Trifocal tensor as a tool for modeling an imperceptible structured light sensor[a)]

T. Molinier,[b)] D. Fofi,[c)] F. Meriaudeau, and R. Seulin

Le2i UMR CNRS 5158—IUT Le Creusot, Le Creusot, France
(Submitted September 5, 2006)
Opticheskiĭ Zhurnal **74**, 68–74 (April 2007)

This article presents the first results for the estimation of the trifocal tensor which defines the relations of correspondence between the points pertaining to three image planes. This estimation is achieved for a trinocular sensor based on imperceptible structured light. From the trifocal tensor, projective matrices can be estimated enabling the projective reconstruction of the scene. Moreover, the transfer relations permit easy solution of the correspondence problem. In addition, it is shown that the tensor provides geometrical tools for self-calibration of the sensor. © *2007 Optical Society of America*.

I. INTRODUCTION

The trifocal tensor estimation is a key point in trinocular vision. It is the equivalent of the fundamental matrix[7] for two views. This article reviews some of the well-known techniques used to estimate it. These methods have been programmed and their accuracy analyzed on synthetic images. With the goal of applying trifocal tensor estimation to imperceptible structured light,[1,6,9,10] explanations on how to extract useful information from the tensor in order to perform a complete and automatic 3D reconstruction are given. Also we will show that the tensor contains enough information to both map textures onto the reconstructed surfaces and to get the required information for self-calibration from an imperceptible structured light sensor.

The article is structured as follows: imperceptible structured light, its concepts and principles are explained in the second Section. The trifocal tensor, the estimation of the projective matrices and the point-transfer method are described in the third part of the paper. Results of computations are given and commented in the third part; finally, the paper ends with conclusions and future work.

II. IMPERCEPTIBLE STRUCTURED LIGHT

A. Principle and setup

Imperceptible structured light systems are composed of a unique light source and two cameras (Fig. 1, left). The light source successively projects onto the scene a light pattern and its complement (inverse pattern) at a high frequency rate, resulting in a uniform pattern. The first camera, synchronized with the projection of the first pattern, permits reconstruction of the scene thanks to the capabilities of structured light vision; the second camera, which has a longer exposure time, observes the scene under uniform lighting (as a result of pattern and complement projection) and acquires a classical gray-level or colored image ready for processing (Fig. 1, right).

The coded pattern used was originally proposed by Salvi *et al.*[12] It is composed by a set of vertical and horizontal slits, uniquely color-encoded in a single pattern projection. Six well-spaced colors in the HSI cone have been used. Red, green and blue have been chosen to code the horizontal slits; cyan, magenta and yellow to code the vertical ones. The codification, similar to the one used by Griffin *et al.*,[2] ensures that each slit color with its two neighbor slit colors forms a triplet that exists only once in the whole pattern. Slits are regularly spaced, leading to a regular grid of 29×29 cross-points. The resolution may be increased by incrementing the number of color primitives.

B. Advantages and drawbacks

A vision sensor based on imperceptible structured light permits one to have the advantages of both a structured light vision system (easy correspondence and segmentation, 3D reconstruction) and a traditional vision system (texture and color analysis), and hence it enables one to get a 3D reconstruction of the observed scene and to map colors and textures onto the reconstructed surfaces.

Moreover, this sensor is non-invasive for flicker frequencies higher than the *critical flicker frequency* (defined as the highest frequency at which a person can detect a flicker in a flickering light source; see Ref. 5 for further explanation). Many applications, such as robotics, anthropometrics, metrology, profilometry, and industrial inspection, could benefit from this kind of sensor.

On the other hand, this sensor has some drawbacks:

- it is impossible to find any points of correspondence (Fig. 2, left) between the projected pattern and camera 2, which perceives the scene under a uniform light, nor between camera 1 and camera 2;
- the sensor is made up of heterogeneous components (cameras and projector) with different intrinsic parameters;
- a movement of the projector produces a sliding of the pattern on the measured surfaces (and thus a loss of the 3D points): reconstruction is therefore limited to three views (two images and one pattern).

III. THE TRIFOCAL TENSOR

A. Overview

The epipolar geometry (Fig. 2, right) describes the relations of correspondence between projections on several image planes. The trifocal tensor is a $3 \times 3 \times 3$ cube which describes the epipolar geometry for three views as the fun-

 1070-9762/2007/040282-06$15.00

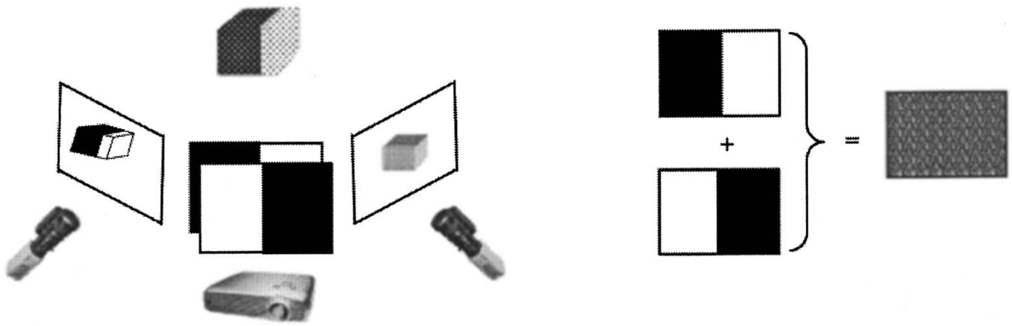

FIG. 1. Principle of imperceptible structured light. (Left) Trinocular sensor based on imperceptible structured light. The pattern is visible to one camera but not to the other nor to the human eye. (Right) Integration over time of the initial pattern and its complement due to high frequency projection.

damental matrix does it for two. It is composed of 27 elements but has only 18 degrees of freedom because it has to satisfy internal constraints. The trifocal tensor permits linking lines and points from one view to two other views: 3 points together, 3 lines together, 2 points and 1 line, 2 lines and 1 point.

Figure 2 (right) depicts how from two correspondences of two different image planes, the two corresponding epipolar lines on the third image plane can be estimated and the coordinates of the third corresponding point computed. A trifocal tensor can be defined by the expression:

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k. \qquad (1)$$

With the three 3×4 camera matrices: $\mathbf{M}=[\mathbf{I}_{3\times 3}|0_{3\times 1}]$, $\mathbf{M}'=[a_i^j]$ and $\mathbf{M}''=[b_i^j]$, where $\mathbf{I}_{3\times 3}$ is the identity matrix (representing the rotational part of the motion) and $0_{3\times 1}$ a null translation vector, i.e., the first camera is placed at the origin of the coordinate frame. Note that, for an imperceptible structured light sensor, there are two camera matrices and one projector matrix. But it is well known that a projector can be seen as a camera acting in reverse by inverting the line of sight and can thus be modeled in the same way a camera is.[12]

The reader interested in trifocal geometry can refer to Ressl[11] for a wide and complete overview.

B. Constraints and properties of the trifocal tensor

The major difficulty lies in computing a *homogeneous* tensor, i.e., a tensor that respects its 18 degrees of freedom. Besides, in order to compute a robust tensor, some other constraints have to be satisfied. This Section presents and discusses these constraints.

As said before, the trifocal tensor is composed of 27 entries but has only 18 degrees of freedom. Consequently 9 constraints must be satisfied by these 27 elements to obtain a valid trifocal tensor. One of these constraints is to fix the value of the norm of the homogeneous tensor (for example, the norm of Frobenius to 1) and the tensor only needs to satisfy 8 more constraints. These constraints are particularly significant when the tensor is computed starting from correspondences of point and/or line from three views. In recent years, many studies have been undertaken to find these constraints. Some of the constraints are not complete (<8), some others are not minimal (>8), and others are rather complicated. All the constraints are related to the properties of the tensor slices (a tensor slice is a 3×3 matrix extracted from the tensor). Papadopoulo and Faugeras[8] proposed the first set of constraints which completely characterizes the trifocal tensor. It is based on a nonminimal set of dependent equations which can be detailed as:

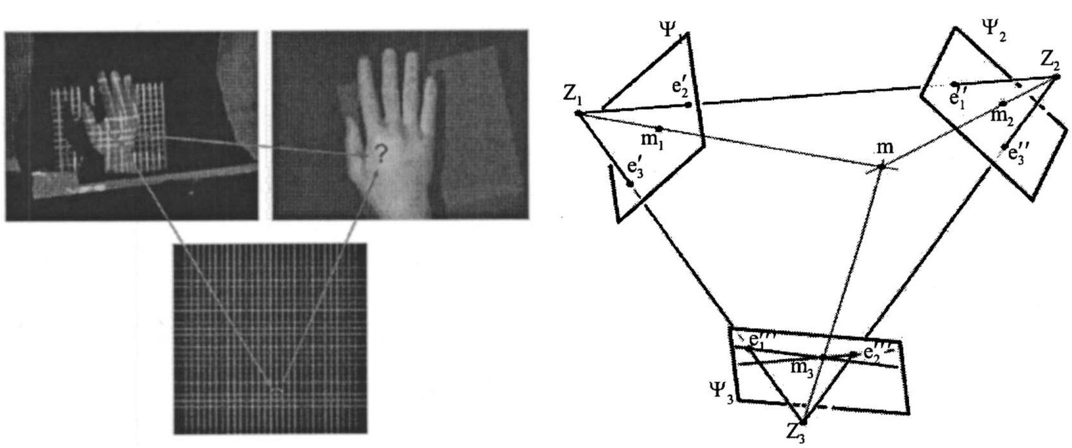

FIG. 2. Left: the correspondence problem. Right: Trifocal geometry, e are the epipoles, Z the projection center, and Ψ the image planes.

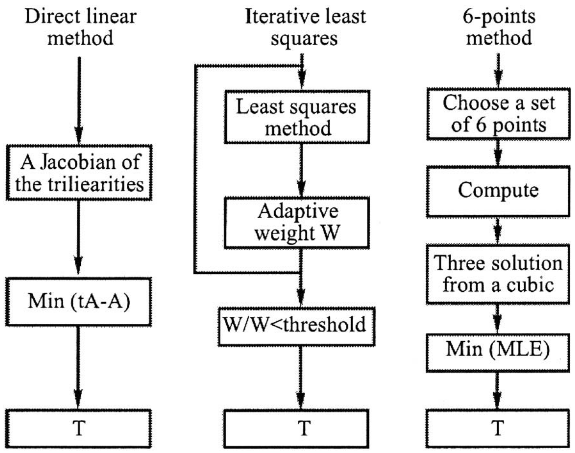

FIG. 3. Methods of estimation of the trifocal tensor.

the three rank constraints:

$$\det(\mathbf{J}_i) = 0, \tag{2}$$

the two epipolar constraints:

$$\det(\mathbf{N}_R) = \det(\mathbf{N}_L) = \mathbf{0}, \tag{3}$$

for $i=1$ to 3. $\mathbf{N}_R$ and $\mathbf{N}_L$ are the right and left null-spaces of $\mathbf{J}_i$;the ten extended rank constraints:

$$\mathrm{rank}\left(\sum_{i=1}^{3} x_i \mathbf{J}_i\right) \leq 2 \tag{4}$$

for all x_i, $i=1,2,3$.

Here $\mathbf{J}_i$ corresponds to the correlation slices of the tensor which are obtained by fixing the i index of the tensor.

C. Three methods of estimation

Part of our work has consisted in implementing and evaluating three methods of estimation: a direct linear method,[11] an iterative least-squares method[3] and the 6-point method[13]—see Fig. 3. Each one will be detailed in the following subsections.

Direct linear method. This method is very simple, but its simplicity raises its imperfection. The elements of the tensor must satisfy 8 internal constraints to represent a valid trifocal tensor. These constraints are usually not satisfied by the direct linear solution. Another drawback is that the direct linear solution does not reduce the errors in the original measurements of point and line, also called errors of reprojection or residual errors.

The equation of the direct linear solution is

$$\mathbf{A}\mathbf{t} = 0. \tag{5}$$

where $\mathbf{t}$ is a 27-vector given as:

$$\mathbf{t} = \begin{pmatrix} vec(\mathbf{J}_1) \\ vec(\mathbf{J}_2) \\ vec(\mathbf{J}_3) \end{pmatrix} \text{ and}$$

$$\mathbf{A} = (\mathbf{S}^{\mathrm{red}}(\mathbf{x}'') \otimes \mathbf{S}^{\mathrm{red}}(\mathbf{x}'''))(\mathbf{x}'^T \otimes \mathbf{I}_{9\times 9}),$$

$\otimes$ is the symbol of the Kronecker product, $\mathbf{S}^{\mathrm{red}} = \begin{bmatrix} \mathbf{0} & \mathbf{-1} & \mathbf{y} \\ \mathbf{1} & \mathbf{0} & \mathbf{x} \end{bmatrix}$ is the reduced axiator and the operator $vec(\,)$ is defined as follows: let $\mathbf{A}$ be an $m\times n$ matrix, the operation $vec(\mathbf{A})$ returns a $mn\times 1$-vector by stacking together the columns of $\mathbf{A}$; $\mathbf{x}'$, $\mathbf{x}''$ and $\mathbf{x}'''$ are three corresponding points.

The solution is given by a 27-element vector ($\mathbf{t}$ is the vector of tensor slices).

Iterative least-squares method. This method is based on the estimation by a least-squares method adjusted by a weighted variable at each iteration.[3]

From a 3-vector $\mathbf{x}_1=[x_1 y_1 z_1]^T$ representing a point, the tensor is estimated by the following equations:

$$x_3(x_2\alpha_{33}^T - z_2\alpha_{13}^T)\cdot \mathbf{x}_1 = z_3(x_2\alpha_{31}^T - z_2\alpha_{11}^T)\cdot \mathbf{x}_1$$

$$y_3(x_2\alpha_{33}^T - z_2\alpha_{13}^T)\cdot \mathbf{x}_1 = z_3(x_2\alpha_{32}^T - z_2\alpha_{12}^T)\cdot \mathbf{x}_1$$

$$x_3(y_2\alpha_{33}^T - z_2\alpha_{23}^T)\cdot \mathbf{x}_1 = z_3(y_2\alpha_{31}^T - z_2\alpha_{21}^T)\cdot \mathbf{x}_1$$

$$y_3(y_2\alpha_{33}^T - z_2\alpha_{23}^T)\cdot \mathbf{x}_1 = z_3(y_2\alpha_{32}^T - z_2\alpha_{22}^T)\cdot \mathbf{x}_1, \tag{6}$$

where α_{ij} is a 3-vector and the tensor is computed from: $\alpha_{ijk}=-T_i^{jk}$; $\mathbf{x}_1$, $\mathbf{x}_2=[x_2 y_2 z_2]^T$ and $\mathbf{x}_3=[x_3 y_3 z_3]^T$ are corresponding points.

To solve this set of equations, at least 7 points are required. The iteration is done by an adaptive weight, specific for each of the four equations. For the first equation, the formula of the weight is given by:

$$W = \frac{1}{(x_2\alpha_{33}^T - \alpha_{13}^T)\cdot \mathbf{x}_1}. \tag{7}$$

The 6-point method. The 6-point method[13] requires the minimum set of points to compute the tensor. A maximum likelihood estimation (MLE) method is used to choose the tensor that minimizes the residual error. It must be remembered that this method also allows detection of the outliers; therefore a tensor computed with an outlier will not be selected by the MLE.

The method is based on a basis transformation: the six points are used to create a canonical basis in 3D space and its projections in the three planes. The equation which comes out from this basis change is a quadratic equation that gives 3 solutions. Each set of six points gives three equations, and the MLE is used to determine the best solution, which minimizes the residual error. The equation of the MLE is:

$$D_M = \sum_{ij} (\bar{x}_i^j - x_i^j)^2. \tag{8}$$

where $\bar{x}_i^j$ and x_i^j are, respectively, the estimated and real coordinates of the point.

D. Point-transfer method

Let us consider that a point $\mathbf{x}$ in space is seen in three images through the three camera matrices $\mathbf{M}=[I_{3\times 3}|0_{3\times 1}]$, $\mathbf{M}'=[a_i^j]$, and $\mathbf{M}''=[b_i^j]$, giving the three image points $\mathbf{u}$

$=[u^1 u^2 u^3]^T$, $\mathbf{u}'$ and $\mathbf{u}''$, respectively, so that $\mathbf{u}=\mathbf{M}\mathbf{x}$, $\mathbf{u}'=\mathbf{M}'\mathbf{x}$, and $\mathbf{u}''=\mathbf{M}''\mathbf{x}$ (note that $\mathbf{x}$, $\mathbf{u}$, $\mathbf{u}'$, and $\mathbf{u}''$ are expressed in homogeneous coordinates).

Because of the form of $\mathbf{M}$, we may write $\mathbf{x}=[\mathbf{u}\ \mathbf{t}]^T$, with unknown t. Projecting this point onto the second image plane gives (the $\approx$ sign denotes equality up to an unknown scale factor)

$$u'^i \approx a^i_k u^k + a^i_4 t. \tag{9}$$

This is obtained because the last column of $\mathbf{M}'$ (but also of $\mathbf{M}$ and $\mathbf{M}''$) represents the translation from the origin and the 3×3 left matrix corresponds to the rotation.

Rearranging this equation,[4] the scale factor can be eliminated, and point $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ u^k(u'^i a^j_k - u'^j a^i_k)/u'^j a^i_4 - u'^i a^j_4 \end{bmatrix} \approx \begin{bmatrix} (u'^j a^i_4 - u'^j a^j_4)\cdot \mathbf{u} \\ u^k(u'^i a^j_k - u'^j a^i_k) \end{bmatrix}. \tag{10}$$

Now, projecting this point through the third camera gives

$$\begin{aligned} u''^l &\approx b^l_k u^k(u'^j a^i_4 - u'^i a^j_4) + b^l_4 u^k(u'^i a^j_k - u'^j a^i_k) \\ &\approx u^k u'^i(a^j_k b^l_4 - a^j_4 b^l_k) - u^k u'^j(a^i_k b^l_4 - a^i_4 b^l_k) \\ &\approx u^k(u'^i T^{jl}_k - u'^j T^{il}_k). \end{aligned} \tag{11}$$

Then, after eliminating the scale factor implied by $\approx$, we finally obtain

$$u^k(u'^i u''^m T^{jl}_k - u'^j u''^m T^{il}_k - u'^i u''^l T^{jm}_k + u'^j u''^l T^{im}_k) = 0^{ijlm}. \tag{12}$$

By knowing the trifocal tensor and the correspondence between two image planes, it is thus possible to determine the coordinates of the corresponding point within the third image plane.

The reader who is unfamiliar with projective geometry or tensor notation can refer to Refs. 3 and 4.

E. Practical implementation

Not a single point of correspondence can be determined between the images of the projector and camera 2 by a common use of the sensor. For the estimation of the trifocal tensor and to have a good precision on the point correspondence of the three planes, the two cameras must see the coded pattern (the projector is switched on, without flickering). Once the correspondence is solved, the estimation of the sensor can be achieved directly.

A homogeneous trifocal tensor is very useful because it contains the projection matrices and all the epipolar geometry of the system. It is possible to extract the projection matrices, the fundamental matrices, and the epipoles from the trifocal tensor. It also permits one to solve the correspondence problem between points, whereas the fundamental matrix does so only between points and lines.

IV. EXPERIMENTAL RESULTS

The three previously exposed methods have been programmed with Matlab and their accuracy analyzed based on synthetic images to which different noise levels were added, as well as on real images.

Figure 4 (left) demonstrates that without noise all the methods obviously give perfect results, but both the linear and the iterative methods are shown to be extremely sensitive to noise.

A. Real images

Our tensor estimation methods were also applied to some real images. The correspondence points were detected manually, the noise is therefore relatively important. The set of constraints to ensure a valid trifocal tensor is the set created by Papadopoulo and Faugeras.[8]

Some results were confirmed: the weakness of the direct linear method, since the reprojection error is not minimized, but also the sensitivity of the iterative method, which made both of them unsuitable for our purpose.

Table I gathers the results; "Head," "Duck," "Kermit," and "Geometrical planes" refer to images of real objects.

The 6-point method gives homogenous tensor and quite accurate results (note that for "Geometrical planes" image, the 6-point method failed as it requires at least 4 non-coplanar points).

B. Transfer relation

If the tensor is known, it can be used to estimate the pixel on the third image when two corresponding points are known on the two other images. In this Section we have tested the transfer relation extracted from the estimated tri

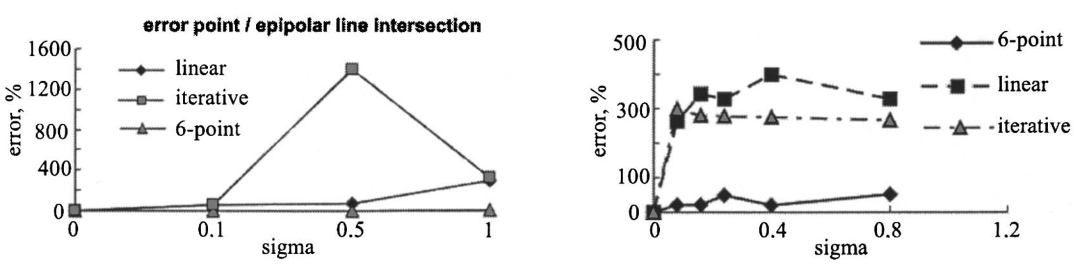

FIG. 4. Left: distance between intersection of epipolar lines and real point. Right: Difference between real points and transferred points, estimated by the transfer relation.

TABLE I. Results of homogeneous tensor.

	Head	Duck	Kermit	Planes
Linear method	No	No	No	No
Iterative method	No	No	No	No
6-points method	Yes	Yes	Yes	

focal tensor. The third pixel coordinates are estimated from the epipolar lines as depicted in Fig. 3.

Figure 4 (right) shows the results performed on synthetic data with varying Gaussian noise level. As expected, the direct linear method and the iterative one are not accurate at all. However, the 6-point method gave satisfactory results and enough accuracy to consider its use in order to solve the correspondence problem.

Figure 5 shows the results of transferred points based on the 6-point method. A point was selected on the projected pattern and its correspondent point on the image 1. We compute the third correspondent point on the second image by means of the transfer relation.

V. CONCLUSIONS

Trifocal tensor is a mathematically elegant method to model a trinocular sensor which unfortunately is very sensitive to noise. In this paper, three methods to estimate the trifocal tensor have been implemented and evaluated. As our aim was to model an imperceptible structured light sensor, and also to use the transfer relations principle to solve the correspondence problem for which a robust method was required. Experiments show that only an estimation that satisfies the internal constraints of the tensor and which has sufficient robustness against noise and outliers is usable.

We are currently investigating a RANSAC method to further improve the results in terms of accuracy. It is based on the 6-point method, as it takes the minimum set of points to calculate the tensor, then it uses the maximum likelihood estimation (MLE) as a selection criterion to select the tensor which minimizes the errors of reprojection. It is necessary to recall that this method also allows detecting the outliers.

Future work will deal with self-calibration of an imperceptible structured light sensor based on the geometrical information contained in the trifocal tensor: extract projection matrices, perform a projective reconstruction, and upgrade it,

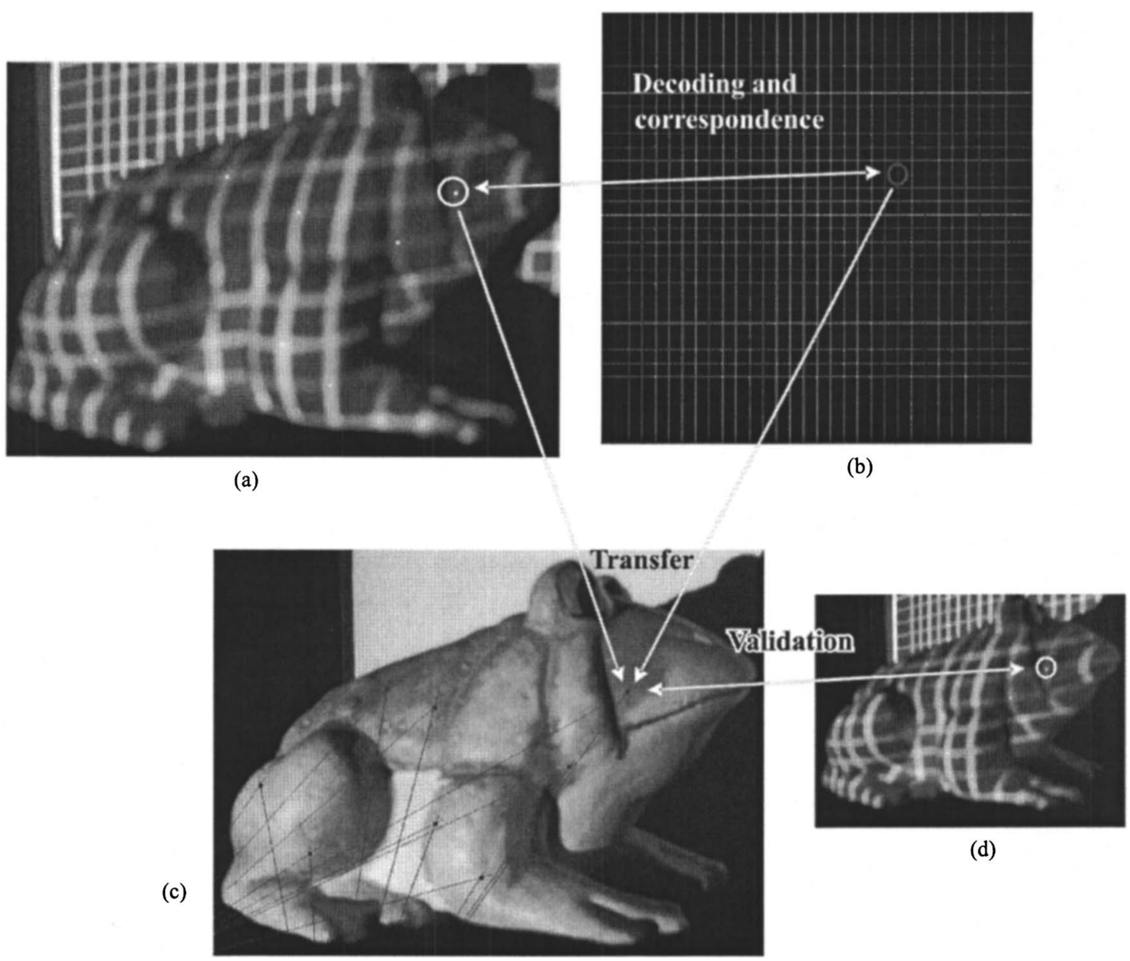

FIG. 5. (a) View as perceived by camera 1. (b) Initial projected pattern. (c) Transferred points in view 2. (d) View 2 with projector switched on so as to validate the transfer: the transferred point lies exactly on the good code.

by means of some approximations, into a Euclidean one. The ultimate goal is to map textures onto the reconstructed surfaces by using images from camera 2.

[a]Materials in this article were presented at the Third International Conference "Fundamental Prolems of Optics" OSAV'2004 (Section "Optical Sensors and Artificial Vision"). St. Petersburg, 2004.

[b]E-mail: t.molinier@iutlecreusot.u-bourgogne.fr
[c]E-mail: d.fofi@iutlecreusot.u-bourgogne.fr

[1]J. Battle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," Pattern Recognition **31**, No. 7, 963–982 (1998).

[2]P. M. Griffin, L. S. Narasimhan, S. R. Yee, "Generation of uniquely encoded light patterns for range data acquision," Pattern Recognition **25**, No. 6, 609–616 (1992).

[3]R. Hartley, "Lines and points in three views—an integrated approach," *Proceedings of the ARPA Image Understanding Workshop*, DARPA, Morgan Kaufmann (1994).

[4]R. Hartley, "Lines and points in three views and the trifocal tensor," International Journal of Computer Vision **22**, No. 2, 125–140 (1997).

[5]L. O. Harvey, Jr., "Critical flicker frequency as a function of viewing distance, stimulus size and luminance," Vision Research **10**, No. 1, 55–63 (1970).

[6]M. Livingston, *Vision-Based Tracking with Dynamic Structured Light for Video See-through Augmented Reality*, PhD Thesis, University of North Carolina (1998).

[7]Q.-T. Luong and O. D. Faugeras, "The fundamental matrix: theory, algorithms, and stability analysis," International Journal of Computer Vision **17**, No. 1, 43–76 (1996).

[8]T. Papadopoulo and O. Faugeras, "A new characterization of the trifocal tensor," *Proceedings of the 5th European Conference on Computer Vision*, Vol. 1405 of Lecture Notes in Computer Science (1998).

[9]T. Molinier, D. Fofi, R. Seulin, and F. Mériaudeau, "Trifocal tensor estimation applied to imperceptible structured light," *International Topical Meeting on Optical Sensing and Artificial Vision*, Saint Petersburg, Russia, October 18–21, 2004, State University ITMO, Vol. 1, pp. 32–39.

[10]R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: a unified approach to image-based modeling and spatially immersive displays," *Proceedings of ACM Siggraph*, Orlando, Florida (1998), pp. 179–188.

[11]C. Ressl, *Geometry, Constraints and Computation of the Trifocal Tensor*, PhD Thesis, Instituüt für Photogrammetrie und Fernerkundung, Technische Universität Wien (June 2003).

[12]J. Salvi, J. Battle, and E. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," Pattern Recognition Letters **19**, No. 11, 1055–1065 (1998).

[13]P. Torr and A. Zisserman, "Robust parameterization and computation of the trifocal tensor," Image and Vision Computing **15**, No. 8, 591–605 (1997).