Publicación en Conferencia

Artículo	
Título	Dynamic Bayesian networks for gesture recognition
Autores	R. Cabañas de Paz, J. Martínez-Gomez, and M.J. Flores
Año	2011

Detalles de la conferencia	
Libro	XII Workshop de Agentes Físicos - WAF 2011, Albacete, 5 y 6 de septiembre de 2011, Proceedings
Lugar	Albacete, Spain

Dynamic Bayesian Networks for Gesture Recognition

Rafael Cabañas de Paz, M. Julia Flores, Jesús Martínez-Gómez

Abstract—**In this paper a gesture recognition system using visual information is described. The system relies on probabilistic graphical models and recognizes up to six gestures performed by a human who interacts with a computer. All these gestures can be used to manage a picture viewer. The main novelty of the system is the use of dynamic Bayesian networks and the use of both hands to perform the gestures. The proposed system was evaluated using real video sequences and the results obtained proved the goodness of the proposal.**

Index Terms—**Computer vision, robotics, probabilistic graphical models, human robot interaction.**

I. INTRODUCTION

GESTURE recognition has become one of the most important elements in human-computer interaction. For example, it can be used to create new types of interfaces based on vision [1][2], where the gesture is used to send a command. In fact, the interest on gesture recognition has increased due to the release of new game consoles devices (Sony Eye Toy[1] and Microsoft Kinect[2]) that use the human body as the main interface. Another interesting application is the interaction with robots when vision is the only available channel of communication, for example, in noisy environment [3][4][5].

Computer vision techniques for gesture recognition have to extract punctual information from the separate images acquired by the robot. The relationship between the information extracted from a sequence of consecutive images is used to identify the gesture. For this purpose, several techniques can be use, such as Neural Networks (NNs) [6], Dynamic Time Wrapping (DTW)[7] or Hidden Markov Model (HMM) [8][9]. Our hypothesis is that gesture recognition can be performed using dynamic Bayesian networks. The input for this classifier is the position of the hands that will be tracked for consecutive images. The position of the hands in this proposal is obtained thanks to the use of colour filtering techniques, but other approaches could be applied.

We evaluated the current proposal by developing a gesture recognizer for a picture viewer where six gestures are recognized. Thanks to this system, any user could control the viewer just by using his own hands. Several experiments on real scenarios proved the goodness of the proposal.

The article is organized as follows: Bayesian networks are outlined in Section II. We describe the image processing and tracking in Section III. In Sections IV and V we explain the use of Bayesian networks for tracking and gesture recognition. The experiments and results are shown in Section VI and finally, conclusions and future work are discussed in Section VII.

University of Castilla-La Mancha, Albacete, Spain

[1]http://www.eyetoy.com/

[2]http://www.xbox.com/Kinect/

II. DYNAMIC AND STATIC BAYESIAN NETWORKS

Bayesian networks (BNs) are an increasingly popular paradigm for representing problems. A Bayesian network [10][11] is a directed acyclic graph (DAG) whose nodes represent the random variables in the problem. A set of directed edges connect pairs of vertices, representing the direct dependencies (which are often causal connections) between variables. The set of nodes pointing to X are called its parents, and is denoted $pa(X)$. The relationship between variables is quantified by conditional probabilities, usually tables (CPTs), which are associated with each node, namely $P(X|pa(X))$. The CPTs together compactly represent the full joint distribution.

Figure 1 gives an example BN for a small robotics problem [12]. It shows how to model a robotic-based problem, we have chosen a simple application in order to illustrate better the BN elements. The kinematic model describes the relationship between the configuration of the robot, i.e., the joint angles, and the body posture, i.e., the positions of the body parts in space. Fig. 1.(a) shows an example of a simple 2-DOF (Degrees Of Freedom) robotic manipulator. The robot consists of two rotary joints a_1 and a_2, and five body parts $X_1, \ldots, X_5$. The first two body parts are connected rigidly. The shoulder X_2 and the upper arm X_3 are connected by the shoulder joint a_1, and so forth. This is formalised in a BN, Fig. 1.(b).

The graph does not only offer a visualisation of the system, but also, and even more important, attempts to capture (in)dependences in the model. In the simple case (two nodes) an edge indicates a probabilistic dependency, while the absence of an edge indicates probabilistic independency. However, these dependences can be found and express for any pair of non-adjacent nodes, for example by the d-separation concept [11].

The capability of a BN to *express* relationships, dependencies and independencies by its associated graph lies on its qualitative side. But these relationships are also modelled with a second element, quantitative, that forms a BN: probability distributions, as already introduced. Notice that in order to complete the BN definition for 1.(b) we will