ELSEVIER

Pattern Recognition Letters 22 (2001) 1371–1378

Pattern Recognition Letters

www.elsevier.com/locate/patrec

A novel visual landmark matching for a biologically inspired homing

A. Rizzi *, D. Duina, S. Inelli, R. Cassinis

Department of Electronics for Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy

Received 26 June 2000

Abstract

This paper presents a homing algorithm for an autonomous robot that uses only visual information. An image grabbed at the target position is compared with the perceived one to determine the position of the robot and its target. Visual landmarks are extracted autonomously from the images and a correlation criterion, based on a novel visual landmark descriptor equalization, is presented. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Homing; Mellin transform; Visual landmark; Robot navigation

1. Introduction

The term *homing* indicates the navigation process by means of which an autonomous robot drives itself towards a precise location. The approach proposed in this paper derives from a biological homing model developed by Cartwright and Collett (1983, 1987). Without requiring any preconditioning of the environment, it estimates the robot and the target relative positions by comparing an image grabbed at the target position with the currently perceived one. The navigation is performed using exclusively the visual information grabbed at the two positions. Analyzing only the chromatic and geometric characteristics of the segmented images, stable chromatic areas, used as visual landmarks, are chosen autonomously.

The main difference between the Cartwright and Collett model and the implemented navigation system is that the camera used in this work cannot take omnidirectional images. To overcome this limitation, the robot learns and then approaches the homing point keeping its heading constant. This is not a critical limitation since different images of the same goal position, grabbed with different heading, can be taken into account by a higher level control module able to switch among them.

Using this approach, visual changes on the image plane can be described with a simplified affine model. This model is applied only on particular regions, automatically extracted from the images, called *Visual References* (VRs). By computing the translation and scale parameters of each VR in the different scenes it is possible to estimate the robot displacement in the environment.

The proposed method can be applied only for a final homing phase, where the system can find corresponding VRs in the two images. A higher

* Corresponding author. Tel.: +39-030-3715457; fax: +39-030-380014.

E-mail address: rizzi@ing.unibs.it (A. Rizzi).

PII: S0167-8655(01)00066-6

level navigation module is supposed to drive the robot until a valid VR can be found.

The most common approach used to compare the different VRs is the one based on the intercorrelation technique (Vanderburg and Rosenfeld, 1977; Pratt, 1978; Segman, 1992). Unfortunately, this approach produces a very wide maximum of the intercorrelation function. It is therefore difficult to estimate the maximum in presence of noise. Moreover, this technique is not invariant with respect to neither the scaling nor the landmark rotation and computing the intercorrelation function of every possible transformation is too expensive (Pratt, 1978; Goshtasby, 1985; Segman, 1992). Another algorithm that uses the phase of the image Fourier transform to implement more selective and robust filters is suggested in (Chen et al., 1994). An earlier work developed by our research group (Bianco et al., 1997; Rizzi et al., 1998) compared different VRs using three color parameters and four geometrical parameters, for each VR (Hu, 1962; Reiss, 1991).

In the research presented here, a comparison method based on the use of Fourier–Mellin transform (Oberhettinger, 1974), invariant with respect to translations and scale changes (Ravichandran and Trivedi, 1995), is proposed to carry out the VRs coupling. A VR descriptor containing its Fourier–Mellin transform, is computed for each VR. This step is performed using a gray-scale VR representation. A distance index based on the intercorrelation function is used to estimate the VRs descriptors coupling between different images. A new VR descriptor equalization has been devised to increase the intercorrelation selectivity. In a second phase, the relative scale and rotation factor obtained from the intercorrelation between two VR descriptors, is used to reinforce the displacement estimation, increasing the system robustness.

2. Computing the VR descriptors

Given a generic VR $r(x,y)$, the Fourier spectrum of its rotated, shifted, scaled version $s(x,y)$ is

$$|S(u,v)| = \sigma^{-2} \cdot \left|R[\sigma^{-1}(u \cdot \cos\alpha + v \cdot \sin\alpha), \sigma^{-1}(-u \cdot \sin\alpha + v \cdot \cos\alpha)]\right|, \quad (1)$$

where α is the rotation angle, σ the scaling factor, (x_0, y_0) the translation and $R(u,v)$ the $r(x,y)$ Fourier transform. As it can be noticed from Eq. (1), $|S(u,v)|$ is shifting invariant. Moreover, $r(x,y)$ rotation and scale can be separated defining R_p and S_p, the $r(x,y)$ and $s(x,y)$ Fourier spectrum in polar coordinates (θ,ρ):

$$S_p(\theta,\rho) = \sigma^{-2} \cdot R_p(\theta - \alpha, \rho/\sigma).$$

In this way rotating $s(x,y)$ is equivalent to shifting $S_p(\theta,\rho)$ along θ. Using a radial logarithmic scale $s(x,y)$ scaling can be mapped in the S_{pl} shifting as follows:

$$S_{pl}(\theta,\psi) = \sigma^{-2} \cdot R_{pl}(\theta - \alpha, \psi - \kappa),$$

where $\psi = \log(\rho)$ and $\kappa = \log(\sigma)$. Thus, $s(x,y)$ rotation and scaling correspond to $S_{pl}(\theta,\sigma)$ shifting. Summarizing, a transform along ρ and then a logarithmic remapping are equivalent to the Mellin transform along the same direction.

To obtain a VR descriptor two steps are performed: the computation of the VR DFT module and its transformation into polar–logarithmic coordinates. Since the DFT module is even, only the first half is taken into account. An example of VR is shown in Fig. 1(a) and the relative DFT module in polar–logarithmic coordinates is shown in Fig. 1(b).

In order to increase the intercorrelation performance a new amplitude equalization function, with a simple non-linear filter, is proposed. Fig. 1(c) shows the proposed equalization function and Fig. 1(d) shows the results of its application on Fig. 1(b).

3. VR descriptors coupling

To achieve the coupling between two descriptors from different images a coupling reliability index is used:

$$CR_{r,s} = \frac{W_r \cdot W_s - \max[\varphi_{rs}(\theta,\psi)]}{W_r \cdot W_s}, \quad (2)$$

where $\varphi_{rs}(\theta,\psi)$ is the intercorrelation function between the descriptors r and s, having energy W_r and W_s, respectively.

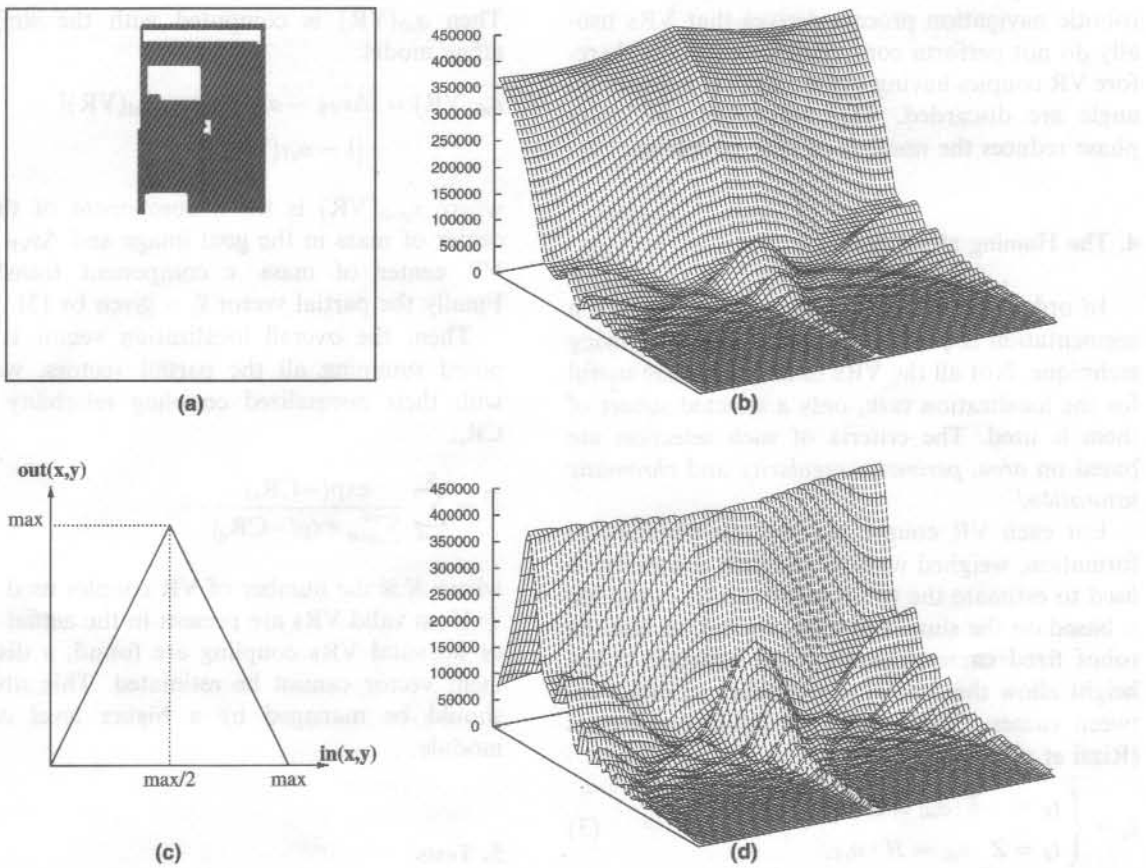

Fig. 1. (a) A VR example, (b) the VR's DFT in polar–logarithmic coordinates, (c) the proposed equalization function, and (d) the final VR's equalized descriptor.

Let $\mathscr{VRS}_{\text{goal}}$ and $\mathscr{VRS}_{\text{act}}$ be the VR sets of the goal and actual images, with m and n elements, respectively. A correlation matrix $\boldsymbol{C}_{m\times n}$ is used to contain the coupling reliability values of the $\mathscr{VRS}_{\text{goal}}$ and $\mathscr{VRS}_{\text{act}}$ elements. From $\boldsymbol{C}$, two boolean matrices are computed, the first $\boldsymbol{B}_{\text{goal,act}}$ links each VR in $\mathscr{VRS}_{\text{goal}}$ with the VRs in $\mathscr{VRS}_{\text{act}}$. Each link refers to the couple that for each element in $\mathscr{VRS}_{\text{goal}}$ maximizes CR with $\mathscr{VRS}_{\text{act}}$. Only the values under a given threshold are considered. The second matrix $\boldsymbol{B}_{\text{act,goal}}$ is computed conversely.

The final VR couples between the images are obtained looking for the index where both $\boldsymbol{B}_{\text{goal,act}}$ and $\boldsymbol{B}_{\text{act,goal}}$ have a link. If this does not happen a VR in $\mathscr{VRS}_{\text{goal}}$ is not coupled and will not affect the localization process.

Once the VR coupling is made it is possible to obtain the related rotation and scale factors. The maximum of the intercorrelation function of the VR descriptor along the radial axis gives the scale factor, while the VR rotation factor can be found from the position of the intercorrelation maximum along the phase axis. The resulting expressions are

$$S_{\text{VR}} = \exp\left(-\frac{y_{\max}}{n}\right), \qquad R_{\text{VR}} = \frac{x_{\max}}{n}\cdot 180^{\circ},$$

where S_{VR} is the scale parameter, R_{VR} the VR rotation parameter in degrees, $x_{\max}$ and $y_{\max}$ the position of the intercorrelation function maximum, and n is the image dimension in pixels.

With these additional parameters a new VR coupling validation phase can be added. The

robotic navigation process derives that VRs usually do not perform considerable rotations, therefore VR couples having a rotation over a threshold angle are discarded. This additional validation phase reduces the number of false couplings.

4. The Homing algorithm

In order to extract the VRs from the image, a segmentation is performed with a region growing technique. Not all the VRs in the image are useful for the localization task, only a selected subset of them is used. The criteria of such selection are based on *area*, *perimeter regularity* and *chromatic saturation*.

For each VR couple, the relative position information, weighed with its coupling reliability, is used to estimate the robot position. This estimate is based on the simplified affine model. In fact the robot fixed camera heading and constant robot height allow the following simplified relations between camera translation and affine parameters (Rizzi et al., 1998):

$$\vec{v}_i = \begin{cases} t_X = -\frac{Z}{f} \cdot a_{x0} = K \cdot a_{x0} \\ t_Z = Z \cdot a_{x1} = H \cdot a_{x1}, \end{cases} \tag{3}$$

where t_X and t_Z are the robot displacement components, Z is the VR distance and a_{x0}, a_{x1} are the translation and dilation/compression factors of the simplified affine model.

In the test presented in this paper H and K have been set with a tuning phase. However their fine tuning is not critical since they give only a displacement estimate proportionality factor. In fact, the biologically inspired approach of the proposed method uses iteratively a qualitative vector estimate instead of a precise camera calibration (Bianco et al., 1999).

To estimate the robot position a displacement vector $\vec{v}_i$ for each VRs couple is computed using the two parameters a_{x0} and a_{x1}, estimated from the affine model.

The scale parameter obtained from the VR descriptors intercorrelation function (3) can be used to reinforce the a_{x1} estimate:

$a_{x1}(\mathrm{VR}) = 1 - S_{\mathrm{VR}}$.

Then $a_{x0}(\mathrm{VR})$ is computed with the simplified affine model:

$$a_{x0}(\mathrm{VR}) = [\Delta x_{\mathrm{VR}} - a_{x1}(\mathrm{VR}) \cdot x_{\mathrm{goal}}(\mathrm{VR})] \cdot [1 - a_{x1}(\mathrm{VR})]$$

where $x_{\mathrm{goal}}(\mathrm{VR})$ is the x component of the VR center of mass in the goal image and Δx_{VR} is the VR center of mass x component translation. Finally the partial vector $\vec{v}_i$ is given by (3).

Then, the overall localization vector is computed summing all the partial vectors, weighed with their normalized coupling reliability value CR_i:

$$\vec{V} = \sum_{i=0}^{N} \frac{\exp(-\mathrm{CR}_i)}{\sum_{i=0}^{N} \exp(-\mathrm{CR}_i)} \cdot \vec{v}_i,$$

where N is the number of VR couples used.

If no valid VRs are present in the actual image or no valid VRs coupling are found, a displacement vector cannot be estimated. This situation should be managed by a higher level control module.

5. Tests

Different tests have been carried on in order to verify the system performance.

Some images, grabbed during an indoor navigation, have been segmented using a classic region-growing technique. Examples are shown in Fig. 2 and 3. The VR couplings of the two examples of Fig. 2 using the Fourier–Mellin correlation method are shown by the arrows in Fig. 3. In Table 1 the coupling reliability (CR) values of the examples shown in Fig. 3 are presented. The first row represents the valid VRs extracted from the actual image and the first column represents the valid VRs extracted from the goal image. The selected results of the coupling method described in Section 3 are written in bold. As it can be noticed, all the matchable VRs have been tested and all the possible correct coupling have been found.

In order to verify the reliability of the rotation and scale factors derived from the Fourier–Mellin

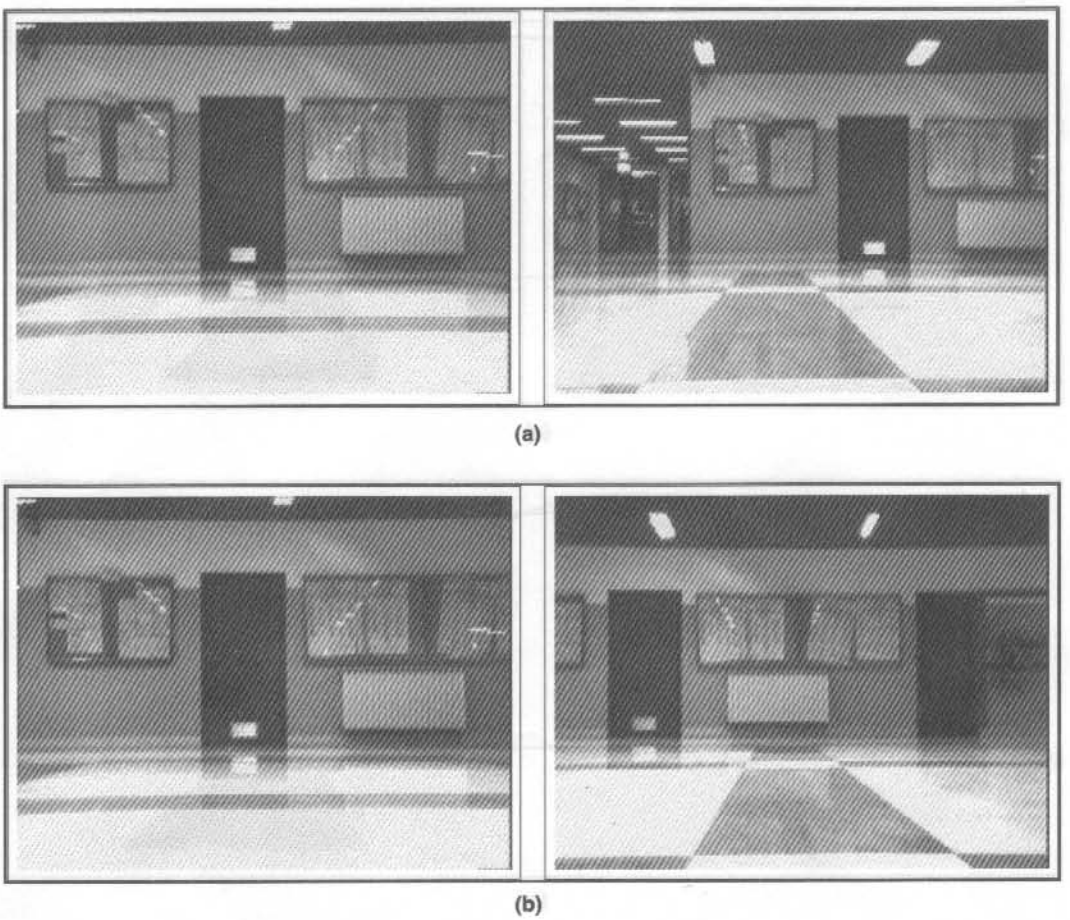

Fig. 2. Example images taken from indoor navigation.

transform, used to reinforce the displacement vector estimate, the following test has been done. The synthetic vectorial images, shown in Fig. 4 have been scaled from a $0.5\times$ to a $1.5\times$ factor with a $0.1\times$ step and rotated from $-180°$ to $+180°$ with a $5°$ step. Synthetic images have been used in order to avoid aliasing problems in the generation of the scaled and rotated test images. Statistics of the Fourier–Mellin estimate error of these artificial rotation and scaling are reported in Table 2. As it can be noticed, the errors are negligible and thus these values can be used to correct the displacement of the affine model.

The homing algorithm has been tested computing the first step estimates in different positions inside a wide area around the goal position (Figs. 5(a) and (b)) and performing complete navigations (Fig. 5(c)). For a better visualization, vectors are shown with halved modules. In Fig. 5(b) the rotation and scale parameters obtained from Fourier–Mellin are used to correct the displacement estimate. They reduce the number of false couplings and give a more precise first movement estimate.

Finally, Fig. 5(c) shows complete navigation examples using the two above described methods. They easily reach the goal position and have comparable performance. In fact, the precision in reaching the goal position can be set as a navigation parameter and the number of steps required is function of the conservativeness of the H and K setting.

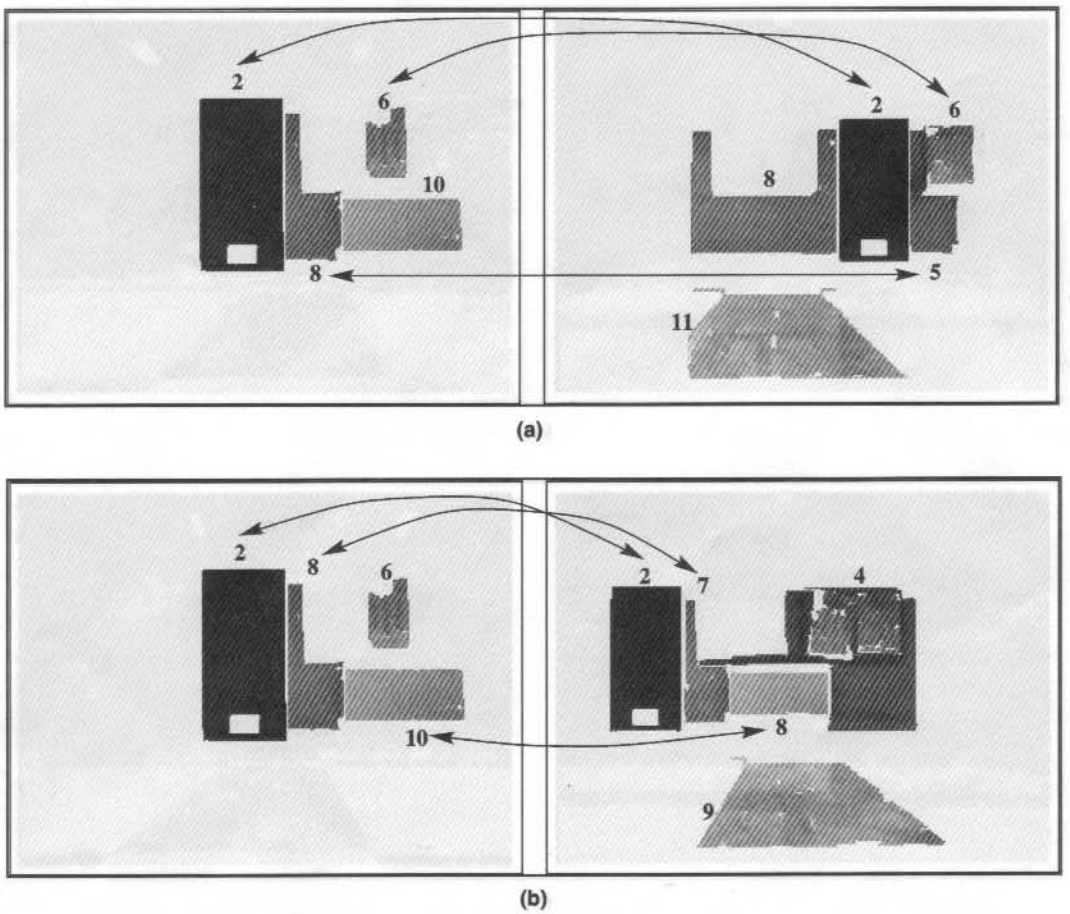

Fig. 3. VR coupling obtained from images in Fig. 2.

Table 1
Coupling values obtained from the VRs coupling examples in Fig. 3

CR	a					b				
	2	5	6	8	11	2	4	7	8	9
2	**0.632**	5.379	5.971	3.679	1.111	**0.683**	6.380	4.502	1.748	1.883
6	3.032	6.501	**1.522**	5.020	3.684	2.845	3.343	5.392	6.031	4.743
8	4.840	**0.479**	7.109	5.699	4.081	4.817	7.858	**0.336**	5.447	4.139
10	1.907	4.468	9.187	2.913	1.398	2.152	8.561	4.995	**0.282**	1.890

6. Conclusions

In this paper a homing algorithm for autonomous robots that uses only visual information is presented. The proposed homing method uses VRs autonomously extracted from the environment images and computes a descriptor for each VR using the Fourier–Mellin transform. The choice of this operator derives from its invariance to both scale and orientation.

A new method is proposed to equalize the VR descriptors and, a distance measurement is used to couple descriptors across images taken from different positions. From the comparison of

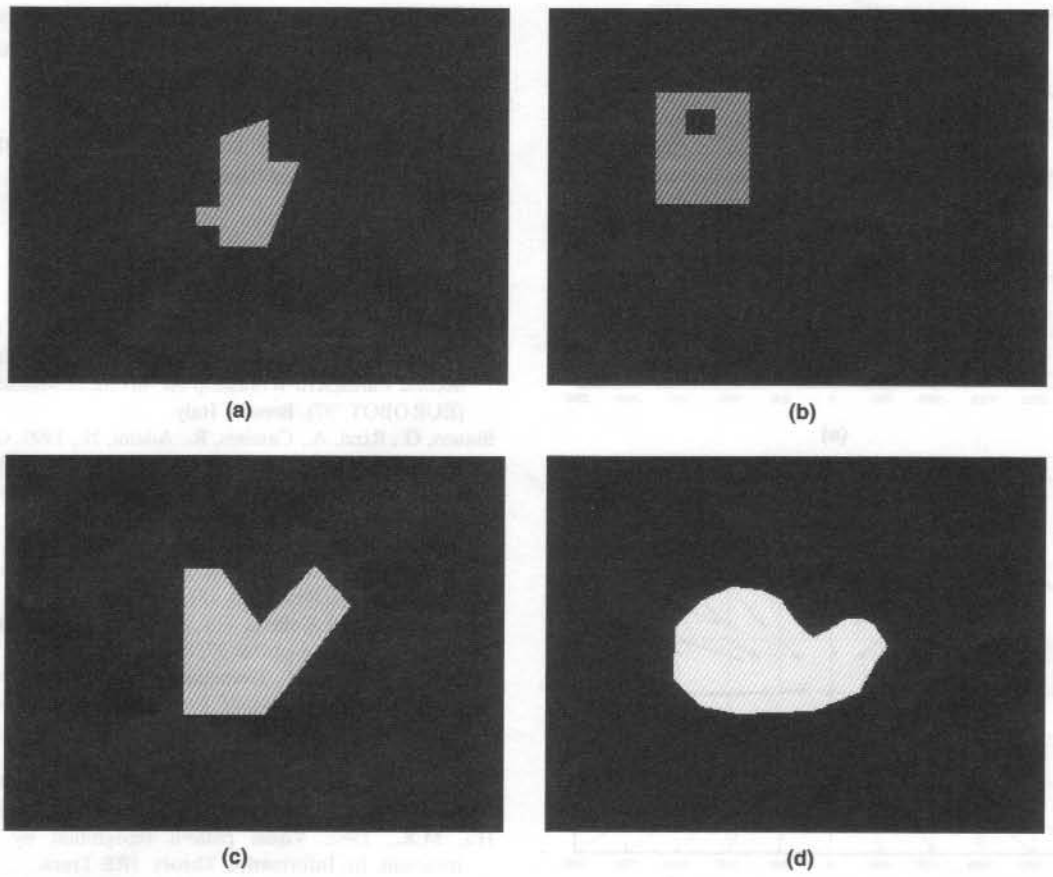

Fig. 4. Test images for scale and rotation estimate.

Table 2
Estimated scale and rotation parameters

Image	Max	μ	σ ($\cdot 10^{-3}$)
	Scale		
A	0.132	0.0063	1.331
B	0.098	0.0071	0.843
C	0.132	0.0018	1.204
D	0.016	0.0000	0.098
	Rotation		
A	1.312	0.0067	225.6
B	0.875	0.0000	119.7
C	0.938	0.0000	151.2
D	0.562	0.0000	60.9

the coupled VRs an estimate of the robot displacement is made. Using this information the robot can navigate toward the goal position. The use of VR Mellin equalized descriptors performs almost in every situation correct couplings.

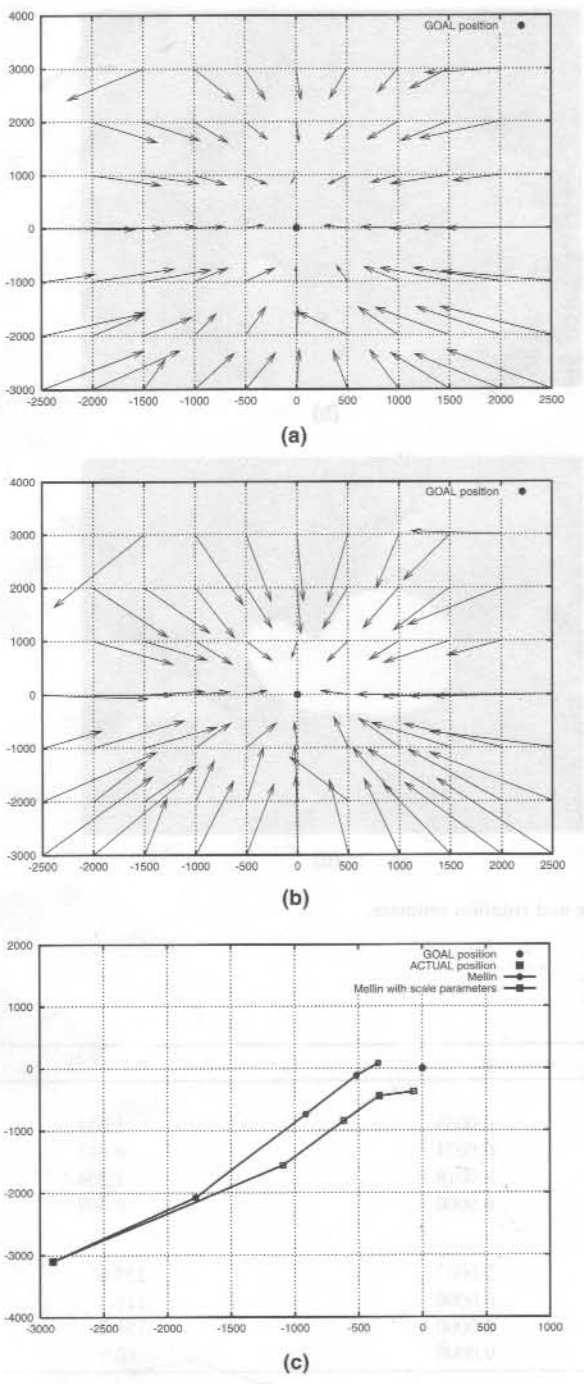

Fig. 5. Movement estimate using (a) Fourier–Mellin, (b) Fourier–Mellin with scale parameter and (c) navigation example using the two methods.

An improvement of the displacement estimate robustness is obtained using the relative scale and rotation factors extracted from the Mellin Transform.

Results prove the effectiveness of the method for indoor robot navigations.

References

Bianco, G., Cassinis, R., Rizzi, A., Adami, N., Mosna, P., 1997. A bee-inspired robot visual homing method. In: Proc. Second Euromicro Workshop on Advanced Mobile Robots (EUROBOT '97), Brescia, Italy.

Bianco, G., Rizzi, A., Cassinis, R., Adami, N., 1999. Guidance principle and robustness issues for a biologically-inspired visual homing. In: Proc. EUROBOT99 Third Euromicro Workshop on Advanced Mobile Robots 99, Zurich, Switzerland.

Cartwright, B.A., Collett, T.S., 1983. Landmark learning in bees. J. Comp. Physiol. A 151, 521–543.

Cartwright, B.A., Collett, T.S., 1987. Landmark maps for honeybees. Biol. Cybern. 57, 85–93.

Chen, Q.S., Defrise, M., Deconinc, F., 1994. Symmetric phase-only matched filtering of Fourier–Mellin transforms for image registration and recognition. IEEE Trans. Pattern Anal. Machine Intell. 16 (n.12).

Goshtasby, A., 1985. Template matching in rotated images. IEEE Trans. Pattern Anal. Machine Intell. 7 (3), 338–344.

Hu, M.K., 1962. Visual pattern recognition by moment invariant. In: Information Theory. IRE Trans.

Oberhettinger, F., 1974. Tables of Mellin Transforms. Springer, Berlin.

Pratt, W.K., 1978. Digital Image Processing. Wiley, New York, pp. 526–566.

Ravichandran, G., Trivedi, M.M., 1995. Circular-Mellin features for texture segmentation. IEEE Trans. Image Process. December.

Reiss, T.H., 1991. The revised fundamental theorem of moment in variants. IEEE Trans. Pattern Anal. Machine Intell., 830–834.

Rizzi, A., Bianco, G., Cassinis, R., 1998. A bee-inspired visual homing using color images. Robot. Autonomous Systems 25, 159–164.

Segman, J., 1992. Fourier cross correlation and invariant transformations for an optimal recognition of functions deformed by affine groups. J. Opt. Soc. Amer. A. 9 (6), 895–902.

Vanderburg, G.J., Rosenfeld, A., 1977. Two-stage template matching. IEEE Trans. Comput. 26, 384–393.