Stereo Inverse Perspective Mapping: Theory and Applications

Massimo Bertozzi, Alberto Broggi, Alessandra Fascioli
Dipartimento di Ingegneria dell'Informazione
Università di Parma, Italy
e-mail: {bertozzi,broggi,fascal}@CE.UniPR.IT

Abstract

This paper discusses an extension to the Inverse Perspective Mapping geometrical transform to the processing of stereo images and presents the calibration method used on the ARGO autonomous vehicle. The article features also an example of application in the automotive field, in which the stereo Inverse Perspective Mapping helps to speed up the process.

1 Introduction

The processing of images is generally performed at different levels, the lowest of which is characterized by the preservation of the data structure after the processing. Different techniques have been introduced for low-level image processing and can be classified in three main categories: *Pointwise operations*, *Cellular Automaton operations*, and *Global operations* [1].

In particular Global operations are transforms between different domains; their application simplifies the detection of image features which, conversely, would require a more complex computation in the original domain. They are not based on a *one-to-one* mapping between pixels of the original and the processed image (like Pointwise operators) nor on a *few-to-one* mapping (like Cellular Automata operators): some of them are *image-wise* operations, such as the FFT [9] or the Hough [6] transforms, meaning that the whole image is taken as input for the computation of the new status of each element in the new domain, others, such as resampling filters [12], are based on the processing of portions of the original image.

The *Inverse Perspective Mapping* (IPM) geometrical transform [7] belongs to the resampling filters family: the initial image is non-homogeneously resampled, in order to produce a new image that represents the same scene as acquired from a different position.

This work is organized as follows: next section describes the theoretical basis of the IPM image transform. Section 3 describes the extension of the IPM transform to the processing of stereo images, while its application to the detection of obstacles in the automotive field is presented in section 4. Section 5 show the calibration methods used on the ARGO autonomous vehicle. Some final remarks end the paper in section 6.

2 Inverse Perspective Mapping

The angle of view under which a scene is acquired and the distance of the objects from the camera (namely the perspective effect) contribute to associate a different information content to each pixel of an image. The perspective effect in fact must be taken into account when processing images in order to weigh each pixel according to its information content, this differentiate processing turns the use of a SIMD machine to a knotty problem.

To cope with this problem a geometrical transform (*Inverse Perspective Mapping* [7], IPM) has been introduced; it allows to remove the perspective effect from the acquired image, remapping it into a new 2-dimensional domain (the *remapped domain*) in which the information content is homogeneously distributed among all pixels, thus allowing the efficient implementation of the following processing steps on a SIMD machine. Obviously the application of the IPM transform requires the knowledge of the specific acquisition conditions (camera position, orientation, optics,...) and some assumption on the scene represented in the image (here defined as *a-priori* knowledge). Thus the IPM transform can be of use in structured environments [10], where, for example, the camera is mounted in a fixed position or in situations where the calibration of the system and the surrounding environment can be sensed via other kind of sensor [11].

2.1 Removing the Perspective Effect

Assuming the vision system acquires an image of an object with a known surface, the IPM transform produces an image that represents the texture of the framed surface. In the following the discussion will be limited to the case of a planar surface $\mathcal{S}$: in this case the use of IPM allows to obtain a bird's eye view of the scene.

The acquisition process can be devised as a transform from the 3D Euclidean space $\mathcal{W}$ to the 2D Euclidean space I, where:

- $\mathcal{W} = \left\{(x,y,z)\right\} \in E^3$ represents the 3*D* world space (*world-coordinate system*), where the real world is defined;
- $I = \left\{(u,v)\right\} \in E^2$ represents the 2*D* image space (*image-coordinate system*), where the 3D scene is projected.

The I space corresponds to the acquired image while, according to the flatness assumption, the remapped image is defined as the *xy* plane of the $\mathcal{W}$ space, namely the $\mathcal{S} \triangleq \left\{(x,y,0) \in \mathcal{W}\right\}$ surface.

2.1.1 $I \rightarrow \mathcal{S}$ mapping

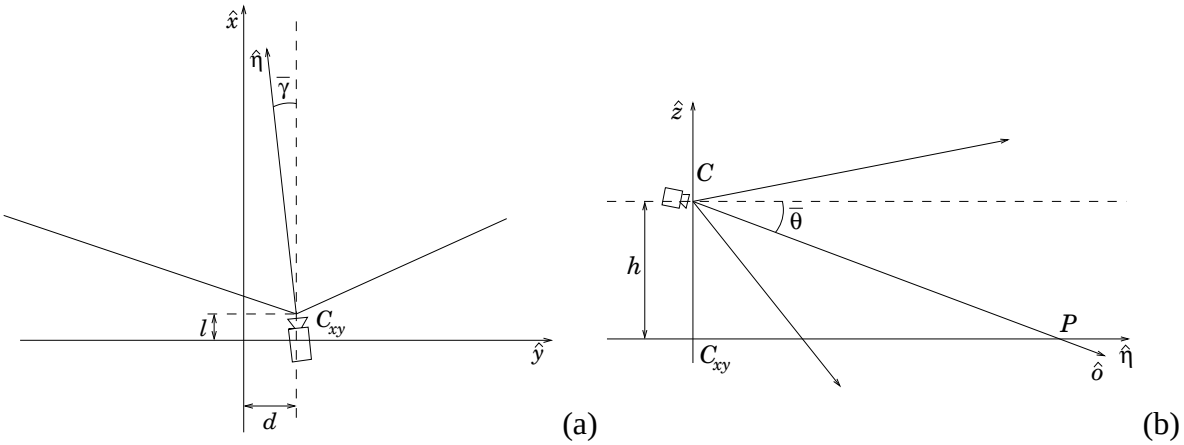


Figure 1: a) The *xy* plane in the $\mathcal{W}$ space, namely the $\mathcal{S}$ surface; and b) the $z\eta$ plane, assuming the Origin translated onto the projection C_{xy} of C on $\mathcal{S}$

The use of the IPM transform requires the knowledge of the following parameters [8]:

- *viewpoint*: the camera position is $C = (l, d, h) \in \mathcal{W}$;
- *viewing direction*: the optical axis $\hat{o}$ is determined by the following angles:
  - $\overline{\gamma}$: the angle formed by the projection (defined by versor $\hat{\eta}$) of the optical axis $\hat{o}$ on the plane $z = 0$ and the x axis (as shown in figure 1.a);
  - $\overline{\theta}$: the angle formed by the optical axis $\hat{o}$ and versor $\hat{\eta}$ (as shown in figure 1.b);
- *aperture*: the camera angular aperture is 2α;
- *resolution*: the camera resolution is $n \times n$.

Thanks to few algebraic and trigonometric operations the final mapping $f : I \to \mathcal{S}$ as a function of u and v is given by:

$$\begin{cases} x(u,v) &= h \times \cot\left[(\overline{\theta} - \alpha) + u\,\frac{2\alpha}{n-1}\right] \times \cos\left[(\overline{\gamma} - \alpha) + v\,\frac{2\alpha}{n-1}\right] + l \\ y(u,v) &= h \times \cot\left[(\overline{\theta} - \alpha) + u\,\frac{2\alpha}{n-1}\right] \times \sin\left[(\overline{\gamma} - \alpha) + v\,\frac{2\alpha}{n-1}\right] + d \\ z(u,v) &= 0 \end{cases} . \qquad (1)$$

Equations (1) return the point $(x, y, 0) \in \mathcal{S}$ corresponding to point $(u, v) \in I$.

2.1.2 $\mathcal{S} \to I$ mapping

Following similar algebraic and trigonometric manipulations, the inverse transform $g : \mathcal{S} \to I$ (the dual mapping) is given as follows:

$$u(x,y,0) = \frac{\arctan\left\{\frac{h\,\sin\left[\arctan\left(\frac{y-d}{x-l}\right)\right]}{y-d}\right\} - \left(\overline{\theta} - \alpha\right)}{\frac{2\alpha}{n-1}} \quad \text{and} \quad v(x,y,0) = \frac{\arctan\left[\frac{y-d}{x-l}\right] - \left(\overline{\gamma} - \alpha\right)}{\frac{2\alpha}{n-1}} \qquad (2)$$

The use of the projection transform defined by equations (2) allows to remove the perspective effect and to recover the texture of the $\mathcal{S}$ surface (the $z = 0$ plane of the $\mathcal{W}$ space). The array of pixels of coordinates $(x, y, 0) \in \mathcal{W}$ (which form the remapped image) is scanned and each pixel is assigned the value of the corresponding pixel of coordinates $\Big(u(x,y,0), v(x,y,0)\Big) \in I$.

3 Extension of IPM to Stereo Vision

As a consequence of the depth loss caused by the acquisition process, the use of a single two-dimensional image does not allow a three dimensional reconstruction of the world without the

use of any a-priori knowledge. In addition, when the target is the reconstruction of the 3D space, the solution gets more and more complex due to the larger amount of computation required by well-known approaches, such as the processing of *stereo images*.

The *traditional* approach to stereo vision [5] can be divided into four steps:

1. *calibration* of the vision system;
2. localization of a *feature* in an image;
3. identification and localization of the *same* feature in the other image;
4. 3D *reconstruction* of the scene.

The problem of three dimensional reconstruction can be solved by the use of triangulations between points that correspond to the same feature (*homologous points*). Unfortunately, the determination of homologous points is a difficult task, however the introduction of some domain specific constraints (such as the assumption of a flat road in front of the cameras) can simplify it. In particular, when a complete 3D reconstruction is not required and the *verification* of the match with a given surface model suffices, the application of IPM to stereo images plays a strategic role. For example the complete 3D reconstruction can be replaced with a match with a model in many cases, such as:

- *Obstacle detection*: the model represents the environment without obstacles; any deviation from the model detects a potential obstacle.
- *Object identification*: the model encodes the shape of a specific object; the match allows to recognize the object by means of its shape.
- *Object localization*: the model encodes a position of a specific object in the environment; the measure of the deviation from the model allows to determine the actual object position.

The set of points $\mathcal{H} = \left\{ (x,y,z) \mid u^{(L)} = u^{(R)} \text{ and } v^{(L)} = v^{(R)} \right\}$, where $\left(u^{(L)}, v^{(L)}\right)$ and $\left(u^{(R)}, v^{(R)}\right)$ correspond to the projection of (x,y,z) in the I space of the left and right camera respectively, represents the zero disparity surface of the stereo system, called *horopter* [4]. Whenever a stereo system acquires images of an object that matches the horopter *size*, *shape*, and *position*, the two stereo images are identical: in this case the search for homologous points becomes trivial, since they have the same coordinates in both stereo images. This property is extremely useful when it is possible to overlap the horopter onto the model of the feature of interest: in fact, the detection

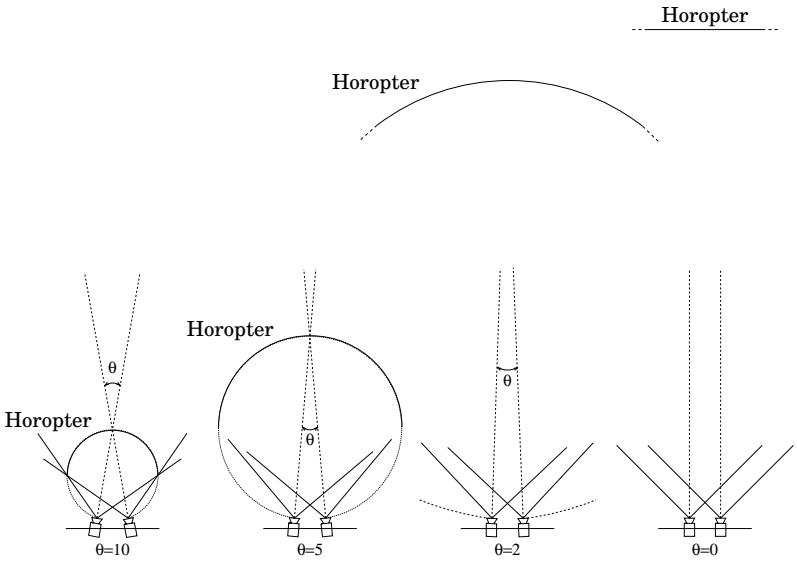


Figure 2: Relationship between the angle formed by the coplanar optical axes of two stereo cameras and the horopter

of disparities between the real scene and the model in this case is reduced to a simple pixel-wise comparison between the two acquired images.

As shown in figure 2, when the optical axis are coplanar, the horopter has a spherical shape, the smaller the angle formed by the cameras optical axes (*cameras vergence*), the larger the radius [4]; when the cameras vergence is small the horopter can be considered planar. Unfortunately, since only the horopter's *size* (radius) can be modified acting on cameras vergence, the horopter cannot be overlapped with a *generic* surface, because this would require changing also its *position* and *shape*.

IPM can solve this problem, acting as an *electronic vergence*, thus allowing to overlap the horopter with any surface. More precisely, since IPM can be used to recover the texture of a specific surface (the *xy* plane in the previous discussion), when it is applied to both stereo images (with different parameters reflecting the different acquisition setup of the two cameras) it provides two instances, namely two partially overlapping patches, of the given surface. These two patches, thanks to the knowledge of the vision system setup, can be brought to correspondence, so that both the *position* and the *shape* of the horopter are changed, and homologous points share the same coordinates in the two remapped images.

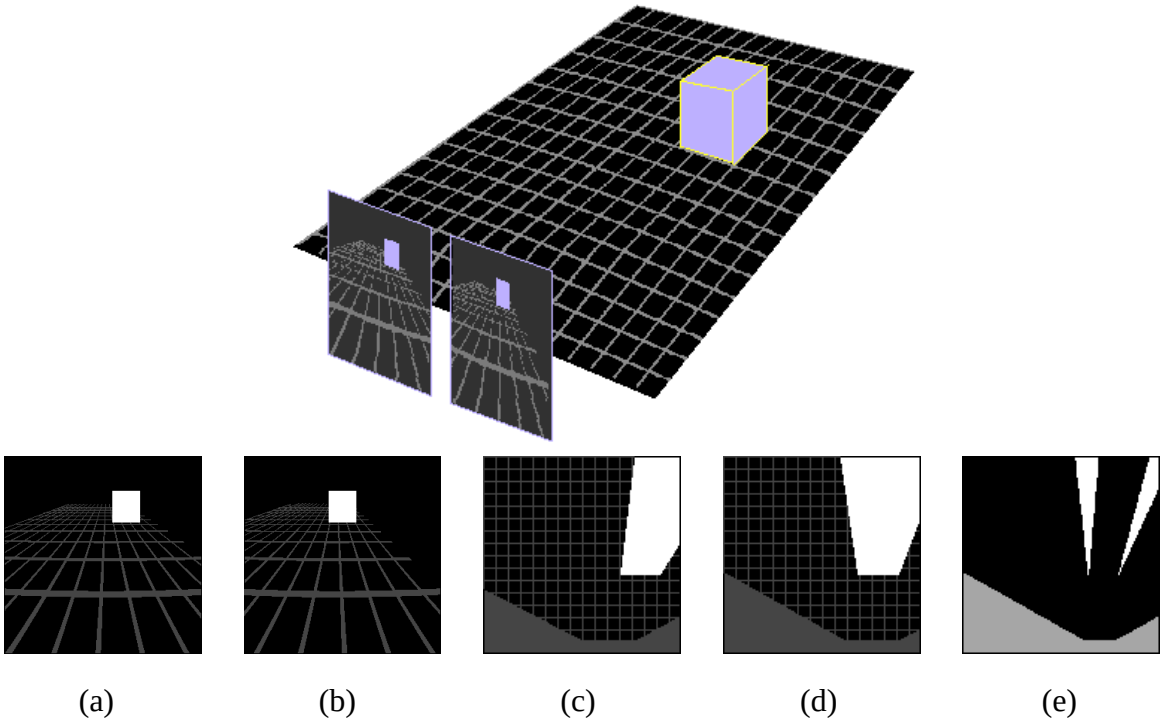
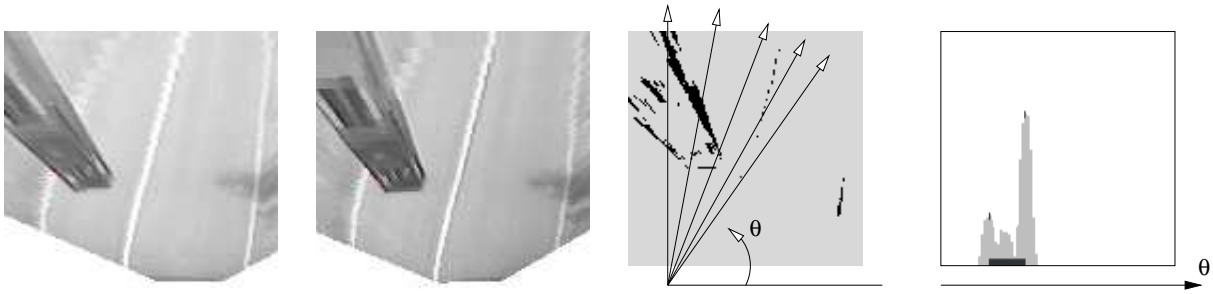


Figure 3: The framing of an ideal square homogeneous obstacle: a) left image; b) right image; c) left remapped image; d) right remapped image; e) difference image in which the gray area represents the region of the road not seen by both cameras

4 An Application of Stereo IPM to the Automotive Field

As mentioned in the previous section, when obstacle detection means the mere *localization* of objects that can obstruct a vehicle's path without their complete *identification* or *recognition*, stereo IPM can be used in conjunction with a geometrical model of the road in front of the vehicle [2]. Assuming a *flat road* hypothesis, IPM is performed using the relations expressed by equations (2). The flat road model is checked using a pixel-wise difference between the two remapped images: in correspondence to a *generic obstacle* in front of the vehicle, namely anything rising up from the road surface, the difference image features sufficiently large clusters of non-zero pixels that have a specific shape.

In fact it is easily demonstrable that the IPM transform maps straight lines perpendicular to the road surface (such as vertical edges of obstacles) into straight lines passing through the projection



(a) (b) (c) (d)

Figure 4: Obstacle detection: (a) and (b) left and right stereo remapped images respectively, (c) the difference image and the angles of view (θ), (d) the polar histogram.

$C_{xy} \equiv (l, d, 0)$ of the camera onto the $z = 0$ plane (see figure 1). Since in a stereo vision system the projections of the cameras on $z = 0$ do not coincide, due to the different angles of view of the stereo cameras, an ideal square homogeneous obstacle produces two clusters of pixels with a triangular shape in the difference image, in correspondence to its vertical edges (see figure 3). Triangles found in real cases are not so clearly defined and often not distinctly disjoint because of the texture, irregular shape, and non-homogeneous color of real obstacles. Nevertheless clusters of pixels having an almost triangular shape are anyway recognizable in the difference image (see figure 4.c). The obstacle detection process is thus reduced to the localization of these *triangles* in the difference image.

The potential presence of two or more obstacles in front of the vehicle at the same time as well as partially visible obstacles, complicate the process. Thus a further processing step is needed in order to classify triangles that belong to the same obstacle.

4.1 Obstacle Detection

A *polar histogram* is used for the detection of triangles, which is obtained scanning the difference image with respect to a point F in the $z = 0$ plane of the $\mathcal{W}$ space, called *focus*. The polar histogram is computed considering every straight line originating from the focus F and counting the number of overthreshold pixels lying on that line (see figure 4.c). The values of the polar histogram are then normalized and a low-pass filter is applied in order to decrease the influence of noise.

Since each of the triangle edges prolongations intersects one of the projections of the two cameras onto the road plane, the focus is placed in the middle of the two projections: in this case the polar histogram presents an appreciable peak corresponding to each triangle. The position of a peak within the histogram determines the angle of view under which the *obstacle edge* is seen. Peaks generated by same obstacle, i.e. by its left and right edges, must be joined in order to consider the whole area in between as occluded.

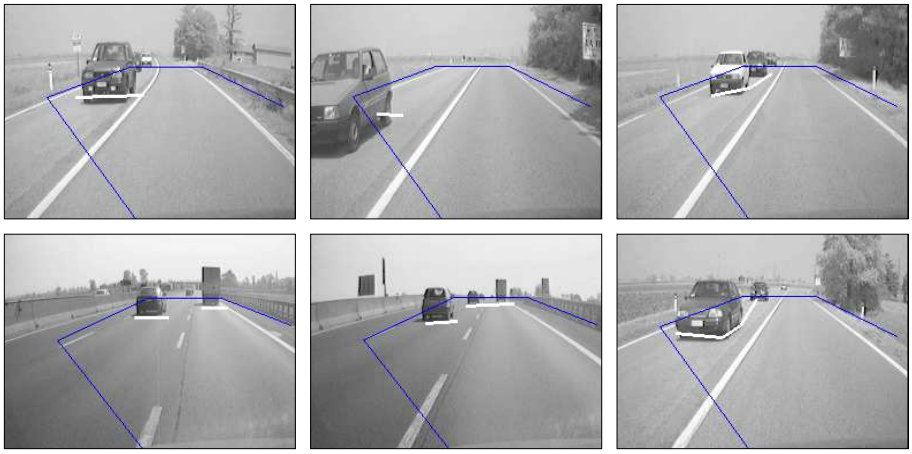


Figure 5: Obstacle detection: the result is shown by a white marker superimposed on the image captured by the left camera; a black thin line limits the portion of the road seen by both cameras

Starting from the analysis of a large number of different situations a criterion has been found, aimed to the grouping of peaks, that takes into account several characteristics such as the peaks amplitude and width, the area they subtend, as well as their distance. After the peaks joining phase, the angle of view under which the *whole obstacle* is seen is computed considering the peaks position, amplitude, and width [2].

In addition, the obstacle distance can be estimated by a further analysis of the difference image along the directions pointed out by the maxima of the polar histogram, in order to detect the triangles' corners. In fact they represent the contact points between obstacles and the road plane and thus hold the information about obstacle distance. For each peak of the polar histogram a *radial histogram* is computed scanning a specific sector of the difference image whose width is determined as a function of the peak width [2]. The number of overthreshold pixels lying in the sector is computed for every distance from the focus and the result is normalized. A simple

Figure 6: The grid used for the vision system calibration and the ARGO vehicle

threshold applied to the radial histogram allows to detect the triangles corners position and thus the obstacle distance.

The result is displayed with white markers superimposed on the left image; the markers position and size encode both the distance and width of obstacles (see figure 5).

5 System Calibration

Stereo IPM has been implemented on ARGO, the autonomous test vehicle developed at the University of Parma. This approach demonstrated to be robust with respect to vehicle movements (pitch and roll) and to a slightly imprecise system calibration. Nevertheless, camera calibration plays a basic role for the success of the approach. It is divided in two steps.

5.1 Supervised Calibration

The first part of calibration is an interactive process, a grid with known size (see figure 6) has been painted onto the ground and two stereo images are captured and used for the calibration. Thanks to a graphical interface a user selects the intersections of the grid lines into the two images using a mouse; these intersections represent a small set of homologous points whose world coordinates are known to the system; this mapping is used to compute the calibration parameters. The set of homologous points is used to minimize different cost functions, such as, the distance between each point and its neighbors and line parallelism.

This first step is intended to be performed only once when the orientation of the cameras or the

vehicle trim have changed. Since the set of homologous points is small this calibration represents only a rough guess of the parameters, and a further process is required.

5.2 Automatic Parameters Tuning

After the supervised phase the computed calibration parameters have to be refined. Moreover small changes in the vision system setup or in the vehicle trim suggest a periodic tuning of the calibration. For this purpose an automatic tool has been developed. Since this step is only a refinement, a structured environment, such as the grid, is no more required and a mere flat road in front of the vision system suffices.

Iteratively small deviations from the coarse parameters computed during the previous step are used to remap the captured images; the target is to have the remapped images as similar as possible. All the pixel of the remapped images are used to test the correctness of the calibration parameters through a least square difference approach.

6 Conclusions

In this paper an extension to the Inverse Perspective Mapping geometrical transform has been presented and an application to the automotive field discussed. This extension has been used as an alternative method for obstacle detection.

The obstacle detection functionality has been tested on board of the ARGO experimental vehicle; although the assumptions of a flat road and of a fixed camera are critical issues in the automotive field (since the system moves), the system has demonstrated to be robust with respect to vehicle movements, and obstacles can be detected with a high confidence even in case the camera orientation and camera height have drifts up to $\pm 1^o$ and ± 10 cm respectively [2]. In addition, a specific hardware support [3] has been developed, allowing to perform the resampling operation in a real-time fashion.

The main advantage of this method is that, instead of performing an exhaustive search for homologous points, a simpler match with a model encoding a-priori knowledge is performed. In addition this approach is well suited for an implementation on a SIMD architecture.

References

[1] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, 1982.

[2] M. Bertozzi and A. Broggi. GOLD: a Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection. *IEEE Transactions on Image Processing*, 1997. In press.

[3] A. Broggi, G. Conte, F. Gregoretti, C. Sansoè, and L. M. Reyneri. The Evolution of the PAPRICA System. *Integrated Computer-Aided Engineering Journal - Special Issue on Massively Parallel Computing*, 4(2):114–136, 1997.

[4] P. J. Burt, P. Anandan, and K. J. Hanna. An electronic front end processor for active vision. In *Proceedings SPIE Conference on Intelligent Robotics*, November 1992.

[5] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, 1993.

[6] J. Illingworth and J. Kittler. A survey of the Hough transform. *CVGIP*, 44:87–116, 1988.

[7] H. A. Mallot, H. H. Bülthoff, J. J. Little, and S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.

[8] W. M. Newman and R. F. Sproull. *Principles of Interactive Computer Graphics*. McGraw-Hill, Tokyo, 1981.

[9] J. H. Nussbaumer. *Fast Fourier Transform and Convolution Algorithms*. Springer-Verlag, 1990.

[10] D. A. Pomerleau. RALPH: Rapidly Adapting Lateral Position Handler. In I. Masaky, editor, *Proceedings IEEE Intelligent Vehicles'95*, pages 506–511, Detroit, September 25-26 1995. IEEE Computer Society.

[11] K. Storjohann, T. Zielke, H. A. Mallot, and W. von Seelen. Visual Obstacle Detection for Automatically Guided Vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume II, pages 761–766, 1990.

[12] M. L. P. van Lierop. Geometrical transformations on pictures represented by leafcodes. *CVGIP*, 33:81–98, 1986.

Contents