A Scale Invariant Local Image Descriptor for Visual Homing

Andrew Vardy and 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 descriptor is presented for characterizing local image patches in a scale invariant manner. The descriptor is biologically-plausible in that the necessary computations are simple and local. Two different methods for robot visual homing based on this descriptor are also presented and tested. The first method utilizes the common technique of corresponding descriptors between images. The second method determines a home vector more directly by finding the stationary local image patch most similar between the two images. We find that the first method exceeds the performance of Franz et. al's *warping method*. No statistically significant difference was found between the second method and the warping method.

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 repertoire of insects such as bees and ants [1]. We present here two methods for visual homing which employ a novel image descriptor that characterizes a small patch of an image such that the descriptor is invariant to scale changes. Scale change is a prevalent source of image distortion in visual homing where viewed landmarks generally appear larger or smaller than in the snapshot image. The image descriptor developed here has a simple structure 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. [2]) to those concerned with fidelity to biological homing (e.g. [3]). 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 [4,5], edges in 1-D images [3], image windows around distinctive points

in 1-D images [2], coloured regions in 2-D images [6], and Harris corners in 2-D images [7, 8]. 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 schemes do exist. Notable examples include Lowe's scale invariant keypoints [9], and a visual homing method using scale invariant features based on the Fourier-Mellin transform [10]. However, it is currently unclear how complex these schemes might be for implementation in the neural hardware of an insect. The descriptor presented here is partially invariant to scale changes and has a direct and simple neural implementation.

The first of our two homing methods operates in a manner quite similar to that described above in that it searches for correspondences between descriptors in the snapshot image and descriptors in the currently viewed image. However, the second method takes advantage of the structure of the motion field for pure translation to avoid this search process. This method only pairs descriptors at the same image position. Very similar pairs ideally correspond to one of two stationary points in the motion field, known as the focus of contraction and focus of expansion. Finding either of these foci is equivalent to solving the visual homing problem.

An alternate approach to visual homing is Franz et. al's *warping method* [11]. 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 reports in [12, 13]) we use it here for comparison with our methods.

The images used in this paper are panoramic and were taken from a robot equipped with a panoramic imaging system. The results we present were obtained on a database of images collected within an office environment. We compare the performance of our two methods with the warping method on these images. Note that we make the assumption that all images were captured at the same compass orientation. A robot homing by one of our methods would require a compass to allow the differing orientation of images to be corrected. The warping method does not share this requirement. However, it has been found that the warping method performs better when it can be assumed that all images are taken from the same orientation [14].

In the next section we define a model of image scaling which is employed in the subsequent section on the development of our scale invariant image descriptor. We then present the two homing methods based on this descriptor. Next is a results section which shows the performance of these two homing methods and the warping method on a database of panoramic images. This is followed by a discussion section. The main content of the chapter ends with concluding remarks and references. An appendix includes a derivation of one of the principles underlying the image descriptor.

2   Scaling Model

We define a model of image scaling applicable to a local image patch. Let **p** be the coordinates of the centre of an image patch. The effect of scaling is to change the distance of image features to **p** by a factor k. Features nearby to **p** will shift by a smaller amount then distant features, yet the same scaling factor of k is applied for all image features. Hence, we refer to this as linear scaling.

Assume we have an image I which has been subject to linear scaling about point **p** by factor k. A point **a** in the original image I now corresponds to a point **a**$'$ in the scaled image I'. That is, a pixel in the original image at **a** will have the same value as a pixel in the scaled image at **a**$'$.

$$I(\mathbf{a}) = I'(\mathbf{a}') \tag{1}$$

Note that $I(\mathbf{a})$ is shorthand for the value of the pixel in image I with coordinates (a_x, a_y). Also, for simplicity we ignore pixel discretization and treat **a** as real-valued.

We now formulate an expression for **a** which involves the centre of scaling **p**. The following parametric equation of a line represents **a** with respect to its distance l from **p**, and with respect to the direction from **p** to **a** indicated by the unit vector **v**.

$$\mathbf{a} = \mathbf{p} + l\mathbf{v} \tag{2}$$

$$\mathbf{v} = \frac{\mathbf{a} - \mathbf{p}}{||\mathbf{a} - \mathbf{p}||} \tag{3}$$

The point **a**$'$ corresponding to **a** after linear scaling is similarly represented.

$$\mathbf{a}' = \mathbf{p} + kl\mathbf{v} \tag{4}$$

Note that this scaling model assumes the scaling factor k to be constant across the whole image. This is generally not true for the panoramic images employed here. However, linear scaling is a reasonable model for the scaling that occurs within local image patches of a panoramic image.

3   The Scale Invariant Descriptor

In this section we develop a local image descriptor which is partially invariant to scale changes. Figure 1 shows an image I and two increasingly scaled variants I' and I''. The figure also plots the value of each image along the ray $\mathbf{p} + l\mathbf{v}$ where $l > 0$ and **v** is arbitrarily set on a diagonal. We refer to this ray as a *channel*. The house image consist only of edges so the plots show isolated pulses where the channel crosses an edge.

It can be observed that while the positions of edge pulses along the channel have changed between I and I', the same two pulses are still found. Hence, the
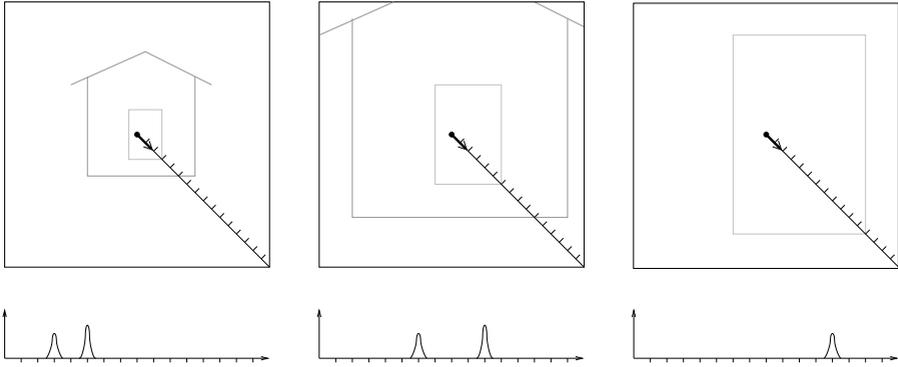
Fig. 1. Scaling of an image I and the value of I along a channel. Three images are shown with scale doubling incrementally from left to right. Beneath each image is a plot of the image values along the indicated channel. The images consist only of edges with darker edges having a higher value than lighter edges

area underneath these two pulses is the same. This observation prompts our first proposal for an invariant measure, which is the sum of image values along the channel

$$f^{\mathbf{p},\mathbf{v},I} = \int_0^{l_{\max}} I(\mathbf{p} + l\mathbf{v})\, dl \tag{5}$$

If indeed the same pulses are found along the same channel of I and I' then the following is true

$$f^{\mathbf{p},\mathbf{v},I'} = f^{\mathbf{p},\mathbf{v},I} \tag{6}$$

However, if the scaling factor k is too large then this condition will not hold. For example, in image I'' of figure 1 the outside edge of the house has been scaled entirely out of the frame. The channel now shows only a single pulse. Thus, $f^{\mathbf{p},\mathbf{v},I} \neq f^{\mathbf{p},\mathbf{v},I''}$. The same problem occurs for contraction ($k < 1$). If I'' was the original image and had been scaled down to I, the pulse representing the outside of the house would have appeared—and again $f^{\mathbf{p},\mathbf{v},I} \neq f^{\mathbf{p},\mathbf{v},I''}$. To mitigate the problem of appearance/disappearance of image features we propose a new invariant measure which includes a decay function

$$g^{\mathbf{p},\mathbf{v},I} = \int_0^{l_{\max}} w(l) I(\mathbf{p} + l\mathbf{v})\, dl \tag{7}$$

The purpose of the decay function $w()$ is to reduce the impact of outlying features on g. The appendix includes a derivation which places some constraints on $w()$. One obvious function which satisfies these constraints is

$$w(l) = \frac{1}{l^\zeta} \tag{8}$$

where $\zeta < 1$.

The objective now is to determine the relationship between $g^{\mathbf{p},\mathbf{v},I}$ and $g^{\mathbf{p},\mathbf{v},I'}$. This relationship is explored in the appendix and found to be as follows:

$$g^{\mathbf{p},\mathbf{v},I'} \approx kw(k)g^{\mathbf{p},\mathbf{v},I} \tag{9}$$

The presence of the factor $kw(k)$ implies that g is not scale invariant. We will deal with this problem momentarily. More fundamentally, however, is the fact that a scalar quantity such as $g^{\mathbf{p},\mathbf{v},I}$ is likely insufficient to describe a local image patch robustly. A richer descriptor is required to allow image patches to be disambiguated. We obtain such a descriptor by forming a vector $\mathbf{g}$ of g values computed from the same point $\mathbf{p}$ but at different directions

$$\mathbf{g}^{\mathbf{p},I} = \begin{pmatrix} g^{\mathbf{p},\mathbf{v}_0,I} \\ g^{\mathbf{p},\mathbf{v}_1,I} \\ \vdots \\ g^{\mathbf{p},\mathbf{v}_n,I} \end{pmatrix} \tag{10}$$

The length of the vector $\mathbf{g}$ is n. An obvious choice for the channel direction vectors $\mathbf{v}_i$ is to arrange them evenly in a radial pattern. For example, if $n = 4$ we would choose left, up, right, and down. If $n = 8$ we would add the four diagonals as well.

For the first algorithm presented below we will not be concerned with the length of $\mathbf{g}$, but only its direction. Therefore we define a normalized vector $\mathbf{h}$

$$\mathbf{h}^{\mathbf{p},I} = \frac{\mathbf{g}^{\mathbf{p},I}}{||\mathbf{g}^{\mathbf{p},I}||} \tag{11}$$

By normalizing we remove the factor $kw(k)$, hence

$$\mathbf{h}^{\mathbf{p},I'} \approx \mathbf{h}^{\mathbf{p},I} \tag{12}$$

and we can say that $\mathbf{h}$ is a scale invariant image descriptor. For the second algorithm it will be necessary to know whether k is greater or less than one. Thus, in the description for this algorithm we will also make reference to $\mathbf{g}$.

3.1   Conditions

The image descriptor $\mathbf{h}$ is invariant to scale changes given the following qualitative conditions:

1. The scale factor k is neither too great nor too small. The decay function can offset the impact of edge pulses being scaled in and out of range, but the scaling of outlying edge pulses will still generally distort the direction of $\mathbf{h}$.
2. If image edges are particularly dense then the edge pulses along a channel may interfere with each other in the summation of equation (7). Thus, it is advantageous for image edges to be relatively sparse.
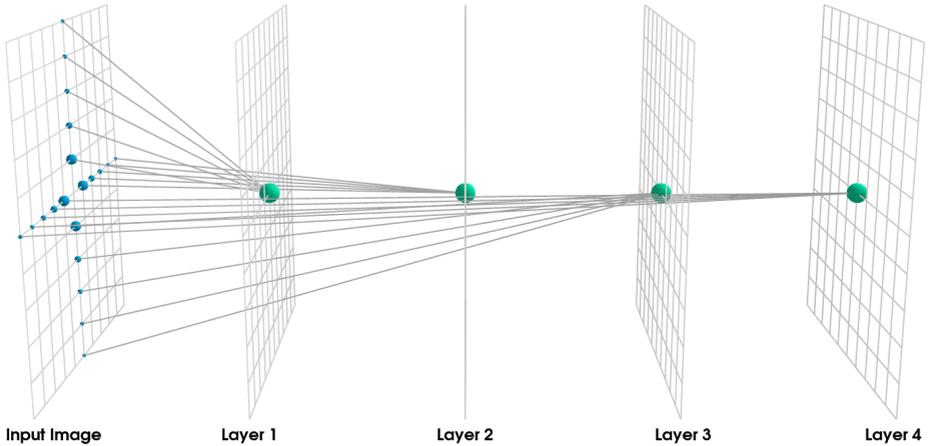
Fig. 2. Structure to compute **g**. This descriptor has $n = 4$ and $l_{\max} = 5$. The grid on the far left represents the input image. Large spheres indicate elements of the descriptor which sum their weighted inputs. Small spheres show connections to the input image. The radii of the small spheres are proportional to weights given by the decay function in equation (8)

3.2   Structure

The computation of **g** involves only the weighted summation of input image values. Figure 2 illustrates the structure that would be required to compute **g**.

This structure is repeated across all image positions that we wish to characterize. Such repetition of structure is similar to the retinotopic arrangement of columns of neurons in the visual systems of insects such as the honeybee [15, 16] and vertebrates such as cats [17]. Further, the computation for **g** consists only of local weighted sums. This style of processing is characteristic of artificial neural networks and is generally believed to be within the space of the processing operations that biological neural networks are capable of. Thus, while our image descriptor is not a model of any known neural structure in the animal kingdom, it is at least *plausible* that this descriptor could be implemented in an animal's brain.

4   1:N Matching Method

We present here the first of two homing methods which use the image descriptor developed above. This method is based on matching each descriptor in the snapshot image to N descriptors in the current image at neighbouring image positions. The coordinates of the best matches are then used to generate correspondence vectors. These correspondence vectors are then mapped to home vectors using the method described in [14]. The average home vector is the final output from this method.

We refer to the positions of descriptors in the snapshot image S as *source positions*. Each source position is matched with descriptors in the current image

C at N *candidate positions*. These candidate positions are located in a block surrounding the source position.

For each source position $\mathbf{p}$ in S we search to find the candidate position $\mathbf{p}'$ in C which is at the centre of the image patch most similar to the image patch at $\mathbf{p}$. To judge the degree of match between these two image patches we compute the scale invariant descriptors $\mathbf{h}^{\mathbf{p},S}$ and $\mathbf{h}^{\mathbf{p}',C}$ and find the dot product between them:

$$\mathrm{DP}(\mathbf{p}, \mathbf{p}') = \mathbf{h}^{\mathbf{p},S} \cdot \mathbf{h}^{\mathbf{p}',C} \tag{13}$$

A high value of DP indicates a good match.

To reduce computational complexity we do not consider all positions in S as source positions, but only a sampling of positions at integer multiples of the horizontal step size m_x and the vertical step size m_y, where m_x and m_y are also integers. Given images of width w pixels and height h pixels we define the number of horizontal and vertical sampling points

$$n_x = \lfloor w/m_x \rfloor \tag{14}$$
$$n_y = \lfloor h/m_y \rfloor \tag{15}$$

The total number of source positions is $n_x n_y$. Each source position requires a search for the best candidate position. This search involves computing DP for N candidate positions. The candidate positions are located within a radius of q pixels from the source position $\mathbf{p}$. Hence, $N = (2q+1)^2$.

We select $\check{\mathbf{p}}$ as the candidate position with the highest DP:

$$\check{\mathbf{p}} = \arg \max_{\mathbf{p}' \in E_q(\mathbf{p})} \mathrm{DP}(\mathbf{p}, \mathbf{p}') \tag{16}$$

$$E_q([p_x, p_y]) = \{(p_x + i, p_y + j) \mid i, j \in \mathbb{Z}, |i| \le q \wedge |j| \le q\} \tag{17}$$

There is an additional constraint made on the correspondence search whereby source positions in the snapshot image will only be paired with candidate positions which are on the same side of the horizon. The horizon of the panoramic image is the line which does not undergo vertical translations under movements of the robot in the plane. As long as the robot moves purely within a single plane, no image features should cross the horizon. Therefore we constrain our search to avoid any such spurious matches.

The candidate position $\check{\mathbf{p}}$ with the highest DP is used to compute the correspondence vector $\boldsymbol{\delta}$

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = \begin{pmatrix} \Delta_x(\check{p}_x - p_x) \\ \Delta_y(\check{p}_y - p_y) \end{pmatrix} \tag{18}$$

where Δ_x represents the inter-pixel angle in the horizontal direction and Δ_y represents vertical inter-pixel angle. These multipliers are required so that $\boldsymbol{\delta}$ is expressed as a pair of angles.

We now have a set of correspondence vectors which ideally describe the movement of image features in S to their new positions in C. From each of these correspondence vectors we can determine an individual home vector. We use the 'vector mapping' method presented in [14] for this purpose. Finally, the average of these home vectors is computed, normalized, and used as the final home vector.

5   1:1 Pairing Method

We now present our second homing method based on the scale invariant image descriptor. While the method above repeatedly searches for correspondences between source positions in S and candidate positions in C, the method below considers only one candidate position for each source position. Only positions at the same image position are compared and the best matching pair is used to compute a final home vector directly.

In general, a landmark seen in the snapshot image will either move to a new position in the current image, or will disappear. However, there is an important exception to this rule. Landmarks at the focus of contraction (FOC) or focus of expansion (FOE) will maintain the same image position if the displacement of the agent from the goal consists of pure translation. For pure non-zero translation the flow field (field of correspondence vectors) exhibits two foci separated by 180°. We assume here that the world is imaged onto the unit sphere, hence both foci are always visible. All correspondence vectors are parallel to great circles passing through the foci. Correspondence vectors are oriented from the FOE to the focus of contraction FOC [1]. Figure 3 shows an ideal flow field for an agent within a simulated environment where all surfaces are equidistant from the agent. It can be observed that the amplitude of flow (the length of correspondence vectors) approaches zero at the foci.
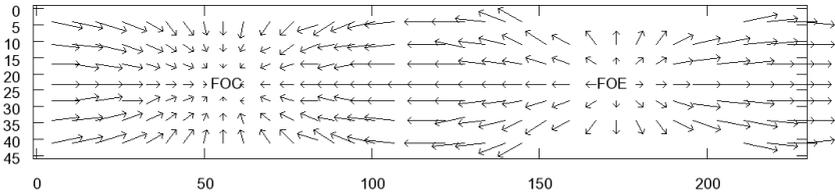

Fig. 3. An ideal flow field for pure translation. Vectors were generated by tracking the displacement of unique markers on the surface of a sphere, where the sphere was centred on the agent for the snapshot image and then shifted to the right for the current image

The 1:1 method computes descriptors $\mathbf{g}$ and $\mathbf{h}$ for positions along the horizon of the snapshot image S and the current image C. Descriptors at the same image position in both images are then compared by computing the dot product between them. The coordinates $\hat{\mathbf{p}}$ of the pairing with the highest DP value is determined

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \mathrm{DP}(\mathbf{p}, \mathbf{p}) \tag{19}$$

If our descriptor truly provides a unique means of characterizing local image patches then $\hat{\mathbf{p}}$ represents a local image patch that is stationary between the

---

[1] See [18] for a more thorough discussion.

snapshot image and current image. Such an image patch could either represent a very distant landmark, or else it could represent one of the foci. Here we assume the latter. In the experiments described below the range of distances to objects remains rather small. However, if distant landmarks were present then some sort of filtering scheme might be employed to remove them from the image [19].

We determine which of the two foci $\hat{\mathbf{p}}$ corresponds to by comparing the length of the vector $\mathbf{g}^{\hat{\mathbf{p}},S}$ with $\mathbf{g}^{\hat{\mathbf{p}},C}$. Growth of the descriptor vector $\mathbf{g}$ from the snapshot to the current image occurs in the neighbourhood of the FOC. By definition, image features move in towards the FOC, and as they do, become weighted more heavily by the decay function $w()$. The opposite situation occurs at the FOE where features become weighted less heavily as they expand away from the FOE. The quantity b equals 1 for the case of contraction and -1 for expansion

$$b = \begin{cases} 1 & \text{if } ||\mathbf{g}^{\hat{\mathbf{p}},S}|| < ||\mathbf{g}^{\hat{\mathbf{p}},C}|| \\ -1 & \text{if } ||\mathbf{g}^{\hat{\mathbf{p}},S}|| > ||\mathbf{g}^{\hat{\mathbf{p}},C}|| \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Finally, the computed home vector is given by converting the image coordinate $\hat{\mathbf{p}}$ into a vector and using b to reverse that vector if appropriate

$$\mathbf{w} = b \begin{pmatrix} \cos(\Delta_x \hat{p}_x) \\ \sin(\Delta_y \hat{p}_y) \end{pmatrix} \quad (21)$$

The vector $\mathbf{w}$ above is the final estimated home vector.

6   Results

6.1   Image Database

A database of images was collected in the robotics laboratory of the Computer Engineering Group of Bielefeld University. Images were collected by a camera mounted on the robot and pointed upwards at a hyperbolic mirror[2]. The room was unmodified except to clear the floor. The capture grid had dimensions 2.7 m by 4.8 m, which covered nearly all of the floor's free space. Further details on the collection and format of these images has been reported in [14].

The images used for homing are low-resolution (206×46) panoramic images. Figure 4 shows sample images along a line from position (6,4) to position (0,4).

6.2   Methods

Both homing methods require edges to be extracted from input images. A Sobel filter is applied for this purpose. Parameters described below control the level of low-pass filtering applied prior to the Sobel filter.

Some parameters of the homing methods are method-specific while others are shared by both methods. The method-specific parameters were set to values

---

[2] The camera was an ImagingSource DFK 4303. The robot was an ActivMedia Pioneer 3-DX. The mirror was a large wide-view hyperbolic mirror from Accowle Ltd.
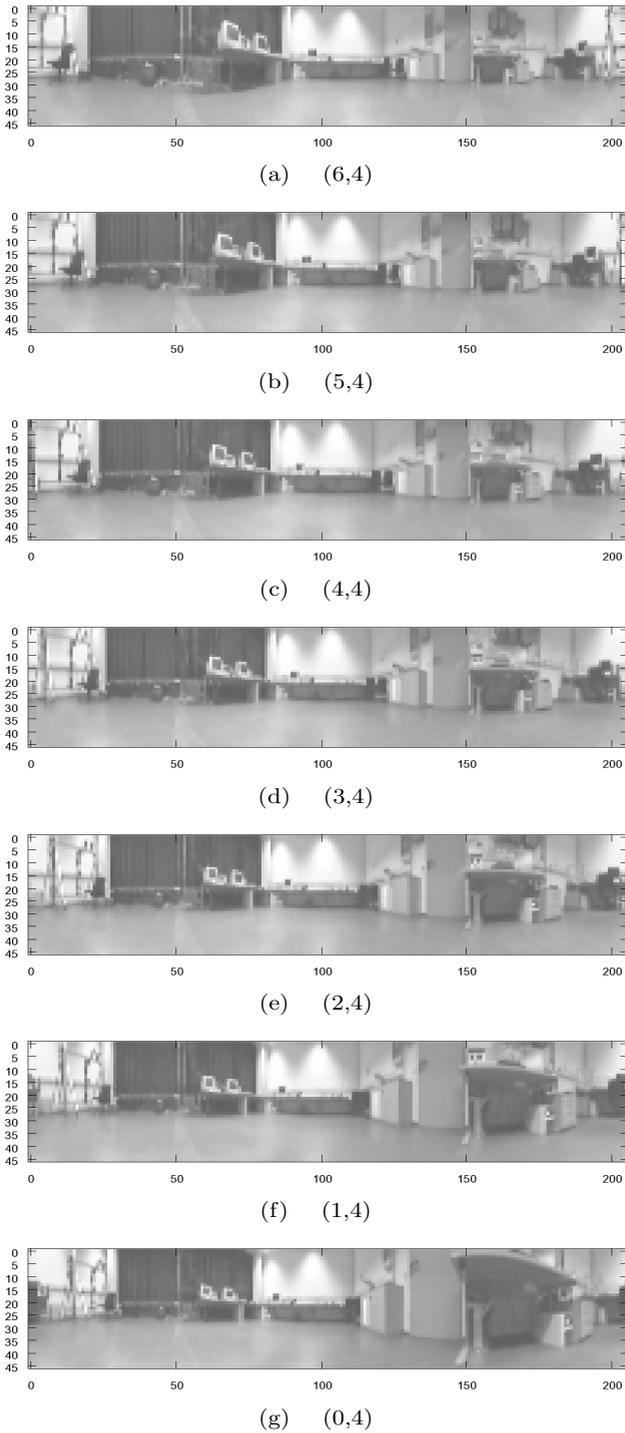
(a)   (6,4)

(b)   (5,4)

(c)   (4,4)

(d)   (3,4)

(e)   (2,4)

(f)   (1,4)

(g)   (0,4)

Fig. 4. Sample images from image database along a line of positions from (6,4) to (0,4)

which generally appeared to provide good results. For the 1:N matching method these included the search radius q (set to 30), the horizontal step size m_x (4), and the vertical step size m_y (4). Another parameter excluded points in the specified number of image rows at the top and bottom of the image from being used as source points. This parameter (set to 10) was introduced upon observing that image patches in the top and bottom portions of the image tended to be relatively indistinct. For the 1:1 pairing method the only method-specific parameter is the height of the window around the horizon (9).

One important shared parameter is the exponent of the decay function, ζ. This parameter was set to 0.75 which appeared to work well for both methods. For the remaining shared parameters a search was carried out to find the best settings for each homing method. This search scored parameter combinations according to the average angular error over 20 snapshot positions, as described below. Four parameters were varied in this search process:

- The length of descriptor vectors, n, was set to either 8 or 32 for the 1:N matching method and 8, 32, or 64 for the 1:1 pairing method.
- The length of channels to sum over, $l_{\max}$, was set to either 20 or 50.
- Prior to edge extraction, the input images are smoothed by a Gaussian operator [3]. The number of applications of this operator was set to either 0 or 4.
- As described in section 3.1, it is not advantageous for image edges to be excessively dense. The density of edges can be reduced by passing the image through a power filter, which raises each pixel's value to exponent τ. τ was set to either 1, 2, or 4.

The best found shared parameters for the 1:N matching method were: $n = 32$, $l_{\max} = 50$, 0 Gaussian operations, and $\tau = 4$. The best shared parameters for the 1:1 pairing method were: $n = 64$, $l_{\max} = 50$, 4 Gaussian operations, and $\tau = 4$.

Franz et. al's warping method was also tested for comparison [11]. Parameters for this method were found using a parameter search similar to that described above. Further details can be found in [14].

Before continuing, it is interesting to examine some of the internal workings of our two homing methods. We begin by examining the correspondence vectors generated by the 1:N matching method. Figure 5 shows these vectors as computed for the images shown in figure 4 with the goal position at (6,4). The flow fields generally appear correct (compare with figure 3 which shows the ideal flow for the same movement—albeit within a different environment). However, there are a number of clearly incorrect vectors embedded within these flow fields.

For the 1:1 pairing method we look at the variation in $DP(\mathbf{p}, \mathbf{p})$. This quantity should show peaks at the FOC and FOE. Figure 6 shows $DP(\mathbf{p}, \mathbf{p})$ for the

---

[3] The Gaussian operator convolves the image by the kernel

$$[0.005 \ 0.061 \ 0.242 \ 0.383 \ 0.242 \ 0.061 \ 0.005]$$

applied separately in the x and y directions.

(a)   (5,4)

(b)   (4,4)
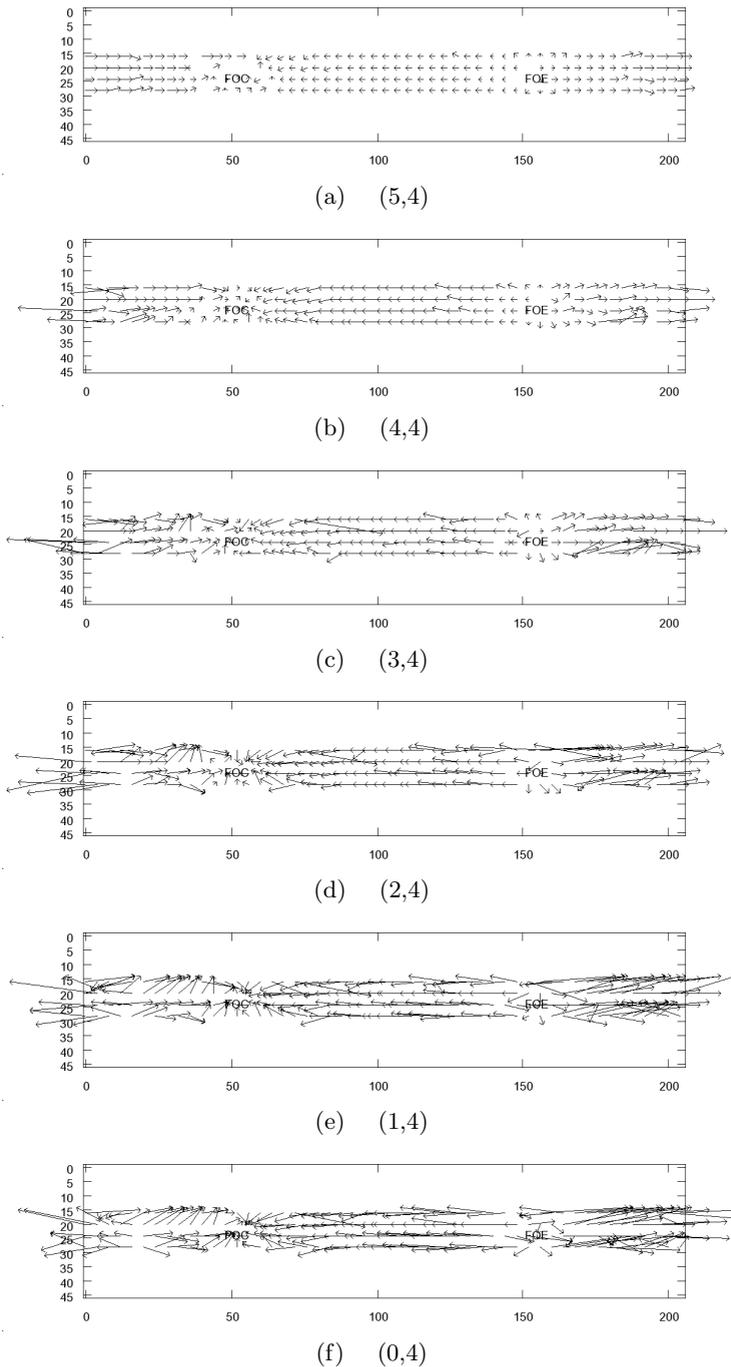
(c)   (3,4)

(d)   (2,4)

(e)   (1,4)

(f)   (0,4)

Fig. 5. Correspondence vectors for the 1:N matching method. The snapshot image was captured at (6,4), which is to the right of the positions indicated above. Hence, the correct FOC should be around (52,23) while the correct FOE should be near (155,23)
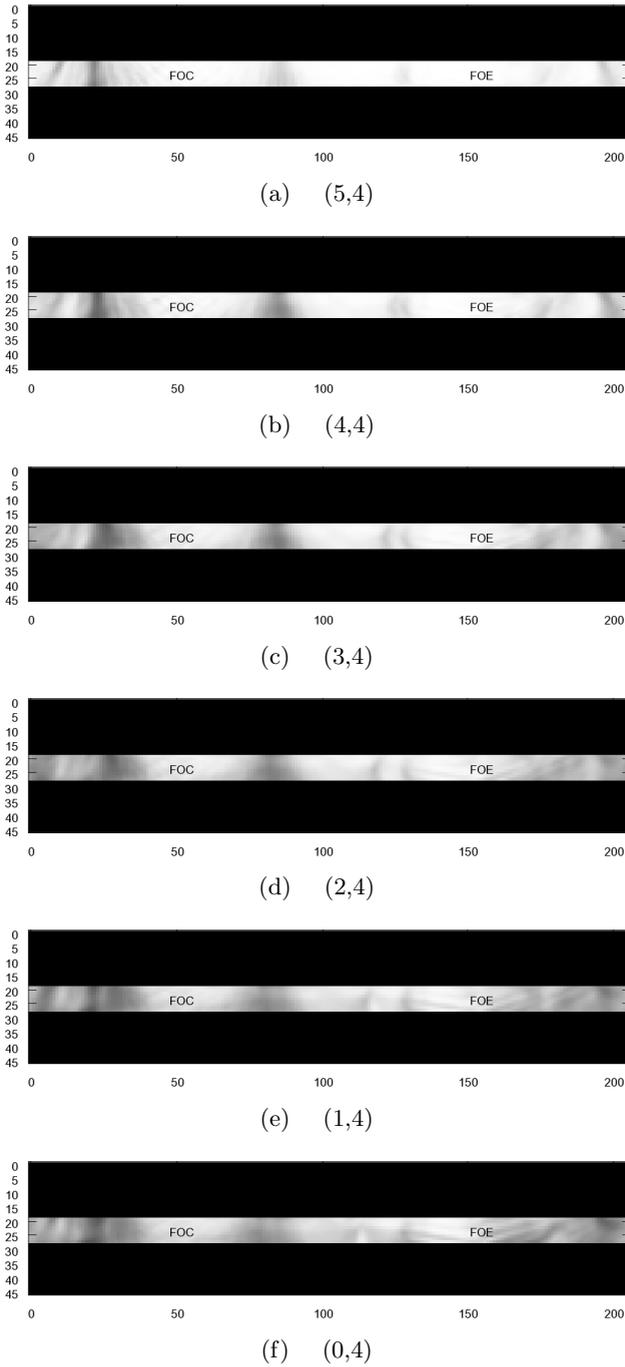
(a)   (5,4)

(b)   (4,4)

(c)   (3,4)

(d)   (2,4)

(e)   (1,4)

(f)   (0,4)

Fig. 6. Variation in $DP(\mathbf{p}, \mathbf{p})$ for the 1:1 pairing method. Snapshot position and labels are as in figure 5

images shown in figure 4 with the goal again at (6,4). Indeed, two major peaks near the ideal locations of the FOC and FOE are found.

In the first set of experiments, a single image is selected from the capture grid as the snapshot image. The method in question then computes home vectors for all other images. Figure 7 shows the home vectors generated by our two methods and the warping method for snapshot positions (6,4) and (0,16).

Both the 1:N matching method and the warping method perform quite well at position (6,4). It is evident for both of these methods that the home vectors for all positions would tend to lead the robot to the goal from all start positions, although the paths taken by the warping method would be somewhat shorter. For the 1:1 matching method, however, there are a number of incorrect vectors embedded within the otherwise correct vector field. At position (0,16) it is apparent that the 1:N matching methods yields the best results of the three methods. The 1:1 pairing method exhibits appropriate home vectors for some positions, but also generates vectors which are directed 180° away from the correct direction, as well as others which point in a variety of incorrect directions. The warping method generates appropriate vectors only within a small neighbourhood around the goal position.

For a more qualitative determination of the success of homing we compute the *average angular error* (AAE) which is the average angular deviation of the computed home vector from the true home vector. We indicate the average angular error for snapshot position (x, y) as $AAE_{(x,y)}$. Values for AAE are shown in the captions for figure 7. These values generally reflect the qualitative discussion above.

It is clear from figure 7 that homing performance is dependent on the chosen snapshot position. To assess this dependence we tested all homing methods on a sampling of 20 snapshot positions and computed AAE for each position. Figure 8 shows these snapshot positions which were chosen to evenly sample the capture grid. Figure 9 shows the computed AAE for all methods over these 20 snapshot positions. All methods exhibit higher error for snapshot positions near the fringes of the capture grid. The captions in this figure show the angular error averaged over all test positions, and all snapshot positions.

To obtain a more quantitative understanding of the difference between these methods we performed statistical tests on AAE_*. A repeated measures ANOVA with Tukey-Kramer multiple comparisons test was carried out between all three methods. Table 1 presents the results of this test. The test indicates that the 1:N matching method exhibits a significantly lower error than the warping method. No significant difference was found between the error of 1:N matching and 1:1 pairing. Nor was a significant difference found between the error of 1:1 pairing and the warping method.

While it is interesting to compare the performance of these homing methods against each other, it is useful also to compare them to an absolute standard. As described in [11], a homing method with an angular error that is always less than $\pi/2$ will yield homing paths that converge to the goal—perhaps taking a very inefficient route, but arriving eventually. Having an average error below $\pi/2$ does
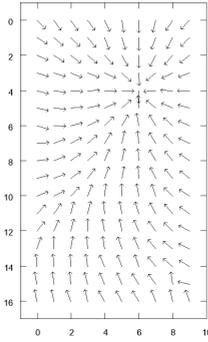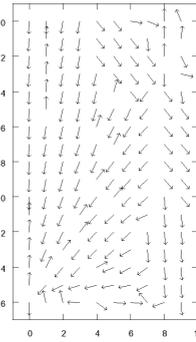
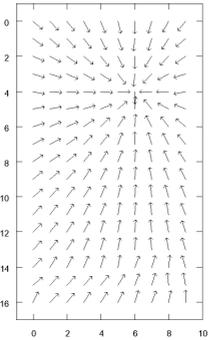(a) 1:N Matching: $AAE_{(6,4)}$ = 0.256
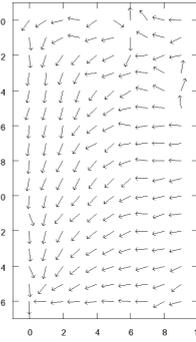
(b) 1:N Matching: $AAE_{(0,16)}$ = 0.488
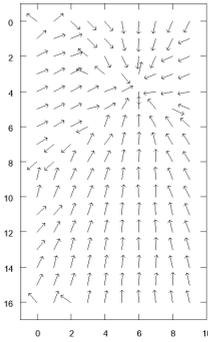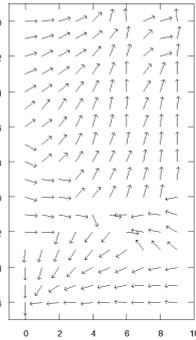

(c) 1:1 Pairing: $AAE_{(6,4)} = 0.368$

(d) 1:1 Pairing: $AAE_{(0,16)} = 0.888$

(e) Warping: $AAE_{(6,4)} = 0.114$

(f) Warping: $AAE_{(0,16)} = 1.793$

Fig. 7. Home vector fields for 1:N matching (a,b), 1:1 pairing (c,d), and warping (e,f), for snapshot positions (6,4) (a,c,e) and (0,16) (b,d,f)
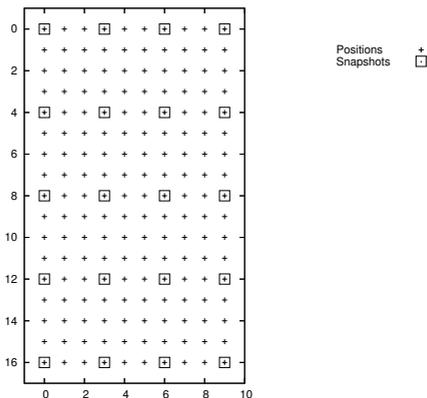
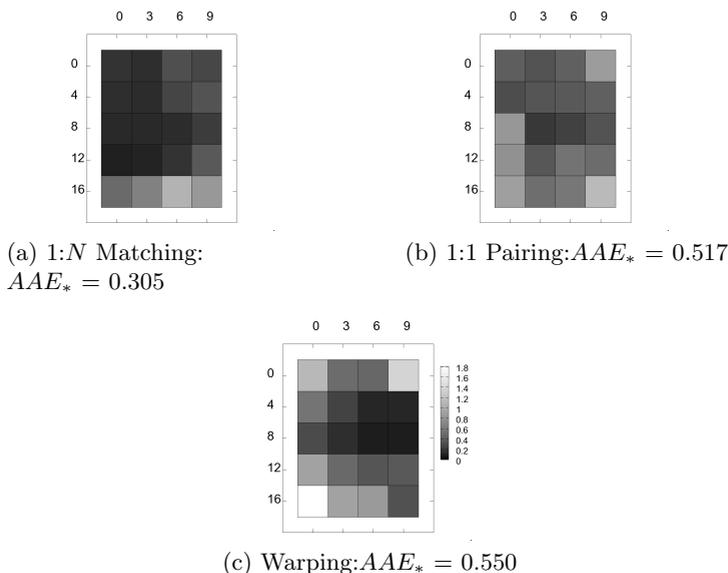**Fig. 8.** Positions of images and snapshots within the capture grid

(a) 1:N Matching: $AAE_* = 0.305$

(b) 1:1 Pairing:$AAE_* = 0.517$

(c) Warping:$AAE_* = 0.550$

Fig. 9. AAE for the twenty snapshot positions shown in figure 8 for all methods on image collection `original`

not imply convergent homing but it is a useful threshold. We have performed a statistical analysis of the difference between the angular error and $\pi/2$ using the Wilcoxon rank sum test. Table 2 presents the results of this test for $\pi/2$, and also for increasingly small angles $\pi/4$, $\pi/6$, and $\pi/8$. The test indicates whether each method exhibits an AAE smaller than the threshold. All three methods exhibit error significantly less than both $\pi/2$ and $\pi/4$. However, only the 1:N matching method exhibits an error significantly less than $\pi/6$.

Table 1. Statistical significance of the difference in AAE_* between homing methods. Significance for each cell is indicated if the method on the vertical axis is significantly better than the method on the horizontal axis. Empty fields indicate no significant difference. *Legend:* * = ($p < 0.05$), ** = ($p < 0.01$), *** = ($p < 0.001$), **** = ($p < 0.0001$), X = self-match. *Test:* Repeated measures ANOVA with Tukey-Kramer multiple comparisons

	1:N Matching	1:1 Pairing	Warping
1:N Matching	X		*
1:1 Pairing		X	
Warping			X

Table 2. Statistical significance of AAE_* being less than the reference angles $\pi/2$, $\pi/4$, $\pi/6$, and $\pi/8$. Significance for each cell is indicated if the method on the vertical axis has an angular error significantly less than the threshold on the horizontal axis. See table 1 for legend. *Test:* Wilcoxon rank sum test

	$\pi/2$	$\pi/4$	$\pi/6$	$\pi/8$
1:N Matching	****	****	**	
1:1 Pairing	****	***		
Warping	****	*		

7   Discussion

Of the three homing methods tested above, the 1:N matching method exhibits the lowest error and overall best results. According to our statistical tests the 1:1 pairing method performs equivalently to the warping method. The pairing method is of interest because it does not require repeated searching to find correspondences between images. In theory, the computational complexity of the 1:1 pairing method should be considerably lower than the 1:N matching method. However, the pairing method appears to be less robust to parameter settings and requires the most expensive parameters (high n, and $l_{\max}$) in order to perform well.

8   Conclusions

This chapter introduced a new descriptor for local image patches which is partially invariant to scale changes. The descriptor has a simple structure that is suitable for neural implementation. Two homing methods based on this descriptor were presented. The first method employed the standard technique of matching descriptors between images. The second method, however, employed the novel notion of extracting one of the foci of motion, and using the position of that focus to compute the home vector directly. The performance of the 1:N matching method was found to exceed that of the warping method. No statistically significant difference was found between the 1:1 pairing method and the warping method. Future work will look at improvements to our descriptor as well as possibilities for using other scale invariant descriptors for the 1:1 pairing method.

Acknowledgments

References

1. Collett, T.: Insect navigation en route to the goal: Multiple strategies for the use of landmarks. Journal of Experimental Biology **199** (1996) 227–235
2. 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
3. Möller, R.: Insect visual homing strategies in a robot with analog processing. Biological Cybernetics **83** (2000) 231–243
4. Cartwright, B., Collett, T.: Landmark learning in bees. Journal of Comparative Physiology A **151** (1983) 521–543
5. 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
6. 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: Biologically Inspired Robotics **17** (2002) 154–162
7. 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
8. 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)
9. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
10. 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
11. 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
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. Vardy, A., Möller, R.: Biologically plausible visual homing methods based on optical flow techniques. Connection Science, Special Issue: Navigation (2005 (to appear))
15. Hertel, H.: Processing of visual information in the honeybee brain. In Menzel, R., Mercer, A., eds.: Neurobiology and Behavior of Honeybees. Springer Verlag (1987) 141–157

16. Ribi, W.: The structural basis of information processing in the visual system of
    the bee. In Menzel, R., Mercer, A., eds.: Neurobiology and Behavior of Honeybees.
    Springer Verlag (1987) 130–140
17. Hubel, D., Wiesel, T.: Receptive fields, binocular interaction and functional archi-
    tecture in the cat's visual cortex. Journal of Physiology **160** (1962) 106–154
18. Nelson, R., Aloimonons, A.: Finding motion parameters from spherical motion
    fields (or the advantage of having eyes in the back of your head). Biological Cy-
    bernetics **58** (1988) 261–273
19. Cartwright, B., Collett, T.: Landmark maps for honeybees. Biological Cybernetics
    57 (1987) 85–93

Appendix

The goal of this appendix is to determine the relationship of $g^{\mathbf{p},\mathbf{v},I'}$ to $g^{\mathbf{p},\mathbf{v},I}$.
We begin with $g^{\mathbf{p},\mathbf{v},I'}$

$$g^{\mathbf{p},\mathbf{v},I'} = \int_0^{l_{\max}} w(l)I'(\mathbf{p}+l\mathbf{v})\,dl \tag{22}$$

From equations (1), (2), and (4) we obtain

$$I(\mathbf{p}+l\mathbf{v}) = I'(\mathbf{p}+kl\mathbf{v}) \tag{23}$$

With a change of variables we have the following:

$$I(\mathbf{p}+\frac{l}{k}\mathbf{v}) = I'(\mathbf{p}+l\mathbf{v}) \tag{24}$$

We insert the above into the right hand side of equation (22) to obtain

$$g^{\mathbf{p},\mathbf{v},I'} = \int_0^{l_{\max}} w(l)I(\mathbf{p}+\frac{l}{k}\mathbf{v})\,dl \tag{25}$$

Next we replace the integration variable l with $j = \frac{l}{k}$

$$g^{\mathbf{p},\mathbf{v},I'} = \int_0^{l_{\max}/k} w(jk)I(\mathbf{p}+j\mathbf{v})k\,dj \tag{26}$$

Now we place our first assumption on $w()$. We assume this function has the
property

$$w(xy) = w(x)w(y) \tag{27}$$

Utilizing this property on expression (26) and renaming the integration variable
j back to l gives

$$g^{\mathbf{p},\mathbf{v},I'} = kw(k)\int_0^{l_{\max}/k} w(l)I(\mathbf{p}+l\mathbf{v})\,dl \tag{28}$$

To proceed further we must place another constraint on $w()$. The intention of
this decay function is to reduce the impact of outlying features on g. Therefore it

makes sense that $w(l)$ should be small for large values of l. We first define a new constant $l^*_{\max} = \min(l_{\max}, l_{\max}/k)$. The second constraint on $w()$ is as follows:

$$w(l) \approx 0 \quad \text{for } l > l^*_{\max} \tag{29}$$

Therefore

$$g^{\mathbf{p},\mathbf{v},I} \approx \int_0^{l^*_{\max}} w(l)I(\mathbf{p} + l\mathbf{v})\, dl \tag{30}$$

and

$$g^{\mathbf{p},\mathbf{v},I'} \approx kw(k) \int_0^{l^*_{\max}} w(l)I(\mathbf{p} + l\mathbf{v})\, dl \tag{31}$$

Combining these two approximations gives us the desired relationship

$$g^{\mathbf{p},\mathbf{v},I'} \approx kw(k)g^{\mathbf{p},\mathbf{v},I} \tag{32}$$