5th International Conference on Artificial Intelligence (ICAI-V)
Cancun, Mexico December 1992

# INTEGRATING SHAPE FROM SHADING IN A GRADIENT HISTOGRAM AND ITS APPLICATION TO ENDOSCOPE NAVIGATION

**L. Enrique Sucar, Duncan F. Gillies**
Department of Computing
Imperial College of Science, Technology and Medicine
180 Queen's Gate
London SW7 2BZ
England

**Haroon U. Rashid**
National Institute of Automation and Control
Islamabad
Pakistan

## ABSTRACT

Shape from shading consists in finding the shape of an illuminated opaque object from a shaded video image. Research in this area started with the work of Horn and has developed into an important area in low-level vision for the last 15 years. But there have not been many practical applications of this research. In this paper we develop a method for integrating the local shape from shading information obtained form a single image into a *gradient histogram*. This allows us to summarize the local shading data in a global structure, opening the door for practical applications of shape from shading techniques.We illustrate this technique in a medical application, using it for endoscope navigation in colonoscopy. For this have implemented the histogram technique in a parallel architecture, and applied it to finding the *lumen* in real colon images. We have tested it in a wide sample of colon images and the results are encouraging. We consider that there could be other applications in computer vision in which this technique could be useful.

## INTRODUCTION

### Endoscopy

Endoscopy is one of the tools available for diagnosis and treatment of gastrointestinal diseases. It allows a physician to obtain direct information of the inside surface of the human digestive system. Beside its diagnostic capabilities, endoscopes have therapeutic applications. They allow the removal of colonic polyps and other foreign bodies, and direct attack on bleeding lesions The endoscope is a flexible tube with viewing capability. It consists of a flexible shaft which has a manoeuvrable tip. The orientation of the tip can be controlled by pull wires that bend it in 4 orthogonal directions (left/right, up/down). It is connected to a cold light source for illumination of the internal organs and has an optical system for viewing directly through an eye piece or on a TV monitor. The instrument has channels for transmitting air to distend the organ, for a water jet to clean the lens and for sucking air or fluid. Additionally, it has an extra "operating" channel that allows the passage of flexible instruments.

The consultant controls the instrument by steering the tip with two mechanical wheels, and by pushing or pulling the shaft . The shaft is relatively torque stable so that he can also apply rotatory movements to the tip. He can control the air supply (inflate or aspirate) for good vision but without creating excessive air pressure. He can use the water jet for cleaning the lens when it is dirty, and aspirate excess fluid. In addition to control actions he makes diagnostic decisions or therapeutic actions in each particular case. We are interested primarily in colonoscopy, which is especially difficult due to the complexity and variability of the human colon. The doctor inserts the instrument estimating the position of the colon centre (*lumen*) using several visual clues such as the darkest region, the colon muscular curves, the longitudinal muscle and others. If the tip is not controlled correctly it can be very painful and dangerous to the patient, and could even cause perforations on the colon wall. This is further complicated by the presence of many difficult situations such as the contraction and movement of the colon, fluid and bubbles that obstruct the view, pockets (*diverticula*) that can be confused with the *lumen* and the paradoxical behaviour produced by the endoscope looping inside the colon. This requires a high degree of skill and experience that only an "expert" endoscopist will have.

We are developing a navigation and advisory system for colonoscopy [6]. Its primary objective is to help the doctor with the navigation of the endoscope inside the colon by controlling the orientation of the tip via the right/left and up/down controls. The control of the translation of the instrument (push/pull) and possible shaft rotation will still be carried out manually by the doctor. As well as a navigation system, it will also serve as an advisory system for learning endoscopists suggesting correct actions. This will free the physician from some of the tasks he has to do concurrently and allow him to concentrate on the diagnostic and therapeutic aspects of colonoscopy.

### Image Interpretation for Endoscope Navigation

The first step for endoscope navigation and advice is to recognize the important features in the images, which can be used to guide the endoscope inside the human colon. Due to the type of illumination of the endoscope, the darkest region generally corresponds to the centre of the colon (*lumen*) because there is a single light source close to the camera.

*Figure 1.* Examples of colon images In (a) the endoscope is in a straight section of the colon, so there is a clear "dark region" that corresponds to the *lumen*. But in (b) the tip is close to the wall so there is not an obvious *lumen*, still we can infer the possible direction of the *lumen* (to the lower-right of the image) by the shading information.

(a) Dark region (lumen)

(b) Tip close to the wall

Khan [2] has developed a technique for extraction of the *dark region* in colon images. Although this algorithm finds the *lumen* in most of the colon images (fig. 1-a), it has problems when the tip of the endoscope is pointing towards the colon wall or close to a sharp bend (fig. 1-b). In these cases the algorithm could give an incorrect dark region, or no information at all. So for these "difficult" images, we need another way to extract the information from the image to find the direction of the *lumen*.

Rashid [5] developed an algorithm for the determination of shape from shading using a single image. He considers a point light source which is at the same point as the camera and near the surface of the object, which closely approximates the illumination arrangement inside the colon. Using a local method he developed a linear algorithm that can estimate the gradient or slope of small patches in the surface. The method is very efficient and could be implemented in parallel, so it has the potential for being used in endoscope navigation. The problem is how to use this shape information to find the *lumen*.

We have developed a technique to infer the approximate position of the *lumen* in a colon image based on local shape information. First we explain briefly the shape from shading algorithm. Then we develop a *gradient histogram* for finding the lumen by considering an ideal tube. We then show some results from real colonoscopy images. Finally we discuss how the gradient histogram can be implemented in a real-time using a parallel architecture, and integrated with other image features in the high-level part of the endoscope navigation system.

## SHAPE FROM SHADING

Shape from shading consists of finding the shape of an illuminated opaque object from a shaded video image. Research in this area started with the work of Horn [1]. He derived an image irradiance equation assuming uniform illumination (distant light source), orthographic projection and a *Lambertian* surface. He shows that the image intensity $E(x,y)$ depends on the surface orientation and its reflective characteristics:

$$E(x,y) = S_0\, \rho(x,y) \cos \theta_i \qquad (1)$$

where $S_0$ is the light source intensity constant, $\rho(x,y)$ is the reflection coefficient or albedo, and $\theta_i$ is the angle between the source direction and the surface normal. We can express (1) in vector notation as:

$$E(x,y) = S_0\, \rho(x,y)\, (\boldsymbol{n} \cdot \boldsymbol{s}) / |\boldsymbol{n}| \qquad (2)$$

where $\boldsymbol{n}(x,y)$ is the surface normal vector, $s$ is a unit vector which specifies the light source direction, and '·'

means dot product. In component form:

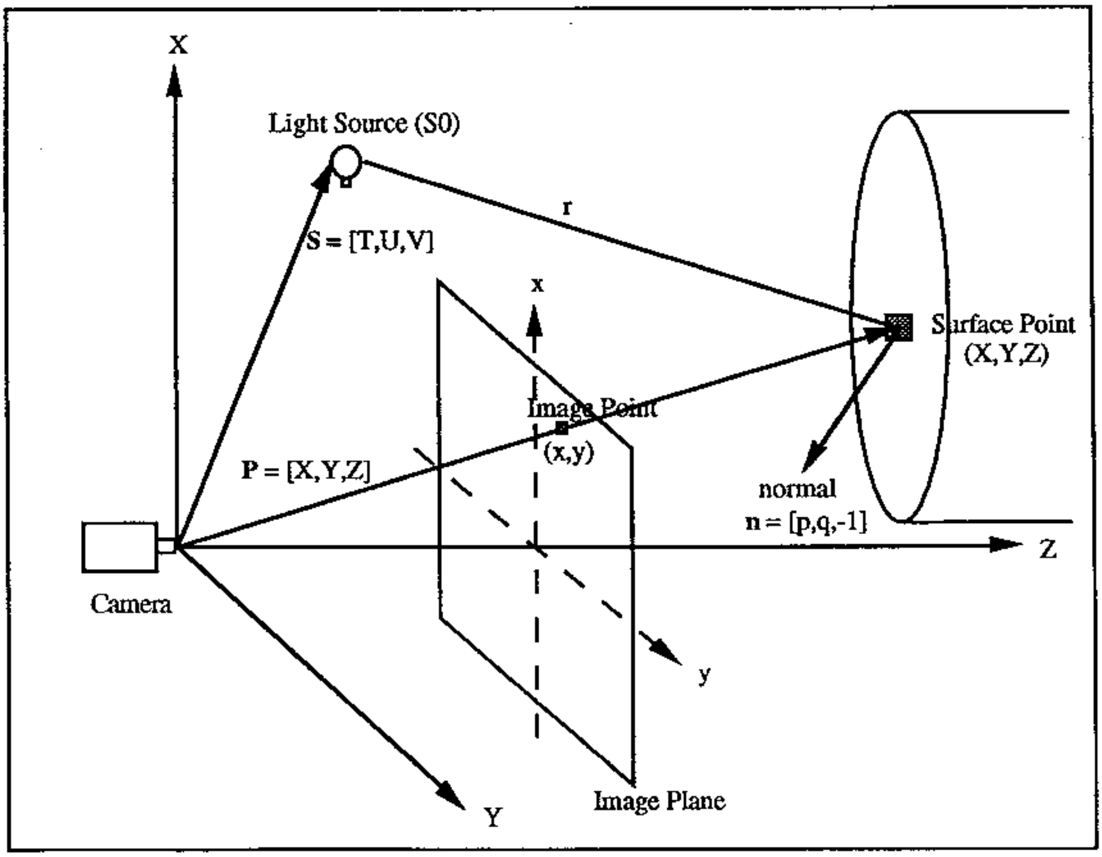
$$n = (p, q, -1), \qquad p = \partial Z/\partial X \text{ and } q = \partial Z/\partial Y$$
$$s = (t, u, v)$$

So (2) can be expressed as:

$$E(x,y) = S_0\, \rho(x,y)\ \frac{pt + qu - v}{\sqrt{p^2 + q^2 + 1}} \tag{3}$$

Assuming that $S_0$, $\rho$ and $s$ are all known, there is still only one equation with two unknowns ($p$ and $q$), so in general it is an ill-posed problem. It can be solved by using multiple light sources (photometric stereo technique) or multiple images.

*Figure 2.* World and image coordinate systems for a near point light source and perspective projection. The image plane is considered conceptually in front of the lens to avoid the perspective inversion.




Considering a near point light source illumination, now denoted by position vector $S$, introduces another factor. The intensity varies according to the inverse square distance between the light source and the surface point, so the surface absolute position ($Z$) is another parameter in this problem. For a near point light source (1) becomes:

$$E(x,y) = S_0\, \rho(x,y) \cos\theta_i / r^2 \tag{4}$$

where $r$ is the distance between the light source and the surface point. Considering a camera-centered coordinate system (fig. 2) we have $r = S - P$, where $S = [T, U, V]$ is the source position vector and $P = [X, Y, Z]$ is the surface position vector. Thus we can write (4) in vector notation as:

$$E(x,y) = S_0\, \rho(x,y)\ \frac{(S - P) \cdot n}{|n|\, |S - P|^3} \tag{5}$$

For orthographic projection we have $X = x$ and $Y = y$, thus (5) in component form is:

$$E(x,y) = S_0\, \rho(x,y)\ \frac{(T - x)p + (U - y)q - (V - Z)}{\sqrt{p^2+q^2+1}\ [(T-x)^2 + (U-y)^2 + (V-Z)^2]^{3/2}} \tag{6}$$

In the case of perspective projection (as in fig. 2) we substitute $x$ by $xZ/f$ and $y$ by $yZ/f$, where $f$ is the focal length of the camera.

Rashid [5] considered the case of a single light source at the origin (T=U=V=0) and perspective projection. So (6) is simplified to:

$$E(x,y) = S_0\, \rho(x,y) \frac{1 - xp/f - yq/f}{Z^2 \sqrt{p^2+q^2+1}\, [1 + (x/f)^2 + (y/f)^2]^{3/2}} \tag{7}$$

Normalizing by the focal length (replacing $x/f$ by $x$ and $y/f$ by $y$) we obtain:

$$E(x,y) = S_0\, r(x,y) \frac{1 - xp - yq}{Z^2 \sqrt{p^2+q^2+1}\, [1 + x^2 + y^2]^{3/2}} \tag{8}$$

Now we have three unknowns $p$, $q$ and $Z$ and still only one equation. To solve this problem Rashid [5] takes a local approach, similar to the one originally proposed by Pentland [4]. He considers a smooth lambertian surface which can be represented by small planar patches with constant albedo ($\rho_{av}$). In this case the surface can be approximated locally by its Taylor series expansion to the first degree as:

$$Z = Z_0 + \partial Z/\partial X|_{X_0,Y_0} (X - X_0) + \partial Z/\partial Y|_{X_0,Y_0} (Y - Y_0) \tag{9}$$

around some point $[X_0, Y_0, Z_0]$. Representing it in terms of image coordinates and perspective projection ($X=xZ$ and $Y=yZ$), and using the previous definitions of $p$ and $q$ we obtain:

$$Z = \frac{Z_0\, (1 - p_0x_0 - q_0y_0)}{1 - p_0x - q_0y} \tag{10}$$

Substituting in (8) we have:

$$E(x,y) = S_0\, \rho_{av} \frac{(1 - xp_0 - yq_0)^3}{Z_0^2 \sqrt{p_0^2+q_0^2+1}\,[1-x_0p_0-y_0q_0]^2[1+x^2+y^2]^{3/2}} \tag{11}$$

Rashid [5] obtains further information by considering the directional derivatives in the $x$ and $y$ directions, which can be obtained from the partial derivatives of the image irradiance $E$. The total differential change in image irradiance at an image point can be evaluated as:

$$dE = E_x\, dx + E_y\, dy,$$

where:

$$E_x = \partial E/\partial x \text{ and } E_y = \partial E/\partial y$$

By considering the normalized derivatives $R_x = E_x/E$ and $R_Y = E_y/E$, and from (11) he obtains two independent equations:

$$R_x = E_x/E = -3 \left[ \frac{p_0}{1-p_0x-q_0y} + \frac{x}{1+x^2+y^2} \right] \tag{12}$$

$$R_y = E_y/E = -3 \left[ \frac{q_0}{1-p_0x-q_0y} + \frac{y}{1+x^2+y^2} \right] \tag{13}$$

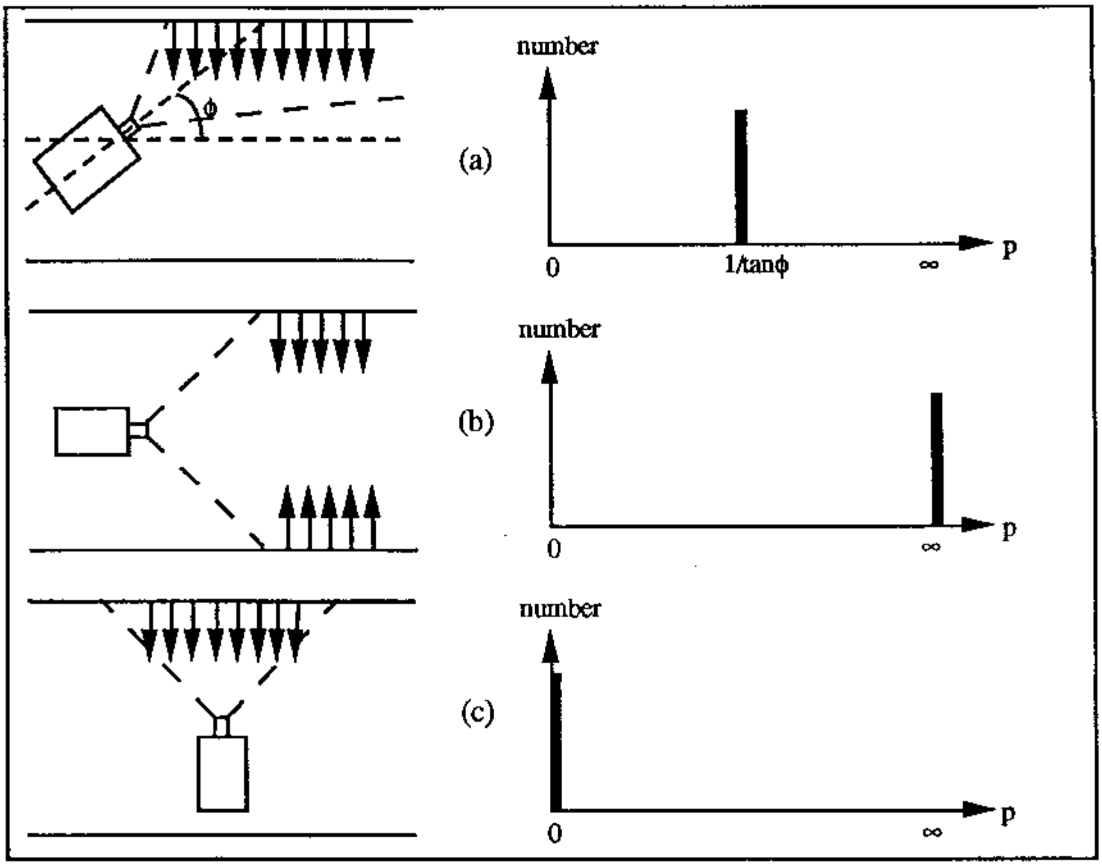
This is an important result in which we have two equations with two unknowns ($p$, $q$), and the absolute position ($Z_0$), average albedo ($\rho_{av}$), and light source ($S_0$) constant have been cancelled out. Equations (12) and (13) can be written as two linear equations with two unknowns, and can be solved directly to obtain the surface gradients $p$ and $q$ for each image point. We only need to know its intensity ($E$) and its gradient in two orthogonal directions ($E_x$ and $E_y$), assuming a smooth surface in which the albedo varies slowly.

This method provides a direct and fast way to obtain the orientation of a 3-D surface from an image. It assumes that the camera and light source are practically at the same point and close to the surface. This model is a good approximation to the situation in endoscopy, and the results from real colon images are good as we will see later. Next we consider how we can make use of this local shape information in a global way to find the *lumen*.

## GRADIENT HISTOGRAM

The depth map we obtain from the shape from shading algorithm consists of the surface gradient $[p,q]$ per pixel that gives the orientation of the surface at this point with respect to two orthogonal axis ($X$, $Y$) that are perpendicular to the camera axis ($Z$). This can be represented as the surface normal vector $\mathbf{n} = [p, q, -1]$, using the coordinate system in fig. 2. The surface orientation can vary from perpendicular to the camera with $p$ and $q$ equal to zero, to parallel to the camera with $p$ or $q$ close to infinity. If we assume that the colon has a shape similar to a tube and in the image a section of the internal wall of this tube is observed, then a reasonable approximation of the position of the centre of the colon (*lumen*) will be a function of the direction in which the majority of the normal vectors are pointing to. This is clear if we consider an infinite ideal tube (fig. 3) and take a cross section along the $X$ axis so we have only the $p$ component of the normal vectors. If there is an angle $\phi$ between the camera axis ($Z$) and the tube's main axis (fig. 3-a), then the normal vectors in the cross section will have all a $p$ value that corresponds to this angle $\phi$, .i.e. $p = 1/\tan(\phi)$. We can illustrate this if we construct a *p-histogram* as shown in figure 3, that represents the number of $p$ vectors for each slope. In this case the histogram will have "$n$" entries for $p$ such that $\tan(\phi) = 1/p$ and zero for all other values. There are two special cases we have to take into account. One is if the camera is at the centre of the tube and parallel to its axis (fig. 3-b) then all the normal vectors are parallel to the camera axis and $[p, q] \rightarrow \infty$. The other extreme is if the camera is close to the wall of the tube and pointing directly into it (fig. 3-c) so the normal vectors are all nearly perpendicular to the camera axis and $[p, q] \approx 0$.

*Figure 3.* Normal vectors and the corresponding *p-histogram* in a cross section of an ideal tube. Each case shows different orientations of the camera with respect to the tube's main axis. In (a) we depict the general case with an angle $\phi$ between them, and (b) and (c) the two extreme cases, with $\phi = 0$ and $\phi = 90°$, respectively.



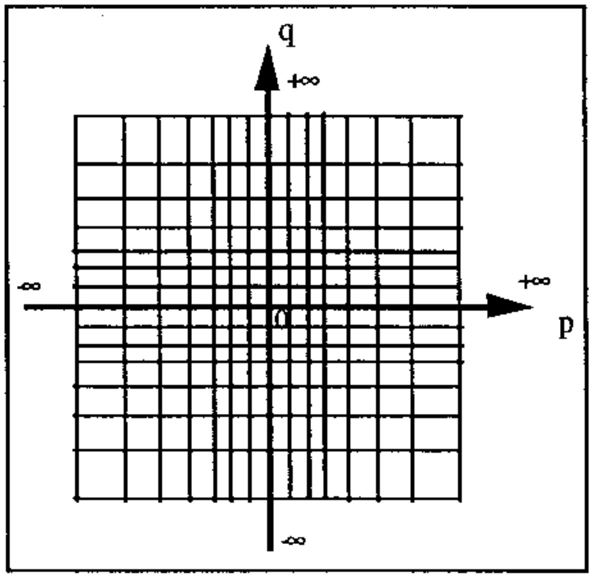

If we extend this analysis to the 3-D surface of the tube, we will have a distribution of $[p, q]$ which will not be a single value. The tube's curvature will produce different values for $[p, q]$, but if its ratio is relatively large respect to the distance from the camera, these values will be concentrated around the corresponding angle $\phi$. So from the $[p, q]$ distribution we can estimate $\phi$ and from that the relative position of the centre of the tube (the *lumen*). For this we divide the space of possible values of $[p, q]$ into a small number of slots and obtain a 2-D histogram of this space (fig. 4), which we denominate a gradient histogram or *pq-histogram*. From this histogram we find the values of [p, q] which occur more frequently, that is we find the largest peak in the histogram. The position of this peak will give an estimate of the position of the *lumen*, that is, there is a direct mapping between the location of peak in the histogram and that of the *lumen* in the image.

## APPLICATION TO ENDOSCOPE NAVIGATION

### Application of the Gradient Histogram

In the case of colonoscopy we do not have, of course, an ideal tube. The shape of the human colon is complex and variable, and there are irregularities caused by muscular transverse folds (rings), veins, polyps, etc. Yet its basic shape is similar to that of a tube, and the principle of the algorithm described in the previous section is applicable. These irregularities, and some incorrect [*p, q*] values due to image artifacts such as texture, edges and specularities, will distort the distribution of the *pq-histogram*. But we think that in most cases these will be reflected as "noise" in the histogram, and the main peak will still correspond to the direction of the *lumen*. That is, an advantage of this technique is that it is a global method, so the local variations do not have a big effect on the overall result.

*Figure 4.* Gradient (*pq*) Histogram. The *p* and *q* spaces are divided into a finite number of slots . The size of these slots increases exponentially from the origin to take into account the non-linear nature of the relation between [p,q] and $\phi$ (tangent function).




We have tested this technique with many real colon images with encouraging results. In figure 5 we present a typical case. It shows an example of a colon image (a) and the corresponding depth map represented as a needle diagram (b), the *pq-histogram* (c), and the position of the *lumen* obtained from the histogram (d).

In practice we have observed that this technique tends to give more reliable results when the *lumen* is toward the edge of the image, that is there is a significant deviation from the centre of the colon. When the lumen is at or near the centre the distribution is more uniform and the "peaks" are usually due to irregularities in the colon surface, so these are "false peaks". It is in these situation when there is a clear dark region in the image, and the dark region extraction algorithm [2] will give better results. So both techniques complement each other and if we can combine their results it should improve the reliability of *lumen* recognition. In the next section we discuss how we combine them in the high-level part of endoscope navigation system.

## High-level Processing

High-level vision seeks to find a consistent interpretation of the features obtained during low-level processing. It is based on recognition, that is matching our internal representation of the world with the sensory data obtained from the images. It involves the use of high-level specific models to infer from visual features the information required for subsequent tasks. Sucar [7] developed a framework for representing visual knowledge in high-level vision. It is based on a *probabilistic network* model that represents the visual knowledge in the domain. A probabilistic or Bayesian network represents the probability distribution of a set of variables and makes explicit the dependency information between them [3]. For a visual knowledge-base the network is divided into a series of layers, in which the nodes of the lower layer correspond to the feature variables and in the upper layer to the object variables. The intermediate layers have nodes for other visual entities, such as parts of an object or image regions; or represent relations between features/objects. The links point from nodes in the upper layers toward nodes in the lower layers, expressing a causal relationship. The network can be partitioned as a series of trees or *multitree* [7], with each tree having its root as one object. In this way we can do probability propagation in each tree independently to obtain the posterior probability of each object at the root of the tree.

We can think of the *pq-histogram* and dark region algorithms as two different ways of obtaining an *estimate* of the position of the *lumen*, which we should be able to combine in a probabilistic way. For representing this information in a probabilistic network we make two important considerations:

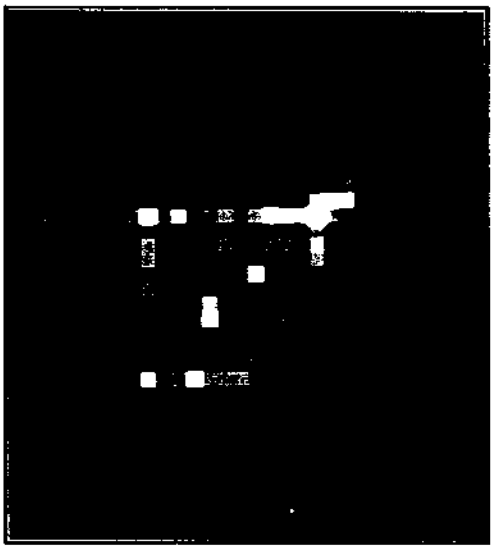
(a) The estimates provided by *pq-histogram* and dark region are *independent*.
(b) The proximity of the different estimates is an important factor for object recognition, so we add their *distance relation* as a node in the network [8].

Based on these assumptions, the probabilistic network for *lumen* recognition will have the structure given in fig. 6. The nodes below the shape estimate correspond to the measured parameters that affect the probability of it being correct: the number of *pq* vectors in the slot (*size*) and its location in the image (*position*).

*Figure 5.* Example of the use of shape information to estimate the *lumen* location. In (a) we show the original colon image, with the relative shape information as a needle diagram in (b). The p-q histogram in (c) is an "image" of the 2-D histogram, with the intensity being proportional to the number in each slot. Finally in (d) we depict the infered position of the lumen as a rectangle overlaid in the original image.

(a) Colon Image

(b) Depth map (needle diagram)



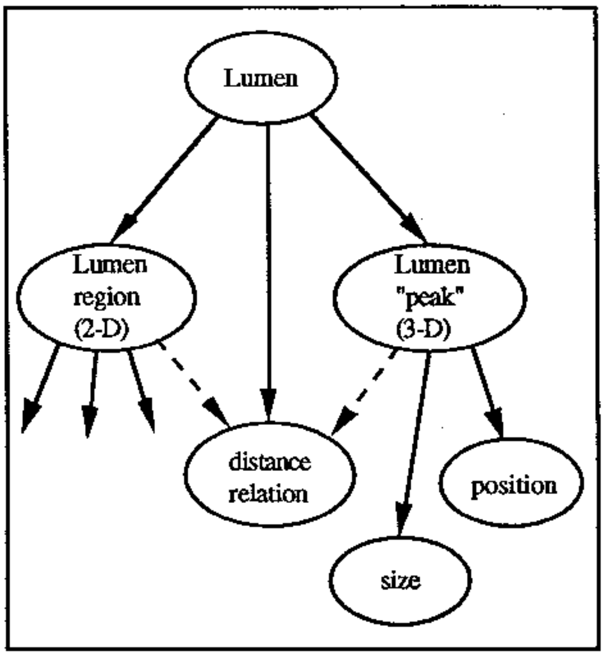
(c) pq-histogram

(d) Lumen location

Implementation

The prototype endoscope navigation system has been implemented on a parallel architecture based on a personal computer (PC) and five *Transputers* (IMS 800). The host is an *IBM-PC AT* compatible machine. The Transputers are on two PC boards, one with one Transputer and a high speed frame grabber and two 512x512 video buffers, and the other with four Transputers. The five Transputers are interconnected in a *pyramid* structure where the "root" Transputer communicates with the PC host. The root Transputer is integrated with the frame grabber which captures the images from a *υ-matic* videotape recorder and stores them in the video buffer. This set-up allows us to test the image interpretation and advisory systems with images recorded from colonoscopy sessions of many different patients.

*Figure 6.* Probabilistic tree for *lumen* recognition.




The local nature of the shape from shading algorithm lends itself to a parallel implementation in a pyramid architecture. For this the image is divided into $N * N$ quadrants, where $N$ is the number of levels in the pyramid. The shape shading algorithm is applied independently in each quadrant and a *local pq-histogram* is obtained. Then each processor sends its local histogram to its parent processor in the pyramid, which just accumulates the histograms it receives. This is done successively until the root is reached, which will have the global histogram. The root processor just has to find the peak in the histogram and send it to high-level vision. In our current implementation we have a two level pyramid, with 4 son transputers and one at the root. So the image is divided in 4 quadrants which are processed in parallel by the son transputers. With this implementation we have a processing rate of approx. 5 images/second for the pq-histogram algorithm alone, and 3 images/second with both algorithms.

## CONCLUSIONS AND FUTURE WORK

We have developed a method for integrating shape from shading information in a gradient histogram. This allows us to summarize the local shading data in a global structure, opening the door for practical applications of shape from shading techniques. One application is for endoscope navigation in colonoscopy. For this have implemented the histogram technique in a pyramid architecture, and applied it to finding the *lumen* in real colon images. We have applied it to a wide sample of colon images and the results are encouraging. We consider that there could be other applications in computer vision in which this technique could be useful.

We are currently implementing the high-level part of the system to integrate the features obtained from the different feature extraction processes. We expect that this will enable us to have a more reliable system that can be useful to the endoscopists. At the same time we are working on two aspects to improve the gradient histogram technique: pre-processing of the image and post-processing of the histogram. In the first aspect, we want to filter the high gradient parts of the image and the specularities, so that only smooth lambertian patches will remain for the shape from shading algorithm. In the post-processing aspect, we are considering different ways to take into account the distribution of the histogram so that "false" peaks can be eliminated.

## REFERENCES

[1] Horn, B., "Obtaining Shape from Shading Information", in *The Psychology of Computer Vision*, P. Winston (Ed.), McGraw-Hill, New York, N.Y., (1975), 115-156.

[2] Khan, G.N. and Gillies, D.F., "A Highly Parallel Shaded Image Segmentation Method", in P.M. Dew and R.A. Earnshaw (Eds.), *Parallel Processing for Vision and Display*, Addison-Wesley, England, (1989), 180-189.

[3] Pearl, J., "On Evidential Reasoning in a Hierarchy of Hypothesis", *Artificial Intelligence* 28, (1986), 9-15.

[4] Pentland, P.A., "Local Shading Analysis", *IEEE Trans. PAMI* 6, No. 2, (1984), 170-187.

[5] Rashid, H. and Burger, P., "Differential Algorithm for the Determination of Shape from Shading Using a Point Light Source", *Image and Vision Computing* 10, No. 2, (1992), 119-127.

[6] Sucar, L.E. and Gillies, D.F., "Knowledge-Based Assistant for Colonoscopy" in *Proc. Third International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems* II, ACM, (1990), 665-672.

[7] Sucar, L.E., Gillies, D.F., and Gillies, D.A., "Handling Uncertainty in Knowledge-Based Computer Vision", in R. Kruse and P. Siegel (Eds.), *Symbolic and Quantitative Approaches to Uncertainty*, Springer-Verlag, Berlin, (1991), 328-332.

[8] Sucar, L.E. and Gillies, D.F., "Representing Relational and Temporal Knowledge in Visual Probabilistic Networks", *Proc. Eighth Conf. on Uncertainty in Artificial Intelligence*, Morgan-Kaufmann, (1992), 303-309.