Visual attention and active vision
Kootstra, Geert Willem

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Paying Attention to Symmetrical Regions of Interest
Abstract

In the previous chapter, the MUlti-scale Symmetry Transform (MUST) is proposed for the detection of interest points based on symmetry. The performance of the method is compared to the SIFT interest-point detector. The results showed that symmetrical points are more stable over small changes of viewpoint and more robust to noise and changing light conditions. Moreover, the proposed method resulted in better SLAM performance when used to select visual landmarks to represent the environment. However, Figure 8.4 reveals that even for MUST the stability of the interest points drops quickly as a function of the displacement of the robot. Regions of interest are expected to be more stably detected than points, because they are supported by larger areas and thus less susceptible to noise. Therefore, a Symmetrical Region-of-Interest Detector (SymRoID) is proposed in this chapter to improve the stability and with that the SLAM performance. The results show that symmetrical regions-of-interest are less susceptible to noise, are more stable, and above all, result in better SLAM performance.

This chapter is based on:

9.1 Introduction

One of the challenges in visual SLAM is to find high-quality visual landmarks to represent the environment (Frintrop & Jensfelt, 2008). A good and reliable landmark is one that is detectable despite noise and changing light conditions, that is stable over a sequence of observations, and that is detectable when the robot revisits the location. In this chapter, a Symmetrical Region-of-Interest Detector (SymRoID) is proposed to improve the detection of landmarks. This research extends the work presented in the previous chapter. Like in the previous chapter, the method exploits the inherent redundancy of symmetrical patterns and the presence of many symmetrical patterns in man-made environments. However, instead of detecting interest points, SymRoID detects symmetrical regions-of-interest.

Many current approaches to visual SLAM detect interest points based on contrast features, for instance using the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004; Se et al., 2002), Speeded-Up Robust Features (SURF) (Bay et al., 2006; Murillo et al., 2007) or Harris corners (Davison & Murray, 2002). SIFT has been proven to be one of the best performing interest-point detectors for SLAM (Mozos et al., 2008), as well as for object recognition (Moreels & Perona, 2007). In Chapter 8, it was shown that using local symmetry instead of contrast results in more robust interest points. Moreover, it was shown that the performance of the SLAM system using local symmetry for landmarks selection outperformed the system using SIFT interest points.

However, the problem with interest points is that a large number of points are found in the image, but many of them are unstable. This can be appreciated in Figure 8.4. The high number of points results in a high computational load. Moreover, the poor stability of the points reduces the quality of the map of the environment. Although the use of symmetry reduces the number of interest points and improves the stability compared to the use of SIFT, the large number of unstable points remains a problem. The hypothesis is that using regions-of-interest will result in fewer and more stable landmarks, since larger areas contribute to the regions-of-interest, providing more evidence and making the method less susceptible to noise. Similar findings were done by Frintrop & Jensfelt (2008) and Mikolajczyk et al. (2005). The disadvantage of using regions instead of points, however, could be that they are more susceptible to changes in viewpoint, which deteriorates the stability of the landmarks. In this chapter, the robustness and stability of the symmetrical regions-of-interest are investigated and compared to those of the
Chapter 9. Paying Attention to Symmetrical Regions of Interest

Figure 9.1: The Symmetrical Region-of-Interest Detector. The symmetry operator (a) is at the basis of the detector. All pixel pairs that lie in the symmetry kernel (the gray marked area) contribute to the symmetry value at position $p$ (a1). The symmetry contribution of a pixel pair is determined by comparing the gradients of both pixels (a2). In the multi-scale symmetry model (b), an image pyramid is constructed. The symmetry operator is applied to all images in the pyramid. The resulting symmetry maps are rescaled, and summed up to result in the multi-scale symmetry map. The complete SymRoID model (c) first calculates the symmetry map for the input image. In this map, the local maxima are found. These local maxima serve as the seeds for the region-growing algorithm that clusters all symmetry pixels. Next, the highest contributing radius of each symmetry pixel in a cluster is found and marked with a circle in the regions-of-symmetry map. The regions-of-interest are finally determined by the bounding box of the marked regions. Subsequently, the regions are described using histograms-of-gradients (HoGs).

MUST and SIFT interest points. Moreover, the SLAM performance is tested using the same annotated SLAM database as used in the previous chapter.

The results show that the detection of regions-of-interest based on symmetry results in landmarks that are more robust and stable than interest points detected using contrast features. Moreover, we test the performance of our landmark selection method on a SLAM database that we annotated with ground-truth positions. The results show that our model results in better SLAM performance.
9.2 Symmetrical Region-of-Interest Detector

The proposed Symmetrical Region-of-Interest Detector, SymRoID, is not based on the symmetrical interest-point detector presented in Chapter 8, but is an extension of the symmetry-saliency model that we proposed in Chapter 3 to predict human eye fixations. In Section 9.2.1, the symmetry operator of Reisfeld et al. (1995) is described, which forms the basis of the symmetry-saliency model. The developed multi-scale symmetry model is described in Section 9.2.2. Next, Section 9.2.3 describes the SymRoID model, which uses the saliency maps that result from the symmetry-saliency model to find symmetrical regions-of-interest in the image. Finally, the descriptor to create a scale-invariant representation of the regions is discussed in Section 9.2.4. The description of the symmetry operator and the multi-scale symmetry model is largely a recapitulation of the description of the symmetry-saliency model in Chapter 3. However, there are a few differences.

9.2.1 The Symmetry Operator

For every position, $p = (x, y)$, in the image, a symmetry kernel is applied that calculates the amount of symmetry by comparing the intensity gradients of the surrounding pixels. Pixel pairs in the neighborhood contribute to the symmetry value. A pixel pair consists of pixel $p_i$ and $p_j$, so that $p = (p_i + p_j)/2$ (see Figure 9.1a). The contribution of the pair is calculated by comparing the intensity gradient $\overrightarrow{g_i}$ at $p_i$ and gradient $\overrightarrow{g_j}$ at $p_j$ according to:

$$s(i, j) = d(i, j, \sigma) \cdot c(i, j) \cdot \log(1 + m_i) \cdot \log(1 + m_j)$$

(9.1)

where $m_i$ is the magnitude of the gradient, and $d(i, j, \sigma)$ is a Gaussian weighting function on the distance between $p_i$ and $p_j$ with a standard deviation of $\sigma$. The multiplication with the gradient magnitudes assures that only strong edges contribute. The logarithm attenuates the influence of large magnitude values, which might be a result from noise. The symmetry measurement is:

$$c(i, j) = (1 - \cos(\gamma_i + \gamma_j)) \cdot (1 - \cos(\gamma_i - \gamma_j))$$

(9.2)
where $\gamma = \theta_i - \alpha$ is the angle between the orientation of gradient, $\theta_i$, and the angle, $\alpha$, of the line between $p_i$ and $p_j$ (see Figure 9.1a). The first term in (9.2) has a maximum value when $\gamma_i + \gamma_j = \pi$, which is true for gradients that are mirror symmetric with respect to $p$. Using only this term would result in high value for points that lie on a straight edge. Since we are not interested in edge detection, but in finding the centers of symmetrical patterns, the second term demotes pixel pairs with similar gradient orientations.

The symmetry value at position $p$ is calculated by summing up the contributions of all pixel pairs in the neighborhood. Differently from the model in Chapter 3, this neighborhood is defined by an inner and an outer square centered around $p$. The size of the sides of the squares are respectively $r_1$ and $r_2$ (see Figure 9.1a). All pixels that lie inside the outer square, but outside the inner square are considered. $\Gamma(p)$ is the set of contributing pairs. In our experiments we used $r_1 = 5$ and $r_2 = 17$. Smaller values of $r_1$ result in too small symmetry patterns, and larger values of $r_2$ are too computationally expensive, and make the operator view the image with too much detail. The total symmetry value at $p = (x, y)$ is then:

$$S_l(x, y) = \sum_{(i,j) \in \Gamma(p)} s(i, j) \quad (9.3)$$

where $S_l$ is the symmetry map at scale $l$. The different scales are discussed in the next section.

### 9.2.2 The Multi-Scale Symmetry Model

A region-of-interest detector for visual SLAM needs to be able to detect structures of various sizes since the appearance of landmarks changes drastically when the robot moves around in the environment. Although the symmetry operator can detect symmetry within the neighborhood radius, it cannot detect patterns on larger scales. Increasing the radius is not a good idea due to the quadratic complexity of the operator. Moreover, at larger radii, the operator takes into account too much detail, making the operator more susceptible to noise. Therefore, we propose a multi-scale symmetry model, similar to that used in Chapter 3.

In Figure 9.1b, the multi-scale symmetry model is depicted. The scale space consists of an image pyramid that is built by progressively applying a Gaussian filter to the
image, followed by a downscaling of the image by a factor of two, where scale zero is the image in its original resolution. Secondly, the symmetry operator is applied to all images in the pyramid, resulting in a pyramid of symmetry maps. Finally, the symmetry maps at the different scales are resized to the size of the first scale, and then summed up to result in the overall symmetry map:

\[ S(x, y) = \bigoplus_{l=L_1}^{L_2} S_l(x, y) \]  \hspace{1cm} (9.4)

where \( L_1 \) is the first, and \( L_2 \) is the last scale. The operator \( \bigoplus \) rescales all maps to the first scale, and subsequently sums the values of the different scales.

Since we are interested in all symmetrical regions in our robotic system, we do not apply the normalization that we used in Chapter 3, because that promotes symmetry maps with only one dominant salient point. Instead, the values in the saliency map are normalized between 0 and 1.

### 9.2.3 The SymRoID Model

A simplified flow chart of the complete SymRoID model is given in Figure 9.1c. It consists of a number of steps:

1. The symmetry map is calculated by the multi-scale symmetry model as described earlier.

2. Local maxima. A pixel \((x_m, y_m)\) is a local maximum if it has the highest value in its \(3 \times 3\) pixels neighborhood, and its symmetry value \(S(x_m, y_m) \geq \tau\), where we used \(\tau = 0.5\) in our experiments.

3. The local maxima are seeds for a region-growing algorithm. The flood-fill algorithm that we applied, takes a local maximum, and grows the area to add all neighboring pixels that have a symmetry value of \(S(x, y) \geq \lambda \cdot S(x_m, y_m)\), where the threshold is a ratio, \(\lambda\), of the symmetry value of the local maximum. Connecting regions are merged. The region growing results in clusters of symmetry pixels. In our experiments, we used \(\lambda = 0.5\).
4. The symmetry-pixel clusters contain the pixels that are the centers of symmetry. Since we are interested in symmetrical regions-of-interest, the complete symmetrical pattern that contributed to these symmetry centers needs to be found. To do so, the radius that has the highest contribution to its symmetry value is stored for every pixel. If $p_i$ and $p_j$ form the pixel pair with the highest symmetry contribution $s_{\text{max}}(i,j)$, then $r_s = \|p_i - p_j\|/2$ is the maximally contributing symmetry radius. A circle with center $p$ and radius $r_s$ is then marked in the regions-of-symmetry map.

5. Finally, the regions-of-interest are determined by taking the bounding box of the different regions in the regions-of-symmetry map. The regions-of-interest can overlap.

Some examples of regions-of-interest found by SymRoID can be found in Figure 9.2. It shows two pairs of subsequent images from the SLAM database.

### 9.2.4 The Region-of-Interest Descriptor

The detected regions-of-interest are described using a histograms-of-gradients (HoGs) descriptor, similar to the SIFT descriptor (Lowe, 2004). A region is first resampled to a $16 \times 16$ pixels descriptor window. This window is then divided into 16 squares (see Figure 9.1c). For each square, a histogram-of-gradients is calculated from the intensity gradients of the $4 \times 4$ pixels that are in the square. Such a histogram contains 8 bins, for the different gradient orientations, i.e., $[0, \frac{1}{4}\pi), [\frac{1}{4}\pi, \frac{1}{2}\pi), [\frac{1}{2}\pi, \frac{3}{4}\pi)$, etc.

The values of the 8 bins in each of the 16 histograms form the 128-dimensional region-of-interest descriptor. Following (Lowe, 2004), the descriptor is normalized to achieve invariance to changes in intensity, resulting in a vector of unit length.

Since the descriptor window adapts to the size of the region-of-interest, and the size of the region-of-interest itself is determined by the observed symmetrical pattern, the SymRoID model is scale invariant. This makes it possible to detect a landmark from different distances. Moreover, the descriptor is relatively invariant to small translational changes due to noise and affine transformations due to change of perspective, as discussed in (Lowe, 2004). Unlike the standard SIFT descriptor (Lowe, 2004) and the symmetrical interest-point detector proposed in Chapter 8, we did not add rotational
invariance, since our robot drives on flat surfaces, and will therefore not encounter rotational transformations of the stimulus.

9.3 The Visual SLAM System

To ensure that only landmarks are added to the map of the environment that are stably detectable over a number of sequential observations, we use a visual buffer that tests the stability of the regions-of-interest. When a region passes the buffer, it is added as a landmark to the SLAM system.

9.3.1 The Visual Buffer

Like in the work presented in the Chapter 8, a visual buffer is used to test the stability of the regions-of-interest to make sure that only stable landmarks are added to the map of the environment. Although the functionality of the visual buffer is the same as that discussed in Section 8.2.2, it is recapitulated here for reasons of completeness.

The visual buffer contains the $N$ most recent camera images. The regions in the current image are compared to those in the $N-1$ previous images. A region, $i$, passes the buffer if it is matched in at least $M$ of the previous images. Two regions, $i$ and $j$, match when the descriptors, $d_i$ and $d_j$, are sufficiently similar. This is true when the Euclidean distance is below the threshold $\tau_1$,

$$\|d_i - d_j\| < \tau_1,$$

and when the best-to-next-best ratio is smaller than the threshold $\delta_1$,

$$\|d_i - d_j\|/\|d_i - d_l\| < \delta_1,$$

where $d_l$ is the descriptor of the second most-similar region in the previous image. This ratio ensures uniqueness.

The parameter settings used for SymRoID compared with those of MUST and SIFT are:
Figure 9.2: Examples of regions-of-interest found in the images. Both pairs contain two sequential images. Note that the bounding box of a region is used to calculate the descriptor.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>$\tau_1$</th>
<th>$\delta_1$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SymRoID</td>
<td>7</td>
<td>5</td>
<td>0.6</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>MUST</td>
<td>12</td>
<td>4</td>
<td>0.3</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>SIFT</td>
<td>14</td>
<td>8</td>
<td>0.3</td>
<td>0.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$\nu$ is not used in the visual buffer for SymRoID.

9.3.2 The SLAM system

We use a standard implementation of the Extended Kalman Filter (EKF) as basis of the SLAM system (see Appendix C). Our methods and results, however, should also be valid for other SLAM approaches. In this section, the incorporation of the landmark observations in EKF-SLAM is discussed. The method is largely similar to that presented in Section 8.2.3.

The position of a landmark in the environment that pass the visual buffer is obtained by comparing the different observations of the region. Estimates of the position are made by triangulation using the bearings of the observations and the displacement of
9.3. The Visual SLAM System

the robot, and by inferring depth information from the change in area of the regions-of-interest and the displacement of the robot. This results in a set of \( K \) estimations:

\[
P = \{ p_k | p_k = \langle r_k, \theta_k \rangle \land 1 \leq k \leq K \}, \tag{9.7}
\]

where \( r_k \) and \( \theta_k \) are respectively the range and bearing of the estimation. The position of the landmark is then determined by the mean of \( P \), and the uncertainty by its covariance matrix.

A landmark \( i \) with descriptor \( d_i \) that results from the buffer is classified as either a new landmark, or a previously observed landmark that is already in the map. It concerns a previously observed landmark if the landmark in the database with the most similar descriptor, \( d_j \), fulfills three criteria:

1. Similarity in descriptors:

\[
||d_i - d_j|| < \tau_2 \tag{9.8}
\]

2. A small best-to-next-best ratio to only match unique landmarks:

\[
||d_i - d_j|| / ||d_i - d_l|| < \delta_2 \tag{9.9}
\]

where \( d_l \) is the second most similar descriptor in the database.

3. A small distance in the EKF map, measured by the Mahalanobis distance:

\[
\sqrt{(x_i - x_j)^T S_j^{-1} (x_i - x_j)} < \eta \tag{9.10}
\]

where \( S_j \) is the uncertainty covariance matrix, discussed in the next paragraph.

The parameter values used for SymRoID compared to those used for MUST and SIFT are:

<table>
<thead>
<tr>
<th></th>
<th>( N )</th>
<th>( M )</th>
<th>( \tau_1 )</th>
<th>( \delta_1 )</th>
<th>( \nu )</th>
<th>( \tau_2 )</th>
<th>( \delta_2 )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SymRoID</td>
<td>7</td>
<td>5</td>
<td>0.6</td>
<td>0.8</td>
<td>-</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>MUST</td>
<td>14</td>
<td>8</td>
<td>0.3</td>
<td>0.8</td>
<td>0.8</td>
<td>0.3</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>SIFT</td>
<td>14</td>
<td>8</td>
<td>0.3</td>
<td>0.5</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The landmark is classified as new only if none of the three criteria is fulfilled. For a new landmark, the state matrix and covariance matrix are augmented using the observation,
9.4 Experiments and Results

9.4.1 Experimental Setup

To test the stability and robustness of the symmetrical regions-of-interest, and to test the SLAM performance using SymRoID, the same SLAM database is used as introduced in Chapter 8. The database is recorded with a Pioneer II DX robot, equipped with a Sony D31 camera. The database contains the camera images and the odometric information of ten different runs, in which the robot drove four laps through an office environment. Each lap was approximately 35 meters, and the robot drove at an average speed of 0.3 m/s. Camera images of $320 \times 240$ were stored at 5Hz. At intervals of one meter, the true location of the robot was logged by hand. This enabled us to quantify the SLAM.
9.4. Experiments and Results

In the experiments, we tested the performance of SymRoID, and compared it to MUST (Chapter 8) and SIFT Lowe (2004). The stability, robustness, and SLAM-performance experiments are setup identically to those discussed in the previous chapter in Section 8.3.

9.4.2 Stability

To test the stability of the regions-of-interest, multiple sequences of images recorded by the robot are taken. A region or point in the last image of the sequence is considered stable if it is matched in all the previous images of the sequence. Two regions match when the Euclidean distance between the two descriptors, $d_i$ and $d_j$, is $\|d_i - d_j\| < 0.6$.

Figure 9.3 shows the stability of the symmetrical regions-of-interest compared to the stability of interest points obtained by MUST (Chapter 8) and SIFT (Lowe, 2004). It can be appreciated that the stability of the symmetry methods is higher than that of SIFT. Moreover, the use of symmetrical regions-of-interest gives considerably higher stability than the use of interest points. These results show that the symmetrical regions-of-interest are stably tracked over a number of sequential images, suggesting that they are more robust to small changes in viewpoint than MUST and SIFT.

9.4.3 Robustness

In Mikolajczyk et al. (2005) a benchmark for testing the robustness of detectors is presented. Unfortunately, it is not suited for testing landmark selection methods for SLAM in indoor environments since it contains outdoor scenes. To test the robustness of SymRoID in indoor environments, a subsample of our SLAM database is used. Images are taken from one of the runs in the database with intervals of 3 meters, so that the complete environment is represented in the subsample. To test the noise robustness of SymRoID, we smoothed the image with a Gaussian kernel and added Gaussian noise to the pixels. In addition, we manipulated the contrast and brightness of the images to test the robustness to changing light conditions (see Section 8.3.2 for a full description of the experiments).

The robustness is measured by the proportion of matching regions between the original and the manipulated images. Two regions match when (9.5) and (9.6) are met, where
Figure 9.4: Robustness to noise and changing light conditions. The lines give
the mean proportion of matched regions-of-interest or interest points between the
distorted and the original images of 57 runs. The gray areas around the lines depict
the 95% confidence intervals on the mean. Note that these intervals for MUST and
SIFT are sometimes small, and therefore hardly visible.

$\tau_1 = 0.6$ and $\delta_1 = 0.75$. Additionally, the distance between the two regions in the image
should be less than 3 pixels.

The results in Figure 9.4 show the robustness results. The lines give the mean performance over the 57 images, with the 95% confidence intervals on the mean given by the gray areas. The symmetry models are significantly less affected by the two types
of noise than SIFT (Figure 9.4a and b). Moreover, using regions-of-interest instead of interest points gives a significant improvement. The reason for this success is likely to be the larger area that is used as evidence for the presence of symmetry. SymRoID is therefore less vulnerable to noise.

For the contrast manipulation, SymRoID performs significantly better when there is low contrast. With enhances contrast, on the other hand, SIFT performs better (Figure 9.4c). Using regions-of-interest results in worse performance for the brightness manipulation. MUST is the best method when the brightness is enhanced (Figure 9.4d). The main influence of the contrast manipulations is on the magnitudes of the image gradients. This influences the symmetry value in Equation (9.1). Since a threshold is used on the symmetry values to determine which pixels are part of the symmetrical region and which not, the change of light conditions can thus result in larger or smaller regions. This explains the reduced performance of SymRoID in changing light conditions.

9.4.4 SLAM Performance

To test the SLAM performance, the first three laps of each run are used to train the EKF. The fourth and last lap is used to determine the estimation error for that run. This is done by calculating the Euclidean distance between the EKF estimation of the robot's position and the ground-truth position. The estimation error is the average distance in the last lap.

The parameters for the buffer and for the matching of regions with the landmark \((\tau_1, \tau_2, \delta_1, \delta_2)\) are optimized for all three landmark selection methods. The overall best settings for SymRoID, MUST, and SIFT are given in section 9.3.

Figure 9.5 shows the estimation error. The bars give the mean over the 10 runs, and the error bars depict the 95% confidence intervals. The horizontal dashed line and the horizontal gray bar show the mean and 95% confidence interval of the odometry error. It can be appreciated that the use of symmetry by SymRoID and MUST gives significantly better SLAM performance. Moreover, the use of symmetrical regions-of-interest significantly outperforms the use of interest points. This is true for both the best settings per run and the overall best settings.

Also computationally, SymRoID outperforms the other models. SIFT selects on average 120 interest points, MUST 40, but SymRoID selects only half a dozen regions per
Chapter 9. Paying Attention to Symmetrical Regions of Interest

Figure 9.5: The slam performance. The bars give the mean estimation error in the last lap through the environment, after a map has been established in the previous three laps. The error bars are the 95% confidence intervals on the mean. The first group of bars give the results where the best buffer and matching settings per run are used, the second group shows the performance for the overall best settings. The horizontal dashed line gives the mean error when only the odometry information is used. The height of the horizontal gray bar represents the 95% interval.

image, thereby greatly reducing the computations in the buffer. Moreover, both SIFT and MUST take in the order of a second to calculate interest points from an image, while SymRoID finds regions about four times as fast. This improvement is due to the coarse scale space used by SymRoID, in contrast with the detailed scale space used by SIFT and MUST.

9.5 Discussion

In this chapter, a symmetrical region-of-interest detector, SymRoID, is presented. The model selects regions-of-interest in the image using local symmetry. The model is used to select landmarks for a visual SLAM system. The stability, robustness, and SLAM performance of the model is tested and compared with that of MUST, a symmetrical interest-point detector proposed in Chapter 8, and SIFT, an interest-point detector based on contrast features Lowe (2004). The results showed that the use of symmetry improves the stability and robustness to noise, and yields significantly better SLAM
performance. Moreover, the use of symmetrical regions-of-interest outperforms the use of interest points. SymRoID also is more robust to decreased contrast in the image. However, for enhanced contrast and brightness manipulation, the model scores worse than the others.

The higher stability of the symmetrical regions of interest shows that the regions are more robust to small changes in perspective. This is probably due to the descriptor window, which is adapted to the size of the region of interest. When there is a change in perspective, the size of the region will change and the descriptor region will adapt. This makes it more robust to perspective changes than the interest points, which always have a squared descriptor window.

The improved robustness to noise of the symmetry model can be explained by the fact that symmetrical regions are intrinsically redundant. Both sides of a mirror-symmetrical pattern contain the same information. This redundancy makes the detection less susceptible to noise. Furthermore, by using regions, instead of points, more evidence of the existence of symmetry is gathered, making the model more robust. Robustness to noise is an important property, since robots usually operate under noisy conditions.

Due to the fact that symmetrical patterns are redundant, the descriptor of the symmetrical regions of interest is redundant as well. It should therefore be possible to decrease the dimensionality of the descriptor by disregarding one half of the symmetrical pattern, while maintaining the same descriptive power. Future work needs to show if this is the case.

The worse robustness to changing light conditions can be explained by the role of the gradient magnitudes in the SymRoID model. When light conditions change, the gradient magnitudes in the image are influenced. This results in different symmetry values calculated by Equation (9.1), which has a consequences for the size of the symmetrical regions, due to the fixed threshold used to determine which pixels are member of the symmetrical regions. More research is required to overcome this problem, since robustness to changing light conditions is an important property when a robot needs to operate in an environment over extended periods.

As can be seen in Figure 9.2, the symmetrical regions-of-interest often coincide with objects. The umbrella, radiator, and blinds, for instance, are detected as interesting regions. This shows that symmetry can be used as a bottom-up object detector. This confirms the Gestalt notion that symmetry is a cue for figure-ground segregation (see Chapter 5 for a full discussion). However, it can also be seen in the figure that non-
objects are selected. A clear example is the area between the two blinds. This area
is indeed symmetrical, however, it does not coincide with an object, but rather with
the background. To solve these kind of situations, other Gestalt principles need to be
incorporated in the region-of-interest detector. The principle of *closure*, for instance,
will reject the area in between the blinds as an object. We propose to use more Gestalt
principles for figure-ground segregation to develop bottom-up object detectors.
9.6 Visual Attention and Active Vision in Machines

Part II of the dissertation discussed visual attention and active vision in artificial vision systems. In Chapter 7, an active approach was proposed for object recognition. The results showed that recognition greatly improves when a robot can explore an object. The active exploration gives it the possibility to observe the object from different viewpoints. This not only builds better three-dimensional object models, it also greatly simplifies the segmentation of the object from its background and enables the robot to test the stability of interest points so that only robust points are included in the object representation.

Inspired by the findings in Part I, symmetry is used in Chapter 8 and Chapter 9 to detect respectively interest points and regions of interest in camera images to serve as visual landmarks, thereby exploiting the many symmetrical patterns present in most indoor environments. The results show that landmarks selected by symmetry are more stable, have a higher repeatability, and are more robust to noise and changing light conditions than those selected by contrast features. Moreover, the SLAM performance improves when symmetry is used, showing that valuable landmarks are selected.

The next chapter concludes the thesis with a discussion on the results and the main insights gained from the multi-disciplinary study to visual attention and active vision in natural and in artificial systems.