Spatiotemporal Context in Robot Vision: Detection of Static Objects in the RoboCup Four Legged League

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

Department of Electrical Engineering, Universidad de Chile,

{pguerrer, jruizd, ropalma}@ing.uchile.cl

VISAPP 2007

Agenda

- Motivation
- Proposed System
- Application: RoboCup 4Legged League
- Results
- Conclusions

Motivation

- Object visual perception in complex and dynamical scenes with cluttered backgrounds is a very difficult task.
- Humans solve it satisfactorily, but computer and robot vision systems not!.
- One of the reasons of this large difference in performance is the use of context by humans (our main hypothesis).

Motivation
Proposed System
Application: RoboCup 4Legged League
Results
Conclusions

How is context useful?

- Reducing the perceptual aliasing:
  - 3D objects are projected onto 2D sensors.
- Increasing the perceptual abilities in hard conditions:
  - Context can facilitate the perception when the local intrinsic information about the object structure is not sufficient
- Speeding up the perceptions:
  - Contextual information can speed up the object discrimination by cutting down the number of object categories, scales and poses that need to be considered.

Types of Context

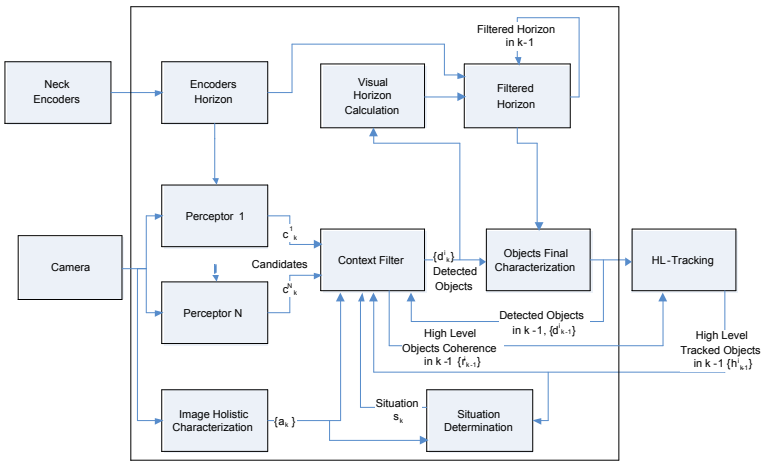
From the visual perception point of view, it is possible to define at least six different types of context:

- Low-Level Context
- Physical Spatial Context
- Temporal Context
- Object´s Configuration Context
- Scene Context
- Situation Context

Proposed System

- Use of several kinds of context.
- The main stages of the system are:
  - Object perceptors
  - A holistic characterization of the scenes
  - context coherence filtering between current and past detections
  - encoder-based, visual-based and filtered horizon information; and
  - high-level tracking of objects' poses.

Proposed System



Perceptors

- Perceptors are specific stages for detecting specific objects in an image.
- They make use of only local information.
- Every detection at this level, c_k^i , is called an object candidate.
- An a priori probability α_k^i is calculated as a measure of confidence on the detection.

Visual Horizon

- Line in the image corresponding to objects having the same altitude than the camera.
- Estimate from encoders: very noisy (depending on the robot's complexity).
- We add an estimation from objects detections candidates.

High Level Tracking

- HL-Tracking stage maintains information about the objects detected in the past.
- This tracking stage is basically a state estimator for each object of interest:
  - For fixed objects, the relative pose of the object with respect to the robot.
  - For mobile objects, the relative velocity may be added to the state vector.
- This module can be implemented using standard state estimation algorithms as Kalman or Particle Filters.

Image Holistic Characterization

- A single glance to a complex, real world scene is enough for an observer to comprehend a variety of perceptual and semantic information.
- There are several works that use different alternatives of representations of the global information contained in an image (e.g. spatial frequency orientations and scales, color, texture density).

Context Filter

- Context information is employed for filtering candidate objects.
- Each candidate must coherent with:
  - The current situation.
  - The holistic image characterization.
  - Every other candidate object.
  - Detections in the last image.
  - Past HL-Tracked poses of objects (including itself).

Context Vector

- For each object candidate, c_k^i, it is defined its context vector:

candidates
HH objects
holistic characterization
situation
past detections

$$\mathbf{C}_k^i = \left(\mathbf{K}_k^{i\,T}, \mathbf{D}_{k-1}^{\;T}, \mathbf{H}_{k-1}^{\;T}, a_k, s_k \right)^T$$

- To measure the importance of each context element, a context weight vector is defined:

a priori prob.
a posteriori prob.
accumulated prob.

$$\Omega_k^i = \left(\Omega_k^{\mathbf{C},iT}, \Omega_{k-1}^{\mathbf{D}\;T}, \Omega_{k-1}^{\mathbf{H}\;T}, p(a_k), p(s_k) \right)^T$$

$$= \left(\omega_{k,1}^i, \ldots, \omega_{k,L}^i \right)^T$$

Candidate Coherence

- For each candidate c_k^i, it is also defined its coherence:

$$q_k^i = p\left(c_k^i \mid \mathbf{C}_k^i\right) = \frac{\sum_{j=1}^{L} p\left(c_k^i \mid \left[\mathbf{C}_k^i\right]_j\right) \omega_{k,j}^i}{\sum_{j=1}^{L} \omega_{k,j}^i}$$

- Its a posteriori probability, is defined as:

$$p_k^i = \alpha_k^i q_k^i$$

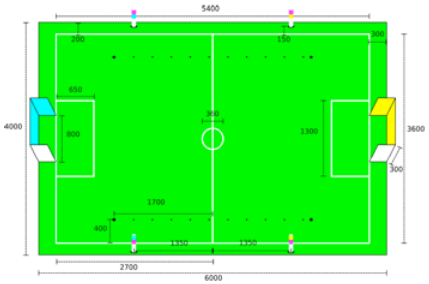
a priori prob.

Relationships between Objects

- (In our RoboCup appl.) there are four kind of relationships that can be checked between physical objects.
- Between objects in the same image:
  - Horizontal Position Alignment
  - Horizon Orientation Alignment
- Between objects in different images:
  - Relative Position or Distance Limits
  - Speed and Acceleration Limits

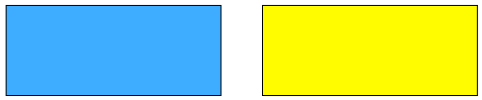
RoboCup 4L League

- 4 AIBO Robots per team, no external processing allowed.
- AIBO: 15 DOF (3 per leg), 1 color camera, 1 embedded RISC processor.

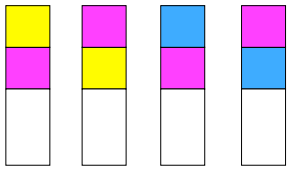


Objects

- Objects are colored to allow their easy detection.
  - Two colored goals:



  - Four colored beacons:



Relationships Considered

- With other object candidates:

$$p\left(c_k^i \mid c_k^j\right) = p_{Hor}\left(c_k^i \mid c_k^j\right) p_{Lat}\left(c_k^i \mid c_k^j\right) p_{Dist}\left(c_k^i \mid c_k^j\right)$$

horizontal coher. lateral coher. dist. coher.

- With detections in the last image:

$$p\left(c_k^i \mid d_{k-1}^j\right) = p_{Hor}\left(c_k^i \mid d_{k-1}^j\right) p_{Lat}\left(c_k^i \mid d_{k-1}^j\right) p_{Dist}\left(c_k^i \mid d_{k-1}^j\right)$$

- With HL-Tracked poses:

$$p\left(c_k^i \mid h_{k-1}^j\right) = p_{Lat}\left(c_k^i \mid h_{k-1}^j\right) p_{Dist}\left(c_k^i \mid h_{k-1}^j\right)$$

- With the holistic characterization of the image:

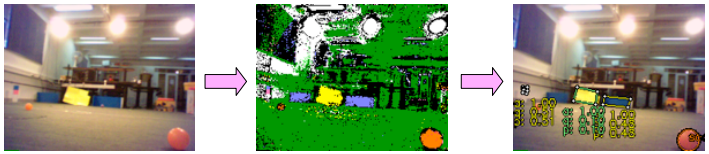
$$p\left(c_k^i \mid a_k\right) = p\left(\mathbf{y}_k^i, \eta_k^i \mid a_k\right)$$

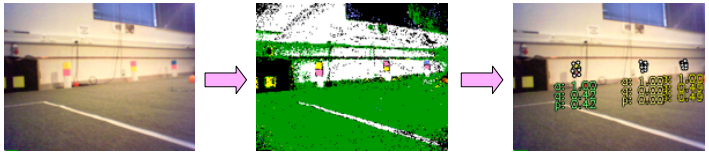
Results

- We have tested our vision system using real video sequences obtained by an AIBO Robot inside a soccer field.
- The detection rates were measured in different situations having different quantities of false objects.

Examples (1)

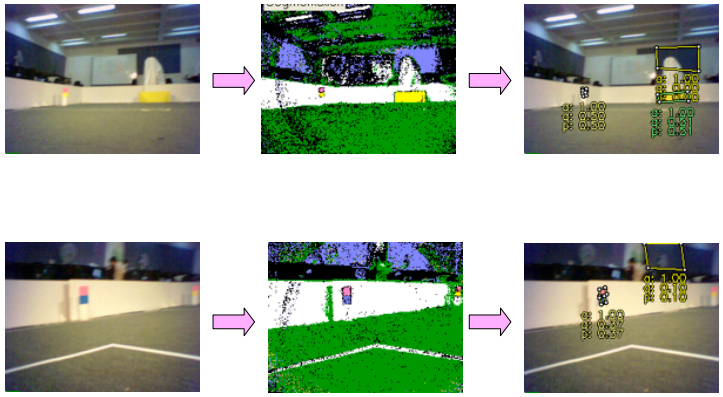
- Some examples of the use of context:





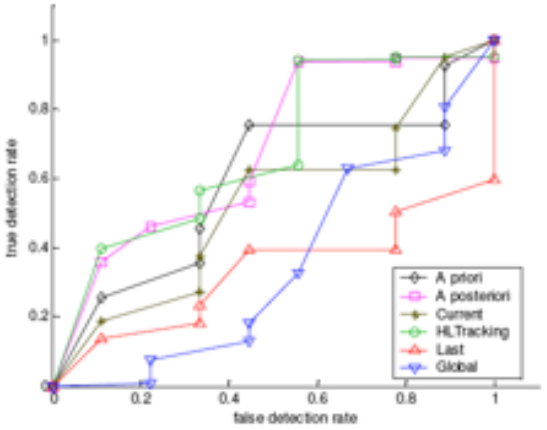
Examples (2)

- Some more examples of the use of context:



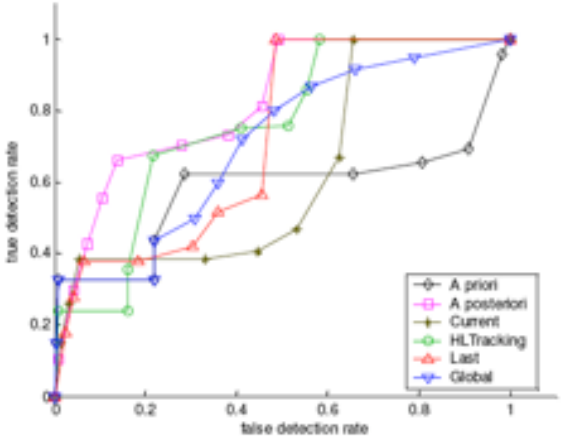
ROC Curves

- Low Noise Situation: false objects are “natural” objects, like the cyan blinds and some other objects of our laboratory .



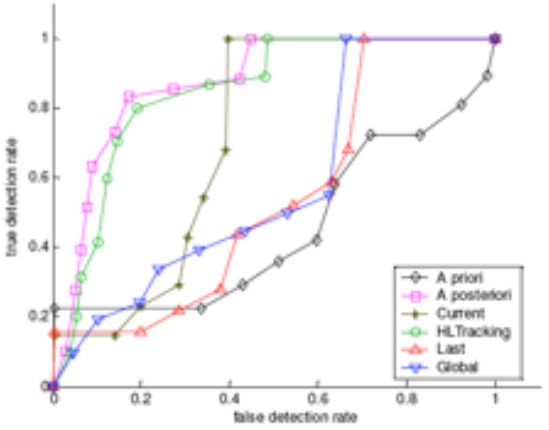
ROC Curves

- Medium Noise Situation: two new false objects.
- Improvement ~ 25%



ROC Curves

- High Noise Situation: two additional new objects
- Improvement of ~40%



Conclusions

- General context-based vision system for a mobile robot having a mobile camera.
- Experimental results confirm that the use of spatiotemporal context improves the performance of object detection in a noisy environment.
- We are currently working in the inclusion of mobile objects to our system.