Bayesian Spatiotemporal Context Integration Sources in Robot Vision Systems*

Rodrigo Palma-Amestoy, Pablo Guerrero, Javier Ruiz-del-Solar, and C. Garretón

Department of Electrical Engineering, Universidad de Chile
{ropalma,pguerrer,jruizd}@ing.uchile.cl

Abstract. Having as a main motivation the development of robust and high performing robot vision systems that can operate in dynamic environments, we propose a bayesian spatiotemporal context-based vision system for a mobile robot with a mobile camera, which uses three different context-coherence instances: current frame coherence, last frame coherence and high level tracking coherence (coherence with tracked objects). We choose as a first application for this vision system, the detection of static objects in the RoboCup Standard Platform League domain. The system has been validated using real video sequences and has presented satisfactory results. A relevant conclusion is that the last frame coherence appears to be not very important in the tested cases, while the coherence with the tracked objects appears to be the most important context level considered.

1 Introduction

Visual perception of objects in complex and dynamical scenes with cluttered backgrounds is a very difficult task which humans can solve satisfactorily. However, computer and robot vision systems perform very badly in this kind of environments. One of the reasons of this large difference in performance is the use of context or contextual information by humans. Several studies in human perception have shown that the human visual system makes extensive use of the strong relationships between objects and their environment for facilitating the object detection and perception [1][3][5][6][12].

Context can play a useful role in visual perception in at least three forms: reducing the perceptual aliasing, increasing the perceptual abilities in hard conditions, speeding up the perceptions. From the visual perception point of view, it is possible to define at least six different types of context: low-level context, physical spatial context, temporal context, objects configuration context, scene context and situation context. More detailed explanation can be found in [17]. Low-level context is frequently used in computer vision. Most of the systems performing color or texture perception use low-level context in some degree (see for example [13]). Scene context have been also addressed in some computer vision [10] and image retrieval [4] systems. However, we believe that not enough attention has been given in robotic and

* This research was partially supported by FONDECYT (Chile) under Project Number 1090250.

L. Iocchi et al. (Eds.): RoboCup 2008, LNAI 5399, pp. 212–224, 2009.

computer vision to the other relevant context information here mentioned, especially in spatiotemporal context levels.

Having as main motivation the development of a robust and high performing robot vision system that can operate in dynamic environment in real-time, in this work we propose a generic vision system for a mobile robot with a mobile camera, which employs spatiotemporal context. Although other systems, as for example [1][3][5][12], use contextual information, to the best of our knowledge this is one of the first work in which context integration is addressed in an integral and robust fashion. We believe that the use of a bayesian-based context filter is the most innovative contributions of this work.

We choose as a first application for our vision system, the detection of static objects in the RoboCup Standard Platform (SP) League domain. We select this application domain mainly because static objects in the field (beacons, goals and field lines) are part of a fixed and previously known 3D layout, where it is possible to use several relationships between objects to calculate the defined context instances.

This paper is organized as follows. The proposed spatiotemporal context based vision system is described in detail in section 2. In section 3, the proposed system is validated using real video sequences. Finally, conclusions of this work are given in section 4.

2 Proposed Context Based Vision System

The proposed vision system is summarized in the block diagram shown in figure 1. The first input used is the sensor information given by the camera and encoders (odometry). Odometry is used in several stages to estimate the horizon position and to correct the images between the different frames (see [18] for more details). The image of the camera is given to the preprocessor module, where color segmentation is performed and blobs of each color of interest are generated. These blobs are the first object candidates. We will call $\{C^k\}$ to the object candidates at time step k.

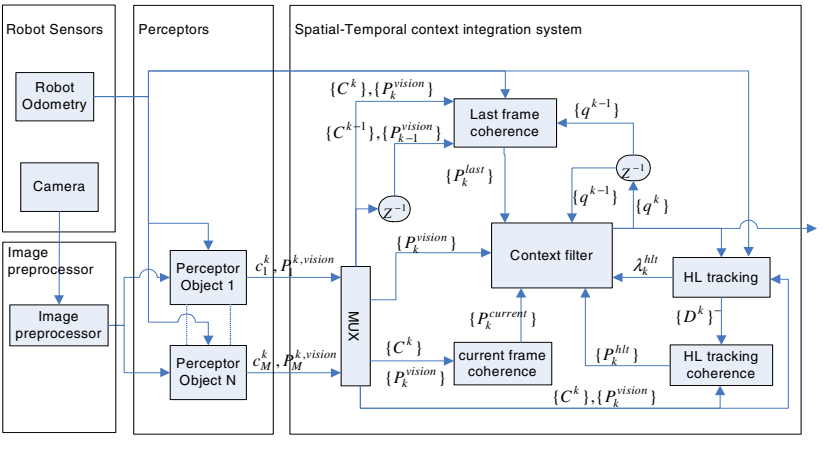


Fig. 1. Block diagram of the proposed general vision system

Each perceptors module evaluates the blob candidates with a model of the real objects. This module selects the best candidates c_i^k, and calculates an a priori probability that the candidate is correctly detected. These probabilities in the time k are called $\{P_k^{vision}\}$.

The spatiotemporal context integration stage has five modules. Current frame coherence, last frame coherence and high level tracking coherence modules give a measurement of the coherence of each current candidate with the respective context instance: with all other current detections, with last detections, and with high-level tracking estimations. The output of these modules are the probabilities $\{P_k^{current}\}$, $\{P_k^{last}\}$, and $\{P_k^{hlt}\}$. The HLT (High-Level Tracking) module maintains an estimation of the objects' pose based on the information given by all detected objects along the time. This module calculates a confidence of these estimations, which is called λ_k^{hlt}. The context filter module uses the information $\{P_k^{current}\}$, $\{P_k^{last}\}$, $\{P_k^{hlt}\}$, and λ_k^{hlt} to calculate an a posteriori probability for each current candidate given all the context instances mentioned before. The context filter module is the more relevant contribution of this work. It implements a bayesian filter to integrate all context information given by each module exposed above. This module can be represented by a function of all context instances whose result is the a posteriori probability that an object is correctly detected given all past detections, which is called ($\{q^k\}$).

2.1 Perceptors

Let c_i^k be the observation of the object i at time step k defined by $c_i^k = \left(\mathbf{x}_i^k, \mathbf{y}_i^k, \eta_i^k, \sigma_{\mathbf{y}_i^k}, \sigma_{\eta_i^k}\right)^T$, where $\mathbf{x}_i^k$ is the relative pose of the object with respect to the robot, and ($\mathbf{y}_i^k, \eta_i^k$) and ($\sigma_{\mathbf{y}_i^k}, \sigma_{\eta_i^k}$) are the horizon position and angle with their corresponding tolerances. Each object of interest has a specialized perceptor that evaluates some intrinsic characteristic of the candidate c_i^k related with the class K^i. We define $[c_i^k]^{OK}$ and $[c_i^k]^{NO.OK}$ as the events where c_i^k has been generated or not by the object i. The output of the preceptor of the candidate c_i^k can be defined as the probability of the event $[c_i^k]^{OK}$ given the observation c_i^k:

$$P^{vision} = P\left([c_i^k]^{OK} \mid c_i^k\right) \tag{1}$$

This definition has a term not explicitly mentioned in the equation. All candidates in this work have passed through binary filters, and have been characterized with some degree of error in perceptors stages. We have shelved this part of the perceptors in these equations, but that is not a problem, because all the probabilities have the same conditional part in this work, and all algebraic developments have the same validity.

2.2 HLT Module

The HLT module is intended to maintain information about all the objects detected in the past, although they are currently not observed (for instance, in any moment you have an estimation of the relative position of the objects that are behind you). This tracking stage is basically a state estimator for each object of interest; where the state to be estimated, for fixed objects, is the relative pose $\mathbf{x}_k^i$ of the object with respect to the robot and not in the camera space. For this reasons it is possible to say that the HLT module needs a transformation of the coordinated system. We define $F^k = T(C^k)$ and $f_j^k = t(c_j^k)$, where $T()$ and $t()$ correspond to the transformation functions from the camera point of view to the field point of view. The relative pose of the objects respect to the robot, is less dynamic and more traceable than the parameters in the camera point of view.

2.3 Context instances Calculation in the RoboCup SP League

We will consider three different context instances separately. The first one is the coherence filtering between all detected objects in the current frame. The second one is the coherence filtering between current and last frame´s detected object, and the third one is the coherence filtering with high level tracking estimator.

We have preferred consider last frame coherence and HLT coherence separately, because last detections may have very relevant information about objects in the current frame. Due that the HLT has an estimation of the object´s pose, which is given by a bayesian filter that integrates the information of the all detected objects in the time; the information of the last frame has a low importance in HLT. In the other hand, we think that to considerate more than one past frame is too noisy and it is better to have an estimation with HLT in these cases.

In this approach we have used two kinds of relationships that can be checked between physical detected objects. The first one, *Horizon Orientation Alignment*, must be checked between candidates belonging to the same image, or at most between candidates of very close images, when the camera's pose change is bounded. The second one, *Relative Position or Distance Limits*, may be checked between candidates or objects of different images, considering the movements of the camera between images:

- Horizon Orientation Alignment. In the RoboCup´s environment, several objects have almost fixed orientation with respect to a vertical axis. Using this quality, it is possible to find a horizon angle that is coherent with the orientation of the object in the image. Horizontal angles of correct candidates must have similar values, and furthermore, they are expected to be similar to the angle of the visual horizontal obtained from the horizontal points.
- Relative Position or Distance Limits. In some specific situations, objects are part of a fixed layout. The robot may know this layout a priori from two different sources: previous information about it, or a map learned from observations. In both cases, the robot can check if the relative position between two objects, or at least their distances (when objects has radial symmetry), is maintained.

2.3.1 Current Frame Coherence

We can define the current frame context coherence as the probability of the event $[c_i^k]^{OK}$ given all other detection in the current frame. If $C^k = \left\{c_i^k\right\}_{i=0}^{M}$ is the vector of observations in time step *k*, then the current frame context coherence may be defined like $P^{curr} = P\left([c_i^k]^{OK} \mid C^k\right)$.

However, this probability must be calculated with comparisons between pairs of objects given that they are correctly detected $P\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right)$. In section 2.4 we will show the relation established between these probabilities.

In a RoboCup SP League soccer field, there are many objects that have spatial relationships between them. These objects are goals, beacons and field lines. This static objects in the field are part of a fixed and previously known 3D layout, thus it is possible to use several of the proposed relationships between objects to calculate a candidate's coherence (for more details about object configuration in RoboCup Four Legged League, see description in [14]).

We consider three terms to calculate the coherence between two objects in the same frame:

$$\begin{aligned} &P\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right) = \\ &\quad P_{hor}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right) \cdot P_{dist}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right) P_{lat}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right) \end{aligned} \tag{2}$$

In this equation, horizontal coherence is related with horizontal position and orientation alignment. In the sense of the relative position and distance limits, we are able to use distances between the objects and laterality. Laterality and distances information comes from the fact that the robot is always moving in an area that is surrounded by the fixed objects. For that reason, it is always possible to determine, for any pair of candidates, which of them should be to the right of the other and their approximated distances.

We define the horizontal coherence term using a triangular function:

$$\begin{aligned} &P_{hor}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right) = \\ &\quad tri\left(\Delta\eta_k^{i,j}, \sigma_{\Delta\eta_k^{i,j}}\right) \cdot tri\left(\Delta\eta_k^{j,i}, \sigma_{\Delta\eta_k^{j,i}}\right); tri(\Delta x, \sigma) = \begin{cases} 1 - \dfrac{\Delta x}{\sigma} & \Delta x < \sigma \\ 0 & otherwise \end{cases} \\ &\Delta\eta_k^{i,j} = \left|\eta_k^i - \eta_k^{i,j}\right|; \eta_k^{i,j} = \eta_k^{j,i} = \measuredangle\left(\mathbf{y}_k^i - \mathbf{y}_k^j\right) \\ &\sigma_{\Delta\eta_k^{i,j}} = \sigma_{\Delta\eta_k^{j,i}} = \sigma_{\eta_k^i} + \sigma_{\eta_k^j} + \tan^{-1}\left(\frac{\sigma_{\mathbf{y}_k^i} + \sigma_{\mathbf{y}_k^j}}{\left|\mathbf{y}_k^i - \mathbf{y}_k^j\right|}\right) \end{aligned} \tag{3}$$

The distance coherence $P_{dist}\left([c_i^k]^{OK} \mid c_j^k\right)$ is also approximated using a triangular function:

$$P_{dist}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right) = tri\left(\Delta\mathbf{x}_k^{i,j}, \sigma_{\Delta\mathbf{x}_k^{i,j}}\right); \Delta\mathbf{x}_k^{i,j} = \left|\mathbf{x}_k^i - \mathbf{x}_k^j\right| \tag{4}$$

where $\mathbf{x}_k^i$, $\mathbf{x}_k^j$ are the relative detected positions of c_i^k and c_j^k respectively.

The lateral coherence $P_{lat}\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right)$ is defined as binary function, which is equal to 1 if the lateral relation between c_i^k and c_j^k is the expected one, and 0 otherwise.

2.3.2 Last Frame Coherence

Analogously to the previous subsection, we can define the coherence between the candidate and the objects in the past frame as $P^{last} = P\left([c_i^k]^{OK} \mid C^{k-1}\right)$. However as well as the previous subsection, we just can calculate the relationship between a pair of objects given that they are correctly detected. We assume the same model that in the current frame:

$$\begin{aligned} &P\left([c_i^k]^{OK} \mid [c_j^{k-1}]^{OK}\right) \\ &\quad = P_{hor}\left([c_i^k]^{OK} \mid [c_j^{k-1}]^{OK}\right) \cdot P_{dist}\left([c_i^k]^{OK} \mid [c_j^{k-1}]^{OK}\right) \cdot P_{lat}\left([c_i^k]^{OK} \mid [c_j^{k-1}]^{OK}\right) \end{aligned} \tag{5}$$

The calculation of these terms is totally analogous with the current frame coherence, with only two differences: $\mathbf{y}_k^j$ and η_k^j are modified using the encoder´s information and the tolerances $\sigma_{\eta_k^j}$ and $\sigma_{\mathbf{y}_k^j}$ are increased to meet the uncertainty generated by the possible camera and robot movements.

2.3.3 High Level Tracking Coherence

The HLT module maintains an estimation of the objects with the information given by all time steps from zero until *k-1*. Let $\left\{F^n\right\}_{n=0}^{k-1}$ be the information of all frames from zero to *k-1*, we call $\{D^k\}^- = \{d_i^k\}_{n=0}^M$ the estimation calculated by the HLT using $\left\{F^n\right\}_{n=0}^{k-1}$. The HLT coherence will be defined as $P^{hlt} = P\left([f_k^i]^{OK} \mid \{D^k\}^-\right)$.

Again, the relation between two objects needs to be calculated.

$$P\left([f_i^k]^{OK} \mid [d_j^k]^{OK}\right) = P_{lat}\left([f_i^k]^{OK} \mid [d_j^k]^{OK}\right) P_{dist}\left([f_i^k]^{OK} \mid [d_j^k]^{OK}\right) \tag{6}$$

In this case we can not consider the terms related with horizon alignment but just the term related with relative position and distances limits. The calculus of P_{lat} and P_{dist} are the same that in the current coherence, but the observations c_i^k must be converted to the field point of view as was written on the equation.

When an object is detected and it is not being tracked, the HLT module creates a new state estimator for it and initializes it with all the values coming from the detection process. In particular, the coherence is initialized with the a posteriori probability obtained by the candidate that has generated the detection. However, as

the robot moves, odometry errors accumulate and high-level estimations become unreliable. If a set of high-level estimations is self-coherent, but moves too far from real poses of tracked objects, then all the new observations may become incoherent and will be rejected. To avoid this situation, high-level estimations are also evaluated in the coherence filter. In order to inhibit the self-confirmation of an obsolete set of estimations, the confidence HLT_k^{conf} is only checked with respect to the current observations, but it is smoothed to avoid a single outlier observation discarding all the objects being tracked. Thus, the confidence of a tracked object is updated using:

$$\{\lambda_k^{conf}\}_i = \beta \cdot \{\lambda_{k-1}^{conf}\}_i + (1-\beta) \cdot \frac{\sum_{j=1}^{N} P\left([d_i^k]^{OK} \mid [f_j^k]^{OK}\right) \cdot P\left([f_j^k]^{OK} \mid f_j^k\right)}{\sum_{j=1}^{N} P\left([f_j^k]^{OK} \mid f_j^k\right)} \tag{7}$$

where β is a smoothing factor.

2.4 Context Filter

Let us define the probability a posteriori that we are interested. The most general spatiotemporal context that we can define is the probability that an object is correct, given all other detections from init frame to current frame *k*. Then we define q_i^k as:

$$q_i^k = P\left([c_i^k]^{OK} \mid \left\{C^n\right\}_{n=0}^{k}\right) \tag{8}$$

We can assume independence between detections in different times as is shown in [19]. Then we have $P\left(\left\{C^n\right\}_{n=0}^{k} \mid [c_i^k]^{OK}\right) = P\left(C^k \mid [c_i^k]^{OK}\right) \cdot P\left(\left\{C^n\right\}_{n=0}^{k-1} \mid [c_i^k]^{OK}\right)$. We apply Bayes theorem in a convenient way:

$$\begin{aligned} q_k^i &= \frac{P\left(C^k \mid [c_i^k]^{OK}\right) \cdot P\left(\left\{C^n\right\}_{n=0}^{k-1} \mid [c_i^k]^{OK}\right) \cdot P\left([c_i^k]^{OK}\right)}{P\left(\left\{C^n\right\}_{n=0}^{k}\right)} \\ q_k^i &= \frac{P\left([c_i^k]^{OK} \mid C^k\right) \cdot P\left([c_k^i]^{OK} \mid \left\{C^n\right\}_{n=0}^{k-1}\right)}{P\left([c_i^k]^{OK}\right)} \end{aligned} \tag{9}$$

Here, $P\left([c_i^k]^{OK} \mid C^k\right)$ is the coherence between objects in the current frame P^{curr} and $P\left([c_k^i]^{OK} \mid \left\{C^n\right\}_{n=0}^{k-1}\right)$ have the information about all other detections in the past. In our case we will separate it into the last frame coherence and HLT coherence.

2.4.1 Current Frame Coherence Integration

To calculate the current frame coherence, we decompose $P\left([c_i^k]^{OK} \mid C^k\right)$ in:

$$P\left([c_i^k]^{OK} \mid C^k\right) = \frac{P\left(C^k \mid [c_i^k]^{OK}\right) \cdot P\left([c_i^k]^{OK}\right)}{P\left(C^k\right)}$$
$$P\left(C^k \mid [c_i^k]^{OK}\right) = \prod_{j=1}^{M} P\left(c_j^k \mid [c_i^k]^{OK}\right) \tag{10}$$

$$P\left(c_j^k \mid [c_i^k]^{OK}\right) = \left[P\left(c_j^k \mid [c_j^k]^{OK}\right) \cdot P\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right)\right] + \left[P\left(c_j^k \mid [c_j^k]^{NO.OK}\right) \cdot P\left([c_j^k]^{NO.OK} \mid [c_i^k]^{OK}\right)\right] \tag{11}$$

Note that we have applied total probabilities theorem to obtain the probability that we need as a function of $P\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right)$ and $P\left(c_j^k \mid [c_j^k]^{OK}\right)$. Note that $P\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right)$ is symmetric, then $P\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right) = P\left([c_i^k]^{OK} \mid [c_j^k]^{OK}\right)$ is the output of the calculus of current context coherence defined in (2). $P\left(c_j^k \mid [c_j^k]^{OK}\right)$ is the a posteriori probability of perceptor modules, so we can apply Bayes and obtain the a priori probability of perceptor modules:

$$P\left(c_j^k \mid [c_j^k]^{OK}\right) = \frac{P\left([c_j^k]^{OK} \mid c_j^k\right) \cdot P\left(c_j^k\right)}{P\left([c_j^k]^{OK}\right)} \tag{12}$$

where $P\left([c_j^k]^{OK} \mid c_j^k\right)$ is directly the output of perceptor module defined in (1). Clearly, $\Pr\left([c_j^k]^{NO.OK} \mid [c_i^k]^{OK}\right) = 1 - \Pr\left([c_j^k]^{OK} \mid [c_i^k]^{OK}\right)$ and applying Bayes and complementary probabilities, the term $\Pr\left(c_j^k \mid [c_j^k]^{NO.OK}\right)$ can be calculated as $P\left(c_j^k \mid [c_j^k]^{NO.OK}\right) = \frac{P\left(c_j^k\right) \cdot \left(1 - P\left([c_j^k]^{OK} \mid c_j^k\right)\right)}{P\left([c_j^k]^{NO.OK}\right)}$.

All other probabilities no explicitly calculated here, can be estimated statistically.

2.4.2 Past Frames Coherence Integration

The term $P\left([c_k^i]^{OK} \mid \left\{C^n\right\}_{n=0}^{k-1}\right)$ considers the information of all detected objects along the time. Each candidate can be represented into the camera coordinate system, or into the field coordinate system. Assuming independence between the probabilities

calculated in both coordinate systems, the problem was decomposed considering both coordinate systems separately. In the camera coordinate system, just the last frame detections are considered, because more than one past frame would introduce too much noise to the problem, due to the highly dynamical nature of the objects. Hence, we just need to calculate the term $P\left([c_k^i]^{OK} \mid C^{k-1}\right)$. In future works, it is possible to face the problem with more details, considering an estimation of the objects in the camera coordinate system to take into account more than one past frame. On the other hand, the HLT module gives an estimation of the objects in the field coordinate system, considering all detections along the time. The HLT module performs a bayesian estimation of the objects; therefore, we can assume the Markov principle, which say that the probability $P\left([f_k^i]^{OK} \mid \left\{F^n\right\}_{n=0}^{k-1}\right)$ can be substituted by $P\left([f_k^i]^{OK} \mid \{D^k\}\right)$ (see subsection 2.3.3). Applying Bayes and assuming F^k, C^k statistically independent, we obtain:

$$P\left([c_i^k]^{OK} \mid F^{k-1}, C^{k-1}\right) = \frac{P\left(\{D^k\} \mid [f_i^k]^{OK}\right) \cdot P\left(C^{k-1} \mid [c_i^k]^{OK}\right) \cdot P\left([c_i^k]^{OK}\right)}{P\left(D^k\right) \cdot P\left(C^{k-1}\right)} \tag{13}$$

where, $P\left(C^{k-1} \mid [c_i^k]^{OK}\right) = \prod_{j=1}^{M} P\left(c_j^{k-1} \mid [c_i^k]^{OK}\right)$ and as in (11):

$$\begin{aligned} &P\left(c_j^{k-1} \mid [c_i^k]^{OK}\right) = \\ &\quad \left[P\left(c_j^{k-1} \mid [c_j^{k-1}]^{OK}\right) \cdot P\left([c_j^{k-1}]^{OK} \mid [c_i^k]^{OK}\right)\right] + \\ &\qquad \left[P\left(c_j^{k-1} \mid [c_j^{k-1}]^{NO.OK}\right) \cdot P\left([c_j^{k-1}]^{NO.OK} \mid [c_i^k]^{OK}\right)\right] \end{aligned} \tag{14}$$

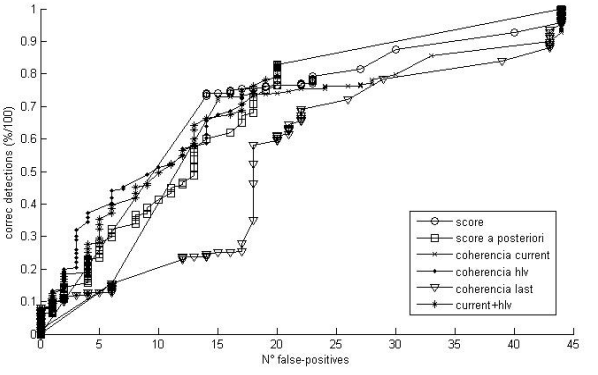
and $P\left(c_j^{k-1} \mid [c_j^{k-1}]^{OK}\right)$ is the a posteriori probability q_{k-1}^j, calculated in the past frame. $P\left([c_j^{k-1}]^{OK} \mid [c_i^k]^{OK}\right) = \Pr\left([c_i^k]^{OK} \mid [c_j^{k-1}]^{OK}\right)$ is the last frame coherence defined in (5). All other terms, can be calculated analogously to the current frame case. On the other hand, $P\left(D^k \mid [f_i^k]^{OK}\right) = \prod_{j=1}^{M} P\left(d_j^k \mid [f_i^k]^{OK}\right)$, then, applying total probabilities theorem we obtain:

$$\begin{aligned} &P\left(d_j^k \mid [f_i^k]^{OK}\right) = \\ &\quad \left[P\left(d_j^k \mid [d_j^k]^{OK}\right) \cdot P\left([d_j^k]^{OK} \mid [f_i^k]^{OK}\right)\right] + \\ &\qquad \left[P\left(d_j^k \mid [d_j^k]^{NO.OK}\right) \cdot P\left([d_j^k]^{NO.OK} \mid [f_i^k]^{OK}\right)\right] \end{aligned} \tag{15}$$

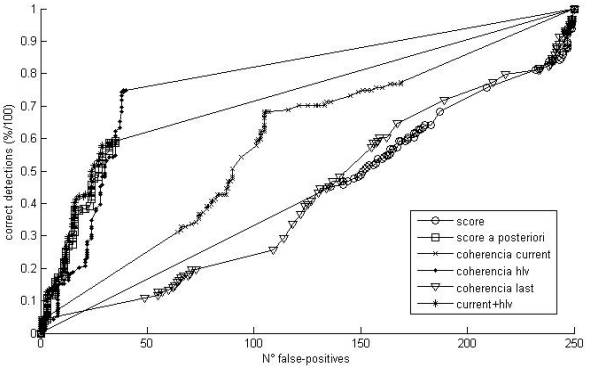
where $\Pr\left(d_j^k \mid [d_j^k]^{OK}\right)$ is the confidence $\{\lambda_k^{conf}\}_j$ defined in (7) by HLT module, and $P\left([d_j^k]^{OK} \mid [f_i^k]^{OK}\right) = P\left([f_i^k]^{OK} \mid [d_j^k]^{OK}\right)$ is the coherence with the HLT module´s estimation defined in (6). All other terms can be calculated in the same way already explained.

3 Experimental Results

Our vision system was tested using real data sequences obtained by an AIBO Robot inside a RoboCup Four Legged Soccer field. The detection rates were measured in



(a)



(b)

Fig. 2. ROC curves using different context instances. Score: it is the a priori probability given by perceptors modules. Score a posteriori: it is the a posteriori probability calculated by the proposed context integration system. Coherence instances: are the a posteriori probability given by each context instance.

two different situations: a low noise situation with few false objects, and a noisier situation, with much more false objects. In the first situation, false object presents were "natural" objects, like the cyan blinds and some other real, colored objects of our laboratory, which are naturally placed around the field. These objects appear in approximately 20% of the frames. In the second situation, additional false objects were added: one false goal and one false beacon over the ground plane, and one false goal and one false beacon in the border of the field. Both situations can be observed in real games of the RoboCup due to the non-controlled conditions of the environment. The public can wears with the same colors of the interesting objects and several other objects of different colors can be founded around the field.

In this work, ROC curves with the number of false-positives in the x-axis have been used to evaluate the system. These ROC curves permit to compare the utility of the different context instances proposed, measuring the rate of correct detection given a number of false positives that indicates the noise degree of the environment. The results are shown in Fig. 2. Note how the a priori and the a posteriori ROC curves evolve as the quantity of noise is increased. When the system is facing situations with low amount of noise (i.e. false objects), the use of context is not very important to improve the performance of the system. However, as the quantity of false objects grows, the use of context increases noticeably the detection rate for a given false positive rate.

An important observation is the fact that last frame coherence appears not to be very important compared with HLT coherence and with the current frame coherence. In fact, if we only consider the current frame coherence and HLT coherence instances, the a posteriori probability calculated is very near to the a posteriori probability calculated when the last frame coherence is included. Hence, the last frame coherence is irrelevant.

4 Conclusions

We have presented a general-propose context based vision system for a mobile robot having a mobile camera. The use of spatiotemporal context is intended to make the vision system robust to noise and high performing in the task of object detection.

We have presented a general-purpose context based vision system for a mobile robot having a mobile camera. The use of spatiotemporal context is intended to make the vision system robust to noise and high performing in the task of object detection.

We have first applied our vision system to detect static objects in the RoboCup SP League domain, and preliminary experimental results are presented. These results confirm that the use of spatiotemporal context is of great help to improve the performance obtained when facing the task of object detection in a noisy environment. The reported results encourage us to continue developing our system and to test it in other applications, where different physical objects and lighting conditions may exist.

As future work, we propose to include some other context instances, and integrate these to the bayesian context filter. In the other hand, it is possible to research about:

what is the best way to calculate the different context instances and how to extend the bayesian approach to the HLT estimation.

Although we have satisfactory results, we believe that the system may be improved considerably by facing these issues.

References

1. Torralba, A., Sinha, P.: On Statistical Context Priming for Object Detection. In: International Conference on Computer Vision (2001)
2. Torralba, A.: Modeling global scene factors in attention. JOSA - A 20, 7 (2003)
3. Cameron, D., Barnes, N.: Knowledge-based autonomous dynamic color calibration. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS, vol. 3020, pp. 226–237. Springer, Heidelberg (2004)
4. Oliva, A., Torralba, A., Guerin-Dugue, A., Herault, J.: Global semantic classification of scenes using power spectrum templates. In: Proceedings of The Challenge of Image Retrieval (CIR 1999), Newcastle, UK. BCS Electronic Workshops in Computing series. Springer, Heidelberg (1999)
5. Jüngel, M., Hoffmann, J., Lötzsch, M.: A real time auto adjusting vision system for robotic soccer. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS, vol. 3020, pp. 214–225. Springer, Heidelberg (2004)
6. Oliva, A.: Gist of the Scene. Neurobiology of Attention, pp. 251–256. Elsevier, San Diego (2003)
7. Foucher, S., Gouaillier, V., Gagnon, L.: Global semantic classification of scenes using ridgelet transform. In: Human Vision and Electronic Imaging IX. Proceedings of the SPIE, vol. 5292, pp. 402–413 (2004)
8. Torralba, A., Oliva, A.: Statistics of Natural Image Categories. In: Network: Computation in Neural Systems, (14), pp. 391–412 (August 2003)
9. Spillman, L., Werner, J. (eds.): Visual Perception: The Neurophysiological Foundations. Academic Press, London (1990)
10. Oliva, A., Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. International Journal of Computer Vision 42(3), 145–175 (2001)
11. Potter, M.C., Staub, A., Rado, J., O'Connor, D.H.: Recognition memory for briefly presented pictures: The time course of rapid forgetting. Journal of Experimental Psychology. Human Perception and Performance 28, 1163–1175 (2002)
12. Strat, T.: Employing contextual information in computer vision. In: Proceedings of DARPA Image Understanding Workshop (1993)
13. Ruiz-del-Solar, J., Verschae, R.: Skin Detection using Neighborhood Information. In: Proc. 6th Int. Conf. on Face and Gesture Recognition – FG 2004, 463 – 468, Seoul, Korea (May 2004)
14. RoboCup Technical Comitee, RoboCup Four-Legged League Rule Book (2006), `http://www.tzi.de/4legged/bin/view/Website/WebHome`
15. Stehling, R., Nascimento, M., Falcao, A.: On 'Shapes' of Colors for Content-Based Image Retrieval. In: Proceedings of the International Workshop on Multimedia Information Retrieval, pp. 171–174 (2000)
16. Zagal, J.C., Ruiz-del-Solar, J., Guerrero, P., Palma, R.: Evolving Visual Object Recognition for Legged Robots. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS, vol. 3020, pp. 181–191. Springer, Heidelberg (2004)

17. Guerrero, P., Ruiz-del-Solar, J., Palma-Amestoy, R.: Spatiotemporal Context in Robot Vision: Detection of Static Objects in the RoboCup Four Legged League. In: Proc. 1st Int. Workshop on Robot Vision, in 2nd Int. Conf. on Computer Vision Theory and Appl. – VISAPP 2007, pp. 136 – 148, March 8 – 11, Barcelona, Spain (2007)
18. Ruiz-del-Solar, J., Guerrero, P., Vallejos, P., Loncomilla, P., Palma-Amestoy, R., Astudillo, P., Dodds, R., Testart, J., Monasterio, D., Marinkovic, A.: UChile1 Strikes Back. In: Team Description Paper, 3rd IEEE Latin American Robotics Symposium – LARS 2006, October 26-27, Santiago, Chile (CD Proceedings) (2006)
19. Torralba, A., Murphy, K., Freeman, W., Rubin, M.: Context-based vision system for place and object recognition. In: Proc. Intl. Conf. on Computer Vision - ICCV 2003, October 13-18, Nice, France (2003)