Unsupervised Matching of Visual Landmarks for Robotic Homing using Fourier-Mellin Transform

A.Rizzi, D.Duina, S.Inelli, R.Cassinis,
Dept. of Electronics for Automation - University of Brescia
E-mail: rizzi@ing.unibs.it

Abstract. This paper presents an algorithm that uses visual information to achieve the homing of an autonomous agent inside a previously visited environment. An image grabbed at the target position is compared with the currently perceived one to determine the position of the robot and its target. Only particular regions of the image called *Visual Reference* are taken into account. A visual reference correlation criterion that uses the Fourier–Mellin transform to match the Visual References in different images is chosen. This transform in fact allows to compute Visual References invariant to rotation, scaling and translation (RST). Robustness due to the use of the Mellin Transform in the Visual References selection and coupling leads to more precise navigation. Tests and results are presented.

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 [1, 2], does not require 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. Stable chromatic areas used as landmarks are chosen automatically, without user supervision, using only the chromatic and geometric characteristics of the segmented images.

The main difference between the Cartwright and Collett model and the implemented navigation system is that the camera used in this work can not take omnidirectional images. To overcome the limitation due to a small angle of view the robot learns and then approaches the homing point keeping its heading constant. Doing the same from several directions simply requires an image for each chosen direction and a software program to switch from one to the other. Using this approach, the visual landmark changes on the image plane can be described with a simplified affine model. Only particular visual landmarks, automatically extracted from the image, called *Visual References* (VRs), are used for the position estimate. The robot movements can be mapped into VRs translations and apparent dimension changes on the camera acquisition plane. 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 level navigation module is supposed to drive the robot around the target image, until a valid displacement vector can be found.

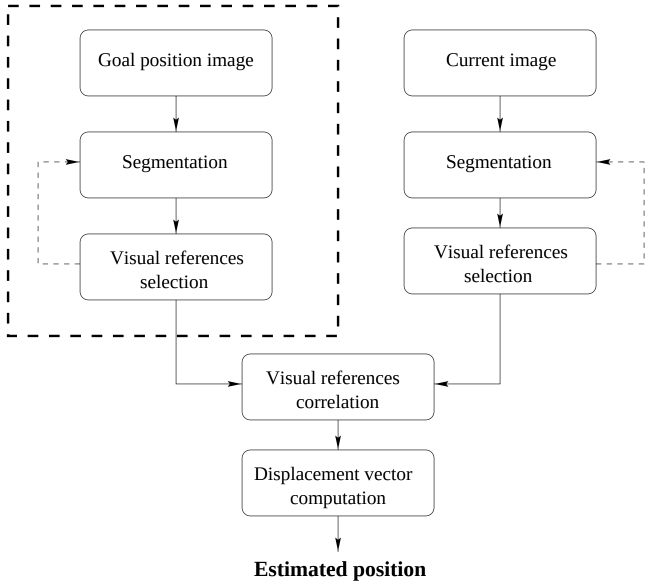


Figure 1: Block diagram

The original approach developed by our research group [3] originally compared these different VRs by using three color parameters and four geometrical parameters, for every VR [4, 5], obtaining a likelihood function. The coupling was then obtained by searching the maximum of the likelihood function. In this work, the use of Fourier-Mellin transform [6], invariant with respect to translations and scale changes [7] , is proposed to carry out the VRs coupling. A VR descriptor containing its Fourier-Mellin transform, a polar–log version of the bi-dimensional Fourier transform, is computed for each VR. This step is performed using a gray–scale VR representation. A distance index based on the inter-correlation function is used to estimate the VRs descriptors coupling. A novel VR descriptor equalization has been inserted to increase the inter-correlation selectivity. Moreover the inter-correlation function between two VR descriptors gives the relative scale and rotation factor, that are used to perform a better displacement estimation, increasing the system robustness.

2 The Homing Algorithm

The proposed algorithm can be split as shown in the block diagram of Fig. 1. This method contains an unsupervised visual reference selection phase that allows the system to work in both conditioned and unconditioned environments. In this phase, the system automatically selects the VRs according to their shape and their chromatic components. A selected region is called *visual reference* instead of *landmark* because its position in the environment is unknown. The VRs in the actual image have to be correlated with the VRs in the goal position image and in this phase a measure of *coupling reliability* is introduced. Finally for each couple, the VR relative position information, weighed with its coupling reliability, is used to estimate the robot position. This estimate is based on the following simplified affine model.

As mentioned above, the robot camera heading and height are kept constant. These constraints allow introducing a simplified affine model [3] and the following relations between camera translation and affine parameters are obtained:

$$\begin{cases} t_X &= -\frac{Z}{f} \cdot a_{x0} &= K \cdot a_{x0} \\ t_Z &= Z \cdot a_{x1} &= H \cdot a_{x1} \end{cases} \quad (1)$$

where t_X and t_Z are the components of the robot displacement, Z is the depth component of the distance between an object in the scene and the robot camera and a_{x0}, a_{x1} are the translation and dilation/compression factors of the simplified affine model.

For the test presented in this paper H and K have been set with a tuning phase inside the navigation environment but H and K values are not critical, giving only a displacement estimation proportional factor. In fact the proposed method uses iteratively a qualitative vector estimate instead of a single precise self-localization.

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*.

3 Using Mellin Transform

Considered a reference VR $r(x, y)$ and its rotated, shifted, scaled version $s(x, y)$, then its Fourier spectrum is:

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

where α is the rotation angle, σ the scaling factor, (x_0, y_0) the translation. It is shifting invariant with respect to $r(x, y)$. Rotation and scale can be separated by defining the $r(.)$ and $s(.)$ Fourier spectrum in polar coordinates (θ, ρ), obtaining the following relationship between the transforms:

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

Rotating $r(x, y)$ is equivalent to shifting $R_p(\theta, \rho)$ along θ. By using a radial logarithmic scale the $r(x, y)$ scale can be mapped in the R_{pl} shifting:

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

where $\lambda = log(\rho)$ and $\kappa = log(\sigma)$. Rotation and scaling corresponds to $R_{pl}(\theta, \sigma)$ shifting. A transform along ρ and then a logarithmic remapping is equivalent to the Mellin transform along the same direction.

To obtain a VR descriptor two step are necessary, 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. 2-a and the relative DFT modulus in polar–logarithmic coordinates is shown in Fig. 2-b.

In order to increase inter–correlation performance a further amplitude equalization phase, with a simple non–linear filter, is inserted. Fig. 2-c shows the filter function and Fig. 2-d shows the results of its application on Fig. 2-b.

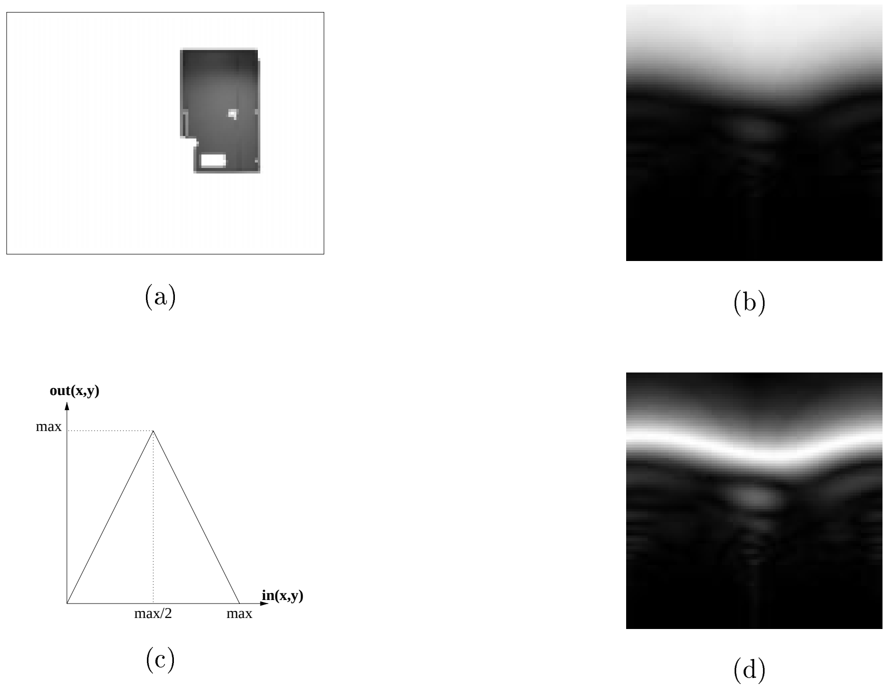


Figure 2: a) an example of valid VR, b) the VR's DFT in polar–log coordinates, c) the used equalization function, d) the resulting VR's descriptor

4 Coupling of VR descriptors

To achieve the coupling between the descriptors the following distance is used:

$$CRV_{r,s} = \frac{W_r \cdot W_s - \max\left[\varphi_{rs}(\tau, \psi)\right]}{W_r \cdot W_s} \tag{2}$$

where $\varphi_{rs}(\tau, \psi)$ is the intercorrelation function between the two descriptors having energy W_r and W_s respectively; τ and ψ are the intercorrelation function variables.

Let $\mathcal{W}_{gp}$ and $\mathcal{W}_{act}$ be the VR sets of the goal position and of the actual image, with N and M elements respectively. A correlation matrix $C_{M \times N}$ is build to contain the coupling reliability values of the $\mathcal{W}_{gp}$ and $\mathcal{W}_{act}$ elements. Starting from C two boolean matrices are computed, the first $B_{gp,act}$ links each VR in $\mathcal{W}_{gp}$with the VR in $\mathcal{W}_{act}$. Each link is the couple that for each element in $\mathcal{W}_{gp}$ maximizes the coupling reliability value with $\mathcal{W}_{act}$; only the values under a threshold of 2 are considered. The second matrix $B_{act,gp}$ is computed conversely.

The final VR couples are obtained looking for the positions with the same relative index where both $B_{gp,act}$ and $B_{act,gp}$ have a link. If this does not happen the VR in $\mathcal{W}_{gp}$ is not coupled and will not affect the localization process.

Once the correct VR coupling is found it possible to obtain the relative rotation and scale factors. The position of the maximum of the VR descriptors inter–correlation function along the radial axis gives the scale factor, while the rotation factor can be found from the position along the phase axis. The resulting expressions are:

$$S = exp(-\frac{y_{MAX}}{n}), \qquad R = \frac{x_{MAX}}{n} \cdot 180°$$

where x_{MAX} and y_{MAX} are the position of the inter–correlation function maximum, S the scale parameter, R the rotation parameter in degree and n the image dimension.

Figure 3: Goal and actual test images

With these additional parameters a new VR coupling validation phase is added. From the used affine model we can observe that VR in different positions can only be scaled but cannot rotate, therefore VR couples having a not null rotation angle are discarded. This new validation phase reduces the number of false couplings.

5 Displacement estimation

To estimate the robot position a displacement vector $\vec{v}_i$ for each VRs couple is computed. From the affine model presented in Par. 2 two parameters a_{x0} and a_{x1} must be estimated. The previous method computes a_{x1} in the following way:

$$a_{x1} = \begin{cases} \frac{A(\mathcal{R}^i_{gp}) - A(\mathcal{R}^j_{act})}{A(\mathcal{R}^i_{gp}) + A(\mathcal{R}^j_{act})} & \text{if } A(\mathcal{R}^i_{gp}) \leq A(\mathcal{R}^j_{act}) \\ \\ \frac{A(\mathcal{R}^j_{act}) - A(\mathcal{R}^i_{gp})}{A(\mathcal{R}^i_{gp}) + A(\mathcal{R}^j_{act})} & \text{if } A(\mathcal{R}^i_{gp}) \geq A(\mathcal{R}^j_{act}) \end{cases}$$

where $A(\cdot)$ is the VR area, observing that a robot movement in Z-direction corresponds to a compression/dilation in the VR area.

Using the Fourier-Mellin Transform the a_{x1} expression can be replaced with the scale parameter obtained from the VR descriptors intercorrelation function (4):

$$a_{x1} = 1 - S$$

Then a_{x0} is computed with the simplified affine model:

$$a_{x0} = (\Delta x_g - a_{x1} \cdot x_g^{goal}) \cdot (1 - a_{x1})$$

where (x_g^{goal}, y_g^{goal}) is VR center of mass position in goal image and $(\Delta x_g, \Delta y_g)$ the VR center of mass translation. Finally the partial vector $\vec{v}_i$ is given by (1).

The overall localization vector is computed summing all the partial vectors, weighed with their normalized coupling reliability value $CRV(i)$:

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

where N is the number of VR couples used.

If no valid VRs are present in actual image or no valid VRs coupling are found a displacement vector can't be estimate. This situation should be managed by a higher–level navigation module.

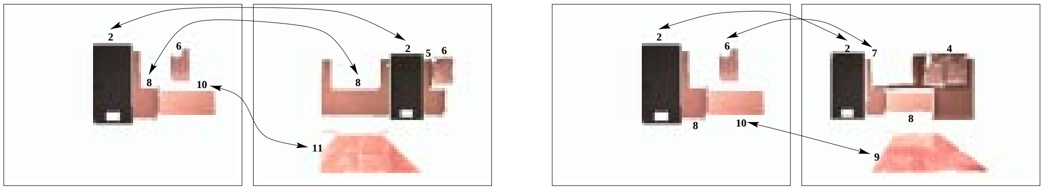


Figure 4: VR coupling with min-distance method

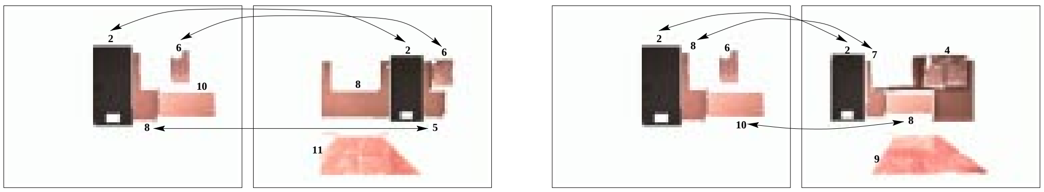


Figure 5: VR coupling with Mellin

6 Tests

Different tests have been performed in some indoor environment. In Fig. 3 two examples obtained from a robot navigation are obtained. The first and third are the same goal images. The second image is been taken 1,5 meter left and 2 meter backward from the goal image and the fourth image 2 meter right and 2 meter backward.

Computed VR couplings of the two examples of Fig. 3 are shown in Fig. 4 for the min-distance method and in Fig. 5 for the Fourier-Mellin correlation method. The estimated displacements resulting for the two techniques are given in Fig. 6.

As it can be seen from the figures, the use of VR Fourier-Mellin descriptors performs a better coupling with respect to the min–distance technique. Moreover VRs scale factor and rotation angle are automatically obtained using the Fourier-Mellin coupling algorithm.

In Fig. 7a and 7b the first steps estimate from a set of points around the goal position are shown. In Fig. 7a VRs are coupled using the min-distance technique and in Fig. 7b using the Fourier-Mellin matching algorithm. For better visualization, vectors are shown with modules reduced by a $\frac{1}{2}$ factor. Using the rotation parameter obtained from Fourier-Mellin Fig. 7c reduces the number of false couplings and gives a more precise first movement estimate. Fig. 7d shows a complete navigation example

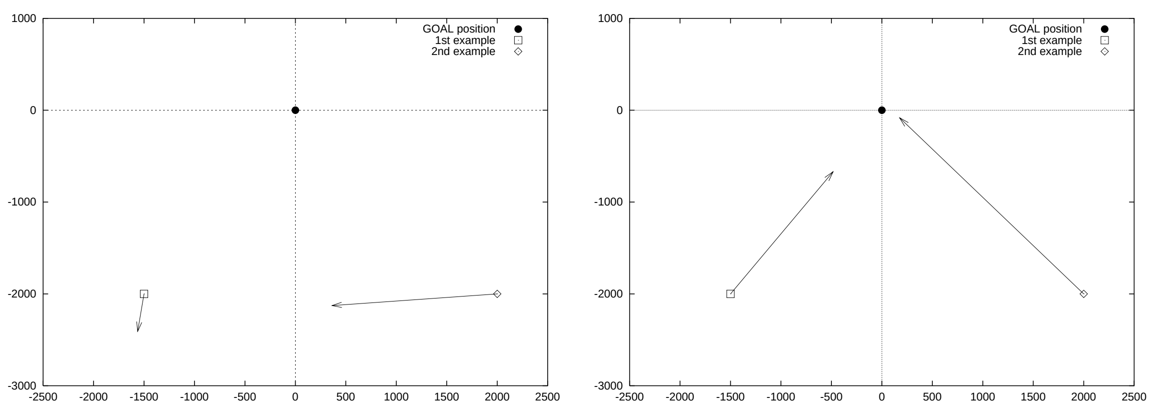
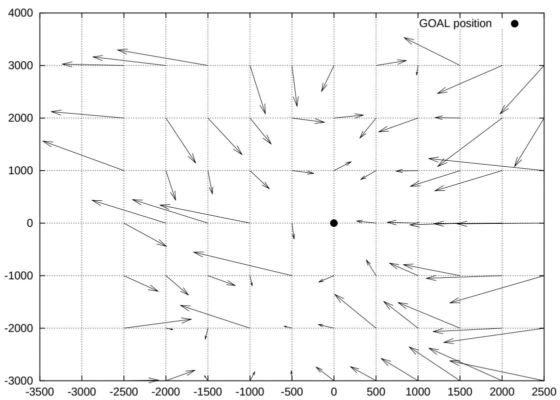
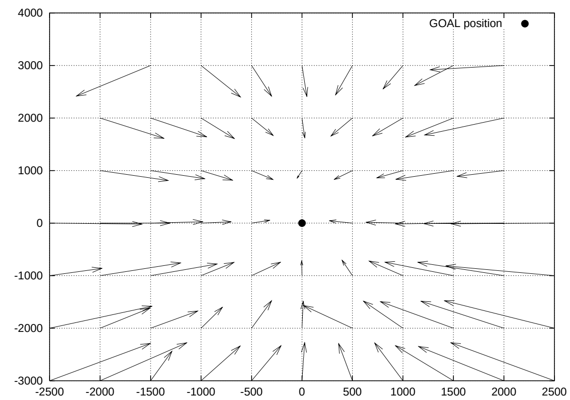


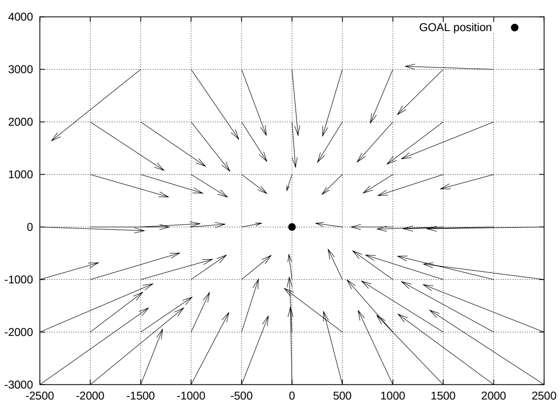
Figure 6: Displacement estimate using the min-distance algorithm and the Fourier-Mellin algorithm



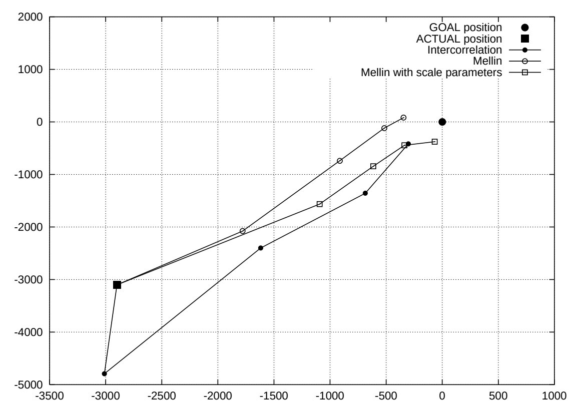
(a)



(b)



(c)



(d)

Figure 7: Movement estimate using a) min-distance technique, b) Fourier-Mellin, c) Fourier-Mellin and its scale parameter and d) navigation example using the three methods

using all the three methods. As it can be noticed, the min-distance algorithm yields to a first step failure due to false VRs couplings performed by the matching algorithm. The use of the Fourier-Mellin algorithm reduces the number of false VRs couplings, leading to more precise displacement estimate, and finally higher proportion displacement is achieved introducing the Fourier-Mellin scale factor.

Tests were performed using an Activemedia Pioneer I, driven by an Intel Pentium II 333 MHz, running Linux RedHat 5.2. The complete navigation step is performed in about 10 sec using the Fourier-Mellin algorithm(2 seconds for the segmentation process and 8 seconds for the navigation algorithm) and about 5 sec using the min-distance algorithm, those times are obtained from different navigations with four VRs in each image. Improvement in computation time can be obtained using dedicated hardware for FFT computation and software optimization.

7 Conclusions and perspectives

The proposed homing method uses *visual references* 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 distance measurement is used to couple descriptors across images taken from different positions. From the comparison of 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 descriptors performs better coupling with respect to the min-distance technique and the use of colour information allows to extract significant VRs in an easier way. An improvement of the displacement estimate robustness is obtained using the relative scale and rotation factors extracted from the Mellin Transform, decreasing the number of false couplings. In this work the Fourier-Mellin algorithm is applied to the VRs gray scale representation and is more dependent on the VR shape, rather than VR color properties. However color unconstancy, due to different environment lighting, heavily affects the VRs extraction phase. Depending on the scene illumination the segmentation process could give very different results, the test of color recovery techniques to overcome this problem are in progress.

References

[1] B. A. Cartwright and T. S. Collett. Landmark learning in bees. *J. of Comp. Physiol. A 151*, pages 521–543, 1983.

[2] B. A. Cartwright and T. S. Collett. Landmark maps for honeybees. *Biol. Cybern. 57*, pages 85–93, 1987.

[3] A. Rizzi, G. Bianco, and R. Cassinis. A bee-inspired visual homing using color images. *Robotics and Autonomous Systems 25*, pages 159–164, 1998.

[4] M. K. Hu. Visual pattern recognition by moment invariant. In *Information Theory.* IRE Transactions, February 1962.

[5] T. H. Reiss. The revised fundamental theorem of moment invariants. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pages 830–834, August 1991.

[6] F. Oberhettinger. *Tables of Mellin transforms.* Springer-Verlag, 1974.

[7] Gopalan Ravichandran and Mohan M. Trivedi. Circular-mellin features for texture segmentation. *IEEE Transaction on Image Processing*, December 1995.