MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1265 OCTOBER 1990

Estimation of Discontinous Displacement Vector Fields with the Minimum Description Length Criterion

Joachim Dengler

Abstract

A new noniterative approach to determine displacement vector fields with discontinuities is described.

In order to overcome the limitations of current methods, the problem is regarded as a general modelling problem. From this point of view the imaging field consists of a set of regions with common properties and their boundaries. The strategy proposed is an analysis of consistency of the displacement estimators between different levels of regularization.

A fully regularized family of model based displacement vector fields is constructed by successive smoothing and subsampling. By measuring the difference in description length the compatibility between different levels of regularization is measured. This gives local but noisy evidence of possible model boundaries at multiple scales.

With the two constraints of continous lines of discontinuities and the spatial coincidence assumption consistent boundary evidence is found. Based on this combined evidence the model pyramid is updated, now describing homogeneous regions with sharp discontinuities.

Copyright © Massachusetts Institute of Technology, 1990

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institue of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defence under Office of Naval Research contract N00014-85-K-0124. The author has been on leave from the German Cancer Research Center in Heidelberg for this project and has been supported by a fellowship grant of the Max-Kade-Foundation.

1 Introduction

The problem of determining discontinuities in displacement vector fields arises in both Binocular Stereo and Motion analysis. It has received much attention in computer vision research. There are two ways of viewing this problem. It can be seen as an "edge" detection problem in a vector field, or it can be seen as a problem of finding regions with common properties. Both views are complementary to each other. The latter view relates the question of discontinuities in vector fields to the old problem of segmentation, with the important difference that regions are not defined by the the image property "gray value", but by the object property "displacement". In the field of displacement vector field estimation there are, however, few attempts to relate this to segmentation approaches. The domain of segmentation techniques have been traditionally and still are gray value related features, but hardly vector fields as they appear in Stereo or motion.

The reason for this is the fact that local displacement vector field (DVF) estimates have large variations in their variance, which is not only due to noise effects, but is inherent to the displacement estimation itself, depending on the image geometry. In "normal" segmentation tasks the variance of a modelled regional property is mainly determined by noise, and therefore the task can in principle be solved with an adequate noise model. An excellent example of this is described in [Lec89].

A number of approaches have been applied to solve the problem of finding displacement information and the discontinuities of it. It seems reasonable to divide them into three different strategies.

1.1 Current Approaches

1.1.1 Discontinuities from Displacement Vector Field

The first approach separates the problems of the determination of the displacements and the detection of discontinuities. In the first step it is assumed that there are no discontinuities, and a fully regularized solution is determined. Examples of this strategy are the "Optical Flow" algorithm [HS81], surface interpolation applying the model of the thin plate [Gri81, Ter83], or the membrane model in matching [Bro81], and a series of motion estimation schemes [Nag83, Hil84, Den86, NE86, Ana87, TWK87]. The detection of discontinuities follows as a second step on the basis of "tensions" in the determined vector field. These can in principle be found by significant gradients [DS88] or by the zero-crossings of the second derivative of the flow field [TMB85]. It must be stressed that up to now only this type of approach for finding a dense displacement vector field has a time complexity that allows a real-time implementation when an appropriate control strategy in combination with an efficient multi-grid solver is used [SD89].

It has been shown that the optical flow allows qualitative statements about

objects' motion, but that the quantitative optical flow is not a reliable measure of the real motion [VP87, SU87]. This applies even more to the derivatives of the flow field. It can be expected, and indeed it can be demonstrated, that discontinuities determined in this manner deviate considerably from the actual ones in real-world sequences.

Another approach was the segmentation of the optical flow field with a model based Hough transform, using the constraints of 3D rigid body motion [Adi85].

1.1.2 Simultaneous estimation of Discontinuites and Vector Field

An appealing approach from a theoretical point of view is to minimize a single functional, where different processes are combined. Typical elements of this functional are a similarity function between potentially corresponding image patches, the regularization conditions, and a line or edge process, that limits the regularization and therefore allows "continous discontinuites". Following this approach the determination of the displacements and the detection of discontinuities are interwoven [KMY86, WW85].

The first problem with this approach is that the constructed functional is no longer convex, it has local minima that can be far away from the global minimum. (Strictly speaking, the functional of the continous approach is not convex either, the convexity is enforced by a local approximation of the gray values, and most local minima are avoided by a coarse-to-fine control strategy [NE86, Ana87, TWK87, DS88]).

A more severe problem is the open question of how to weigh the different parts of the functional. Typically these coupling constants are determined heuristically. It has been shown that the optimal choice depends on the image content, and can vary within an image [Yui87]. In the extreme case so many free parameters are introduced that the entire approach is led ad absurdum, since its elegance is based upon the idea of reducing a complex problem to a simple principle. The approach can only be succesful when these parameters can be found from apriori principles and the image data themselves. Whether the proposed min-max principle [GY89], where the functional is maximized with respect to the coupling parameters and minimized with respect to the displacements, is adequate for real image sequences, is an open question, especially with space-variant parameters. Also, the computational costs are immense, because for every configuration of parameters the entire minimization problem has to be solved.

1.1.3 Discontinuities without full Displacement Vector Field

This approach also separates the two problems, the discontinuities are, however, detected first. When this segmentation problem is solved, then the computation of the displacements becomes simple, often just a least squares estimation or the

minimization of a simple quadratic functional. This segmentation is also a precondition to the direct determination of 3D structure or motion [TH84, NH85]. It may seem to be an impossible task to estimate the discontinuities before the vector field itself, but there is already encouraging evidence that such information can indeed be found. One approach in this direction is the analysis of local histograms of possible displacements and the search for "vanishing bars" [SU87], and another is the detection of significant non-vanishing divergence of the local non-iteratively determined displacement vector field [NA89]. While this type of approach is appealing due to its noniterative character, these results are at best preliminary, and a precise analysis of their reliability is missing. These approaches indicate, however, that it is indeed possible to extract the important information about discontinuities at a very early stage of the vision process.

1.2 Observations which motivate a new scheme

For the developement of a new scheme to the detection of displacement discontinuities and the segmentation of motion fields the following observations will be of importance:

- The vector field determined by a regularization approach such as the "optical flow" or a gaussian weighted least squares fit is more or less correct in the interior of a displaced region, apart from the distorting effects discussed in [VP87]. The size of the region determines how much smoothing of the vector field is appropriate. Within these limits the noise is better suppressed and thus the vector field is more reliable, if more smoothing is applied. The vector field changes dramatically, however, when the smoothing is too strong.

- In the interior of a region the measured local data are compatible with the regularized vector field. The only deviations are due to noise.

- Near the occlusion boundaries the estimates of the regularized fields deviate significantly from the actual displacement. Near occlusion boundaries a less regularized vector field reflects the actual displacement better than a smoother vector field, while at the same time it is more sensitive to noise. In particular the measured data themselves are significantly incompatible with the regularized vector field. The consequence of this is that close to occlusion boundaries essentially all members of a family of regularized solutions are incompatible with each other.

- An observation made by David Marr is that discontinuities of any given dimension are relatively sparse and they are continuous in the next lower dimension [Mar82]. This means that discontinuites of 2D regions are essentially continous lines. These lines can have again relatively few directional

discontinuities in the form of endstops and breakpoints. These are, however, not considered in this paper, but a smooth and closed occlusion boundary is assumed.

- If there is an actual boundary, it shows up at multiple scales. This was formulated by David Marr as the spatial coincidence assumption [Mar82].

- The detection of boundaries due to motion alone is instantaneous for the human visual system. This can be easily found out with a moving random dot pattern. Therefore there must be a fast mechanism with (at most) a few iterations to find these discontinuities.

As a consequence of these observations the following strategy is followed.

1.3 General strategy

At the beginning of the process a family of regularized displacement vector fields is determined. According to [VP87] it is not critical which particular regularization method is applied. Therefore a computationally cheap noniterative approach of local weighted least squares estimates is followed. With this approach the calculation of the family of regularized vector field can be decomposed into the construction of a kind of Gaussian pyramid of moments and a pointwise calculation at each node of the pyramid. With this procedure not only the actual displacement estimates but also the sum of squared deviations can be determined. Each level of this pyramid represents one level of regularization, with the measured data at the bottom, and the global average displacement at the top.

The most critical part of the procedure is to calculate the compatibilites between the different levels of the pyramid. It will be shown that the traditional Euclidian distance between properties corresponds to the difference of the sums of squared deviations, when only mean values are estimated. For comparing regression coefficients as in the case of displacement information euclidian distances are not suitable.

The difference of sums of squared deviations with respect to the displacement model is a candidate for a general compatibility measure, with the nice properties of a metric. There are two drawbacks, however. First there is no inherent interpretation in this measure, which also means that for quantitative use a heuristic scaling parameter needs to be introduced. This occurs, wherever the sums of squared deviation or euclidian distances are used as compatibility measures. Secondly it is impossible to use the sums of squared deviations for comparing different models which all explain the data. In motion interpretation this could, for example, be pure translation versus translation with rotation and scaling. The sums of squared deviations is a monotonously decreasing function of the number of parameters, and it vanishes when there are as many parameters as observations.

In order to overcome these two drawbacks, the criterion of minimal description length (MDL) can be used as a compatibility measure to compare displacement models at different levels of the pyramid. Although this compatibility measure is not a metric, it has a probabilistic interpretation and it can be used to compare models of different complexity. Even when only a single model is used, compatibility defined by the MDL criterion is less sensitive to structural variations in the images than the other measures.

The noisy evidence of each individual scale leads to the problem of combining evidence from multiple locations and scales. Due to the weak contrast of local measurements of displacement discontinuities, the "continuity of discontinuities" and the "spatial coincidence assumption" [Mar82] serve as guiding principles, allowing displacement discontinuities only if there is combined evidence from the neighborhood and from several scales.

A pyramidal structure with the local links between its levels opens the possibility of constructing a model that makes use of these constraints.

1.4 Outline of the paper

The next section discusses the minimum description length approach, in particular with its application to the regression problem.

In section 3 the question of compatibility of two sets of data with respect to a given model is discussed. The displacement estimation problem is described as a regional regression problem. Three different compatibility measures are presented.

Section 4 describes the actual scheme, where the individual steps are described and illustrated with an example of a random dot stereogram. At the end the result of applying the algorithm to a real-world image pair from a sequence is shown.

Section 5 concludes the paper with a discussion of the approach and an outlook to possible extensions.

2 The Minimum Description Length Approach

The extension of the widely applied maximum likelihood principle is motivated by the question what an overall best model for a given set of data is. In order to have a clear definition of what a "best" model is, not only the goodness of fit of a model, but also the complexity of the model itself have to be considered. In this sense, the "best" model is the simplest that can explain a given set of data. While this as a general statement has been known at least since "Occam's Razor" and has been a philosophical guideline in the developement of the natural sciences, it is far from trivial to formalize this principle in such a way that it is applicable to determine a specific model.

2.1 General background of minimum description length

The foundations of a formalization were laid in information theory, by the introduction of an algorithmic definition of information [Sol64, Kol65, Cha66], which defines the information content of a string by the minimal length of a program to describe this string. A limitation to the application of this idea was already set by one of its inventors, stating, that no program can automatically generate such a program or description of minimal length [Kol65]. This essentially rules out a fully automated process of general induction. It does not rule out, however, the testing of any number of models—invented by human ingenuity—by this criterion. This latter approach mirrors the philosophy of critical rationalism widely adopted in the field of science [Pop59].

The practical applications and further theoretical justifications of this concept were developed independently by C.S. Wallace and his coworkers [WB68, BW73, WB75, GW84, WF87] and by J. Rissanen [Ris76, Ris78, HR82, Ris83, Ris86, Ris87]. Both envisage primarily the statistical problem, how to choose an optimal set of parameters within a given model class. The length of the description of the parameters is then interpreted as the negative logarithm of a Bayesian prior, and the best model is found by maximizing the product of conditional and prior probabilities. Putting it another way, the optimal model is selected by minimizing the sum of the negative binary loglikelihood of the data given the model, and the length of the shortest code to describe the parameters. Although the approach appears to be formally Bayesian, there are fundamental differences. The most prominent difference is that the relevant distribution for the MDL concept—the "marginal prior distribution" [WF87] or the "stochastic complexity" [Ris87]–is a distribution on the data, and not on a model.

One key question to make the approach feasible is, how to encode and therefore how to optimally truncate the—typically real-valued—parameters. The answer was found by both researchers independently, although the resulting formulas look rather different, and appear to be partly inconsistent. The parameters are truncated in such a way, that the truncation error equals the statistical estimation error of the model (see e.g. [Ris83, WF87]).

2.2 Minimum Description Length of Regression

One important motivation to apply the principle of Minimum Description Length to the regression problem is related to the question of the best choice of regressor variables. Specifically when modelling a 1D signal or a 2D surface with piecewise polynomials naturally one has to somehow determine the optimal degree of the elements. Interweaved with this is the problem of finding the optimal breakpoints in 1D or the region boundaries in 2D. Both problems can be formulated in terms of a descriptive language. By minimizing the length of the code in this language

the optimal description is found. The gray value segmentation problem has been treated in this spirit by Y. Leclerc [Lec89].

The main emphasis here is laid on measuring the statistical compatibility of two sets of data with respect to a given model. This also means to find the optimal degree of a polynomial. The partitioning problem is here approached in a way that is based on fundamental insights about discontinuities in the visual input data, thereby avoiding the difficult and time-consuming minimization of functionals that include explicit boundary models.

2.2.1 Selection of variables

Estimating the best degree of a polynomial for a given set of data points is a special case of the general selection of variables problem. The model likelihood with the maximum likelihood estimates of the regressor variables naturally increases with the number of variables used. Therefore the otherwise powerful maximum likelihood principle fails to give an answer to the important question, which variables explain the observed data best. The description length approach extends the maximum likelihood principle by including a term that measures the model complexity.

2.2.2 A simple measure of model complexity

An intuitive approach to capture the notion of complexity is the following [Foe89]. When the variables explain the data optimally, then the residual variance is entirely due to noise. Assuming Gaussian noise, the standard deviation of the noise is proportional to $\sqrt{n}$, with n the number of observations. A "good" variable, which "explains" the data, should therefore reduce the residual standard deviation by more than $\sqrt{n}$. Thus a good term to express the cost of k parameters is

$$\frac{k}{2}\log_2 n$$

The coding length interpretation of this term is that in the presence of Gaussian noise for the required precision each parameter needs to be encoded with $\frac{1}{2}\log_2 n$ bit. The complete description length $L(x|\theta)$ for a model with the parameter vector θ and the n observations x therefore is

$$L(x|\theta) = -\log_2 p(x|\theta) + \frac{k}{2}\log_2 n \tag{1}$$

This criterion has been derived by J. Rissanen [Ris78] and independently by G. Schwarz [Sch78], who derived it for the exponential family of distributions without referring to description length.

Despite its apparent simplicity, for large enough n the term $\frac{k}{2}\log_2 n$ can be proven to represent the asymptotically optimal model complexity [Ris89].

2.2.3 Minimum description length for multiple regression

For a finite number of observations, however, another approximation of the description length is needed. This criterion has been independently derived by C.S. Wallace [WF87] and J. Rissanen [Ris87]. Let the general linear model of multiple regression be of the form

$$\mu_i = \sum_{j=1}^{k} g_{ij} \cdot \theta_j \tag{2}$$

which is in matrix notation

$$\mu = G \cdot \theta \tag{3}$$

where G represents the set of regressor variables and θ is the parameter vector to be determined. The n measurements x_i are assumed to be independent samples from a normal distribution

$$p(x|\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{(x-\mu)^T(x-\mu)}{2\sigma^2}} \tag{4}$$

The least squares estimator θ^* of the parameter vector θ is determined by minimization of the sum of squared deviations S with respect to θ:

$$S = (x-\mu)^T(x-\mu) \tag{5}$$

$$\theta^* = (G^TG)^{-1}G^Tx \tag{6}$$

Therefore the optimized sum of squared deviations S^* is

$$S^* = x^Tx - \theta^{*T}G^TG\theta^* \tag{7}$$

and the probability at the least squares estimate θ^* is

$$p(x = G \cdot \theta^*) = \left(2\pi e \frac{S^*}{n}\right)^{-\frac{n}{2}} \tag{8}$$

According to both authors the generalized description length becomes

$$L(x) = \frac{n}{2}\log_2 2\pi e + \frac{n}{2}\log_2 \frac{S^*}{n} + \frac{1}{2}\log_2 |G^TG| \tag{9}$$

The first two terms are the negative log of the probability in equation 8. The last term, the log of the determinant of the curvature of the log-likelihood function at its maximum, represents the complexity of the parameters. The interpretation in terms of description length is, that the curvature of the negative log-likelihood function provides a yardstick by which the precision of the parameters is determined. The parameters are considered to be of optimal precision, when the error caused by truncation of the parameters is the same as the statistical error of the measurement.

It needs to be mentioned that this formulation of the description length is not invariant with respect to the scale of the data. It also becomes undefined, when the so called design matrix $G^T G$ becomes singular, which can easily happen in a local analysis of image properties. In the framework of "Stochastic Complexity" there is, however, a formulation of the description length, which is always defined and which is completely scale invariant [Ris89]. It is briefly described in Appendix D. From there it can be seen that equation 9 is an approximation to the exact description length.

3 A model based homogenity criterion

Here a criterion of homogenity is developed that is based on the minimal description length of data with respect to a model. It is shown to be a logical extension of euclidian distance used in gray value segmentation and the sum of squared deviations applied as a merging criterion in the context of hierarchical Cluster Analysis and traditional Split-and Merge algorithms.

In order to show the correspondence to existing models for gray value segmentation, the simple case of the ordinary gray value segmentation problem is analysed first with a model based view. Then the 1D displacement estimation with a locally constant vector field is discussed. It will be shown that this extends easily to 2D and to more general and complex models.

3.1 Homogenity for gray value segmentation

The model of the image I is a set of regions R_k with constant gray values within the regions, each region consisting of n_k pixels with the gray value x_i. With a Gaussian noise model the mean values μ_k and the variances σ_k^2 are a sufficient statistics to describe each region:

$$\begin{aligned} p(\{x_i\}|\mu_k) &= \prod_{i \epsilon R_k} \frac{1}{\sigma_k \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\frac{(x_i-\mu_k)^2}{\sigma_k^2}} \\ &= \frac{1}{\left(\sigma_k \cdot \sqrt{2\pi}\right)^{n_k}} \cdot e^{-\frac{1}{2}\sum_{i \epsilon R_k}\frac{(x_i-\mu_k)^2}{\sigma_k^2}} \end{aligned} \tag{10}$$

The parameters of interest, the estimated mean value μ_k^* and the sums of squared deviations S_k are determined:

$$\mu_k^* = \frac{1}{n_k} \cdot \sum_{i \epsilon R_k} x_i \tag{11}$$

$$S_k = \sum_{i \epsilon R_k} (x_i - \mu_k^*)^2 \tag{12}$$

$$= \sum_{i \epsilon R_k} x_i^2 - n_k \cdot \mu_k^{*2}$$

From the definition of S_k it is obvious that it grows monotonously with each additional x_j that is added to the region. Therefore it is possible to define a feature distance measure between a region R_k and an x_j as the increase of S_k when x_j is included in R_k. This generalizes to the distance between two regions R_k and R_l:

$$D(k,l) = S_{k \cup l} - S_k - S_l \tag{13}$$

where $S_{k \cup l}$ are the sum of squared deviations, when R_k and R_l are joined. Another view of this distance measure is the decrease of $S_{k \cup l}$ when a member or a subregion, say R_k, of region $R_{k \cup l}$ is taken out of that region. Trivially the sum of squared deviations of a single item is 0.

With the simple model discussed here, the S_k don't have to be calculated explicitly in order to determine the distance measure. As can be easily shown, the distance $D(k,l)$ between two regions R_k and R_l becomes

$$D(k,l) = \frac{n_k \cdot n_l}{n_k + n_l} \cdot (\mu_k^* - \mu_l^*)^2 \tag{14}$$

This relates the introduced distance measure to the traditional Euclidian distance between mean values [Pav82], the only difference being the weighting term, which depends on the size of the regions involved.

3.2 Homogenity of displacement vector fields

The homogenity measure is now generalized to stereo and motion models. The first model is a regionally constant displacement u_k and the 1D "images" are locally approximated with first order polynomials. The intuitive understanding of this model is that the intensity differences between the two images are "explained" by the displacement, when a locally planar image model is assumed. The expression to be minimized with respect to u_k is (See details in Appendix A)

$$S_k = \sum_{i \epsilon R_k} w_i (h_i - u_k \cdot g_i)^2 \tag{15}$$

with h_i the regularized intensity difference at location i and g_i the regularized estimate of the local gradient of the average of both images at location i. The w_i are the weights of the different locations within the support R_k. The weighted least squares estimate for u_k is

$$u_k^* = \frac{\sum_{i \epsilon R_k} w_i h_i g_i}{\sum_{i \epsilon R_k} w_i g_i^2} \tag{16}$$

Therefore the sum of squared deviations of region k at the optimal estimated value u_k^* becomes

$$\begin{aligned} S_k &= \sum_{i \epsilon R_k} w_i (h_i - u_k^* \cdot g_i)^2 \\ &= \sum_{i \epsilon R_k} w_i h_i^2 - \frac{(\sum_{i \epsilon R_k} w_i h_i g_i)^2}{\sum_{i \epsilon R_k} w_i g_i^2} \end{aligned} \tag{17}$$

With this residual sum of squares the same kind of distance measure $D(k,l)$ as in equation 13 can be generated.

Although the sum of squared deviations is formally suitable as a compatibility measure, it has no obvious interpretation nor any invariance properties. In particular it is a monotonously decreasing function of the number of parameters, and it vanishes when there are as many parameters as data points.

A better measure is the negative log-likelihood at the maximum likelihood estimate

$$M_k = n_k \log \left(\frac{S_k}{n_k} \right) \tag{18}$$

This has a probability interpretation for a given model, and in the context of the Neyman-Pearson test it has been successfully used for gray value region segmentation [Yak75].

The likelihood also does not take into account the complexity of a model, however. Therefore by choosing a sufficiently complex model the likelihood can be made as large as needed. Only the minimum description length makes it possible to compare models of different complexity in a consistent way.

As described in the last section, the part of the description length of a region that is not related to the geometry, is essentially

$$DL_k = \frac{n_k}{2} \cdot \log_2 \left(\frac{S_k}{n_k} \right) + \frac{1}{2} \cdot \log_2 |C_k| \tag{19}$$

where C_k is the design matrix of the statistical model. In the simple 1D flow model it is $\sum_{i \in R_k} g_i^2$.

The actual compatibility calculation is very simple. Suppose there are two neighbouring regions R_k and R_l. When an item or a region j that is not yet part of R_k is inserted into R_k, then the statistical part of the description length changes by the amount

$$\Delta DL(k,j) = DL_{k \cup j} - DL_k - DL_j \tag{20}$$

If region j is to be compared with different regions, then the compatibility measure can be simplified due to the constant contribution of region j

$$D(j,k) = DL_{k \cup j} - DL_k \tag{21}$$

This "pseudo-distance" (it is not symmetric and may well be negative) can now be converted into a relative assignment probability $Q(j,k)$ to a region k of a set K:

$$Q(j,k) = \frac{2^{-D(j,k)}}{\sum_{k' \in K} 2^{-D(j,k')}} \tag{22}$$

This obviously does not include the probability, that j may not belong to any of the possible groups.

4 Description of the complete scheme

4.1 The pyramidal architecture of the scheme

A key to the whole concept is a pyramidal architecture, which can represent both a family of regularized model estimates as well as a partition of the image with piecewise homogeneous regions and sharp discontinuities between them. This architecture has first been used for gray value segmentation and the applied algorithm has become known as pyramidal linking [BHR81, PR81, Bur84]. Although the pyramidal linking algorithm applied there has proven to be inadequate for displacement vector field segmentation, the underlying pyramidal architecture will be used to develop a new scheme.

The architecture of this pyramid can best be explained in one dimension. The extension to two (or more) dimensions can be done in a straightforward way by applying the 1D concept in each dimension. Fig. 1 shows the graph of this pyramid with 16 nodes at the bottom level, representing the primary image data.

The key property of this pyramid is that each node of a lower level (= son node) is connected to two nodes (4 in 2D) at the next higher level (= father nodes), except at the image boundary, where each son node is connected to only one father node. Each father node has links to 4 son nodes (16 in 2D), which define the maximal support for any estimation at the father node. The links are represented by weight factors $w(i,j)$ between 0 and 1, linking son node i to father node j.

Through these links information can be propagated in both directions. The bottom-up propagation of signal properties $P_i^{(m)}$ from pyramid level (m) to level $(m+1)$ is by weighted summation:

$$P_j^{(m+1)} = \sum_{i \in \mathrm{son}(j)} w(i,j) P_i^{(m)} \tag{23}$$

The top-down propagation from level $(m+1)$ to level (m) is analogous:

$$P_i^{(m)} = \sum_{j \in \mathrm{father}(i)} w(i,j) P_j^{(m+1)} \tag{24}$$

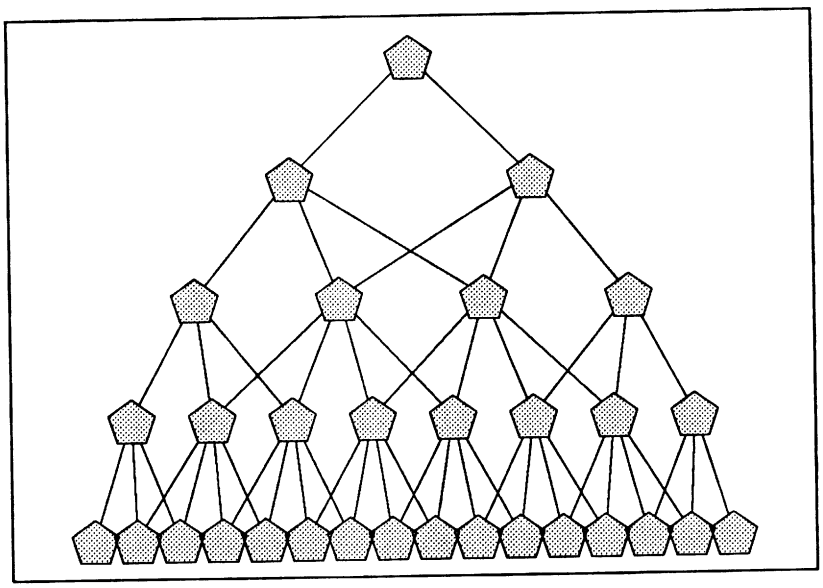

Figure 1: Architecture of the pyramid structure used in 1D

For normalization purposes, the sum of the weight factors is constrained

$$\sum_{i \in \text{son}(j)} w(i,j) = 1 \quad \forall j \tag{25}$$

This means that each son node contributes with the total weight of 1 to the next level of the pyramid. When links to father nodes are cut completely in order to define a region, then obviously all weights are 0. This aspect will only be dealt with in such way here, that a maximum pyramid level is chosen a priori. All nodes at this level are assumed to define regions.

Both these processes are in general space-variant linear operators, and they can be interpreted in such a way, that each father node represents a region of those son nodes with non-zero weight factors. In the general case these regions are overlapping.

With the necessary properties propagated up, a weighted least squares estimate can be obtained for the region corresponding to a father node. In the simplest case of a model with regionally constant gray values, the grayvalues themselves and the number of contributing pixels (one at the bottom level) are propagated up according to equation 23, recursively from level to level, up to the top level of the pyramid, which consists of only one node. At each node of this pyramid a weighted mean value can be estimated, which corresponds to the image region that has non-zero links to this node.

When each father node has the same set of binomially distributed weights, then a kind of Gaussian pyramid is created. When the weights are strictly 0 or 1, then the image is segmented into non-overlapping regions, and the weights define the

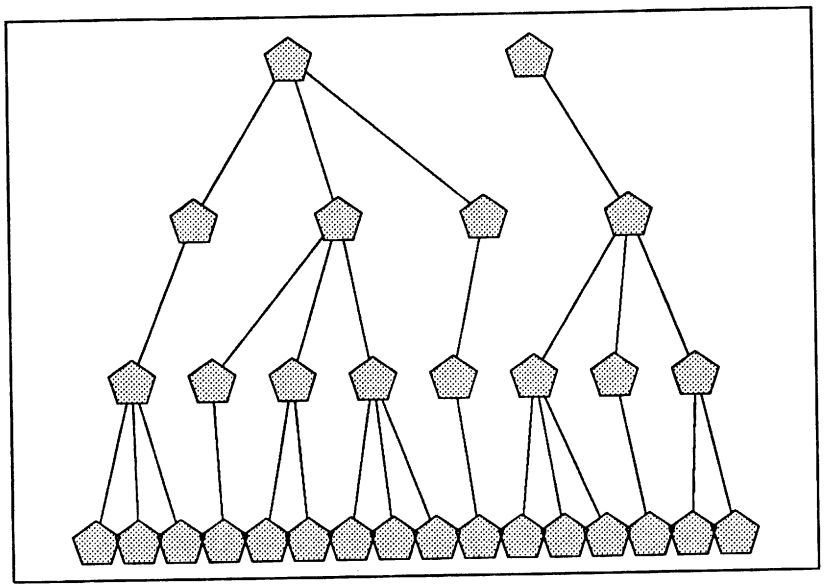

Figure 2: Graph of a complete segmentation in 1D

partition.

The top-down propagation process can best be understood when the image is fully segmented, so that each son node has exactly one father node, to which it is linked with a non-zero weight. The corresponding graph is a tree, indicated in Fig. 2.

In this case the top-down propagation is particularly simple. Starting from the root nodes of the tree, which represent the region properties, according to equation 24 each son node inherits these properties from its father node. Therefore the properties of the whole region are distributed to all contributing pixels, which is, for example, relevant to show the result of a segmentation.

The other extreme is the situation with those space-invariant weights, by which the Gaussian pyramid was created. Then the top-down propagation from a certain level of the pyramid effectively means a smoothing operation of the properties that were propagated up from the bottom level. The higher the referred level of the pyramid, the larger the effective filter size.

The segmentation task can be formulated within this pyramidal architecture. It is assumed that the image consists of reasonably large coherent regions. Within these regions the properties are assumed to be homogeneous. As a consequence the linking weights in the interior of regions can be assumed to be arbitrarily distributed as long as the corresponding father nodes also belong to the same region. In particular the weights can be binomially distributed with respect to a particular father node, which is the initial weight configuration of the scheme. The critical parts of the image are its discontinuities. Discontinuities are locations where the linking weights from neighboring son nodes should be non-zero only

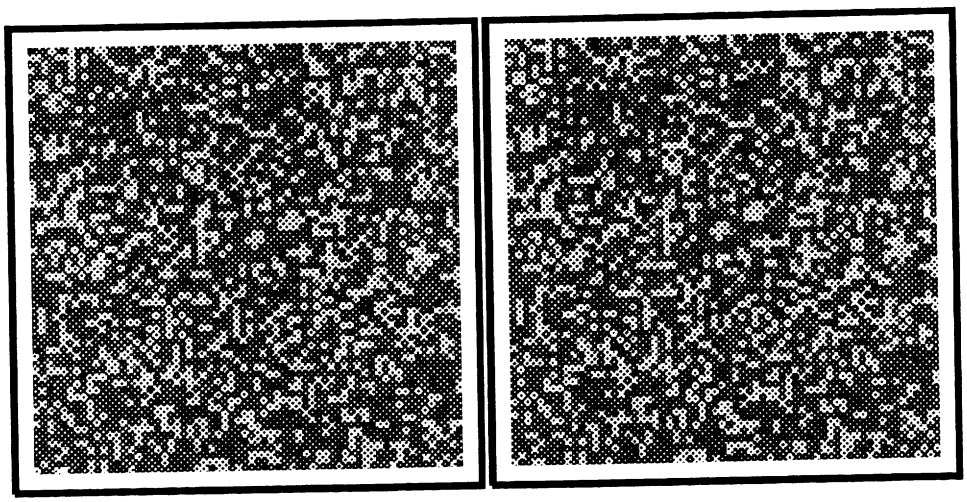

Figure 3: Random dot stereo pair

in the direction pointing away from the discontinuity towards the interior of the region. This applies even in the case when the region is represented by just one father node. On the other hand, if evidence can be found that the links from two neighboring son nodes should point in opposite directions, a local hypothesis of a discontinuity can be made. The segmentation task thus consists of gathering evidence about potential discontinuites by analysing the initial pyramid which has space-invariant binomially distributed linking weights, making this evidence consistent within and between the pyramid levels, and finally changing the weights in such a way that a segmented tree or at least something close to it results.

For the purpose of illustration an example of a random dot stereo pair will be used to show the different stages of the method. The original pair is displayed in Fig 3. In the center of the second image a square is displaced by 2 pixels. Both images also contain random noise.

4.2 Initialization

At the beginning of the process a family of regularized displacement estimates is constructed.

The basic procedure is demonstrated with the simple model of locally constant 1D displacements. This model is described above in section 3.2 (see also Appendix A). As with any least squares model, the estimator as well as the minimal sum of squared deviations can be calculated by summing the required moments over the region of interest and performing a few scalar operations. Given the set of moments at each pixel any degree of regularization can be achieved by smoothing all the moments before calculating the desired properties with scalar operations at each level of smoothness. Therefore by creating the Gaussian pyramid described above with all the moments a family of increasingly regularized vector fields is

available.

Taking the local model of constant 1D flow in an image with size $n^{(0)}$ pixels (for simplicity only one index is used here) from equations 16 and 17 it follows that for u_k^* and for S_k the following four moments are required:

$$\left(\sum_{i\epsilon R_k} 1\right), \left(\sum_{i\epsilon R_k} h_i^2\right), \left(\sum_{i\epsilon R_k} h_i g_i\right), \left(\sum_{i\epsilon R_k} g_i^2\right)$$

This means that the elements

$$P_i^{(0)} = \{1, h_i^2, h_i g_i, g_i^2\} \quad (i = 1 \ldots n^{(0)}) \tag{26}$$

form the elements of the bottom level of the pyramid.

At the initialization stage a family of fully regularized solutions to the displacement estimation problem is created. This means that the moments of the successive levels of the pyramid are formed by bottom-up propagation of the moments of their 16 son nodes at the level below (see equation 23). These are the 4 closest nodes in each dimension. The weights $w(i,j)$ define the support over which the least squares model is estimated at level j. The weighting factors are such that each son node distributes its information among its 4 father nodes according to equation 25. This guarantees that the total sum of contributing data elements is constant at each level of the pyramid. The initial weights of the 16 son nodes are binomially distributed in each dimension, which is the optimal 4×4-pixel approximation to a 2D Gaussian distribution. This optimizes the localization of the model properties at the coarser levels of the pyramid.

Fig 4 shows the family of regularized displacement fields. As there is only a one dimensional displacement, the scalar fields are gray level encoded. The different levels of the pyramid show displacement fields based on binomially weighted least squares with an increasing support.

4.3 Measurement of compatibility

Due to the way the pyramid is constructed it can be interpreted as a graph, where each son node is connected to 4 father nodes (in 2D) at the next higher level of the pyramid, which represent 4 directions. This representation has been used before for gray value and texture segmentation [BHR81, PR81, Bur84], but the algorithm described there is not adequate for the segmentation of displacement vector fields.

By measuring the compatibility to all 4 father nodes, there are three possible outcomes:

1. There is approximately the same compatibility to all father nodes, possibly with minor variations. This means that the son node is in the interior of a homogeneous region.

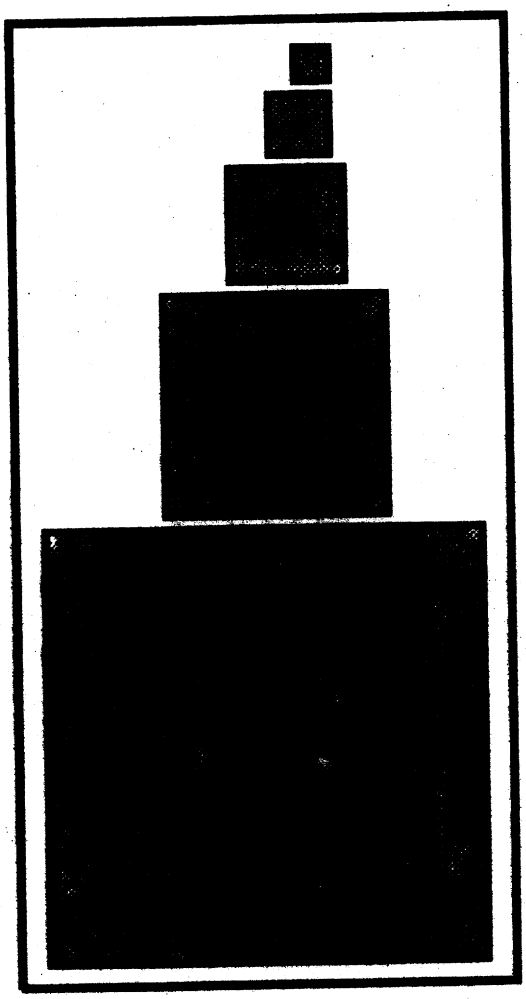

Figure 4: Family of regularized displacement estimates

2. There is a significant preference towards a certain direction. This is an indication that the son node is close to a displacement discontinuity, especially when the corresponding neighbor node shows a preference in the other direction. Significance in this context has the interpretation that the creation of a piece of occlusion boundary is "cheaper" in terms of coding length than the enforcement of the model across this boundary.

3. There is total incompatibility to any of the father nodes. This is an indication that at the pyramid level above there is too much smoothing with respect to this particular son node. Therefore this node (with all nodes "below" it) represents an independent region.

Although the last case is of conceptual interest and is important when an actual labelling of regions is performed, it can in many cases be ignored when the pyramid is cut at a level that contains more nodes than the expected number of distinct regions. This typically is a level of $4 - 12$ pixels square. At the "computationally expensive" levels of the pyramid only the first two alternatives are relevant.

The compatibility $D(i,j)$ betweeen a node i and its father node j is measured by the difference of description length between node j with the content of i included and node j with the content of node i removed. Assuming that i contributes with the weight $w(i,j)$ to j before the measurement, then in analogy to equation 21 $D(i,j)$ is given by

$$D(i,j) = DL(P_j^{(m+1)} + (1 - w(i,j))P_i^{(m)}) - DL(P_j^{(m+1)} - w(i,j)P_i^{(m)}) \quad (27)$$

The description length has an interpretation as the negative log of a probability, so $D(i,j)$ can be converted to a normalized assignment probability as in equation 22

$$Q(i,j) = \frac{2^{-\beta D(i,j)}}{\sum_{j'=\text{father}(i)} 2^{-\beta D(i,j')}} \quad (28)$$

The denominator makes sure, that the total assignment probability of i to the next level is exactly 1.

The introduction of β becomes necessary because it is desirable to have a similar range of assignment probabilities at each level. This is achieved, when β is the inverse of the number of pixels contributing to the father node j. The interpretation of this choice is that the "per pixel description length" is measured. As the population in the pyramid increases by a factor 4 with each level, the choice for β is 4^{-m}, where m is the pyramid level (The bottom level of the pyramid is 0).

4.4 Regularization of boundary evidence

After the measurement of compatibilities between the pyramid levels at each node a set of 4 probabilities is given, measuring the local preference of each node to-

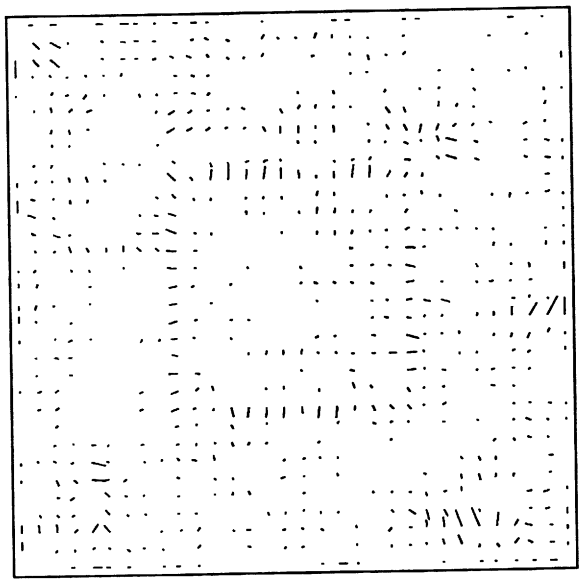

Figure 5: Measured assignment probability field at 2nd level

wards one of 4 directions. These 4 probabilities are related to the 4 directions NW,NE,SW,SE. They weigh the directional preference of each node in the pyramid. This quadruple of directional preference can be transformed into a vector field which directly shows the directional preference of a node and the strength of this preference. Associating the S-N direction with the y-axis, and the W-E direction with the x-axis, the vector field components (F_x, F_y) for directional preference are determined by

$$F_x = (Q_{NE} + Q_{SE}) - (Q_{NW} + Q_{SW}) \quad (29)$$

$$F_y = (Q_{NE} + Q_{NW}) - (Q_{SE} + Q_{SW}) \quad (30)$$

The vector field has only 2 components, whereas the quadruple representation has 3 (the normalization determines the 4th). This is due to the possibility that the quadruple could represent preferences that are inconsistent with a directional interpretation. These are, however not considered as relevant. Therefore it is possible to convert both representations into each other.

Fig. 5 shows this vector field at the second level of the pyramid. When the vector field is observed, it can be easily seen that it is very noisy in comparison to the ideal situation, where there should only be nonzero vectors on both sides of the displacement discontinuities.

The discontinuities appear as sinks in this vector field. Therefore a good operator to measure the strength of discontinuities is the divergence of this field. The lines of discontinuity are to be found where the negative valleys of the divergence of the directional preference field F are. The negative part of the divergence of the

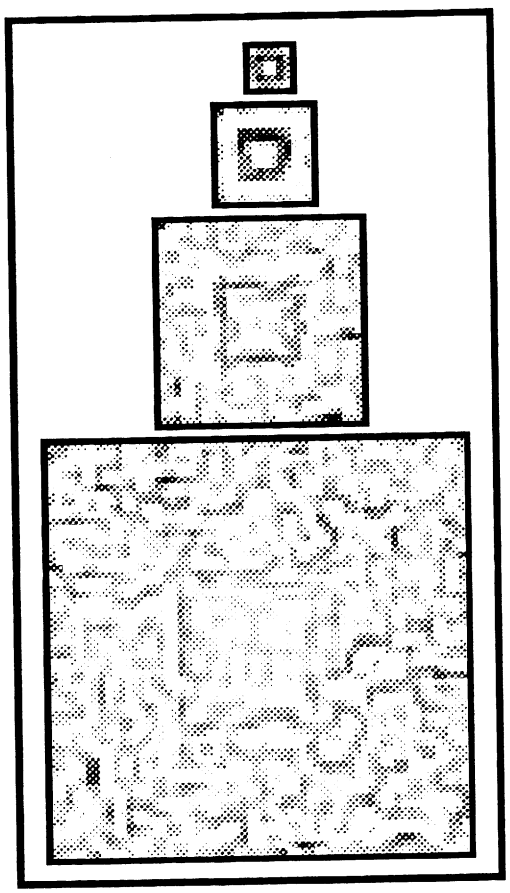

Figure 6: Measured boundary evidence at all levels

field at all levels is shown in Figure 6, indicating the measured boundary evidence. The sources of the field have no such meaningful interpretation, so for clarity the positive part of the divergence is set to 0.

As can be seen from this figure, at the lower levels of the pyramid the assignment probability vector field is very distorted. By sufficiently smoothing the vector field, however, the errors are averaged out, but the field preserves its qualitative essential properties. This can be seen at Fig 7, where the vector field has been smoothed with a 5×5 Gaussian filter. The different levels of the pyramid are smoothed differently. Obviously the lowest level has to be smoothed with the largest filter mask. The sizes of these masks have been determined heuristically.

Smoothing of the probability quadruples has the same effect. While the local directional contrast decreases the vector field becomes more coherent. The location of the sinks is also preserved, unless there is another sink nearby. In that case the two sinks will be attracted and with more smoothing they will merge into one.

This means that by smoothing the probability field, smooth and stable boundary evidence can be found. A boundary's location will be correct, if there is no other boundary close by. In other words, the maximal smoothing area should be

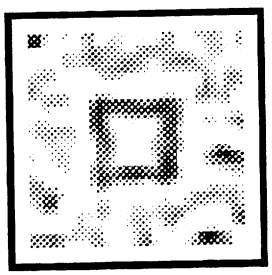

Figure 7: Boundary evidence from smoothed probability field

smaller than the smallest expected distance to another boundary.

4.5 The spatial coincidence constraint

An actual discontinuity should show up at the same location at multiple scales. This essentially is the spatial coincidence assumption formulated by David Marr [Mar82].

This constrains the possible discontinuities to those that coincide at multiple scales. In the proposed approach the spatial coincidence constraint is used to combine the evidence of several levels of the pyramid.

Although the smoothing at a single level improves the coherence of the probability field considerably, there are still undesirable fluctuations, especially at the highest level of resolution. This is a natural consequence of constructing local estimates from only few measured data points.

By summing over the probabilities of several pyramid levels, the fluctuations at locations of no discontinuity can be expected to average out. Where there are discontinuities, however, the probabilities will reinforce each other in their directional preference.

In order to relate the probabilities of different levels, the coarse level probability quadruples must be projected down one or more pyramid levels according to the formula, which is analogous to equation 24:

$$Q^{(m)}(i,j) = \sum_{l \in \mathrm{father}(i)} w(i,l) Q^{(m+1)}(l,j) \quad \forall i,j \tag{31}$$

This is optimally done with the binomially distributed weights $w(i,j)$ that were used to create the initial pyramid.

Fig. 8 shows the downprojected divergence of the field shown in Fig. 7. This shows that the downprojection effectively performs another smoothing step.

When the smoothed probability quadruples of the first three levels are added in the above-described way, the resulting field has the divergence shown at the bottom of the pyramid display in Fig. 9. The two intermediate levels are constructed by collecting evidence from the levels themselves and from one level above. The fourth level only uses its own evidence.

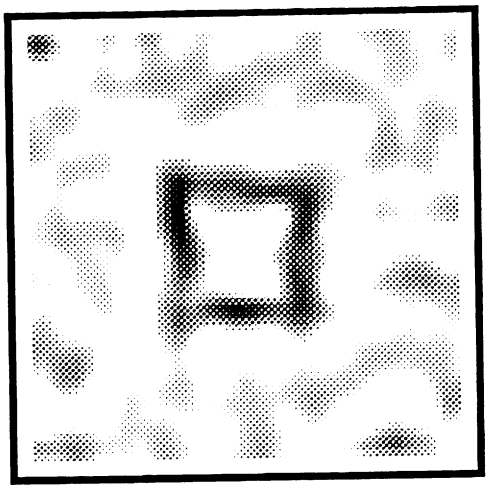

Figure 8: Boundary evidence from downprojected probability field

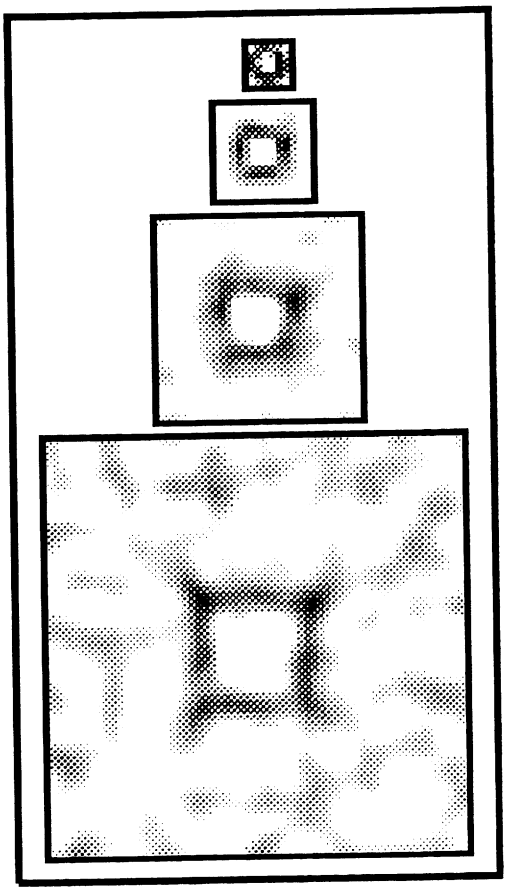

Figure 9: Final boundary evidence at all pyramid levels

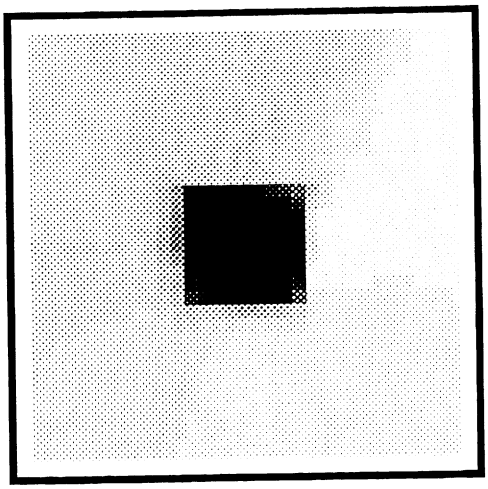

Figure 10: Final estimate of displacement field

4.6 Constructing the final model

As the probability quadruples express a directional preference, the coherent probability field determined by the regularization step and the spatial coincidence constraint can now serve as a new set of new weights that define the properties of each nodes father nodes.

Due to the smoothing the "contrast" of the probabilities has decreased, so that they can't be used as new linking weight factors directly. Their qualitative trend, however, is correct in the sense, that the directional preference goes to opposing directions at the hypothetical lines of discontinuity. The conversion of this trend to actual linking weight factors is done up to now in a very heuristic way. The final normalized probability quadrupels are taken to a high power, say 20, and the result is normalized again. This greatly enhances the "contrast" of the probabilities.

With these weights a new pyramid is created according to the bottom-up propagation scheme of equation 23, up to the pre-chosen top level of the pyramid. The moments of this level are then again propagated down to the bottom level. When the model is actually determined at the bottom level, the result can be visualized and used for further computations. Fig. 10 shows this final estimation of the random dot example. Although this is not yet a segmentation in the strict sense, it is obvious, that the displacements have a clear discontinuity around the black square.

4.7 A real example

Finally the method is tested on a real image sequence. The two frames used from this sequence are displayed at Fig. 11. One can see that the can is on a table. Although this is an example of 2D flow, the model of which is described in Appendix B, the vertical component of the displacement is so small that it can be

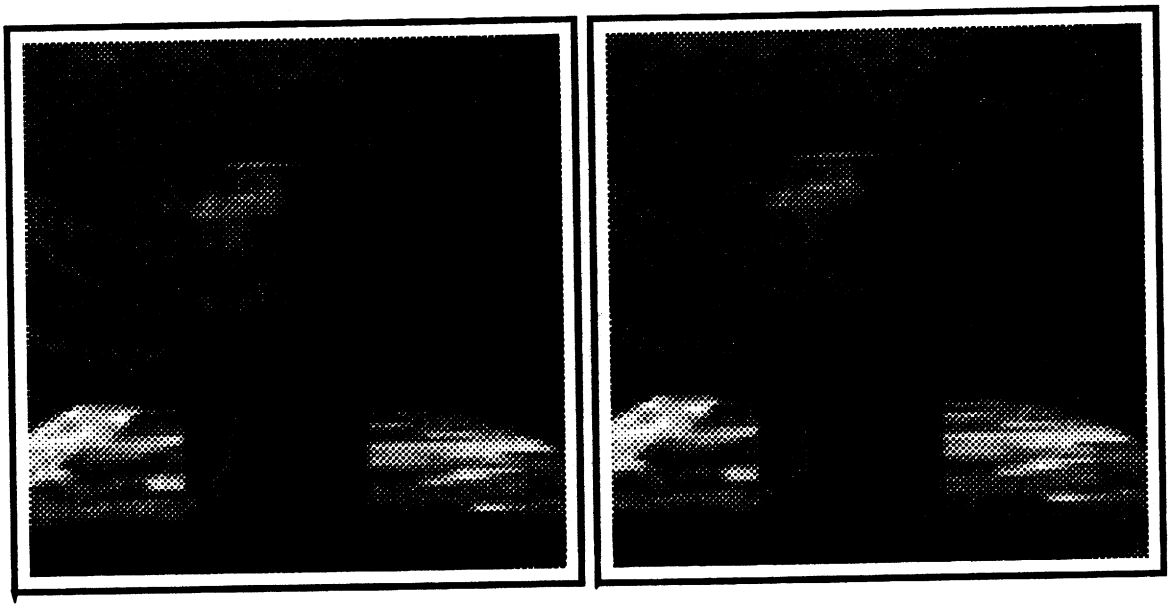

Figure 11: Pair from real image sequence

neglected. Therefore the horizontal component can be displayed the same way as in the example before. Obviously a model of regionally constant flow will not be able to capture the field of the table motion. The necessary extension is to use a planar model for the displacement as described in Appendix C. The result of Fig. 12, which is based on the regionally constant model, shows clearly, however, the discontinous motion of the can in comparison to the background, except at the top right corner of the can, where the contrast between object and background is so low, that a good estimate is just not possible.

5 Conclusions and discussion

A new approach has been presented for the fast determination of discontinuos displacement vector fields, which is closely related to the problem of finding a segmentation of the vector field. The last step of the actual segmentation has not yet been taken, this will be a natural follow-up project with its own specific problems.

The approach is based on regional models of the vector field, and it uses gray level data as primary input, either the original gray value images or appropriately filtered images.

The boundaries and the discontinous displacement vector field can be found within a single step, no iterations of any procedure are required. The elegance of the scheme is that boundaries are found instantaneously by demanding consistency of local model compatibility measures within a neighborhood and at multiple scales. In that way the "continuity of discontinuities" and the "spatial coincidence

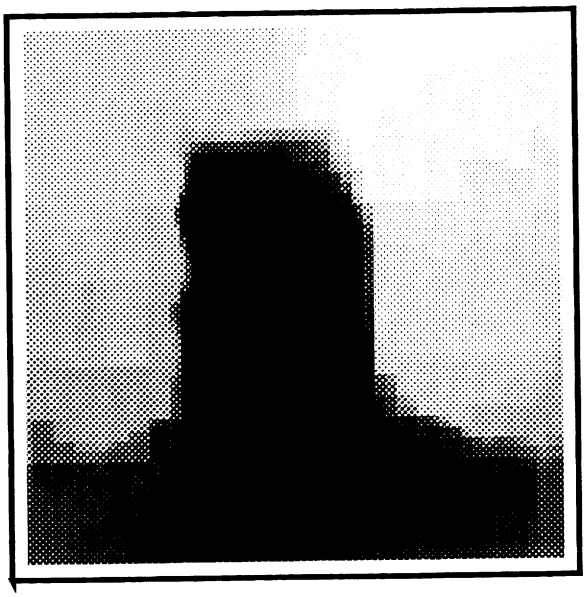

Figure 12: Resulting estimate of 1D displacement field

constraint" are made practical.

The current constraint on the boundary is simply its smoothness, This is similar to the chain-code boundary model as it is common in Markov Random Field models. Obviously there are more efficient boundary models such as polygons, quadratic curves or Bezier curves. Those will most probably result in a shorter code and will therefore allow more easily the segmentation of extremely thin objects, which will hardly be found by the current model.

At this point only very simple models have actually been tried. Although the description length criterion allows the comparison of different models, the current implementation will need to be extended to allow this comparison. In particular the planar and the second order models will have to be tested with real stereo and motion data. One aspect to consider is the increased storage requirement, when a more complex model is applied.

The method is not limited to segmentation based on displacement vector fields, but can be applied with any model that can be set up as a regional least squares problem. Therefore the same method also might be used, for example, for texture based segmentation.

The fact that the information about discontinuities is represented as a probability field allows a relatively easy integration of different vision modules. The same way as the probabilities from different scales, those from different vision modules can be integrated. There is no inherent problem of coupling constants, because the probabilities are all normalized.

Acknowledgements

The author gratefully acknowledges the fellowship grant from the Max-Kade Foundation, by which this work has been made possible. The stimulating discussions with Eric Grimson, Shimon Ullman, and in particular with Berthold Horn were very helpful in clarifying the relevant questions. Their comments on various versions of the paper are greatly appreciated . Thomas Breuel and Joachim Heel helped a great deal by providing the programs for visualization of graphics and images. Joachim Heel also provided the image sequence that was used for the final result. The company Dyadic Systems helped by giving a temporary free license for the Dyalog-APL interpreter during this project.

References

[Adi85] G. Adiv. Determining three–dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(4):384–401, 1985.

[Ana87] P. Anandan. A unified perspective on computational techniques for the measurement of visual motion. In *Proceedings Image Understanding Workshop*, pages 219–230, Los Angeles, CA, February 1987. Morgan Kaufmann, San Mateo, CA.

[BHR81] P.J. Burt, T.H. Hong, and A. Rosenfeld. Segmentation and estimation of image region properties through cooperative hierarchical computation. *IEEE Transactions on Systems, Man and Cybernetics*, 11:822–825, 1981.

[Bro81] C. Broit. *Optimal Registration of Deformed Images.* PhD thesis, University of Pennsylvania, 1981.

[Bur84] P. Burt. The pyramid as a structure for efficient computation. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis.* Springer, Berlin, 1984.

[BW73] D.M. Boulton and C.S. Wallace. An information measure for hierarchic classification. *The Computer Journal*, 16(3):254–261, August 1973.

[Cha66] G.J. Chaitin. On the length of programs for computing finite sequences. *J. Ass. Comp. Mach.*, 13:547–549, 1966.

[Den86] J. Dengler. Local motion estimation with the dynamic pyramid. In *Proceedings Int. Conf. on Pattern Recognition*, pages 1289–1292. Proceedings of the IEEE, 1986.

[DS88] J. Dengler and M. Schmidt. The dynamic pyramid - a model for motion analysis with controlled continuity. *Int. Journal for Pattern Recognition and Artificial Intelligence*, 2(2):275–286, 1988.

[Foe89] W. Förstner. Image analysis techniques for digital photogrammetry. In *Photogrammetrische Woche, Stuttgart*, 1989.

[Gri81] W.E.L. Grimson. *From Images to Surfaces.* MIT Press, Cambridge, Mass., 1981.

[GW84] M.P. Georgeff and C.S. Wallace. A general selection criterion for inductive inference. In *ECAI 84: Advances in Artificial Intelligence*, pages 473–482. Elsevier Science Publishers, 1984.

[GY89] M. Gennert and A. Yuille. Determining the optimal weights in multiple objective function optimization. In *Proceedings of the International Conference on Computer Vision*, pages 87–89, 1989.

[Hil84] E.C. Hildreth. The computation of the velocity field. *Proceedings of the Royal Society of London B*, 221:189–220, 1984.

[HR82] E.J. Hannan and J. Rissanen. Recursive estimation of arma order. *Biometrika*, 69:81–94, 1982.

[HS81] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[KMY86] C. Koch, J. Marroquin, and A. Yuille. Analog 'neuronal' networks in early vision. *Proceedings of the National Academy of Science*, 83:4263–4267, 1986.

[Kol65] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965.

[Lec89] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[Mar82] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W.H. Freeman and Company, San Francisco, 1982.

[NA89] R.C. Nelson and J. Aloimonos. Using flow field divergence for obstacle avoidance: Towards qualitative vision. In *Proceedings of the International Conference on Computer Vision*, pages 188–196, 1989.

[Nag83] H.-H. Nagel. Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics, and Image Processing*, 21(1):85–117, January 1983.

[NE86] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:565 593, 1986.

[NH85] S. Negahdaripour and B.K.P. Horn. Direct passive navigation. Technical report, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1985.

[Pav82] T. Pavlidis. *Algorithms for Graphics and Image Processing.* Computer Science Press, Rockville, MD, 1982.

[Pop59] K.R. Popper. *The Logic of Scientific Discovery.* Hutchinson, London, 1959.

[PR81] M. Pietikainen and A. Rosenfeld. Image segmentation by texture using pyramid node linking. *IEEE Transactions on Systems, Man and Cybernetics*, 11:822–825, 1981.

[Ris76] J. Rissanen. Parameter estimation by shortest description of data. In *Proc. JACE Conference RSME*, page 593, 1976.

[Ris78] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[Ris83] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Stat.*, 11:416–431, 1983.

[Ris86] J. Rissanen. Stochastic complexity and modeling. *The Annals of Statistics*, 14(3):1080–1100, 1986.

[Ris87] J. Rissanen. Stochastic complexity. *Journal of the Royal Statist. Society B*, 49(3):223–239 and 252–265, 1987.

[Ris89] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, Teaneck, NJ, 1989.

[Ris90] J. Rissanen. personal communication. 1990.

[Sch78] G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[SD89] M. Schmidt and J. Dengler. Adapting multi-grid methods to the class of elliptic differential equations appearing in the estimation of displacement vector fields. In V. Cantoni, R. Creutzburg, S. Levialdi, and H. Wolf, editors, *Recent Issues in Pattern Analysis and Recognition, Lecture Notes in Computer Science 399*, pages 266–274. Springer, Berlin-Heidelberg-New York-Tokio, 1989.

[Sol64] R. Solmonoff. A formal theory of inductive inference i,ii. *Inf. and Control*, 7:1–22,224–254, 1964.

[SU87] A. Spoerri and S. Ullman. The early detection of motion boundaries. A.I. Memo No. 935, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1987.

[Ter83] D. Terzopoulos. Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 24:52–96, 1983.

[TH84] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:13–27, 1984.

[TMB85] W.B. Thompson, K.M. Mutch, and V.A. Berzin. Dynamic occlusion analysis in optical fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):374–383, 1985.

[TWK87] D. Terzopoulos, A. Witkin, and M. Kass. Stereo matching as constrained optimization using scale continuation methods. In *SPIE Conference on Optical and Digital Pattern Recognition*, 1987.

[VP87] A. Verri and T. Poggio. Against quantitative optical flow. In *Proceedings of the International Conference on Computer Vision*, pages 171–180, London, England, June 1987. IEEE, Washington, DC.

[WB68] C.S. Wallace and D.M. Boulton. An information measure for classification. *Comput. J.*, 11(2):185–194, 1968.

[WB75] C.S. Wallace and D.M. Boulton. An invariant bayes method for point estimation. *Classification Soc. Bull.*, 3:11–34, 1975.

[WF87] C.S. Wallace and P.R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statist. Society B*, 49(3):240–265, 1987.

[WW85] A.M. Waxman and K. Wohn. Contour evolution, neighborhood deformation and global image flow: Planar surfaces in motion. *International Journal of Robotics Research*, 4(3):95–108, Fall 1985.

[Yak75] Y. Yakimovsky. Boundary and object detection in real world images. In *Proceedings IJCAI*, pages 695–704, 1975.

[Yui87] A.L. Yuille. Energy functions for early vision and analog networks. A.I. Memo No. 987, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, October 1987.

Appendix

A Estimation of locally constant 1D flow

The idea behind this model is that all intensity differences between image $I^{(1)}(x)$ and image $I^{(2)}(x)$ are caused by the constant displacement u. In order to minimize the approximation error, both images are treated in a symmetric way. Therefore the following integral has to be minimized with respect to u:

$$\int w(x)\left(I^{(1)}(x-\frac{u}{2})-I^{(2)}(x+\frac{u}{2})\right)^2 dx \tag{32}$$

The weighting function $w(x)$ defines the support of the functional and the weights of the individual measurements within this support. By a local linear expansion the $I^{(k)}(x)$ are approximated:

$$I^{(k)}(x+u) \approx I_0^{(k)}(x) + u \cdot I_x^{(k)}(x) \tag{33}$$

Now it is possible to restrict x to the grid coordinates i. With

$$h_i = I_0^{(1)}(i) - I_0^{(2)}(i)$$

and

$$g_i = \frac{1}{2}\left(I_x^{(1)}(i) + I_x^{(2)}(i)\right)$$

the integral to be minimized is transformed into a sum over the discrete support R and the weights w_i, the sum of squared deviations S_R:

$$S_R = \sum_{i\epsilon R} w_i \left(h_i - u \cdot g_i\right)^2 \tag{34}$$

This has the least squares solution

$$u^* = \frac{\sum_{i\epsilon R} w_i h_i g_i}{\sum_{i\epsilon R} w_i g_i^2} \tag{35}$$

and the minimal sum of squared deviations S_R^*

$$S_R^* = \sum_{i\in R} w_i h_i^2 - \frac{(\sum_{i\in R} w_i h_i g_i)^2}{\sum_{i\in R} w_i g_i^2} \tag{36}$$

$$= \sum_{i\in R} w_i h_i^2 - u^*(\sum_{i\in R} w_i g_i^2)u^* \tag{37}$$

B Estimation of locally constant 2D flow

The derivation is very similar to the 1D case, with the difference that now the displacement has the two components u and v and the gray value gradients have the components g_{xi} and g_{yi}. The weighted sum of squared deviations becomes

$$S_R = \sum_{i \epsilon R} w_i \left(h_i - u \cdot g_{xi} - v \cdot g_{yi}\right)^2 \tag{38}$$

with the least squares solution

$$\begin{pmatrix} u^* \\ v^* \end{pmatrix} = \begin{pmatrix} \sum_i w_i g_{xi}^2 & \sum_i w_i g_{xi} g_{yi} \\ \sum_i w_i g_{xi} g_{yi} & \sum_i w_i g_{yi}^2 \end{pmatrix} \begin{pmatrix} \sum_i w_i h_i g_{xi} \\ \sum_i w_i h_i g_{yi} \end{pmatrix} \tag{39}$$

and the minimal sum of squared deviations

$$S_R^* = \sum_i w_i h_i^2 - (u^* \;\; v^*) \begin{pmatrix} \sum_i w_i g_{xi}^2 & \sum_i w_i g_{xi} g_{yi} \\ \sum_i w_i g_{xi} g_{yi} & \sum_i w_i g_{yi}^2 \end{pmatrix} \begin{pmatrix} u^* \\ v^* \end{pmatrix} \tag{40}$$

C Estimation of planar 1D flow in a 2D image

This is the planar surface model relevant for stereo, when the images are given in normal coordinates and the epipolar line is parallel to the x-axis. The displacement model now depends on the image coordinates (x_i, y_i)

$$u(x_i, y_i) = u_0 + a x_i + b y_i \tag{41}$$

The weighted sum of squared deviations S_R becomes

$$S_R = \sum_i w_i \left(h_i - (u_0 + a x_i + b y_i) g_i\right)^2 \tag{42}$$

and the least squares estimates are

$$\begin{pmatrix} u_0 \\ a \\ b \end{pmatrix} = \begin{pmatrix} \sum_i w_i g_i^2 & \sum_i w_i g_i^2 x_i & \sum_i w_i g_i^2 y_i \\ \sum_i w_i g_i^2 x_i & \sum_i w_i g_i^2 x_i^2 & \sum_i w_i g_i^2 x_i y_i \\ \sum_i w_i g_i^2 y_i & \sum_i w_i g_i^2 x_i y_i & \sum_i w_i g_i^2 y_i^2 \end{pmatrix} \begin{pmatrix} \sum_i w_i h_i g_i \\ \sum_i w_i h_i g_i x_i \\ \sum_i w_i h_i g_i y_i \end{pmatrix} \tag{43}$$

The minimal sum of squared deviations is analogous to the previous cases.

D Stochastic Complexity and regression

Unfortunately the criterion described in equation 9 is not invariant with respect to scale change of the the data. If both the observations as well as the regressor variables are multiplied with a factor s, then the additional term

$$(k + n) \log_2 s$$

appears, although the parameters are identical in both cases. This can influence the decision, which set of regressor variables is optimal, which is certainly not desirable.

J. Rissanen has introduced the concept of "Stochastic Complexity", which eliminates all possible redundancy of a coding system [Ris87, Ris89]. For a model class M_k it is defined as

$$I(x|M_k) = -\log \int P(x|\theta) d\pi(\theta) \tag{44}$$

Essentially the difference to the minimal description length (MDL) is the integration over all possible "priors" or models instead of chosing the one, where the product under the integral is minimal. As the integral is larger than the maximum of its elements, naturally $I(x|M_k)$ is smaller than the minimal description length. Although asymptotically the Stochastic Complexity is identical with the criteria above, it is optimal for data sets of any size and has the required invariance properties.

The problem is, that the Stochastic Complexity can't be calculated analytically in many cases. Fortunately there is a closed form solution for the regression problem. It requires the parametrization of the the prior distribution for the variance σ in the form of the conjugate priors

$$\pi(\sigma) = \pi^{-\frac{1}{2}} \left(\frac{a}{2}\right)^{\frac{3}{2}} e^{-\frac{a}{2\sigma}} \sigma^{-\frac{3}{2}} \tag{45}$$

and for the parameters as a normal distribution $\pi(\theta_i)$ with zero mean and variance $\frac{\sigma}{c}$, with a and c as nuisance parameters. The Stochastic Complexity is optimized with respect to the nuisance parameters. The derivation is described in [Ris89, page 126-130], here just the result is reported. The design matrix $\Sigma = G^T G$ is extended to $\Sigma_c = cI + G^T G$, which prevents it from becoming singular. Then the minimum sum of squared deviations also depends on c:

$$S_c = x^T x - x^T G(cI + G^T G)^{-1} G^T x \tag{46}$$

The complete Stochastic Complexity is

$$\begin{aligned} I(x|g,c) &= \frac{n}{2}\log_2 S_c + \frac{1}{2}\log_2 |\Sigma_c| \\ &\quad -\frac{k}{2}\log_2 c + \frac{n+1}{2}\log_2 \frac{\pi(n+1)}{n} - \frac{1}{2}\log_2 \Gamma(\frac{n+1}{2}) \end{aligned} \tag{47}$$

The minimization with respect to c has to be done numerically with the iteration:

$$c = \frac{S_c}{S_c tr\Sigma_c + (n+1)x^T G\Sigma_c^{-2} G^T x} \tag{48}$$

It has been proven [Ris90], that the Stochastic Complexity found this way is invariant with respect to scale changes of regressor variables and observations.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AIM 1265	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A235512
4. TITLE *(and Subtitle)* Estimation of Discontinuous Displacement Vector Fields with the Minimum Description Length Criterion		5. TYPE OF REPORT & PERIOD COVERED memorandum 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Joachim Dengler		8. CONTRACT OR GRANT NUMBER*(s)* N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE October 1990 13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* Office of Naval Research Information Systems Arlington, Virginia 22217		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

correspondence problem
pyramids
motion stereo
discontinuities

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A new noniterative approach to determine displacement vector fields with discontinuities is described.

In order to overcome the limitations of current methods, the problem is regarded as a general modelling problem. From this point of view the imaging field consists of a set of regions with common prop-

(continued on back)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Block 20 continued:

erties and their boundaries. The strategy proposed is an analysis of consistency of the displacement estimators between different levels of regularization.

A fully regularized family of model based displacement vector fields is constructed by successive smoothing and subsampling. By measuring the difference in description length the compatibility between different levels of regularization is measured. This gives local but noisy evidence of possible model boundaries at multiple scales.

With the two constraints of continous lines of discontinuities and the spatial coincidence assumption consistent boundary evidence is found. Based on this combined evidence the model pyramid is updated, now describing homogeneous regions with sharp discontinuities.

CS-TR Scanning Project
Document Control Form

Date : 10/06/94

Report # AIM-1265

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 34
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form 2(PGS) ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number): 16, 18, 24, 25, 26

Other (note description/page number):

Description : Page Number:

_____
_____
_____
_____

Scanning Agent Signoff:

Date Received: 10/06/94 Date Scanned: 10/13/94 Date Returned: 10/13/94

Scanning Agent Signature: _____