ELSEVIER

Robotics and Autonomous Systems 40 (2002) 131–138

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Unsupervised matching of visual landmarks for robotic homing using Fourier–Mellin transform

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

Department of Electronics for Automation, University of Brescia, Via Branze 38, I-25100 Brescia, Italy

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 relative position of the robot and of its target. Only particular regions of the image, called *Visual References*, are taken into account. A Visual References correlation criterion that uses the Fourier–Mellin Transform to match the Visual References in different images is employed. This transform in fact allows computing Visual References that are invariant to rotation, scaling and translation. 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.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Robot navigation; Visual navigation; Robot sensors; Landmark 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 [1,2], and does not require any preconditioning of the environment. It estimates the relative positions of the robot and of its target by comparing an image grabbed at the target position with the currently perceived one. Navigation is performed exclusively using 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 among the Cartwright and Collett model and the navigation system presented in this paper is that the camera used in this work is not an omni-directional one. To overcome the limitations 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 among different images. 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

* Corresponding author.
E-mail address: riccardo.cassinis@unibs.it (R. Cassinis).

0921-8890/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0921-8890(02)00238-5

navigation module is supposed to drive the robot until it is sufficiently close to the target position, where a valid displacement vector can be found.

The first approach developed by our research group [3] originally compared these different VRs using three color parameters and four geometrical parameters for every VR, obtaining a likelihood function [4,5]. 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 and 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 VR 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.

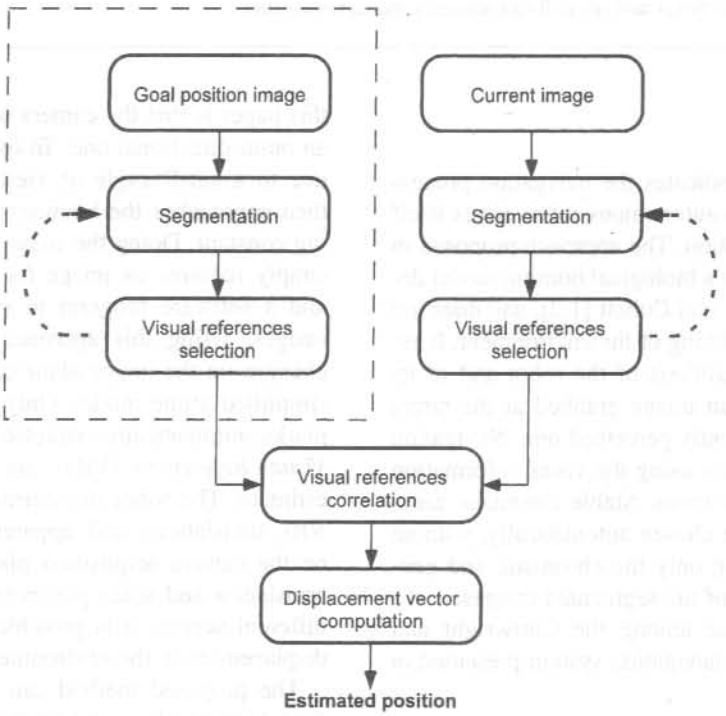


Fig. 1. System block diagram.

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:

$$t_X = -\frac{Z}{f}a_{x0} = Ka_{x0}, \qquad t_Z = Za_{x1} = Ha_{x1}, \quad (1)$$

where t_X and t_Z are the components of the robot displacement, Z is depth component of the distance between an object in the scene and the robot camera, and a_{x0}, and 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, 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 is used. The criteria of such selection is based on area, perimeter regularity and chromatic saturation.

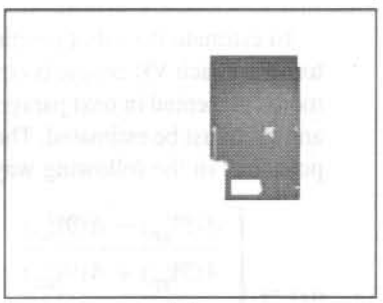
3. Using Mellin Transform

Consider VR $r(x, y)$ and its rotated, shifted, and scaled version $s(x, y)$, then its Fourier spectrum is

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

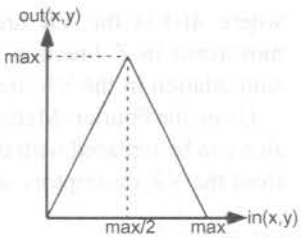
where α is the rotation angle, σ the scaling factor, and (x_0, y_0) the translation. It is the shifting invariant with respect to $r(x, y)$. Rotation and scale can be separated by defining the $r(\cdot)$ and $s(\cdot)$ Fourier spectrum in polar coordinates (θ, ρ), obtaining the following relationship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

$$S_p(\theta, \rho) = \sigma^{-2}R_p\left(\theta - \alpha, \frac{\rho}{\sigma}\right).$$

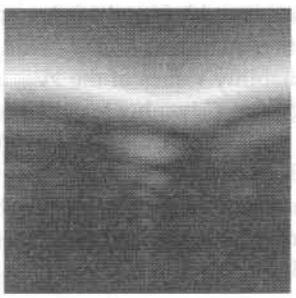


(a)

(b)



(c)



(d)

Fig. 2. (a) An example of a valid VR; (b) the VR's DFT in polar-logarithmic coordinates; (c) the used equalization function; (d) the resulting VR's descriptor.

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:

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

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 re-mapping is equivalent to the Mellin Transform along the same direction.

To obtain a VR descriptor, two steps 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).

4. Coupling of VR descriptors

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

$$\mathrm{CRV}_{r,s} = \frac{W_r \cdot W_s - \max[\phi_{rs}(\tau, \psi)]}{W_r W_s}, \tag{2}$$

where $\phi_{rs}(\tau, \psi)$ is the inter-correlation function between the two descriptors having energy W_r and W_s respectively; τ and ψ are the inter-correlation 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 built 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 is 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 yields the scale factor, while the rotation factor can be found from the position along the phase axis. The resulting expressions are

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

where x_{MAX} and y_{MAX} are the position of the inter-correlation function maximum, S the scale parameter, R the rotation parameter in degrees and n the image dimension. 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 $\mathbf{v}_i$ for each VR couple is computed. From the affine model presented in next paragraph two parameters a_{x0} and a_{x1} must be estimated. The previous method computes a_{x1} in the following way:

$$a_{x1} = \begin{cases} \dfrac{A(\Re^i_{gp}) - A(\Re^j_{act})}{A(\Re^i_{gp}) + A(\Re^j_{act})} & \text{if } A(\Re^i_{gp}) \le A(\Re^j_{act}), \\ \dfrac{A(\Re^i_{act}) - A(\Re^j_{gp})}{A(\Re^i_{gp}) + A(\Re^j_{act})} & \text{if } A(\Re^i_{gp}) > A(\Re^j_{act}), \end{cases}$$

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 inter-correlation function:

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

a_{x0} is then 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 the VR center of mass position in goal image and $(\Delta x_g, \Delta y_g)$ is the VR center of mass translation. Finally, the partial vector $\boldsymbol{v}_i$ is given by (1).

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

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

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 cannot be estimated. This situation should be managed by a higher-level navigation module.

6. Tests

Different tests have been performed in indoor environments. In Fig. 3 two examples from a real robot navigation are shown. The first and the third frame were taken from the goal position, while the second one was taken from a position 1.5 m left and 2 m behind the goal position, and the fourth image from a position 2 m right and 2 m behind.

Automatically computed VR couplings of the two examples of Fig. 3 are shown in Fig. 4 (min-distance method) and in Fig. 5 (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 results in a better coupling with respect

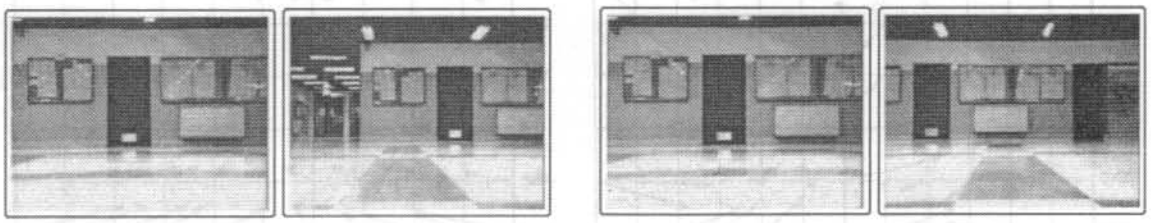


Fig. 3. Goal and actual test images.

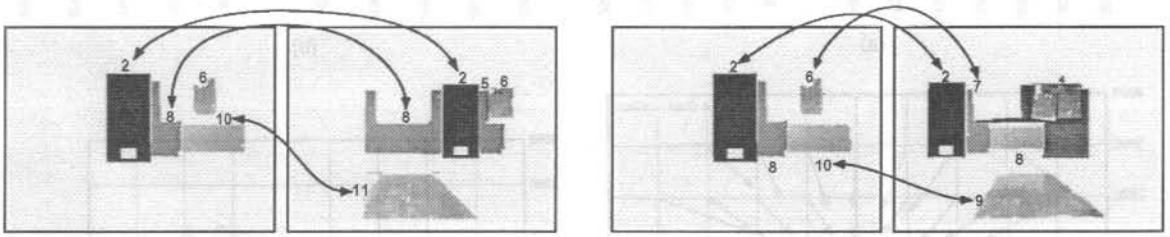


Fig. 4. VR coupling with min-distance method.

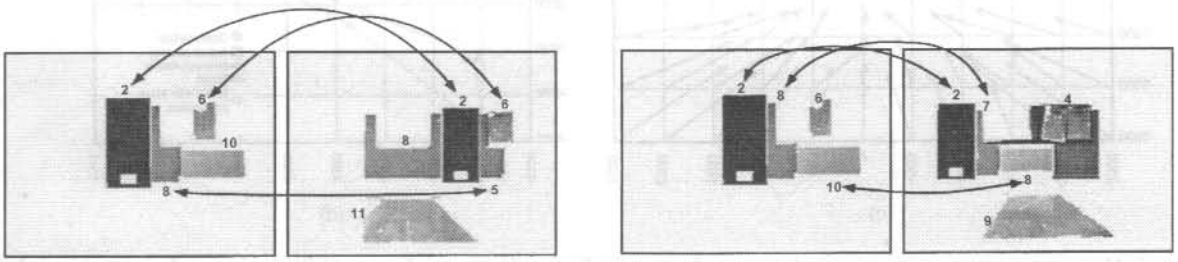


Fig. 5. VR coupling with Fourier–Mellin.

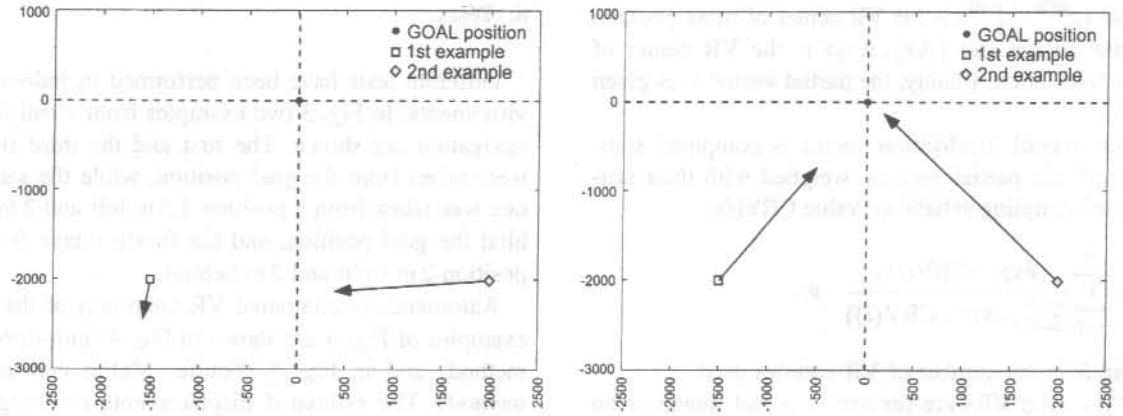


Fig. 6. Displacement estimate using the min-distance algorithm and the Fourier–Mellin algorithm.

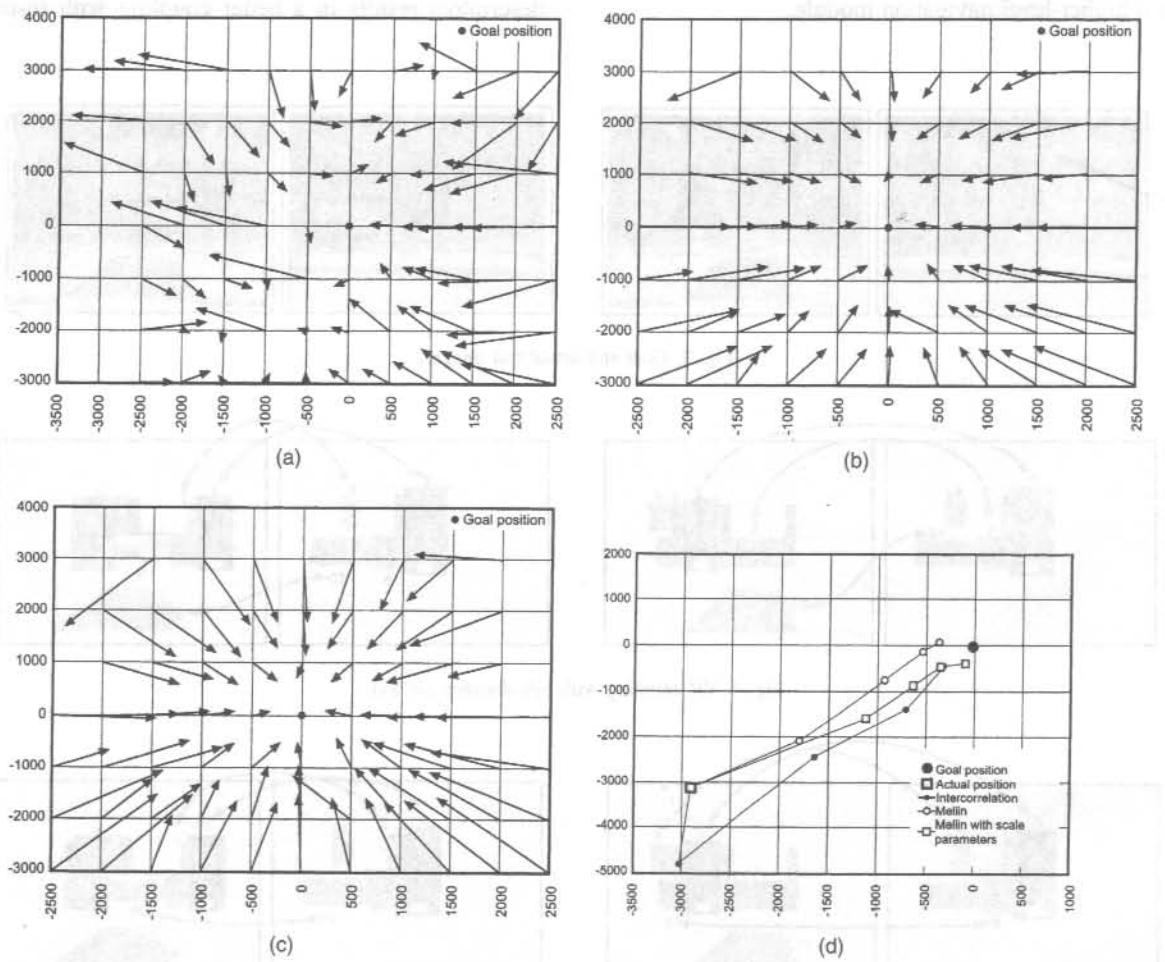


Fig. 7. Movement estimates using (a) min-distance technique, (b) Fourier–Mellin, (c) Fourier–Mellin and its scale parameters and (d) comparison of the three methods.

to the min-distance technique. Moreover, VRs scale factor and rotation angle are automatically obtained using the Fourier–Mellin coupling algorithm.

In Fig. 7(a) and (b) the first step estimates from a set of points around the goal position are shown. In Fig. 7(a) VRs are coupled using the min-distance technique and in Fig. 7(b) using the Fourier–Mellin matching algorithm. For better visualization, modules of vectors have been reduced by a factor of 0.5 using the rotation parameter obtained from Fourier–Mellin Fig. 7(c) reduces the number of false couplings and gives a more precise first movement estimate. Fig. 7(d) shows a complete navigation example using the three methods. As it can be noticed, the min-distance algorithm yields to a first step failure due to false VR couplings performed by the matching algorithm. The use of the Fourier–Mellin algorithm reduces the number of false VR couplings, leading to more precise displacement estimates, and finally higher proportion displacement is achieved, introducing the Fourier–Mellin scale factor. Tests were performed using an Activ-Media Pioneer I robot driven by an Intel Pentium II 333 MHz running Linux RedHat 5.2. The complete navigation step was performed in about 10 s using the Fourier–Mellin algorithm (2 s for the segmentation process and 8 s for the navigation algorithm) and in about 5 s using the min-distance algorithm. Such times were obtained averaging several navigations with four VRs in each image. Improvements in computation time, if at all needed, could be obtained using dedicated hardware for FFT computation and performing a general software optimization.

7. Conclusions and perspectives

The proposed homing method uses VRs extracted from the environment images in a completely autonomous way, and computes a descriptor for each VR using the Fourier–Mellin Transform. The main reason for choosing this operator is its invariance to scale and orientation. A distance measurement is used to couple descriptors in 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 towards the goal position. The use of VR Mellin descriptors performs better couplings with respect to the min-distance technique and the use of color information allows extracting 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 inconstancy, due to different environment lighting, heavily affects the VRs extraction phase, because, depending on the scene illumination, the segmentation process could give very different results. Color recovery techniques to overcome this problem are in progress.

References

[1] B.A. Cartwright, T.S. Collett, Landmark learning in bees, Journal of Comparative Physiology A 151 (1983) 521–543.
[2] B.A. Cartwright, T.S. Collett, Landmark maps for honeybees, Biology of Cybernetics 57 (1987) 85–93.
[3] A. Rizzi, G. Bianco, R. Cassinis, A bee-inspired visual homing using color images, Robotics and Autonomous Systems 25 (1998) 159–164.
[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 Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 830–834.
[6] F. Oberhettinger, Tables of Mellin Transforms, Springer, Berlin, 1974.
[7] G. Ravichandran, M.M. Trivedi, Circular-Mellin features for texture segmentation, IEEE Transactions on Image Processing 4 (12) (1995) 1629–1640.

Riccardo Cassinis received his degree in Electronic Engineering in 1977 from the Polytechnic University of Milan. In 1987 he was appointed Associate Professor of Robotics and of Numerical Systems Design at the University of Udine. Since 1991 he is Associate Professor of Computer Science and of Robotics at the University of Brescia. He has established and directed robotics laboratories at the Polytechnic University of Milan, at the University of Udine, and is now the Director of the Advanced Robotics Laboratory of the University of Brescia. Since 1975 he has been working on several topics related to industrial robots, and since 1985 he is involved in navigation and sensing problems for advanced mobile robots. His current interests include mobile robots for exploring unknown environment, with particular attention on problems related to humanitarian de-mining.

Alessandro Rizzi graduated in information science at the State University of Milan in 1992 and got his Ph.D. in Information Engineering from the University of Brescia (Italy). From 1992 to 1994 he worked as a contract researcher at the University of Milan in the Image Laboratory, studying the problem of color constancy and its application in computer graphics. From 1994 to 1996 he worked as a contract professor of Information Systems at the University of Brescia. From 1994 to 2000 he has been working with Professor Riccardo Cassinis at the Advanced Robotics Laboratory, studying problems of self-localization and navigation using color information, omni-directional devices and biologically inspired models. In 1999 he also worked as a contract professor of Computer Graphics at Politecnico Milano. He is currently Research Associate at the State University of Milan.