# Robust Self-Localization using Elastic Templates

Christian Balkenius
Lars Kopp

Lund University Cognitive Science
Kungshuset, Lundagård
S-222 22 Lund

christian.balkenius@fil.lu.se
lars.kopp@fil.lu.se

## 1 Introduction

A visually orienting mobile robot must cope with a number of changes to its environment. Most importantly, it must be able to identify its location even when objects in the environment have been moved or when the illumination conditions have changed. We report experiments with a template-based self-localization method that operates in real time on a Pentium 133 MHz PC (Balkenius and Kopp 1996a, b; 1997). The algorithm has been implemented in an autonomous mobile robot as an important part of its navigational system (Balkenius and Kopp 1996a, b; 1997).

The main task for the algorithm is to recognize landmarks robustly around the robot. These landmarks are represented as elastic templates and are automatically selected by the robot during learning. Our methods differ from other vision-based localization techniques in a number of respects. First, it does not require artificial landmark-symbols (Adorni, et al. 1996). Second, it can derive the exact angle toward a landmark which allows precise positioning using triangulation. This differs from similar methods that only produces an approximate location (von Wichert 1996). Other precise angle estimators are typically very slow (Suburo and Shigang 1993). Third, it is very insensitive to changes in illumination. This can be a large problem for methods that tries to recognize light intensities or colors directly (Cassinis, et al. 1996). Fourth, our method does not require an environment that generates clean motion vectors (Arkin 1987) or a floor with homogenous texture (Horswill 1993).

In a typical indoor environment, our method can generalize a learned landmark to locations at approximately a half a meter radius around the learned location. We call this region a view-field (Balkenius and Kopp 1997; Zipser 1985). A critical ability of the algorithm is to make view-fields robust against changes in the environment.

Below, we investigate how well the algorithm generalizes between different locations and angles toward the landmark when the robot is moved around. We also report experiments with different illumination and partial occlusion of the landmarks. We show that elastic template matching is very robust against changing environments without any preparation such as artificial landmark-symbols that have to be identified.

## 2 The Template-Matching Algorithm

The experiments described below were based on the landmark recognition system in the XT-1 architecture (Balkenius and Kopp 1996a, b, 1997). In this section we briefly describe the main features of the template matching system. A more complete description of the algorithm can be found in (Balkenius and Kopp 1997).

The main idea behind the template representation is that the image is divided into a number of regions which contain features that are as distinct as possible. These features are selected automatically when the template for a landmark is learned (Balkenius and Kopp 1997). To hold the features together, the template also records the spatial relations between the features. In the simplest case, these spatial relations are stored as the x and y coordinates in the image of each feature (figure 2.1).

Each landmark is represented in this way as a set of features together with their spatial relations. When the visual system tries to identify a landmark in the image, it attempts to locate each of the features at the approximately correct relative positions. In the figure, the spatial relations are represented as springs to signify that a distortion of the template is allowed. If sufficiently many of the features are found at the approximately correct location in the image, they are used to estimate the center of the target in the image.

Elastic template matching allows landmarks to be identified even when parts of the landmark is occluded or removed. For example, objects may change their spatial position and objects may move around and occlude parts of the landmark. The matching process also compensates for expansion of the landmark-view and incorrect orientation toward the landmark.

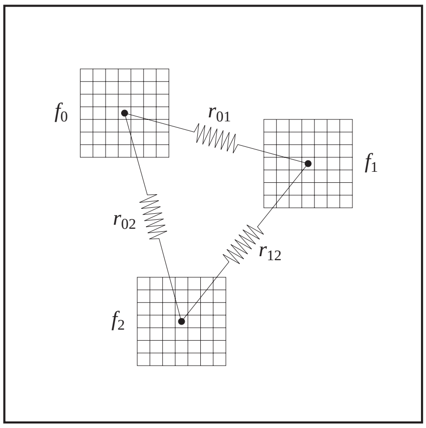


**Figure 2.1** An elastic template is coded by a number of features $f_0$, $f_1$, $f_2$ together with their spatial relations $r_{01}$, $r_{02}$, $r_{12}$.

To optimize the matching process we view the spatial relations among the features as expectations of where one feature should be located given the locations of the already found features (Balkenius and Kopp 1996b). The expectations can be generated during search, either from the last feature found, or from all features which have already been localized. In either case, the search process keeps an estimate of the location of the reference position for the template, and this estimate is used to set the center of the search field.

Since a feature is only considered found if it lies within the expected search-field, it is not necessary to search the regions of the image that falls outside the search-field. Since the search field is considerably smaller than the input image, this speeds up the template matching considerably.

However, the search-field can only be used when some other feature has already been found. Since this is not the case for the first feature, it is necessary to do a total search for this feature. If the first feature is found, its location will subsequently constrain the search for the next feature. When a feature is not found, a total search is performed for the next feature since no expectation will be present.

In out current robot, POLUCS (Balkenius and Kopp 1996a, b), the camera is not allowed to move in the vertical direction. This means that features within a template do not move very much in the vertical directions. As a consequence, a total search is limited to a small horizontal band across the image where the feature can potentially be located.

To determine whether a template matches the image, we check whether a sufficiently large fraction of the features are found with a sufficiently large correlation within the expected location in the image. The correlation of each feature with its corresponding search–field is calculated as the maximum of the Moravec distance between the feature and the edge-filtered input image within the search–field (Lawton, et al. 1988; Balkenius and Kopp 1997). To determine whether the feature matches or not, it is compared to a correlation threshold. If the correlation is higher than this threshold, the feature will contribute to the template match. Otherwise, it will be assumed missing or occluded.

A template matches the input image if the number of found features are larger than a certain hit threshold. For example, if the hit threshold is 50%, then more than half of the features must be found for the template to match. The lower the hit threshold, the more occlusion is possible before the system fails to recognize the landmark.

## 3 Changes in Position

In this section we describe how well the algorithm generalizes when the robot is moved away from the location where a landmark-template was learned. Figure 3.1 shows nine sample views of our lab. The central image shows the input image where the landmark template with 32 features was learned. The other images show locations around the learned position. A total of 120 positions where tested with different correlation and hit thresholds to see how these parameters would influence the matching process.

In figure 3.2, the average correlation among all features are shown as a function of the location of the robot. In all cases, the robot looks in the same direction relative to the coordinate system in which it is moved. Ideally, this would give rise only to translation and expansion of the input image. However, since the camera on the robot uses a 110° lens, the transformation of the input image is more complex.

The graph shows that the average correlation stays above 0.3 within an area of approximately 1m$^2$ around the learned position. While this indicated a large degree of generalization, it is clearly not sufficient for discrimination between different views. It can be used, however, to determine when it is time to learn a new template.

To increase discrimination, we raised the hit threshold to 20% and the correlation threshold to 0.7. In this case, the result is very different (figure 3.3). The two interacting thresholds serve to elevate the response of the template within the view-field. This mechanisms has two important properties. First, it makes the localization of the landmark more precise since features with a too low correlation are discarded. Second, it enhances discrimination between different view-fields, and hence locations. Typically, two or less features were incorrectly identified in the distractors used.

These experiments show that localization of a template is possible when the robot is moved while looking in the same direction. However, it is also necessary that it can generalize between different angles toward the landmark. This property was investigated in a second experiment where the robot was moved from left to right and was allowed to turn 15° left and right to try to localize the landmark.

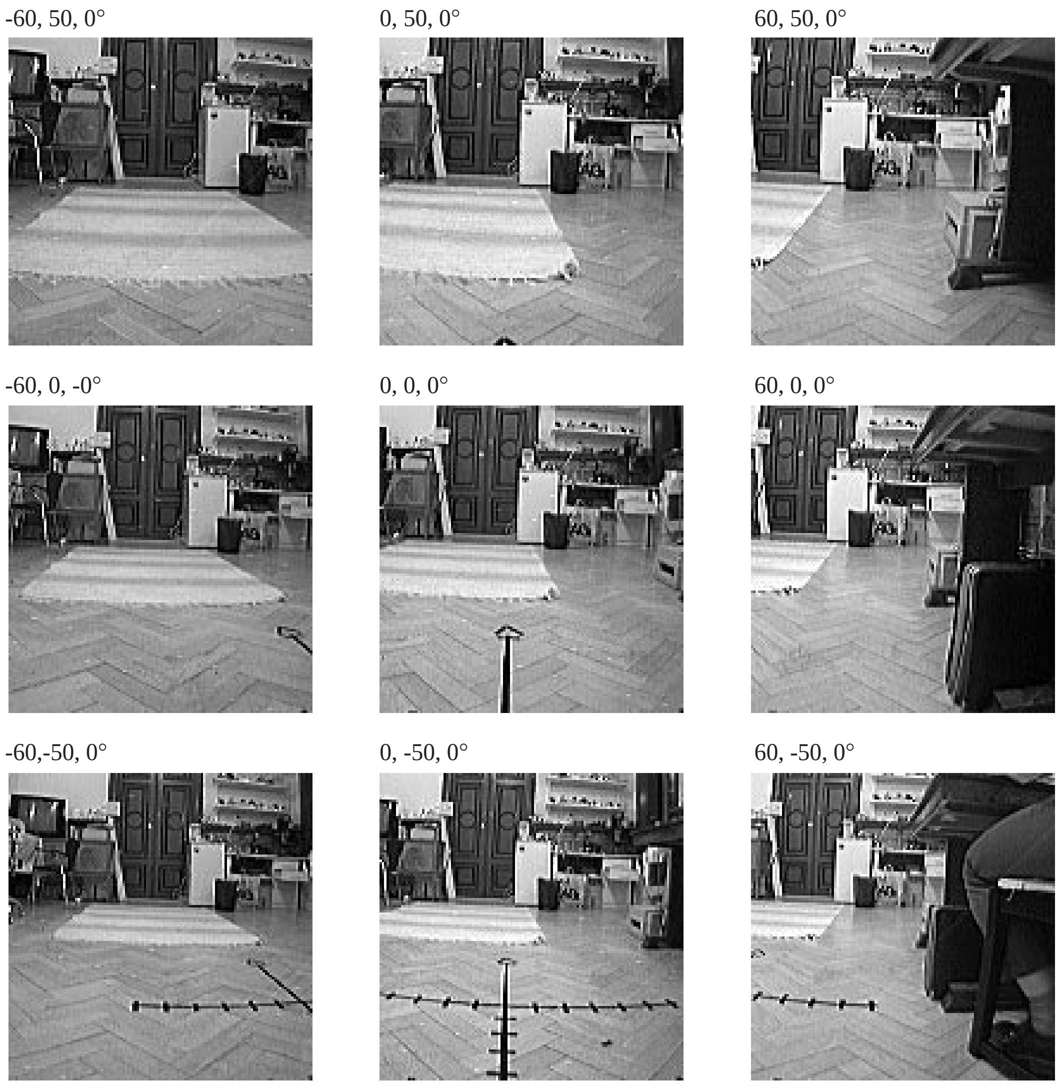


**Figure 3.1.** Translation sensitivity within a view-field. The figure consists of nine views within a single view-field with a size of approximately 1m$^2$. The coordinates indicate distance in cm from the central position. The central image shows the position where the template was learned.

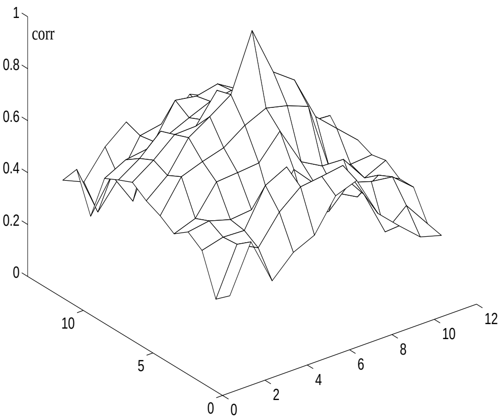


**Figure 3.2** Sensitivity to translation when all feature are allowed to contribute to the average correlation. The match level decreases smoothly as the robot is moved away from the location where the template was learned. (The axes indicates the coordinates of thedifferent samples.)

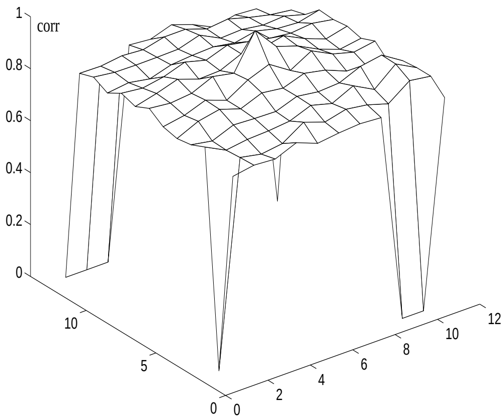


**Figure 3.3** Sensitivity to translation when a correlation threshold of 0.7 was used together with a hit threshold of 20%. A strong match is supported even when the robot is moved 50 cm in either direction away from the learned template position. (The axes indicates the different samples.)

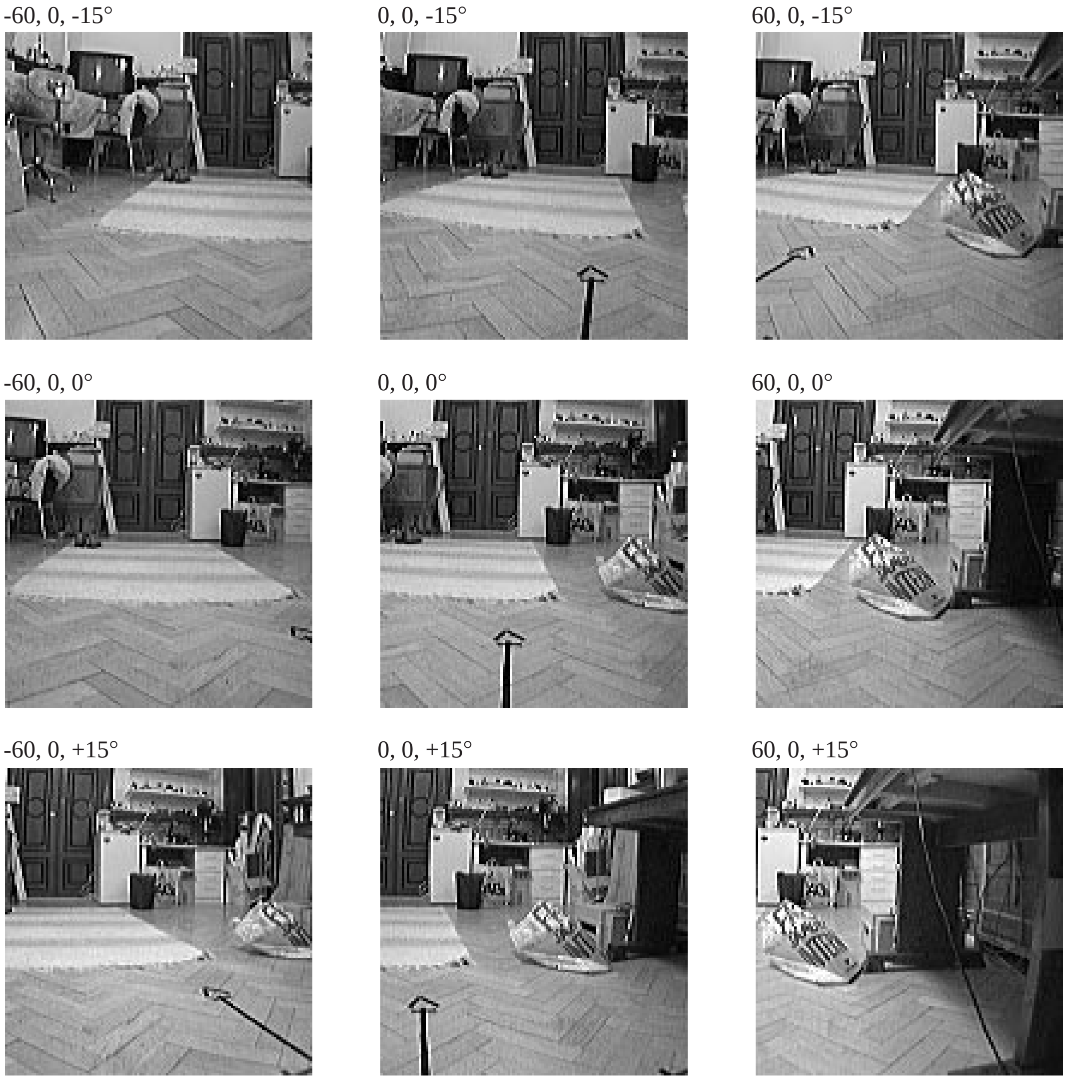


**Figure 3.5**. Rotations sensitivity within a view-field. The coordinates indicate the distance from the central position and the angle away from the forward direction of the view-field.

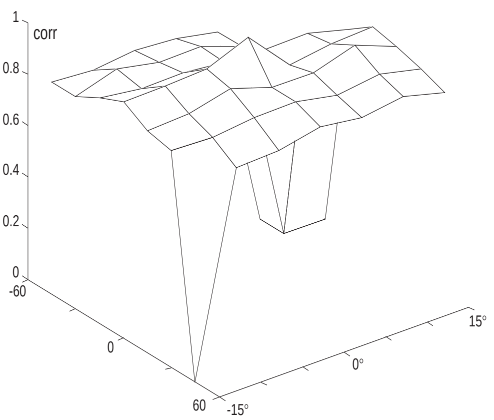


**Figure 3.5.** Sensitivity to translation combined with rotation of the robot. The template matches the input image as long as the angle is less than 15° *or* the robot is less than 60 cm from the learned location. Parameters were set as in figure 3.3.

Figure 3.4 shows nine views from this experiment. As above, the central image shows the learned location and the surrounding images show the extreme positions. The result is shown in figure 3.5. The landmark is successfully recognized as long as the angle to the landmark is less than 15° *or* the distance to the learned location is less than 60 cm. Again, the matching process is shown to be quite insensitive to fairly large transformations of the input image.

## 4 Environmental Change

Apart from changes in the position of the robot, the environment may also change on its own. People may come and go, and move object around. A robust localization method must be relatively insensitive to such changes. Figure 4.1 shows how a template can match an input image where several changes have been made. The top image shows the original image and the small squares shows the 32 features used by the template. In the middle, the door has been opened and a chair has been placed in the scene. In this case, 23 of the 32 feature were found, and the average correlation of the features was 0.87.

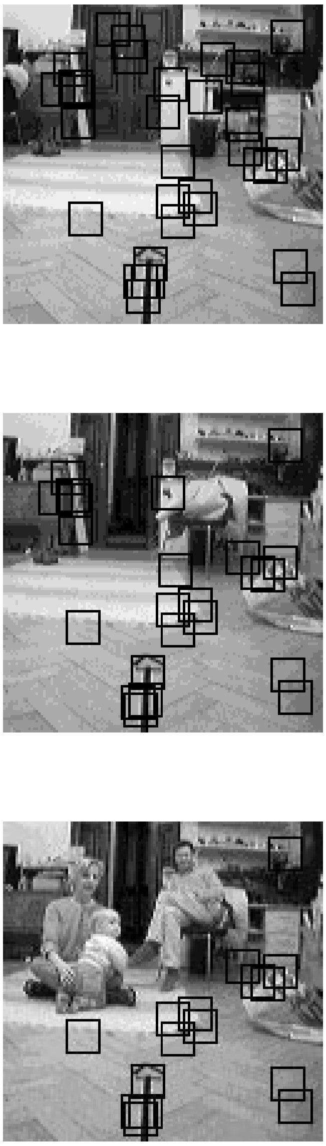

**Figure 4.1.** Sensitivity to changes in the environment with a hit threshold of 0.7. TOP. The original template with 32 features. MIDDLE. Matching when the door has been opened and a chair inserted. The hit rate is 16/32 with the average correlation of 0.85. BOTTOM. Matching with people present in the image. The hit rate is 23/32 with average correlation of 0.73.

In the lower image, the environment has changed even more. Three people have entered the scene and occludes a large portion of the template. However, 16 of the 32 feature are still found with an average correlation of 0.85. Given that the threshold for the hit rate is low enough, and that there are sufficiently many features in the template, changes such as these do not obstruct the localization process.

## 5 Changes in Illumination

Another important aspect of any visual self-localization techniques is that it must be insensitive to changes in illumination. To test the robustness of the template matching method we tested the matching algorithm under different lighting conditions: (1) only overhead light, (2) only light from the left wall, (3) only light from the right wall, and (4) light from the walls but no overhead light. The original template had been learned with all the light sources active. The different lighting conditions tested do not cause dramatically changes in the input image, but they do span all possible illuminations of our lab and, subjectively, they feel quite different.

Under the different illuminations, the template matching acquired a hit rate between 17/32 and 30/32 which is more than enough for localization purposes. The average correlation ranged from 0.80 to 0.84 for the different lighting conditions which implies that in our test environment, the different possible changes in illumination could not disrupt the performance of the template matching process.

## Acknowledgements

The support from the Swedish National Board for Industrial and Technical Development (NUTEK) is gratefully acknowledged.

## References

Adorni, G., Destri, G., Mardonini, M., Zanichelli, F. (1996). Robot self-localization by means of vision. In Proceedings of the first euromicro workshop on advanced mobile robots. Los Alamitos, CA: IEEE Computer Society Press. 160-165.

Arkin, R. C. (1987). Motor schema based navigation for a mobile robot: an approach to programming by behavior. In Proceedings of the 1987 IEEE international conference on robotics and automation. 264-271.

Balkenius, C., Kopp, L. (1997). Elastic template matching as a basis for visual landmark recognition and spatial navigation. *submitted*.

Balkenius, C., Kopp, L. (1996a). Visual tracking and target selection for mobile robots. In Proceedings of the first euromicro workshop on advanced mobile robots. Los Alamitos, CA: IEEE Computer Society Press.

Balkenius, C., Kopp, L. (1996b). The XT-1 Vision Architecture. In Linde, P., Sparr, G. (eds) Symposium on image analysis. Lund 1996.

Cassinis, R., Grana, D., Rizzi, A. (1996). Using colour information in an omnidirectional perception system for autonomous robot localization. In Proceedings of the first euromicro workshop on advanced mobile robots. Los Alamitos, CA: IEEE Computer Society Press. 172-176.

Horswill, I. (1993). A simple, cheap, and robust visual navigation system. In Mayer, J.-A., Roitblat, H. L., Wilson, S. W. (eds) From animals to animats 2. Cambridge, MA: MIT Press.

Lawton, D. T., Rieger, J., Steenstrup, M. (1988). Computational techniques in motion processing. In Arbib, M. A., Hanson, A. R. (eds) Vision, brain and cooperative computation. Cambridge, MA: MIT Press. 419–488.

Suburo, T., Shigang, L. (1993). Memorizing and representing route scenes. In Mayer, J.-A., Roitblat, H. L., Wilson, S. W. (eds) From animals to animats 2 Cambridge, MA: MIT Press.

von Wichert, G. (1996). Selforganizing visual perception for mobile robot navigation. In ed. Proceedings of the first euromicro workshop on advanced mobile robots. Los Alamitos, CA: IEEE Computer Society Press. 194-200.

Zipser, D. (1985). A computational model of hippocampal place fields. Behavioral Neuroscience, 99, 1006-1018.