

**Multiple Resolution Representation
and Probabilistic Matching of
2-D Gray-Scale Shape**

James L. Crowley and Arthur C. Sanderson

CMU-RI-TR-84-29

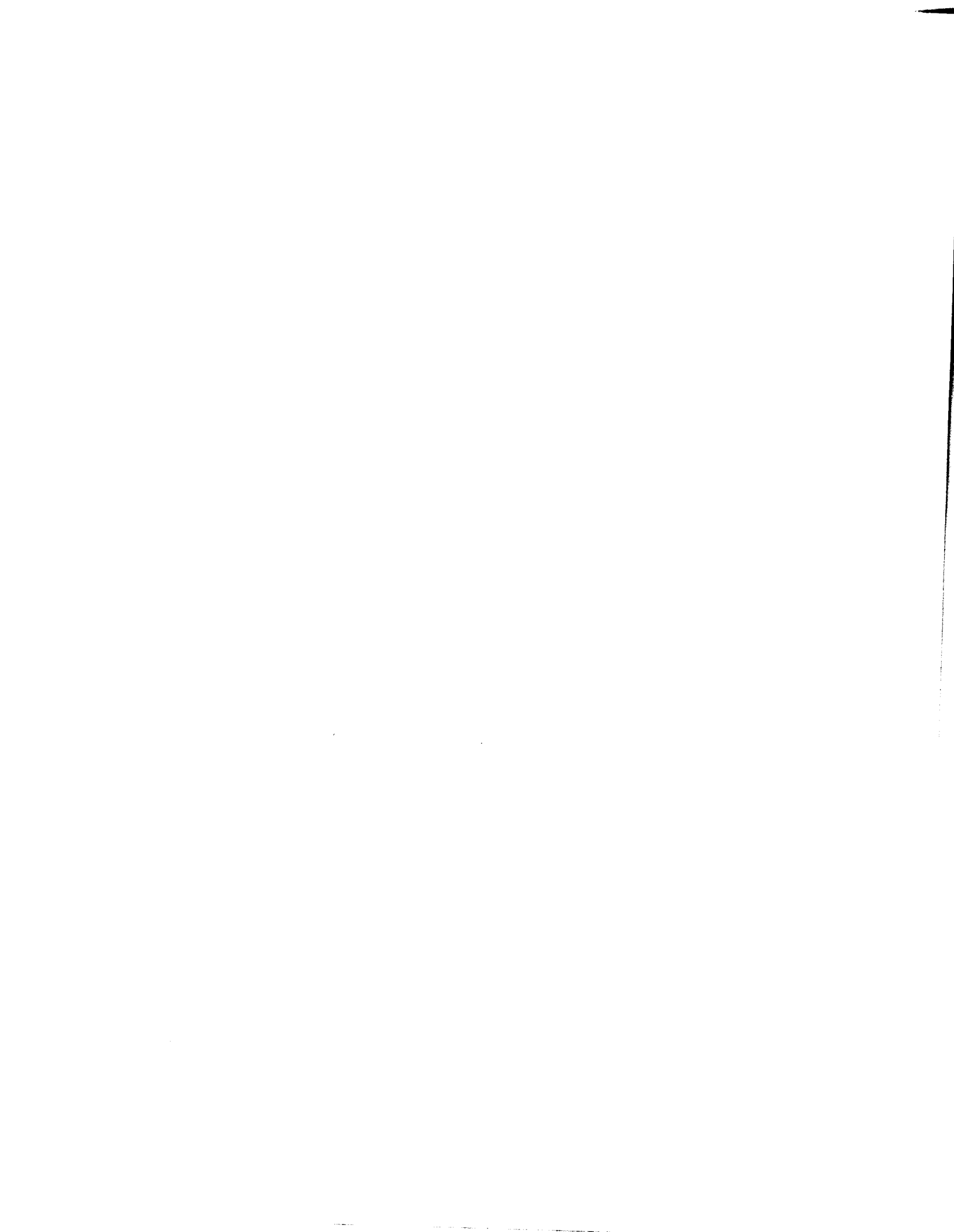
The Robotics Institute
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

December 1984

Copyright © 1985 Carnegie-Mellon University

This research was funded in part by the Westinghouse Electric Corporation, by the Air Force Office of Scientific Research, Contract No. F49620-83-C-0100, and The Robotics Institute.





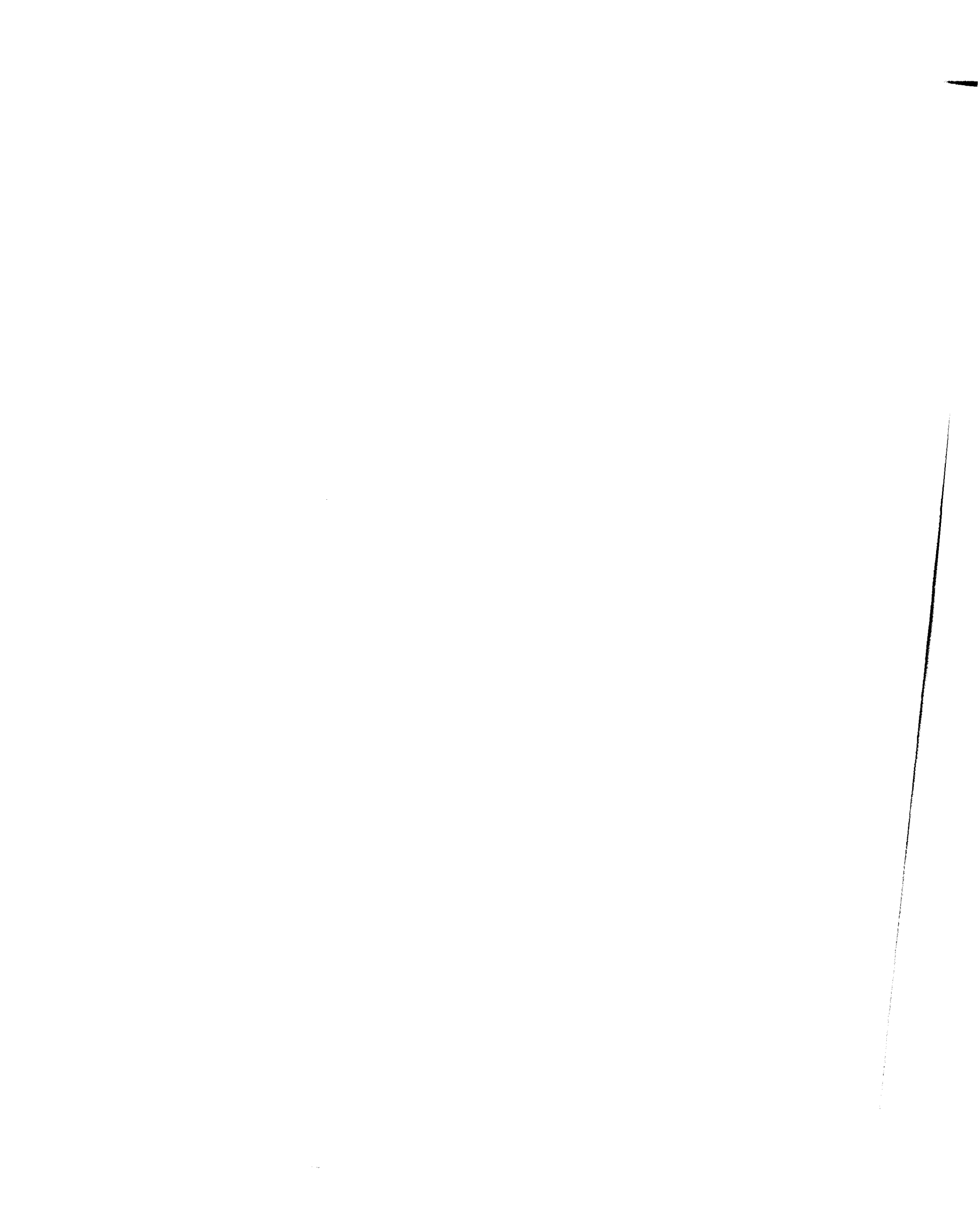


Table of Contents

1	Introduction	1
1.1	The Structural Pattern Recognition Problem	1
1.2	Summary of Solution	1
1.3	The Multiple Resolution Representation of Patterns	2
2	The Difference of Low-Pass Transform	2
2.1	Definition of the DOLP Transform	3
2.2	Fast Computation Techniques: Resampling and Cascade Convolution	4
2.3	An Example of a DOLP Transform	4
3	A Multiple Resolution Representation for Gray Scale Patterns	5
3.1	Definition and the Symbols Set	5
3.2	Peak Paths	8
4	Probabilistic Models of Gray Scale Patterns	11
4.1	Definition of Model Symbols	11
4.2	Peak Path Labels in a Model	12
4.3	Attribute Probabilities and Likelihoods	12
4.4	Probabilistic Attributes in the Model	13
4.5	The No-Match Likelihood	14
4.6	Model Transformation	14
5	Multiple Resolution Probabilistic Matching	14
5.1	Finding the Pose	15
5.2	Matching Top Peak Paths	16
5.3	Determining the Pose by Pairwise Search	18
5.4	Matching Children Peak Paths	18
5.5	Similarity Measure	19
6	Training	19
6.1	The Training Algorithm	20
6.2	The Training Program	20
6.3	Correspondence Matching in Training	21
7	Conclusions	23
8	Acknowledgements	24



List of Figures

Figure 1:	A sampled DOLP transform of an image with two bolts	4
Figure 2:	An enlargement of band-pass levels 5 through 11	5
Figure 3:	Subset relations of node types	7
Figure 4:	Circle representation of the negative peaks from levels 11 through 4 of the larger bolt from figure 1. Darker circles represent M-nodes, lighter circles represent P-nodes.	7
Figure 5:	The negative peaks from levels of 11 through 4 which describe the larger bolt from figure 1	8
Figure 6:	A circle representation of the M-nodes of the Peak Paths for the larger bolt in figure 1. The scale of the M-Node is represented by the radius of the circle. The radius is chosen to approximate the radius of the inner positive lobe of the Difference of Gaussian filter. The position of the M-node is represented by the position of the circle.	10
Figure 7:	The M-node attributes of the lowest resolution negative Peak-Paths in the larger bolt shown in figure 5	10
Figure 8:	The interactive training program. The upper left window contains the model image (currently the first training sample). The upper right window contains the current training sample. The circles indicate the location and scale of the principal Peak Path. The lower window is used for text messages.	21
Figure 9:	The interactive training program. The pose orientation is determined from the correspondence of the "second" Peak Path, displayed graphically for approval.	22
Figure 10:	The interactive training program. The correspondence of top level Peak Paths are displayed graphically for approval.	23

Abstract

One approach to pattern classification is to match a structural description of a pattern to models which describe the structural properties of pattern classes. The central problem in structural pattern matching is to determine the correspondence between the symbols which comprise a model and symbols which describe a pattern. The difficulty of determining this correspondence depends critically on the representation that is used to describe patterns.

This paper presents a probabilistic representation for structural models of pattern classes. Both pattern descriptions and models for pattern classes are based on symbols which represent gray-scale information at multiple resolutions. A pattern description is given by a tree of symbols with attribute values. Structural models are represented by a tree of symbols with probabilistic attributes. The position and scale (resolution) of the symbols, as well as other "features," are represented by these attributes.

An algorithm is presented for determining the correspondence between symbols in a description of a pattern and symbols in a model of a pattern class. This algorithm uses the connectivity between symbols at different scales to constrain the search for correspondence. An interactive training program for learning models of pattern classes is described, and some conclusions from the work are presented.

1 Introduction

This paper describes a system for classifying two dimensional gray scale patterns which is based on structural pattern matching using a multiple resolution representation. The mathematical basis for this representation is the "Difference of Low Pass" (DOLP) transform [3]. The DOLP transform is defined, followed by the definition of a symbolic structure which is used to represent patterns. A matching algorithm is described which exploits the multiple resolution structure to efficiently match a model for a pattern class to an observed pattern. A training algorithm for deriving pattern models from observation sets is also presented.

1.1 The Structural Pattern Recognition Problem

The purpose of a pattern recognition system is to label a pattern as an instance of a predefined class. One approach to pattern classification is structural pattern matching. In this approach, a structural description of a pattern is matched to a set of models which describe predefined pattern classes. A similarity measure is computed for the match for each pattern class. The pattern is then classified as an instance of the class for which the similarity measure is maximized.

The most important aspects in the design of a structural pattern matching system are

- *the representation* which is used for the pattern and models,
- *the matching algorithm*, and
- *the similarity measure*.

This paper describes the use of a multiple resolution representation for pattern descriptions and object models. It then presents a matching algorithm which uses the results of matching at low resolution to guide and simplify the matching at higher resolutions. A similarity measure is presented based on the logarithm of the probability that each symbol in the description is an instance of the corresponding symbol in the model.

1.2 Summary of Solution

A structural description for patterns typically has the form of a network of symbols. In the system describes below each symbol has an associated set of attributes. We call such a description an "attributed graph" [12]. This paper describes an attributed graph representation for image patterns in which the symbols are derived from peaks and ridges in the Difference of Low Pass (DOLP) transform [3].

A structural model of a pattern represents the ensemble of descriptions of patterns in the training set. Thus, while a model has the same form as a description, symbols in the model are assigned a probability of occurrence and the symbol attributes are described by probability distributions. In its general form, such a pattern model is a "probabilistic graph model."¹ In this paper we describe techniques for learning and representing probabilistic graph pattern models from examples of multiple resolution descriptions of patterns. These technique are restricted to a subset of the multiple resolution representation described in [3], given by local peaks. This subset has the form of a multiple resolution tree. Thus, in this paper we will describe a structural modeling technique based on a "Probabilistic Tree Model."

¹Also called a "Random Graph Model."

Matching a pattern description to a model is a problem of determining the most likely correspondence between symbols in the pattern description and symbols in each model. One of the properties of a multiple resolution representation is that it permits the matching process to be decomposed into a sequence of very small matching problems. That is, matching may begin with a small number of low resolution symbols. The connectivity of symbols at adjacent resolutions permits the correspondence at low resolutions to be used to constrain the possible correspondences at higher resolutions. This paper describes such a matching algorithm.

The most likely correspondence provides the basis for a similarity measure. In this paper we describe a similarity measure based on the log likelihood that the most likely corresponding symbol is an instance of the model symbol. The sum of the log likelihoods is used as a global similarity measure for the match between the description and the model. The pattern is assigned the label of the class for which the match produces the largest similarity measure above a threshold.

1.3 The Multiple Resolution Representation of Patterns

Earlier papers [3] have described a representation for shape based on peaks and ridges in the DOLP transform. A description of a pattern expressed in this representation was shown to retain its structure despite changes in size or orientation, and such a description was shown to degrade gracefully when boundaries are blurred and when the image is corrupted by various forms of image noise [2]. Experiments have been performed in matching motion stereo images using the DOLP transform. An algorithm for matching rows from co-planar stereo images has recently been described using a one dimensional form of DOLP transform [4]. A fast computation technique for the DOLP transform has also been defined [5].

The representation presented in this paper is based on connected sequences of peaks called "Peak Paths" and on connected sequences of ridges called "ridge paths." This paper described a matching algorithm that only makes use of Peak Paths. Extending this algorithm to use ridge paths is believed to be straight forward. Each such symbol has a set of attributes associated with it. These symbols are connected through the resolution levels by connectivity relations which give it the form of a tree. Using these higher level symbols further decreases the complexity of the pattern description, and thus simplifies the matching problem.

2 The Difference of Low-Pass Transform

This representation is based on a reversible transform referred to as the "Difference of Low-Pass" (DOLP) transform [2]. The DOLP transform is a reversible transform which converts an image (or signal) into a set of band-pass images (or signals). Each band-pass image is equivalent to a convolution of the original image with a band-pass filter, b_k . Each band-pass filter is formed by a difference of two size scaled copies of a low-pass filter, g_{k-1} and g_k .

$$b_k = g_{k-1} - g_k$$

Each low-pass filter g_k is a copy of the low pass filter g_{k-1} scaled larger in size. These band-pass images comprise a three space (the DOLP space) in which the third dimension is scale (or resolution).

The DOLP transform expresses the image information at a discrete set of resolutions in a manner which preserves all of the image information. This transform separates local forms from more global forms in a manner that makes no assumptions about the scales at which significant information occurs. The DOLP filters overlap in the frequency domain; thus there is a smooth variation from each band-pass level to the next. This

"smoothness" makes size-independent matching of forms possible and makes it possible to use the symbols from one band-pass level to constrain the correspondence of symbols at the next (higher resolution) level.

2.1 Definition of the DOLP Transform

The DOLP transform expands an $N = M \times M$ image signal $p(x,y)$ into K band-pass images, $\mathfrak{B}_k(x,y)$. Each band-pass image is equivalent to a convolution of the image $p(x,y)$ with a band-pass impulse response $b_k(x,y)$.

$$\mathfrak{B}_k(x,y) = p(x,y) * b_k(x,y) \quad (1)$$

For $k=0$, the band-pass filter is formed by subtracting a circularly symmetric low-pass filter $g_0(x,y)$ from a unit sample positioned over the center coefficient at the point $(0,0)$.

$$b_0(x,y) = \delta(x,y) - g_0(x,y) \quad (2)$$

The filter $b_0(x,y)$ gives a high-pass image, $\mathfrak{B}_0(x,y)$. This image is equivalent to the result produced by the edge detection technique known as "unsharp masking" [10].

$$\begin{aligned} \mathfrak{B}_0(x,y) &= p(x,y) * (\delta(x,y) - g_0(x,y)) \\ &= p(x,y) - (p(x,y) * g_0(x,y)) \end{aligned} \quad (3)$$

For band-pass levels $1 \leq k < K$ the band-pass filter is formed as a difference of two size-scaled copies of the low-pass filter.

$$b_k(x,y) = g_{k-1}(x,y) - g_k(x,y) \quad (4)$$

In order for the configuration of peaks in a DOLP transform of a form to be invariant to the size and orientation of a form, it is necessary that each low-pass filter, $g_k(x,y)$ be a copy of the circularly symmetric low-pass filter $g_0(x,y)$ scaled larger in size by a scale factor raised to the k^{th} power. [2]. Thus for each k , the band-pass impulse response, $b_k(x,y)$, is a size scaled copy of the band-pass impulse response, $b_{k-1}(x,y)$. For two-dimensional circularly-symmetric filters which are defined by sampling a continuous function, size scaling increases the density of sample points over a fixed domain of the function. The change in scale between filter $g_k(x,y)$ and filter $g_{k+1}(x,y)$ is denoted by the the scale factor, denoted S_s . For Gaussian filters, S_s is the ratio of the standard deviations for the k^{th} and the $(k+1)^{\text{th}}$ filters.

$$S_s = \sigma_{k+1} / \sigma_k$$

It is possible to define a DOLP transform with any scale factor for which the difference of low-pass filter provides a useful pass band. Marr, for example, argues that a scale factor of $S_s = 1.6$ is optimum for a difference of Gaussian filters [9]. We have found that a scale factor $S_s = \sqrt{2}$ yields effectively the same band-pass filter and provides two other interesting properties [2].

In principle the DOLP transform can be defined for any number of band-pass levels K . A convenient value of K is

$$K = \text{Log}_{S_s}(N) - R_0 \quad (5)$$

Where R_0 is the radius of $g_0(x,y)$. With a scale factor of S_s , $\text{Log}_{S_s}(N)$ is the level at which the original distance between filter samples becomes equal to the width of the image. Subtracting R_0 gives the level at which the filter $g_k(x,y)$ becomes larger than the image.

The DOLP transform is reversible which proves that no information is lost. The original image may be recovered by adding all of the band-pass images, plus a low-pass residue. This low pass residue, which has not been found to be useful for describing the image, is the convolution of the lowest frequency (largest) low-pass filter, $g_K(x,y)$ with the image.

$$p(x,y) = (p(x,y) * g_K(x,y)) + \sum_{k=0}^{K-1} \mathfrak{B}_k(x,y) \quad (6)$$

2.2 Fast Computation Techniques: Resampling and Cascade Convolution

A full DOLP transform of an image composed of N samples, produces $K = \log_5(N)$ band-pass images of N samples each, and requires $O(N^2)$ multiplies and additions. Two techniques can be used to reduce the computational complexity of the DOLP transform: "resampling" and "cascaded convolution with expansion" [3]. Combining these two techniques gives an algorithm which will compute a DOLP transform of an N sample signal in $O(N)$ multiplies, producing $3N$ sample points.

2.3 An Example of a DOLP Transform

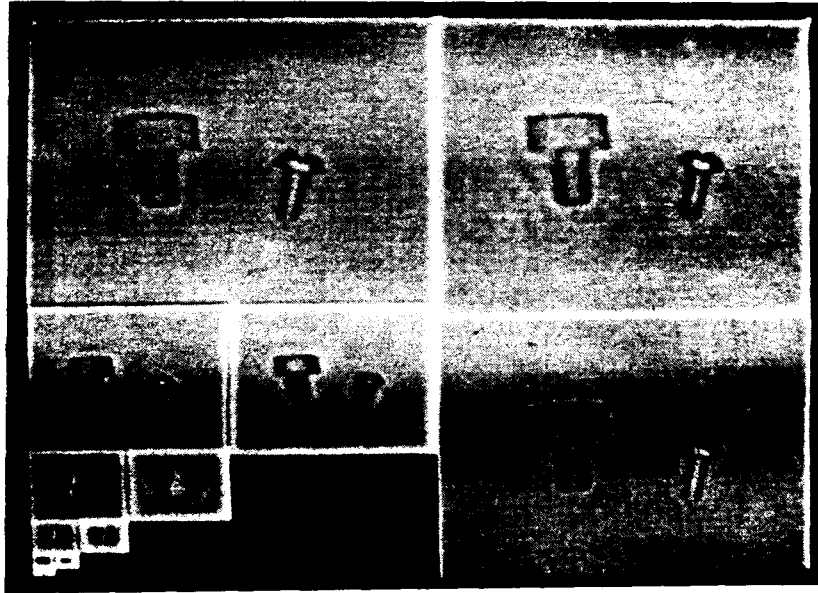


Figure 1: A sampled DOLP transform of an image with two bolts

Figure 1 shows a DOLP transform of an image of two bolts that was produced using the fast computation techniques described above. In this figure the image at the lower right is the high frequency image, $\mathfrak{B}_0(x,y)$. The upper left corner shows the level 1 band-pass image, $\mathfrak{B}_1(x,y)$, while the upper right hand corner contains the level 2 band-pass image, $\mathfrak{B}_2(x,y)$. Underneath the level 1 band pass image are levels 3 and 4, then 5 and 6, etc.

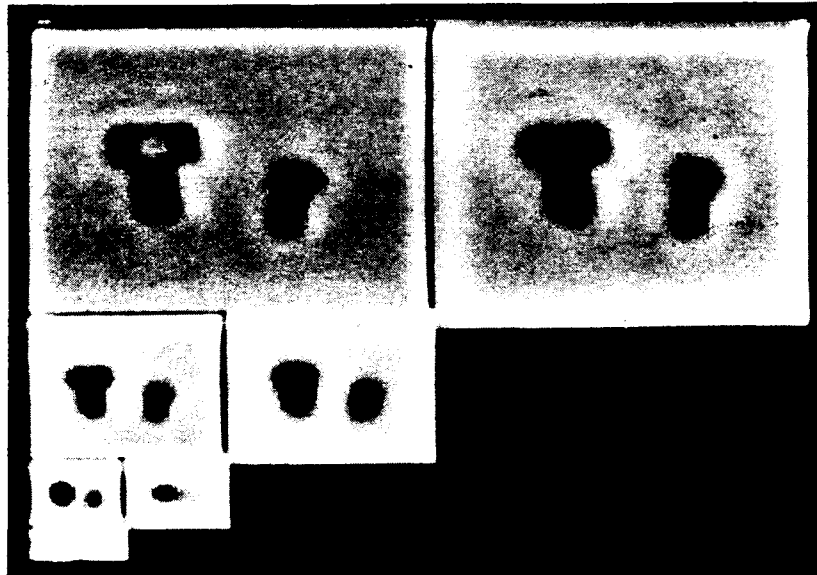


Figure 2: An enlargement of band-pass levels 5 through 11

Figure 2 shows an enlarged view of band-pass levels 5 through 13. This enlargement illustrates the unique peaks in the low frequency images that occur for each gray-scale form. Examples will be given below using the symbolic description which was produced from this image.

3 A Multiple Resolution Representation for Gray Scale Patterns

The amount of information required to represent the structure of a pattern may be greatly reduced by encoding a symbolic description of the DOLP band-pass images. We have developed two levels of symbolic representation for patterns based on the peaks and ridges from the DOLP transform. The first level is composed of symbols derived directly from the individual DOLP samples. These are DOLP samples which are found to be local positive maxima or negative minima in one, two or three dimensions within a DOLP band-pass image. The second level symbols exploit a connectivity between peak and ridge samples which is inherent in the DOLP transform. Sequences of connected peaks and ridges are grouped to form symbols called "Peak Paths" and "Ridge Paths." These two levels of symbols are described below.

3.1 Definition and the Symbols Set

The "local neighborhood" of a DOLP sample includes the nearest eight neighbors on the sample grid at its band-pass level. A "peak" (or P-node) is a local positive maximum or negative minimum within a two-dimensional band-pass image. A "ridge-node" (or R-node) is a local one-dimensional positive maximum or negative minimum within a two-dimensional band-pass image.

The DOLP shape representation is based on four types of symbols. Two of these are detected in each DOLP Band-Pass image. These are:

- R-nodes: DOLP samples which are on a ridge (1-D maxima or negative minima within a band-pass image).
- P-nodes: DOLP samples which are local 2-D maxima or negative minima within a band-pass image. These are R-nodes which are a maxima in every direction.

The other two symbols are detected within the three dimensional space (x, y, k) defined by the DOLP transform.

- M-nodes: P-Nodes which have a DOLP sample of larger magnitude than P-nodes at adjacent positions in adjacent band-pass images.
- L-nodes: DOLP samples which are on a ridge across levels (i.e. in the three space (x,y,k)). These are R-nodes which are larger than their neighbors at adjacent band-pass levels above and below in the DOLP space.

The symbol set provides the basis for structural representation of gray-level shapes in images. To complete the representation, each symbol, P, M, L, R, retains four attributes, (x,y,k,d) :

- (x,y) the image coordinates of the symbol;
- k the band-pass level of the symbol;
- d the intensity of the DOLP image at the sample.

These attributes attach quantitative information to each symbol. The attributes are used to improve the efficiency of matching by constraining possible structural correspondences, as well as improvement in matching accuracy and reliability.

Let us define the the first level symbols as a list of nodes $\{v_i\}$, which are connected between levels by a set of links, $\{e_{ij}\}$. A node consists of a type from the set $\{P, R, M, L\}$ and a set of attributes $\{x, y, k, d\}$.

A link, e_{ij} is made between node v_i at level k and node v_j at level $k-1$ if and only if

1. v_i and v_j have DOLP values of the same sign, and
2. v_i and v_j are at adjacent band-pass levels and are within a predefined distance of each other. The neighborhood size is approximately equal to the inner positive lobe of the larger DOLP filter.

We have implemented peak linking with an algorithm that starts at the highest resolution level (1), and steps up through the band-pass images to the lowest resolution level (K-1). For each peak at level k , a search is made over a 5 by 5 sample region in level $k+1$. The peak at level k is linked to the nearest peak at level $k+1$ within this region.

The node types of R and P are assigned based on the value of a DOLP sample relative to its local

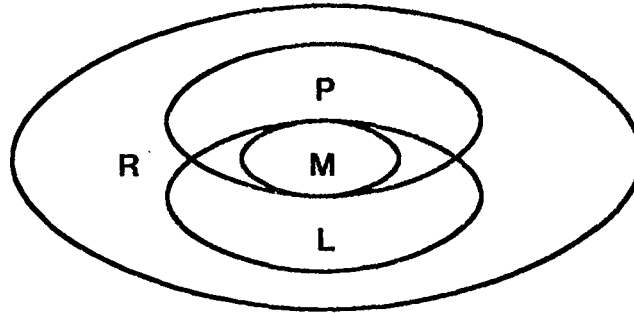


Figure 3: Subset relations of node types

neighborhood. A DOLP band-pass sample which is a local 1-D positive maximum or negative minimum within an band-pass image is labeled as an R. A node of type R which is a positive maximum or negative minimum in all four directions in an image is labeled as a P.

The node types of L and M are assigned based on the relative values of the DOLP sample attributes to linked nodes of type R or P. A node at level k of type R may be labeled as a node of type L if it is connected to nodes at level $k+1$ of type R which have a smaller DOLP sample, and it is not connected to a node at level $k-1$ of type R with a larger DOLP sample. Similarly, a node of type P may be labeled as a node of type M if it is connected to nodes at level $k+1$ of type P which have a smaller DOLP sample, and it is not connected to a node at level $k-1$ of type P with a larger DOLP sample.

Thus the node types can be grouped into subsets, based on context, as shown in figure 3. All nodes are type R. Some nodes of type R can be type L. Some nodes of type R or L can also be type P. Some nodes which are both type P and type L can also be type M.

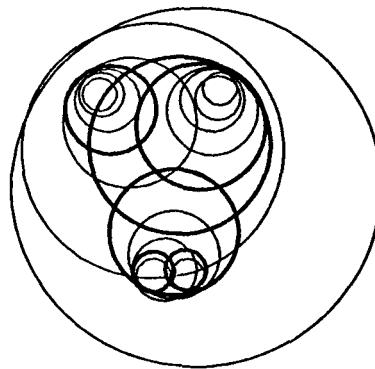


Figure 4: Circle representation of the negative peaks from levels 11 through 4 of the larger bolt from figure 1. Darker circles represent M-nodes, lighter circles represent P-nodes.

An example of the use of peaks to represent a gray scale pattern is given in figures 4 and 5. Figure 4 shows the negative peaks from band-pass levels 11 through 4 for the larger bolt in the DOLP transform image shown

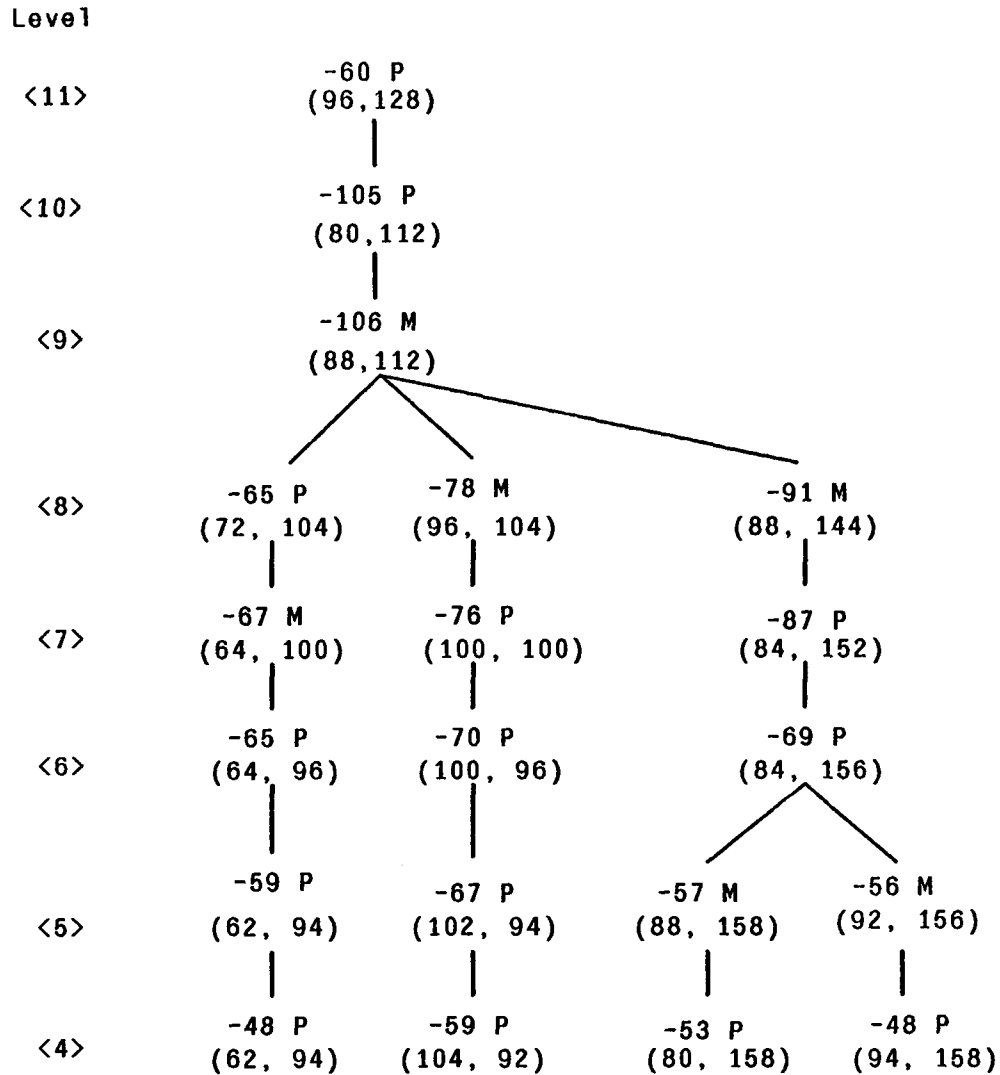


Figure 5: The negative peaks from levels of 11 through 4 which describe the larger bolt from figure 1

in figures 1 and 2. For simplicity, only negative P-nodes over this bolt at the low resolution levels are shown. Figure 5 shows some of the attributes of these peaks. At the top of each symbol, the DOLP sample value, and the symbol type (M or P) are shown. In the second row, the coordinates are shown in parentheses as (x, y). L-nodes and R-nodes, and the opposite signed peaks are not shown in these figure.

3.2 Peak Paths

In general, when a DOLP filter has a high correlation with an image pattern, a peak node will occur at similar locations in the set of adjacent band-pass images. These peaks will be assigned a connectivity as described above. Except in rare cases, the magnitude of DOLP sample attributes of these peaks will rise monotonically through the levels to a local maximum and then decrease. It is possible to use this regularity to

group collections of connected peaks into a second level symbol called a Peak Path. This second level encoding can greatly reduce the complexity of pattern description.

A connected set of P-nodes form a tree which spans from low resolution to high resolution. Peak Paths represent branches in this tree. A connected set of P-nodes from level i to level j ($i > j$) are grouped into a Peak Path whenever:

- a P-node at level j is a local minimum along a Peak Path, or
- a P-node at level j has more than one P-node at level $j-1$ attached to it.

When a Peak Path is formed by cutting a connected set of peaks with one of these rules, the top (low-resolution) Peak Path is said to be a *Parent* Peak Path, while the bottom (high resolution) Peak Path is said to be a *child* Peak Path. Peak Paths formed in this manner retain their connections. These connections are the basis for the tree structure of both the shape descriptions and probabilistic models described in this paper.

A Peak Path, P_n , is defined by

\vec{a} a vector of attributes,

l a set of connections to the parent and children Peak Paths, if they exist.

The attributes for a Peak Path are the set:

$$\vec{a} = \{\text{top}, x, y, k, d, t, l, c\}$$

These attributes are sometimes referred to as a_0 through a_7 . As defined above, every Peak Path contains an M node. The M-node plays a crucial role in defining the attributes of a Peak Path. The attributes (x, y, k, d) are simply the attributes of the M-node. The attributes t and l reflect the length of the Peak Path, while the attribute c describes the number of children.

Top A boolean which is true if the Peak Path is not the child of a lower resolution Peak Path.

(x, y) The image coordinates of the M-node of the Peak Path.

k The band-pass level of the M-node.

d The DOLP amplitude of the M-node.

t The distance (in band-pass levels) from the top most node in the Peak Path to the M-node.

l The number of band-pass levels spanned by the Peak Path.

c The number of children Peak Paths which descend from this Peak Path.

The connections, l , include a connection to a parent Peak Path if it exists, and a connection to each of the children Peak Paths.

It is also possible to define path structures for R nodes and for L nodes. A ridge path, "R-Path" is a sequence of ridges of the same sign from adjacent locations in the same band-pass image. R-Paths are characteristic of boundary segments at a given resolution level. R-paths were used in the shape matching technique described in [2]. An "L-Path" is a connected sequence of L-Nodes of the same sign connected to adjacent locations at the same or adjacent band-pass levels. L-Paths are characteristic of elongated shapes. An L-Path usually terminates in an M-Node. In the near future we will extend the training and matching techniques described below to include L-paths. Inclusion of L-paths into descriptions and models will create a network structure between the branches of the Peak Path trees.

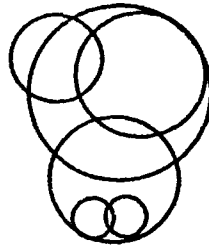


Figure 6: A circle representation of the M-nodes of the Peak Paths for the larger bolt in figure 1.

The scale of the M-Node is represented by the radius of the circle. The radius is chosen to approximate the radius of the inner positive lobe of the Difference of Gaussian filter. The position of the M-node is represented by the position of the circle.

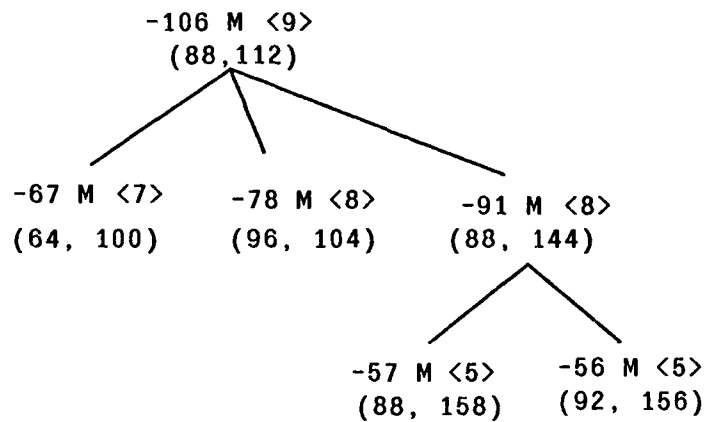


Figure 7: The M-node attributes of the lowest resolution negative Peak-Paths in the larger bolt shown in figure 5

Figures 6 and 7 shows the Peak Paths that result from the peaks shown in figure 5. Figure 6 shows a circle representation from the M-nodes from the low resolution Peak Paths. Figure 7 shows the connectivity between Peak Paths and the attributes of DOI.P sample intensity and level for the M-node, and the locations of the M-nodes. Only the negative Peak Paths at the lower resolution levels (levels 11 through 4) are shown.

4 Probabilistic Models of Gray Scale Patterns

A pattern model serves as a definition for a class of patterns. In structural pattern recognition, the pattern model defines the symbols, attributes and connectivity that have been observed in an ensemble of training samples. In most cases the members of a training set will exhibit some variability in structure. This variability may be expressed in a pattern model by attaching a probability of occurrence to each symbol, and expressing the attributes of each symbol as a probability distribution.

The probability of occurrence and the attributes for each symbol are learned by "training" on a sample set for each class. In those cases where the structure of the training samples are the same, the probability of occurrence for the symbols will be high, and the attributes will be expressed by a narrow probability distribution. In those cases where the structure varies, the probability of occurrence for each symbol will be decreased and the attributes will be expressed by probability distributions with a large variance. A training process for multiple resolution probabilistic graph models is described in section 6. This section defines a technique for representing a probabilistic graph model based on Peak Paths. The extension of this to L paths has not been performed as of this writing, but is thought to be straight forward.

4.1 Definition of Model Symbols

A multiple resolution probabilistic graph model, M , is composed of a position transformation (or "pose") \vec{T} , and a list of labeled Peak Paths, π_n for $n = 1, \dots, N$, with probabilistic attributes and with connections to other Peak Paths. Each Peak Path, π_n , contains:

1. a label from the set {PP, TP, CP} (described below),
2. a probability of occurrence, $P(\pi_n)$,
3. a set of attribute probabilities, $\vec{\alpha}_n$, and,
4. a connection to a parent Peak Path (if it exists) and a set of connections to children Peak Paths λ_n .

Each attribute probability is represented by a Gaussian probability density function. The mean and standard deviation of the density function are determined incrementally during training. In principle, the connections, λ_n , can be implemented with a probability attribute. In the techniques described in this paper this is not done.

It is desirable to be able to represent classes of objects which can undergo transformations such as scaling, translation, rotation, various forms of stretching, and articulation of rigid components. The probabilistic models defined here are designed to accommodate such transformations. This is accomplished by defining the position, orientation and scale of each symbol relative to a parent symbol. Only a root symbol, called the "Principal Peak Path" has attributes in absolute coordinates. For other Peak Paths in the model the absolute attributes at a particular pose are obtained by following the tree structure from the Principal Peak Path.

Thus, for example, in the training samples when a part of a pattern occurs scaled or rotated with respect to the rest of the pattern, the probability distributions for position or scale of the lowest resolution symbol which describes that part is given a larger variance, but the distributions for higher resolution children symbols are kept small. Defining attribute probabilities relative to parent symbols requires the definition of labels for Peak Paths.

4.2 Peak Path Labels in a Model

Peak-Paths are divided into three classes according to their position in the model:

- Principal Peak-Paths (PP) contain the lowest resolution M-Node in the model. It is the only Peak Path whose attribute probabilities are expressed in absolute coordinates.
- Top Peak-Paths (TP), other than the Principal Peak Path, have no parent peak-paths (i.e., are not children). Attribute probabilities of TP symbols are defined with respect to the PP.
- Child Peak-Paths (CP) are directly connected to lower-resolution "parent" Peak Paths. The attribute probabilities are defined with respect to the attributes of the parents.

4.3 Attribute Probabilities and Likelihoods

Matching a description to a model requires finding the correspondence mapping which maximizes the joint Bayesian probability that each description symbol is an instance of the corresponding model symbol. In the techniques below, the matching algorithm is designed to maximize the sum of the log likelihood of the conditional probability [6].

The probability that an observed description Peak Path, P_j , is an instance of model Peak Path, π_n , is given by Bayes law to be

$$P(\pi_n | P_j) = \frac{P(P_j | \pi_n) P(\pi_n)}{P(P_j)}$$

The probability of the description Peak Path, $P(P_j)$, is constant for all possible model Peak Paths, and thus may be ignored in a search for a maximum. The probability that a model symbol exists, $P(\pi_n)$, is the probability of occurrence. Thus, the model symbol which has the highest probability of being an instance of a description symbol may be found by finding the model symbol, π_n , for which similarity, $S_{j,n}$, is maximized.

$$S_{j,n} = P(\pi_n) P(P_j | \pi_n)$$

The expression $P(P_j | \pi_n)$ may be evaluated as the probability of obtaining the attribute vector, \vec{a}_j of P_j given the the attribute probabilities, $\alpha_{i,n}$ of the model Peak Paths π_n .

$$S_{j,n} = P(\pi_n) \prod_{i=1}^I P(a_{ij} | \alpha_{i,n}).$$

This equation assumes that each of the attributes are independent. While this may not be strictly true, it significantly simplifies the implementation of the probability calculations, without seriously affecting the performance of the system.

The probability distribution for each attribute α_i is represented by the mean, μ_i , and the variance, σ_i^2 of a Gaussian distribution. Thus, given the assumption that the attributes have a covariance which is the identity matrix, the probability of observing an attribute value, a_i , given a probabilistic attribute, α_i , is given by

$$P(a_i | \alpha_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(a_i - \mu_i)^2}{2\sigma_i^2}}$$

Note that π in this equation refers to the well known constant.

Representing these probabilities with a Gaussian distribution provides the opportunity to replace the calculation of a probability for each attribute, with the calculation of a computationally cheaper log likelihood.

$$\ln(P(a_{ij} | \alpha_{i,n})) = -\frac{(a_{ij} - \mu_{i,n})^2}{2\sigma_{i,n}^2} + \ln\left(\frac{1}{\sigma_{i,n}\sqrt{2\pi}}\right)$$

The term

$$\left(\frac{1}{\sigma_{i,n}\sqrt{2\pi}}\right)$$

scales the probability distribution of each attribute so that it has an area of one. This term is constant for a given model Peak Path attribute, and is not needed when searching for the description Peak Path with the highest likelihood of correspondence. It is, however, needed for the global similarity measure.

Thus, the likelihood, $L_{j,n}$, that an observed description Peak Path, P_j , is an instance of a model Peak Path π_n is given by

$$L_{j,n} = \ln(S_{j,n}) = \ln(P(\pi_n)) + \sum_{i=1}^I \left[-\frac{(a_{ij} - \mu_{i,n})^2}{2\sigma_{i,n}^2} + \ln\left(\frac{1}{\sigma_{i,n}\sqrt{2\pi}}\right) \right].$$

This formula replaces the use of true probabilities with a log likelihood measure which returns a negative value near 0 for attributes near the mean and a more negative number for less likely values. Because the logarithm is a monotonic function, the sorted order and the maximum of the log likelihoods will be the same as for probabilities. The log likelihood is also considerably less expensive to compute.

4.4 Probabilistic Attributes in the Model

For Principal Peak Paths, the attribute set is identical to those used for description, except that each attribute has a random variable with an associated probability distribution in place of a value.

For TP and CP Peak Paths, the attribute probabilities used in the model divide into two classes:

1. attributes which are *dependent* on the parent Peak Path,
2. attributes which are *independent* of the parent Peak Path.

The *dependent* attributes are:

α_1 θ : the angle of the vector from the parent Peak Path. The Principal Peak Path is the parent for Peak Paths of type TP.

- α_2 R: the distance from the parent Peak Path
- α_3 Δk : the difference of level from the parent Peak Path
- α_4 Δd : the ratio of the DOI.P sample intensity of the M node of this Peak Path to the parent Peak Path

Model independent attributes have been defined by

- α_5 t: the number of levels between the M node and the top-most peak in the path,
- α_6 l: the number of levels between the top most and the bottom-most peak in the path,
- α_7 c: the number of children Peak Paths.

4.5 The No-Match Likelihood

A lower bound on the log likelihood measure, called the "no-match" likelihood, $L_{n, none}$, is provided by the probability of occurrence of a model Peak Path. This is the likelihood that will be used in a global similarity measure if no correspondence is found for a model Peak Path.

$$L_{n, none} = \ln(1.0 - P(\pi_n)) + \beta_n$$

The term β_n is a constant determined from the attribute distributions in order to scale likelihoods of observed and unobserved symbols.

4.6 Model Transformation

Except for the Principal Peak Path, each calculation of a log likelihood function requires that description Peak Path attributes which correspond to model dependent attributes be converted to a "relative" coordinate system. The attributes of distance, R, scale, Δk , and intensity, Δd , can be computed relative to the attributes of a *parent* Peak Path. However, orientation requires additional information. This information is supplied by the "pose" of the model.

During matching, the pose of the model is given by a "transformation," \bar{T} , which is hypothesized to translate, scale, and rotate the model Peak Paths so that they can best correspond to description Peak Paths. The first step in matching is to determine the position, scale and intensity parts of \bar{T} from the Peak Path in the description which is hypothesized to match to the Principal Peak Path in the model. The second step is to determine an estimated orientation from the correspondence of the model Peak Paths of type TP to the description Peak Paths of type TP.

5 Multiple Resolution Probabilistic Matching

Matching a probabilistic graph or tree model to a structural description is a problem of finding a correspondence relation between the symbols in the model and the symbols in the description which maximizes some global similarity measure. In the algorithm described here, the global similarity measure is the sum of the log likelihoods of the individual Peak Path correspondences.

This algorithm uses the hierarchical structure of the multiple resolution description and model to successively constrain possible correspondences. Thus the algorithm assumes that the lowest resolution Peak Path can be reliably found in the data. This may not be an appropriate assumption in situations where the 2-D image of a 3-D object can have a significant variation in form due to changes in 3-D orientation or photometric effects.

The matching algorithm consists of the following steps:

1. Determine an initial "Pose" which translates, scales the model onto the description.
2. Determine a correspondence between the list of Peak Paths of type TP in the model and the Peak Paths in the description. This correspondence provides the orientation part of the pose.
3. Determine the correspondence for the list of Peak Paths of type CP.

As each new correspondence is obtained, a likelihood estimate is updated. This partial estimate may be used to halt the matching if "reasonable" correspondences are not being found.

5.1 Finding the Pose

The matching process begins with a hypothesis that a low resolution Peak Path, P_j , in a description is an instance of the Principal Peak Path, π_1 , from the model. The similarity measure, S_e , is initialized by evaluating the absolute attributes of P_j with the sum of the log likelihoods for the Principal Peak Path, π_1 , for the set of attributes $i = 2$ to I . The orientation attribute, α_1 , is not meaningful for the Principal Peak Path.

$$S_e = - \sum_{i=2}^I \frac{(a_{ij} - \alpha_{i1})^2}{\sigma_{i1}^2} + \ln\left(\frac{1}{\sigma_{i1} \sqrt{2\pi}}\right)$$

If this initial estimate is above a threshold, then matching may continue.

Accepting the hypothesis for the Principal Peak Path specifies an estimate of the position, intensity, and scale for matching. It does not provide any information about the orientation. In our initial matching algorithm the orientation estimate is determined from the correspondence for a Second Peak Path. In the section below we describe an algorithm in which a set of possible orientations are hypothesized as the first stage in a search for the best correspondence mapping.

The Peak Paths of type TP in the model are sorted by probability of occurrence, resolution level, and distance from the Principal Peak Path. Our initial matching algorithm, uses the Peak Path at the top of this list to determine the orientation of the Pose. This Peak Path is sometimes referred to as the "Second Peak Path." The log likelihood similarity measure for the Second Peak Path is computed for each Peak Path of type TP in the description, using attributes α_2 through α_7 . The Peak Path that maximizes this similarity measure is chosen as the corresponding description Peak Path. The pose orientation is determined by the difference in angle of the vector to this Peak Path in the model and in the description.

5.2 Matching Top Peak Paths

Given a pose, the matching problem becomes a problem of finding a correspondence mapping for the Top (type TP) Peak Paths that maximizes some global measure of similarity. This is perhaps the most important and most difficult stage of the matching process. The similarity function for a correspondence between an individual model Peak Path and a description Peak Path is measured by the log likelihood of obtaining the description Peak Path's attributes given the model Peak Path and the current pose. A normalized sum of such likelihoods is used as a global similarity measure.

If there are N_M model Top Peak Paths and N_D description Top Peak Paths then there are as many as $N_M N_D$ log likelihoods to be evaluated. Such a correspondence problem has become a classic problem in machine vision, for which a variety of approaches have been investigated and reported. Some examples include the maximal cliques algorithm [1], relaxation labeling [15], pose clustering [13], hypothesis of rigid transformation [8] and [7] and heuristic search [14]. Most of these algorithms could be adapted for this task, although in some cases it would be difficult or expensive to incorporate a generalized similarity measure.

A crucial aspect in the application of any correspondence matching algorithm is the representation in which the information is expressed. We initially believed that the expressive power of the Peak Path representation would permit us to determine the best correspondence without backtracking. In particular we sought to use the lowest resolution Peak Paths to determine the pose transformation for the model, instead of using a long search process. Thus we implemented a very simple correspondence matching algorithm which we have come to call the "greedy algorithm."

The greedy algorithm is sequential, and does not guarantee an optimal match. The algorithm is based on the idea of forming the list of $N_M N_D$ possible correspondences and sorting this list based on the likelihood value for each pair. Correspondences are then assigned in the order of the likelihoods with no possibility of multiple matches. To keep the problem small, we only include Top Peak Paths within 6 resolution levels (a factor of 8 in scale) of the Principal Peak. For our bolts images, this typically limited the number of Top Peak Paths to less than 10.

Stated more precisely, the greedy algorithm operates as follows:

1. For all model Peak Paths of type TP, π_n , for $n = 1$ to N , and for all description Peak Paths, P_m , for $m = 1$ to M , construct the triple $(n, m, L_{n, m})$, where $L_{n, m}$ is the likelihood that π_n is an instance of P_m .
2. Sort this list of triples on the field $L_{n, m}$.
3. Starting with the top of the sorted list, retrieve the triple, $(n, m, L_{n, m})$.
 - a. If the correspondence for π_n or for P_m has been found, then discard this triple and advance to the next.
 - b. If neither correspondence has been found, then mark both symbols and save the correspondence.
 - c. Add the likelihood to the partial similarity estimate, i.e. $S_e = S_e + L_{n, m}$.

- d. If at least one model symbol remains with no correspondence, and at least one description symbol remains with no correspondence, and the most recent likelihood measures retrieved from the sorted list is above a threshold, then continue retrieving triples from the list.

4. If the partial estimate S_e is below a threshold, halt.

The principal benefit of the greedy algorithm is that it is fast; it has a computational complexity of $O(N_M N_D \text{Log}(N_M N_D))$. It is also very easy to implement.

Although, the greedy algorithm has been found to work well in most cases, there are some situations in which it fails. Most of these situations fall into the following three categories:

The comb effect: If a pattern contains a sequence of identical small patterns spaced at close regular intervals, then if the correspondence for one is not correctly assigned, the correspondence for all its neighbors is also affected.

Noise Effects: If there are spurious Peak Paths of type TP in the description, these can sometimes block the correspondence of a model Peak Path to the correct description Peak Path. In this case, there is usually no correspondence or a very poor correspondence for the children Peak Paths in the model, and the overall likelihood is severely degraded. The biggest source of noise was middle resolution photometric effects (highlights and shadows).

Low Resolution Photometric Effects:

The position attributes of the Peak Paths of type TP are specified with respect to the principal peak path. If highlights or shadows are large enough and severe enough to interfere with the Principal Peak Path in a description, then the pose transformation will be incorrect. In this situation, the log likelihoods for all the TP Peak Paths will be distorted and frequent mismatches will occur.

Most of our experience with this matching algorithm was obtained with an interactive training algorithm described below. With this training program, each match was displayed graphically, and we were able to intervene by hand to reject an incorrect correspondence. The most frequent error (~5%) in our training examples, was an incorrectly chosen correspondence for the Second Peak Path which results in an incorrect model orientation. Given the fact that the matching can fail completely if the wrong orientation is selected, we now believe that this orientation part of the pose is best determined by evaluating the correspondence of Top Peak Paths at a set of hypothesized model orientations.

The consequences of both noise in the description Peak Paths and the Comb Effect can be minimized by only seeking a match for the model Peak Paths, and by permitting more than one model Peak Path to correspond to the same description Peak Path. This prevents an error in one correspondence from creating errors in other correspondences. Alternatively it is possible to determine the set of correspondences which are globally best using a heuristic search algorithm.

The problems caused by low resolution photometric effects are fundamental to using one or two lowest resolution Peak Paths to define the model transformation or pose. Any error in the position or scale of the Principal Peak Path will degrade the correspondence likelihoods for all the Peak Paths of type TP, and thus increase the possibility of mismatch. We could avoid this problem only by abandoning the use of a "Principal Peak Path" and using one of the other matching techniques listed above.

5.3 Determining the Pose by Pairwise Search

The comb effect and the effects of middle resolution noise can be minimized by only seeking a match for each model Peak Path and by permitting more than one model Peak Path to correspond to the same description Peak Path. The catastrophic results of an incorrectly chosen Principal Peak path or Second Peak Path can be avoided by hypothesizing a set of possible pose transformations based on pairs of Top Peak Paths. This idea is a generalization of the "hypothesis of rigid transformations" technique used by [7] and [8]. These two ideas are embodied in an algorithm which hypothesizes rigid transformations for the Top Peak Paths in the model based on the correspondences of pairs of Peak Paths, and then evaluates the global similarity for a subset of "most likely" transformations. Such an algorithm would work as follows.

For Top Peak Paths in the model, the pose dependent attributes (α_1 through α_4), which were previously computed with respect to a Principal Peak Path, are each augmented with a list of attribute vectors, computed for a set of other Top Peak Paths. This set can be kept small by only computing relative attributes for Top Peak Paths with a probability of occurrence of 1.0 and by restricting each list to the closest N_p Peak Paths.

Matching then begins by forming a list of possible correspondences between pairs of Peak Paths from the Model and pairs of Peak Paths from the description. Each correspondence of pairs defines a pose hypothesis. For each such correspondence, the log likelihood is computed using the appropriate attribute vector, as well as the pose independent attributes. The list of pair-wise correspondences are sorted by log likelihood. The members at the top of the list provide a set of "most likely" poses. A subset of the "best N" or "all above a threshold tolerance" can be selected as candidates. The model Peak Paths of type TP are then rigidly transformed by the each hypothesized pose and a global similarity is computed.

For each hypothesized pose, the algorithm computes a global similarity as follows. For each model Peak Path of type TP, a list is made of the likelihood of correspondence for each description Peak Path of type TP, given the pose. The best log likelihood is selected for inclusion in the global similarity. If the best log likelihood is more negative than the "no-match" likelihood, $L_{n, none}$, then the no-match likelihood is used in its place. For each pose, the sum of the best log likelihoods are computed. The pose which yields the highest sum of log likelihoods will provide both the pose and the correspondence mapping for Peak Paths of type TP.

The computational complexity for this algorithm need not be prohibitive. If each model Peak Path is restricted to a list of N_p pairwise attribute vectors, then the total number of pairwise combinations in the model will be $N_M N_p$. Pairing each description Peak Path with its nearest N_p neighbors will give $N_D N_p$ pairs. Thus the number of log likelihoods to be evaluated in this stage is $N_M N_D N_p^2$. For each pose hypothesis, computing a global similarity requires $N_D N_M$ likelihoods. If we restrict the search to the N best poses, then the total number of likelihoods to be computed N_C is

$$N_C = N_M N_D N_p^2 + N_M N_D N^2.$$

5.4 Matching Children Peak Paths

If the partial estimate remains above a threshold after finding the best correspondence for the model Peak Paths of type TP, then matching continues for the children Peak Paths (Peak Paths of type CP). Finding the correspondence for the Peak Paths of type CP is organized as a hierarchical sequence of correspondence matches among a very small number of candidates. The organization of this process stems directly from the structure of the Peak Path hierarchy. That is, a model Peak Path can only correspond to a description Peak Path if their parents correspond.

Each step in the matching process for child Peak Paths starts with the correspondence of a model Peak Path, π_n to a description Peak Path, P_m . This correspondence provides a list of children from both the model and the description. The log likelihoods for all possible correspondences are then computed. For N model Peak Paths and M description Peak Paths, this gives MN likelihoods. As M and N are small (typically 2, and maximum 4), it is not unreasonable to consider all possible sets of correspondences. The set of correspondences for which the sum of likelihoods is maximum is selected. Any correspondence in this set for which the likelihood falls below a threshold are rejected.

For each accepted correspondence, the global similarity measure is updated by adding the likelihood.

$$S_e = S_e + L_{n,m}$$

For each model Peak Path for which no correspondence was found, or for which the best correspondence was below the no-match likelihood, the global similarity is updated by adding the no-match likelihood.

$$S_e = S_e + L_{n, \text{none}}$$

The matching algorithm is then applied recursively to each of the correspondences above the rejection threshold.

Matching of Child Peak Paths proceeds through the Top Peak Path in the model in sorted order. For each Top Peak Path, the matching proceeds recursively "depth" first. If the partial similarity measure ever becomes less than the rejection threshold, the process is halted.

5.5 Similarity Measure

The final result of matching is a correspondence list and a numerical measure of the similarity of the model and the description. The log likelihood values are rarely exactly zero. Hence every Peak Path in the model decreases the overall log likelihood, even if it has a close match. The log likelihoods for different models can not be compared unless they are normalized to remove this effect. We normalize the model similarity by dividing by, N , the total number of correspondences included in the similarity measure. Thus the similarity of model of a pattern class to a description of a pattern is given by

$$S_M = \frac{S_e}{N}$$

6 Training

The probabilistic graph models described in section 4 are learned by a training process. An ensemble of observations, called the "training set," is used to derive the structural components and attributes for a "composite" model of the training samples. Training is performed on a set of "clean" images; that is, the images contain only an example of the pattern class to be learned. The presence of spurious forms in the training images is minimized. This helps to assure that only the "important" Peak Paths are found in the model.

6.1 The Training Algorithm

The training process begins with an empty model and repeats the following stages for each sample image in the training set:

1. The next image from the training set is expressed as a tree of Peak Paths in absolute image coordinates.
2. A "pose" is obtained which places the model at the position, orientation, and scale of the pattern in the sample image. This pose is hypothesized by the system and confirmed by the human supervisor.
3. For each model Peak Path of type TP, the correspondence to the most likely Peak Path in the description is determined and displayed. In our experimental implementation, the human supervisor may intervene and examine the log likelihoods of alternate matches.
4. As each correspondence is found, the mean and standard deviation of the probability distributions for the attributes are incrementally determined, and the probability of occurrence for the Peak Path is updated.
5. The most likely correspondences are then found for the Peak Paths of type CP, and their probability distributions and probability of occurrence are also updated.
6. Any Peak Path in the description for which there was no corresponding Peak Path in the model, and which is within the appropriate bounds of resolution and position, is added to the model as a Peak Path of type TP. When a new Peak Path is added to the model, the value of each attribute is assigned to the mean, and a default value is assigned to the standard deviation.
7. The Peak Paths of type TP are sorted with the following precedence: probability of occurrence (highest to lowest), average level of the M-node (lowest resolution to highest), average distance from the Principal Peak Path (furthest to closest).

6.2 The Training Program

Training is currently done with an interactive program that uses a raster graphics monitor. Our philosophy has been to begin with a program in which the user must verify each step, and to incrementally automate the training process as confidence and experience are gained with each stage.

The screen of the monitor is divided into three windows. In the upper left is a "model" window, in which a copy of the first training sample is shown. In the near future we plan to display a "synthesized" image of the model in this window. This synthesized image will be created by summing the impulse response of the DOLP filters for the mean value and location of the M-node of each Peak Path. The image of the current training sample is shown in the upper right window. At the bottom of the screen a set of text field are maintained that provide information about training.

Most of training involves specifying Peak Paths that match. Peak Paths in both the model and the description are indicated by drawing circles in the overlay plane, over the the model window or the training images window. These circle act as a cursor. The radius of the circle is given by the M-node level of the Peak Path. It is the radius of the inner "positive" lobe for the DOLP filter at that level in the DOLP transform.

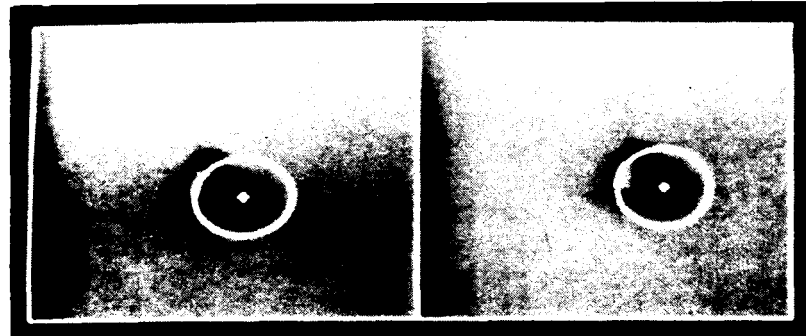


Figure 8: The interactive training program. The upper left window contains the model image (currently the first training sample). The upper right window contains the current training sample. The circles indicate the location and scale of the principal Peak Path. The lower window is used for text messages.

The position of the circle is given by the mean of the position for the M-node in the Peak Path. Examples of the screen of the training program are shown in figures 8, 9 and 10.

A cursor support mechanism permits the user to step the "current" cursor through the description or the model in a number of ways. The cursor may be stepped through the list of TP Peak Paths, up or down a Peak Path tree, or through the list of siblings at a given level.

6.3 Correspondence Matching in Training

As with structural matching, the basic problem in training is to find a correspondence between symbols in a description and symbols in the model. The problem is easier than the general structural matching problem because it is known a-priori that the training sample is an instance of the model, and because restrictions can be made about the "cleanness" of the images used for training. The problem is harder, however, because during training the model is only partially constructed and may not be useful in finding the best correspondence.

Our initial implementation of the training process is based on the assumption that the training images are "clean," that is, the example of the pattern to be learned is the only thing in the image. In this way, the Principal Peak Path (PP) and the second TP can be easily found by finding the lowest resolution M node. The

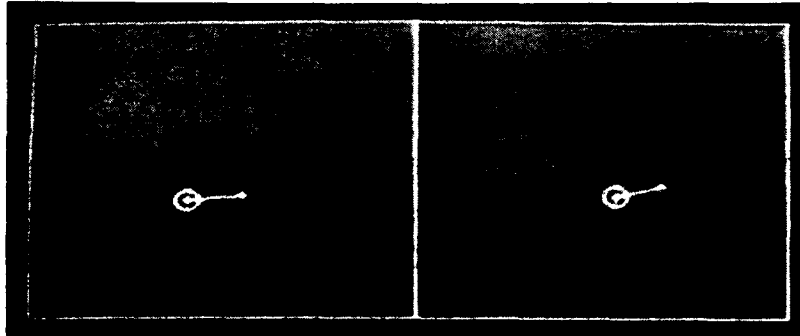


Figure 9: The interactive training program. The pose orientation is determined from the correspondence of the "second" Peak Path, displayed graphically for approval.

user is asked to approve this correspondence. The correspondence for the Principal Peak Path specifies the position and scale of the pose.

Determining the orientation of the pose requires finding the correspondence for a Second Peak Path. The model Peak Path which acts as the Second Peak Path is the Peak Path at the top of the sorted list. A search is made for the description Peak Path which is most likely to correspond to this Second Peak Path, based on all of the attributes except orientation. This search uses the similarity function described in the previous section. The user is asked to verify this correspondence. The angle to this Peak Path is then used to determine the orientation of the pose. An example of determining the Second Peak Path is shown in figure 9.

Given the pose, the training program advances through the list of model Peak Paths of type TP. The "current" model Peak Path is illustrated by a circle drawn over the model image. For each model Peak Path of type TP, the system locates the description Peak Path of type TP with the highest correspondence likelihood. The likelihood calculation is based on the mean and standard deviation obtained incrementally in previous training. The description Peak Path with the highest correspondence likelihood is also presented to the user as a circle over the training sample image. An example of this is shown in figure 10. The user has the option to accept the match, select a different description Peak Path, or abort the search for a match to the current model Peak Path. Each time the user indicates the selection of a match, the statistics of the model Peak Path are updated and the correspondence is saved.

After specifying the correspondence for all of the model Peak Paths of type TP, the search is made for the

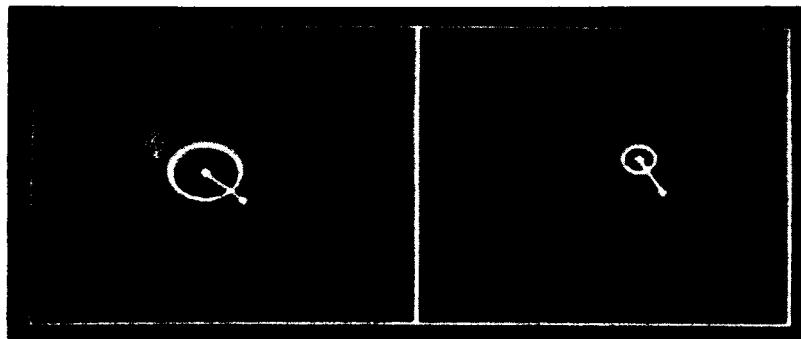


Figure 10: The interactive training program.
The correspondence of top level Peak Paths are displayed graphically for approval.

correspondence for the model symbols of type CP. This search is made "depth-first," from the children Peak Paths from each TP Peak Paths. The matching process proceeds as with model Peak Paths of type TP, except that the user is forbidden to specify a match in a case where the parents do not match. That is, a model Peak Path of type CP can only be matched to a description Peak Path of type CP if their parents were matched. The model path of type CP may, however, be matched to a previously unmatched description Peak Path of type TP.

If a model Peak Path has children, but the corresponding description Peak Path does not, then the user is informed. The user may elect to not match those Peak Paths, or to search among the unmatched model Peak Paths of type TP.

After all of the model Peak Paths have been processed, the description Peak Paths which remain are added to the model. Each Peak Path is shown to the user before it is added to the model. If a model Peak Path which has no children is matched to a description Peak Path which has children, these children are added as children of the model Peak Path to which their parent was matched.

7 Conclusions

This paper has presented a multiple resolution representation technique for gray scale patterns based on the DOLP transform. A description in this representation has the form of a connected tree of symbols with attributes. Information about the position and scale of a symbol are encoded in the attribute list along with

other features. While the connectivity relation is based on position and scale, this information is not explicitly encoded in the connectivity.

A representation for a probabilistic model was presented, based on the multiple resolution description. In the model, the attributes of symbols are replaced by probability distributions. The connectivity structure of the model has the form of a tree of trees. In the implementation described here, the top layer is a tree with 2 levels and a large branching factor. All symbols are encoded relative to a "principal" symbol, given by the lowest resolution Peak Path. The "pose dependent" attributes of all of the second level symbols are relative to the principal symbol. At the second level, each top level symbol is the root of a strict tree structure of symbols. At this level, the tree may have an arbitrary depth, but the maximum branching factor is 4.

The connectivity information in the model and the description provides a structure for controlling the search for correspondence between model symbols and description symbols. In the matching technique presented here, the correspondences between a model and a description are found first for the principal symbol, and this correspondence is used to determine a "pose" for further matching. Correspondence are then found for the symbols at the top level, and finally for the tree under each top level symbols.

If the correct "principal" Peak Path does not occur in the description, then this matching algorithm fails. This weakness can be avoided by structuring the top layer as a graph. However, this introduces a much more difficult control problem for the search for correspondence. Indeed, one of the interesting aspects of this representation is that the hierarchical structure greatly simplifies the correspondence matching process.

8 Acknowledgements

It is important that we acknowledge the contribution of Tom Wood to this work. Not only has Tom implemented the several versions of these algorithms as they evolved, but he was an active participant in the brain storming sessions which lead to much of what has been described here. In particular, it was Tom Wood who first suggested the greedy algorithm described in section 5.

Credit also goes to Amy Lowrie who first demonstrated the expressive power of Peak Paths in a parallel project concerning matching of co-planar stereo images [4] with a multiple resolution representation. Amy was an early and strong advocate of the use of Peak Paths for representing patterns.

Credit also goes to John Schlag, who tackled many of the same problems in a parallel effort toward developing an edge based vision system [11]. We would also like to thank Jerry Agin for his many excellent comments on the techniques during their development, as well as his comments on early drafts of this paper. Finally, we would like to thank Robin Wallace for her help in typing parts of the manuscript.

References

- [1] Bolles, R. C. .
Robust Feature Matching Through Maximal Cliques.
In *Proc. SPIE Technical Symposium on Imaging and Assembly*. April, 1979.
- [2] Crowley, James L.
A Representation for Visual Information.
PhD thesis, Carnegie-Mellon University, November, 1981.
- [3] Crowley, J. L. and R. M. Stern.
Fast Computation of the Difference of Low-Pass Transform.
IEEE Trans. on P.A.M.I. , March, 1984.
- [4] Crowley, J. L. and A. L. Lowrie.
Multi-Resolution Stereo Correspondence Matching Using the 1-D DOLP Transform.
In Preparation , 1984.
- [5] Crowley, J. L. and A. C. Parker.
A Representation for Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform.
IEEE Trans. on P.A.M.I. , March, 1984.
Also available as C-MU Robotics Institute Technical Report CMU-RI-TR-83-4.
- [6] Duda, R. O. and Hart P. E.
Picture Processing and Scene Analysis.
Wiley, New York, 1973.
- [7] Ayache, N. and Faugeras, O. D.
A New Method for the Recognition and Positioning of 2-D Objects.
In *Proc. Seventh Int. Conf. on Pattern Recognition*. The IEEE Computer Society Press, August, 1984.
- [8] Hebert, M. and Faugeras, O. D.
A 3D Recognition and Positioning Algorithm using Geometric Constraints Between Primitive Surfaces.
In *Proc. Eight Int. Conf. on Artificial Intelligence*. Kaufmann Press, August 1984.
- [9] Marr, D. L., and Hildreth, E.
Theory of Edge Detection.
Proc. R. Soc. Lond. B. (207):187-217, 1980.
- [10] Pratt, William K.
Wiley-Interscience: Digital Image Processing.
John Wiley & Sons, 1978.
page 322.
- [11] A. Sanderson and J. L. Crowley.
Probabalistic Graph Matching.
In Preparation , 1984.
- [12] Sanderson, A. C. and G. Perry.
Sensor Based Robotic Assembly Systems: Research and Applications in Electronic Manufacturing.
Proceedings of the IEEE 71(7), July. 1983.

- [13] Stockman, G. and J. C. Esteva.
Use of Geometrical Constraints and Clustering to Determine 3D Object Pose.
Technical Report TR 84-002, Dept of Computer Science, Michigan State University, April, 1984.
- [14] Winston, P. H. .
Artificial Intelligence.
Addison Wesley, Reading, Mass, 1977.
- [15] Hummel, R. and S. Zucker.
On the Foundations of Relaxation Labeling Processes.
IEEE Trans. on PAMI 5, 1983.