Visual Recognition of Gestures using Dynamic Naive Bayesian Classifiers

Héctor Hugo Avilés-Arriaga* Luis Enrique Sucar† Carlos Eduardo Mendoza‡
Blanca Vargas⋆
Tec de Monterrey Campus Cuernavaca
Av. Paseo de la Reforma No. 182-A Col. Lomas de Cuernavaca
C.P. 82589 Cuernavaca Morelos México
*00374765@academ01.mor.itesm.mx, †esucar@itesm.mx
‡cmendoza@alumni.princeton.edu, ⋆blanca.vargas@itesm.mx

Abstract

Visual recognition of gestures is an important field of study in human-robot interaction research. Although there exist several approaches in order to recognize gestures, on-line learning of visual gestures does not have received the same special attention. For teaching a new gesture, a recognition model that can be trained with just a few examples is required. In this paper we propose an extension to naive Bayesian classifiers for gesture recognition that we call dynamic naive Bayesian classifiers*. The observation variables in these combine motion and posture information of the user's right hand. We tested the model with a set of gestures for commanding a mobile robot, and compare it with hidden Markov models. When the number of training samples is high, the recognition rate is similar with both types of models; but when the number of training samples is low, dynamic naive classifiers have a better performance. We also show that the inclusion of posture attributes in the form of spatial relationships between the right hand and other parts of the human body improves the recognition rate in a significant way.*

1 Introduction

Visual recognition of gestures applied to command mobile robots provides a natural form of communication with mobile robots, and an alternative to speech in particular in noisy environments. When using gestures, it is possible to communicate spatial information of the type of "go there" or "go to the right" [1, 2, 3]. In the literature there exist several approaches in order to recognize gestures in terms of their motion [4, 2, 5, 6]. Hidden Markov models (*HMM*) are the most widely used technique [7, 8, 9]. Recently, dynamic Bayesian networks have been used for gesture recognition with good performance [10, 11]. However, on-line learning of visual gestures does not have received the same special attention. For teaching a robot a new gesture, a recognition model that can be trained with just a few examples is required.

In the case of HMM, the number of parameters needed to define the model grows exponentially as we increase the number of states or observation variables and their possible values [12]. Naive Bayesian classifiers (*NBC*), a special case of Bayesian networks [13], are well-known probabilistic classifiers that bypass this shortcoming, due to their inherent conditional independence assumptions. Also, in many cases, they perform better than more sophisticated non-probabilistic classification approaches, -*e.g.*, neural networks and decision trees [14, 15]. However, they are not a suitable alternative to describe stochastic domains with a dynamic nature, -*i.e.*, processes containing variables that change over time.

In this document we propose an extension to naive Bayesian classifiers that we call *dynamic naive Bayesian classifiers* (*DNBC*). We apply such a classifier to recognize a set of five dynamic gestures executed with the user's right-hand, and intended to command a mobile robot. Gestures are characterized by using four simple motion features and three posture features in the form of spatial relationships between the right hand, face and torso. This kind of information increases the recognition rate of our system in comparison with our previous work using motion features only [16]. We show that using DNBC, we can decrease the number of gestures samples that are needed for training the model. We tested the model with a set of gestures for commanding a mobile robot, and compare it with hidden Markov models. When the number of training samples is high, the recognition rate is similar with both types of models; but when the number of training samples is low, the dynamic classifiers have a better performance.

Section 2 explains briefly the visual techniques of our system. Section 3 presents naive Bayesian classifiers and some extensions, and discusses the problems posed by using them to represent dynamic processes. Section 4 describes dynamic naive Bayesian classifiers and how we used them in the gesture recognition problem. Experimen-

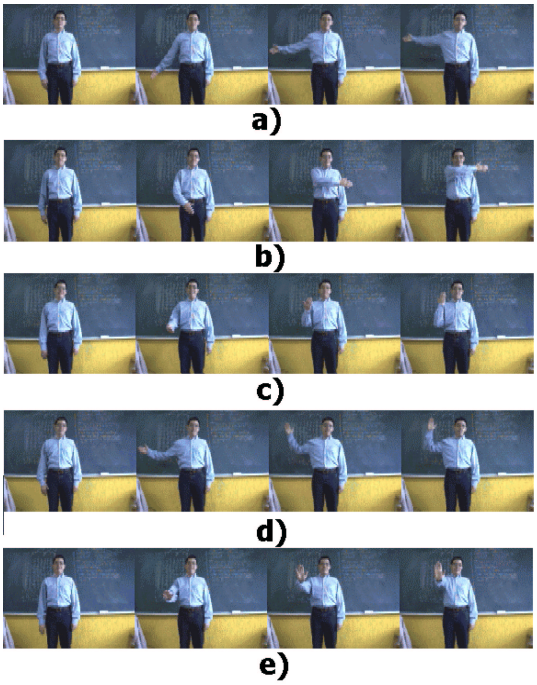

Figure 1: *Gestures considered by our system: a) go-right, b) go-left, c) come, d) attention and e) stop.*

tal results in gesture recognition are presented in section 5. Section 6 is a discussion of the experimental results. Finally, section 7 shows our conclusions.

2 Vision techniques

In order to locate and track the motion of the user, we use a skin pixel classifier and a radial scan algorithm for skin segmentation [17, 18]. The skin pixel classifier is based on histogram color models and Bayes rule for skin or non-skin pixel detection. The segmentation algorithm traces lines over the image with certain angular distance among them, from the center of the image to its edges, classifying pixels over these lines, as skin or non-skin pixels. At the same time, it uses some segmentation conditions to grow skin regions. These algorithms are applied over the image to locate the user's face and his right-hand. After the right-hand is localized, it can be tracked in the image sequence using a search window around its previous position. Some images that show face segmentation and hand tracking are presented in figure 2. We have tested this visual system under different lighting conditions (natural and artificial lighting in the laboratory) and more than 20 different users. An extended explanation of these vision techniques can be found in [19].

From the image sequence we obtain two sets of attributes that are the used as motion features to describe the gestures: *motion attributes* and *posture attributes*.

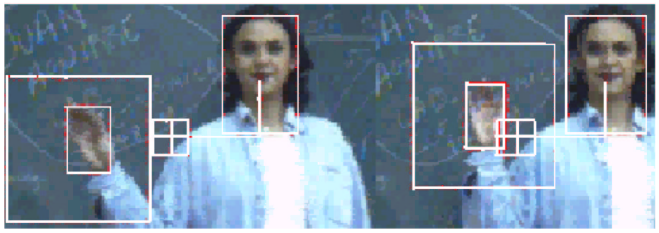

Figure 2: Tracking of the user's right hand.

Motion attributes correspond to four simple features used to describe the hand displacement: $\Delta area$ or changes in area of the hand region (rectangle), Δx or changes in hand position on the *x-axis* of the image, Δy or changes in hand position on the *y-axis* of the image and *form* or comparison between sides of the square region that segments the hand. To estimate depth motion in a simple way, we use the $\Delta area$ feature. To evaluate the hand motion between two images, each of these features takes only one of three possible values: (+), (-) or (0) that indicate increment, decrement or no change, depending on the area, position and form of the hand in the previous image.

Posture attributes represent spatial relations between the hand position and other parts of the body, such as the face and torso. These are obtained by comparing the coordinates of the regions of interest directly from the image, without discretization. Each attribute is a binary variable, that indicates if the relation is satisfied –true– or not –false. We are initially considering 3 relations: (i) hand is to the right of the head -called *right*-, hand is above of the head -*above*-, (iii) hand is over the user's torso -*torso*. The combination of the evaluation of these relations provides spatial information about the arm posture. Given that these relations implicitly establish a reference system based on the user, it is less sensible to the distance between the user and the camera -or different users- than other systems based on relative motion [19].

3 Naive Bayesian classifiers

For many years, researchers in *pattern recognition*, classification and *Machine Learning* have been interested on naive Bayesian classifiers. This is a supervised probabilistic algorithm used to determine the most likely instance c_i of a class variable C, given an instantiated set $A = \{A_1 = a_1, ..., A_n = a_n\}$ of attributes or observation variables. It is based on the prior probabilities of the class and the conditional probabilities of each attribute given the class [15]. Different evaluation tests [20, 14] have shown that naive Bayesian classifier is useful in many domains, and a simple and accurate algorithm when is compared to other more complex probabilistic and non-probabilistic classifiers.

The naive Bayesian classifier can be defined as follows:

$$P(C=c_i|A_1=a_1,...,A_n=a_n) = \frac{P(C=c_i)\prod_{j=1}^{n} P(A_j=a_j|C=c_i)}{P(A_1=a_1,...,A_n=a_n)}$$

where $P(A_1 = a_1,...,A_n = a_n) > 0$. $P(C = c_i|A_1 = a_1,...,A_n = a_n)$ is the desired probability of the class c_i given the observed data, $P(C = c_i)$, $P(A_1 = a_1,...,A_n = a_n)$, and $P(A_j = a_j|C = c_i)$ are *a priori* probabilities of the class, observation variables, and each observation variable given the class, respectively. The product $\prod_{j=1}^{n} P(A_j = a_j|C = c_i)$ corresponds to 'naive' assumptions of conditional independence among observation variables given the class.

We can consider naive Bayesian classifier as a special case of a Bayesian network when the former is described in terms of the joint distribution of the class variable C and the observation variables A_i given the class:

$$P(C,A_1,...,A_n) = P(C)\prod_{j=1}^{n} P(A_j|C).$$

This corresponds to a factored form of the joint distribution of a star-like Bayesian network [13].

Different approaches have been proposed in order to improve the accuracy of naive Bayesian classifiers. In some applications, conditional independence cannot be assumed. In order to relax the assumptions, Pazzani [21] proposes an improvement to naive Bayesian classifiers, joining pairs of variables into the same conditional distribution using an exhaustive search algorithm. Friedman *et al.* [22] explore adding edges between observation variables to reflect correlations among them. Díaz de León and Sucar [23] use this latter extension to recognize activities using Bayesian classifiers. In their approach, motion observations are recorded only once, and the class of the activity that best explains these observations is found. However, this approach loses information, because of the need of discretizing motion observations on a constant number of samples. Complete motion information is particularly important in the recognition of activities that share similar motions. Moreover, this approach does not consider effects of previous activities in the recognition of the current one. For example, if a person is walking, it is probable that this person would remain walking. For these reasons, a model that describes explicitly the temporal evolution of an activity or gesture is desirable.

4 Dynamic naive Bayesian classifiers

We propose an extension to naive Bayesian classifiers for dynamic processes that we call *dynamic naive Bayesian classifiers*. This model is composed by the set $\mathcal{A} = \{A_n^1, A_n^2, ..., A_n^T\}$, where each A_n^t for $t = 1,...,T$ is a set of n instantiated attributes or observation variables generated by some dynamic process, and $\mathcal{C} = \{C_1, C_2, ..., C_T\}$ the set of T class variables C_t generated by the same process at each time t.

We define the pair $\{\mathcal{A}, \mathcal{C}\}$ as a dynamic naive Bayesian classifier *iff* it has the following general probability distribution function:

$$P(\mathcal{A},\mathcal{C}) = P(C_1)\prod_{t=1}^{T}\prod_{j=1}^{n} P(A_j^t|C_t)\prod_{t=2}^{T} P(C_t|C_{t-1})$$

where:

- $P(C_1)$ is the initial probability distribution for the class variable C_1,
- $P(A_j^t|C_t)$ is the probability distribution of an attribute given the class, and
- $P(C_t|C_{t-1})$ is the class transition probability distribution among class variables over time.

The product $\prod_{j=1}^{N} P(A_j^t|C_t)$ stands for the naive assumptions of conditional independence among attributes given the class, as described in section 3. To represent our model, we use two standard assumptions: i) the *Markovian property*, that establishes independence of the future respect to the past given the present, and ii) the *stationarity* of the process, -*i.e.*, that transition probabilities among states are all the same through time.

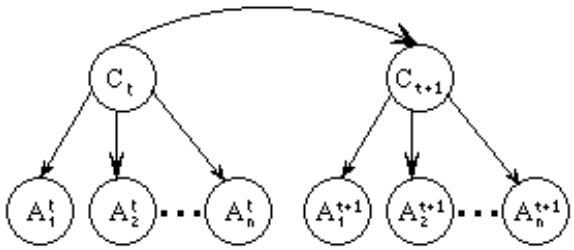

Figure 3: Dynamic naive Bayesian classifier unrolled two times.

Following the graphical representation of probabilistic independence [24], a DNBC model unrolled two times can be depicted as is shown in figure 3. Although it is possible to describe these models using an analytical form, it is simple and clearer to describe them in terms of its graph. This representation allows us to consider well-known techniques for probability propagation in Bayesian networks [24] and the *EM* algorithm for training with missing data [12].

Dynamic naive Bayesian classifiers relax the problems described in section 4 when using their "static" version. For example, in order to avoid the loss of temporal information, we can consider all the information generated by the dynamic process as attributes in a sequence, without discretizing activity observations on a constant number of samples. Then, the class that best explains the observations at each time t can be found. The effects of previous classes on the recognition of the current class is described in terms of the transition probability distribution $P(C_t|C_{t-1})$.

For gesture recognition we consider a DNBC model for each gesture $\mathcal{A}$, in a similar way as with HMM. So the *class* node C_t in the DNBC corresponds to the hidden state, S_t, at each time, t. Then we obtain the probability $P(\mathcal{A})$ of each model given the observation sequence, and select the one with higher probability.

5 Experiments and gesture recognition results

In this section we present recognition results using DNBC that use motion and posture attributes -ΔX, Δy, $\Delta area$, $form$, $right$, $above$ and $torso$- described above. These results are compared with hidden Markov models that use the same posture and motion attributes. We present also the recognition results when using DNBC with motion observations only.

In the case of dynamic naive Bayesian networks, with and without posture information, Δx and Δy are *joined* in a single node because with this topology we obtained better results than using one node per variable [16]. This operation has a direct relationship with Pazzani's work described above. An intuitive explanation about this improvement is that Δx and Δy are not independent given the $state$. This does not hold for the other attributes, that we can consider independent among them. We used a two states *ergodic* probabability transition model distribution [12] for hidden Markov models and dynamic naive Bayesian networks.

To train HMM and DNBC we used the EM algorithm. To test these models a modified version of the *Forward* algorithm [12] is used. The initial probability distribution for each model was a uniform distribution. We used the same error threshold for each model in order to define when a model has converged to a local maxima.

Our gesture data set is composed of about 150 samples of each gesture, taken from one user. The samples were taken in our laboratory, with different illumination conditions. The distance of the person to the camera varies between 3.0 and 5.0 meters. In the experiments, we randomly divided 60% of this data set for training, and 40% for testing. We trained each model by varying the number of training samples from 5% to 100% of the training data set. For testing we used the complete test data set. Recognition results of the three models are shown in figure 4. This results

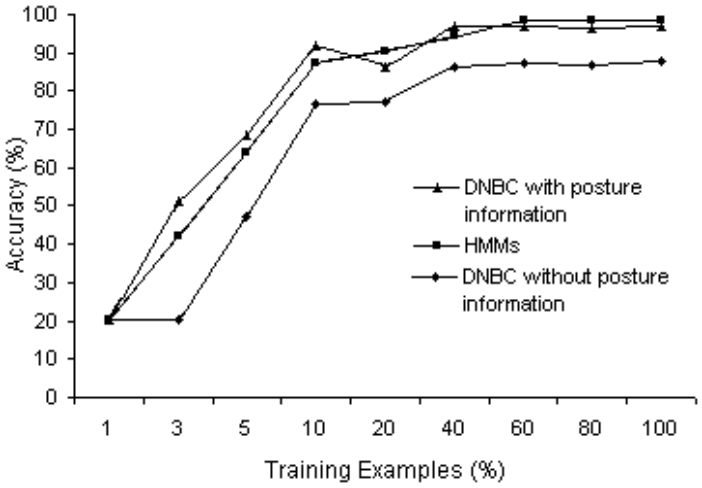

Figure 4: Average recognition rates of the dynamic naive Bayesian classifiers with and without posture information and for hidden Markov models as a function of the percent of training data.

are an average of the recognition rates of five repetitions of the same experiment. The figure presents the average of the recognition rates of the dynamic naive Bayesian classifiers (with and without posture information) and hidden Markov models as a function of the percent of training data.

Tables 1 to 3 present the confusion matrix for the 3 models, DNBC with and without posture information, and HMM, for 100 % of the training data. In the models with posture information, there is just some confusion for the "stop" gesture. Without posture information, there is more confusion for the "stop", "come" and "attention" gestures.

	Attention	Come	Go-right	Stop	Go-left
Attention	100%				
Come		100%			
Go-right			100%		
Stop	6.35%	11.11%		82.5%	
Go-left					100%

Table 1: Gestures recognition rates using dynamic naive Bayesian classifier with posture information. The average recognition rate is 96.75%

6 Discussion

Hidden Markov models and dynamic naive Bayesian networks with posture information obtained better recognition results than using DNBC without posture information. This shows that an explicit inclusion of simple spatial

	Attention	Come	Go-right	Stop	Go-left
Attention	100%				
Come		100%			
Go-right			100%		
Stop		7.41%		92.59%	
Go-left					100%

Table 2: Gestures recognition rates using hidden Markov models. The average recognition rate is 98.47%.

relationships among hands and other body parts is important to improve recognition results.

To specify DNBC model with posture information only 21 attribute parameters are required per state, 9 permutations of possible values of Δx and Δy, 6 possible values of $\Delta area$ and $form$, plus 6 parameters for *above*, *head* and *torso*. This is a reduction of 99.96% of the parameters needed to define a single state of the hidden Markov models. Reductions in the number of parameters is useful to increase the recognition results when a small training data set is available, as it is shown in figure 4. There is a significant improvement when the training samples is between 1 and 10 % of the training data set (between 2 and 15 samples per gesture). We will also expect that if the number of attributes increases, for instance if we consider other spatial relations, this performance difference for few training samples will be higher. Although there are different *parameter tieing* techniques to reduce the number of parameters and training data needed on HMM [12], with DNBC these extra calculations are not necessary, maintaining at the same time the model expresiveness and clarity.

7 Conclusions and future work

This document describes an online system to recognize dynamic gestures making use of dynamic naive Bayesian classifiers. In comparison with hidden Markov models, dynamic naive Bayesian classifiers represent dynamic process with a small number of parameters, without sacrifying the model recognition rates. When the number of training samples is high, the recognition rate is similar with both types of models; but when the number of training samples is low, the dynamic classifiers have a better performance. The observation variables in these classifiers combine motion and posture information of the user's right hand. We showed that posture information increases the recognition rates for a set of *natural* gestures intended to command a mobile robot.

As a future work we plan to conduct experiments to evaluate recognition rates of models with different transition distributions such as *left-right* models with different number of states. We also plan to increase the number gestures and to test the models with different users.

	Attention	Come	Go-right	Stop	Go-left
Attention	86.77%	2.65%	10.58%		
Come		98.44%			1.56%
Go-right			100%		
Stop	38.62%	4.23%		56.08%	1.06%
Go-left					100%

Table 3: Gestures recognition rates using dynamic naive Bayesian classifier without posture information. The average recognition rate is 87.75%.

References

[1] David Kortenkamp, Eric Huber, and Peter Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proceedings of the AAAI-96*, AAAI Press/The MIT Press, pages 915–921, 1996.

[2] Stefan Waldherr. Gesture recognition on a mobile robot. Master's thesis, Carnegie Mellon University. School of Computer Science, 1998.

[3] Roger E. Kahn. *Perseus: An Extensible Vision System for Human-Machine Interaction*. PhD thesis, The University of Chicago, 1995.

[4] James W. Davis and Aaron F. Bobick. The representation and recognition of action using temporal templates. Technical Report 402, MIT Vision and Modeling Group, 1996.

[5] Andrea Corradini and Horst-Michael Gross. Implementation and comparison of three achitectures for gesture recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2000)*, volume IV, pages 2361–2364, 2000.

[6] Andrea Corradini. Dynamic gestures as an input device for directing a mobile platform. In *ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 82–89, 2001.

[7] Thad Eugene Starner. Visual recognition of american sign language using hidden markov models. Master's thesis, MIT. Program in Media Arts and Science, 1995.

[8] Jr. Donald O. Tanguay. Hidden markov models for gesture recognition. Master's thesis, MIT. Department of Electrical Engineering and Computer Science, 1995.

[9] Stefan Becker. Sensei: A real-time recognition, feedback and training system for t'ai chi gestures. Master's thesis, MIT. Program in Media Arts and Science, 1997.

[10] Andrew David Wilson. *Adaptive Models for the Recognition of Human Gestures*. PhD thesis, MIT Program in Arts and Sciences, 2000.

[11] Vladimir Ivan Pavlovic. *Dynamic bayesian networks for information fusion with applications to human-computer interfaces*. PhD thesis, University of Illinois at Urbana-Champaign, 1999.

[12] Lawrence R. Rabiner. *Readings in Speech Recognition*, chapter A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Morgan Kaufmann Publishers, 1990.

[13] Christian Borgelt and Rudolf Kruse. *Graphical Models. Methods for Data Analysis and Mining*. John Wiley & Sons, E. U. A, 2002.

[14] Donald Michie, D.J. SpeigelHalter, and C.C. Taylor, editors. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Series In Artificial Intelligence, England, 1994.

[15] Tom M. Mitchell. *Machine Learning*. McGraw Hill Series in Computer Science, U.S.A, 1997.

[16] Héctor Avilés, Enrique Sucar, and Víctor Zárate. Dynamical arm gestures visual recognition using hidden markov models. In *IBERAMIA/SBIA, Workshop on Probabilistic Reasoning in Artificial Intelligence*, 2000.

[17] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. Technical Report CRL 98/11, Cambridge Research Laboratory, 1996.

[18] Stereo Active Visual Interface Group. Available at: http://ww.cs.toronto.edu/˜herpers/projects.html, May 28, 1999.

[19] Héctor Avilés. Reconocimiento de gestos dinámicos aplicado a robots móviles. Master's thesis, Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Cuernavaca, 2000.

[20] Pat Langley, Wayne Iba, and Kevin Thompson. An analysis of Bayesian classifiers. In *National Conference on Artificial Intelligence*, pages 223–228, 1992.

[21] M. Pazzani. Searching for dependencies in bayesian classifiers. In *Fifth International Workshop on AI and Statistics*, 1995.

[22] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[23] Rocío Díaz de León and Luis Enrique Sucar. Continuous activity recognition with missing data. In *International Conference on Pattern Recognition (ICPR'02)*, pages 92–97. Morgan Kaufmann Publishers, 2002.

[24] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan, 1988.