec{p_1}+\vec{p_2}) \wedge \vec{n}.\vec{x}]^2\sigma_2^2}$$

Substituting for $\sigma_1^2 = \frac{2\sigma_n^2 l_0}{l}$ and $\sigma_2^2 = \frac{6\sigma_n^2 l_0}{l}$:

$$\rho = \frac{l}{l_0\sigma_n^2}\frac{(\vec{p_1} \wedge \vec{p_2}.\vec{x})^2}{[(\vec{p_1}-\vec{p_2}) \wedge \vec{n}.\vec{x}]^2 + 3[(\vec{p_1}+\vec{p_2}) \wedge \vec{n}.\vec{x}]^2}$$
$$(\beta\sigma)^2 = \frac{12\sigma_n^2\frac{l_0}{l}\left[(\vec{p_1} \wedge \vec{n}.\vec{x})^2 + (\vec{p_2} \wedge \vec{n}.\vec{x})^2\right]}{[(\vec{p_1}-\vec{p_2}) \wedge \vec{n}.\vec{x}]^2 + 3[(\vec{p_1}+\vec{p_2}) \wedge \vec{n}.\vec{x}]^2}$$

We can now express the likelihood in equation (16) as a likelihood conditioned on $\vec{x}$, the vanishing direction.

For ease of computation we can reformulate the above expressions in a matrix form:

$$\rho = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$$
$$(\beta\sigma)^2 = \frac{\mathbf{x}^T\mathbf{A'}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$$

The matrices need only be computed once for each line segment.

REFERENCES

[1] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans.*, 3:370-418, 1763.

[2] J. W. Dickson. *Forthcoming D.Phil. Thesis.* Oxford University, Department of Engineering Science, 1990.

[3] Radu Horaud. Combining Image and Spatial Reasoning for Model Retrieval. In *Proc. European Conference on A.I.*, 1986.

[4] David W. Jacobs. The Use of Grouping in Visual Object Recognition. Technical Report Tech.Rep. 1023, Massachusetts Inst. Technol., 1988.

[5] D. G. Lowe. Three-Dimensional Object Recognition from Single Two-Dimensional Images. *Artificial Intelligence*, 31:pp.355–395, 1987.

[6] Alan M. McIvor. An Analysis of Lowe's Model-based Vision System. In *Proc. of Alvey Vision Conference*, 1988.

[7] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, Inc., 1988.