Novel View Synthesis for Projective Texture Mapping on Real 3D Objects

Thierry MOLINIER, David FOFI, Patrick GORRIA

Le2i UMR CNRS 5158, University of Burgundy

ABSTRACT

Industrial reproduction, as stereography or lithography, have a lack in texture information, as they only deal with 3D reconstruction. In this paper, we provide a new technique to map texture on real 3D objects, by synthesizing a novel view from two camera images to a projector frame, considered as a camera acting in reverse. No prior information on the pose or the shape of the 3D object is necessary, however hard calibration of the complete system is needed.

Keywords: texture mapping, stereomatching, epipolar geometry, trifocal tensor

1. INTRODUCTION

Texture mapping on virtual objects is widely developped[1] and is generally easier than texture mapping on real objects, because virtual 3D information and environment are perfectly known. Since our goal is to texture a real 3D object with minimal prior information about it, we have to be able to complete stereomatching, view synthesis and finally texture mapping onto the three-dimensional surface. This article focus on the first two steps.

We aim to estimate a novel view from two captured views by means of the trifocal tensor in order to do augmented reality by projecting a synthesized view on a 3D real object (for virtual texturing, virtual painting or mixing physical and virtual textures[2]). This can be helpful in many applications such as rendering architecture, reverse engineering, industrial designing and so on. For instance, it may be useful for engineers or architect to visualize how the object they have created will look with different material or under different conditions before real painting or mass production.

Our machine vision system is composed of two (same or not) cameras and one digital light projector but can be generalized to any projector-camera system. Note that the system can be used for both acquiring a 3D model of an object and texturing a blank 3D object.

In the article, the complete methodology will be described from the basic formulation of the system geometry. Calibration, trifocal geometry and novel view synthesis will be detailed as well as the application of virtual texturing. First experimental results on real data will prove the feasibility of the method. This article is structured as follows. The first part is dedicated to the calibration of the system. The second part describes the estimation of the trifocal geometry and novel view synthesis. The third part shows the results; and finally, the paper ends with conclusions and future work.

E-mail: molinier@iutlecreusot.u-bourgogne.fr

2. CALIBRATION OF THE SYSTEM

2.1. Description of the system

Our system (see fig 2) is composed of two cameras (same or not) and one projector. As the projector parameters can be estimated as those of a camera, the two cameras and the projector are based on the same geometrical model (see fig 1).

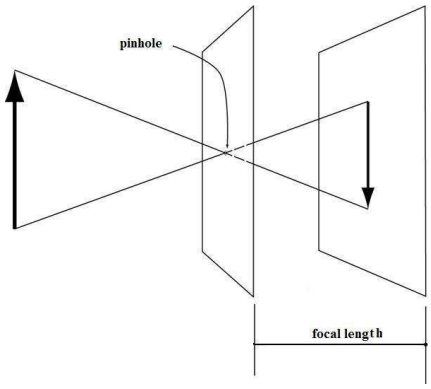

Figure 1. Pinhole model of the camera

Cameras and projector are based on the well known model used to describe cameras: the pinhole model (see fig 1). Moreover, the lens distortion will be estimated and corrected to improve the view synthesis transfer. We do not have any information about the scene: position and orientation of the real object with respect to the cameras and 3D geometry of the object are unknown. To ensure precise projection of the texture onto the object surface, hard calibration is required.

Hard-calibration of a data projector is usually more complicated than that of a camera, because projectors can not image the surface that they illuminate, so that the correspondence between the 2D projected points and the 3D illuminated points can not be made without the use of a camera and moreover it is difficult to retrieve the coordinates of the 3D points because the calibrating pattern is projected and not attached to the world coordinate frame. As a consequence, we propose a simple calibration method by performing in two steps: firstly both cameras are calibrated, then the projector is calibrated by triangulation of projected points.

Figure 2. System composed of two cameras and one projector

2.2. Stereo-calibration of cameras

A camera is characterized by two types of parameters : the extrinsic parameters, which describe the position and the orientation of the camera with respect to the world coordinates system, and the intrinsic parameters, which describe the optical characteristics of the camera. Extrinsic and intrinsic paramaters of the two cameras and the projector must be estimated. The two cameras are first calibrated by using a plane-based method (see fig 3) with homography computation.[3]

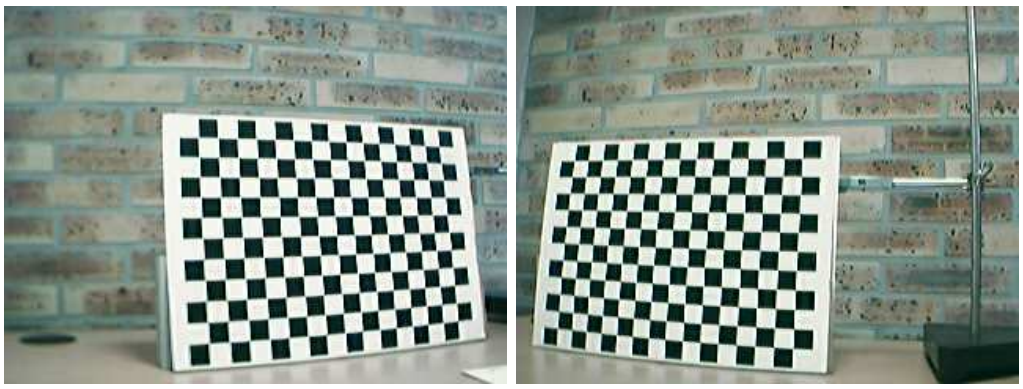

Figure 3. Pattern used to calibrate the cameras

In our work, we use the *Camera Calibration Toolbox* for Matlab implemented by J.-Y. Bouguet[4] and available on line. We perform a stereo calibration of the system (see fig 4). The choice of one camera, in our case the left camera, as the origin of the system as no consequence on the results. Only the extrinsic parameters, the matrices of rotation R and translation T of one camera with respect of the other one, have to be adapted. If the origin moves to the right camera, the new extrinsic parameters will be R^t and $-T$.

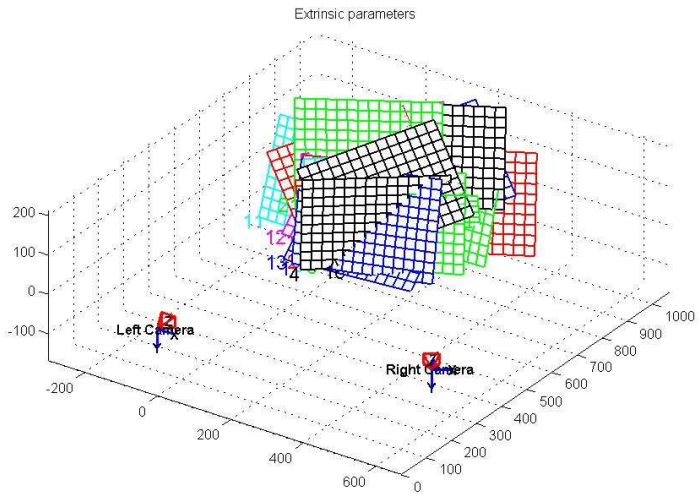

Figure 4. Calibration of the camera

The intrinsic parameters , that compose the calibration matrix, must be estimated :

$$K = \begin{bmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where α_u and α_v are focal lengthes in the principal directions, u_0 and v_0 are the coordinates of the principal point, and s is the skew factor.

The extrinsic parameters R and T express the position and the orientation of the right camera from the origin (the left camera).

$$X_{right} = R.X_{left} + T$$

with X_{right} 3D position of the center of the right camera and X_{left} 3D position of the center of the left camera

The calibration is based on two steps : firstly we approximate the parameters with the common initialisation: optical center is on the center of the view, the skew factor is null. An estimation of the homography between the points of the pattern and the views is calculated and then α_u and α_v are estimated .

For a pixel of size $k_u.k_v$, the focal length can be estimated by $\alpha_u = f_c.k_u$ and $\alpha_v = f_c.k_v$, with f_c focal length, α_u and α_u focal length along the horizontal and vertical axis

The second step is based on the minimisation of the error of retroprojection following the least square method between many pairs of images (see fig 4).

2.3. Calibration of the projector

We can not use the previous method because the projected pattern depends on the scene and is unknown. After having performed the hard calibration of the cameras, a pattern is projected onto a blank planar object and captured by the two calibrated cameras. Points of the pattern are acquired and classified (see fig 5).

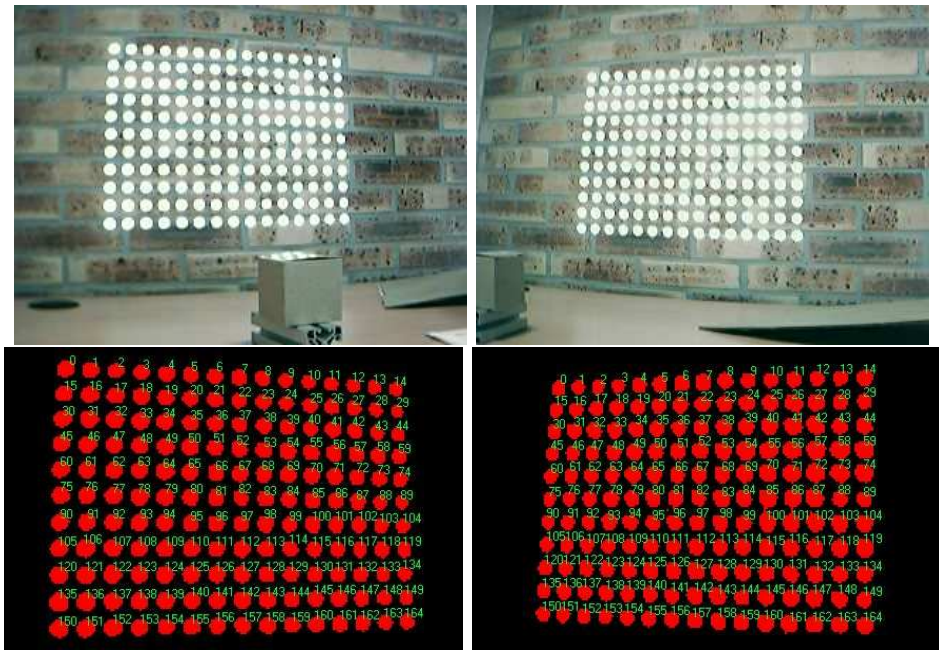

Figure 5. Left and right images of the projected pattern, and below same images classified

The classification of the points are crucial because an error will considerably decrease the quality of the calibration of the projector as the classification allows us to guess the points of correspondence between the views of the cameras.

The 3D points are reconstructed by triangulation of the 2D classified points of correspondence. A projector can be calibrated with the 3D position of the points and 2D position of the same points on the plane of the camera. The 3D position of the points are estimated from the triangulation with the calibrated cameras :

$$\begin{bmatrix} s.X_u \\ s.Y_u \\ s \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \begin{bmatrix} y_{11} & y_{12} & y_{13} & t_x \\ y_{21} & y_{22} & y_{23} & t_y \\ y_{31} & y_{32} & y_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} w.X_w \\ w.Y_w \\ w.Z_w \\ 1 \end{bmatrix}$$

t_x, t_y, t_z, y_{ij} extrinsic parameters

α_u, α_v, u_0, v_0, intrinsic parameters

The Faugeras method[5] is used to compute intrinsic and extrinsic parameters of the projector. The figures 4 and 6 show that our results are close to the real distances between the cameras and the projector (a distance of 45 centimeters is measured, and the software estimates a length of 46 cm). Several images are used to improve the precision of our parameters, each one is projected onto a blank planar surface that is moved at eacj acquisition.

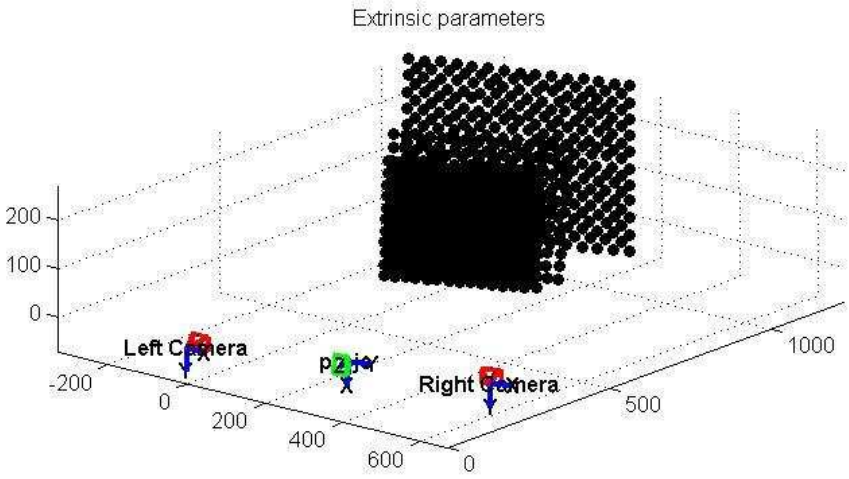

Figure 6. Calibration of the projector

3. VIEW SYNTHESIS AND EPIPOLAR GEOMETRY

Once the system is calibrated, the trifocal tensor can be computed (see fig 8) .The trifocal tensor T is a valency 3 tensor with two contravariants and one covariant indices.

3.1. Trifocal geometry

The epipolar geometry (see fig 7) describes the relations of correspondence between projections on several image planes. The fundamental matrix can only estimate the correspondance between a point and a line between two views. The trifocal tensor is 3x3x3-cube which describes the epipolar geometry for three views as the fundamental matrix does for two views. It is composed of 27 elements but has only 18 degrees of freedom because it has to satisfy internal constraints. The trifocal tensor allows to link lines and points from one view to the other two views: 3 points together, 3 lines, 2 points and 1 line, 2 lines and 1 point.[6]

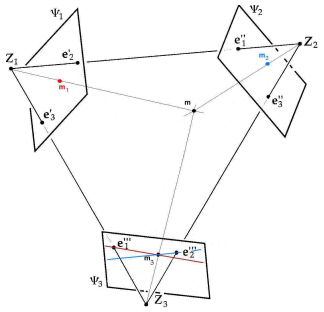

Figure 7. Epipolar geometry

e are the epipoles on the plane, Z the projection center of the image plane 2. The figure 7 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. The reader interested in trifocal geometry can refer to the thesis of Ressel.[7]

The projection matrices can be written as :

$$P = [I_d|0]$$
$$P' = K'[R'|T']$$
$$P'' = K''[R''|T'']$$

The left camera is choosen as the origin of our local system. That's why its rotation matrix T_{left} and translation vector R_{left} can be set to $T_{left} = 0$ and $R_{left} = I_d$. But actually the calibration matrix of the left camera is not I_d .We need to write $P = (I_d|0)$ to estimate the trifocal tensor, so we apply a homography on the points of the first image with K^{-1}, K matrix of calibration of the left image.

When dealing with the trifocal tensor, it is convenient to use standard tensor notation :

- images points and lines are represented by homogeneous column and vectors, respectively $x = [x^1x^2x^3]^t$, $l = [l_1l_2l_3]^t$.
- the ij-entry of a matrix A is denoted by a^i_j, with index i being the contravariant (row) index and index j being the covariant (column) index.
- the Einstein convention is observed for which indices repeated in the contravariant and convariant positions imply summation over the range of the index; for example : $x^{'i} = \sum_j a^{'}_j.x^j = a^i_j.x^j$.

Then the trifocal tensor is expressed by:

$$T^{kj}_i = v^{'j}b^k_i - v^{''k}a^j_i$$

3.2. Novel view synthesis

The projector is considered as a camera acting in reverse. We estimate the view that would be seen by a camera if this camera had the same intrinsic and extrinsic parameters. The trifocal tensor enables to estimate the position of a point on the third image, in our case the view of the projector, if the points of correspondence on the two first views are known.

The novel view can be estimated from

$$p^i s_j^\mu T_i^{kj} \cong p''^k$$

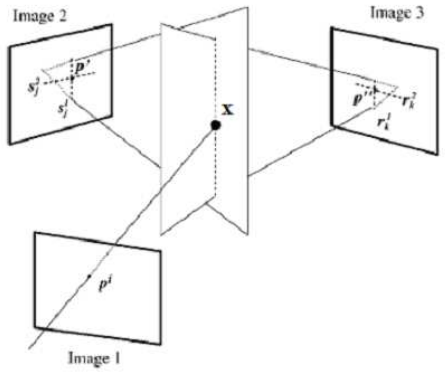

Figure 8. Transfer on the third image from the first and second images

p et p'' are the points on the first and third image, s is a line on the second image, perpendicular on the epipolar line which is the projection of the point p with the fundamental matrix F_{21}, which describes the epipolar geometry between the first and second images and is extracted from the trifocal tensor. It is better to use perpendicular lines than epipolar lines to avoid problems of critical configurations (see Hartley[6])

The first step of the method is to compute the corresponding points between the two views. Since a dense correspondence is required in order to compute a novel view, the disparity map between the views has to be estimated. The pair of images has to be previously rectified, which means that a point in the first image and its correspondence in the second image lie on the same scanline: as a result, the epipolar lines are horizontal and the epipoles are at infinity. Moreover, the system has to be calibrated in order to ensure that the projective texture maps exactly onto the real 3D objects. We have now a set of corresponding points and the calibration matrices of the two cameras and the projector. The next step consists in estimating a novel view, which can be considered as a third image of the object grabbed by a virtual camera that would have the same parameters and position than the data projector. Since the system is fully calibrated, the trifocal tensor can be estimated from the three calibration matrices directly. Through the trifocal tensor and the dense correspondence, a novel view can be synthesized by means of point-transfer method. This novel view can be projected onto on real 3D object in order to map virtual texture on its surface.

4. RESULTS

4.1. View synthesis

We only try to simulate view with simple objects, like cube. The images are easy, so the correspondance is not dense: centers of the circles, vertices of cube. This points are tranferred with the method described in[8, 9] .

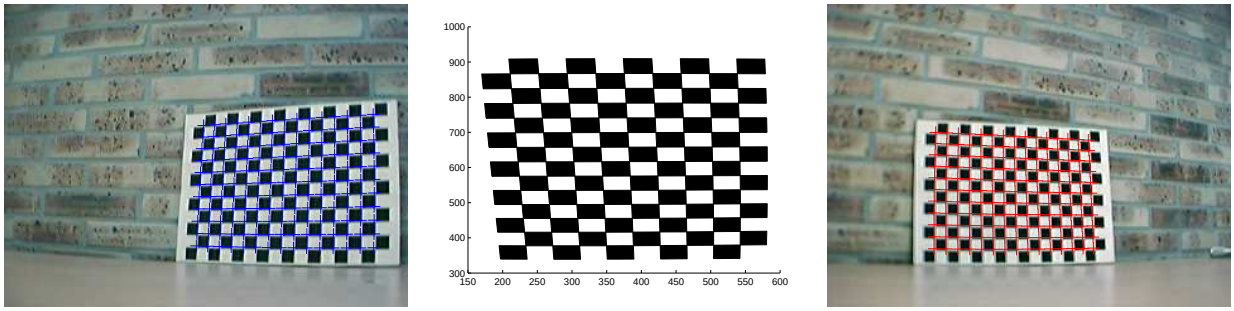

Figure 9. View synthesis : a) left, b) synthesized, c) right images

The pattern (see fig 9) is supposed to be in front of the projector, so we guess that the projected image must be composed of squares. By analizing the synthesized image, it can be notice that the left part is more distorted than the right part.

4.2. Synthesis of the pattern of the projector

Then we synthesize the pattern projected by the projector (see fig 11).

We know the projected pattern (see fig 10), so a comparison between the true pattern and the simulated pattern is possible (see fig 11) .

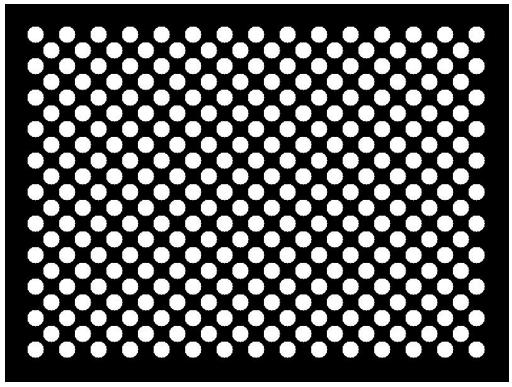

Figure 10. Projected pattern

The figure 11 shows the same type of skew of the synthesized image. The left part of the simulated image is better than the right part. As we know the true position of the centers of the pattern circle, we can estimate the error of retroprojection

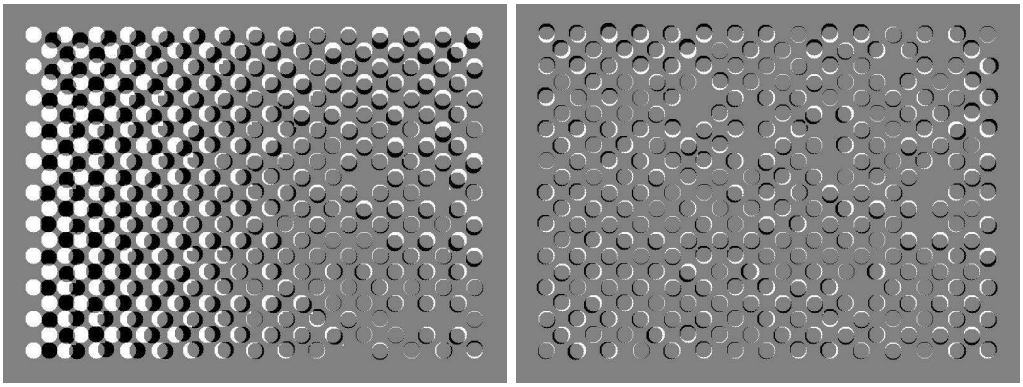

Figure 11. fusion : a)and b)fusion of projected and synthesed images before and after the correction of the distorsion

4.3. Error of retroprojection

The error of retroprojection is the distance between a known point on the third view and the tranfer of the points of correspondence of the first and second views on the third view.We have triplets of points between the three cameras when the projector is calibrated, so we can estimate the precision of the transfer method. The table shows 1 the results :

Mesure	dist mean	dist max	dist min	std deviation
distance in pixel	9	31	0.8	7.5

Table 1. Error in pixels between projected view and corresponding synthesized view

If the mean error is acceptable and have the same range than Habed[10] but the deviation on the left part is not. The mean error of retroprojection on the left part of the image is 3 pixels, but the error on the right part is about 40 pixels.

4.4. Improving the transfer

We try many common techniques to reduce the error: change of the origin of the base, normalisation of Hartley, but the reduction is not noticeable. The distorsion factor (radial and tangential) is estimated with the calibration of the cameras.[11] The points are modified on the twos views of the cameras, then the trifocal tensor must be adjusted, following[6]:

$$T_i^{kj} = (F^{-1})_i^r . G_s^j . H_t^k . T_r^{st}$$

with F, G, H homographies applied on the three views to correct the three images.

The results after correction of the distorsion are showed in the table 2

Mesure	dist mean	dist max	dist min	std deviation
distance in pixel	2	6	0.2	1.2

Table 2. Error after correction of the distorsion due to the cameras

4.5. Texture mapping on real object

Vertices of the cube on the left and right images are manually selected and then the points of the projector view can be estimated, and finally the squares can be drawn. The result can be seen in figure 12.

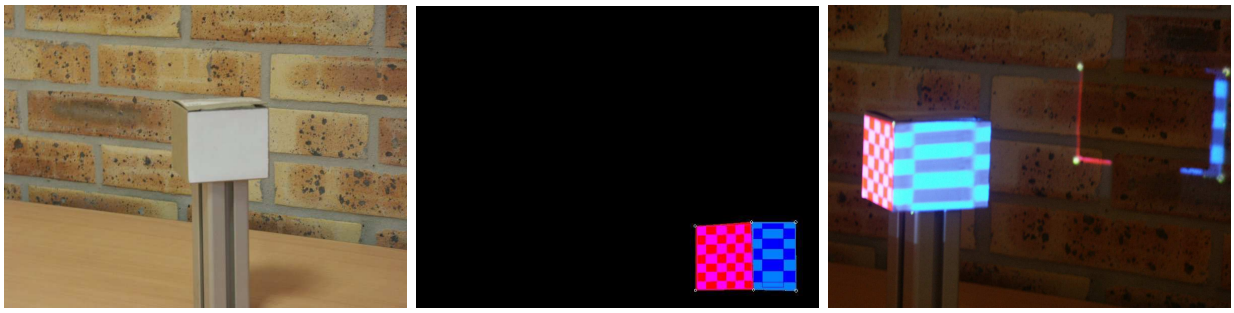

Figure 12. from left to right : view from camera, view synthesized from projector, texture mapping on real object

The edge of the cube, between the blue and red facet is visually near perfect, but we can see behind the cube on the wall that the projected texture is closed to the edges of the cube. This error comes from the error of retroprojection.

5. CONCLUSION

In this paper, a method to calibrate a system composed of two cameras (same or not) and a projector, considered as a reverse camera is presented and validated. It has been shown that, from the estimation of the trifocal tensor, it is possible to synthesize a novel view on the plane of the projector from points of correspondence of the views of the two cameras. We can then projecte this view on a real 3D object with a good precision. Unfortunately an automatic stereomatching don't give today enough interesting points. Two problems are antagonist : On one hand, it is easier to do stereomatching if the two cameras are closed to each other, as the two images will be near the same; On the other hand, precision is greater if the two cameras are not closed one to other.

Future work will concentrate on a new methodology to avoid this antagonism. By using prior information as the 3D shape of the object and by virtualizing the whole system with the 3D library OpenGL with the paramaters extracted from the stereo calibration , the view synthesis of the image of the cameras enables to pose the 3D object with respect to the system and the adjustement of this virtual view to the projector frame.

REFERENCES

1. P. S. Heckbert, "Survey of texture mapping," *IEEE Comput. Graph. Appl.* **6**(11), pp. 56–67, 1986.
2. R. Raskar, G. Welch, K.-L. Low, and D. Bandyopadhyay, "Shader lamps: Animating real objects with image-based illumination," in *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, pp. 89–102, Springer-Verlag, (London, UK), 2001.
3. P. Sturm and S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, USA*, pp. 432–437, Juin 1999.
4. J. Bouguet, "Camera calibration toolbox," in *http://www.vision.caltech.edu/bouguetj*, 1995.
5. T. Faugeras, "The calibration problem for stereo," in *CVPR86*, pp. 15–20, 1986.
6. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second ed., 2004.
7. C. Ressl, *Geometry, Constraints and Computation of the Trifocal Tensor*. PhD thesis, Institut fr Photogrammetrie und Fernerkundung, Germany.
8. D. F. T. Molinier, "Trifocal tensor estimation applied to imperceptible structured light," in *((OSAV'2004), Saint Petersburg (Russia))*, 2004.
9. F. M. T. Molinier, D. Fofi, "Self-calibration of a trinocular sensor with imperceptible structured light and varying intrinsic parameters," in *(QCAV'2005, Nagoya (Japan)*, 2005.
10. B. Habed, "Novel view synthesis: a comparative analysis study," *(Vision Interface, Montrea(Canada)* , 2000.
11. J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, p. 1106, IEEE Computer Society, (Washington, DC, USA), 1997.