A Visual Landmark Framework for Indoor Mobile Robot Navigation

J.B. Hayet, F. Lerasle, M. Devy

Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS)
7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, FRANCE

E-mail: jbhayet,lerasle,michel@laas.fr

Abstract

This article presents vision functions needed on a mobile robot to deal with landmark-based navigation in buildings. Landmarks are planar, quadrangular surfaces, which must be distinguished from the background, typically a poster on a wall or a door-plate. In a first step, these landmarks are detected and their positions with respect to a global reference frame are learned; this learning step is supervised so that only the best landmarks are memorized, with an invariant representation based on a set of interest points. Then, when the robot looks for visible landmarks, the recognition procedure takes advantage of the partial Hausdorff distance to compare the landmark model and the detected quadrangles. The paper presents the landmark detection and recognition procedures, and discusses their performances.

1 Introduction

Landmark-based environment modeling can be used in different contexts. In [1], we described a navigation method based on the expression of trajectories with respect to landmarks reference frames; in such a context, the robot localization is only relative to a reference landmark. In [2], we described a navigation strategy dedicated for corridor-like environments: in an office building, the corridors topology can be modeled using a Generalized Voronoi Graph, and the visual recognition of landmarks is used to annotate some specific areas, like the corridor crossings or the room entrances; the robot localization is only qualitative. In this paper, we describe another application of our visual landmark framework, the enhancement of a localization function based mainly on laser segments; the environment model is mainly built thanks to the laser sensor, and the visual landmarks are used in order to recover the robot position in situations where the laser sensor is not efficient.

The contribution of this paper also consists in an improved landmark detection method, based on edge grouping by a relaxation algorithm, and a new landmark recognition method based on the matching of a set of interest points using the partial Hausdorff distance. The stability of such a method with respect to several parameters is discussed.

In the section 2, the landmark detection algorithm is described; the method is dedicated to planar and quadrangular landmarks. The section 3 presents our strategy to represent and to recognize such landmarks, from the construction of an invariant appearance model -an icon- at first, and then, from the matching of interest points extracted from this icon. Next, in the section 4, experimental results about the construction and the exploitation of a landmark-based model, are commented. Finally in the section 5, discussions about this work and some future researches are considered.

2 Quadrangle detection by edge grouping

As we focus this work on quadrangular landmarks, we chose a natural way of extracting quadrilaterons that relies upon perceptual grouping on edge segments.

Edges are obtained by classical contour extraction and segmentation methods. All the segments are separated into two classes according to their main directions in the image. The matching process is first applied to the horizontally-oriented edges, which produces additional constraints for the remaining vertically-oriented edges matching process.

The whole process is based on the sequential application of constraints. Constraints between single segments are applied in order to define an initial set of pairs of segments matchings. Then constraints between segments pairs are processed through a continuous relaxation scheme.

2.1 Constraints between single segments

Segment couples corresponding to potential landmarks are formed according to geometric and luminance consistency criteria.

Geometric constraints. Two image segments i and k must have a length ratio superior to a certain value and must overlap coarsely. For each combination (i,k), a global score noted $score(i,k)$ is computed as the sum of two terms respectively depending on the length and overlapping ratios between segments i and k.

We have an additional constraint for the vertically-oriented segments from the already matched horizontally-oriented segments. Indeed, for each couple (i,j) of vertically-oriented segments, there must exist at least one couple (k,l) of horizontally-oriented segments, such as the angular sector delimited by the straight lines containing i and j include segments k and l and vice versa.

Luminance constraints. For each extracted segment, an average grey-level profile is computed in the direction orthogonal to it. For each combination (i,k), we deduce a score resulting from correlation between these profiles. If this score exceeds a certain threshold, the correspondent potential matching is removed.

2.2 Constraints between segments couples

Uniqueness and convexity rules are checked for the remaining potential segments matchings. *Uniqueness rule* implies that for two segments pairs, any segment of the first pair can be associated to at most one segment from the other pair. *Convexity rule* says that two pairs of segments define two quadrangles which must verify rules of full inclusion or (on the contrary) no intersection.

The propagation procedure is applied to all the segments matchings remaining from the first step through continuous relaxation. Like Hummel in [3], we define a confidence level $p(i,k) \in [0,1]$ for the association between segments i and k. $p(i,k)=1$ if k is unambiguously associated with i, 0 if it is not the case.

The relaxation process consists in making the ambiguous matchings ($p(i,k) \in]0,1[$) evolve towards 0 or 1. The $n \times n$ matrix P such as $P_{ik} = p(i,k)$ is the association matrix we must estimate. K is the space of acceptable association matrices, defined as:

$$K = \{P = \{P_{ik}\}_{1 \le i,k \le n} \mid \forall (i,k)\ P_{ik} \ge 0 \ and\ \forall i \sum_{k=1}^{n} P_{ik} = 1\}$$

with n the number of image segments. The relaxation process requires two steps :

• an initial set of segments $L_0(i)$ is formed such as the initial weights $p_0(i,k)$ are zero if the association (i,k) doesn't respect the constraints on single segments. For the others, we compute $p_0(i,k)$ by :

$$p_0(i,k) = \frac{score(i,k)}{\sum_{l \in L_0(i)} score(i,l)},\ with\ L_0(i) = \{l / p_0(i,l) \neq 0\}$$

• an iterative update of the weights $p(i,k)$ to reduce the matching inconsistencies between segments. This iterative algorithm maximizes the global consistency score using gradient ascent in the K space :

$$\begin{cases} \max_{P \in K} A(P) \\ A(P) = \sum_{i,k} \sum_{j,l} r(i,j,k,l) p(i,k) p(j,l) \end{cases}$$

where the $r(i,j,k,l)$ represents the compatibility degree between associations (i,k) and (j,l) : the bigger is the value of $r(i,j,k,l)$, the stronger is the compatibility.

Figure 1 illustrates quadrilateral object detection in a typical indoor environment. The lines and their associated numbers represent the matched pairs of horizontally and vertically-oriented segments that finally define potential landmarks.

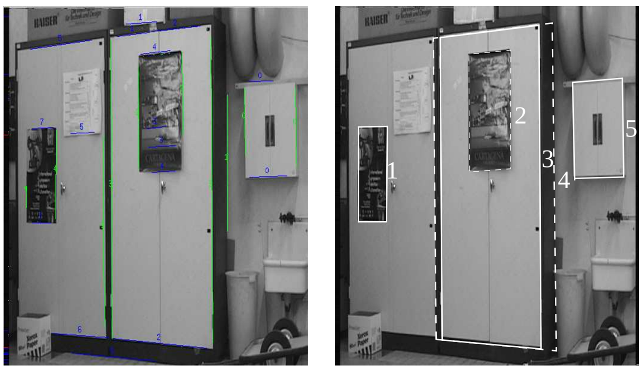

Figure 1: Segment matchings and associated detected landmarks

2.3 Final landmarks extraction

Once the matching process is done, we can use the output data, i.e. straight lines quadruples, to refine the potential landmarks boundaries. This procedure has already been described in [1], and may be rapidly recalled, as in figure 2: (1) computation of grey-level profiles along directions orthogonal to the estimated boundary direction, (2) correlation of these profiles with step-like signals and (3) RANSAC straight line parameters estimation from correlation maxima.

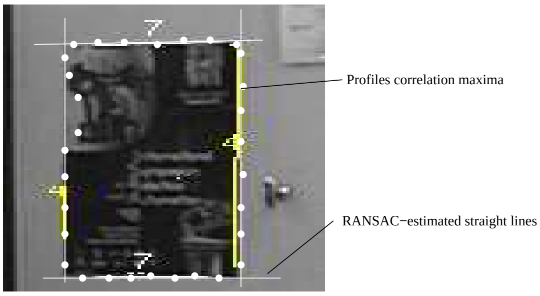

Figure 2: Refined extraction

3 Landmark recognition

The different sources for misrecognition with planar landmarks are light effects, in particular in indoor environments, perspective distortion, scale changes, and bad warping from the detection step.

To perform recognition, the simplest class of methods we may think to are correlation-based methods. Centered, normalized correlation between views of poster-like landmarks may allow to get a distance invariant to scale, perspective, overall light changes. However two drawbacks arise : first, the lack of compactness of the data to be stored, as we would need at least an image of the landmark, and, second, the difficulty to handle partial occlusions, or local light effects.

Among all the other appearance-based methods that have been described in the computer vision literature over these last years, the most promising are surely the ones based on interest points matchings. Schmidt [4] has developed a framework for object recognition based on Harris points matchings thanks to local descriptors based on Gaussian derivative. Recent studies [5] have put into light the remarkable behavior of Harris operator as far as scale, rotation or light changes are concerned.

In our situation, we also use an approach based on interest points in addition to the boundaries of the quadrangular landmark, that makes our task much easier. Indeed, these boundaries allow us to define an homography to rectify the observed pattern. Using such a mapping for recognition allows us (1) to use only *spatial* configuration of interest points, (2) at *one scale only* and (3) to have an *invariant* representation under perspective changes, up to a certain level.

3.1 Landmark iconification

Let us consider, on the one hand, an extracted quadrangular landmark $Q = \{P_i\}_{1 \le i \le 4}$ from an image I, and, on the other hand, a square S, corresponding to a $s \times s$ picture (s typically equal to 75), i.e. at lower scale than in image I. The four matchings this definition induces allow us to define an homography H_{SQ} mapping points from S to Q.

H_{SQ} allows to define a new small-sized image I' from the image I by averaging pixels from I into the pixels in I'. This process is illustrated on figure 3. To handle perspective distortions, averaging is done in order not to lose too much information from the original image to approximate the low-scale front view. If we consider a pixel (a, b) in image I', its grey level value is determined by taking into account all the pixels in image I belonging to a certain neighborhood of $H_{SQ}(a, b, 1)^T$, its image in I. This neighborhood is computed by approximating with simple heuristics the image of a pixel square, i.e. a certain quadrilateron. This process is illustrated in figure 4.

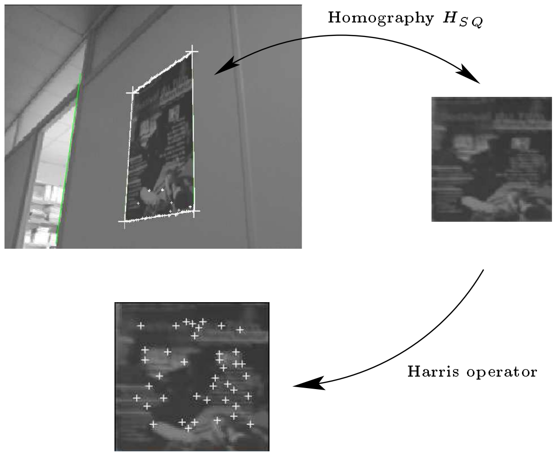

Figure 3: The landmark model construction

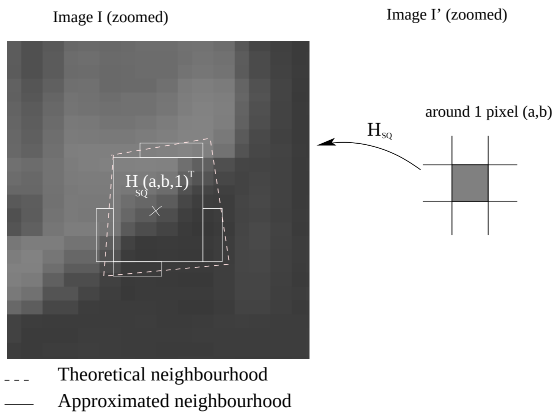

Figure 4: Averaging invariantly to perspective effects

We process this image I' by the Harris operator with a certain cornerness threshold to get a set of n_h Harris points $\{P_i^h\}_{1 \le i \le n_h}$.

3.2 Partial Hausdorff distance

In order to compare two sets of points, a popular distance is the Hausdorff distance, and noticeably its partial version introduced by Huttenlocher[6]. Let be two sets of points $S_1 = \{P_i^1\}$ and $S_2 = \{P_j^2\}$, the Hausdorff distance between S_1 and S_2 is defined by :

$$\begin{cases} d_h(S_1, S_2) = \max(h(S_1, S_2), h(S_2, S_1)) \\ h(S_1, S_2) = \max_{1 \leq i \leq |S_1|} \min_{1 \leq j \leq |S_2|} \|P_i^1 - P_j^2\| \end{cases}$$

A natural way of extending this definition to take into account outliers and to perform robust pattern recognition is to relax this definition and take the k^{th} greatest minimum distance or, equivalently, a fraction r of $\min(|S_1|, |S_2|)$, the minimum of the two sets cardinals, to define :

$$\begin{cases} d_h^r(S_1, S_2) = \max(h^r(S_1, S_2), h^r(S_2, S_1)) \\ h^r(S_1, S_2) = k^{th}_{1 \leq i \leq |S_1|} \min_{1 \leq j \leq |S_2|} \|P_i^1 - P_j^2\| \\ k = r. \min(|S_1|, |S_2|) \end{cases}$$

We set a threshold τ on the computed distance to decide which class the current object belongs to. Physically speaking, an object is recognized provided that we can find a correspondent point in the first set for at least k points in the second, and vice versa. From this definition, we see that we can stop computation as soon as $\min(|S_1|, |S_2|) - k$ outliers have been detected, so recognition tests can be performed in relatively short time. However, given a threshold τ, all fractions r are not desirable for recognition purposes and vice versa. Indeed, if we estimate the distribution of the number of points from a set S_1 for which we can find at least one point from another set S_2 within a given distance τ, among all couples of sets of n_h points, we see that there is an inherent bias in this distribution. Figure 5 illustrates this fact and shows that with $\tau = 3$ pixels and $n_h = 50$ points we must consider a fraction of points $r > 0.6$ to avoid the "intrinsic" bias zone.

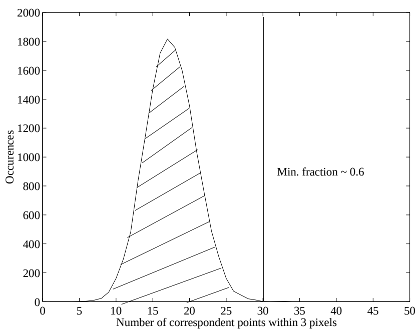

Figure 5: Constraint on the choice of Hausdorff fraction

3.3 Learning appearance models

In our implementation, the learning phase is performed on a totally supervised way, as we define, for each landmark, a set of representative images I_i from which we extract iconified views I'_i. A Principal Component Analysis on the data extracted from each landmark allows us to keep only the three most representative models that we will use for further recognition.

Saliency is guaranteed at different levels : *edges* detected at the segmentation step, *covariance level* of the iconified view I' and number of interest points extracted from I'. All of them define a saliency criterion to be satisfied satisfied.

In addition of this saliency level, we define a visibility level that indicates how far from each other are the extreme positions at which the landmark is seen. To achieve this goal, we consider the n_l learning images I_i.

For all the couples $(i, j) \in [1, n_l]^2$, we can define an inter-image homography H^{ij} mapping the vertices of the landmark in image I_i to its corresponding vertices in image I_j. Let us consider the normalized homography $\hat{H}^{ij}$, such as $\hat{H}^{ij}_{33} = 1$. Then we define the visibility confidence by :

$$v_c = max_{ij} \|\hat{H}^{ij} - I\|$$

I is the identity matrix. The greater is v_c, the more extended is the area on which the landmark can be seen, provided that the camera parameters remain constant in this phase.

3.4 Appearance model evaluation

We have tried to analyze this representation discriminating power through the distribution of the distances we get between a given landmark and other ones from a 150 images database done while the robot wandered around the lab, illustrated through some examples on figure 6.

A poster we find in this sequence has been selected and learned as landmark, and figure 6 now represents the distributions of the distance values we get (a) for the objects not corresponding to this learned landmark and (b) for the objects corresponding to it. We can see on these graphs that classes are well separated. Another point visible in (a) arises from the choice of a partial distance and echoes to what we noticed in 3.2 : distances distribution between sets in $\mathcal{S}$ can be approximated by a Gaussian function, which center and variance depends on the Hausdorff fraction and on the sets cardinals.

Then, figure 7 shows the same kind of histogram of the partial Hausdorff distances between pictures selected over

Figure 6: Distances histogram over real images database

a large synthetic database of 120 movie posters, with some examples of posters on the right. Note that here, all posters are different and that the values are centered on a similar value, around 3.0, and that the variance is comparable to the one in figure 6.

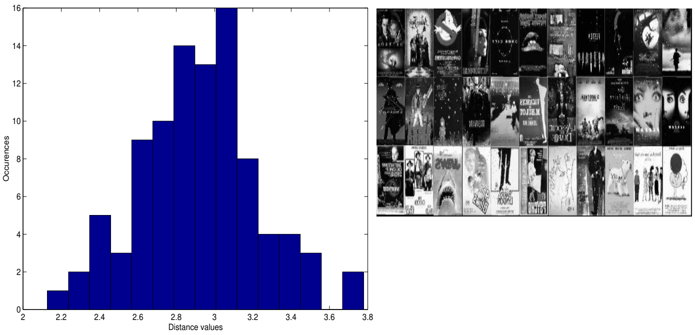

Figure 7: Distances histogram over synthetic database

An important issue we must care about is the way the algorithm behaves with light, scale and perspective changes. Especially, does the Harris detector processing on very low resolution icons, give stable enough representations ? To investigate this problem, we have created a synthetic database of images of different landmarks with different light, scale or perspective conditions. The figure 8 shows some results about it. The graphs correspond to values $r = 0.4$ and $\tau = 1.7$.

The top left graph shows that it is possible to have good recognition results until light saturation appears in the image. On the same way, if we stay under a critical scale corresponding to the one of the model, we can notice on the top right graph, that scale changes do not affect recognition results. As expected, results are deteriorated as soon as the pattern apparent size is below the size of the square we use for representation, i.e. 75.

As far as perspective distortions are concerned, the third graph in figure 8 represents the evolution of the partial Hausdorff distance when performing a planar rotation in the horizontal plane of a quadrangular landmark. It shows that the combination of segments and interest points is a powerful tool to achieve recognition of planar objects, as distances remain reliable up to 75 degrees in this case, a situation that may occur in corridor-like environments. Other studies show that we can consider partial occlusions of the landmark up to 20% of its area, as far as both detection (thanks to RANSAC) and recognition (thanks to the partial Hausdorff distance) are concerned.

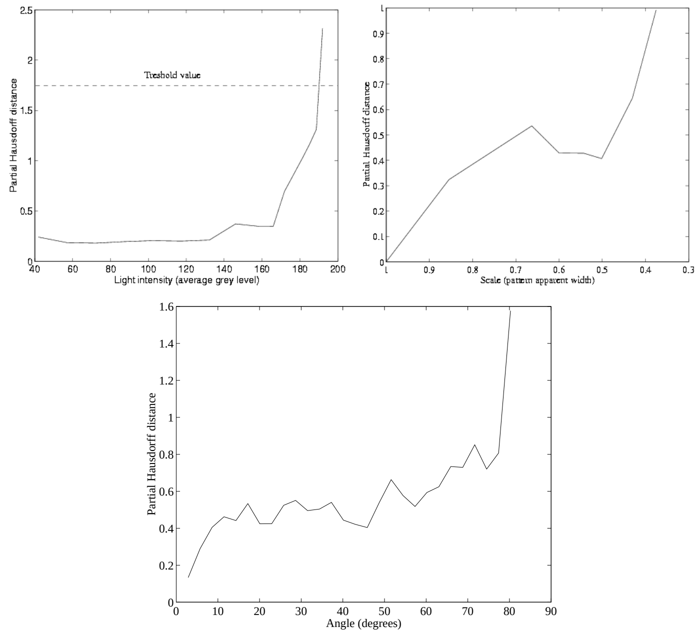

Figure 8: Behavior under variable conditions

4 Application to robot navigation

Our experimental setup consists in a XR4000 Nomadic robot equipped with 2D laser, US and visual sensors mounted on a pan/tilt rig. [7] had presented a full integration of various algorithms for the robot navigation and control under the modular architecture Genom. Localization was achieved thanks to odometry and a laser model-based approach. Such a localization method has very good results provided that enough data is available in. However, in some cases, ambiguities appear inevitably : as an example, the robot may get lost in long corridors, as it has no possibilities to match segments orthogonal to the corridor direction. In such parts of the environment, visual landmark-based navigation may be useful.

4.1 Visibility zones

Our approach has consisted in defining an influence zone in the map for each landmark so that, systematically at this moment, whenever the robot enters such a zone, according to the current localization estimate, it stops and tries to search for the associated landmark. As we may be in situations where the estimates given by odometry and laser localizations are noticeably different from the ground truth, we have implemented a spiral-like systematic search to find the landmark associated to the current visibility zone.

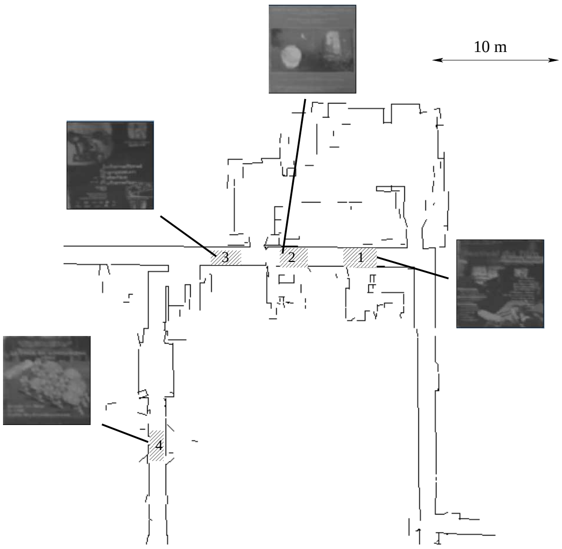

Figure 9: Definition of landmarks zones

4.2 Localization

When the landmark is found, it is used to perform localization. As the landmark model has been defined in the learning phase and the camera is supposed to be fully calibrated, localization on this planar pattern can be performed using well-known methods, as the one in [1]. Uncertainty evaluation is achieved by propagating uncertainties on boundaries detection, on model dimensions and camera calibration through the localization scheme. Computed pose and its uncertainty are then used to warp the continuous localization process based on laser segments.

The robot navigates using these different localization abilities in a corridor-like environment represented in figure 9. Recognition rates are about 80 %, which is important considering the variability in light or pose. Furthermore, uncertainty on localization along corridors is noticeably reduced thanks to visual landmarks, the one on y along the first corridor (landmarks #1,#2,#3) dropping to less than $10cm$ after a landmark is seen, whereas it could climb to $50cm$ otherwise.

5 Discussions and future works

We have presented an original framework for the use of visual landmarks, planar objects in this case. A new recognition method for this kind of landmarks is presented and is used in robot navigation experiments. In this article, localization is performed with a fully calibrated camera. However, considering approaches like [8] and a priori information on the camera and on the scene (no roll, known relative height from the landmark, planar movement . . .), we think we will be able to achieve auto-calibration-and-localization on a single landmark, and perform multi-modal navigation strategies with zooming abilities.

Acknowledgments

Special thanks to Sara Fleury, Benoît Morisset and Thomas Nemeth for their precious help.

References

[1] V. Ayala Ramirez, J.B. Hayet, F. Lerasle, and M. Devy, "Visual localization of a mobile robot in indoor environments using planar landmarks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2000.

[2] P. Ranganathan, J.B.Hayet, M. Devy, S. Hutchinson, and F.Lerasle, "Topological navigation and qualitative localization for indoor environment using multisensory perception," in *9th International Symposium on Intelligent Robotic Systems*, July 2001.

[3] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes," *IEEE Trans. on PAMI*, vol. 5, no. 3, pp. 267–287, May 1983.

[4] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. on PAMI*, , no. 19, pp. 530–535, May 1997.

[5] C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and evaluating interest points," in *International Conference on Computer Vision*, 1998.

[6] D.P. Huttenlocher, A. Klanderman, and J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Trans. on PAMI*, vol. 15, no. 9, 1993.

[7] R. Alami, R. Chatila, S. Fleury, M. Herrb, F. Ingrand, M. Khatib, B.Morisset, P. Moutarlier, and T. Simeon, "Around the lab in 40 days...," in *IEEE International Conference on Robotics and Automation*, San Francisco (USA), April 2000, pp. 88–94.

[8] D. Liebowitz and A. Zisserman, "Combining scene and auto-calibration constraints," in *7th International Conference on Computer Vision*, September 1999.