Self Calibration and 3D Reconstruction from Lines with a Single Translating Camera

E. Thirion and C. Ronse
Universite Louis Pasteur
Departement d'Informatique
7, rue Rene Descartes
67000 Strasbourg - France
thirion@dpt-info.u-strasbg.fr

Abstract

This paper presents a simple method for the 3D reconstruction of a scene with a single translating camera and without calibration. The self-calibration method is based on lines and requires three images of the scene. The reconstruction also includes error bounds on the reconstructed lines.

1 Introduction

The first 3D reconstruction techniques always involved a calibration stage, i.e. an off-line evaluation of the camera parameters from a specially designed scene (a calibration grid for example). In fact these parameters can also be recovered on-line, from the scene which has to be reconstructed. This is sometimes called "self-calibration".

Several types of reconstruction are possible without calibration: euclidean, affine and projective. Euclidean (respectively projective, affine) reconstructions differ from the true reconstruction by an arbitrary euclidean (respectively projective, affine) transformation. Euclidean reconstructions preserve angles, proportions and shape. Affine reconstructions do not preserve shape but they preserve parallelism. Projective reconstructions preserve none of these properties and are the poorest type of reconstructions.

The first investigations in self-calibration have been done in the case of point correspondences. In [4] Faugeras and Maybank showed theorically that when all the cameras have the same intrisic parameters, an euclidean reconstruction is possible with at least three images. For a projective reconstruction, only two images are necessary [3], [9]. The methods of T. Moons [11] and Koenderink [7] produce an affine reconstruction from two views in restricted cases: Koenderink assumes weak perspective effects and Moons supposes a translating camera.

The case of lines has been studied more recently. Up to our knowledge, there is only the method of Hartley which gives a projective reconstruction from at least three images [6], [5] and the method of Quan [8] (not yet published) which

BMVC 1996 doi:10.5244/C.10.59

produces an affine reconstruction under unconstrained camera motion and weak perspective effects (affine camera model).

2 Definition of the problem

Our goal is to get a 3D reconstruction made of 3D line segments with the following hypotheses:

1. We have a single translating camera taking several images of a scene.
2. Each image is segmented into line segments approximating the edges.
3. The correspondence between the segments in each image is given.
4. Nothing is known about the camera.
5. The translations are unknown.

It can be shown that under translation, affine reconstruction is the reachest type of reconstruction achievable. In order to have a self-calibration method which is not perturbated by occultation or over-segmentation problems, we decided to ignore line segment extremities. The determination of the camera parameters (the two translations in our case) is only based on the infinite lines containing the line segments. In this condition we need at least three images. We are now going to show that our goal can be achieved with this minimum number of images.

3 Camera modelisation and terminology

We assume that the camera behaves like a pure perspective projection (pinhole model). The projection is defined by two 3D points: F, O and two non parallel 3D vectors $\vec{I}$, $\vec{J}$. We prefer this representation to the usual matrix representation because it is more explicit and easier to "visualise mentally". The point F represents the focal centre of the camera. The plane containing the point O and parallel to $\vec{I}$ and $\vec{J}$ is the image plane. The triple $(O, \vec{I}, \vec{J})$ is the *image coordinate system.*

Let P be a 3D point and let L be the line containing P and F. The image of a 3D point P is defined as the couple of coordinates (x, y) in the image coordinate system of the point P', intersection of L with the image plane (figure 1). Inversely, given any point (x, y) in the image, there is a unique line L going through F and $P' = O + x\vec{I} + y\vec{J}$. This line is the *interpretation line* of (x, y). It is the set of all 3D points which can have the image (x, y).

The *camera coordinate system* is the coordinate system defined by $(F, \vec{I}, \vec{J}, \vec{K})$, with $K = O - F$. Note that this coordinate system is not necessarily euclidean: the angles between the vectors $\vec{I}, \vec{J}, \vec{K}$ can take any value and their norms do not have to be identical. It is an affine coordinate system.

The coordinates of P' in the camera coordinate system are $(x, y, 1)$. Thus, in $(F, \vec{I}, \vec{J}, \vec{K})$ we have directly the parametric equation of the interpretation line of a point (x, y). It is the set of 3D points $\lambda(x, y, 1)$, for any real number λ. The 3D point P', associated to an image point p, will be denoted by $\tilde{p}$.

The *interpretation plane* of a 2D line l in the image is the set of all the 3D points which have an image in l. Let us consider a line segment $[a, b]$ in l. The interpretation plane of l contains the interpretation line of a and the interpretation line of b (fig. 2). Therefore it is the plane that contains F, $\tilde{a}$ and $\tilde{b}$. Since F is the origin of the camera coordinate system, the equation of the interpretation plane of l in $(F, \vec{I}, \vec{J}, \vec{K})$ is $N.P = 0$ with $N = \tilde{a} \wedge \tilde{b}$. N is the normal of the interpretation plane of l.

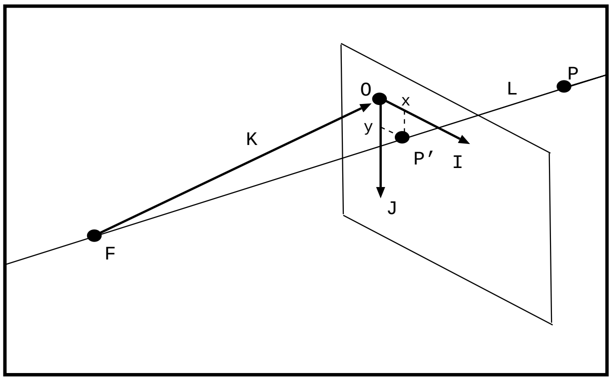

Figure 1: *Camera model*

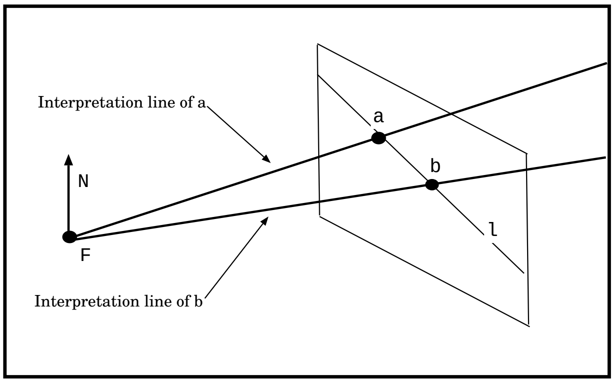

Figure 2: *Interpretation plane of a line*

4 Method

In this section, we give a method for determining the two translations T_1 (from first to second position), T_2 (from first to third position) and the 3D coordinate of the line segments in the camera coordinate system (up to a scale factor).

4.1 Self calibration constraint

Let us consider a correspondence (s, s', s'') where s (respectively s',s'') is a segment from the first (respectively second, third) image. We suppose here that the scene translates instead of the camera. Obviously this does not change anything to the problem.

Let S, S', S'' be the three positions of the 3D line segment corresponding with s (see figure 3). s, s' and s'' are defined in the same image plane. Let L, L', L'' be the infinite 3D lines containing respectively S, S' and S''. Let l,l', l'' be the infinite 2D lines containing respectively s, s' and s''. The normals of the interpretation planes of l, l', l'' are denoted by N,N',N''. Remember (section 3) that we know the coordinates of these vectors in the camera coordinate system: they can be derived from the extremities of s, s' and s''.

For starting, we suppose that the images of L, L' and L'' are exactly l,l' and l''. Note that this allows partial occultation and over-segmentation. The extremities (s, s', s'') do not have to be the projection of the extremities of (S, S', S'').

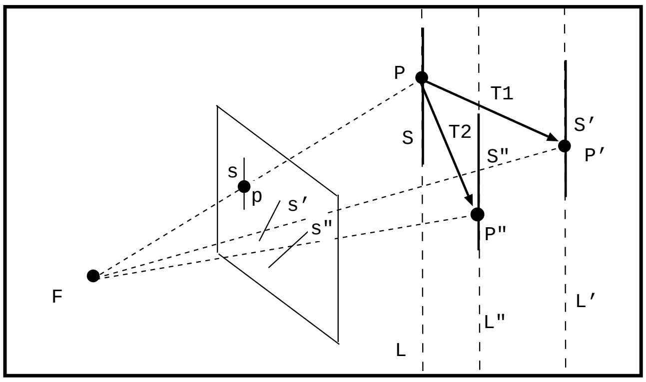

Figure 3: *Self calibration constraint from three segments.*

Let us take any p point in s. This point is the image of a 3D point P in L. P is also in the interpretation line of p. Thus, in the camera coordinate system, $P = \lambda\tilde{p}$. After the first translation, P moves to $P' = P + T_1 = \lambda\tilde{p} + T_1$. P' belongs to L'. Consequently it is in the interpretation plane of l'. This condition can be written as :

$$(\lambda\tilde{p} + T_1).N' = 0 \tag{1}$$

The third position of P is $P'' = P + T_2 = \lambda\tilde{p} + T_2$. This point belongs to L''. Therefore, it is in the interpretation plane of l'' and:

$$(\lambda\tilde{p} + T_2).N'' = 0 \tag{2}$$

From equation 1 we get:

$$\lambda = \frac{-T_1.N'}{\tilde{p}.N'} \tag{3}$$

By replacing λ by this expression in equation 2, we get:

$$-(T_1.N')(\tilde{p}.N'') + (T_2.N'')(\tilde{p}.N') = 0 \tag{4}$$

This is our self calibration constraint from one correspondence. It is a linear constraint on the six dimensional vector $U = (X_1, Y_1, Z_1, X_2, Y_2, Z_2)$ with $T_1 = (X_1, Y_1, Z_1)$ and $T_2 = (X_2, Y_2, Z_2)$. We need at least five of these constraints for determining U up to a scale factor (setting $\|U\| = 1$)

4.2 Determination of the translations

With inexact lines, we can solve the problem by least square. We consider n segment correspondences $(s_i, s_i', s_i''), i = 1, \ldots, n$. p_i is any point in the line l_i containing s_i. N_i' (respectively N_i'') is the normal of the interpretation plane of the line containing s_i' (respectively s_i''). With $n > 5$, we have an overconstrained linear system $AU =$ where A is a matrix of n lines and six columns. Each line a_i of A is the six dimensional vector $[-(\tilde{p}_i.N_i'')N_i', (\tilde{p}_i.N_i')N_i'']$. The "optimal" (statistically) solution is the unit vector U minimizing $\|AU\|^2$. This is the eigen vector associated with the smallest eigen value of the 6×6 symmetric matrix $A^T A$. We use the Jacobi method for solving this problem. Another possibility is to do a singular value decomposition of A and to keep the singular vector associated with the smallest singular value.

4.3 Reconstruction

Once the translations are determined, the reconstruction is very simple. We reconstruct the end points of each line segment $s_i = [a_i, b_i]$ of the first image. The 3D points A_i, B_i associated with a_i, b_i are simply given by:

$$A_i = \lambda_i^a \tilde{a}_i \qquad B_i = \lambda_i^b \tilde{b}_i$$

with

$$\lambda_i^a = \frac{-T_1.N_i'}{\tilde{a}_i.N_i'} \qquad \lambda_i^b = \frac{-T_1.N_i'}{\tilde{b}_i.N_i'}$$

5 Bounding reconstruction errors

It can be useful to know which lines are the most reliable and even better: to bound them by some kind of uncertainty domain.

Such an error bounding technique has been integrated in our affine reconstruction method. In this technique, 3D vectors are bound by convex domains like for example plane sectors (fig.4-A) or solid angles (fig. 4-B).

We suppose that we have a maximal error E_{max} on the position of the 2D line segments. The definition of E_{max} is illustrated by figure 4-A. $[e_1, e_2]$ is a 2D line segment included in a line l. We suppose that the "true" line l^* passes between the two line segments $[e_1^-, e_2^-]$ and $[e_1^+, e_2^+]$ at distance E_{max} from l. Using this hypothesis we can compute, for each line l, a solid angle S_l that bounds the normal of the interpretation plane of l^*. We bound also the interpretation line of

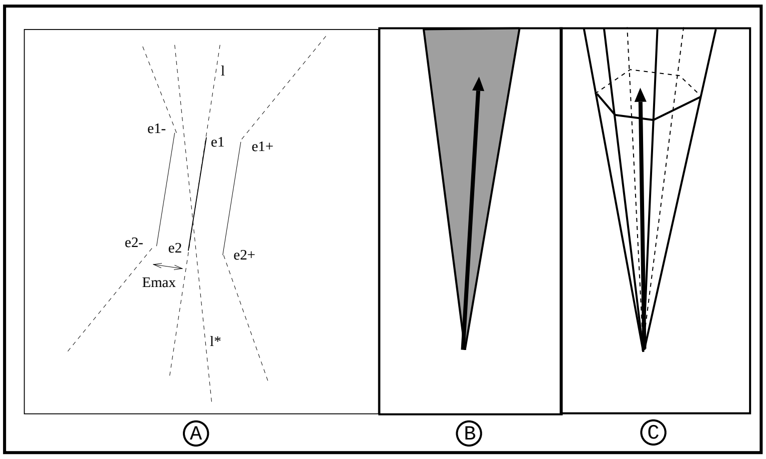

Figure 4: *A) Error bound on a line segment. B) Plane sector. C) Solid Angle*

e_1 (resp. e_2) or in other words, the vector $\tilde{e_1}$ (resp. $\tilde{e_2}$) by a plane sector P_1 (resp. P_2).

The 3D points associated with e_1 and e_2 are given by:

$$E_1 = \lambda_1 \tilde{e_1} \qquad E_2 = \lambda_2 \tilde{e_2} \qquad \lambda_1 = \frac{-T_1.N'}{\tilde{e_1}.N'} \qquad \lambda_2 = \frac{-T_1.N'}{\tilde{e_2}.N'}$$

We suppose also that T_1 is bound by a solid angle S_1. The problem is to find the lower and upper bounds of λ_1 and λ_2, knowing that $T_1 \in S_1$, $\tilde{e_1} \in P_1$, $\tilde{e_2} \in P_2$ and $N' \in S'_l$ (solid angle bounding the normal of the interpretation plane of l'). This problem is solved with procedures computing the extremes values of the scalar product of two uncertain 3D vectors and a procedure computing the bounds of the quotient of two uncertain scalars. These procedures are described in detail in [2] and [1].

6 Results

6.1 Real data

Figure 8 and 9 show two views of a reconstruction from three 256×380 images (fig 5, 6, 7). The object is a dodecahedron with five branch stars drawn on each face. For this example, 61 line correspondences have been entered by hand.

6.2 Statistical evaluation of robustness on simulated data

The graphs presented here show the influence of various parameters on the robustness of the program. For each parameter value, the program was executed E times. The ordinate represents the proportion P of program executions for which a certain level of precision is reached. For simulating the scene N 3D line segments are randomly generated inside of a sphere. The scene is then translated twice in two random directions and the 3D segments are projected onto the image plane. All the 2D segments obtained in this way are then bound by a minimal rectangle representing the image frame. Then noise is added to the 2D line segments by moving slightly their extremities by ϵ pixels (a pixel is the maximum dimension of the image frame divided by 512). More details on the simulation process are given in [1].

Figure 10 shows how much precision is required on the translation estimation in order to get a reconstruction of "reasonable" quality. For each reconstructed point P_i , we measured the relative error on the depth coefficient λ_i. To get the relative error, we divided the absolute error by the difference between the maximum and minimum value of all the λ_i's of the exact reconstruction. The ordinate represents here the proportion of points (over 6000) reconstructed with a relative error not exceeding a threshold. In this experiment $N = 20$,$\epsilon = 0.5$ and $E = 150$. The abscissa is the error on the direction of the first translation T_1. Note that even with a perfect translation, there is still 10 percent of points with a relative error larger than 10 percent. This is due to the presence of unstable lines.

The influence of the noise on the translation estimation is shown in figure 11 ($N = 20$ and $E = 220$). P is the proportion of executions for which the error on the first translation (in degrees) is lower than a threshold. Figure 12 shows that the robustness of the translation estimation increases significantly with the number of lines ($\epsilon = 0.5$, $E = 100$).

We also compare the results with those of our previous method [10] which was restricted too parallel translations. The conclusion is quite interesting. First of all, if we apply the current method with parallel translations, the robustness is slightly lower than the robustness of the previous method . This is shown in figure 13 ($N = 20$, and $E = 100$). In this figure P is the proportion of program executions for which the error on the translation was smaller than two degrees. This phenomenon can be explained by the fact that in the previous method, we had only to estimate two parameters (direction of the translation) instead of five for the current method.

But it is quite surprising to see that, in the case of non parallel translations the results are much better than before. This seems to indicate that translating twice the camera in the same direction leads to a degenerated situation which is numerically unstable. We did another experiment to confirm this. It is shown in figure 14. In this experiment we measured the precision reached for various values of the angle between the two translations (with $\epsilon = 0.5$, $N = 20$, $E = 400$). P is the percentage of program executions for which the error on T_1 is smaller than a threshold. The graph shows clearly that the robustness increases when the angle between the two translations increases also and tends toward 90 degrees.

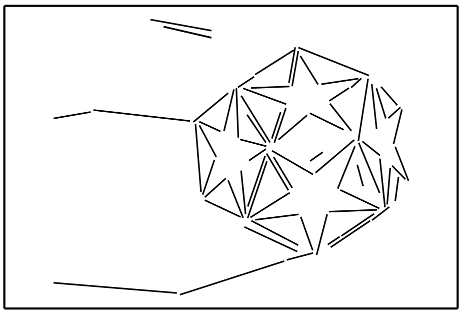

Figure 5: *Segments from first image.*

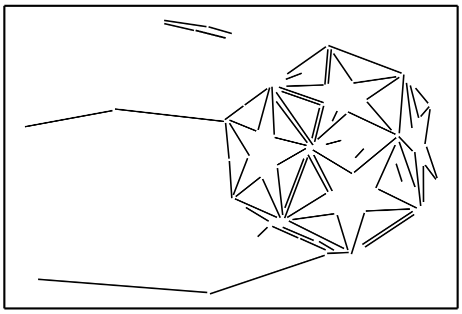

Figure 6: *Segments from second image.*

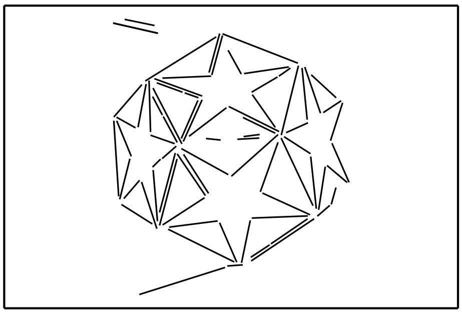

Figure 7: *Segments from third image.*

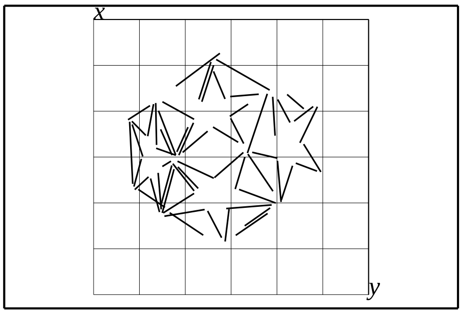

Figure 8: *3D reconstruction, first view.*

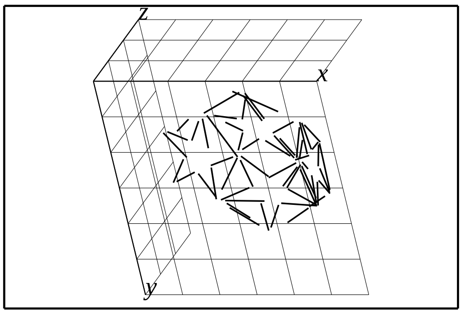

Figure 9: *3D reconstruction, second view.*

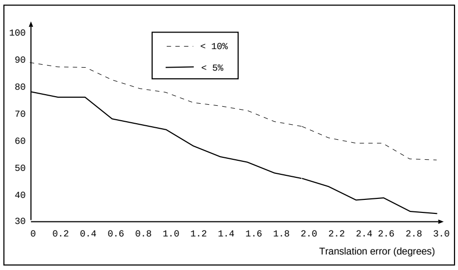

Figure 10: *Influence of translation errors on reconstructions.*

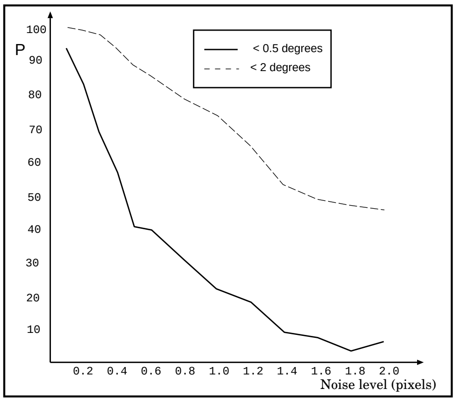

Figure 11: *Noise influence on translation estimation.*

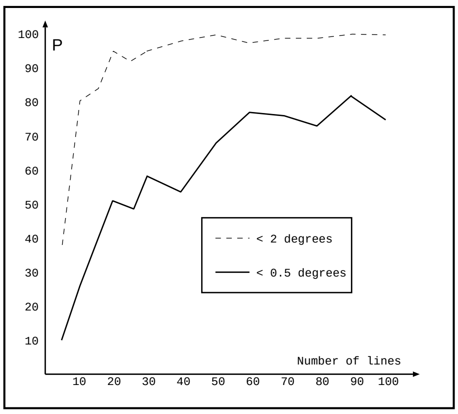

Figure 12: *Influence of the number of lines.*

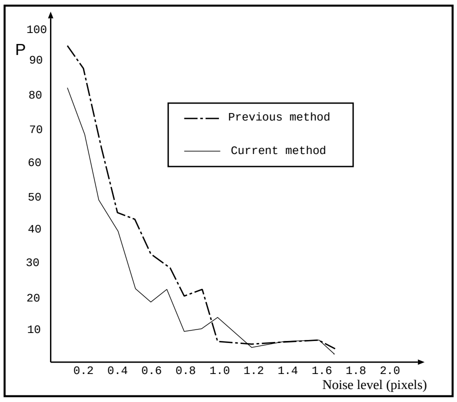

Figure 13: *Noise influence with parallel translations.*

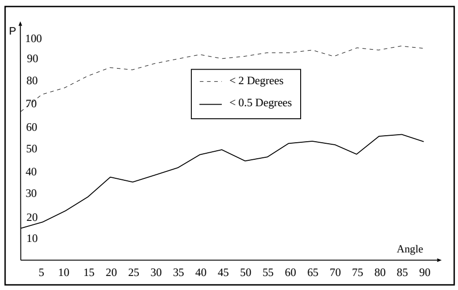

Figure 14: *Influence of the angle between the two translations.*

References

[1] E.Thirion. Affine reconstruction from lines. Technical Report KUL/ESAT/MI2/9415, Katholieke Universiteit Leuven,Departement Elektrotechniek, Kardinaal Mercierlaan 94, B-3001 Heverlee - Belgium, 1994.

[2] E.Thirion. Bounding functions of uncertain scalars, 3d vectors and points. Technical Report KUL/ESAT/MI2/9505, Katholieke Universiteit Leuven,Departement Elektrotechniek, Kardinaal Mercierlaan 94, B-3001 Heverlee - Belgium, 1995.

[3] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceeding of ECCV92*, pages 563–578, 1992.

[4] O.D. Faugeras, Q.T. Luong, and S.J Maybank. Camera self-calibration: Theory and experiments. In *Proceedings of the 2nd European Conference on Computer Vision*, pages 321–334, Santa Margherita, Italy, 1992.

[5] R. I. Hartley. Lines and points in three views - an unified approach. In *ARPA IU Worshop Proceedings*, 1994.

[6] R. I. Hartley. Projective reconstruction from line correspondences. In *Proceeding of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 903–907, 1994.

[7] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, 1989.

[8] L. Quan. A factorisation method for affine structure from line correspondence. In *To appear in Proceedings of Computer Vision and Pattern Recognition*, 1995.

[9] R.Mohr, L.Quan, F. Veillon, and B. Boufama. Relative 3d reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG-12, LIFIA-IMAG, Grenoble, 1983.

[10] E. Thirion, T. Moons, and L. Van Gool. Affine reconstruction from lines. In *Proceedings of the 6th British Machine Vision Conference-Vol. 1*, pages 267–276, Birmingham, 1995.

[11] T.Moons, L. Van Gool, M. Van Dienst, and E.Pauwels. Affine structure from perspective image pairs under relative translations between object and camera. Technical Report 9306, KUL/ESAT/MI2, Katholieke Universiteit Leuven, Departement Elektrotechniek, Kardinaal Mercierlaan 94, B-3001 Heverlee - Belgium, 1993.