A Scale Invariant Neural Feature Detector for Visual Homing

Andrew Vardy Franz Oppacher

School of Computer Science
Carleton University
Ottawa, K1S 5B6, Canada
Fax: +1 (613) 520-4334
avardy@scs.carleton.ca
http://www.scs.carleton.ca/~avardy

Abstract

A novel feature detector is presented for extracting local image features which are invariant to scale. This detector is composed of simple neuron-like elements whose activation is given by the sum of their weighted inputs. These features are employed for the visual homing task and found to yield superior performance to the *warping method* [1] on panoramic images of an unstructured environment.

1 Introduction

Visual homing is the act of returning to a place by comparing the image currently viewed with an image taken when at the goal (the snapshot image). While this ability is certainly of interest for mobile robotics, it also appears to be a crucial component in the behavioural repetoire of insects such as bees and ants [2]. We present here a novel method for visual homing employing a feature detector which might plausibly be implemented in the limited hardware of the insect brain.

Approaches to visual homing range from those purely interested in robotic implementation (e.g. [3]) to those concerned with fidelity to biological homing (e.g. [4]). Both camps have proposed methods which find correspondences between image features and use these to compute a home vector. These feature-based methods rely on visual features such as regions in 1-D (one-dimensional) images [5, 6], edges in 1-D images [4], image windows around distinctive points in 1-D images [3], coloured regions in 2-D images [7], and Harris corners in 2-D images [8, 9]. Any visual feature is subject to distortions in scale, illumination, and perspective, as well as distortions from occlusion. The ability to correspond features in the presence of these distortions is critical for feature-based homing. Scale invariant feature detector schemes do exist. Notable examples include Lowe's scale invariant keypoints [10], and a visual homing method using scale invariant features based on the Fourier-Mellin transform [11]. However, it is currently unclear how complex these schemes might be for implementation in the neural hardware of an insect. The feature detector presented here is partially invariant to scale and has a direct and simple neural implementation.

An alternate approach which relies upon global image properites as opposed to features, is Franz et. al's *warping method.* This method warps 1-D images of the environment according to parameters specifying displacement of the agent. The parameters of the warp generating the image most similar to the snapshot image specify an approximate home vector. As the warping method is known for its excellent performance (see comparative studies in [12, 13]) we use it here for comparison with our homing method.

In the next section we describe this scale invariant feature detector. We then give details on the overall homing algorithm and compare performance with the warping method.

2 A Scale Invariant Neural Feature Detector

2.1 Structure

This feature detector is composed of a set of neuron-like processing elements with wedge-shaped receptive fields. Figure 1 depicts a 4-layer detector. Connections are feedforward from the input image layer. The receptive fields are shaped like a wedge or piece of pie, with the cusp being defined as the narrow pointed end. Connection weights are not uniform but decrease as the distance from the cusp increases. Justification for this decay in connection weights is provided below.

2.2 Function

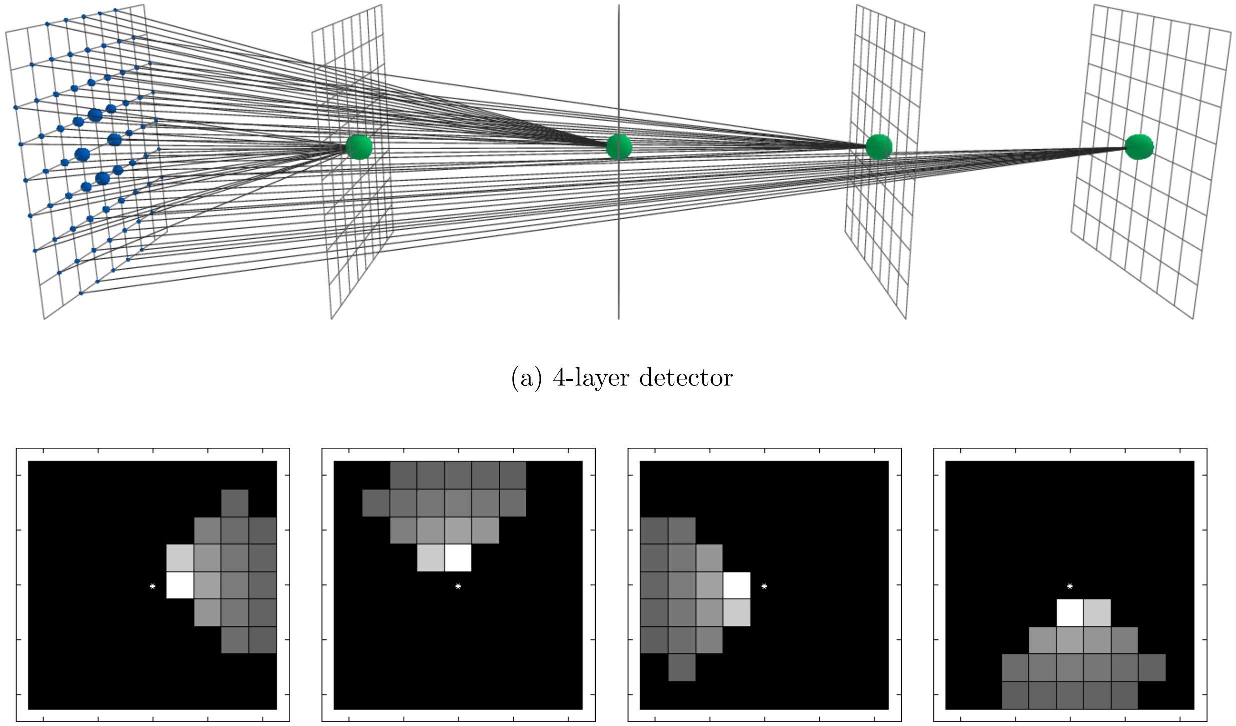
The input layer holds an image of lower-level features such as edges or corners. Each detector element computes its activation level by summing the weighted inputs from the wedge of cells in the input layer covered by its receptive field. The vector of activation levels is $\boldsymbol{f}$. We denote the i-th element by f_i, and the i-th receptive field by R_i. The set of triples (x, y, w) in R_i defines the connectivity of element i's receptive field to the input layer, where x and y are image coordinates and w is the connection weight. The value of pixel (x, y) in the input image is given by $I(x, y)$. Equation (1) gives the function for f_i.

$$f_i = \sum_{(x,y,w)\in R_i} wI(x, y) \tag{1}$$

Let a feature detector be centred about a point p in the input image. Connection weights decay out inversely in proportion to distance from p in the plane of the input image. This relation is given by equation (2).

$$w(d) = \frac{1}{d}, \qquad \text{where } d = \sqrt{(x - p_x)^2 + (y - p_y)^2} \tag{2}$$

This decay equation ensures that the ratio of weights from two distances, d_1 and d_2, stays the same after a scaling by factor k of both distances.



(a) 4-layer detector

(b)

(c)

(d)

(e)

Figure 1: 4-layer scale invariant feature detector of radius 4. **a:** Three-dimensional view of detector. The grid on the far left represents the input image. Large spheres indicate elements of the detector, while small spheres show connection points with the input image. The radius of the small spheres is proportional to connection weight, which is given by equation (2). **b-e:** Receptive fields of the individual elements. The level of whiteness is proportional to connection weight. The '*' marks the center of each receptive field.

$$\frac{w(d_1)}{w(d_2)} = \frac{w(kd_1)}{w(kd_2)} \tag{3}$$

This property will ensure that a pure scaling of the local image pattern will affect only the length of a feature vector, and not its direction. Thus, the normalized feature vector $\hat{\boldsymbol{f}} = \boldsymbol{f}/\|\boldsymbol{f}\|$ is invariant to scale changes. The example given below should help to illustrate this invariance.

To compare two feature vectors we find the distance between them in n-dimensional vector space. These vectors will have been extracted from two image positions a and b. For our homing algorithm, position a will be a point in the snapshot image while position b will be in the current view image. To denote the distance between two vectors extracted from positions a and b, we use following notation,

$$DIST_{\zeta}^{a,b} = \|\hat{\boldsymbol{f}}^a - \hat{\boldsymbol{f}}^b\|, \quad \text{where } \hat{\boldsymbol{f}}^a, \hat{\boldsymbol{f}}^b \text{extracted using paramater settings } \zeta \tag{4}$$

The ζ term just indicates that parameter settings for the $DIST$ function will be shown in subscript. Note that as $DIST$ finds the distance between normalized vectors its range is $[0, 2]$.

2.3 Example

The following example illustrates the scale invariance property of this feature detector. Consider the case of an image with just three non-zero image pixels. Figure 2 shows three different arrangements of pixels. In Figure 2a we have the original arrangement about point p. Figure 2b shows the same arrangement after doubling all distances to p. The centre point is now called p'. Finally, figure 2c shows an arrangement of three distractor pixels. The distance between feature vectors from the first two arrangements should be quite small, but distance between the first and third should be large. Along with varying the arrangement of pixels, we also vary the parameters of the feature detectors used in this figure. We denote each feature vector by $\hat{\boldsymbol{f}}^{point}_{radius,condition}$ where *point* refers to a position such as p, *radius* gives the detector radius, and *condition* is either u for uniform weights or d for decayed weights. This same notation applies to the $DIST$ function described in equation (4).

We first consider the correlation of the original feature vectors with those of the distractor pattern. Note that pixels D, E, F in figure 2c have the same values and general orientation as A, B, and C in figure 2a. However, D, E, and F, are at distances of 8, 1, and 1 units as opposed to 1, 2, and 4. For infininte radius detectors and uniform weights, the distance between the vectors, $DIST^{p,q}_{\infty,u}$, is zero indicating that the two configurations are the same. This error occurs because no information about the relative distance distributions has been retained to disambiguate these two configurations. However, with decayed weights the distance becomes 1.12, indicating dissimilarity. With decayed weights and radius 6 vectors the distance is 1.25, also indicating dissimilarity. Thus, the decayed weights allow discrimination between configurations based upon distance.

Another reason for decayed weights is to minimize the influence of outlying pixels. Pixel C in the scaled configuration of figure 2b falls outside of the radius 6 detector. $DIST^{p,p'}_{6,u}$ is 0.56 whereas for the infinite radius detector the distance is zero. The loss of C causes a large

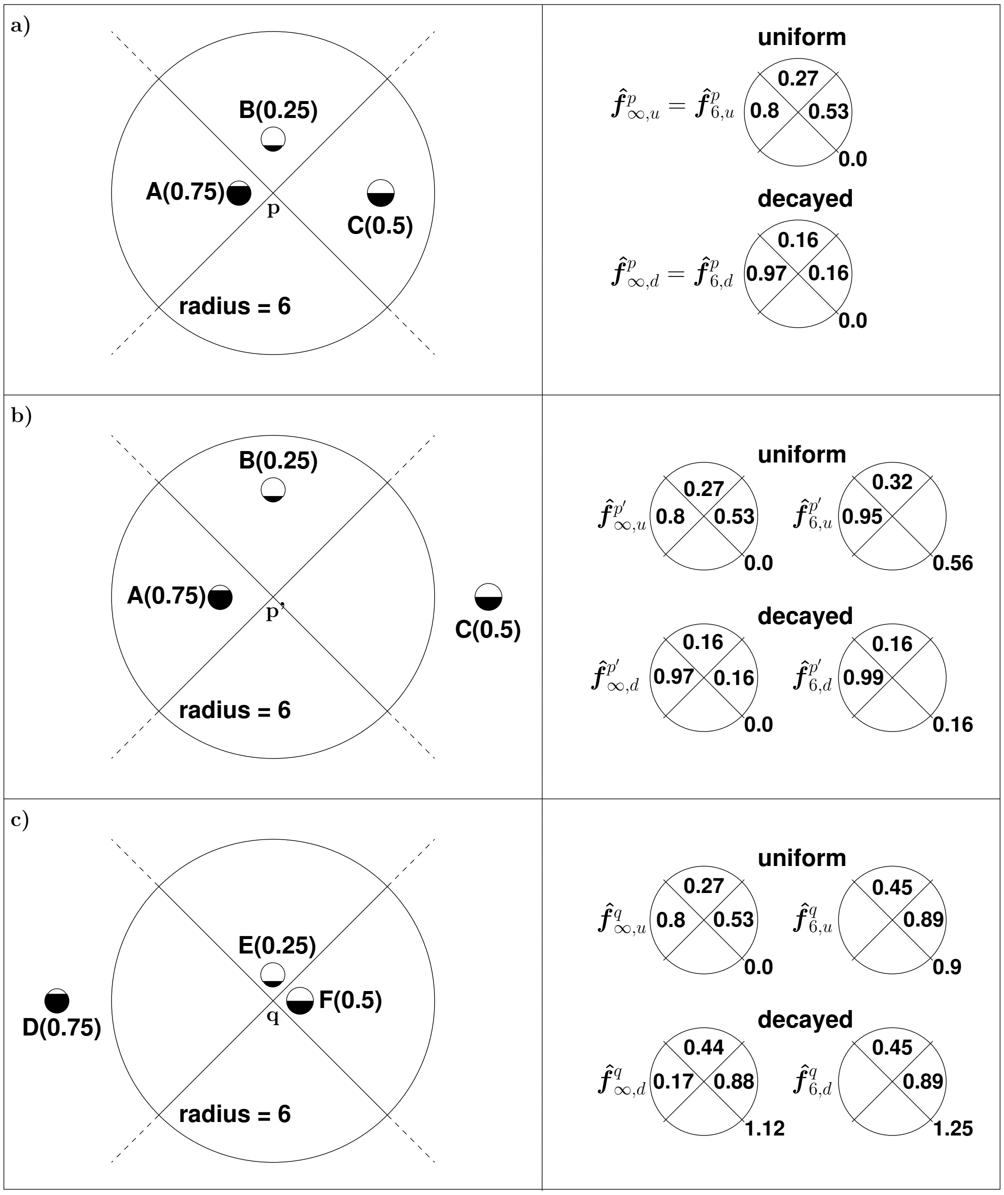

a)
B(0.25)
A(0.75)
p
C(0.5)
radius = 6
uniform
$\hat{\boldsymbol{f}}^{p}_{\infty,u} = \hat{\boldsymbol{f}}^{p}_{6,u}$
0.27
0.8
0.53
0.0
decayed
$\hat{\boldsymbol{f}}^{p}_{\infty,d} = \hat{\boldsymbol{f}}^{p}_{6,d}$
0.16
0.97
0.16
0.0
b)
B(0.25)
A(0.75)
p'
C(0.5)
radius = 6
uniform
$\hat{\boldsymbol{f}}^{p'}_{\infty,u}$
0.27
0.8
0.53
0.0
$\hat{\boldsymbol{f}}^{p'}_{6,u}$
0.32
0.95
0.56
decayed
$\hat{\boldsymbol{f}}^{p'}_{\infty,d}$
0.16
0.97
0.16
0.0
$\hat{\boldsymbol{f}}^{p'}_{6,d}$
0.16
0.99
0.16
c)
E(0.25)
F(0.5)
D(0.75)
q
radius = 6
uniform
$\hat{\boldsymbol{f}}^{q}_{\infty,u}$
0.27
0.8
0.53
0.0
$\hat{\boldsymbol{f}}^{q}_{6,u}$
0.45
0.89
0.9
decayed
$\hat{\boldsymbol{f}}^{q}_{\infty,d}$
0.44
0.17
0.88
1.12
$\hat{\boldsymbol{f}}^{q}_{6,d}$
0.45
0.89
1.25

Figure 2: Motivating example for decayed weights. **a:** Three active pixels (A, B, and C) arranged in a particular pattern about a point p. The receptive fields of a 4-layer detector are superimposed onto the image plane. These receptive fields all meet at p and extend out to a radius of 6 units. Dotted lines extending the receptive fields outwards indicate an alternate detector with an effectively infinite radius. On the right are the feature vectors detected from the pattern on the left, as computed by equation (1), for two different scenarios. In the top scenario all connection weights w are unity. In the bottom scenario, the values of w are given by equation (2). **b:** Shows A, B, and C after being scaled about p by a factor of two. On the right four different scenarios are depicted. The top row is for uniform weights while the bottom is for weights decayed according to equation (2). The left column shows feature vector values for a detector of infinite radius. The right column is for a detector of radius 6. The value of DIST is shown at the bottom right of each feature vector, where each feature has been compared with its corresponding feature from figure 2a (e.g. $\hat{\boldsymbol{f}}^{p'}_{\infty,d}$ with $\hat{\boldsymbol{f}}^{p}_{\infty,d}$). **c:** shows a distractor pattern of three pixels D, E, and F with detectors centred at q.

change. However, with decayed weights the loss of C has much less impact: $DIST_{6,d}^{p,p'} = 0.16$. Figure 2b is an example of expansion. There will often be non-zero pixels expanded out of the detector's range. Thus it makes sense to partially discount outlying pixels. A similar argument can be made for contraction.

3 Results

A database of panoramic images was collected using a robot-mounted camera pointed upwards at a hyperbolic mirror. These images were captured within the robotics lab of the Computer Engineering Group, Bielefeld University on a 10x17 grid mapped out on the floor of this room. They have been transformed into cylindrical images and downsampled to a relatively low-resolution (202x46). This level of resolution is roughly consistent with that of honeybee eyes[1].

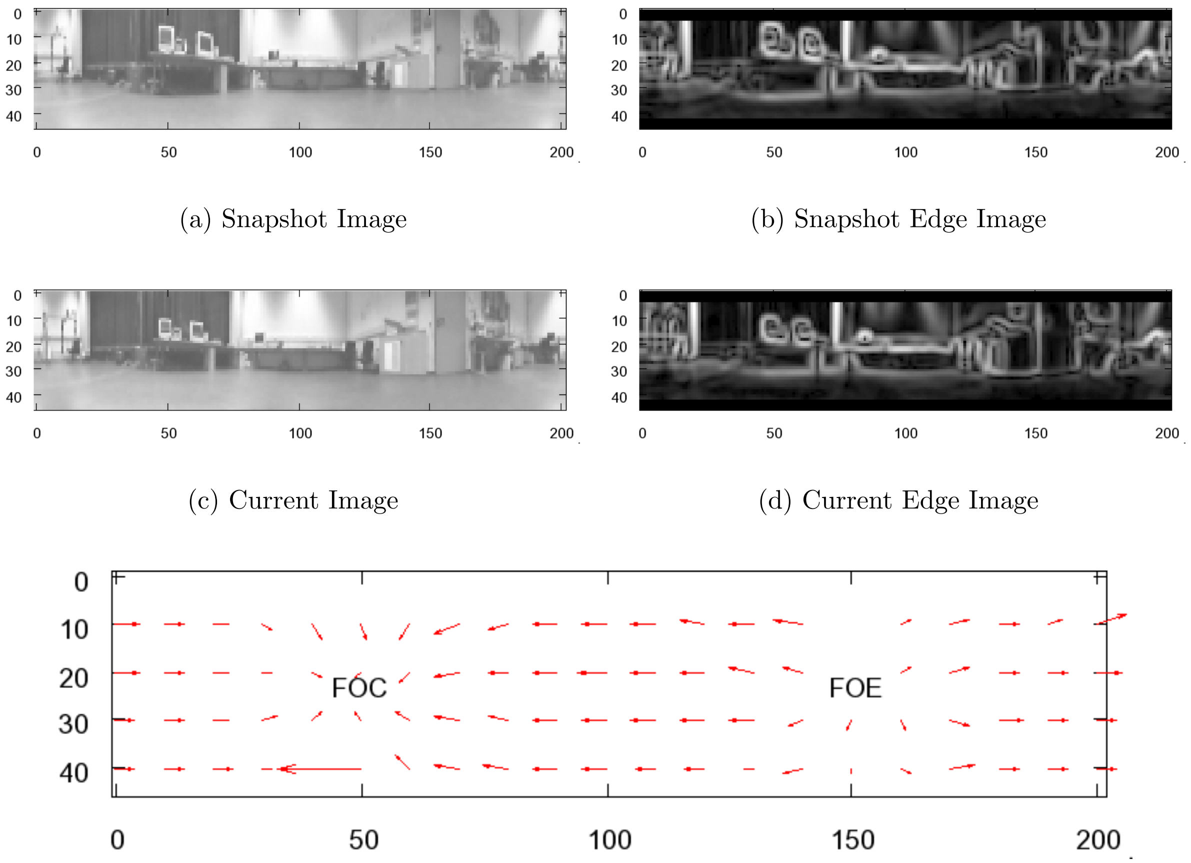
We present now the details of the homing algorithm. For both snapshot and current images we first extract edges as features using the Sobel filter. Note that these edge images are not thresholded or otherwise post-processed. Next our detectors are applied to generate feature vectors. To speed the correspondence process features are extracted only from every 10th image position in the snapshot image. Feature vectors are extracted from all positions in the current image. The detectors used for homing have a radius of 20 pixels, with each wedge subtending 10^o. The wedges are adjacent and non-overlapping, thus there are $360/10 = 36$ layers in each feature detector. After feature vectors are extracted they are all individually normalized. Finally, each candidate feature vector from the snapshot image is paired with all feature vectors of the current image lying within a 30-pixel radius. The closest pairing is used to generate a correspondence vector.

Figure 3 gives an example of all processing mentioned thus far. The correspondence vectors in figure 3(e) have the correct overall structure for this leftwards movement from the snapshot position to the current position. The labels of FOC and FOE indicate the ideal focus-of-contraction and focus-of-expansion for this movement. There are obviously incorrect vectors such as the one located at (50, 40). However, the success of this method, as revealed below, shows that the incorrect correspondences are generally outnumbered by the correct correspondences, at least within this environment.

Each correspondence vector is mapped individually to an approximate home vector using the method presented in [8]. This method requires that snapshot and current view be taken from the same orientation which is the case here. In a free-roving robot experiment some manner of compass would have to be employed to correct for orientation changes. The overall motion vector is computed by averaging all home vectors.

We compare the homing performance of this method now with the warping method. One way to compare these two methods is to assign one particular grid position as the goal. Figures 4(a) and 4(b) depict the motion vectors generated when position (5,8) is selected as the goal for both the warping method and our method. For this position both methods

[1] Land reviews the visual acuity data of many insects in [14] and presents four differing estimates of *Apis Mellifera* (honeybee) interommatidial angles. Larger interommatidial angles give lower resolution images. The low resolution used here is approximately consistent with the largest interommatidial angle reported by Land ($\Delta\phi = 1.7^o$).



(a) Snapshot Image

(b) Snapshot Edge Image

(c) Current Image

(d) Current Edge Image

(e) Correspondence Vectors

Figure 3: Examples of the images input to the homing algorithm (**a** and **c**), the edge images extracted from these (**b** and **d**), and the correspondence vectors found by correlating feature vectors extracted from the two edge images (**e**). The snapshot image here is from position (5,8) of the image database and the current image is from position (3,8). The actual focuses of contraction and expansion for a leftwards movement from position (5,8) to (3,8) are labelled FOC and FOE.

exhibit a perfect *return ratio* of 1.0. Return ratio is measured for a particular position by taking that position as the goal and counting the number of other positions from which the goal can be reached. Each such homing attempt is simulated by allowing an agent to process the image from its current virtual position, compute a motion vector, and then step to its new position—repeating until it reaches the goal, leaves the grid, or gets caught in a loop.

A more thorough test, however, is to evaluate the return ratio across all positions. Figure 4 shows the return ratios for both methods across the 10x17 image database. The minimum and average return ratios for our method are higher. Grid positions are paired and the differences between pairs subject to statistical analysis. The median difference was found to be significantly greater than zero, indicating superior return ratio of our method within the test environment (sign test, P=0.0011).

The region at the bottom of these grids is problematic for both methods. At the end of the room which this region represents there is a large bench against a white wall. Images taken here show the bench as large and dark. From further out, however, the white wall is more prevalent. This change in appearance upsets the warping method's global matching. The minimum return ratio for the warping method is at position (1, 16), adjacent to the bench. There the return ratio is 0.064, meaning that this position can only be returned to from 6.4% of the room. The minimum for our method is at position (9, 16), which is also adjacent to the bench. The return ratio here, however, is 0.541, indicating that this worst position can still be reached from 54.1% of the room.

4 Conclusions

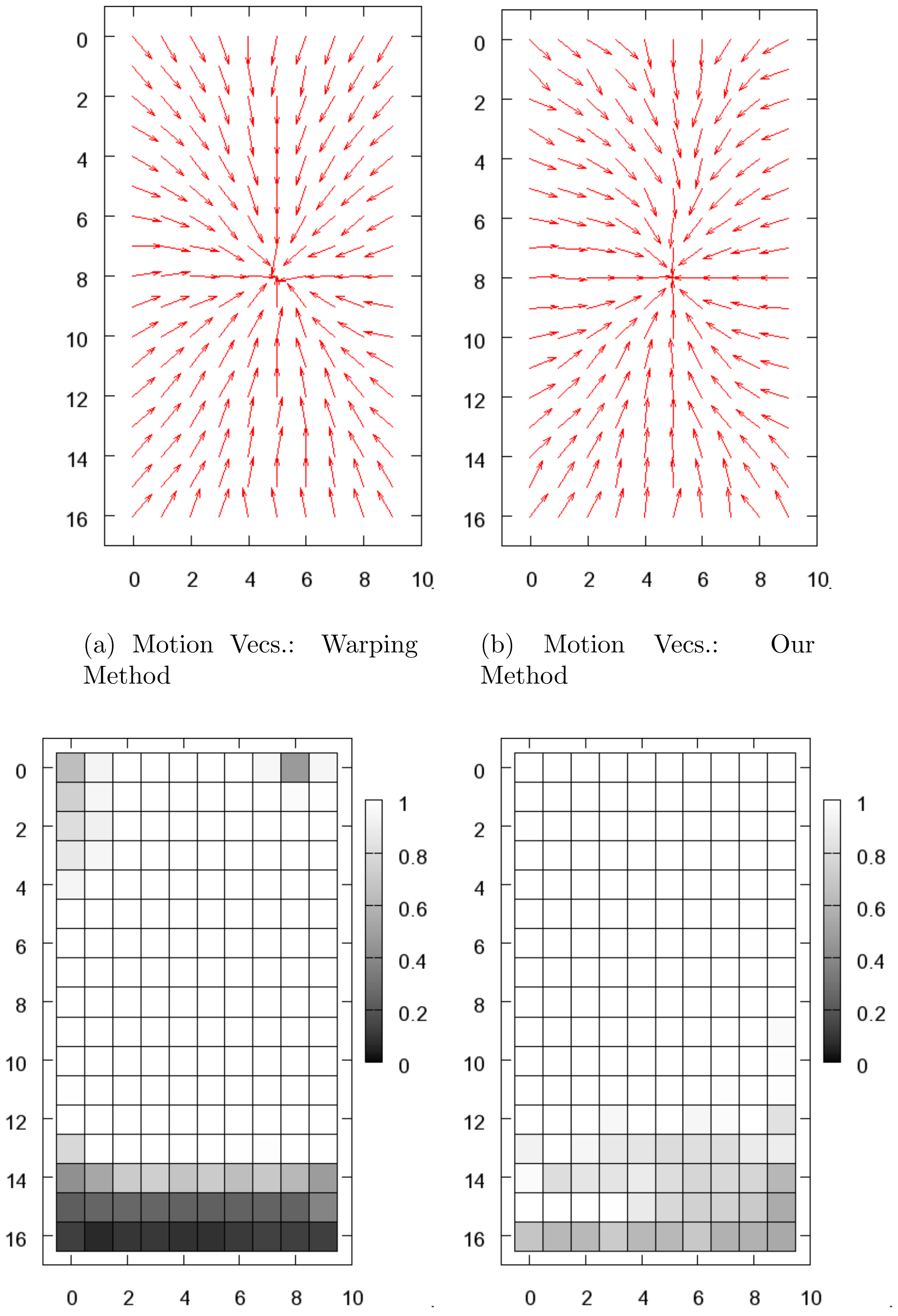
The purpose of this paper was to introduce a novel scale invariant feature detector and show how this detector could be employed for visual homing. Current work is focussed upon understanding the detector's parameter space, developing comparisons with other scale invariant feature detectors (e.g. [10]), and creating a more local process for finding correspondences between feature vectors.

Acknowledgments

Many thanks to Ralf Möller for the use of his lab and equipment, along with helpful reviews and the C++ implementation of the warping method. Thanks also to his students Frank Röben, Wolfram Schenck, and Tim Köhler for discussion and technical support. This work has been supported by scholarships from NSERC (PGS-B 232621 - 2002), and the DAAD (A/04/13058).

References

[1] Franz, M., Schölkopf, B., Mallot, H., Bülthoff, H.: Where did I take that snapshot? Scene-based homing by image matching. Biological Cybernetics **79** (1998) 191–202



(a) Motion Vecs.: Warping Method

(b) Motion Vecs.: Our Method

(c) Return Ratio Map: Warping Method

(d) Return Ratio Map: Our Method

Figure 4: **a** and **b:** Motion vectors to goal position (5,8). **c** and **d:** Return ratio maps. **c:** Average and minimum return ratios are 0.868 and 0.065. **d:** Average and minimum return ratios are 0.947 and 0.541. The maximum value for both maps is 1.0.

[2] Collett, T.: Insect navigation en route to the goal: Multiple strategies for the use of landmarks. Journal of Experimental Biology **199** (1996) 227–235

[3] Hong, J., Tan, X., Pinette, B., Weiss, R., Riseman, E.: Image-based homing. In: Proceedings of the 1991 IEEE International Conference on Robotics and Automation, Sacremento, CA. (1991) 620–625

[4] Möller, R.: Insect visual homing strategies in a robot with analog processing. Biological Cybernetics **83** (2000) 231–243

[5] Cartwright, B., Collett, T.: Landmark learning in bees. Journal of Comparative Physiology **151** (1983) 521–543

[6] Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R., Wehner, R.: A mobile robot employing insect strategies for navigation. Robotics and Autonomous Systems, Special Issue: Biomimetic Robots **30** (2000) 39–64

[7] Gourichon, S., Meyer, J.A., Pirim, P.: Using colored snapshots for short-range guidance in mobile robots. International Journal of Robotics and Automation, Special issue on Biologically Inspired Robotics **17** (2002) 154–162

[8] Vardy, A., Oppacher, F.: Low-level visual homing. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J.T., Ziegler, J., eds.: Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL), Springer Verlag Berlin, Heidelberg (2003) 875–884

[9] Vardy, A., Oppacher, F.: Anatomy and physiology of an artificial vision matrix. In Ijspreet, A., Murata, M., eds.: Proceedings of the First International Workshop on Biologically Inspired Approaches to Advanced Information Technology. (2004) (to appear)

[10] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110

[11] Rizzi, A., Duina, D., Inelli, S., Cassinis, R.: A novel visual landmark matching for a biologically inspired homing. Pattern Recognition Letters **22** (2001) 1371–1378

[12] Weber, K., Venkatesh, S., Srinivasan, M.: Insect-inspired robotic homing. Adaptive Behavior **7** (1999) 65–97

[13] Möller, R.: A biorobotics approach to the study of insect visual homing strategies, Habilitationsschrift, Universität Zürich (2002)

[14] Land, M.: Visual acuity in insects. Annu. Rev. Entomol. **42** (1997) 147–177