World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:11, 2007

Integrating Low and High Level Object Recognition Steps

András Barta, and István Vajk

Abstract**—**In pattern recognition applications the low level segmentation and the high level object recognition are generally considered as two separate steps. The paper presents a method that bridges the gap between the low and the high level object recognition. It is based on a Bayesian network representation and network propagation algorithm. At the low level it uses hierarchical structure of quadratic spline wavelet image bases. The method is demonstrated for a simple circuit diagram component identification problem.

Keywords**—**Object recognition, Bayesian network, Wavelets, Document processing.

I. Introduction

COMPUTER vision has gone through significant advancement during the past few decades. Though several real world applications are created, it is still behind its capabilities. Due to the fast technical development the main limitation is not memory and speed any more. More emphasis should be put on the control and teaching procedures of the object recognition systems. In this research we try to address these problems by combining the low level segmentation problem with the high level object recognition process.

General object recognition and part based structural description has a long history. Several object recognition system with structural object description have been created: VISION (Hanson, Riseman, 1978), SIGMA (Hwang at al., 1986) , SPAM (McKeon at al., 1985), ACRONYM (Brooks, Binford, 1981), SCHEMA (Draper et al., 1989). These systems, their successes and failures are investigated by Draper [3]. He finds that knowledge-directed vision systems typically failed because the control problem for vision procedures was never properly addressed as an independent problem. He also argues that problems are created because adding new features or new object classes solves many problems initially but as the system grows they make the system intractable. This paper searches solutions for these problems. The control problem of the object recognition system is treated in the framework of Bayesian networks. The system learning is addressed by an incremental learning procedure. The Bayesian framework is extended by agent-based programming in order to simplify the control procedures in case of complicated object descriptions. A similar approach is used by Takatsuka et. al [7]. They used Hopfield neural network to solve the sub-graph matching problem.

The algorithm is based on a Bayes network and a recursive orthogonal linear transformation. The method is demonstrated on a document image processing problem, extracting the components of a circuit diagram. Many old blue-prints of electrical equipment are sitting on shelves. Converting them to a meaningful digital representation would make it possible to search and retrieve them by content.

Probabilistic Bayesian networks are investigated extensively in the literature. Perl presented a tree based belief network inference with linear complexity [5]. Dynamic tree structures are gaining popularity, because of their better object representation capabilities [9]. Okazaki at al. proposes a method for processing VLSI-CAD data input [6]. His system is implemented for digital circuitry where the components are mainly loop-structured symbols. Symbol identification is achieved by a hybrid method, which uses heuristics to mediate between template matching and feature extraction. The entire symbol recognition process is carried out under a decision-tree control strategy. Siddiqi at al. present a Bayesian inference for part based representation [8]. The object subcomponents are represented by fourth order polynomials. The recognition is based on geometric invariants, but it does not provide a data-structure for representing the image features. Wavelets and digital filters are used in the literature many ways. Freeman and Adelson provided a framework for steerable filters [13]. They presented an architecture to synthesize filters of arbitrary orientations from linear combinations of basis filters. They, however not addressed the problem of joining the edge pixels. Sung, Bang and Choi constructed hierarchical network of wavelets for handwritten numeral recognition [14]. They used only the imaginary components of Gabor wavelets for the identifications. Deng, Lati and Regentova used cubic splines for document processing [15]. They applied a classification method for segmenting documents. They applied three-means or two-means classification for classifying pixels with similar characteristics after feature estimation of spline wavelet transforms.

Manuscript received December 15, 2005. This work was supported by the fund of the Hungarian Academy of Sciences for control research and by part by the OTKA fund TO42741. The supports are kindly acknowledged.

András Barta is with the Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Budapest, H-1111 Budapest, Goldmann Gy. tér 3., Hungary (phone: 36-1-463-2870; fax: 36-1-463-2871; e-mail: barta@ aut.bme.hu).

István Vajk is with the Department of Automation and Applied Informatics, Budapest University of Technology and Economics.

In the next section we present a Bayesian network for object recognition and show an algorithm to perform the network calculations. In section III we investigate the low level wavelet image component creation. In section IV we define the wavelet bases and show how to detect low level line features. The last section presents a simple simulation example.

II. Model based Bayesian Network

Bayesian networks are well suited for image processing applications and it is used for this research because of the following advantages:

- provides probabilistic representation
- provides a hierarchical data structure
- provides an inference algorithm
- separates the operating code from the data representation
- it is capable of processing both predictive and diagnostic evidence
- provides and inhibiting mechanism that decreases the probabilities of the unused image bases

Bayesian network is defined for this application as follows [1], [2]. An image feature is represented by lower level image bases in a recursive way.

$$\xi_j = \sum_{i=1}^{n} T(\xi_i(\mathbf{a}_i), \mathbf{r}_i) \tag{1}$$

T is an operator that performs an orthogonal linear transformation on the image bases. The parameters of the transformation are stored in the $\mathbf{r}_i$ parameter vector. The image bases may be parameterized by an $\mathbf{a}_i$ attribute vector. Since features belong to parameterized feature classes the $\mathbf{a}_i$ vector is necessary to identify their parameters. This description defines a tree structure. The tree is constructed from its nodes and a library. The T transformation has three components, displacement, rotation and scaling. The four parameters of the transformation of node i are placed in a reference vector

$$\mathbf{r}_i = \begin{bmatrix} \mathbf{x}_i^r & s_i^r & \varphi_i^r \end{bmatrix} \tag{2}$$

where $\mathbf{x}_i^r = \begin{bmatrix} x_i^r & y_i^r \end{bmatrix}$ is the position of the image element in the coordinate system of its parent node, s_i^r is the scaling parameter and φ_i^r is the rotation angle. The conditional probability parameters $\theta_{i,j}$ are learned as relative frequencies. It can be shown that the distribution of the $\theta_{i,j}$ parameters is a Dirichlet distribution [4]. The conditional probabilities of the network can be described

$$p(\theta_1, \theta_2, ..., \theta_{L-1}) = \frac{\Gamma(n)}{\prod_{k=1}^{L} \Gamma(n_k)} \theta_1^{N_1-1} \theta_2^{N_2-1} ... \theta_K^{N_L-1} = \\ = Dir(\theta_1, \theta_2, ..., \theta_{L-1}; n_1, n_2, ..., n_L) \tag{3}$$

where n_k is the number of time node k occurs in the sample data and $n = \sum_{k=1}^{L} n_k$ is the sample size. The $\Gamma(x)$ function for integer values is the factorial function, $\Gamma(x) = (x-1)!$. These conditional probabilities are learned from the training data. In a typical Bayesian network the direction of the edge shows the casual relationships. In image processing applications it can not be said whether the object is causing the feature or the feature is causing the object; the edges of the tree may go in either direction. The direction depends on whether we are using a generative or descriptive model [10].

The recognition process starts by selecting a new image component. This single node tree is expanded by adding a structure shown on Fig. 1. By adding more and more nodes the whole image tree is created.

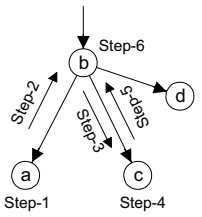

Fig. 1 Steps of the algorithm

Step 1: A new image component (a) is selected randomly based on the node probability distribution. The selection is performed by the roulette-wheel algorithm. In case of new node the prior probability is used.

Step 2: This new evidence starts the belief propagation of the network. Based on the conditional probabilities several parent node hypotheses are created (b, upward hypothesis). These object hypotheses are described by library index and coordinate system of the node. The coordinate system of the object hypothesis (a) $\mathbf{i}_{pi} = \begin{bmatrix} \mathbf{x}_{pi} & s_{pi} & \varphi_{pi} \end{bmatrix}$ can be calculated by the following coordinate transformation:

$$s_{pi} = \frac{s_i}{s_k^r} \tag{4}$$

$$\varphi_{pi} = \varphi_i - \varphi_k^r \tag{5}$$

$$\mathbf{x}_{pi} = \mathbf{x}_i - \mathbf{x}_k^r s_{pi} \begin{bmatrix} \cos\varphi_{pi} & \sin\varphi_{pi} \\ -\sin\varphi_{pi} & \cos\varphi_{pi} \end{bmatrix} = \mathbf{x}_i - \mathbf{x}_k^r s_{pi} \mathbf{R}_{\varphi} \tag{6}$$

where $\mathbf{i}_i = \begin{bmatrix} \mathbf{x}_i & s_i & \varphi_i \end{bmatrix}$ is the coordinate system of the

image component (b) and $\mathbf{r}_k = \begin{bmatrix} \mathbf{x}_k^r & s_k^r & \varphi_k^r \end{bmatrix}$ is the reference vector of child node of the library tree.

Step 3: This parent node hypothesis can be projected back to the image. This projection creates child hypotheses not only for node c, but all of the child nodes of b (for example d).

Step 4: A search is performed to match this projected child node hypotheses. If this object hypothesis matches one of the already identified subtrees then they are combined. If no match has been found then a new hypothesis are created (downward hypothesis) for the child node. If the child hypothesis is one of the lowest level image components then it is compared against the image, based on a distance measure. This distance measure can be, for example, the Euclidean distance. It should be defined for every basic image element independently; in our case for lines, circles and arcs. The results of the child node comparisons are converted to probability by an arbitrarily chosen function.

Step 5: The probability of the child modes propagates upward as new evidence. The upward probabilities are combined to calculate the probability of root b.

Step 6: Only the high probability nodes are processed, the others are neglected. This is true for both the upward and downward object hypotheses. This process creates a structure with several root nodes. These root nodes can be an input to a next level of recognition step. The root nodes are either lower level components that the algorithm will grow further or they are the final solutions.

The search method is adaptive and local; only certain area of the image is processed at a time. This is advantageous for images with noise or clutter. The calculation complexity of the algorithm can be described by the following dependencies:

- The complexity is linear with the number of nodes.
- The complexity does not depend on the size of the library but only on the number of nonzero upward conditional probability values. The complexity is lower if these probability values are concentrated in few high probability entries.
- The complexity is lower if the average object size is higher.
- The complexity is higher if the objects have symmetries.
- The complexity is higher if a node has several child nodes with identical library index.

Based on the probability values child parent relationships are created or terminated. This is similar to the "cut and merge" region segmentation methods. The parent hypotheses are created based on the conditional and prior probabilities the same way as for the Bayesian network.

III. Wavelets for Object Recognition

There is significant difference in using wavelet for image representation and object recognition. In case of image representation the main task is to code the image in denser form and then reconstruct the original image. In object recognition it is not necessary to represent every detail of the image and perfect reconstruction is also not needed. In case of image representation the purpose of the image coding is tile the position-frequency plane as much as possible. In object recognition applications the object descriptions are advantageous either in the frequency or in the position domain, therefore exact tiling is not necessary. A texture pattern may be identified better by the frequency response and therefore frequency domain description is more advantageous. For localized features however position domain description is simpler.

In object recognition applications the selection of wavelets for low level image base representation is a critical step. The following issues should be considered:

- Translation invariant representation
- Support size
- Frequency domain resolution
- Real or complex
- Orthogonal representation

Translation invariant wavelet representation is a critical requirement for object recognition. In case of translation invariant representation if the objects are translated then the wavelet coefficients are also translated without any change in values. If the representation is not translation invariant, even small position variation may result significant coefficient changes. The coefficient variations would make the higher level object recognition difficult.

The support size of the wavelets is an important factor in practical implementations. Compact support is necessary for fast calculations.

The frequency resolution of the wavelets depends on the detectable image feature. For texture identification high frequency resolution wavelets are necessary

Complex wavelets have some advantages to real wavelets. They can provide linear phase filter representation and they are easier to use in translation invariant applications. They provide, however very redundant representation, because the coefficients have imaginary and real components.

For multilevel representation the image element are represented by lower level image bases. In case of wavelet basis the wavelet scaling function can be constructed with higher resolution scaling functions, where $h(n)$ is a linear filter [11][12].

$$\phi(x) = \sum_n h(n)\sqrt{2}\phi(2x-n) \tag{7}$$

Similarly, the wavelet function can be constructed with higher resolution scaling functions.

$$\psi(x) = \sum_n g(n)\sqrt{2}\phi(2x-n) \tag{8}$$

In this paper spline wavelets are used for low level image base representation, because they have several properties that are advantageous for object recognition.

- They are symmetric.
- Exact reconstruction is possible.
- They provide a bi-orthogonal basis system.
- The wavelet decomposition can be done by mirror filters.
- They have compact support.

The one-dimensional quadratic spline wavelet can be calculated by the following mirror filters (in Matlab *rbio3.1*).

low fr. decomp. =	0.1250	0.3750	0.3750	0.1250
hi fr. decomp. =	0.2500	0.7500	-0.7500	-0.2
low fr. reconstr. =	-0.2500	0.7500	0.7500	-0.2500
hi fr. reconstr. =	0.1250	-0.3750	0.3750	-0.1250

Fig. 2 shows the scaling function and Fig. 3 the wavelet function decomposition. In case of hierarchical object structure this recursive decomposition is important.

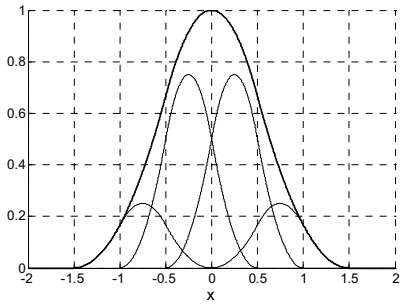

Fig. 2 Quadratic spline scaling function decomposition

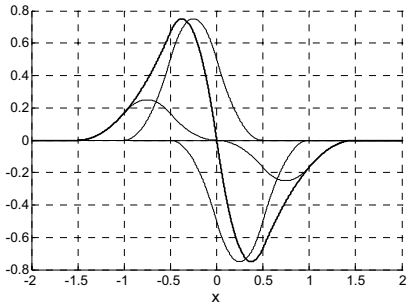

Fig. 3 Quadratic spline wavelet function decomposition

Quadratic spline wavelets provide good frequency localization. Its scaling function frequency response can be given in analytical form. Its Fourier transform is:

$$\Phi(\omega)=\left(\frac{\sin(\omega/2)}{\omega/2}\right)^3 e^{-i\omega/2} \tag{9}$$

The fastest method of calculating the wavelet coefficients is the *fast wavelet transform* (FWT). Unfortunately the FWT is not translation invariant because of dyadic sampling of the position parameter. At higher scale due to this fixed re-sampling the sampling grid and the features positions are not aligned. The continuous wavelet and the dyadic wavelet transform (à trous algorithm) is translation invariant, but they provide highly redundant representations [11]. Adaptive re-sampling is a way to solve this problem. A translation invariant representation can be achieved by optimization. The image base location is optimized in a continuous position parameter then the surrounding of the image feature is re-sampled relative to this maxima position.

IV. Low Level Representation by Wavelet Disks

The presented Bayesian network representation and algorithm can be used for low level image element detection. The recognition process is based on local continuous wavelet decomposition. The purpose of the wavelet decomposition in case of object recognition is to represent the image with as few as possible wavelet coefficients. This can be achieved by selecting wavelet bases which are very similar in shape to the detectable image elements. We introduce a new object, the disk. The pixel resolution of an image depends on the image creation process and it is independent of the content of the image. On the other hand the resolution of the image elements depends on the image content. The disk can bridge this gap, because it is constructed from pixels and its size is selected to reflect the resolution of the image content. With the introduction of the disk the picture can be represented independently from the pixel resolution. A disk also performs a discrete to continuous conversion, since it is defined on a discrete grid but its position is a continuous variable. In the hierarchical structure of image bases the disk is between the pixel and the low level image features, edges and lines. The disk is an extension of the mask concept. A disk is defined by the disk area and the disk function. Applying this mask to the image provides a matching or cost value.

The disk transform is defined as follows. The disk area A is defined as a circle on the image plain. The disk transform is

$$D(x,y,r)=\int_{x',y'\in \mathsf{A}(x,y,r)} I(x+x',y+y')\gamma(x',y')dx'dy' \tag{10}$$

where $\gamma(x,y)$ is the disk wavelet function and $\mathbf{I}(x,y)$ is the value of the image at x,y position and r is the radius of the disk.

Two-dimensional quadratic splines are used for the disk function. The wavelet coefficients at higher scale can be calculated easily by the mirror filters. The low level image feature detection requires directional sensitivity. Directional sensitivity can be achieved by selecting different scaling parameters for the two main axes. Fig. 7 shows some of the low level spline image bases. The edge detecting two-dimensional wavelet (Fig. 4-b) uses the scaling function for one axis and the wavelet function for the other in order to detect the intensity variation of the image:

$$\xi(x,y) = \psi(x)\phi(y) \tag{11}$$

The image bases can be rotated by applying the transformation equations (6) with the $\mathbf{R}_\varphi$ rotation operator. Directional sensitivity is applied on Fig. 4-c and Fig. 4-d.

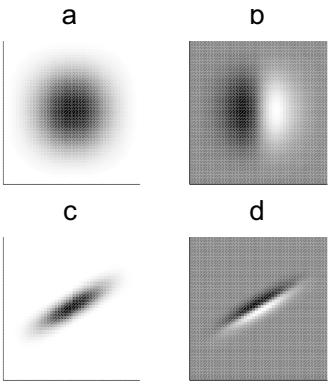

Fig. 4 Quadratic spline wavelet image bases a) Scaling function b) Wavelet function c) Directional ridge filter d) Directional edge filter

The image element is detected by placing the disk on the image in different positions and orientation angles, and calculating its cost value. The position scale and orientation is described by the coordinate system transformation. The higher level image elements are constructed by the combination of these disks. They can be identified by the same the same network calculations. The method has the advantage to other conventional edge detection algorithms that both bottom-up and top-down calculation can be used. The neighborhood of the found image element is searched again for new elements. The relationships of these new disks are calculated and they can be used for higher level image component detection. The curve segments are located and joined together to form a curve. After the first segment is found, the location of the next one can be guessed from the position and orientation of the previous disk. This representation is simpler than a general graph structure, since the points are calculated sequentially. From the coordinate system of a disk the position and the first derivative of the curve can be gained directly. Edge detection can be performed the same way as line detection, except the disk function has to be chosen differently.

With this method lines, circles and parameterized curves can be detected. The method can be used for detecting any other line type features the same way, but in this paper we focus on line detection Disks can also be defined to identify special types of picture elements, for example special types of edge profiles or textures. This can be achieved by defining mask functions that identifies that type of edge profiles. The disk definition can also be extended to use statistical properties of the disk area.

V. Simulation Results

The integrated method is tested for both low and high level image component detection. The system was built in Matlab object oriented environment. A general object graph structure is defined. Every node and edge can contain a user defined object.

A. Low Level Processing

The low level segmentation is carried out by the disk object and the spline ridge wavelets. Fig. 5 shows the disk behavior near a line if the position and the rotation parameter varies.

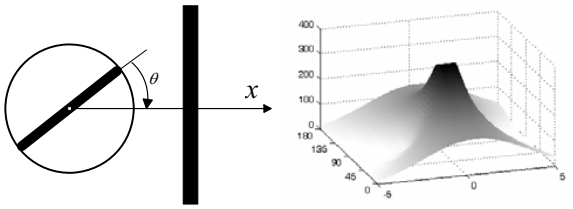

Fig. 5 Disk cost function dependence on position and rotation parameter

Since it is a monotone continuous function it is easy to find its local maxima. Similarly edge can be detected by a directional edge filter.

Region detection can be carried out by a spline wavelet. In order to detect a line type features and regions two special graph structures are defined (Fig. 6). Bidirectional information flow is used for the central node, and unidirectional for all the others. If the previously defined network propagation algorithm (in section II) is applied to these sub-graphs then the graph expands horizontally. The result of this horizontal expansion is the segmentation of the whole image.

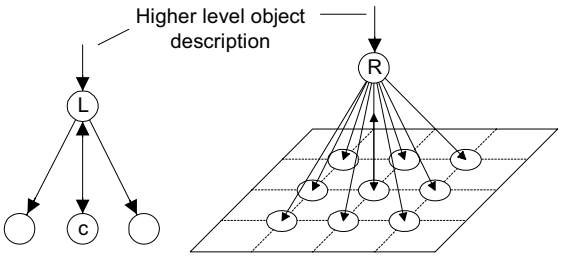

Fig. 6 Graph structures for line and region detection

This low level segmentation was tested on synthetic images containing different types of lines. An evaluation was also carried out for edge detection of real images. In both cases the result was satisfactory. The method works well, since edge pieces can be identified and connected easier if a line or curve hypothesis exists.

B. High Level Processing

A simple document processing example is used to demonstrate the method. The components of a circuit diagram

needs to be identified. Fig. 7 shows a computer generated simple circuit diagram.

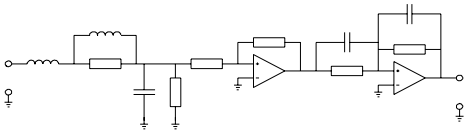

Fig. 7 Sample circuit diagram

A library of the components is created. Based on this component library the graph of the circuit diagram is created (Fig. 8). The identification is carried out by the Bayesian network belief propagation.

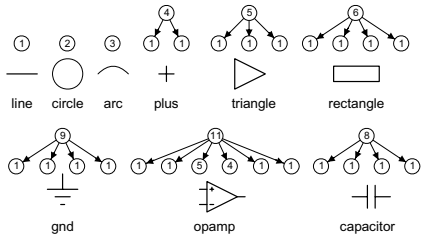

Fig. 8 Library components for the sample circuit diagram

The highest probability nodes are identified as the components of the circuit diagram. Fig. 9 shows the number of unidentified nodes during the identification process.

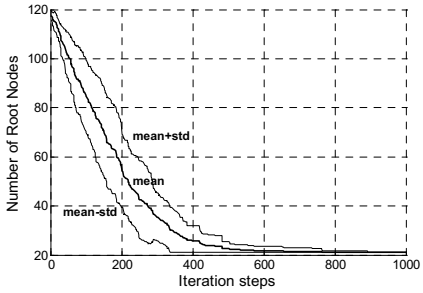

Fig. 9 The state of the network during identification

This demonstration shows that the presented method can be applied to integrate the low and high level object recognition steps. More work is needed to expand the low level image base library in order to able process more complicated images.

VI. Discussion and Conclusion

The presented method is capable of performing both low and high level object recognition. The simulation shows that the same framework can be used for integrating low and high level object recognition. The individual steps of the object recognition can be performed properly if higher level object hypothesis are available. This can be achieved by storing the different levels of object descriptions in a hierarchical graph structure. The graph structure can be fixed or dynamically modified during the recognition. Fix structure is used for example in case of quad tree methods. The advantage of the dynamic structures, that it can be adaptively adjusted to the input image. There are several ways to create a dynamic graph structure. In case of the human visual system the dynamic structure is created by moving the visual attention and focusing the eye. In our method the graph is built up adaptively, by adding small sub-graphs.

In the presented simulation only spatial relationships are used for object description. This is only sufficient to recognize simple objects. More complex object recognition requires symbolic object descriptions. The method with little modification can be also implemented for symbolic description. In the human brain higher level reasoning is the result of the combined effects of the neurons. This kind of behavior can be achieved by an upward and downward propagation on a hierarchical graph structure. The presented method is far from achieving this kind of system, but we believe that this type of research will reveal the true *theory of object recognition.*

References

[1] Barta A, Vajk I., Document Image Analysis by Probabilistic Network and Circuit Diagram Extraction, *Informatica, An International Journal of Computing and Informatics*, 29, pp. 291-301, 2005.

[2] Barta A., Vajk I, Processing Circuit Diagrams with Belief Network and Intelligent Agents., *Transactions on Information Science and Applications*, Issue 9, Vol. 2, September, pp. 1321-1329, 2005.

[3] Draper B., Hanson H., Riseman E., Knowledge-Directed Vision: Control, Learning and Integration,http://www.cs.colostate.edu/ ~draper/ publications/draper_ieee96.pdf Proceedings of the IEEE, 84(11), pp. 1625-1637, 1996.

[4] Neopolitan R. E., *Learning Bayesian networks*, Pearson Prentice Hall, 2004.

[5] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Interference*, Morgan Kauffmann Publishers, 1988.

[6] Okazaki A., Kondo T., Mori K., Tsunekawa S., Kawamoto E., An Automatic Circuit Diagram Reader With Loop-Structure-Based Symbol Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 3, pp. 331-341, May 1988.

[7] Takatsuka M., Caelli T. M., West G. A. W., Venkatesh S., An application of ''agent-oriented'' techniques to symbolic matching and object recognition, Pattern Recognition Letters 23, pp. 419–429, 2002.

[8] Siddiqi K., Subrahmonia J., Cooper D., Kimia B.B., Part-Based Bayesian Recognition Using Implicit Polynomial Invariants, *Proceedings of the 1995 International Conference on Image Processing* (ICIP), pp. 360-363, 1995.

[9] Storkey A.J., Williams C.K.I., Image Modeling with Position-Encoding Dynamic Trees, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 7, July pp. 859-871, 2003.

[10] Zou Song-Chun, Statistical Modeling and Conceptualization of Visual Patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 6, June, pp 691-712, 2003.

[11] Mallat S., *A Wavelet Tour of Signal Processing*, Academic Press, 1999

[12] Burrus C. S, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, 1998.

[13] Freeman T.W., Adelson E.H., The Design and Use of Steerable Filters, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 9, September, pp 891-906, 1991.

[14] Sung J., Bang S.J., Choi S., A Bayesian network classifier and hierarchical Gabor features for Handwritten Numeral Recognition, *Pattern Recognition Letters*, 27. pp 66-75, 2006.

[15] Deng S., Lati S., Regentova E., Document segmentation using polynomial spline wavelets, *Pattern Recognition*, 34. pp. 2533-2545, 2001.