Modality control of an active camera for an object recognition task

Felipe Trujillo-Romero, Víctor Ayala-Ramírez and René Jaime-Rivas
Universidad de Guanajuato FIMEE
Tampico 912, Col. Bellavista
36770 Salamanca, Gto.
México
e-mail: *ayalav*@salamanca.ugto.mx

Abstract

In this paper, we show an active object recognition system. This system uses a mutual information framework in order to choose the optimal parameters of an active camera for recognizing an unknown object. In a learning step, our system builds a database of all objects by means of a controlled acquisition process over a set of actions. These actions are taken from the set of different feasible configurations for our active sensor. Actions include pan, tilt and zoom values for an active camera. For every action, we compute the conditional probability density of observing some features of interest in the objects to recognize. Using a sequential decision making process, our system determines an optimal action that increase discrimination between objects in our database. This procedure iterate until a decision about the class of the unknown object can be done. We use the color patch mean over an interest region in our image as the discrimination feature. We have used a set 8 different soda bottles as our test objects and we have obtained a recognition rate of about 99%. The system needs to iterate about 4 times (that is, to perform 4 actions) before being capable of making a decision.

1. Introduction

A very important task in robot navigation is object recognition. This capability enables a robot to behave autonomously. When a robot needs to identify a preplanified path or evading an object or any other mobile object, this capability is essential[1].

An active camera is very useful in several robot tasks, specially in navigation. For example, using this sensor enables the robot to look the different objects that exist in the neighborhood of its trajectory. A survey on active sensing has been done by Mihaylova *et al*[5].

Actually there are many works that use mutual information framework to determinate the best action to move a smart sensor in order to recognize some objects. We can see an application of mutual information in the work of Denzler *et al* [3] [2]. Another works implementing the sensor control for path servoing in robot navigation are [4] [6].

In this paper, we present an active object recognition system. Our system recognizes a set of objects by using mutual information to choose the best action for an active sensor configuration in order to recognize an object in a database.

The rest of this paper is organized as follows: In section 2, we formulate the mutual information framework for object recognition. Our system implementation is presented in section 3. Test and results for our systems are shown in section 4. Final section (section 5) present our conclusions and the work to be done in the near future.

2. Problem Formulation

The active nature of a sensor is very useful when the robot tries to recognize an object. In his work, Denzler [3] [2] proposed to use a smart sensor and choose successive configuration in order to discriminate the object in a learned database.

Using mutual information in object recognition enable us to reduce uncertainty and ambiguity between objects. To take advantage of this framework, we have to find an action that maximizes mutual information between all feasible actions.

If we define x_t as the estimated state for recognition of Ω_k classes, $k \in \{1, n\}$. At each step, we are interested in computing the true state given an observation o_t. According to the information theory frame-

Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP'04)
0-7695-2074-X/04 $ 20.00 © 2004 IEEE

work, optimal estimation is given by an action a_t that optimizes mutual information. Mutual information is defined as

$$I(x_t; a_t | o_t) = H(x_t) - H(x_t | o_t, a_t) \tag{1}$$

where $H(\cdot)$ denotes the entropy of a probability distribution. Considering

$$H(x_t) = -\int_{x_t} p(x_t) \log p(x_t) dx_t \tag{2}$$

and equation 3, an optimal action a_t^* that maximizes mutual information is given by

$$a_t^* = \max_{a_t} I(x_t | o_t, a_t). \tag{4}$$

3. System Implementation

We can divide the implementation of object recognition in two parts: the learning phase and the recognition phase. In Figure 1 we can observe the flow diagram for the object recognition system.

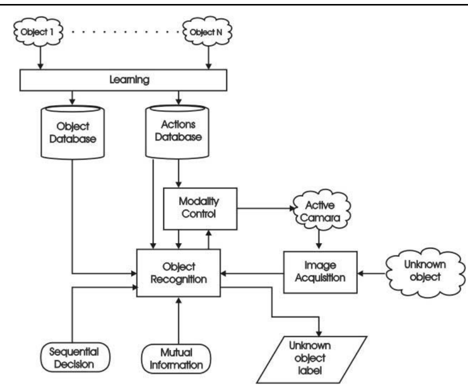

Figure 1. Flow diagram of object recognition system.

3.1. Learning phase

In this phase, the objective is to learn a conditional probability density function database. These probability functions must link actions of the active sensor to objects in database.

We consider a set three different actions: pan, tilt and zoom. These are the configuration values for an active camera. As every parameter can take different values we can divide every configurable action into a set of discrete values, that is, each action a_t is defined as:

$$a_t = (p_{k,t}, t_{k,t}, z_{k,t})^T \tag{5}$$

where $p_k, k \in \{1, n_p\}$ is a pan value,$t_k, k \in \{1, n_t\}$ is a tilt value and $z_k, k \in \{1, n_z\}$ is a zoom value, with n_p, n_t, n_z being respectively the number of discrete steps for the ranges of pan, tilt and zoom values.

We can obtain several features of an object. These features can be the size, form, chromatic intensity, edges, etc. We use these features to characterize the objects in our database. In our system, we use the chromatic intensity for modelling objects in our database. This feature was modelled by a gaussian probability density function. Here we obtain three different gaussian probability density function one for every RGB color. Gaussian distribution takes into account gaussian variations in illumination.

The Equation 6 represents a RGB intensity mean vector. We use these values to characterize the object in the database at a given active sensor modality.

In order to obtain the illumination distribution parameters, we compute mean and variance for several runs under the same sensor configuration.

$$I_p \begin{pmatrix} I_r(p) \\ I_g(p) \\ I_b(p) \end{pmatrix} = \frac{1}{n} \sum_{i=0}^{n-1} I_i \begin{pmatrix} I_r(i) \\ I_g(i) \\ I_b(i) \end{pmatrix} \tag{6}$$

In Equation 6, I_p represent the mean intensity for each image. I_i represent the intensity for each pixel of that image. And n is the total quantity of pixels of the image. With this values, we can compute the probability of observing some characteristic of the object when the sensor has been set to a given configuration.

Figure 2. Images of the objects in our database.

We used 8 different objects with similar properties. These objects are shown in Figure 2. We obtained a set of different images for each object. We use 30 different values for the pan, 5 values for tilt and 3 different values for zoom. The process was repeated 8 times, each time in a different position of the object. This

Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP'04)
0-7695-2074-X/04 $ 20.00 © 2004 IEEE

$$I(x_t; a_t|o_t) = \int_{x_t}\int_{o_t} p(x_t)p(o_t|x_t, a_t)\log\frac{p(o_t|x_t, a_t)}{p(o_t, a_t)}do_t dx_t, \tag{3}$$

procedure let us to capture an object model including distinctive properties. As we have taken images from different viewpoints of each object, we can recognize objects even if they show a different aspect from the learned ones.

3.2. Recognition phase

In this phase we present the main objective of this work: the active object recognition phase.

When we presented an unknown object (like the objects in Figure 3) from our database to the system, a sequential process started. In the beginning, we assume equal *a priori* probabilities for all class objects in our database.

First, mutual information is computed as:

$$I_0(\Omega, c|a) = \sum_{k=1}^{K} e_k(a)P_k \tag{7}$$

where the entropy e_k is defined as follows:

$$e_k(a) = \sum_{c_i} P(c_i|\Omega_k, a)\log\frac{P(c_i|\Omega_k, a)}{P(c_i|a)} \tag{8}$$

To compute the mutual information we obtain the best matching between current estimate state an the observation made in this step. We look then the action value a_0^* that maximizes mutual information.

$$a_0^* = \max_a I_0(\Omega, c|a) \tag{9}$$

Finally we need to execute this action on the sensor. Also we to update *a priori* probabilities P_k for every possible class. This reenforce probability of maybe ambiguous class and in the other side, it will weaken probabilities for non similar classes.

$$P_k = \frac{P(c_0|\Omega_k, a_0)P(\Omega_k|a_0)}{P(c_0|a_0)} \tag{10}$$

This procedure iterate in a sequential way until that system obtain the probability of the most probable class. This probability have to exceeds a given certainty threshold.

4. Test and Results

In this section we are going to present tests and result of our implementations.

In Figure 3 we can see the objects that we use for make this test. The first is the object that the system learn. And the other image is the same object with perturbations. We use a label to make this perturbation.

We presented this alterate object to system in the execution of an object recognition procedure and the results are shown in Figures 4 and 5. In this figures we can observe that system oscillates between two objects classes for the test inmage: the correct one and other with similar features. In Figure 4, the system decided to assigned the correct class label (Some small probability is given to object type 8).

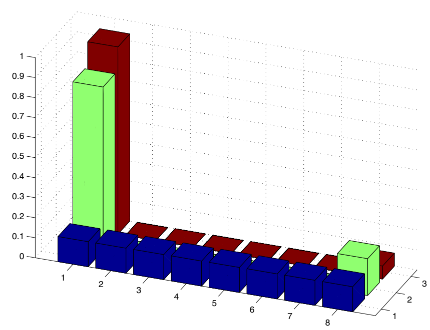

Figure 4. Recognition of object 1.

In other system run, the object was identified as one of type 8. In this case, the system gives a grater probability to object type 1 than that given to object 8 in the previous case. This situation is shown in Figure 5. We can get some important conclusions from test results in Figure 5. We can take more insight into the behavior of our system. We can observe that our system assigns a large probability to object type 1 in the beginning of the recognition step. However, this probability decreases and guides the system to decide for other object with similar features. In this case, the object type 8 is the object that have better similarity with the unknown object.

This test was run several times and approximately 50% the object was recognized as belonging to class type 8. An important feature is that system only oscillates between the real object belonging to class 1 and the object belonging to class type 8.

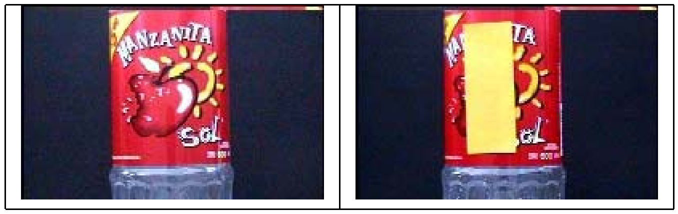

Figure 3. Models used in this test.

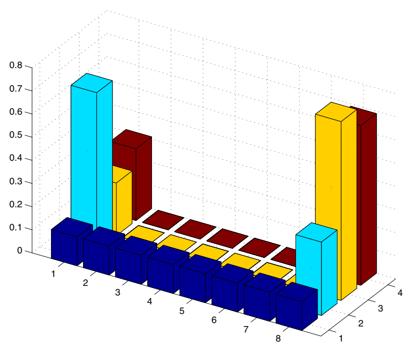

Figure 5. Recognition of object 1 as object 8.

5. Conclusion and Perspectives

In this paper we have presented an object recognition system that have a good recognition rate (99%). It achieves this recognition rate even if it is sensitive to illumination changes. Another problem presented by the system is when the object appearance changes from the learned appearance.

Our future work will be directed towards:

- Increasing the number of features to take into account. With this increase, we could recognize objects in a more reliable way.
- Implementing the use of a kind of active handler in combination with the active sensor. We can use, for example, a turntable like active handler. This active handler will help in the recognition phase to let system choose pairs views-actions where the object can be recognized more easily.
- Modelling the conditional probability functions of normal distributions by using fuzzy logic rules could be useful in two issues: firstly, it could save amounts of memory because we could store the probabilities functions as a set of fuzzy rules. The other advantage is to implement non parametric conditional functions for the pairs objects-actions.
- Using a more suitable color space in order to overcome illumination problems during outdoor visual navigation.
- Finally we can extend this method for its application to landmark recognition in visual navigation.

Acknowledgements

This work has been partially funded by México's CONACyT project I39328A and the LAFMI project "Concepción de funciones de percepción y planificación para la navegación topológica de un robot móvil en ambiente semi-estructurado interior o natural".

References

[1] G. de Souza and A. C. Kak. Vision for mobile robot navigation: a survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(2):237–267, February 2002.

[2] J. Denzler and C. Brown. Optimal selection of camera parameters for state estimation of static systems: An information theoretic approach. Technical Report 732, The University of Rochester, New York, August 2000.

[3] J. Denzler, C. Brown, and H. Niemann. *Pattern Recognition – 23rd DAGM Symposium*, chapter Optimal Camera Parameters for State Estimation with Applications in Object Recognition, pages 305–312. Springer, Berlin, September 2001.

[4] A. Marín-Hernández, V. Ayala-Ramírez, and Michel Devy. Collaborative tracking for robotic tasks. In *Proc. of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NE, October 2003. IEEE.

[5] L. Mihaylova, T. Lefevbre, H. Bruyincxx, K. Gadeyne, and J. De Schutter. A comparison of decision making criteria and optimization techniques for active robot sensing. In *Proc. of the Fifth Int. Conf. on Numerical Methods and Applications NMA'02*, Borovets, Bulgaria, August 2002.

[6] Felipe Trujillo-Romero, Víctor Ayala-Ramírez, Antonio Marín Hernández, and Michel Devy. Control de modalidades en un modelo de funcionalidad visual: aplicación a una tarea de localización de un robot móvil. In *Actas de Conielecomp-03*, pages 206–211, Puebla, México, Febrero 2003. UDLAP.

Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP'04)
0-7695-2074-X/04 $ 20.00 © 2004 IEEE