AttentiRobot: Attentive Vision for Robot Self-Localization

N. Ouerhani•, H. Hugli•, G. Gruener, A. Codourey

This work aims at demonstrating the usefulness of visual features detected by visual attention for robot localization. The most salient visual features are used as landmarks for mapping the navigation environment. During navigation the landmarks are recognized and the robot position is derived.

Vision-based robot localization can be achieved using visual landmarks. Representative visual features of the navigation environment are selected and used to create a topological map. During navigation, the robot recognizes these landmarks and determines its current location. This demonstrates the usefulness of the visual attention mechanism in solving problems related to the localization task. Figure 1 shows CSEM's Pelé robot and its environment.

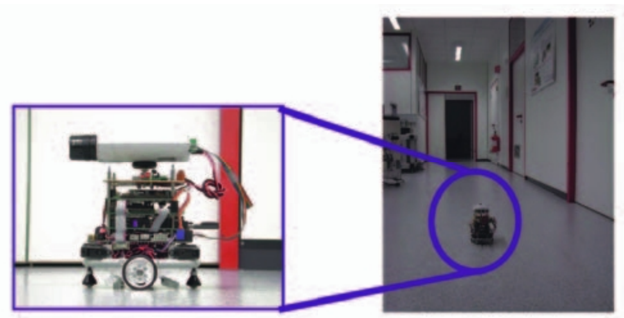

Figure 1: CSEM's Pelé robot with a color web camera mounted on top and its navigation environment.

Reliable visual landmarks must satisfy two crucial conditions for robot navigation: uniqueness and robustness. They must be unique to be easily distinguished from each other. They must be robust to be recognized under different conditions. A landmark selection method was developed. It guarantees the 2 conditions[1]. A saliency-based model of visual attention is used to detect the most salient, and thus unique, features of the environment. The features that satisfy the uniqueness condition are then tracked over time. A selection procedure retains only those features that have been robustly tracked, yielding visual landmarks that satisfy the robustness criteria[2].

The selected landmarks are used to build a topological map of the environment. Mapping consists of dividing the navigation environment into representative portions called key frames. Each key frame is represented by its associated landmarks. Each landmark is characterized by a set of visual descriptors that facilitates their recognition during the navigation phase. Fig. 2 illustrates the mapping method by dividing the environment into three key frames.

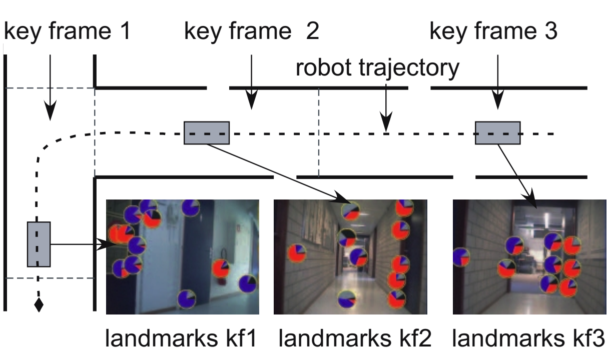

Figure 2: Mapping. The environment is divided into representative key frames, each of which is represented by a set of visual landmarks.

During navigation, the visual attention model is used to detect the most salient features at the current position of the robot. The detected features are compared with learned landmarks using a metric that combines the similarity between visual descriptors and the spatial relationships between the features and the landmarks. More specifically, a set of three features is matched to a set of three landmarks and their spatial configuration is used as an additional constraint.

For each key frame, the matching procedure computes a score that quantifies the likelihood that a key frame is the current position of the robot. This probabilistic estimation of the robot position is necessary to be able to integrate this method into more complete localization frameworks, like stochastic models. Figure 3 shows an example of how this works.

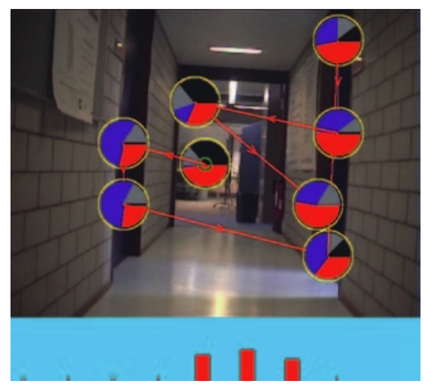

Figure 3: An example of the recognition process. A set of features are detected from a test sequence and matched to landmarks learned from a training sequence. The red vertical bars at the bottom represent the probability that the current robot position is at each of eight key frames. The key frame with the highest probability matches correctly the location of the robot.

In future work the method will be integrated into a more complete stochastic navigation framework. The algorithms will be extended to panoramic images in order to have richer visual input.

This work is co-funded by the cantons of central Switzerland, the MCCS Micro Center Central Switzerland and OFES. Their support is gratefully acknowledged.

• IMT, University of Neuchâtel

[1] N. Ouerhani, H. Hugli, G. Gruener, A. Codourey, "Attention-Based Landmark Detection and Tracking for Visual Robot Navigation", Workshop on Attention Performance in Computer Vision-WAPCV04 (2004) 83

[2] N. Ouerhani, H. Hugli, G. Gruener, A. Codourey, "A Visual Attention-Based Approach for Automatic Landmark Selection and Recognition", Lecture Notes in Computer Science, LNCS 3368, Springer Verlag (2005) 183