Automated Endoscope Navigation and Advisory System from medical imaging

Chee Keong Kwoh[a], Gul Nawaz Khan[a], Duncan Fyfe Gillies[b]

[a]School of Applied Science, Nanyang Technological University,
Blk N4, #2A-36, Nanyang Avenue, Republic of Singapore 639798

[b]Department of Computing, Imperial College of Science, Technology and Medicine,
180 Queen's Gate, London SW7 2BZ, UK

ABSTRACT

In this paper, we present a review of the research conducted by our group to design an automatic endoscope navigation and advisory system. The whole system can be viewed as a two-layer system. The first layer is at the signal level, which consists of the processing that will be performed on a series of images to extract all the identifiable features. The information is purely dependent on what can be extracted from the 'raw' images. At the signal level, the first task is performed by detecting a single dominant feature, lumen. Few methods of identifying the lumen are proposed. The first method used contour extraction. Contours are extracted by edge detection, thresholding and linking. This method required images to be divided into overlapping squares (8 by 8 or 4 by 4) where line segments are extracted by using a Hough transform. Perceptual criteria such as proximity, connectivity, similarity in orientation, contrast and edge pixel intensity, are used to group edges both strong and weak. This approach is called perceptual grouping. The second method is based on a region extraction using split and merge approach using spatial domain data. An n-level (for a 2^n by 2^n image) quadtree based pyramid structure is constructed to find the most homogenous large dark region, which in most cases corresponds to the lumen. The algorithm constructs the quadtree from the bottom (pixel) level upward, recursively and computes the mean and variance of image regions corresponding to quadtree nodes. On reaching the root, the largest uniform seed region, whose mean corresponds to a lumen is selected that is grown by merging with its neighboring regions.

In addition to the use of two-dimensional information in the form of regions and contours, three-dimensional shape can provide additional information that will enhance the system capabilities. Shape or depth information from an image is estimated by various methods. A particular technique suitable for endoscopy is the shape from shading, which is developed to obtain the relative depth of the colon surface in the image by assuming a point light source very close to the camera. If we assume the colon has a shape similar to a tube, then a reasonable approximation of the position of the center of the colon (lumen) will be a function of the direction in which the majority of the normal vectors of shape are pointing. From the above, it is obvious that there are multiple methods for image processing at the signal level.

The second layer is the control layer and at this level, a decision model must be built for endoscope navigation and advisory system. The system that we built is the models of probabilistic networks that create a basic, artificial intelligence system for navigation in the colon. We have constructed the probabilistic networks from correlated objective data using the maximum weighted spanning tree algorithm. In the construction of a probabilistic network, it is always assumed that the variables starting from the same parent are conditionally independent. However, this may not hold and will give rise to incorrect inferences. In these cases, we proposed the creation of a hidden node to modify the network topology, which in effect models the dependency of correlated variables, to solve the problem. The conditional probability matrices linking the hidden node to its neighbors are determined using a gradient descent method which minimizing the objective cost function. The error gradients can be treated as updating messages and can be propagated in any direction throughout any singly connected network to adjust the network parameters. With the above two-level approach, we have been able to built an automated endoscope navigation and advisory system successfully.

Keywords: Probabilistic Network, Bayesian Inference, Unobservable variables, Shape from Shading, Region Segmentation, Armijo's rule.

1. INTRODUCTION

An endoscope is a medical instrument used for observing the inner surfaces of the human body. Particularly, it is used for diagnosing different upper gastrointestinal (UGI), colon and bronchus diseases. Different types of endoscopes are widely used in medical institutions all over the world as an alternative to exploratory surgery. However, navigating an endoscope in the human body requires a great deal of skill and experience.

Navigation of an endoscope inside the human colon is a complex task. The endoscope behaves like a chain of articulated rods being pushed at the rear end into a highly flexible tube-like UGI or colon with many bends, twists and pockets. For instance during colon endoscopy, the endoscope tip is introduced into the rectum and is gradually advanced through the large intestine. When it bears on colon walls it may distort its shape and produce paradoxical behaviors around the tip. The colon cross sections are also not uniform as the colon can completely collapse at certain places, making the lumen difficult to see. The aim has been to make use of machine vision for guiding the terminal portion of the endoscope and spare the endoscopist from this task while the endoscope is advanced[1-3].

The current generation of endoscopes has a single fixed camera and there is no provision for the direct measurement of depth from UGI, colon or bronchus images. It is difficult to define an accurate reflectance function for the inner body surfaces, which are illuminated by a point light source accommodated at the endoscope tip. The light source and camera are assumed at the same point as they are located in the same plane and very close to each other on the endoscope tip as shown in Figure given below. Under these illumination conditions, the deepest area in the colon with respect to the viewer roughly corresponds to the darkest area in the image. Tip direction can be controlled accordingly during the endoscope insertion process. Region extraction is the most appropriate method for detecting dark regions in colon images where they are directly visible.

2. SIGNAL LEVEL PROCESSING

The endoscope navigation and advisory system can be viewed as a two-layer system. The first layer is at the signal level, which consists of the processing that will be performed on a series of images to extract all the identifiable features. The information is purely dependent on what can be extracted from the 'raw' images. At the signal level, the first task is performed by identifying a single dominant feature, lumen. Few methods of identifying the lumen are developed during our course of investigations.

2.1. Contour Extraction

The inner walls of a human UGI or colon contain circular rings of muscle as evident from the colon images shown in Figure 1. These rings are clearly distinguishable as they form occluding edges. When an endoscope is directed along the centerline of a straight section of colon, the muscle rings appear as closed contours in the image. The center of closed contours coincides with the correct insertion direction. For partially visible contours, an estimate of the insertion direction is made from their curvature[2].

The contours are extracted by edge detection, thresholding and linking. This method required images to be divided into overlapping squares (8 by 8 or 4 by 4) where line segments are extracted by using a modified Hough transform. Perceptual criteria, such as proximity, connectivity, similarity in orientation, contrast and edge pixel intensity, are used to group edges both strong and weak. This approach is called perceptual grouping.

2.2. Dark Region Detection

The second method is based on a region extraction using split and merge approach with spatial domain data. An N-level quadtree based pyramid structure is constructed to find the most homogenous large dark region, which in most cases corresponds to the lumen. The algorithm constructs the quadtree from the bottom (pixel) level upward, recursively and computes the mean and variance of image regions corresponding to quadtree nodes. On reaching the root, the largest uniform seed region, whose mean corresponds to a lumen is selected that is grown by merging with its neighboring regions[2].
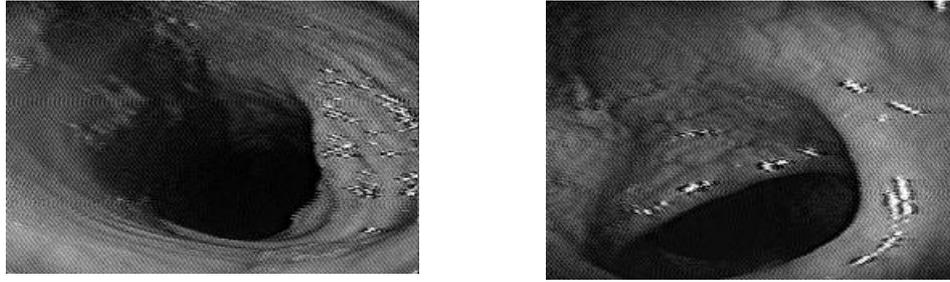
Figure 1 image of the inner wall of colon

2.3.   Shape from Shading

In addition to the use of two-dimensional information in the form of regions and contours, three-dimensional shape can provide additional information that will enhance the system capabilities. Shape or depth information from an image is estimated by various methods. A particular technique suitable for endoscopy is the shape from shading, which is developed to obtain the relative depth of the colon surface in the image by assuming a point light source very close to the camera. If we assume the colon has a shape similar to a tube, then a reasonable approximation of the position of the center of the colon (lumen) will be a function of the direction in which the majority of the normal vectors of shape are pointing.

This approach is based on reconstruction of three-dimensional coordinates at each pixel point in an image by inverting the reflectance equation[4]. In the case of endoscope image formation, there is a useful illumination arrangement where the point light source and viewer are in the same plane and very close to each other. This has enabled researchers at the Imperial College[5] to devise a linear shape from shading algorithm to recover relative depth.

The shape from shading method reconstructs the surface normals $(p,q,-1)$ at a set of points in the image. The normals that we obtain from low-level processing consists of one vector (p,q) per pixel which gives the orientation of the surface at this point with respect to two orthogonal axes (x, y) that are perpendicular to the camera axis (z). The surface normal can vary from $p = q = 0$ when it is perpendicular to the endoscope's tip (camera), to p or q close to infinity when it is parallel to the camera. 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 (p,q) vectors are pointing to (except in the case in which the lumen is in the centre of the image, in which case there will be no dominant direction)[5]. Although it was necessary to assume that the colon has Lambertian surfaces, the results could be still be used to give a reasonable statistical estimate of the lumen position. In our system, a set of 32 x 32 pixel points is selected and the slope of surface at each point is computed. From these, the deepest point in the image is estimated.

2.4.   Template Matching In Fourier Domain

Other techniques investigated includes the uses template matching to identify the lumen position[6] in Fourier domain. The template with the best correlation in magnitude gave an indication of the size and the phase of the first harmonic was used to indicate the position. Real time performance was achieved by computing two one-dimensional transforms on the horizontal and vertical projections of the image, rather than using the full two-dimensional transform, and by restricting the correlations to the low frequencies. The method proved effective in a large number of cases, and indeed would give a correct indication of position in the case where the lumen was not directly visible. However, it was unable to cope well with the artifacts such as diverticula or pockets on the inner walls, which also resemble the lumen.

3. CONTROL LEVEL PROCESSING

In the second layer, known as the control layer, for endoscope navigation, a decision model must be built for endoscope navigation and advisory system.

3.1. Probabilistic Networks from Objective Data

The system that we built is the models of probabilistic networks that create a basic, artificial intelligence system for navigation in the colon. We have constructed the probabilistic networks from correlated objective data using the maximum weighted spanning tree algorithm. Figure 2 shows the information involved in the reasoning about the location of lumen.
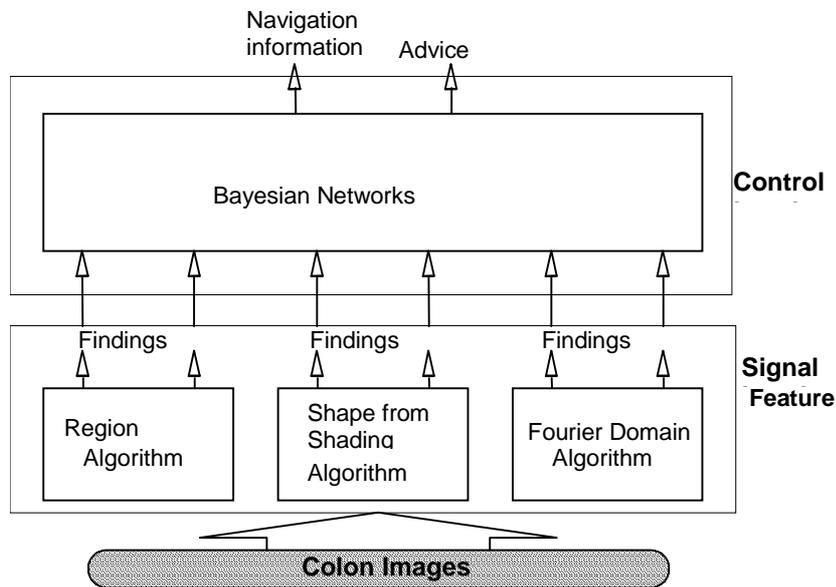

Figure 2 Continuous probabilistic network for estimation of the lumen location in the navigation module.
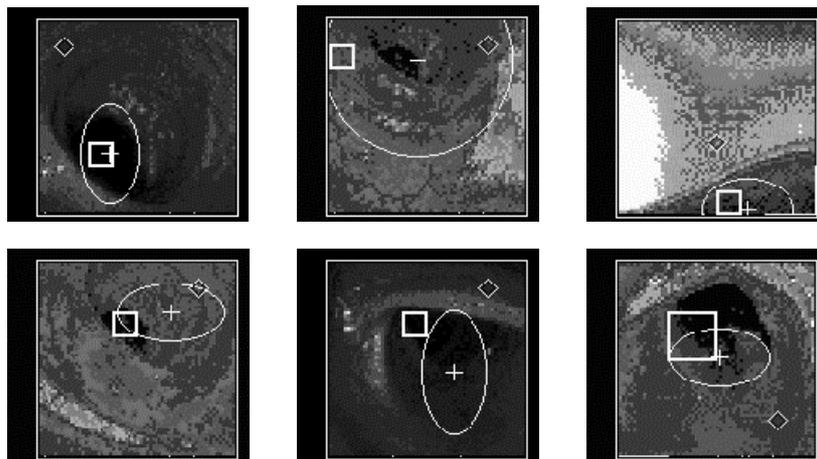

Figure 3 Some images with processed information

Figure 3 show some of the images where all the three models provide estimation of the lumen location. In these images, the cross is the estimated location of the lumen by the Fourier domain model with an ellipse indicating the estimated size of the lumen. The square is the large dark region estimated by the region based segmentation model and its centre will be the estimated location of lumen. Lastly the diamond in the images represent the estimated location of the lumen by the shape from shading algorithm.

In constructing a numerical expert system with many signal level models, we can approach the problem as a system that consists of many sub-systems, each with an associated probabilistic network. In the simplistic sense, they can be considered as branches in the total network[7]. In this simplest approach, we assume that all the interacting variables are observed and we want to construct a probabilistic network for each sub-system, using all the observed variables. The observed data, together with the topology, derived from a knowledge base, will be translated into prior and conditional probabilities for each state of

the variables. In order to determine the mapping from the problem to the solution space, the probabilistic network knowledge based system must be constructed from available data and information.

In the closed world definition, if **V** is the set of all interacting variables $\{V_1, V_2, V_3, \cdots\}$ for a model, then

$$P(\mathbf{V}) = \prod_{V_i \subset \mathbf{V}} P(V_i / pr(V_i)) \quad V_i, pr(V_i) \subseteq V$$

(1)

where $pr(V_i)$ is the parent of the variable V_i.

If we assume that the extracted features, denoted as variables **V**, are all that exist and are required to model a system, then we would expect to have observed data for all nodes in the desired network. Many researchers have developed algorithms for constructing the network topology from empirical observations[8-12]. Most of their algorithms are improvements of the maximum-weighted spanning tree algorithm first formulated by Chow and Liu[13] which utilised a mutual information measure to measure the divergence between the true (measured) distribution P and the tree-dependent distribution P_t as

$$I(P, P_t) = -\sum_{i=1}^{|X|} I(X_i, pr(X_i)) + \sum_{i=1}^{|X|} H(X_i) - H(\mathbf{X})$$

(2)

In the above equation, $H(\ )$ is the entropy measurement of a (marginal) distribution. $I(\ )$ is the cross-entropy measurement between two distributions and $pr(\ )$ represents the casual parent. Since the last two terms of are constants, minimizing the divergence is equivalent to maximize the first term, the total branch weight. Hence, their algorithm is known as the maximum weighted spanning tree (MWST). These types of maximum connection weight algorithms have the big advantage of not needing to consider all the possible trees that could be constructed from purely objective data. However, the possible ignorance of some interacting variables will generate many probabilistic networks that could closely approximate the given observed data.

3.2. Naive Bayes Networks

In the earlier work, Sucar and Gillies[14] utilized both region based segmentation and shape from shading algorithms to implement an advisory module based on Pearl's[15] Bayesian networks and assumed total conditional independence for all the feature variables. The resulting basic artificial intelligence system performed better than a rule base system. Although the system performed well, there were still cases in which images were mis-classified.

3.2.1. First Generation Naive Bayes Advisory Module

One of the reason that the first generation advisory module mis-classified the colon images in certain cases was due to the conditional independence assumption of probabilistic networks had been violated. To overcome the problem, Sucar and Gillies[14] suggested the three strategies to handle correlated data: node deletion, node combination and node creation. In practice we seldom have 100% correlated data, and hence, node deletion will usually throw away some information. For node combination, we have a situation where two variables must be assigned to one object which is the generation of clique as in the paper of Lauritzen and Speigelhalter[16], but it increase the complexity of the problem dramatically. The third strategy, node creation appears to be a powerful solution method. Sucar proposed a methodological solution, namely consultation with experts, to derive a node that makes the two dependent variables conditional independent. However, it will, in general, be a very difficult process for the expert to define a function that will combine the information from the two evidence variables into a coherent variable. Hence, the next stage of the work is to devise a way to create a hidden node based on the statistical distribution of the two evidence variables and an objective function that satisfies the axioms of conditional independence in the framework of the probabilistic methodology. The approach uses the training data, without seeking expert opinion, to define a mapping that will fuse the dependent information.

Figure 4(a) illustrates the naming conventions of one of these models (sub-systems) and the 'naive Bayes' probabilistic networks. The fundamental conditional independence assumption for the above 'naive Bayes' models were found to be weak according to statistical testing.

3.2.2. Introduction of Intermediate Nodes

In order to improve the system performance, intermediate variables were into the probabilistic network as shown in Figure 4(b).
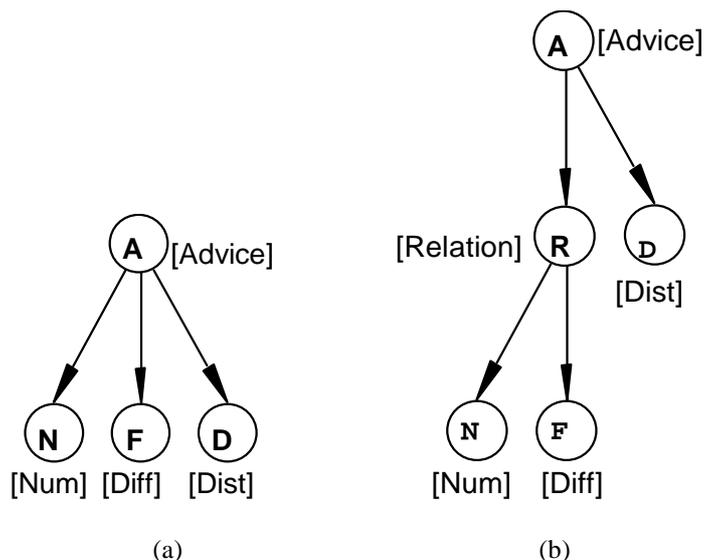
Figure 4 The Bayesian network for Dark Region model,
(a) the naïve bayes model and (b) model with additional node created.

To solicit the states of instantiation for the nodes [Relation] for each set of training data is a difficult task. Besides, even with these intermediate variables, an iterative iterative revision of the value of intermediate node is needed to ensure that the resulted network conforms closer to the conditional independence assumption than the original network. If the states for [Relation] are not well estimated, the performance of the new network can in fact be worsen. Hence, node creation is powerful but subjects the model to more uncertainty and usually requires much iteration to make the created node render its children conditional independent. This limitation motivated the development of a statistical regression method to collect training data and search for a mapping that will fuse the dependent information, without seeking expert opinion, statistically. This approach is the creation of hidden nodes (unobservable variables), to model the dependency[17].

3.3. Hidden Variables

In the construction of a probabilistic network, it is always assumed that the variables starting from the same parent are conditionally independent. However, this may not hold and will give rise to incorrect inferences. Pearl and Verma[18] stated that "Nature possesses stable causal mechanisms which, on a microscopic level are deterministic functional relationships between variables, some of which are unobservable." By adding hidden variables to model those nature hides from us, we should be able to satisfy the axioms of conditional probability and recover a better causal structure. A hidden node is transparent during data collection and its existence may be unknown to the expert. Its probability parameters are found by a search in the world of feasible values for the best value. For simplicity, we use the term unobservable variables to refer to these types of variables. From the derivation of the MWST algorithm, the divergence measurement can be used as a comparative performance measurement to select the best network structures. However, if there are some variables that are unobservable, then direct application of the divergence measurement is not possible.

In our work, we treated the problem of finding the best probabilistic network from different possible structures and different conditional probabilistic matrices, as an estimation problem. If we represent all the parameters in the probabilistic network as S, then the fundamental problem is the estimation of the parameters, S, given the known information, using only training data. Let X be the collection of training data and Y be all data in the environment ($X \subset Y$). If X can be separated into X_{in}, the measured values of input variables, and X_{out}, the desired value of output variables. The most common expression that is used to estimate the models parameters S, is

$$S(\mathbf{X}) = \max_{S} Q\left(\frac{\mathbf{X}_{out}}{S(\mathbf{X}_{in})}\right) \tag{3}$$

In our application, since there is no strict direction of signal flow, queries can be made at any node in the network. Hence we define the training data set X as collective of $\{X_1, X_2, X_3, ..., X_{|X|}\}$ and the above equation is modified to

$$S(\mathbf{X}) = \max_{S}\left\{\bigvee_{i=1}^{|X|} \bigvee_{Z \subseteq \mathbf{X} \backslash X_i} Q\left(\frac{\mathbf{X}_i}{S(\mathbf{Z})}\right)\right\} \tag{4}$$

In this case we search for the parameters that will maximize the predictive ability of the model by allowing any node to be the query node and instantiating all other nodes as evidence nodes.

In Bayseian networks, the estimated posterior probability is denoted as $Bel(X_i)$ and the actual observed value of the query node denoted as $Q(X_i)$ can be used to quantify the ability of the model in estimating the true value of the query variable. If we define a monotonic error cost function for the differences between the desired posterior probability and the estimated posterior probability. The expected cost function will be a measure of the overall network's performance and can be used in a performance search algorithm to find the best estimation of S. There are many choice of cost functions for this application[19,20]. One such function is the sum of squared error cost function that has been widely understood by classical statisticians and widely used in many applications, it is also known as squared-error cost function.

Let the variable A be the decision variable and the training data $\mathbf{E} \in \mathfrak{S}$ (The full set of training data). If we are interested in the posterior probability of A given some evidence $\mathbf{E}$, denoted as $Bel(A)$.

$$\xi = \sum_{i=1}^{|A|} \frac{1}{2}[Q(a_i) - Bel(a_i)]^2 \tag{5}$$

where $E\{.\}$ is the expectation operator, and $Q(a_i)$ is the observed value of a_i.

We then expand the expression for those conditional probabilities by a Taylor's series, so that we can use a gradient search method, such as least mean squares algorithm to find the solution. Updating is done using:

$$\begin{aligned} S(\underline{\underline{n+1}}) &= S(\underline{\underline{n}}) - \eta \Delta S(\underline{\underline{n}}) \\ &= S(\underline{\underline{n}}) - \eta E\left[\frac{\partial \xi}{\partial S(\underline{\underline{n}})}\right] \end{aligned} \tag{6}$$

In the creation of a hidden node, the conditional probability matrices linking the hidden node to its neighbors are determined using a gradient descent method which minimizing the objective cost function. The error gradients can be treated as updating messages and can be propagated in any direction throughout any singly connected network to adjust the network parameters[17,21].

It is well known that the quality and search time of a solution depends a lot on the initial estimate. The commonly used random number approaches often get stuck in an inferior local minimum and require multiple restarts or the process of simulated annealing to try to move them nearer a global minimum. Since the probabilistic network is formally derived, we derived a method of estimating the initial values of hidden using linear average[22]. With good initial starting point, the search algorithm will arrive at a good minimum point in roughly one-third of the training time as compared to a random numbers starting condition. Furthermore, the (first) solution was a good approximation to the global minimum without restart.

3.4. Optimising the Learning Rate (Armijo's Rule)

In this paper, we will discuss one aspect of the iteration technique employed in our minimization problem. The primary difference between most algorithms rests in the rule by which successive directions of descent are selected. The secondary differences lie in the selection of step size (also known as learning rate). For a general non-linear function that cannot be minimized analytically, a quadratic approximation is used and it is desired that at each iteration, we move towards the approximated minimum as quickly as possible. A practical and popular criterion for determining the optimal step size (learning rate) is the Armijo's rule[23]. Let us define the function

$$\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k) \tag{7}$$

where $f()$ is the cost function, α is the step size, $\mathbf{x}_k$ is the current value of parameters and $\mathbf{d}_k$ is the direction of search.

Armijo's rule is implemented by consideration of the function $\phi(0) + \varepsilon \phi(0)'\alpha$ for fixed ε, $0 < \varepsilon < 1$. A step size is considered not too large if the corresponding function value is

$$\phi(\alpha) \leq \phi(0) + \varepsilon \phi(0)'\alpha \tag{8}$$

and α is considered not too small if, for $\eta > 1$

$$\phi(\alpha) > \phi(0) + \varepsilon \phi(0)'\eta \alpha \tag{9}$$

From the above equations we can derive

$$\phi(0) - \phi(\alpha) \geq -\varepsilon \phi(0)'\alpha$$
$$\phi(0)' = \left. \frac{\partial \phi(\alpha)}{\partial \alpha} \right|_{\alpha=0+} = \nabla f(\mathbf{x}_k)^T \mathbf{d}_k \tag{10}$$

If the direction of search is the negative gradient for minimization, we have

$$\phi(0) - \phi(\alpha) \geq \varepsilon \alpha \left[\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k) \right] \tag{11}$$
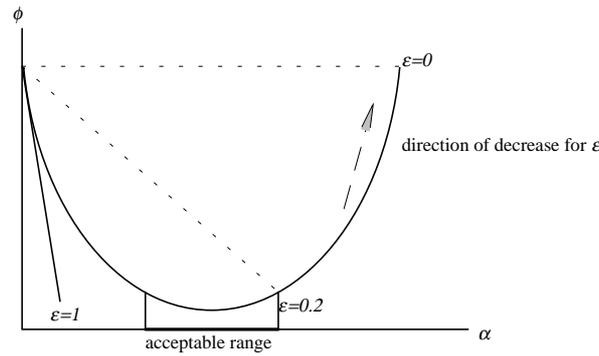

Figure 5 Pictorial illustration of ε in Armijo's rule

Figure 6 indicates that the improvement in performance must be greater than a fraction, ε, of the gradient projected in the search direction, which is the inner product expressed in the [ ].

In practice, our function f is the sum-of-squared error cost function, ξ, and our procedure to determine the optimal learning rate is as follows. We define

$$g(\alpha, \varepsilon) = \xi(S_k - \alpha \nabla \xi(S_k)) - \xi(S_k) + \varepsilon \alpha \nabla \xi(S_k)^T \nabla \xi(S_k) \tag{12}$$

where $0 < \varepsilon < 0.5$

We tested the above rule for choosing the optimal step size. However, it is not easy to select the constant ε. From expert opinion, it is usually taken as constant of 0.2. However when we tested 0.2 for our system, the convergence speed was not satisfied. Hence modification to the choice of ε is critical especially we come close to the performance basin. One solution is to start with demanding constraint (high ε) and slowly relax the constraint (lowering ε) as we are reaching the minimum.

In this case, we treat the constant ε as a function of time, $\varepsilon(t)$. Another approach is to treat the constant ε as a function of the step size α, $\varepsilon(\alpha)$. In this case, we want the gain in performance to be significant (associated with large ε) if we were to allow a large step size and we allow the gain to be marginal (associated with small ε) if we only move with a very small step size α.

3.5. Experimental Results and Comments

Figure 7 provides the comparison of the various options illustrated by sub-graphs. In (a) we plot the sum-of-squared error for the root node. In (b) we plot the step size for each training iteration, and in (c) we plot the correlation performance between the estimated and actual observed data for the root node (solid line) and the two leaf nodes (dashed line and dotted line).
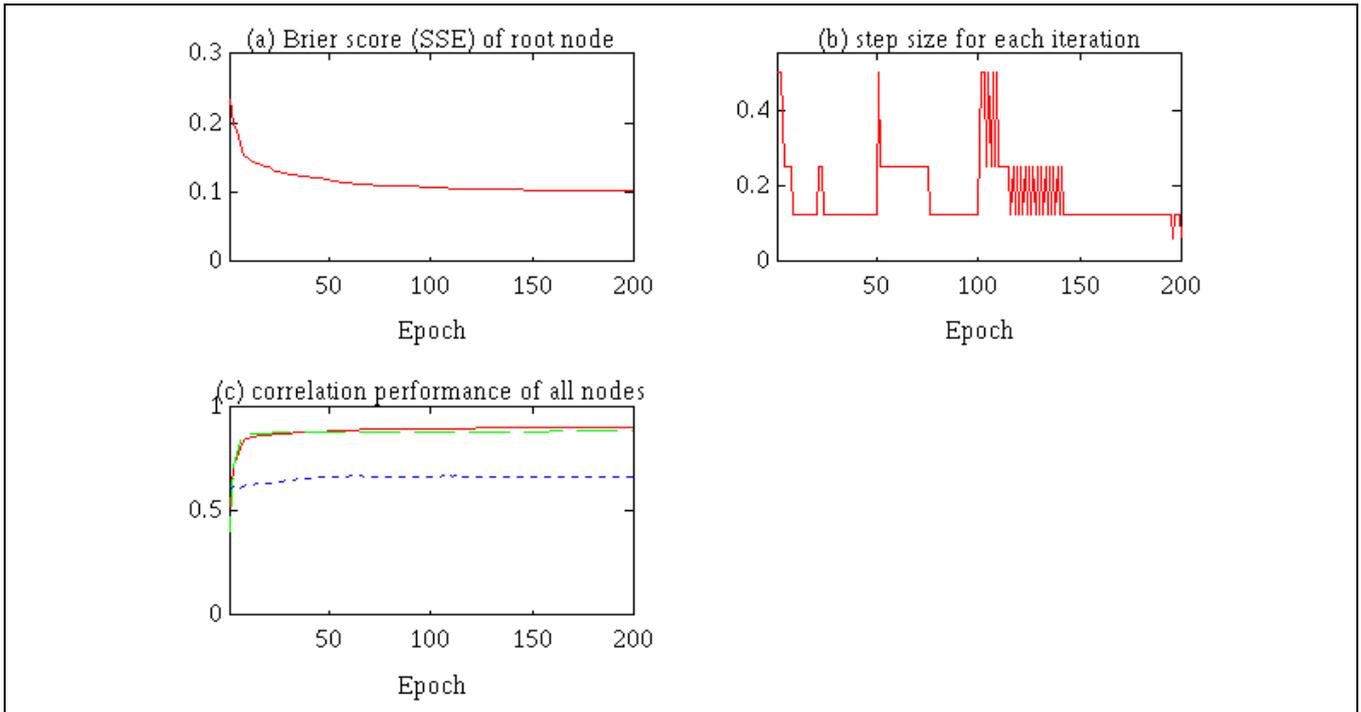

Figure 7 Adaptive Armijo's rule for optimal step size $\varepsilon = 0.5 \ \alpha$

We did a comparison with the following test. (1) Using a fixed step size of 0.1 as is commonly found in literature of connectionist network. (2) Using an adaptive step size as commonly found in connectionist applications with increment ratio of 1.05 and decrement ratio of 0.7. (3) Using Armijo's rule for optimal step size with different ε constant as employed in operations research. (4) Using our adaptive Armijo's rule for optimal step size with different $\varepsilon(\alpha)$ function.

From the experiment (not shown) the constant step size approach is simple but will take a long time to reach the minimum. Furthermore, the solution will oscillate near the solution basin. In the case of adaptive step sizes, there are improvement but the learning time is still long and the ripple effect found near the minimum is reduced but still exists due to the heuristic rule by which the step size is adapted. When we tested the various constant ε values for the original Armijo's rule. Due to the nature of our solution space, we found that the original Armijo's rule brings the solution down very quickly (in 10 to 20 cycles) to near the optimal solution. However, due to the demanding requirement for subsequent improvement, the step size is always clamped to the minimum value. Hence it takes more than 2000 epochs to reach a minimum equivalent to the other method. The best ε constant for our application is 0.1. The above charts shown an adaptive Armijo's rule where $\varepsilon = 0.5 \ \alpha$ provides the best solution. The algorithm brings the error monotonically down to the solution. The training time is reasonable, around 50 epochs to reach near minimum and around 150 to reach a good solution where subsequent cycles provide only minimum gain. In our system, the system performance is around 75% as compared to 64% with the naïve bayes structure.

4. CONCLUSION

In a typical engineering approach, we deal directly with sequence of endoscopic images. We have to understand the cues that a domain expert utilizes for decision making and translate those abstract cues into qualitative features, decide how and what features are to be extracted and construct a numerical knowledge based system for reasoning with evidence and uncertainty.

To prevent building an over complex probabilistic network incorporating all the features and getting trapped in the process of validating and modifying the interaction of many observed variables, we chose to model the system with two-level approach. At the signal level, we dealt directly with finding the feature sets and feature extraction models for influence. The system utilized multiple models to estimate the position of the lumen. These are the region segmentation model, the shape from shading model and a template matching in the Fourier domain.

In the control level, we used Bayesian networks as the numerical expert system for reasoning under uncertainty. Bayesian networks employs a graphical inference structure to capture explicit dependencies among the domain variables. In the advisory module where data collection is not a problem, the maximum weighted spanning tree algorithm (MWST) algorithm is commonly used to construct the inference structure. To improve the performance of our system, we proposed the creation of a hidden node to modify the constructed network topology, which in effect models the dependency of correlated variables, to solve the problem unobservable variables. The conditional probability matrices linking the hidden node to its neighbours are determined using a gradient descent method which minimizing the objective cost function. The error gradients can be treated as updating messages and can be propagated in any direction throughout any singly connected network to adjust the network parameters. Our method utilizes objective probabilities determined from the training data and constrained by the axioms of probability, and performance is maximized without expert intervention during training. Improvement in reducing search time was achieved by an adaptive Armijo's rule to determine the optimal learning rate for each iteration.

With the above two-level approach, we have built an automated endoscope navigation and advisory system successfully.

5. REFERENCES

1. G. N. Khan, Duncan F. Gillies, "A highly parallel shade image segmentation method," *Proc. International Conference on Parallel Processing for Computer Vision and Display,* University of Leeds, 1988.

2. G. N. Khan, Duncan F. Gillies, "Extracting contours by perceptual grouping," *Image and vision computing,* 1992, vol. 10, no. 2, pp. 77-88

3. G. N. Khan, Duncan F. Gillies, "Vision Based Navigation System for an Endoscope," *Image and vision computing,* 1996, vol. 14, pp. 763-772

4. Rashid H. and Burger P, "Differential Algorithm for the determination of Shape from Shading using a point light source" *Image and Vision Computing*, 10(2) 119-127.

5. L. E. Sucar, D. F. Gillies and H. Rashid, "Integrating Shape from Shading in a Gradient Histogram and its application to Endoscope Navigation*." 5th International Conference on Artificial Intelligence (ICAI-V)* Cancun, Mexico (1992).

6. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Using Fourier Information for the Detection of the Lumen in Endoscope images." *IEEE Region 10's Ninth Annual International Conferene, Proceeding of TENCON conference* Aug. 1994, Singapore, pp 981-985.

7. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Probabilistic Reasoning and Multiple-Expert Methodology for Correlated Objective Data." *Artificial Intelligence in Engineering* 12 (1998), pp 21-33.

8. G. Rebane & J. Pearl, "The Recovery of Causal Poly-trees from Statistical Data," *Uncertainty in Artificial Intelligence,* 3, 1989, ed. L. N. Kanal, T. S. Levitt and J. F. Lemmer, North-Holland, Amsterdam, pp. 175-182

9. Dan Geiger, "An Entropy-based Learning Algorithm of Bayesian Conditional Trees," *Uncertainty in Artificial Intelligence*, 1992, ed. Dubois, Wellman, D'Ambrosio & Smerts. pp. 92-97

10. G F Cooper, E Herskovits, "A Bayesian Method for Constructing Bayesian Belief Networks for Databases," *7th Conference on Uncertainty in Artificial Intelligence*, 1991, UCLA, ed. B D D'Ambrosio, P Smets, P P Bonissone, Morgan Kaufmann. pp. 86- 94.

11. G F Cooper, E Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, 9, 1992. pp. 309-347.

12. Rafael Molina, Luis M de Campos, Javier Mateos, "Using Bayesian Algorithms for Learning Causal Networks in Classification Problems," *Uncertainty in Intelligence Systems*, ed. B. BouchonMeunier et al, c1993, pp. 49-58

13. C K Chow, C N Liu, "Approximating Discrete Probability Distributions with Dependence Trees," *IEEE Transactions on Information Theory*, 1986, Vol. 14, No. 3, pp. 462-467.

14. L. E. Sucar, D. F. Gillies, D. A. Gillies, "Objective probabilities in expert systems*." Artificial Intelligence* 61 (1993), pp. 187-208

15. Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, c1988

16. S. L. Lauritzen, D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems." *Journal of Royal Statistics Society series-B Methodological*, 1988, Vol. 50, No. 2, pp. 157-224

17. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Using Hidden Nodes in Bayesian Networks." *Artificial Intelligence* 88 (1996), pp 1-38.

18. J. Pearl, T S Verma, "A Theory of Inferred Causation." *2nd Conference on the Principles of Knowledge Representation and Reasoning*. pp. 441-452.

19. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Choice of error cost function for training unobservable nodes in Bayesian network." *1997 The first International Conference on Knowledge-Based Intelligent Electronic Systems (KES' 97),* 21-23 May, 1997, Adelaide, Australia, Editor L C Jain. pp 565-574.

20. John Binder, Daphne Koller, Stuart Russell, Keiji Kanazawa, "Adaptive Probabilistic Networks with Hidden Variables." *Machine Learning*, 1997.

21. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Research Note: Gradient Updating and Propagation in a singly-connected Baysian Network." *Artificial Intelligence Journal*, accepted.

22. Kwoh, Chee Keong and Gillies, Duncan Fyfe, "Estimating the Initial Values of Unobservable Variables in Visual Probabilistic Networks." *6th International Conference, CAIP' 95*, Prague, Czech Republic, Sept 1995. pp 326-333.

23. Larry Armijo, "Minimization of Functions Having Lipschitz Continuous First Partial Derivatives." *Pacific Journal of Mathematics*, Vol. 16, No. 1, 1966, pp. 1-3