INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

3-D Scene Representation as a Collection of Images and Fundamental Matrices

Stéphane Laveau, Olivier Faugeras

N° 2205

Février 1994

PROGRAMME 4

Robotique,

image

et vision

Rapport de recherche

1994

ISSN 0249-6399

3-D Scene Representation as a Collection of Images and Fundamental Matrices

Stéphane Laveau, Olivier Faugeras

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n°2205 — Février 1994 — 25 pages

Abstract: In this report, we address the problem of the prediction of new views of a given scene from existing weakly or fully calibrated views called *reference views*. Our method does not make use of a three-dimensional model of the scene, but of the existing relations between the images. The new views are represented in the reference views by a viewpoint and a retinal plane, i.e. by four points which can be chosen interactively. From this representation and from the constraints between the images, we derive an algorithm to predict the new views. We discuss the advantages of this method compared to the commonly used scheme : 3-D reconstruction-projection. We show some experimental results with synthetic and real data.

Key-words: 3-D scene representation, multi-view stereo, image synthesis

(Résumé : tsvp)

This work was partially supported by DRET contract No 91-815/DRET/EAR and by the EEC under Esprit project 6448, Viva

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Représentation d'une scène 3-D à l'aide d'une collection d'images et de matrices fondamentales

Résumé : Dans ce rapport, nous étudions le problème de la prédiction de nouvelles vues d'une scène donnée à partir de vues faiblement ou fortement calibrées que nous appellerons *vues de référence*. Notre méthode n'utilise pas le modèle tridimensionnel de la scène, mais les relations existantes entre les différentes images. La nouvelle vue est représentée dans les vues de référence par un point de vue et un plan rétinien, c'est à dire par 4 points qui peuvent être choisis interactivement. A partir de cette représentation et des contraintes entre les images, nous avons construit un algorithme pour prédire de nouvelles vues. Nous discutons les avantages et les inconvénients par rapport au schéma classique: reconstruction 3-D puis projection. Nous présentons des résultats expérimentaux avec des images réelles et synthétiques.

Mots-clé : Représentation des données 3-D, Stéréo multi-vues, Synthèse d'images

1 Introduction

The problem we solve in this paper is the following. Suppose we are given N views of a static scene obtained from different viewpoints, perhaps with different cameras. These viewpoints we call *reference viewpoints* since they are all we know of the scene. We would like to decide if it is possible to predict another view of the scene taken by a camera from a viewpoint which is arbitrary and a priori different from all the reference viewpoints. One method for doing this would be to use these viewpoints to construct a three-dimensional representation of the scene and reproject this representation on the retinal plane of the virtual camera. In order to achieve this goal, we would have to establish some sort of calibration of our system of cameras, fuse the three-dimensional representations obtained from, say, pairs of cameras [2, 26] thereby obtaining a set of 3-D points, the scene. We would then have to approximate this set of points by surfaces, a segmentation problem which is still mostly unsolved, and then intersect the optical rays from the virtual camera with these surfaces. This is the most straightforward way of going from a set of images to a new image using the current computer vision paradigm of first building a three-dimensional representation of the environment from which the rest is derived. We do not claim that there does not exist any simpler way of using the three-dimensional representation than the one we just sketched, but this is just simply not our point.

Our point is that it is possible to avoid entirely the explicit three-dimensional reconstruction process: the scene is represented by its images and by some basically linear relations that govern the way points can be put in correspondence between views when they are the images of the same scene-point. These images and their algebraic relations are all we need for predicting a new image. This approach is similar in spirit to the one that has been used in trinocular stereo. Hypotheses of correspondences between two of the images are used to predict features in the third. These predictions can then be checked to validate or invalidate the initial correspondence. This approach has proved to be quite efficient and accurate [11, 17, 19, 20, 1, 22, 13]. Related to these ideas are those developed in the photogrammetric community under the name of *transfer* methods which find for one or more image points in a given image set, the corresponding points in some new image set. If the camera geometries are known, transfer is done in a straightforward fashion by 3-D reconstruction and reprojection. If the camera geometries are unknown, this can still be done by methods based on using projective invariants [3]. As a third source of correspondence, people interested in the recognition and pose estimation of three-dimensional objects have recognized more recently that the variety of views of the same rigid object under rigid transformations can be expressed as the combinations of a small number of views [25]. The connection between the two problems has only been realized even more recently [4] and in a rather incomplete way even though the central role played by the epipolar geometry of the set of cameras has been acknowledged.

We start with the analysis of the case of two views which are used to predict a third one. The main difference with the trinocular stereo case, for example, is that the user actually chooses the new viewpoint and the direction of the retinal plane in the two views. Once this has been done, the new image of the scene can be constructed with very few further assumptions, for example that the epipolar geometry between the two reference views is known. This is now known as the weak calibration situation [5]. We show that under this assumption, the new image can be constructed up to an unknown projective transformation of the image plane. If some more information about the cameras is known, for example if we assume that our system is fully calibrated (thus allowing in theory a three-dimensional metric reconstruction of the scene, which will not be used directly), then the new image can be constructed as in the standard scheme of reconstruction-projection.

2 The approach

In this section we explain the principle of the method.

2.1 Background and notations

We make heavy use of elementary projective geometry. The reader who is unfamiliar with these notions is referred to the recent computer vision literature on the subject [18, 5, 6, 12]. Given a pair of images noted 1 and 2, we call them weakly calibrated when the epipolar geometry between the images is known. This means that the epipolar geometry of these two views, considered as a stereo pair, is known. Therefore, given a pixel m^1 in the first view, its epipolar line $l^2_{m^1}$ in the second view can be computed. This is done through a special 3×3 matrix, $\mathbf{F}_{12}$, called the fundamental matrix, which has been described at great length in previous publications [8]. The epipolar line of pixel m^1 is represented by the 3×1 vector $\mathbf{F}_{12}\mathbf{m}^1$, where $\mathbf{m}^1$ is a projective representation (i.e. a 3×1 vector) of the pixel m^1. The fundamental matrix is known to be of rank 2 and, being defined up to a scale factor, depends upon 7 parameters. These parameters have a very simple geometric interpretation. Four of them are the coordinates of the two epipoles e_{12} and e_{21} which are represented by vectors in the null-space of $\mathbf{F}_{12}$ and $\mathbf{F}_{21} = \mathbf{F}_{12}^T$:

$$\mathbf{F}_{12}\mathbf{e}_{12} = \mathbf{F}_{12}^T\mathbf{e}_{21} = \mathbf{0}$$

The three remaining parameters describe the collineation between the two pencils of epipolar lines. If (m^1, m^2) is a pair of corresponding points in images 1 and 2, they satisfy the epipolar constraint:

$$\mathbf{m}^{2T}\mathbf{F}\mathbf{m}^1 = 0$$

When we consider $N > 2$ views, we use $\mathbf{F}_{ij}$ to denote the fundamental matrix between views i and j (in that order). e_{ij} (resp. e_{ji}) is the epipole in view i (resp. in view j) with respect to view j (resp. to view i).

In the case where the intrinsic parameters of the two views are known [9, 24, 16] (case that we call the strong-calibration case), the fundamental matrix is simply the essential matrix $\mathbf{E}$ introduced by Longuet-Higgins [14] which depends only upon the 5 parameters describing the relative displacement between the two views (the translation being, as it is well-known, defined only up to a scale factor).

Another important and well-known idea that will be heavily used in the remaining of the paper is the fact that planes induce homographies between images. This means that given two views 1 and 2 and m^1 and m^2 be the images of a point in a fixed plane in the scene, then the relation between m^1 and m^2 is a homography that depends only on the plane and the relative camera geometry. We note $\mathbf{H}_{12}$ the 3×3 matrix describing this homography and we have $\mathbf{m}^2 = \mathbf{H}_{12}\mathbf{m}^1$. Also, clearly, $\mathbf{H}_{21} = \mathbf{H}_{12}^{-1}$.

2.2 Two reference images

Let us consider first the case where two views are available. Since the epipolar constraint is the most important constraint used in many existing stereo algorithms, the knowledge of the fundamental matrix is sufficient to run most of them [7] and obtain a set of point correspondences between the two views. In the weak-calibration case, it is known that these correspondences allow us to reconstruct the scene up to an unknown three-dimensional projective transformation. In the strong-calibration case, the scene can be reconstructed without projective distortion. We stress again that in any case we do not need to reconstruct explicitly the scene in 3-D.

We can set up the third view by selecting two corresponding points e_{13} and e_{23} in the two images. These points may or may not be the images of a real point of the scene, but they must satisfy the epipolar constraint that each point belongs to the epipolar line of the other. Cross-eye stereo fusion of the two images, when possible, may help in selecting the point. We say that this point is the new point of view, it is the optical center of the virtual camera that will be used to predict the new image. In the weakly-calibrated situation, the notion of perpendicularity does not exist and we cannot define the retinal plane as perpendicular to the viewing direction and must therefore define this plane directly from the two views. Since a plane is defined by three points, we can select them interactively like we did for the optical center, making sure that the three pairs of corresponding points each satisfy the epipolar constraint. This is shown in figure 1 in which the three corresponding pairs are denoted by $(m_i^1, m_i^2),\ i = 1, 2, 3$. The upper indexes refer to the view number.

These three pairs of points, plus a fourth one which is determined as explained in Appendix A will be considered as being the images of a projective basis of the retinal plane of the virtual camera. We call them control points. This fourth point can be chosen arbitrarily as long as it is not aligned with two of the previous three. These choices having been made, we are ready to construct the new image. In order to do this, we need the following two ingredients:

- A set of point correspondences between the two reference views, the denser the better. These correspondences can be represented as a function d_{12} from the first view to the second. This function called the disparity

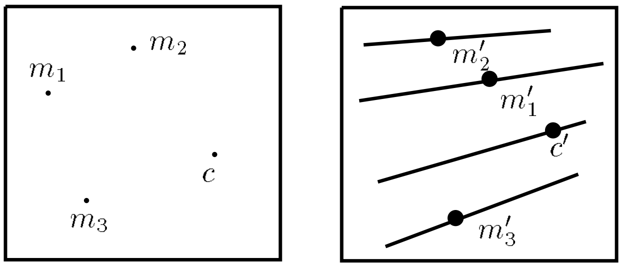

Figure 1: Choosing the new viewpoint and the new retinal plane.

function is such that if pixel m^1 in view 1 has been matched to pixel m^2 in view 2, then $m^2 = d_{12}(m^1)$.

- A way to compute the intersection of the optical rays of the virtual camera (i.e. the lines joining its optical center to a point in the scene) with its retinal plane.

The first ingredient is obtained through the use of standard (binocular in this case) stereo algorithms, for example the one described in [7]. The reason for the second ingredient should be pretty clear by now. For each point correspondence (m^1, m^2) between the two reference views, we consider the two image lines $\langle e_{13}, m^1 \rangle$ and $\langle e_{23}, m^2 \rangle$. They are the images of the optical ray from the virtual optical center to the scene point whose images are m^1 and m^2. Therefore, we need to compute the intersection of this optical ray with the retinal plane which was defined previously. Since we do not want to reconstruct anything in three dimensions, this computation must be done in the reference images. This problem has been solved by several authors [21, 23]. More precisely, if p^1 (resp. p^2) is the image of the point of intersection of the optical ray defined by the two image lines $\langle e_{13}, m^1 \rangle$ and $\langle e_{23}, m^2 \rangle$ with the retinal plane $\mathcal{R}$ of the virtual camera, p^1 (resp p^2) can be computed as the intersection of $\langle e_{13}, m^1 \rangle$ (resp $\langle e_{23}, m^2 \rangle$) and $\mathbf{H}_{21}^T.\langle e_{23}, m^2 \rangle$ (resp $\mathbf{H}_{12}^T.\langle e_{13}, m^1 \rangle$), $\mathbf{H}_{12}$ being the homography defined by the 4 pairs of corresponding points $(m_1^1, m_2^1, m_3^1, m_4^1)$ and $(m_1^2, m_2^2, m_3^2, m_4^2)$.

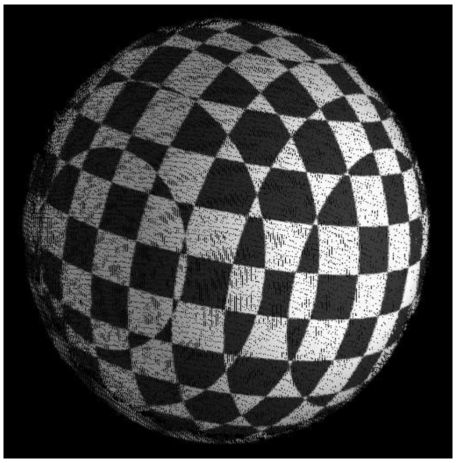

Figure 2: Defaults of the straightforward algorithm

Having built these points, their projective coordinates in the projective basis formed by the four reference points can be read directly from the reference images. The projective coordinates of p^1 in the projective basis $(m_1^1, m_2^1, m_3^1, m_4^1)$ are the same as those of p^2 in the projective basis $(m_1^2, m_2^2, m_3^2, m_4^2)$[1]. This allows us to construct a collection of points in a projective plane defined by the four reference points: the virtual retinal plane. An intensity can be given to these points by combining the intensities of the two corresponding points in the reference images, this is our predicted image. If the choice of the four points has been made in an arbitrary fashion in the retinal plane, it is seen that the image is obtained as a distortion of the "real" image by an arbitrary and unknown planar projective transformation.

It turns out that a straightforward implementation of these ideas does not work well because of the appearance of gaps in the predicted image (see Figure 2) caused by the irregular distribution in 3 of the pixels in 1 and 2 as seen from the new viewpoint. This is why we develop in the next section an alternative approach based on the same principles.

[1] For completeness, the epipoles e_{31} and e_{32} can be found very simply: To determine them is equivalent to determining their image in the reference views 1 and 2. The image of e_{31} in 1 (of e_{32} in 2) is by definition e_{13} (e_{12}). As a consequence, the image of e_{31} in 1 (of e_{32} in 2) is $\mathbf{H}_{12}\mathbf{e}_{13}$ ($\mathbf{H}_{21}\mathbf{e}_{23}$)

As a summary of this section, we can say that in the case of two reference views, the three-dimensional scene is represented by these two views and the fundamental matrix describing the epipolar geometry between these views. This information allows us to predict any view of the scene as seen from any viewpoint without explicitly reconstructing the scene in 3-D. A natural question to ask next is, what is the situation when more than two reference views are available? We answer this question in the next section.

2.3 More than two reference views

We now assume that we are given $N > 2$ reference views and the fundamental matrixes between any two of these views. There are a priori $\frac{N(N-1)}{2}$ such independent matrices and since each matrix depends upon 7 parameters, $O(N^2)$ independent parameters are necessary to represent the full epipolar geometry of these N reference views. It has been shown elsewhere [15] that there are in fact only $18 + 11(N-3)$ of them which are independent and from which the others can be computed.

The procedure for predicting a new view from a viewpoint which is different from the existing N is then very similar to the two views case and is explained in Figure 3. This figure represents the case $N = 3$, double arrows between the retinal planes indicate that the epipolar geometry is known. The images e_{14} and e_{24} of the new viewpoint have been chosen in the first two reference views as described in the previous section, as well as the three pairs of corresponding points (m_i^1, m_i^2), $i = 1, 2, 3$ defining the new retinal plane. The epipolar geometry between the reference views 1 and 3, and 2 and 3 being known, we can in fact propagate this information very simply. For example, the image e_{34} of the new viewpoint is obtained by intersecting the epipolar lines $l^3_{e_{14}}$ of e_{14} and $l^3_{e_{24}}$ of e_{24} in the third reference view. Similar computations can be done for the points m_i^1, m_i^2, as well as for the fourth point defining the projective frame in the new retinal plane.

2.4 Strongly calibrated images

In this paragraph, we assume that our system is fully calibrated. The algorithm described previously is still valid. But since we know the internal parameters

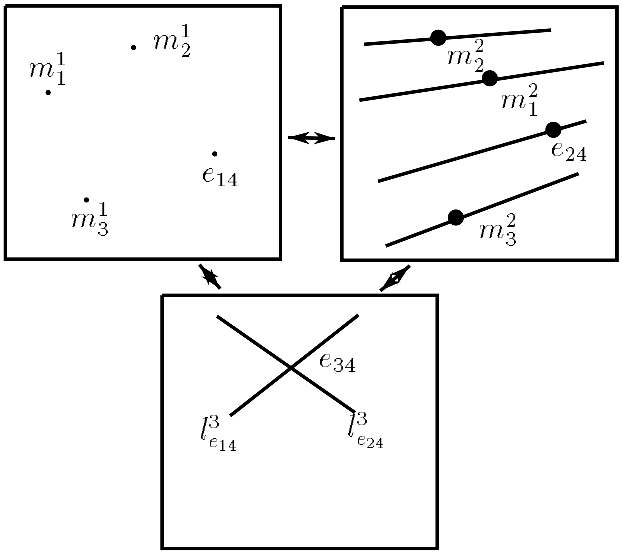

Figure 3: Propagating the viewpoint and the retinal plane among the reference views.

of the cameras, we can use them to eliminate the unknown projective transformation mentioned previously. First, the epipoles of the new camera in the reference views must be the images of the optical center. Second, we choose the four points of the projective basis as the images of the four corners of the desired image in the retinal plane. It is obvious that once we know the 3-D coordinates of our 5 control points, we can project them into every other camera. Let P be the projection matrix of the camera n. We then need the points corresponding to $(0,0)$,$(511,0)$,$(0,511)$ and $(511,511)$ in the plane defined by $s=\mu$, μ being an arbitrary distance between the optical center and the retinal plane. Let us consider the matrix A such that:

$$A = \begin{pmatrix} & & P & \\ & & & \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

then we have:

$$\begin{pmatrix} sx \\ sy \\ s \\ 1 \end{pmatrix} = A \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Therefore, we can get $(X,Y,Z)^T$ by multiplying A^{-1} with $(sx,sy,s,1)^T$. s will be meaningful only if the norm of the 3 first numbers of the last row is normalized to one. The optical center is easily obtained by setting s to zero.

It should be noted that the retinal plane is only defined by its normal (the direction of viewing). The choice of a distance μ between the retinal plane and the optical center will not affect the final appearance of the new image[2].

A very interesting point of this method compared with the reconstruction is that we need only two strongly calibrated images. The images of the control points in every other reference views can be inferred from the weak calibration[3].

[2]We must still bear in mind that only the points in front of the retinal plane are seen.

[3]It is consistent with the conclusion of [15].

Figure 4: Defaults due to the trifocal plane

2.5 In the tri-focal plane

The trifocal plane is defined as the plane passing through the optical center of three cameras. For every trifocal plane containing the optical center of the new camera, this algorithm fails: The epipolar line created by the first image and the epipolar line created by the second image are undiscernable. We cannot therefore get a precise positioning by intersecting them. It is not in general a problem, because the trifocal plane does not intersects the images. Figure 4 shows such a problem.

3 Implementation of the ray-tracing like algorithm

In this section we describe how we apply the relationships described in the previous section. First, we limited ourselves to the case of the synthesis of one view from two other reference views. We refer to the new image as 3 and to the known images as 1 and 2. $\mathcal{R}_3$ is the retinal plane of 3. We then extend the ideas to the prediction for an arbitrary number of views N ($N > 2$).

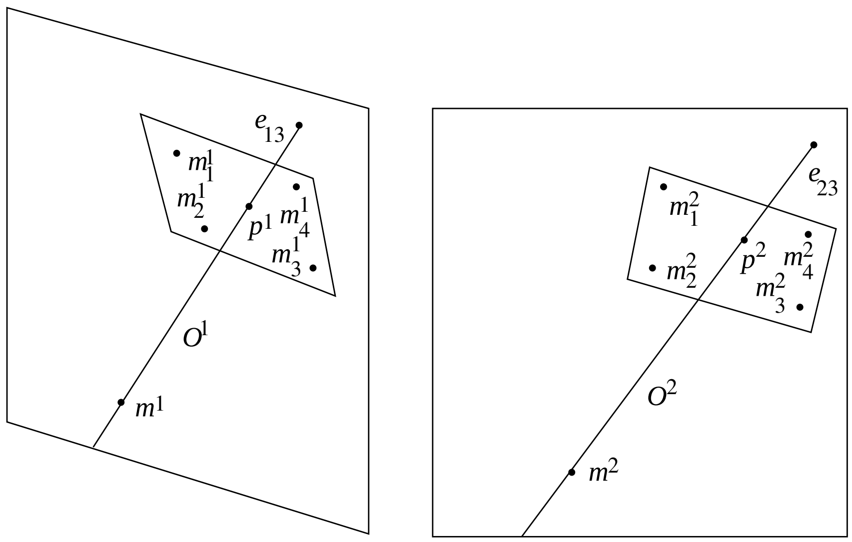

Figure 5: Transfer: information available in the reference views..

The solution to the problem discussed in section 2.2 is well known in image synthesis: the scanning must take place directly in the target image. The ray-tracing methods follow the optical ray coming from a pixel into the world instead of projecting a world point onto the image. It is slightly more complicated in our case, since the direct prediction (as in section 2.2) involves a projection onto $\mathcal{R}_3$ which is not bijective.

Given a point m^3 in $\mathcal{R}_3$, we can draw p^1 (see Figure 5) and p^2 in $\mathcal{R}_1$ and $\mathcal{R}_2$ since they have the same projective coordinates. It is equivalent to say that they are images by an homography (unknown in the weakly calibrated case, known in the strongly calibrated). The epipoles e_{13} and e_{23} are given by the user. The projections of the optical ray $\langle C_3, m^3\rangle$ in $\mathcal{R}_1$ and $\mathcal{R}_2$ (O^1 and O^2) are $\langle p^1, e_{13}\rangle$ and $\langle p^2, e_{23}\rangle$.

The problem now is first to find where the scene points are located on the optical ray from the virtual camera seen as O^1 and O^2 and, second, among the possibly several points, which one leads to the correct interpretation, i.e. is the closest to the virtual camera (we assume objects are opaque).

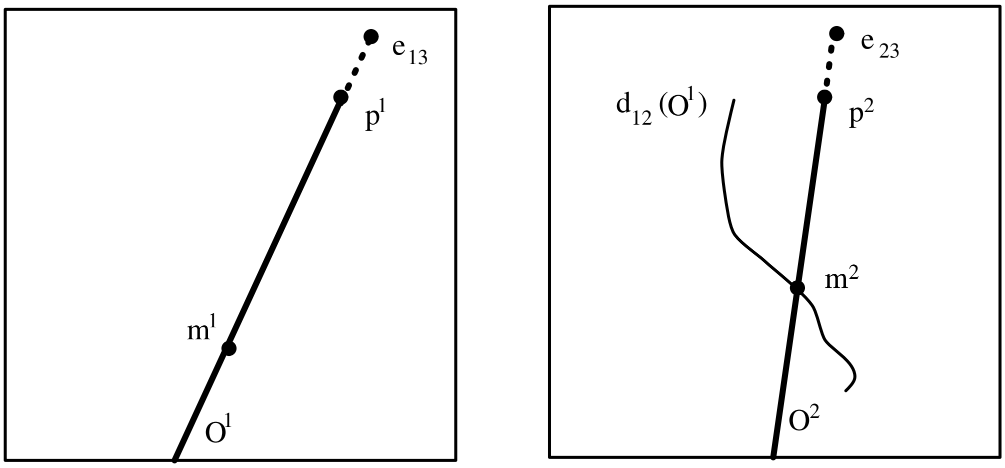

Figure 6: Determination of m^1 and m^2.

3.1 Physical points

If $m^1 \in O^1$ is the image of a physical point M, there are constraints on its correspondent m^2 in $\mathcal{R}_2$. First, m^2 lies on O^2 because M belongs to the line $\langle C_3, m^3 \rangle$. Second, M being a physical point, m^2 lies on $d_{12}(O^1)$ the image of O^1 by the disparity map. $d_{12}(O^1)$ is in general not a line, but a curve. $d_{12}(O^1)$ and O^2 can meet in more than one point. m^2 is one of the intersections between $d_{12}(O^1)$ and O^2 (see Figure 6).

3.2 Disambiguating whenever possible

- If $d_{12}(O^1)$ and O^2 do not meet, it means that the point is occluded in either 1 or 2 or is not correlated and therefore no information can be obtained.
- If $d_{12}(O^1)$ and O^2 meet once, there is no ambiguity.
- If $d_{12}(O^1)$ and O^2 meet in more than one point, the optical ray intersects an object in the physical world in more than one point too (see figure 7).

We want to sort the corresponding points M_1, M_2 with respect to the distance to the optical center C_{N+1} in the world from the information contained

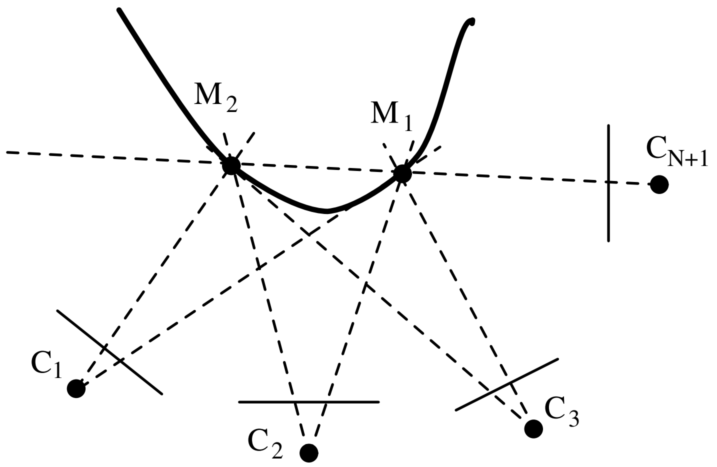

Figure 7: Ambiguities: multiple solutions..

in the reference views. Let us choose arbitrarily a view i in which M_1 and M_2 are seen. They are physical points, therefore they are in front of the focal plane $\mathcal{F}_i$ of this reference view. If we know whether C_{N+1} is in front of $\mathcal{F}_i$, we are able to decide on the order of the points in the world. If C_{N+1} is in front of the focal plane, the order of M_1, M_2 and C_{N+1} will be preserved by perspective projection, whereas it will be inverted if C_{N+1} is behind F_i.

If we are fully calibrated, We know the 3-D position of C_{N+1} and $\mathcal{F}_i$. If we are weakly calibrated, but we have identified the plane at infinity, that is to say, we know the set of vanishing points, then we can determine the vanishing point v_∞ of the line passing through C_{N+1}, M_1 and M_2. Then, if v_∞ is between C_{N+1} and M_1 and M_2, C_{N+1} is behind the focal plane, otherwise C_{N+1} is in front of the focal plane[4]. The knowledge of the position of C_{N+1} with respect to the focal plane of the current image can easily propagate through the sequence. It is enough that one ray intersects the object in 2 different points. If the order of the points on the image changed, so did the position of C_{N+1} with respect to the focal plane. If none of the above is applicable, we are not helpless. The focal plane position does not change for every pixel, but only for every image. therefore, we get at most 2 possible images. In the case of a sequence where

[4]The presence of C_{N+1} in the focal plane could be detected by an epipole at infinity

no propagation is possible, we have 2^{N-1} possible images. This extremal case is very improbable.

This reasoning seems to be somehow related to the difficult problems raised in [10].

Note that it is always possible to disambiguate in the strongly calibrated case.

3.3 Generalization to the case of an arbitrary number of views

Suppose now that we are given N views of the scene and wish to predict an $(N+1)$st view. The user chooses the control points in say views 1 and 2. We know from section 2.3 that the knowledge of the fundamental matrices enables us to compute the control points for every other image. We estimate then for each pair of images. It would probably be better to estimate the whole epipolar geometry of the sequence in one pass, since this take all the constraints into account and therefore give us better results.

The algorithm proceeds as follows:

- For each image i do:
  - Compute the disparity function $d_{i\,i+1}$ between views i and $i+1$.
  - From $\mathbf{F}_{i-1\,i+1}$ and $\mathbf{F}_{i\,i+1}$ compute the control points of image $i+1$ from the control points in image i and $i-1$.
- For each pixel m^{N+1} in image $N+1$ do:
  - Compute O_i in every image i.
  - Iteratively, scan O_i and its image in $i+1$ to find the physical points M on the optical ray. If possible, disambiguate.

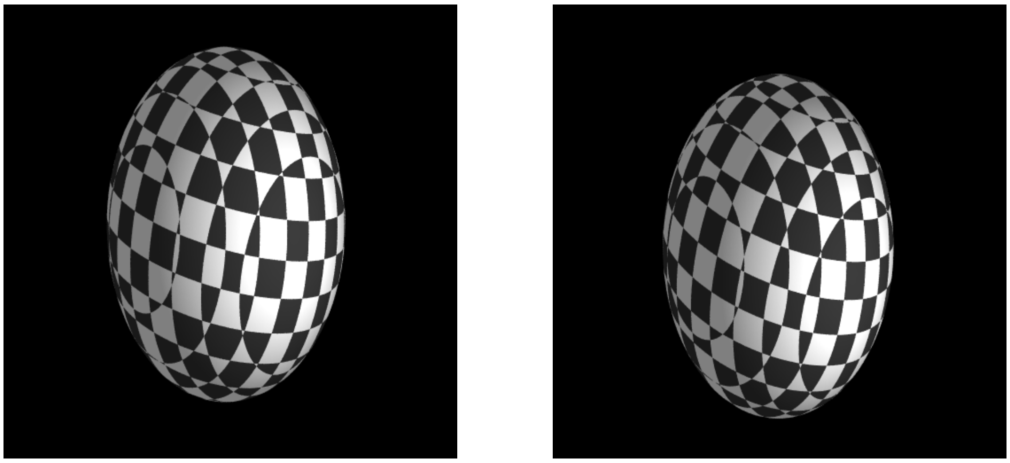

Figure 8: Some reference views synthesized by *rayshade*

4 Experimental Results

4.1 Comparison between the transferred image and the projection of a 3D reconstruction

In order to validate our algorithms and their accuracy, we used *rayshade*[5] to synthesize five views of the same object (a checkered ellipsoid). We took four of these views (two are shown in Figure 8) as reference views for the computation to the fifth. The results are presented in Figure 9. The correspondences between the points in the reference views were extracted by correlation.

The imperfection of the edges of the sphere is caused by uncorrelated points at the boundaries of the object. The prediction is very accurate: the curves drawn on the ellipsoid present the same curvature, and the shape is very well preserved. The error on the positioning of the lines' intersections is on the order of the pixel.

[5] *rayshade* is a public domain ray-tracer

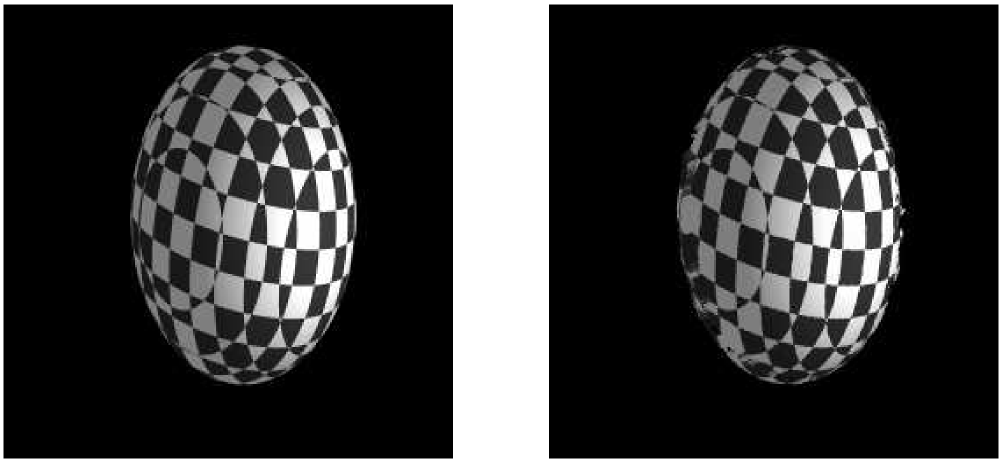

Figure 9: Left: view synthesized by *rayshade*, Right: view predicted by transfer.

4.2 Comparison between the transferred image and a reprojected image

We used a strongly calibrated pair of images as reference views (Figure 10). The transferred image and the reconstruction of the third image through a 3-D model are shown in figure 11.

The missing parts are uniquely uncorrelated areas. If there is no 3-D information given by the point correspondences, we cannot be able to predict the appearance of the object neither by transfer nor by reconstruction. The 3-D representation we use consists of a set of triangles drawn between neighboring points. This representation leads to some false interpolation between non-adjacent points. The image obtained by transfer does not present any false interpolation between the data because we are not restricted by a given representation of 3-D objects.

4.3 From a very different viewpoint

We took 2 front images of our mannequin (Figure 12) and we predicted what a side view would be (Figure 13). The angle of rotation between the reference views and the predicted view is 70 degrees.

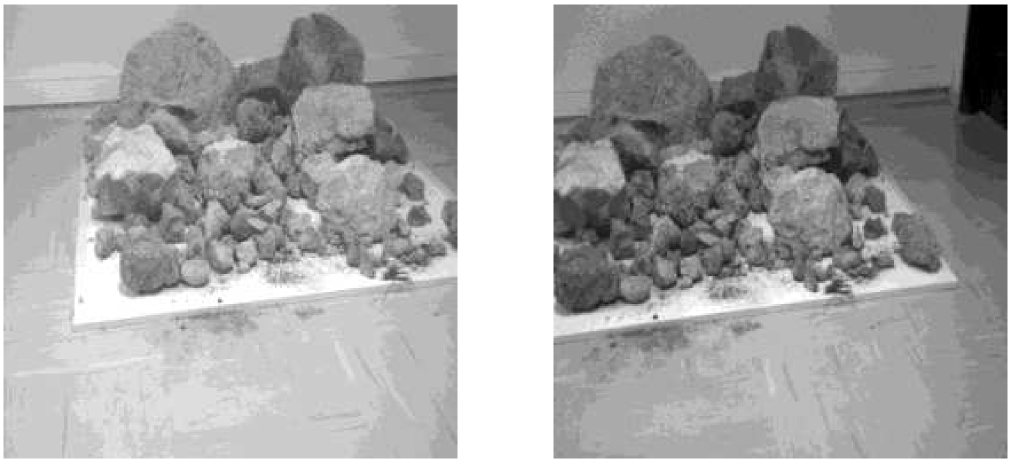

Figure 10: Reference views.

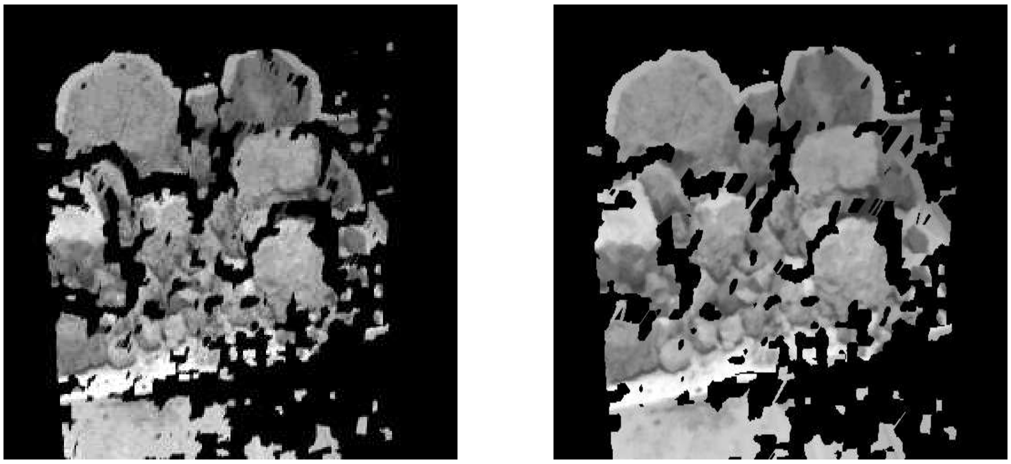

Figure 11: Predicted image by transfer (left) and 3-D reconstruction-projection (right).

Figure 12: Face images of the mannequin

Figure 13: Predicted side view of the mannequin

The occlusions are well dealt with as can be seen on the left breast and on the neck. The shape does not present any irregularities. The strip on the right is an error due to false matches given by our correlation. We are currently working on improving our correlation algorithms to avoid these false matches. Of course, the unseen areas in the reference views are not visible in the transferred image (the right breast, the right part of the neck)

5 Conclusion

We have proposed a method for representing a 3-D scene which does not involve an explicit reconstruction. It rather considers the scene as a collection of images related by simple algebraic relations. We have shown elsewhere [5] that these relations allow us to compute 3-D information about the scene. We show here that they allow the prediction of an image of the scene from an arbitrary viewpoint in a fairly efficient manner. We believe that representing 3-D data as images could be as powerful as using a complete 3-D model.

One advantage over existing methods of reconstruction and projection is that we do not need calibration for all reference views, but only for two of them. No complicated triangulation is needed either.

In the future, we plan to investigate further this representation for image compression and synthesis. We also want to integrate a more sophisticated reflectance model.

A Construction of a fourth point

Since the property of being collinear is preserved by perspective projection, it can be checked directly from the two reference images. We thus choose a point m_4^1 in the first reference image not aligned with any pair of the other three points. We want to find the corresponding point in the second reference image such that the pair represents a point in the retinal plane of the virtual camera. This corresponding point must be on the epipolar line $l_{m_4}^2$ determined by the point chosen in the first reference image. If we consider the point of intersection m^1 of two of the diagonals of the quadrilateral formed by the four points in the first reference image, because the four points are coplanar,

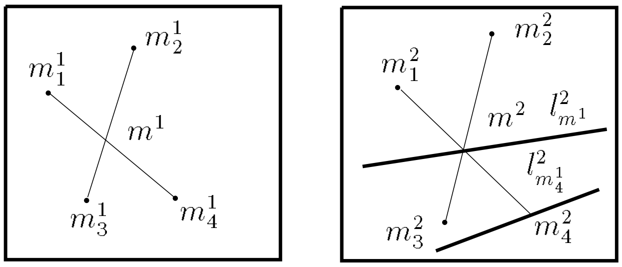

Figure 14: Constructing a fourth point in the new retinal plane.

this point is the image of a real point in the retinal plane and therefore the epipolar constraint can be used to construct its correspondent in the second reference image as shown in figure 14: m^2 is at the intersection of the epipolar line $l^2_{m^1}$ of m^1 with the image line $\langle m_2^2, m_3^2 \rangle$. m_4^2 is then obtained as the point of intersection of $\langle m_1^2, m^2 \rangle$ and the epipolar line $l^2_{m_4^1}$ of m_4^1 $\square$

References

[1] N. Ayache and F. Lustman. Fast and Reliable Trinocular Stereovision. In *Proceedings of the International Conference on Computer Vision*, pages 422–427. IEEE, June 1987.

[2] Nicholas Ayache and Olivier D. Faugeras. Maintaining Representations of the Environment of a Mobile Robot. *IEEE Transactions on Robotics and Automation*, 5(6):804–819, December 1989. also INRIA report 789.

[3] Eamon B. Barett, Michael H. Brill, Nils N. Haag, and Paul M. Payton. *Invariant Linear Methods in Photogrammetry and Model-Matching*, chapter 14. MIT Press, 1992.

[4] Ronen Basri. On the Uniqueness of Correspondence under Orthographic and Perspective Projections. In *Proc. of The Image Understanding Workshop*, pages 875–884, 1993.

[5] O.D. Faugeras. What Can be Seen in Three Dimensions with an Uncalibrated Stereo Rig ? In Giulio Sandini, editor, *Proc. European Conference on Computer Vision*, pages 563–578, Santa Margherita Ligure, Italy, 1992. Springer-Verlag.

[6] O.D. Faugeras. *Three-Dimensional Computer Vision: a geometric viewpoint.* MIT Press, 1993.

[7] Olivier D. Faugeras, Bernard Hotz, Hervé Mathieu, Thierry Viéville, Zhengyou Zhang, Pascal Fua, Eric Théron, Laurent Moll, Gérard Berry, Jean Vuillemin, Patrice Bertin, and Catherine Proy. Real Time Correlation Based Stereo: Algorithm, Implementations and Applications. *The International Journal of Computer Vision*, 1993. Submitted.

[8] Olivier D. Faugeras, Tuan Luong, and Steven Maybank. Camera Self-Calibration: Theory and Experiments. In Giulio Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 321–334. Springer-Verlag, Lecture Notes in Computer Science 588, 1992.

[9] Olivier D. Faugeras and Giorgio Toscani. The Calibration Problem for Stereo. In *Proceedings CVPR '86, Miami Beach, Florida*, pages 15–20. IEEE, June 1986.

[10] R. I. Hartley. Cheirality Invariants. In *Proc. DARPA Image Understanding Workshop*, pages 745–753, Washington, DC, April 1993.

[11] M. Ito and A. Ishii. Three-View Stereo Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:524–532, 1986.

[12] Kenichi Kanatani. Computational Projective Geometry. *CVGIP: Image Understanding*, 54(3):333–348, November 1991.

[13] Yoshifumi Kitamura and Masahiko Yachida. Three-Dimensional Data Acquisition by Trinocular Vision. *Advanced Robotics*, 4(1):29–42, 1990. Robotics Society of Japan.

[14] H.C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, 293:133–135, 1981.

[15] Q.-T. Luong and T. Viéville. Canonic Representations for the Geometries of Multiple Projective Views. In *3rd E.C.C.V., Stockholm*, 1994.

[16] S.J. Maybank and O.D. Faugeras. A Theory of Self-Calibration of a Moving Camera. *The International Journal of Computer Vision*, 8(2):123–152, August 1992.

[17] V.J. Milenkovic and T. Kanade. Trinocular Vision Using Photometric and Edge Orientation Constraints. In *Proceedings of DARPA Image Understanding Workshop*, pages 163–175, , December 1985.

[18] Joseph L. Mundy and Andrew Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[19] M. Pietikainen and D. Harwood. Depth from Three-Camera Stereo. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 2–8. IEEE, 1986. Miami Beach, Florida.

[20] M. Pietikainen and D. Harwood. Progress in Trinocular Stereo. In *Proceedings NATO Advanced Workshop on Real-time Object and Environment Measurement and Classification, Maratea, Italy*, August 31 - September 3 1987.

[21] L. Robert and O.D. Faugeras. Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair. In *Proc. First International Conference on Computer Vision*, pages 540–544, Berlin, Germany, May 1993.

[22] Luc Robert and Olivier D. Faugeras. Curve-Based Stereo: Figural Continuity And Curvature. In *CVPR91*, pages 57–62. IEEE, June 1991. Maui, Hawai.

[23] Amnon Shashua. Projective Depth: A Geometric Invariant for 3D Reconstruction From Two Perspective\Orthographic Views and For Visual Recognition. In *Proc. First International Conference on Computer Vision*, pages 583–590, 1993.

[24] Roger Tsai. An Efficient and Accurate Camera Calibration Technique for 3-D Machine Vision. In IEEE, editor, *Proceedings of the International Conference on Computer Vision And Pattern Recognition*, pages 364–374, Miami Beach, Florida, June 1986.

[25] Shimon Ullman and Ronen Basri. Recognition by Linear Combinations of Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):992–1006, 1991.

[26] Zhengyou Zhang and Olivier Faugeras. A 3D World Model Builder with a Mobile Robot. *International Journal of Robotics Research*, 11(4):269–285, August 1992.

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399