# A Representation for Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform

JAMES L. CROWLEY, MEMBER, IEEE, AND ALICE C. PARKER, MEMBER, IEEE

*Abstract*—This paper defines a multiple resolution representation for the two-dimensional gray-scale shapes in an image. This representation is constructed by detecting peaks and ridges in the difference of low-pass (DOLP) transform. Descriptions of shapes which are encoded in this representation may be matched efficiently despite changes in size, orientation, or position.

Motivations for a multiple resolution representation are presented first, followed by the definition of the DOLP transform. Techniques are then presented for encoding a symbolic structural description of forms from the DOLP transform. This process involves detecting local peaks and ridges in each bandpass image and in the entire three-dimensional space defined by the DOLP transform. Linking adjacent peaks in different bandpass images gives a multiple resolution tree which describes shape. Peaks which are local maxima in this tree provide landmarks for aligning, manipulating, and matching shapes. Detecting and linking the ridges in each DOLP bandpass image provides a graph which links peaks within a shape in a bandpass image and describes the positions of the boundaries of the shape at multiple resolutions. Detecting and linking the ridges in the DOLP three-space describes elongated forms and links the largest peaks in the tree.

The principles for determining the correspondence between symbols in pairs of such descriptions are then described. Such correspondence matching is shown to be simplified by using the correspondence at lower resolutions to constrain the possible correspondence at higher resolutions.

Manuscript received January 26, 1983; revised May 17, 1983 and August 25, 1983. This work was supported in part by the Robotics Institute, Carnegie-Mellon University, in part by the National Science Foundation under Grant APR75-08154, and in part by the Naval Electronics System Command under Grant N00039-79-Z-0169.

J. L. Crowley is with the Robotics Institute, Carnegie-Mellon University, Pittsburgh, PA 15213.

A. C. Parker is with the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90007.

*Index Terms*—Correspondence matching, difference of Gaussian, difference of low-pass transform, image pyramid, multiresolution representation, shape matching.

## I. Introduction

A REPRESENTATION is a *formal system* for making explicit certain entities or types of information, and a specification of how the system does this [20]. Representation plays a crucial role in determining the computational complexity of an information processing problem.

This paper describes a representation for two-dimensional shape which can be used for a variety of tasks in which the shapes (or gray-level forms) in an image must be manipulated. An important property of this representation is that it makes the task of comparing the structure of two shapes to determine the correspondence of their components computationally simple. However, this representation has other desirable properties as well. For example, the network of symbols that describe a shape in this representation have a structure which, except for the effects of quantization, is invariant to the size, orientation, and position of a shape. Thus a shape can be compared to prototypes without having to normalize its size or orientation. An object can be tracked in a sequence of images by matching the largest peak(s) in its description in each image. This representation can also describe a shape when its boundaries are blurred or poorly defined or when the image has been corrupted by various sources of image noise.

This representation is based on a reversible transform referred to as the "difference of low-pass" (DOLP) transform. From

0162-8828/84/0300-0156$01.00 

its definition, the DOLP transform of an image appears to be very costly to compute. However, several techniques can be used to greatly reduce the computational complexity and memory requirement for a DOLP transform. These techniques, together with the definition of the DOLP transform, are presented in a companion paper [14].

The difference of low-pass (DOLP) transform is a reversible transform which converts an image into a set of bandpass images. Each bandpass image is equivalent to a convolution of the original image with a bandpass filter $b_k$. Each bandpass 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 bandpass images comprise a three-space (the DOLP space). The representation is constructed by detecting peaks and ridges in the DOLP space.

### A. Motivation: A Multiresolution Structural Description of Images

Interpreting the patterns in an image requires matching. If the interpretation is restricted to two-dimensional patterns, this matching is between descriptions of shapes in the image and object models. If the interpretation is in terms of three-dimensional objects, then techniques for matching among stereo images or motion sequences may be required to obtain the description of three-dimensional shape. In either case, the matching problem is simplified if descriptions are compared at multiple resolutions. Peaks and ridges in a DOLP transform provide a structural description of the gray-scale shapes in an image.

The motivation for computing a structural description is to spend a fixed computational cost to transform the information in each image into a representation in which searching and matching are more efficient. In many cases the computation involved in constructing a structural description is regular and local, making the computation amenable to fast implementation in special purpose hardware.

Several researchers have shown that the efficiency of searching and matching processes can be dramatically improved by performing the search at multiple resolutions. Moravec [21] has demonstrated a multiresolution correspondence matching algorithm for object location in stereo images. Marr and Poggio [18] have demonstrated correspondence matching using edges detected by filters at four resolutions formed from a difference of Gaussians. Rosenfeld and Vanderbrug [28] have described a two-stage hierarchical template-matching algorithm. Hall has reported using a multiresolution pyramid to dramatically speed up correlation of aerial images [15]. Kelly [17], Pavlidis and Tanimoto [30], Hanson and Riseman [16], and many others have described the use of multiple resolution images for segmentation and edge detection.

There is also experimental evidence that the visual systems of humans and other mammals separate images into a set of "spatial frequency" channels as a first encoding of visual information. This "multichannel theory" is based on measurements of the adaption of the threshold sensitivity to vertical

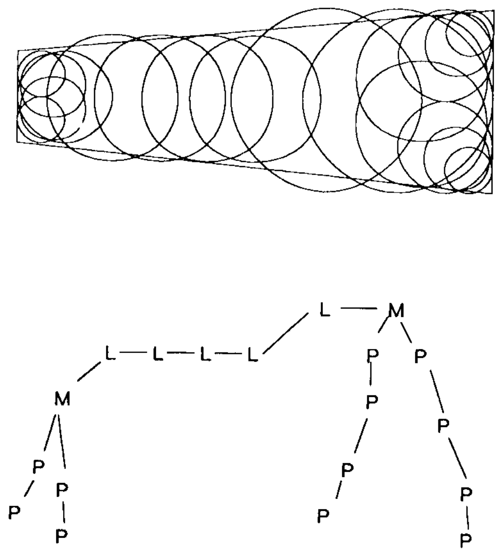


Fig. 1. A rhomboidal form and its representation. In the upper part of this figure the rhomboidal form is outlined in solid straight lines. The description is for such a form which is dark on a light background. Circles indicate the locations and sizes where the bandpass filters from a sampled DOLP transform produced 3-space peaks ($M$-nodes), 2-space peaks ($P$-nodes), and 3-space ridges ($L$-nodes). The structure of the resulting description is shown in the lower part of the figure. The description of the "negative shape" which surrounds this form is not presented.

sinusoidal functions of various frequencies [10], [29]. Adaption to a sinusoid of a particular frequency affects only the threshold sensitivity for frequencies within one octave. This evidence suggests that mammalian visual systems employ a set of bandpass channels with a bandwidth of about one octave. Such a set of channels would carry information from different resolutions in the image. These studies, and physiological experiments supporting the concept of parallel spatial frequency analysis, are reviewed in [9] and [31].

### B. Properties of the Representation

The patterns which are described by this representation are "gray-scale shapes" or "forms." We prefer the term "forms," because the term shape carries connotations of the outline of a uniform intensity region. It is not necessary for a pattern to have a uniform intensity for it to have a well defined description in this representation. In this paper we will use the term "form" to refer to the patterns in an image.

In this representation, a form is described by a tree of symbols which respresent the structure of the form at every resolution. There are four types of symbols $\{M, L, P, R\}$[1] which mark locations $(x, y, k)$ in the DOLP three-space where a bandpass filter of radius $R_k$ is a local "best-fit" to the form.

Fig. 1 shows an example of the use of peaks and ridges for representing a uniform intensity form. This figure shows the outline of a dark rhomboid on a light background. Circles illustrate the position and radii of bandpass filters whose positive center lobes are a local "best-fit" to the rhomboid. Below the rhomboid is part of the graph produced by detecting and linking

[1] In previous writing about this representation, most notably in [13], these symbols were referred to by the names $\{M^*, L, M, P\}$.

peaks and ridges in the sampled DOLP transform. The meaning of the symbols in this graph is described below.

A description is this representation contains a small number of symbols at the root. These symbols describe the global (or low-frequency) structure of a form. At lower levels, this tree contains increasingly larger numbers of symbols which represent more local details. The correspondence between symbols at one level in the tree constrains the possible set of correspondences at the next higher resolution level.

The description is created by detecting local positive maxima and negative minima in one dimension (ridges) and two dimensions (peaks) in each bandpass image of a DOLP transform. Local peaks in the DOLP three-space define locations and sizes at which a DOLP bandpass filter best fits a gray-scale pattern. These points are encoded as symbols which serve as landmarks for matching the information in images. Peaks of the same sign which are in adjacent positions in adjacent bandpass images are linked to form a tree. During the linking process, the largest peak along each branch is detected. This largest peak serves as a landmark which marks the position and size of a gray-scale form. The paths of the other peaks which are attached to such landmarks provide further description of the form, as well as continuity with structure at other resolutions. Further information is encoded by detecting and linking two-dimensional ridge points in each bandpass image and three-dimensional ridge points within the DOLP three-space. The ridges in each bandpass image link the peaks in that image which are part of the same form. The three-dimensional ridges link the largest peaks that are part of the same form and provide a description of elongated forms.

### *C. Correspondence Matching*

The easiest method for determining the correspondence of points in a pair of images is to detect landmarks in the two images and determine the correspondence of these landmarks. The peaks and ridges in a DOLP transform make excellent landmarks for such correspondence matching for several reasons. These peaks and ridges provide a compact set of symbols which denote the presence and describe the shape of forms in an image. Correspondence of symbols of similar shapes and resolutions can be found, even as forms change shape due to motion of an object or the camera. Such peaks and ridges can also be matched when the image has been corrupted by blur or high frequency noise. Matching can also be performed for a shape whose surface is composed of a random texture.

When the DOLP transform is computed with a scale factor of $\sqrt{2}$, there is a continuity between peaks at different levels which provides a description which varies gradually from a few symbols which describe low resolution information to the much larger number of symbols that describe high resolution details. Finding the correspondence between any pair of peaks constrains the possible correspondences of peaks under them at higher resolutions.

Segmentation techniques can be used to produce symbols which represent groupings of pixels and which can act as tokens for later processing. However, the gray-scale forms that occur in an image do not, necessarily, correspond to individual objects, pieces of objects, or surfaces in a 3-D scene. Furthermore, forms which are best described as a single entity at one resolution may be best described as several entities at a higher resolution. The peaks and ridges in a DOLP transform provide tokens for matching without the need for assertions about whether adjacent similar regions be grouped together.

Three-dimensional correspondence matching presents special problems, because the gray-scale appearence of objects can change due to photometric effects. Such correspondence matching is most reliable when the tokens to be matched represent points which may be detected invariant to photometric effects. The presence of such invariant points of three-dimensional shapes must themselves be detected in the gray-scale patterns of the image. These invariant points may be efficiently detected using the representation described below.

The bandpass images in a DOLP transform provide a multiresolution set of symbols for representing the image gray-scale data. These symbols may be detected in each bandpass image as either the closed zero-crossing contours or the peaks and ridges within each contour. In either case, symbols result from regions where the intensity is either darker or lighter than in surrounding regions. Each "region" will have one or more samples which are local "largest peaks" whose position in the DOLP space provides an estimate of the position and size of the region. It it not necessary for a region to be uniform to yield such peaks. Furthurmore, regions which produce a single peak at one resolution can produce more than one peak at another resolution. Finally, there is no guarantee that each peak corresponds to only one physical object, or that a particular physical object will result in a single peak.

We have observed that this representation is useful for correspondence matching to obtain three-dimensional surface information from generalized stereo, motion, or shape from occluding contours. Stereo interpretation assumes that the gray-level patterns whose shapes are compared result from the same physical three-dimensional location. This is not strictly true. Highlights on a shiny surface can move as the position of the light source or viewing angle changes. The position of shadows will change as light sources move. Nevertheless, correspondence matching of gray-level patterns can be a useful source of information about the shape of three-dimensional surfaces. The representation described above can simplify such correspondence matching.

### *D. Contents of this Paper*

The following section describes the DOLP transform. The definition of the DOLP transform is presented, followed by a description of a fast algorithm for computing the DOLP transform. This fast algorithm is based on two independent techniques which are briefly described. An example of a DOLP transform of an image which contains a teapot is also provided in this section. This image will provide the data for examples in later sections.

Section III describes techniques for converting the signals from a DOLP transform into a network of symbols. Processes are described for detecting points in each bandpass image which are on a ridge, or are a local peak. Techniques for linking peaks at adjacent locations in adjacent images are then described, along with a technique for detecting peaks which are local positive maxima and negative minima in the three-dimensional DOLP space. A process is then described for detecting the three-dimensional ridge paths in the DOLP space.

Section IV describes the basic principles of matching descriptions of shape by presenting a simple example in which the lower resolution levels of the descriptions of two teapot images are matched. The teapots in these two images differ in size by approximately 1.36. This section illustrates the use of correspondence between the lowest resolution largest peak to determine an estimate of the relative sizes and positions of the two objects. The constraints in correspondence imposed by lower resolution peaks on higher resolution peaks is then illustrated. An example of the use of the direction and length of the ridge lengths between peaks to determine correspondence is also presented.

## II. The Difference of Low-Pass Transform

This section defines the difference of low-pass (DOLP) transform and demonstrates its reversibility. A fast algorithm is then described for computing the DOLP transform. This fast algorithm is described in greater detail in a companion paper [14].

### *A. The Purpose of the DOLP Transform*

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

The difference of two low-pass filters is a bandpass filter provided that:

1) the two filters are not identical;

2) the two filters have both been normalized so that their coefficients sum to 1.0.

A filter which has a circularly symmetric passband that rises and then falls monotonically will be sensitive to image information at a particular size scale. The DOLP transform employs a set of such filters which are exponentially scaled in size and cover the entire two-dimensional frequency spectrum.

### *B. Definition of the DOLP Transform*

The DOLP transform expands an image signal $p(x, y)$ composed of $N = M \times M$ samples into $\log_S(N)$ bandpass images[2] $\mathfrak{B}_k(x, y)$. Each bandpass image is equivalent to a convolution of the image $p(x, y)$ with a bandpass impulse response $b_k(x, y)$:

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

For $k = 0$, the bandpass 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).

$$b_0(x,y) = \delta(x,y) - g_0(x,y). \tag{2}$$

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" [26]:

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

For bandpass levels $1 \leqslant k < K$ the bandpass filter is formed as a difference of two size-scaled copies of the low-pass filter:

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

In order for the configuration of peaks in a DOLP transform of a form to be invariant to the size of the 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 [13]. Thus for each $k$, the bandpass impulse response $b_k(x, y)$ is a size-scaled copy of the bandpass 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. In the Gaussian filter, this increases the standard deviation $\sigma$, relative to the image sample rate by a factor of $S_2^k$.

The scale factor is an important parameter. For a two-dimensional DOLP transform, this scale factor, denoted $S_2$, has a typical value of $\sqrt{2}$. It is possible to define a DOLP transform with any scale factor $S_2$ for which the difference of low-pass filter provides a useful passband. Marr, for example, argues that a scale factor of $S_2 = 1.6$ is optimum for a difference of Gaussian filters [19]. We have found that a scale factor $S_2 = \sqrt{2}$ yields effectively the same bandpass filter and provides two other interesting properties [13].

First, resampling each bandpass image at a sample distance which is a fixed fraction of the filter's size provides a configuration of peaks and ridges in each bandpass image which is invariant to the size of the object, except for the effects of quantization. Thus the resample distance and the scale factor should be the same value. The smallest distance at which a two-dimensional signal can be resampled is $\sqrt{2}$. Second, a DOLP transform can be computed using Gaussian low-pass filters. The convolution of a Gaussian filter with itself produces a new Gaussian filter which is scaled larger in size by a factor of $\sqrt{2}$. These two properties make $\sqrt{2}$ a convenient value for both the scale factor and the resample distance.

In principle the DOLP transform can be defined for any number of bandpass levels $K$. A convenient value of $K$ is

$$K = \log_S(N) \tag{5}$$

where the value $S$ is the square of the sample distance $S_2$:

$$S = S_2^2. \tag{6}$$

This value of $K$ is the number of bandpass images that result if each bandpass image $\mathfrak{B}_k$ is resampled at a sampling distance of $S_2^k$. With this resampling, the $K$th image contains only one sample.

[2] $S$ is the square of the scale factor.

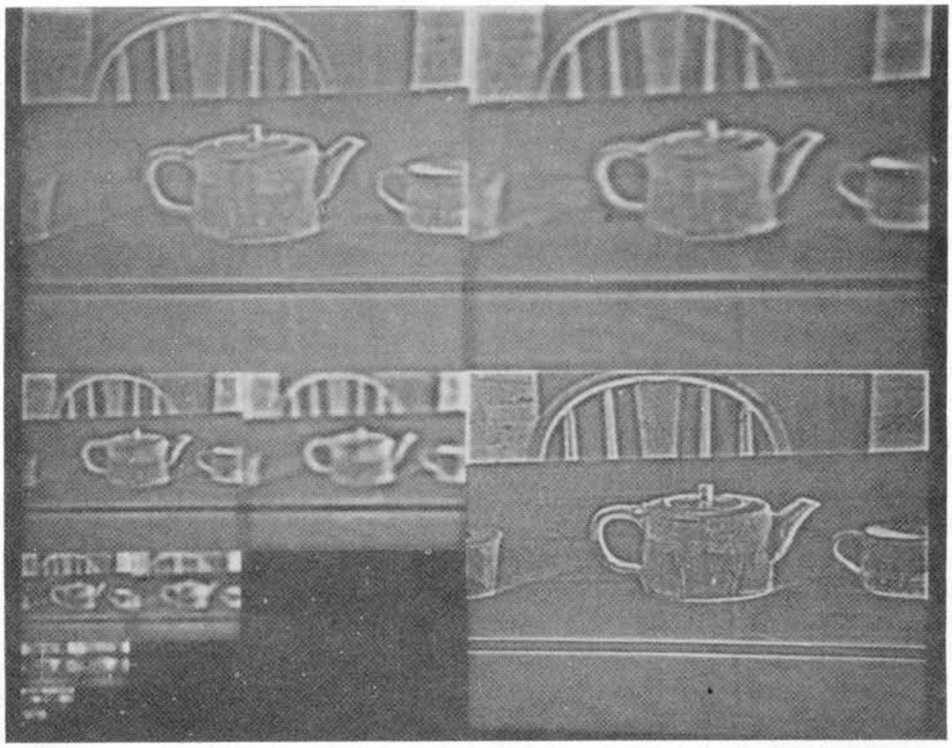

Fig. 2. The resampled DOLP transform of a teapot image.

The DOLP transform is reversible which proves that no information is lost. The original image may be recovered by adding all of the bandpass 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

$$p(x,y) = (p(x,y) * g_k(x,y)) + \sum_{k=0}^{K-1} \mathcal{B}_k(x,y). \quad (7)$$

*C. Fast Computation Techniques: Resampling and Cascade Convolution*

A full DOLP transform of an image composed of $N$ samples, produces $K = \log_S(N)$ bandpass 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."

Resampling is based on the fact that the filters used in a DOLP transform are scaled copies of a band-limited filter. As the filter's impulse response becomes larger, its upper cutoff frequency decreases, and thus its output can be resampled with coarser spacing without loss of information. The exponential growth in the number of filter coefficients which results from the exponential scaling of size is offset by an exponential growth in distance between points at which the convolution is computed. The result is that each bandpass image may be computed with the same number of multiplications and additions. Resampling each bandpass image at a distance of $\sqrt{2}$ reduces the total number of points in the DOLP space from $N \log_S(N)$ samples to $3N$ samples.

Cascaded convolution exploits the fact that the convolution of a Gaussian function with itself produces a Gaussian scaled larger by $\sqrt{2}$. This method also employs "expansion," in which the coefficients of a filter are mapped into a larger sample grid, thereby expanding the size of the filter, at the cost of introducing reflections of the pass region about a new Nyquist boundary in the transfer function of the filter. This operation does not introduce distortion, provided the filter is designed so that the reflections of the pass region fall on the stop region of the composite filter and are sufficiently attenuated so as to have a negligible effect on the composite filter. Thus a sequence of low-pass images are formed by repeatedly convolving the image with each expanded version of the low-pass filter $g_0$. Each expansion of the low-pass filter maps its coefficients onto a sample grid with a spacing between samples increased $\sqrt{2}$. Thus each low-pass image has an impulse response which is $\sqrt{2}$ larger than that of the previous image in the sequence. Each low-pass image is then subtracted from the previous low-pass image to form the bandpass images.

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. This algorithm is described in [14]. In this algorithm, each low-pass image is resampled at $\sqrt{2}$ and then convolved with the low-pass filter $g_0$ to form the next low-pass image. Since each low-pass image has half the number of samples as the previous low-pass image, and the number of filter coefficients is constant, each low-pass image is computed from the previous low-pass image using half the number of mulitplies and additions. Thus, if $C_0$ is the number of multiplies required to compute low-pass image 0, the total number of multiplies needed to compute $K$ bandpass levels is given by

$$C_{tot} = C_0(1 + 1 + 1/2 + 1/4 + 1/8 + 1/16 + \cdots + 1/K) \quad (8)$$
$$\approx 3\, C_0.$$

Each low-pass image is then subtracted from the resampled version of the previous low-pass image to form the bandpass

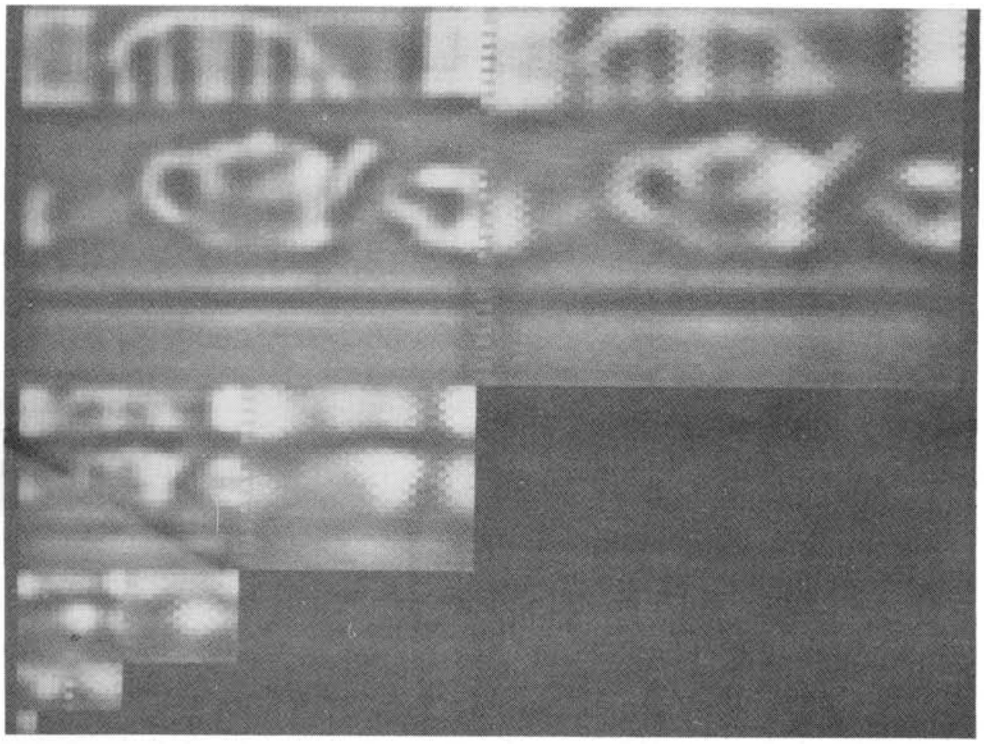

Fig. 3. Levels 5 through 13 of the resampled DOLP transform of a teapot image.

image. Thus each bandpass image has a sample density which is proportional to the size of its impulse response.

### D. An Example: The DOLP Transform of a Teapot Image

Fig. 2 shows a DOLP transform of an image of a teapot that was produced using the fast computation techniques described above. In this figure the image at the lower right is the high frequency image $\mathcal{B}_0(x, y)$. The upper left corner shows the level 1 bandpass image, $\mathcal{B}_1(x, y)$, while the upper right hand corner contains the level 2 bandpass image $\mathcal{B}_2(x, y)$. Underneath the level 1 bandpass image are levels 3 and 4, then 5 and 6, etc. Fig. 3 shows an enlarged view of bandpass levels 5 through 13. This enlargement illustrates the unique peaks in the low frequency images that occur for each gray-scale form.

The use of $\sqrt{2}$ resampling is apparent from the reduction in size for each image from level 3 to 13. Each even numbered image is actually on a $\sqrt{2}$ sample grid. To display these $\sqrt{2}$ images, each pixel is printed twice, creating the interlocking brick texture evident in Fig. 3.

## III. Construction of the Representation from a DOLP Transform

In this section we describe techniques for constructing the representation for gray-scale forms. This construction process is described as a sequence of steps in which peaks and ridges are first detected and linked in each bandpass image, and the resulting symbols are then linked among the bandpass levels.

### A. The Approach

Peaks and ridges mark locations where the DOLP impulse responses are a "best fit" to the image data. This "best-fit" paradigm is based on the observation that, for a circularly symmetric filter, correlation, and convolution are equivalent operations. Furthermore, a correlation is composed of a sequence of inner products between the filter coefficients and neighborhoods (of the same size as the filter support) in the image. Thus peaks in the convolution are locations where the impulse response correlates (is a local best fit) to the image. Ridges are a sequence of locations where the filters are a "good fit" to the image data. We may think of the DOLP bandpass impulse responses as a set of "primitive" functions for representing forms in an image.

The "local neighborhood" of a DOLP sample is the nearest eight neighbors on the sample grid at its bandpass level. A "peak" (or $P$-node) is a local positive maxima or negative minima within a two-dimensional bandpass image. A "ridge-node" (or $R$-node) is a local one-dimensional positive maximum or negative minimum within a two-dimensional bandpass image. Peaks within a form are linked by paths of largest ridge-nodes ($R$-paths).

In order for a DOLP sample to be a local positive maximum or negative minimum in the DOLP three-space, it must also be a local peak within its bandpass level. Furthermore, for a sample to be a peak in its bandpass level, it must be a ridge-node in the four directions given by opposite pairs of its eight neighbors. Peaks and ridge-nodes are first detected within each bandpass image. Peaks are then linked to peaks at adjacent levels to form a tree of symbols (composed of a paths of peaks, or $P$-paths). During this linking it is possible to detect the peaks which are local positive maxima and negative minima in the DOLP three-space. The three-space peaks are referred to as $M$-nodes.

The ridge-nodes are also linked to form ridge-paths in each bandpass image (called $R$-paths) and in the DOLP three-space (called $L$-paths). The ridges in the DOLP three-space ($L$-paths) describe elongated forms and connect the largest peaks ($M$-nodes) which are part of the same form.

The process for constructing a description is composed of the following stages.

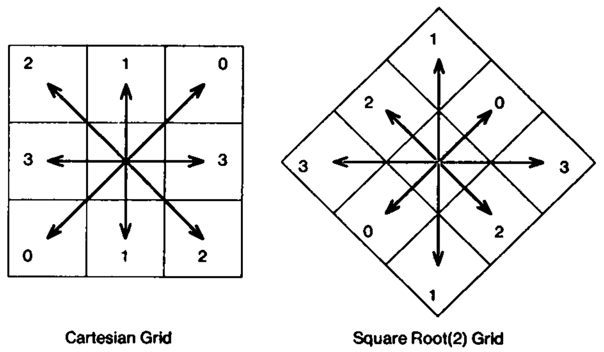


Fig. 4. The four direction tests for ridge-nodes. The four pairs of neighbors for a node in a Cartesian grid (left) and a node in a $\sqrt{2}$ grid (right) are show here. Pairs of neighbors, on opposite sides of a DOLP sample, are numbered 0 through 3, as illustrated by the arrows. The magnitude and sign of a DOLP sample is compared to each pair of neighbors. For each direction, if neither neighbor has a DOLP value with a larger magnitude and the same sign, then the direction flag for that direction is set, marking the sample as a ridge-node.

1) Detect ridge-nodes ($R$-nodes) and peaks ($P$-nodes) at each bandpass level.

2) Link the largest adjacent ridge-nodes with the same direction flags in a bandpass level to form ridges ($R$-paths) which connect the $P$-nodes in that level.

3) Link two-dimensional peaks ($P$-nodes) at adjacent positions in adjacent levels to form $P$-paths.

4) Detect local maxima along each $P$-path ($M$-nodes).

5) Detect the ridge nodes ($R$-nodes) which have larger DOLP values than those at neighboring locations in adjacent images to detect $L$-nodes.

6) Link the largest adjacent ridge points with the same direction among the bandpass levels to form three-dimensional ridge paths ($L$-paths).

The result of this process is a tree-like graph which contains four classes of symbols.

- *R-nodes:* DOLP samples which are on a ridge at a level.
- *P-nodes:* DOLP samples which are local two-dimensional maxima at a level.
- *L-nodes:* DOLP samples which are on a ridge across levels [i.e., in the three-space $(x, y, k)$].
- *M-nodes:* Points which are local maxima in the three-space.

Every uniform (or approximately uniform) region will have one or more $M$-nodes as a root in its description. These are connected to paths of $L$'s ($L$-paths) which describe the general form of the region, and paths of $P$-nodes ($P$-paths) which branch into the concavities and convexities. $L$-paths terminate at other $M$-nodes which describe significant features at higher resolutions. The shape of the boundaries are described in multiple resolutions by the ridges at each bandpass level ($R$-paths). If a boundary is blurry, then the highest resolution (lowest level) $R$-paths are lost, but the boundary is still described by the lower resolution $R$-paths.

### *B. Detection of Peak-Nodes and Ridge-Nodes within Each Bandpass Image*

Peak-nodes and ridge-nodes in each bandpass level are detected by comparing the magnitude and sign of each sample with the magnitude and sign of opposite pairs of its eight nearest neighbors. This comparison is made in four directions, as indicated by Fig. 4, and can result in one of four "direction flags" being set. A direction flag is set when neither neighbor sample in a direction has a DOLP value of the same sign and a larger magnitude.

If any of the four direction flags are set, then the sample is encoded as a $R$-node. If all four direction flags have been set then the sample is encoded as an $P$-node. The direction flags are saved to be used to guide the processes for detecting two-dimensional ridges ($R$-paths) and three-dimensional ridges ($L$-paths).

Two possibilities complicate this rather simple process. When the amplitude of the signal is very small, it is possible to have a small region of adjacent samples with the same DOLP sample value. Such a plateau region may be avoided by not setting direction flags for samples with a magnitude less then a small threshold. A value 5 has been found to work well for 8-bit DOLP samples. Also, it is possible to have two adjacent samples with equal DOLP values, while only one has a neighbor with a larger magnitude. Such cases may be easily detected and corrected by a local two-stage process. The correction involves turning off the direction flag for the neighbor without a larger neighbor.

Fig. 5 shows the direction flags detected in a region from bandpass level 7 of the teapot image. Each direction flag which is set is represented as a pair of short line segments on both sides of a sample. These line segments point in the direction in which the sample is a one-dimensional maxima. Samples which are two-dimensional peaks ($P$-nodes) are marked with a circle. It is possible to implement this detection in parallel or with a fast serial procedure.

### *C. Linking of Ridge-Paths at a Bandpass Level*

There are two purposes for which ridge paths in a two-dimensional bandpass level are detected:

1) to provide a link between $P$-nodes at a level which are part of the same form, and,

2) to construct a description of the boundary of a form.

Linking $P$-nodes of the same sign and bandpass level with ridges provides information about the connectivity of a form and provides attributes of distance and relative orientation which can be used in determining correspondences of $P$-nodes across levels.

In general, when a boundary is not a straight line, the convexities and concavities are described by a $P$-path. However, when the curvature is very gradual $P$-nodes may not occur for the concavities and convexities. In either case, a precise description of the location of the boundary is provided at multiple resolutions by the path of the ridge in a bandpass level.

A ridge is the path of largest $R$-nodes between $P$-nodes. This path can be formed by a local linking process which is executed independently at each $R$-node. The ridge path can be detected by having each $R$-node make a pointer to neighboring $R$-nodes which meet two conditions:

1) the neighbor $R$-node has the same sign and direction flags; and

2) the magnitude of the DOLP sample at the neighboring $R$-node is a local maximum in a linear list of DOLP values of neighbors.

An earlier, more complex algorithm for the same purpose was described in [13]. The result of this process when applied to the level 7 bandpass image is shown in Fig. 6.

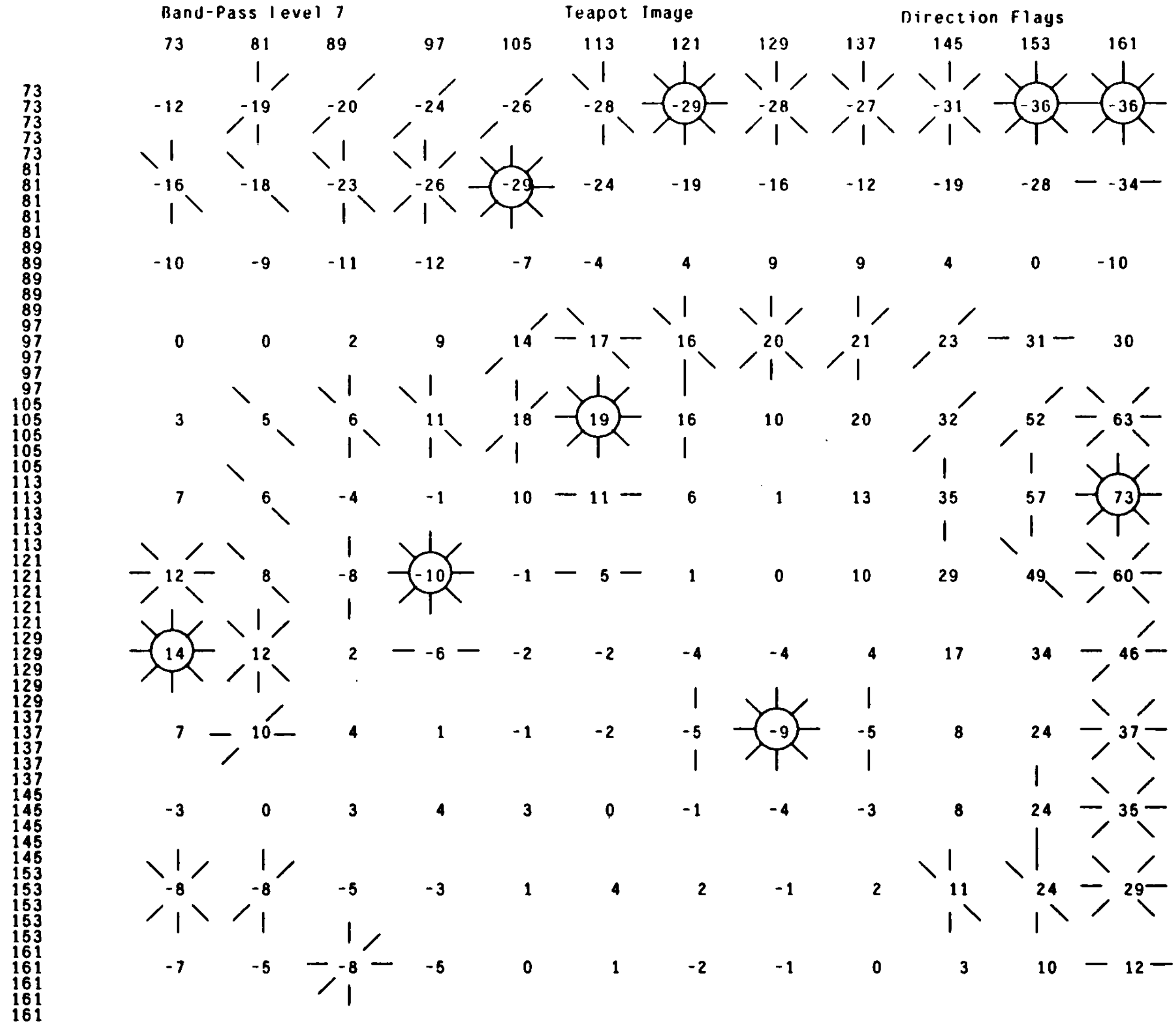


Fig. 5. The direction flags in a bandpass level 7 of the teapot image. This figure shows the direction flags detected in a region of bandpass level 7 of the teapot image. Each direction flag is represented by a pair of bars pointing toward the smaller valued neighbors. Ridges tend to run perpendicular to the direction flags. Peaks (*P*-nodes) are marked with circles. Note that both the positive and negative peaks and ridges are shown. Note also that direction flags are not detected for nodes where the magnitude of the DOLP response is less than 5.

### *D. Linking Peaks Between Levels and Detecting the Largest Peak*

The bandpass filters which compose a DOLP transform are densely packed in the frequency domain. Each filter has a significant overlap in the passband of its transfer function with the bandpass filters from neighboring levels. As a result, when a form results in a two-dimensional peak (or *P*-node) at one bandpass level the filters at adjacent levels will tend to cause a peak of the same sign to occur at the same or adjacent positions. Connecting *P*-nodes of the same sign which are at adjacent locations in adjacent bandpass images yields a sequence of *P*-nodes refered to as a *P*-path. *P*-paths tend to converge at lower resolutions, which gives the description the form of a tree. The branches at higher resolution of this tree describe the form of "roundish" blobs, bar-ends, corners, and pointed protrusions, and the patterns of concavities and convexities along a boundary. Descending the tree of *P*-paths in a description gives an increasingly more complex and higher resolution description of the form.

The magnitude of the DOLP filter response of *P*-nodes along a *P*-path tend to rise monotonically to a largest magnitude, and then drop off monotonically. This largest value is encoded as an *M*-node. Such nodes serve as landmarks for matching descriptions. An *M*-node gives an estimate of the size and position of a form or a significant component of a form. Determining the correspondence of parts of forms in two descriptions is primarily a problem of finding the correspondence between *M*-nodes and the *L*-paths which connect them.

A simple technique may be used to simultaneously link *P*-nodes into a *P*-path and detect the *M*-node (largest *P*-node) along each *P*-path. This technique is applied iteratively for each level, starting at the next to the lowest resolution level of the DOLP transform (level $K$-2). The technique can be implemented in parallel within each level. This technique works as follows. Starting at each *P*-node at level $k$, the nearest upper neighbors at level $k + 1$ are examined to see if they are also *P*-nodes of the same sign. If so, a two-way pointer is made between these two *P*-nodes.

It is possible for *P*-nodes that describe the same form at two

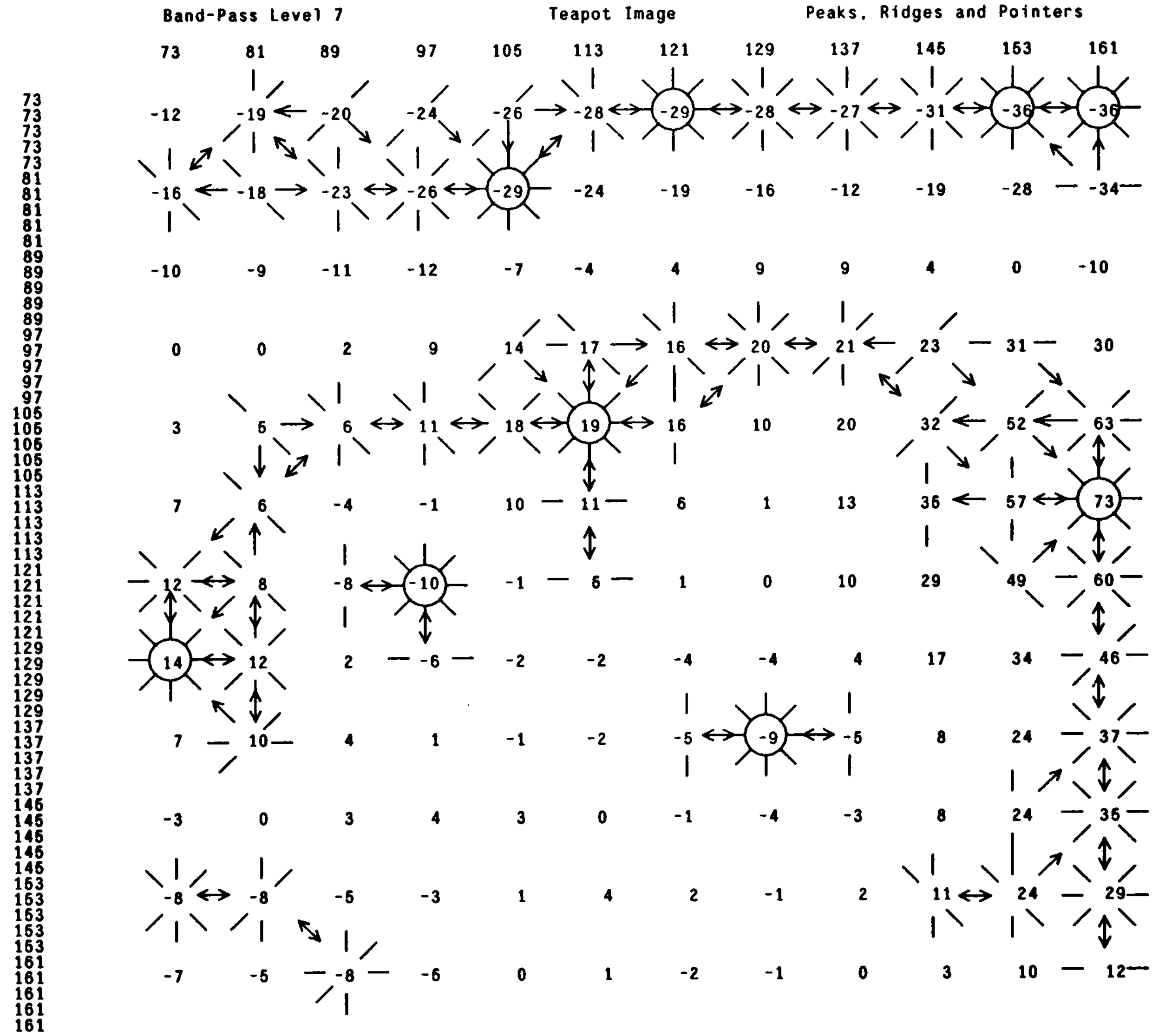


Fig. 6. The ridge paths connecting peaks ($P$-nodes) in bandpass level 7 in the teapot image. This figure shows the pointers connecting adjacent DOLP samples along positive and negative ridges in the crop from bandpass level 7 of the teapot image. Each pointer is represented by an arrow pointing to a neighbor node. A pointer is made from a $R$-node to a neighboring $R$-node if it has a common direction flag and is a local maxima among the nearest eight neighbors. A ridge may be traced between peaks by following the pointers.

adjacent levels to be separated by as much as two samples. Thus, if no $P$-nodes are found in the nearest 4 or 8 neighbors[3] at level $k + 1$ for a $P$-node at level $k$, then the nodes in the larger neighborhood given by the neighbors of the neighbors is examined. A two-way pointer is made for any $P$-nodes found in this larger neighborhood.

During this linking process it is also possible to detect the largest $P$-nodes on a $P$-path by a process referred to as "flag-stealing." This technique requires that $P$-node linking occur serially by level. In the flag stealing process, a $P$-node with no upper neighbor or with a magnitude greater or equal to all of its upper neighbors sets a flag which indicates that it is an $M$-node. Peaks which are adjacent to it at lower levels can "steal" this flag if they have an equal or larger magnitude. When the flag is stolen, the lower node sets its own flag as well as setting a second flag in the upper $P$-node which is then used to cancel the flag. This two stage process permits the $M$-flag to propagate down multiple branches if the $P$-path splits.

Fig. 7 shows the $P$-paths and the $M$-node that occur at level 6 through 1 for a uniform intensity square of $11 \times 11$ pixels, and gray level 96 on a background of 32. The reader can simulate the $P$-node linking and flag stealing process with this figure. The process starts at level 6, where the $P$-node has a value of 19.

### E. Detecting the Largest Three-Dimensional Ridge Path

Three-dimensional ridges are essential for describing forms which are elongated. An elongated form almost always has an $M$-node at each end, and a ridge of large DOLP values connecting the two $M$-nodes. The DOLP values along this ridge tend to be larger than those along the ridges in the bandpass levels above and below, because the positive center coefficients of the bandpass for that level "fit" the width of the elongated form. Where the form grows wider, the largest ridge will move

[3]The two possible upper neighborhoods in the DOLP space with $\sqrt{2}$ sampling.

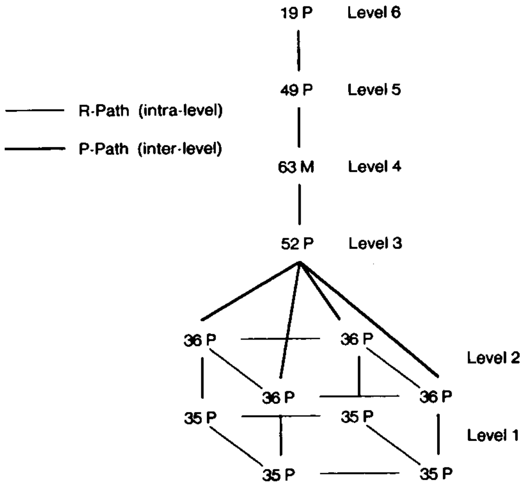

Fig. 7. Positive $P$-paths for square of size $11 \times 11$ pixels.

to a higher (coarser) bandpass level. Where the form grows thinner, the largest ridge will move to a lower (smaller resolution) bandpass level. This ridge of largest DOLP samples is called an $L$-path and the nodes along it are called $L$-nodes. $L$-nodes are $R$-nodes that are larger than their neighbors at adjacent bandpass levels.

$L$-nodes may be detected by a process similar to the flag-stealing process used to detect the largest peak, or $M$-node along a $P$-path. That is, starting at the bandpass level below the lowest resolution, each $R$-node examines a neighborhood in the level above it. An $R$-node is determined to be an $L$-node if is has a larger value than the $R$-nodes in approximately the same place in the ridges above and below it.

Thus each $R$-node scans an area of the bandpass level above it. This area is above and to the sides of its ridge. The magnitudes of DOLP samples of the same sign found in the neighborhood in the upper ridge are compared to that of the $R$-node, and a flag is set in the lower $R$-node and cleared in the upper $R$-node if the lower $R$-node is smaller. In this way, the $L$-flags propagate down to the level with the largest DOLP samples along the ridge. $L$-nodes are linked to form $L$-paths, by having each $L$-node scan its three-dimensional neighborhood and link to $L$-nodes which have the same sign and are local maxima in the three-dimensional DOLP space neighborhood.

## IV. A Simple Example of Matching

There are many applications for shape matching, and each application demands matching algorithms with certain properties. *This section does not provide a matching algorithm.* Instead, it describes some principles about matching forms that have been encoded in the representation described above. Primarily, these principles involve techniques for discovering the correspondence between "landmark" symbols in the two descriptions. A fundamental principle is that the correspondence of $P$-nodes and $M$-nodes in two descriptions is constrained by the correspondence of $P$-nodes and $M$-nodes at coarser resolutions in the same $P$-path.

As an example of correspondence matching using this representation, this section shows the process of discovering the correspondence between the coarsest resolution $P$-nodes in two images of a teapot taken with a change in distance between the teapot and the camera by a factor of 1.36. In this example matching is shown for the $P$-nodes from the most global level (level 12) to the second highest level with more than one $P$-node.

The first image is referred to as teapot image 1. This is the image whose sampled DOLP transform is shown in the examples in Figs. 2 and 3. The $P$-nodes for levels 12 through 6 of teapot image 1 were hand matched to those of the second teapot image, referred to below as teapot 2. Other examples of $M$-node matching for the teapot images are given in [13].

### A. Abstracting the Graph of Connected Peaks at a Level

The algorithms described above are all presented from the point of view of having data which are "embedded" in the DOLP space. To obtain a description of gray-scale shape which is general purpose, it is desirable to construct a graph which not embedded in the DOLP space. Such a description may be stored with much less memory.

The primary skeleton of such a description is the tree of $P$-paths and the interconnecting $L$-paths. The $P$-nodes at each bandpass level are linked to other $P$-nodes of the same sign and level which are part of the same form. This linking is provided by tracing the $R$-paths that connect $P$-nodes at a level. Each link is encoded as a two-way pointer between $P$-nodes.

Each $P$-node and $M$-node has attributes of its DOLP sample value and its position $(x, y, k)$ in the DOLP space. Connected $P$-nodes are "linked" by two way pointers. Each half of a pointer may also be assigned the attributes of distance ($D$) and orientation ($\theta$), which are defined as follows.

*Distance:* The distance between two $P$-nodes is the cartesian distance measured in terms of the number of samples at that level. In levels with a $\sqrt{2}$ sample grid, the distance along the $x$ and $y$ axes are in units of $\sqrt{2}$.

*Orientation:* The orientation between two $P$-nodes is the angle between the line that connects them and the $x$ axis in the positive direction.

The attributes of distance orientation are useful for determining the correspondence between small groups of $P$-nodes from two DOLP transforms.

*Example of Abstracted P-Nodes and R-Paths:* The $P$-nodes and $R$-nodes from level 7 of the teapot image are shown above in Fig. 6. Level 7 is the highest level with more than one $P$-node describing the teapot. The three positive peaks from level 7 of the teapot image are shown abstracted from the bandpass data in Fig. 8. The $R$-path links between these $P$-nodes are illustrated with arrows and labeled with circled numbers, called "link numbers." Links 1 and 2 are examples of "directly" connected $P$-nodes. A pair of $P$-nodes are directly connected when they are connected by an $R$-path with no intervening $P$-nodes between them. The $R$-path link between the rightmost and leftmost $P$-nodes is shown as a dotted arrow labeled as link 3. Link 3 shows an example of a pair of "indirectly" connected $P$-nodes. Including indirect $R$-path links in matching $P$-nodes prevents the matching algorithm from errors caused by missing or extraneous $P$-nodes.

In this early matching experiment, special status was given to the $P$-nodes along the "principal $P$-path." This is the $P$-path which includes the highest $M$-node. Thus arrows and indirect links are shown emanating from the $P$-node from this $P$-path.

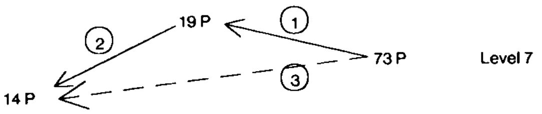


Fig. 8. Positive $P$-nodes and $R$-paths for level 7 of the teapot image.

TABLE I
$R$-PATH LINKS FOR LEVELS 7 AND 6 OF THE FIRST TEAPOT

| $R$-Path | Level | $dx$ | $dy$ | $D$ | $\theta$ |
|---|---|---|---|---|---|
| 1 | 7 | $-6$ | $-2$ | 6.32 | 161.5° |
| 2 | 7 | $-5$ | 3 | 5.83 | 210.9° |
| 3 (1&2) | 7 | $-11$ | 1 | 11.04 | 185.2° |
| 4 | 6 | $-4.0\sqrt{2}$ | $-2.0\sqrt{2}$ | 6.32 | 153.4° |
| 5 | 6 | $-3.25\sqrt{2}$ | $1.5\sqrt{2}$ | 5.06 | 205.8° |
| 6 | 6 | $-3.0\sqrt{2}$ | 0.0 | 4.24 | 180° |
| 7 | 6 | $0.25\sqrt{2}$ | $3.25\sqrt{2}$ | 4.6 | 265.6° |
| 8 (4&5) | 6 | $-7.25\sqrt{2}$ | $-0.5\sqrt{2}$ | 10.2 | 176.1° |
| 9 (4&5&6&7) | 6 | $-10\sqrt{2}$ | $2.75\sqrt{2}$ | 14.6 | 195.3° |

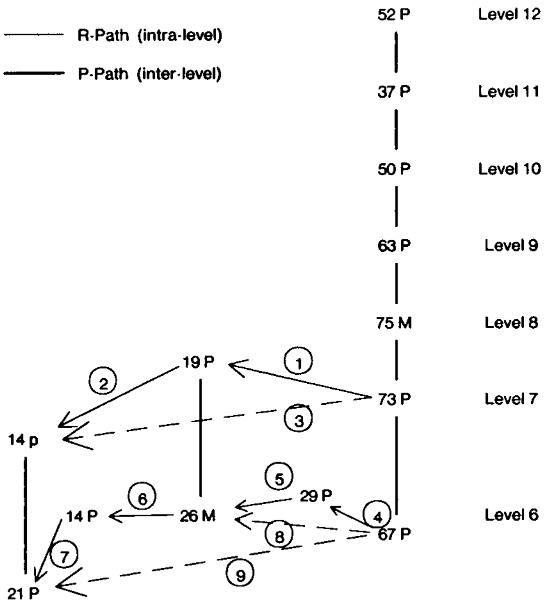


Fig. 9. $P$-nodes and $P$-paths for levels 12 to 6 of the smaller teapot image (teapot 1).

In our more recent experiments, all links are two-way, and indirect links are made for all $P$-nodes which are not at the top of a $P$-path.

The link numbers are also used as an index into a table of attributes. The attributes for these particular links are given in Table I in the next section. This same set of links is included in Fig. 9. These numbers are also used to show the correspondence which was assigned by hand matching between these links and the same links in the larger teapot image.

These attribute tables give the values for $dx$, $dy$, $D$, and $\theta$ for each $R$-path link. The positive directions for $dx$ and $dy$ are the same as used in the image: $+x$ points right, $+y$ points down. Note that $\theta$ increases in the counterclockwise direction. In these tables, in the levels which are at a $\sqrt{2}$ sample grid, the distances $dx$ and $dy$ are recorded in units of $\sqrt{2}$. In cases where a $P$-node spans two adjacent samples, the $P$-nodes position is assigned at the midpoint between them. This results in values of $dx$ or $dy$ that have fractional parts of 0.5 in the Cartesian-sampled (odd) levels, and 0.25, 0.5, or 0.75 in the $\sqrt{2}$ -sampled (even) levels.

TABLE II
$R$-PATH LINKS FOR LEVELS 8 AND 7 OF THE SECOND TEAPOT
(SCALED LARGER IN SIZE BY 1.36)

| $R$-Path | Level | $dx$ | $dy$ | $D$ | $\theta$ |
|---|---|---|---|---|---|
| 3 | 8 | $-7.5\sqrt{2}$ | $1.5\sqrt{2}$ | 10.81 | 191.3° |
| 4 | 7 | $-3.5$ | $-6.0$ | 6.94 | 149.7° |
| 5 | 7 | $-4.0$ | 1.0 | 4.12 | 194.0° |
| 6 | 7 | $-4.5$ | 1.0 | 4.61 | 192.0° |
| 7 | 7 | $-0.5$ | 5.0 | 5.02 | 264.3° |
| 8 (4&5) | 7 | $-10.0$ | $-1.5$ | 10.11 | 171.5° |
| 9(4&5&6&7) | 7 | $-15$ | 3.5 | 15.4 | 193.1° |

In Tables I and II, orientation ($\theta$) is measured in degrees. On a Cartesian grid, at distances that are typically 5 to 10 pixels, angular resolution is typically 5 to 10 degrees. Of course, the longer the distance, the more accurate the estimate of orientation.

The $P$-nodes for levels 12 through 6 of the teapot image are shown in Fig. 9. In levels 12 through 9 of Fig. 9 only a single $P$-node occurs in the teapot. These $P$-nodes all occur within a distance of two samples of the $P$-node above them, and are thus linked into a single $P$-path.[4] This $P$-path is referred to as the principal $P$-path. The $P$-node at level 8 has the largest value along this $P$-path and is thus marked as an $M$-node. This $P$-node corresponds to a filter with a positive center lobe of radius $R_+ \approx 18$ pixels or a diameter of 37 pixels. This corresponds to the form in the image that results from the overlap of the shadow on the right side of the teapot and the darkly glazed upper half of the teapot.[5] At level 7, additional detail begins to emerge. $P$-nodes occur over the upper right corner of the teapot and over the handle region. These $P$-nodes are joined to the $P$-node on the principal $P$-path by an $R$-path.

Five $P$-nodes occur in level 6. Three of these $P$-nodes occur underneath (within 2 samples of) $P$-nodes from level 7. These three $P$-nodes are thus part of three $P$-paths. The remaining two $P$-nodes are in fact the highest levels of two more $P$-paths. The $P$-path that begins at level 12 is referred to as the principal $P$-path. Only the indirect links between the principal $P$-path and a subset of the other $P$-nodes are shown in this figure and used in the matching example.

Note that an $M$-node occurs at level 6. This $M$-node corresponds to the upper left corner of the teapot and marks the left end of the dark region of glaze on the upper half of the teapot. The width of the positive center lobe of the filter which corresponds to this $M$-node gives an approximation of the width of the darkly glazed region.

### B. Initial Alignment to Obtain Size and Position

In matching two forms it is convenient to designate one form as a "reference form" and the other as a "data form." One then speaks of rotating, translating, and scaling the reference form so that its elements are brought into correspon-

[4] The $P$-path links appear as vertical dark lines in Fig. 9 although in fact there can be a lateral shift of up to two samples between their positions.

[5] The teapot images were digitized from negatives. Thus dark forms appear light in Figs. 2 and 3.

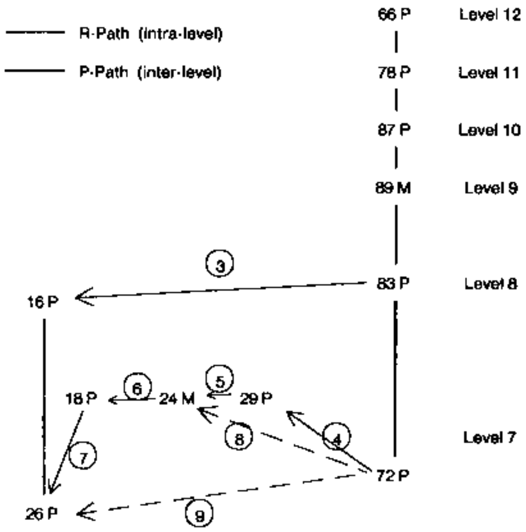


Fig. 10. $P$-nodes and $P$-paths for levels 12 to 7 of second (scaled larger in size by 1.36).

dence with the data form. In the examples presented below, teapot 1 is considered as the reference form which is transformed to match the teapot 2 (the data form).

Initial estimates of the alignment and relative sizes of two gray scale forms may be constructed by making a correspondence between their highest level $P$-nodes. This is illustrated by comparing the $P$-nodes and links in Fig. 9 to those in Fig. 10 shown below. Fig. 10 shows the $P$-nodes and $P$-path links for a teapot from a second image. This size scaling was accomplished by moving the teapot closer to the camera, and was thus accompanied by some changes in lighting. This second teapot is scaled larger in size by a factor of 1.36, which is just less than $\sqrt{2}$. The distance and orientation for each $P$-path link in this second teapot levels 12 through 7 is shown in Table II.

The highest level $M$-node in this second teapot occurs at level 9. The fact that this $M$-node is one level higher than the highest level $M$-node for teapot 1 confirms that this second teapot is approximately $\sqrt{2}$ larger than the first teapot.

The correspondence of the highest level $M$-nodes from these two teapots gives an estimate of the alignment of the two teapots as well as the scaling. The correspondence tells us the position at which the first teapot, scaled by $\sqrt{2}$ in size will match this second teapot. The tolerance of the initial position alignment is ± the sample rate at the level of the $M$-node in the data image. If this second teapot is designated as the data image, then the sample rate at level 9 determines the tolerance. The positioning tolerance at level 9 is $\pm 8\sqrt{2}$ pixels.

The tolerance of the size scaling is less than $\pm\sqrt{2}$. The correspondence of the highest level $M$-nodes provides an estimate of the size scaling factor which is a power of $\sqrt{2}$. Such an estimate is sufficient to constrain the correspondence process. A more accurate estimate can be obtained from the correspondence of higher resolution $P$-nodes and $M$-nodes.

### *C. Determining Further Correspondence and Orientation*

The matching process starts by finding the correspondence for the highest level $M$-nodes. This provides the process with an initial estimates of the size and position of the two forms. The next step is to find the correspondence of lower level $P$-nodes and $M$-nodes to refine the estimates of relative size and position, discover the relative orientations, and discover where one of the forms has been distorted by parallax or other effects.

Let us continue with our example. A $P$-node for the upper left corner of this second teapot does not occur. The change in scale from the first teapot to this second teapot was not enough to bring this $P$-node up to level 8. This may also be a result of the slight difference in shading that resulted from moving the teapot with respect to the lights and camera in order to size scale the object. Such errors are a natural result of changing the relative position between the camera and objects. A matching algorithm must tolerate them to be useful. The fact that the $P$-node of value 16 in level 8 of this second teapot corresponds to the $P$-node of value 14 in level 7 of the first teapot must be discovered from the position relative to their principal $P$-nodes and the distance and orientation from the $P$-node on the principal $P$-path at the same level.

The values for $D$ and $\theta$ for the link attributes in levels 7 and 6 of teapot 1 are compared to the attributes in the corresponding links from levels 8 and 7 of teapot 2 in Table III. All of these links are constrained to begin and end at samples in their respective levels. Because we are dealing with distances of between 4 and 15 samples at arbitrary angles, there is quantization noise in these attributes. The differences in orientation are shown in the column labeled $\theta_1 - \theta_2$. Except for link 3, these values show a consistent small rotation in the counterclockwise direction for the links from teapot 2. A careful measurement of the angle between the line connecting two landmarks and the raster line in the two images confirms that the two teapots actually have a relative change in orientation of approximately 3.3°. The actual values of $\theta$ fluctuate more than this due to quantization error from sampling and changes in shading.

The ratio $D_2/D_1$ shows a factor by which the lengths consistently shift when the teapot is scaled by 1.36. Because the actual values of $D_2$ and $D_1$ are restricted to distances between discrete locations, there is some random error built into this ratio. Since this shift in scale was enough to drive the corresponding $R$-paths in this second teapot up to a new level, but less than the $\sqrt{2} = 1.41$ scale change between levels, an average ratio of $D_2/D_1 = 1.36/1.41 = 0.96$ was anticipated. In Table III we see that this average ratio worked out to 1.02. Our conclusion is that quantization noise and changes in shading accounted for most of this difference. The actual differences in length, $D_2 - D_1$, show that the lengths are always within one sample: except for link 5, the percentage differences $(D_2 - D_1)/D_2$ are generally small ($\leqslant$ 10 percent). The conclusion from this experiment is that the correspondence between $R$-nodes from similar gray-scale forms of different sizes can be found, provided that the matching tolerates variations of the lengths of $R$-paths of up to 25 percent and variations in the relative angles of up to 12°.

## V. Comments

The representation for gray-scale shape which is formed by detecting peaks and ridges in a resampled DOLP transform resembles the representation provided by a medial axis transform (MAT) described by Blum [5]. There are, however, several important differences. It is worth while to compare these two representations and examine their similarities and differences.

TABLE III
COMPARISON OF $D$ AND $\theta$ ATTRIBUTES FOR TEAPOTS 1 AND 3

| | Teapot 1 | | Teapot 2 | | | Difference | | |
|---|---|---|---|---|---|---|---|---|
| $R$-Path | $D_1$ | $\theta_1$ | $D_2$ | $\theta_2$ | $\theta_1 - \theta_2$ | $D_2/D_1$ | $D_2 - D_1$ | $100 \times (D_2 - D_1)/D_2$ |
| 3 | 11.09 | 185° | 10.8 | 191° | -6° | 0.974 | -0.2 | -1.8% |
| 4 | 6.3 | 153° | 6.9 | 148° | 5° | 1.095 | 0.6 | 8.7% |
| 5 | 5.1 | 206° | 4.1 | 194° | 12° | 0.804 | 1.0 | 24.4% |
| 6 | 4.2 | 180° | 4.6 | 192° | 12° | 1.09 | 0.4 | 8.7% |
| 7 | 4.6 | 266° | 5.2 | 264° | 2° | 1.13 | -0.6 | -11.5% |
| 8 | 10.2 | 176° | 10.1 | 171° | 5° | 0.99 | -0.1 | -1.0% |
| 9 | 14.6 | 195° | 15.4 | 193° | 2° | 1.05 | 0.8 | 5.2% |
| Average Error | | | | | 4.57° | 1.020 | 0.257 | 4.3% |

### A. Comparison with Blum's Medial Axis Transform

The MAT (or grass fire transform) is a technique for deriving a spine for a binary shape. The transform is defined as follows: every point on the boundary of the binary shape simultaneously emits a circular wave. The waves propagate in such a manner that waves do not flow through each other. When waves meet head on, they cancel. The point at which they cancel is marked as a point on the MAT spine of the shape. By propagating the waves in discrete time units, and keeping track of the time at which waves cancel, the spine may be encoded with the distance to the boundary. An axis occurs inside every concave curve, whether it is inside of a shape or not.

Rosenfeld [27] has shown a fast two-pass operator which will implement the grass fire transform. This operator is significant on its own right because it makes possible the matching technique of "chamfer matching" [6].

There are at least two fundamental problems which prevent the spine from a MAT from being useful for describing gray-scale shape. The first of these is that the transform only exists for binary shapes. The second problem, first pointed out by Agin [2], is that a small narrow concavity in the boundary will significantly alter the shape of the resulting spine. Similar effects can occur from many other types of noise patterns. Thus the transform and the spine are very sensitive to noise.

In contrast, the representation given by peaks and ridges in a DOLP transform is a representation for gray-scale shape instead of binary shape. The DOLP bandpass filters have a circular positive center lobe which is a best fit to the gray-scale pattern when the DOLP value is large. Thus, as with the MAT spine, the DOLP ridges tend to exist where a circle is a best fit to the pattern. However, the DOLP bandpass filters have a smoothing effect; they are only sensitive to patterns at narrow range of sizes (spatial frequencies). Thus a narrow concavity is described in detail by small DOLP filters, the concavity has almost no effect on the ridge given by large DOLP filters.

The representation given by peaks and ridges in the DOLP transform has many other properties which a MAT spine does not have: For example, there is the existence of a largest peak as a landmark for matching, the fact that the representation can be used to guide matching from course resolution to high resolution, and the important property that the configuration of peaks and ridges can be matched when the pattern occurs at any size.

## VI. SUMMARY AND CONCLUSION

The principal topic of this paper is a representation for gray-scale shape which is composed of peaks and ridges in the DOLP transform of an image. Descriptions of the shape of an object which are encoded in this representation may be matched efficiently despite changes in size, orientation, or position by the object. Such descriptions can also be matched when the object is blurry or noisy.

The definition of the DOLP transform was presented, and the DOLP transform was shown to be reversible. A fast algorithm for computing the DOLP transform based on the techniques of resampling and cascaded convolution with expansion was then described. This fast algorithm is described in greater detail in [14]. This section concluded with an example of the DOLP transform of an image which contains a teapot.

A representation for gray-scale form based on the peaks and ridges in a DOLP transform was then described. This representation is composed of four types of symbols: $\{M, P, L, R\}$. The symbols $R$ and $P$ (ridge and peak) are detected within each DOLP bandpass image. $R$-nodes are samples which are local positive maxima or negative minima among three contiguous DOLP samples in any of the four possible directions. $P$-nodes are samples which are local positive maxima or negative minima in all four directions. $P$-nodes within the same form in a bandpass level are connected by a path of largest $R$-nodes, called an $R$-path (or ridge). An $R$-path is formed by having each $R$-node make a pointer to members of its local neighborhood which are also $R$-nodes and local maxima within a linear list of the neighborhood. $P$-nodes are connected with nearby $P$-nodes at adjacent bandpass levels to form $P$-paths. The skeleton of the description of a form is a tree composed to $P$-paths.

The DOLP values along each $P$-path rise monotonically to a maximum in magnitude and then decrease. The maximum magnitude DOLP sample along a $P$-path is marked as an $M$-node. $M$-nodes serve as landmarks for matching, and provide an estimate of the position and orientation of a form in an image. If the values along an $R$-path are compared to the values along the $R$-paths at nearby locations in adjacent bandpass images, an $R$-path of largest DOLP samples can be detected. These samples are marked as $L$-nodes, and these nodes form an $L$-path. $L$-paths begin and end at $M$-nodes and describe elongated forms. Thus, descriptions in this representation have the structure of a tree composed of $P$-paths, with a distinguished *M*-node along each. The *P*-nodes in each level are connected by *R*-paths, and the *M*-nodes are connected by *L*-paths which can travel among as well as within the levels.

The teapot image was used to illustrate the construction of a description in this representation. In this illustration, the *R*-nodes and *P*-nodes from bandpass level 7 from the DOLP transform of the teapot and the pointers between these *R*-nodes were displayed.

The final section of the paper presented a description and examples of the problem of determining the correspondence between the *M*-nodes and *P*-nodes in two descriptions of the same object. A description of a second teapot image, in which the teapot had been moved so as to be scaled larger by 1.36, was used to illustrate the principles of matching such descriptions. In both teapot images, the *P*-paths, *R*-paths, and *M*-nodes from the coarsest resolution bandpass images were presented. Matching to determine the correspondence of *L*-paths was not described in this paper. Such matching is described in [13].

The teapot matching examples first illustrated the correspondence of the coarsest resolution *M*-nodes in the two descriptions. This correspondence provides an estimate of the position and size at which the two teapot description best match. The principle that *P*-nodes in two descriptions can only correspond if the *P*-nodes above them correspond was also illustrated. An example was then provided for the use of the lengths and directions of the *R*-paths that connect *P*-nodes at each level to further determine correspondence when new *P*-paths are introduced and the orientation has not been determined.

This example addresses only a small part of the general problem of matching descriptions of objects. The problem of matching two descriptions of an object with large differences in image plane orientation was not illustrated. An example of such matching is provided in [13]. The more difficult problems of matching in the presence of motion of either the camera or the object was not discussed. Such matching must be robust enough to accommodate the changes in two-dimensional shape that occur with a changing three-dimensional viewing angle. Similarly, the problems of forming and matching to a prototype for a class of objects was not discussed. We believe that this representation will provide a powerful structural pattern recognition technique for recognizing objects in two-dimensional domain and for dynamically constructing a three-dimensional model of a three-dimensional scene.

## Acknowledgment

We would like to dedicate this paper to F. Quick, who first posed many of the questions which led us to develop the DOLP transform and this representation. We would also like to thank the Dr. R. Reddy and the Carnegie-Mellon University Department of Computer Science who provided computer time and disk space for this research. Their support permitted this research to continue at a critical time during which there was no funding. We would also like to thank R. Stern, T. Kanade, and A. Sanderson who generously provided intellectual critiques and emotional support during the course of this research. Finally, we would like to thank the Westinghouse Electric Corporation which has recently made it possible to continue this research from the point described here.

## References

[1] J. F. Abramatic and O. D. Faugeras, "Sequential convolution techniques for image filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 1-10, Feb. 1982.

[2] G. J. Agin, "Representation and description of curved objects," Ph.D. dissertation, Stanford Univ., 1971.

[3] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *Computer Science and Information Processing: The Design and Analysis of Computer Algorithms.* Reading MA: Addison-Wesley, 1974.

[4] T. O. Binford, "Survey of model-based image analysis systems," *Robotics Res.*, vol. 1, pp. 18-64, Spring 1982.

[5] H. Blum, "A transformation for extracting new descriptors of shape," in *Models for the Perception of Speech and Visual Form.* Cambridge, MA: MIT Press, 1967.

[6] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," SRI Int., Tech. Note 153, 1978.

[7] P. J. Burt, "Fast, hierarchical correlations with Gaussian-like kernels," Comput. Vision Lab., Univ. Maryland, Tech. Rep. TR-860, Jan. 1980.

[8] —, "Fast filter transforms for image processing," *Comput. Graphics Image Processing*, vol. 16, pp. 20-51, 1981.

[9] F. W. Campbell, *The Transmission of Spatial Information through the Visual System.* Cambridge, MA: MIT Press, 1974.

[10] F. W. Campbell and J. G. Robson, "Applications of Fourier analysis to the visibility of gratings," *J. Physiol.*, pp. 551-566, 1971.

[11] J. L. Crowley and A. C. Parker, "The analysis, synthesis, and evaluation of local measures for discrimination and segmentation of textured regions," in *Proc. Conf. Pattern Recognition and Image Processing*, IEEE Comput. Soc., June 1978, pp. 372-378.

[12] J. L. Crowley and A. C. Parker, "Transfer function analysis of picture processing operators," in *Issues in Digital Image Processing*, R. M. Haralick and J. C. Simon, Eds. Sijthoff & Noordhoff, 1980, ch. 1, pp. 3-30.

[13] J. L. Crowley, "A representation for visual information," Ph.D. dissertation, Carnegie-Mellon Univ., Nov. 1981.

[14] J. L. Crowley and R. M. Stern, "Fast computation of the difference of low-pass transform," *IEEE Trans. Pattern Anal. Machine Intell.*, to be published.

[15] E. L. Hall, L. J. Rouge, and R. Y. Wong, "Hierarchical search for image matching," in *Proc. 1976 IEEE Conf. Decision and Control*, Dec. 1976, pp. 791-796.

[16] A. R. Hanson and E. M. Riseman, *Computer Vision Systems.* New York: Academic, 1978, pp. 758-768.

[17] M. D. Kelly, "Edge detection in computers by computer using planning," in *Machine Intelligence*, B. Meltzer and D. Mitchie, Eds. Halsted, 1971.

[18] D. Marr and T. Poggio, "A computational theory of human vision," *Proc. Royal Soc. London B*, 1979.

[19] D. L. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Soc. London B*, vol. 207, pp. 187-217, 1980.

[20] D. Marr, *Vision.* San Francisco, CA: W. H. Freeman, 1982.

[21] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Ph.D. dissertation, Stanford Univ., Sept. 1980.

[22] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, vol. 3, pp. 324-346, Apr. 1924.

[23] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

[24] R.E.A.C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain.* New York: Amer. Math. Soc. Coloq. 1934.

[25] A. Papoulis, *Systems Sciences: Systems and Transforms with Applications in Optics.* New York: McGraw-Hill, 1968.

[26] W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978, p. 322.

[27] A. Rosenfeld and J. L. Pflatz, "Distance functions on digital pictures," *Pattern Recognition*, vol. 1, pp. 33-62, July 1968.

[28] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template

matching," *IEEE Trans. Syst., Man., Cybern.*, vol. SMC-7, pp. 104-107, Feb. 1977.

[29] M. Sachs, J. Nochmias, and J. G. Robson, "Spatial-frequency channels in human vision," *J. Opt. Soc. Amer.*, vol. 61, pp. 1176-1186, 1971.

[30] S. L. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," *Comput. Graphics Image Processing*, vol. 4, pp. 165-174, June, 1975.

[31] J. P. Thomas, *Spatial Resolution and Spatial Interaction.* New York: Academic, 1975.

[32] L. Uhr, "Layered recognition cone networks that preprocess, classify, and describe," *IEEE Trans. Comput.*, vol. C-21, pp. 758-768, 1972.

**James L. Crowley** (S'73-M'82) received the B.S.E.E. degree from Southern Methodist University, Dallas, TX, in 1975, and the M.S.E.E. and Ph.D. degrees from Carnegie-Mellon University, Pittsburgh, PA, in 1977 and 1982, respectively.

He is currently a Research Scientist in the Robotics Institute at Carnegie-Mellon University, where he is directing research projects in three-dimensional scene analysis and in mobile robot navigation. Since joining the Robotics Institute in 1980 he has completed three projects dealing with measuring, representing, and matching three-dimensional shapes for industrial applications. His research interests include techniques for shape representation, 3-D scene analysis, and mobile robot navigation.

**Alice C. Parker** (S'68-M'75) received the M.S.E.E. degree from Stanford University, Stanford, CA, in 1971 and the Ph.D. degree from North Carolina State University, Raleigh, in 1976.

Currently, she is an Associate Professor in the Department of Electrical Engineering-Systems at the University of Southern California. Prior to that she was an Assistant Professor of Electrical Engineering at Carnegie-Mellon University, Pittsburgh, PA. Her research interests include automatic synthesis and optimization of digital hardware, hardware-descriptive languages, and expert systems for VLSI design.