Learning Bayesian Networks for Diverse and Varying Numbers of Evidence Sets

ZuWhan Kim ZUWHAN@IRIS.USC.EDU
Ramakant Nevatia NEVATIA@IRIS.USC.EDU
Institute of Robotics and Intelligent Systems, University of Southern California, 3737 USC-Watt Way, PHE 204, Los Angeles, CA 90089, USA

Abstract

We introduce an expandable Bayesian network (EBN) to handle the combination of diverse multiple homogeneous evidence sets. An EBN is an augmented Bayesian network which instantiates its structure at runtime according to the structure of input. We show an application of an EBN for a multi-view 3-D object description problem in computer vision. The experiments show that the proposed method gives reasonable performance even for an unlearned structure of input data.

1. Introduction

It is common in machine learning that training data and test data have the same structure. An exception is found in Bayesian networks, which allow missing data. But, in some applications, the structure of input data is not determined when the system is developed. Such a case can be found in computer vision applications dealing with multiple images; the number of images to use is not determined when the classifiers are trained. In this paper, we present an expandable Bayesian network which modifies its structure for input at runtime. We show an application of an expandable Bayesian network for a multi-view 3-D object description problem.

Computing a 3-D description (model) of an object from images is a key goal of computer vision. It is common to use more than one image of an object to retrieve 3-D information efficiently. Often times, non-intensity data, such as from radar and hyper-spectral sensors, are used as important cues to reconstruct 3-D structures. The research here has been focused on how to retrieve efficient evidence from these information sources. Evidence from these various sources is combined in a decision making process to generate 3-D models. In spite of the importance of this decision making process, it is usually done in an *ad hoc* way, such as a weighted summation where the weights are determined by a trial and error process of a human developer. Few efforts (Sarkar & Boyer, 1995; Maloof *et al*., 1998; Kim & Nevatia, 1999) are found to use formal reasoning and machine learning in this domain; however, they only deal with evidence from a single image.

Traditional classification techniques can be applied for such classification problems. Difficulties come when we incorporate the evidence from multiple images and multiple sensors. First of all, the number of views (images) is not fixed when the system is developed. Therefore, we need to combine an unknown number of features (evidence) for the classification. We need multiple copies of a classifier which act in a homogeneous way. But simply keeping multiple copies of a classifier is not enough because we may also have evidence from other sources. The evidence features from those sources are view-independent which means that they are independent of the number of views used. Therefore, we need to find a formal way to homogeneously combine the evidence from images (view-dependent) and the evidence from other sources (view-independent). Another difficulty is that we often have missing evidence because some sensor data are not available.

We introduce the use of Bayesian network in this problem domain. A Bayesian network has several advantages for this problem. First of all, it gives a formal representation of the problem and guarantees probabilistic optimum given some assumptions. Secondly, it is *transparent*, as all the parameters are represented as probabilities, making it easier for the system developer to understand their meaning in helping design improvements. Thirdly, it has strong representational power; any causal relationship can easily be represented by a Bayesian network. Finally, it handles missing data; we still get the probabilistic optimum even when some of the evidence is not available. However, the homogeneous combination of view-dependent and view-independent evidence still remains a difficulty. We present an augmentation, an expandable Bayesian network (EBN) which instantiates its structure at runtime according to the structures of input applied.

We first give a brief introduction on Bayesian networks in Section 2. Then, we present expandable Bayesian networks (Section 3). We apply an EBN to a multi-view 3-D object description system called MVS (Multi-View System for building detection and description; Noronha & Nevatia, 1997). In Section 4, we introduce MVS briefly and describe issues on applying an EBN to the system. We show experimental results in Section 5 and present the conclusion.

2. Background

We use boldface letters, such as $\boldsymbol{X}$ and $\boldsymbol{U}$, for random variables or sets of random variables while plane letters represent certain assignments. Boldface letters are also used to represent nodes of a Bayesian network.

A *Bayesian network* (Pearl, 1988) is a directed acyclic graph with random variables as its nodes. Edges represent causal relationships where all the child nodes are conditionally independent given a parent node. Each node has a set of conditional probabilities as a function of its parent nodes (represented as a conditional probability table, CPT, for quantized variables). The process of computing probabilities of some nodes given probabilities of other nodes is the process of Bayesian inference.

The structure of a Bayesian network can be determined by the known causal relations between nodes or learned by statistical methods (Friedman, Geiger, & Goldszmidt, 1997). Generally, when the structure is given by the former method, it is called a causal Bayesian network. When the nodes are discrete, CPT values can be estimated by simply collecting the statistics from examples. When a parametrized curves, such as Gaussian distributions, are used to represent the conditional probabilities, EM algorithm is often used to find the best parameters (Heckerman, 1995).

Sometimes, *hidden* nodes (unobservable nodes) are introduced for several reasons. They are used for semantic reason, compactness of the structure (Binder *et al*., 1997), or to satisfy the axioms of conditional independence (Pearl, 1986). Since the CPT values of hidden nodes cannot be observed directly, several approaches have been proposed to obtain them (Pearl, 1986; Kwoh & Gillies, 1996; Binder *et al*., 1997). Binder *et al.* (1997) introduced a gradient-based approach, which can be applied generally. In this approach, the probability of observed evidence is maximized. In each iteration (or epoch), a conditional probability $w = P(X|U)$ of a node $\boldsymbol{X}$ and its parents $\boldsymbol{U}$ is updated with

$$\Delta w = \frac{\partial}{\partial w}P(D) = \sum_{i}^{n}\frac{P(X, U|D_i)}{w}, \qquad (1)$$

where n is the number of examples, D_i is a set of observable evidence variables on the i th example, and $D = D_1, \ldots, D_n$. There is a constraint that the sum of the conditional probabilities given parent should be 1. Therefore, the resulting parameters are projected onto this constraint space.

In gradient-based approaches, simulated annealing is often used to maximize (or minimize) objective (or error) functions for avoiding local maxima (minima), which requires a significant amount of computation. But, when a reasonable set of initial parameters are given, it is less likely to find a local maximum even with a simple greedy algorithm. For Bayesian networks, we can manually give reasonably good initial parameters since they are transparent. Thus, when we have enough knowledge on the evidence variables, we can simply use a greedy algorithm to find the optimal parameters.

3. Expandable Bayesian Networks

We face several difficulties in applying Bayesian networks in a multi-view 3-D object description problem. The classifier should give an optimal decision for any number of views supplied. For example, say the number of evidence variables collected from each view is ten, the Bayesian network should have 20 nodes when two views are used, and 30 for three views. Moreover, the result should be commutative with respect to the views; it should give the same result regardless of the order of the views. Therefore, the conditional probabilities for the same evidence should be the same through all views. We augment Bayesian networks to allow modifications of their structures at runtime.

In the augmented Bayesian network (we call it an expandable Bayesian network; EBN), we have *repeatable* nodes for view-dependent evidence variables. When an input data set is applied, a repeatable node is instantiated into multiple copies according to the number of the respective input data. Figure 1 shows an example EBN. Figure 1a shows a *template* network and Figure 1b and Figure 1c show example network *instances* for various input structures. With an EBN, homogeneous combination between view-dependent variables and view-independent variables are also possible because we can simply set view-independent variables as non-repeatable nodes.

The implementation is simple; we can easily make an instance by copying repeatable nodes according to the structures of input. For each view, CPT entries and the edges (links) of the nodes are copied. When some of the edges are connected to other repeatable nodes, the instances of the edges should be linked to the instances of those nodes which are for the same view. For some applications, such as in MVS which will be described in Section 4, we only use two or three different network instances for hundreds of data sets. For such a case, we do not generate new network instances for every input data, but reuse an instance for the same configuration of input data.

To learn the parameters of an EBN, we need to first apply learning to each instance and project the learned parameters onto the constraint space that $w_j = w_k$, where w_j and w_k are the conditional probabilities of j th and k th instance of a repeatable node. Suppose $\vec{w}' = (w_1, \ldots, w_m)$, where m is

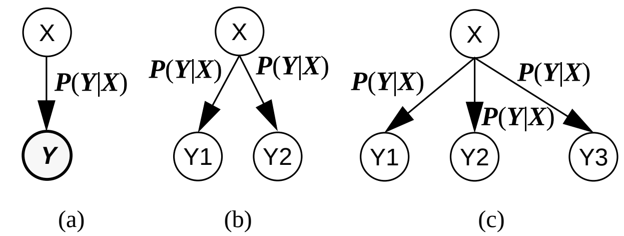


Figure 1. (a) A simple EBN with two nodes. ***Y*** is a repeatable node; (b) an instance when ***Y*** has two values; (c) an instance when ***Y*** has three values.

the number of instances, is projected onto the constraint space and the projected parameter is $\vec{w} = (w, \ldots, w)$. We get

$$(\vec{w}' - \vec{w}) \cdot \vec{w} = 0 .$$

By applying simple algebra, we find that the projection corresponds to averaging all the conditional probabilities of the instances;

$$w = \frac{1}{m}\sum_{i}^{m} w_i . \tag{2}$$

For hidden nodes, we follow the approach of Binder *et al.* (1997); apply the projection, Equation (2), to the result of Equation (1). But, unfortunately, Equation (1) cannot be directly applied to an EBN because the structures of the instances of w are different for each input data D_i. As an approximation, we reverse the order; we apply Equation (1) to the result of Equation (2). Then, we get

$$\Delta w = \sum_{i}^{n}\sum_{j}^{m} \frac{P(X_j, U|D_i)}{m w_j} , \tag{3}$$

where w_j represents the conditional probabilities of the j th instance, $\boldsymbol{X}_j$, of a repeatable node $\boldsymbol{X}$ to the parent node(s) $\boldsymbol{U}$ of $\boldsymbol{X}_j$.

4. An EBN for a Building Description System

We present an application of an expandable Bayesian network for the task of automatic detection and construction of 3-D models of buildings in aerial images. It is a task of interest for many applications such as for radiowave reachability tests for wireless communications, computer graphics, virtual reality, and mission planning. It has proven to be difficult to automate and has been an active research area.

The number of views (images) used for the modeling is usually two (stereo), but some of the recent work show good results with 6 or more images used (Baillard *et al.*, 1999). In addition, non-intensity sensor data, such as HYDICE (HYper-spectral Digital Imagery Collection Experiment) and IFSAR (Inter-Ferometric Synthetic Aperture Radar),

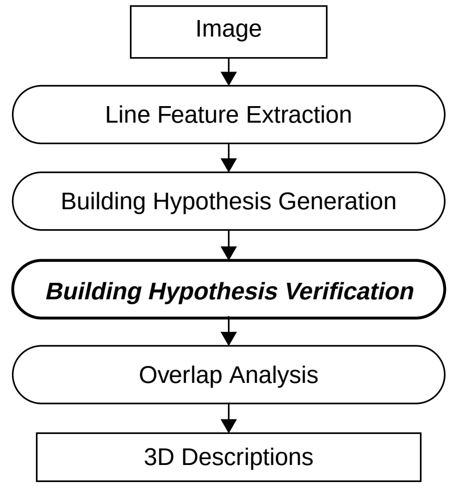


Figure 2. A flow diagram for MVS. An EBN is applied for the *Building Hypothesis Verification* procedure.

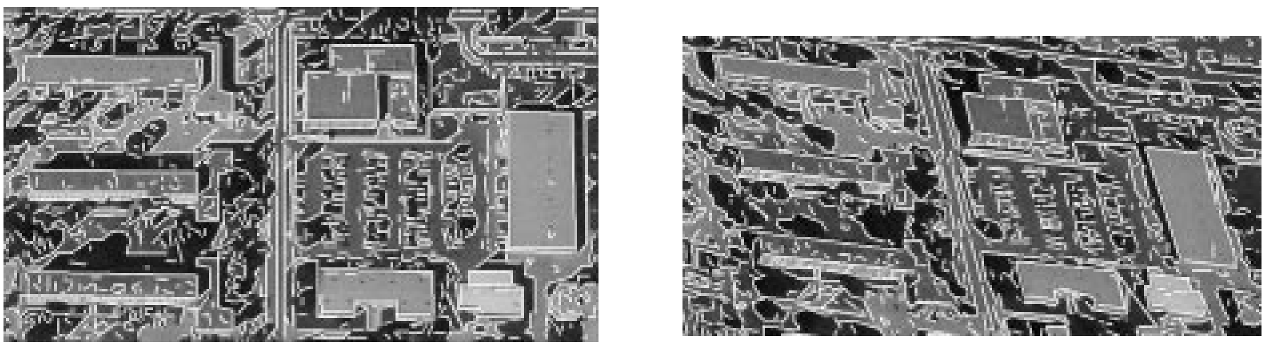
may be used. Also, often times, a DEM (Digital Elevation Map) is generated from stereo analysis and used as a separate input. The decision making is done based on the evidence collected from these various sources.

We apply an EBN to a 3-D building description system called MVS (Multi-View System for building detection and description; Noronha & Nevatia, 1997). MVS uses multiple images and sensor data to detect and describe 3-D models of rectilinear buildings. Figure 2 shows the building description procedure of MVS, and Figure 3 shows example intermediate results displayed on image windows of Fort Hood area. First, image features, such as lines, are extracted (Figure 3b). By grouping those features, 3-D building hypotheses are generated (Figure 3c). The next step is that of the verification which is to classify building hypotheses from non-building ones. Figure 3d shows a desired classification result. Since MVS generates many overlapping hypotheses in the hypothesis generation stage, an overlap analysis procedure is required to get final buildings.

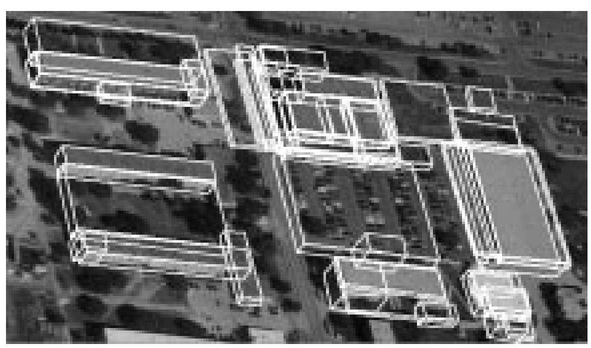
We apply an EBN to the classification task of the *hypothesis verification*. For the classification, we collect evidence support both from view-dependent (images) and view-independent (other sensors and DEM) sources. Table 1 summarizes the evidence variables used for the classification. View-dependent evidence consists of three parts, which are the evidence supports for the roof, the shadow, and the wall of a building, respectively. Some of them have continuous values while the others are either binary or ternary. For details of the evidence, see Noronha and Nevatia (1997), and Huertas, Nevatia, and Landgrebe (1999).

The next subsections describe issues of applying an EBN to this problem; representation of continuous probability distributions and construction of an EBN structure.

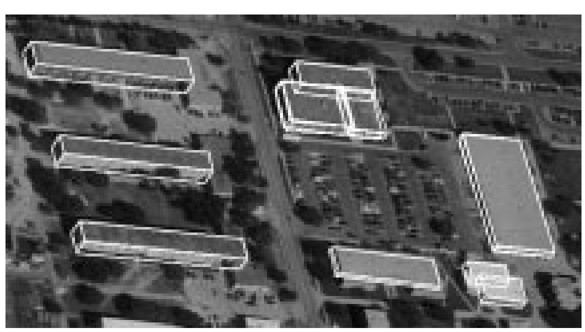
(a) Original Image Windows



(b) Line Segments



(c) 3-D Building Hypotheses



(e) Positive Hypotheses

Figure 3. (a) Aerial images taken from Fort Hood, Texas; (b) line segments extracted from (a); (c) 3-D building hypotheses generated from the line segments of (b), and; (e) a desired classification result (positive hypotheses).

Table 1. Evidence used in MVS.

Category		Evidence
view-dependent	Roof	Supporting lines (RP), Distracting lines (RN), Standard deviation (RS)
	Shadow	Strong junctions (SS), Weak junctions (SW), Horizontal lines (SH), Vertical lines (SV), Darkness (SD)
	Wall	Vertical lines (WV), Baselines (WB)
view-independent		DEM cue coverage (DEM), HYDICE cue coverage (HYDICE), IFSAR cue coverage (IFSAR)

4.1 Representation of Probability Distributions

In MVS, some of the probability distributions are multidimensional continuous ones. There are several considerations for an effective representation of continuous distributions. First of all, the number of parameters should be small enough to avoid overfitting while allowing representations of a variety of distribution shapes. Also, the distribution function should be non-negative and, for continuous node, the total area of the cut of the distribution given arbitrary values of the parent nodes should sum up 1. Moreover, for a discrete node with continuous parents, an interfunctional normalization rule is applied:

$$\sum_X \boldsymbol{f}_{\boldsymbol{X}|\boldsymbol{Y}_1, \boldsymbol{Y}_2, \dots, \boldsymbol{Y}_n}(y_1, y_2, \dots, y_n) = \mathbf{1},$$

where X are particular assignments of a continuous node $\boldsymbol{X}$, $\boldsymbol{Y}_i$ are the parent nodes of $\boldsymbol{X}$, y_i are the function parameters for $\boldsymbol{Y}_i$, $\boldsymbol{f}_{\boldsymbol{X}|\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n}(y_1, \dots, y_n)$ is a conditional probability density function, and $\mathbf{1}$ is a constant function of value 1.

With these considerations, two different approaches can be applied. The first is to use parametric curves such as normal distribution curves. Unfortunately, it is almost impossible to find a parametric multidimensional distribution which satisfies all the conditions. The problem becomes easier if all the continuous nodes have single discrete parents and no children, such as in a naive Bayesian classifier – a Gaussian distribution satisfies the condition, for example. Still, this was not applied for MVS since we could not find any good one-dimensional distribution which fits evidence distributions for MVS. Therefore, we applied the second approach which is to quantize the distributions into several discrete levels. For MVS, we discretized continuous evidence input for 5 different levels. For another approach with parametric distributions, see Pearl (1988), Heckerman (1995), and Binder *et al.* (1997).

4.2 Structuring the EBN

Learning the structure of a Bayesian network has been one of the key issues for the Bayesian network research. Usually, when the causal relationships between the nodes are not

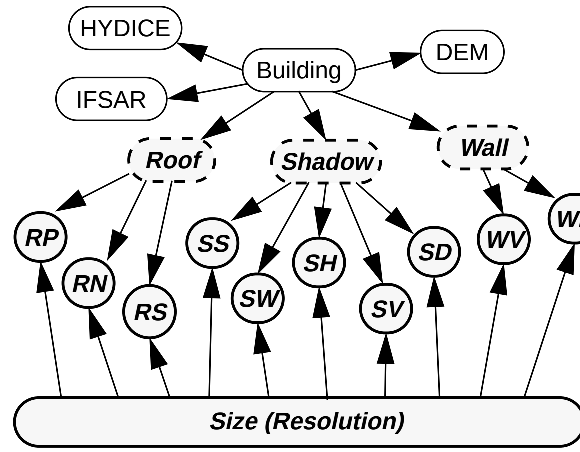


Figure 4. A causal Bayesian network for MVS. Nodes in shadowed circles (italicized letters) are repeatable nodes and nodes with dotted boundaries are hidden (unobservable) nodes.

clear, pure statistical approaches are applied to find the structure connecting the observable nodes (Heckerman, 1995; Friedman, Geiger, & Goldszmidt, 1997). However, most of these approaches follow an iterative technique and it is not easy to find a globally optimal structure; a biased network structure can be obtained according to the initial condition. For example, suppose we know that there is no direct relationship between roof standard deviation (**RS**) and shadow strong junctions (**SS**). However, there is a strong correlation between the prior distributions of **RS** and **SS** because, for positive hypotheses, both **RS** and **SS** have high values. Thus, the pure statistical approach may create a link between **RS** and **SS** which creates biased results. Also, often times, causal explanations for the resulting network are not clear, which, as a result, brings lack of *transparency*.

We designed a causal Bayesian network for MVS according to our knowledge of the evidence structure which is shown in Table 1. Figure 4 shows the causal Bayesian network. Hidden nodes are used for the evidence variables of the same categories. For example, a positive building hypothesis causes strong wall evidence, and strong wall evidence causes strong vertical lines (**WV**) and baselines (**WB**) of the wall. A node **Size** is added because each evidence node has different distributions according to the different sizes of the hypotheses. With the node **Size**, the network is not singly connected; a *conditioning method* (Russell & Norvig, 1995), which is to make multiple copies for **Size**, is used to apply a linear time algorithm.

However, purely depending on our intuitive knowledge may not be reliable either because sometimes it could be wrong. For example, the correlation between roof negative lines (**RN**) and roof standard deviation (**RS**), given **Building** on 940 examples is -0.02 while the correlation of **RS** in view 1 and view 2 is 0.65. In fact, it is matter of course that **RP**, **RN** and **RS** are conditionally independent given **Building** which means that our intuitive knowledge on using hidden node

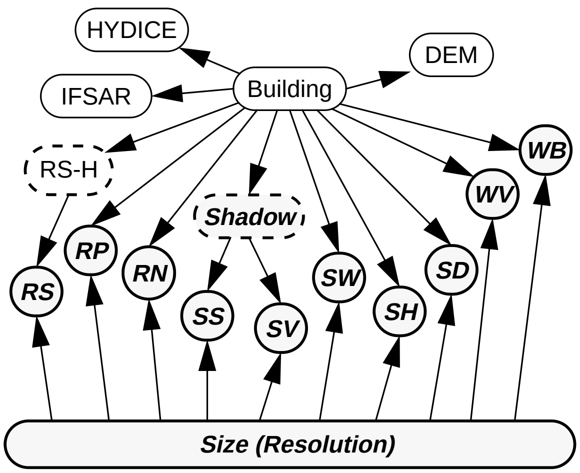


Figure 5. A Bayesian network for MVS constructed based on correlation analysis. Note that roof standard deviation (**RS**) is correlated among instances of views.

Roof was not reasonable. Instead, there is a strong correlation among instances of **RS** for each view - it is easy to figure out because, given a building, the intensity statistics of the roof on each view should be very similar. Therefore, we need to add a hidden node for **RS** to handle the correlation problem. Figure 5 shows a Bayesian network constructed based on the correlation analysis.

5. Experimental Results and Discussion

To evaluate the suggested approach, we created several learning data sets from several aerial image window sets. First, building hypotheses were generated from each window set which consists of two or three views of the same area. From the building hypotheses, we created a learning data set by displaying one hypothesis at a time to a human and asking for a decision on whether it is positive or not. 1077 learning examples were collected from 12 different image combinations of 9 different locations; 705 of them were determined to be positive and 372 to be negative hypotheses. 109 of them had evidence from both HYDICE and DEM, and 64 had only HYDICE evidence.

We implemented the EBN of Figure 5 and tested with the data set described above. Continuous variables were discretized into 5 different levels and hidden nodes were set to be 3-valued. To effectively set the initial parameters for iterative learning, an EBN without hidden nodes was implemented. First, the parameters of the EBN without hidden nodes were learned, then they were used as a guide to generate the initial parameters of the iterative learning. For example, ***P***(**SS**|**Shadow** = 2, **Size**) was set to be ***P***(**SS**|Building, **Size**) of the EBN without hidden nodes, ***P***(**SS**|**Shadow** = 0, **Size**) to be ***P***(**SS**|~Building, **Size**), and ***P***(**SS**|**Shadow** = 1, **Size**) was set to be a uniform distribution initially.

We present three experiments with this EBN. In the first two experiments, we compare the results with a naive Bayesian classifier. Note that naive Bayesian classifiers show good performances for many classification problems (Kohavi, 1995; John & Langley, 1995; Domingos & Pazzani, 1997). For the naive Bayesian classifier, evidence variables with continuous values were also discretized into 5 different levels. Naive Bayesian classifiers do not provide a formal way of dealing with view-dependent evidence; we designed and trained it for a single view, and averaged the outputs of all views and thresholded the decision. Formal combinations of view-dependent and view-independent evidence are not supported in naive Bayesian classifiers either. For a fair comparison, view-independent evidence, such as **HYDICE** and **DEM**, is not used for the first two experiments. The use of the view-independent evidence is presented in the last experiment.

For a binary classification problem, a trade-off exists between missing rate (or detection rate) and false alarm rate. For example, if we change parameters, such as a threshold, to reduce the missing rate, we get more false alarms. On the other hand, if we change the threshold to reduce false alarms, the missing rate goes up. A certain point on the trade-off curve can be emphasized according to an application. Therefore, we show the results in terms of trade-off curves rather than choose certain points on them with fixed parameters. The trade-off curves are commonly called ROC ("Receiver Operating Characteristics") curves, which have been introduced for computer vision applications by Maloof *et al*. (1998). In this paper, missing rate is the ratio of false negatives to all the positive hypotheses and false alarm rate is the ratio of false positives to all the positive hypotheses.

For statistical verification, we apply *k*-fold cross validations (Kohavi, 1995) to get ROC curves. Note that, either for the naive Bayesian classifier or for the EBN, the only parameter that determines the trade-off is the threshold value for the posterior probability, which is not involved in the learning process. Therefore, we can generate ROC curves by simply modifying the threshold. To apply a *k*-fold cross validation to get an ROC curve, we first collect the posterior probabilities of the test data from all the folds. Then, we obtain missing rates and false alarm rates with various threshold values. To get statistical variations, we apply cross validations for a number of times and calculated the means and the standard deviations (confidence intervals) of missing rates for given false alarm rates. The resulting ROC curves for the next subsections have small confidence intervals (mostly under 1% point), which illustrates the reliability of the proposed analysis.

5.1 Testing View-dependent Evidence

First, we compare the performance between the proposed EBN and the naive Bayesian classifier. For all examples (1077), stratified 5-fold cross validations were repeated for 10 times and ROC curves were obtained with confidence intervals. View-independent evidence was not used. Figure 6 shows the resulting ROC curves. The EBN shows a consis-

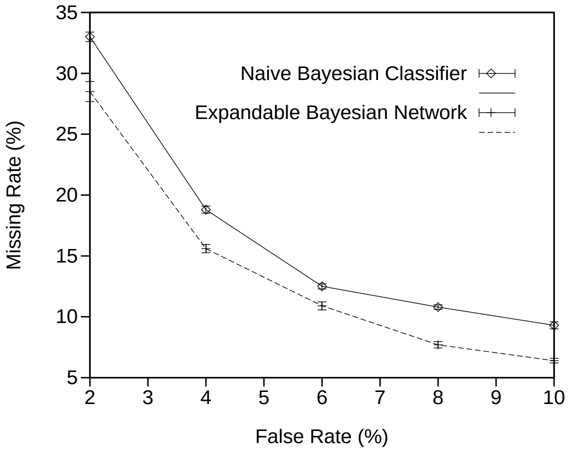


Figure 6. ROC curves with confidence intervals for the naive Bayesian classifier and the EBN. Results of 10 stratified 5-fold cross validations.

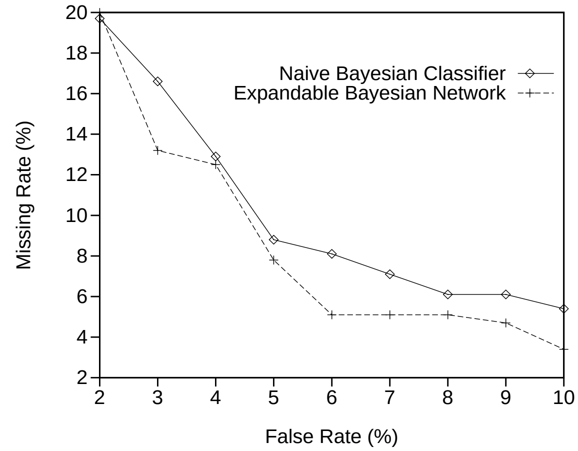


Figure 7. ROC curves for the naive Bayesian classifier and the EBN. Trained with two views and tested on three views.

tent and significant improvement over the naive Bayesian classifier. Given a false alarm rate, the improvement is up to 5 times bigger than the confidence interval. We believe that the reasonable use of the hidden nodes, the stronger representational power of the Bayesian network (i.e., the use of the node, **Size**), and, finally, a formalism for the varying numbers of views explains its higher performance over the naive Bayesian classifier.

5.2 Testing Different Numbers of Views

In this experiment, we trained the classifiers with the data set obtained from two views (804 examples) and tested on the data set obtained from three views (400 examples). Therefore, the structure of the test data is not compatible with that of the training data. Figure 7 shows the resulting ROC curves. It is easy to see that the EBN generalizes the examples of a different number of views better than the naive Bayesian classifier. Figure 8 shows error rate (missing rate) in relation to the number of training examples. To obtain the training curve of Figure 8, we fixed the false alarm rate to be 5%. Although, due to its large parameter space, the EBN does not show a good performance for a small number (80) of training examples, it shows better performance than that of the naive Bayesian classifiers for the most cases. This is highly encouraging because, unlike the previous experiment, the training data set and the test data set have very different characteristics. In the previous experiment with cross validations, the training data and the test data are randomly chosen from a common pool, but, here, they are from different site locations, and they have different structures.

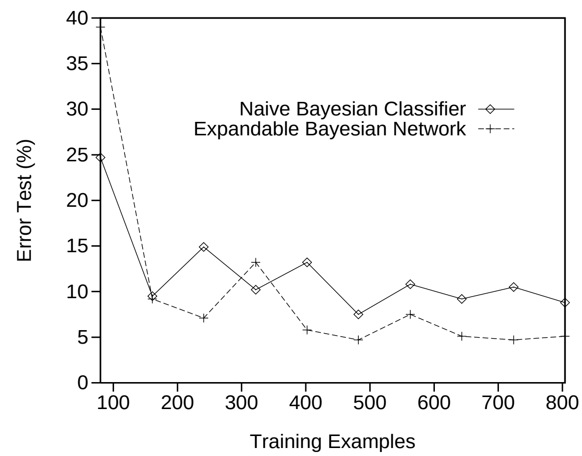


Figure 8. Error versus the number of training examples. Trained with two views and tested on three views (400 examples).

5.3 Adding View-independent Evidence

Finally, to show the homogeneous combination of view-dependent and view-independent evidence in the EBN, we present the result when some of the view-independent evidence variables, **HYDICE** and **DEM**, are incorporated. In Figure 9, we compare the result with the one without using the view-independent evidence (the same curve as the one in Figure 6). It shows an improvement mostly because **HYDICE** and **DEM** provide a good separation between positive and negative hypotheses. Here, we do not present the comparison with other classifiers because they do not provide a formal way to combine view-dependent evidence and view-independent evidence.

6. Conclusion

We have introduced an augmented Bayesian network, an expandable Bayesian network (EBN), which handles the combination of diverse multiple homogeneous evidence sets. We have shown an application of an EBN for a multi-view 3-D object description problem and proved its use. It shows

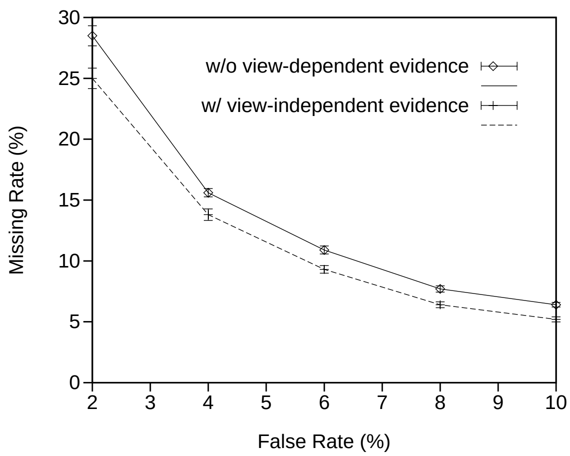


Figure 9. A result when view-independent evidence is included. Stratified 5-fold cross validations were repeated for 10 times. Compared with the result of Figure 6.

reasonable performance even for the previously unlearned structure of input data. We believe that the use of an EBN is not limited to multi-view 3-D object description problems. It provides formal reasoning for any multi-view vision applications. Moreover, it can be applied for other sensor fusion applications, where multiple homogeneous sensors are used.

Acknowledgements

This research was supported by a MURI subgrant from Purdue University under Army Research Office grant No. DAAH04-96-1-0444.

References

Baillard, C., Schmid, C., Zisserman, A., & Fitzgibbon, A. (1999). Automatic line matching and 3D reconstruction of building from multiple views. *Proceedings of ISPRS Automatic Extraction of GIS Objects from Digital Imagery* (pp. 69-80). Munich, Germany: Chair for Photogrammetry and Remote Sensing.

Binder, J., Koller, D., Russell, S., & Kanazawa, K. (1997). Adaptive probabilistic networks with hidden variables. *Machine Learning*, 29, 213-244.

Domingos, P., and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29, 103-130.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29, 131-163.

Heckerman, D. (1995). *A tutorial on learning with Bayesian networks* (Technical Report MSR-TR-95-06). Microsoft Research, Redmond, WA.

Huertas, A., Nevatia, R., & Landgrebe. D. (1999). Use of hyperspectral data with intensity images for automatic building modeling. *Proceedings of the Second International Conference on Information Fusion* (pp. 680-687). Madison, WI: Omnipress.

John, G. H., and Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338-345). San Francisco: Morgan Kaufmann.

Kim, Z., & Nevatia, R. (1999). Uncertain reasoning and learning for feature grouping. *Computer Vision and Image Understanding*, 76, 278-288.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1137-1143). San Francisco, CA: Morgan-Kaufmann.

Kwoh, C. -K., & Gillies, D. F. (1996). Using hidden nodes in Bayesian networks. *Artificial Intelligence*, 88, 1-38.

Maloof, M. A., Langley, P., Binford., T. O., & Nevatia, R. (1998). Generalizing over aspect and location for rooftop detection. *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision* (pp. 194-199). Los Alamitos, CA: IEEE Computer Society.

Noronha, S., & Nevatia, R. (1997). Detection and description of buildings from multiple aerial images. *Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp.588-594). Los Alamitos, CA: IEEE Computer Society.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29, 241-288.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco: Morgan Kaufmann.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice-Hall.

Sarkar, S. & Boyer, K. L. (1995). Using perceptual inference networks to manage vision process. *Computer Vision and Image Understanding*, 62, 27-46.