Detection of surfaces for projection of texture

Thierry Molinier[a], David Fofi[a], Patrick Gorria[a], Joaquim Salvi[b]

[a]Le2i UMR CNRS 5158, University of Burgundy, France
[b]Computer Vision and Robotics Group, University of Girona, Spain

ABSTRACT

Augmented reality is used to improve color segmentation on human's body or on precious no touch artefacts. We propose a technique based on structured light to project texture on a real object without any contact with it. Such techniques can be apply on medical application, archeology, industrial inspection and augmented prototyping. Coded structured light is an optical technique based on active stereovision which allows shape acquisition. By projecting a light pattern onto the surface of an object and capturing images with a camera, a large number of correspondences can be found and 3D points can be reconstructed by means of triangulation.

1. INTRODUCTION

Surface reconstruction is one of the most important topics in computer vision due to its wide field of application. Some examples of applications are range sensoring, industrial inspection of manufactured parts, reverse engineering, object recognition and 3D map building. Amongst all the ranging techniques,[8] stereovision is based on viewing the scene from two or more points of view and then finding correspondences between the different images in order to triangulate the 3D position. Triangulation is possible if cameras have been calibrated beforehand,[5].[11] However, difficulties in finding the correspondences arise, even when taking into account epipolar constraints. Coded structured light consists in replacing one of the cameras by a device that projects a light pattern onto the measuring surface. Today the most commonly used devices are LCD projectors, although previously it was slide projectors.

The projection of texture converts a two-dimensional texture or image into a three-dimensional texture by projecting it in one or several directions to the objects present in the scene. Basically, texture is projected in synthetic objects with many applications in the field of computer graphics, such as virtual reality and visualization. Usually, texture is only projected on real objects as a way to alleviate the correspondence problem in surface acquisition by the projection of coded patterns or pseudorandom sequences. Nevertheless, few papers have been published centered on the projection of texture so that the three-dimensional texture is adapted adequately to a real object. The applications of this technique are various and interesting. Actually, it is the only way to see how it will be seen an object before painting or carving it. There are also applications in clothing designing, virtual museums, movie making and augmented prototyping. Previous methods,[915] are not fully automated. With our method the knowledge of the pose (position and orientation) of the object with respect to the sensor is not required. Such method will be usefull for the control of dangerous material in unwelcome environnement by improving visual aspect of default, or improve the rendering of prototyping objects by adding texture.

The structure of the paper is as follows. First the system we used is presented, Secondly we describe the technique used finding points of correspondence, our results are explaned and finally the conclusion.

2. PROJECTOR-CAMERAS SYSTEM

We have designed a sensor composed of two cameras (same or not) and one digital projector (see figure 1). This system ensures great flexibility and allows both capturing scene structure and displaying above. Indeed, it is possible to use the two cameras to establish correspondence and triangulate the scene or, alternatively, to use the projector and one camera as a structured light system (in case of blank or poor-textured surfaces); it is also possible to simultaneously project a texturing light and a structured light as demonstrated by Cotting et al.[3] So as to model the system we choose to estimate the three fundamental matrices for each couple among the three views to estimate the whole epipolar geometry. In order to project the virtual texture onto the object surface precisely, the calibration of the system is necessary - although the optical characteristics of the projector are not required to project the texture, as we use the epipolar geometry and the structured pattern to estimate points of correspondence between the projector and the two cameras.

Note that this system may easily be reduced to a one-projector and one-camera system. It can also be extended to n-projetors and m-cameras systems, depending on the application and context.

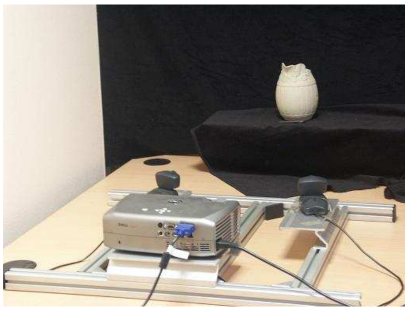

Figure 1. System

2.1. Camera model

Consider the pinhole model for the camera. The projection of a 3D point $\mathbf{X} = [X\ Y\ Z\ 1]^T$; onto the image plane x_c is expressed by the following equation:

$$\mathbf{x_c} \approx \mathbf{P_c}.\mathbf{X}. \tag{1}$$

where projection matrix $\mathbf{P_c}$ can be factorized as follows:

$$\mathbf{P_c} \approx \mathbf{K_c}.[\mathbf{R_c}\ \mathbf{t_c}]. \tag{2}$$

where $\mathbf{R_c}$ and $\mathbf{t_c}$ represent the pose of the camera with respect to the world coordinate frame (rotation matrix and translation vector, respectively), $\mathbf{x_c} = [x\ y\ 1]^t$ is the 2D image point in pixels and $\mathbf{K_c}$ is the matrix of the intrinsic parameters. In general:

$$\mathbf{K_c} = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

where (x_0, y_0) are the co-ordinates of the principal point; f_x and f_y the focal lengths for the x-axis and y-axis and s is the skew (the angle between the x-axis and y-axis of the image plane). Currently, manufactured cameras do not always justify this very general optical model. For example, it is now common to assume rectangular pixels, and thus assume zero skew ($s = 0$). Likewise, the principal point is often assumed to be at the centre of the image plane by translating the origin of the image plane to the centre) and the pixels to be square (i.e.

unity aspect ratio: $f_x/f_y = 1$). In its simplest expression, the matrix of the intrinsic parameters of the camera has only one degree of freedom and may be expressed by:

$$\mathbf{K_c} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

2.2. Projector model

It is well known that a data projector can be seen as a camera acting in reverse, by inverting the line of sight. Thus, equations (1), (2) and (3) of the above sub-section are still valid for a projector. The differences arise when the approximations are considered. The assumptions that the principal point is close to the image centre cannot be held for a projector (Figure 1). Indeed, the principal point of a projector is vertically shifted from the image centre to allow an off-axis projection cone (because a projector is supposed to lie on a table and to project the light over-head). However, as for a camera, it is realistic to assume that a projector has unity aspect ratio and zero skew. Thus, in its simplest expression, matrix of intrinsic parameters of a projector has two degrees of freedom f and y_0 and may be expressed by:

$$\mathbf{K_c} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

2.3. Calibration

In order to project the virtual texture onto the object surface precisely, the calibration of the system is necessary. Calibration provides us with the two camera matrices $\mathbf{P_{c1}} \approx \mathbf{K_{c1}}.[\mathbf{R_{c1}}\ \mathbf{t_{c1}}]$ and $\mathbf{P_{c2}} \approx \mathbf{K_{c2}}.[\mathbf{R_{c2}}\ \mathbf{t_{c2}}]$. The matrix $\mathbf{R_i}$ represents the pose of the camera with respect to the world co-ordinate frame (with $i = c_1$, c_2).

A plane-based calibration method could advantageously be used.[2] This method is based on the observation of a planar pattern shown at differents orientations. The stereo-calibration of the system is estimated by the toolbox developped by Bouguet : Extrinsic and intrinsic parameters and lens distorsion are known. The two cameras can then triangulate the points of the pattern projected by the projector and gives us a set of 3D points and 2D points. The calibration of the projector is performed by means of the method of Faugeras.[4]

The figure 2 shows the whole calibrated system and the triangulated pattern used to calibrate it.

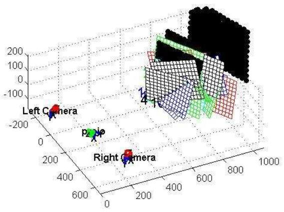

Figure 2. Hard calibration of the system

2.4. Epipolar geometry

The epipolar geometry describes the projective geometry of a system composed by at least two views. It is independant of the scene structure and only depends on the intrinsic parameters of the cameras and its relative pose. The fundamental matrix describes the epipolar geometry for two view and enables the correspondence between a point and an epipolar line where the point of correspondence is. Methods for the estimation of a robust fundamental matrix[12] are well known. The fundamental matrix F is computed from the points of correspondence of two images, m_1 and m_2, by the equation 6 :

$$m_1^t.F.m_2 = 0 \tag{6}$$

The figure 3 shows the relation between the 2D points m_1 and m_2 and the 3D point m.

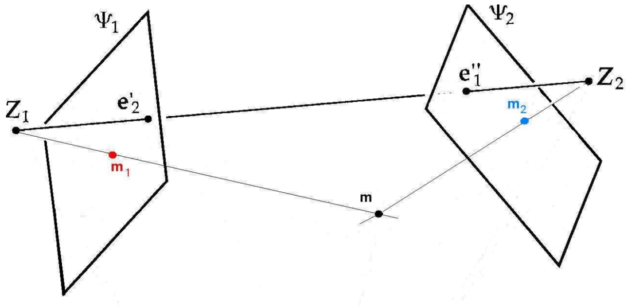

Figure 3. epipolar geometry between two views

The fundamental matrix is computed for two between the three views. We choose to start from the view of the projector, i.e. the projected structured pattern, because it is easier to recognize the element of the pattern. Armangué[13] surveyed up to 19 of the most used methods in fundamental matrix estimation. Experimental results show that: (a) linear methods are quite good if the points are well located in the image and the corresponding problem previously solved; (b) iterative methods can cope with some gaussian noise in the localization of points, but they become really inefficient in the presence of outliers; (c) robust methods can cope with both discrepancy in the localization of points and false matchings. However, LMedS still obtain the best results when a low computing time is not required. In all, LMedS is the most appropriate for outlier detection and removal. Although three views are available, we prefer to use the fundamental matrix rather than the trifocal tensor because the estimation of the fundamental matrices appears more robust against noise than the estimation of the trifocal tensor.

3. CORRESPONDENCE PROBLEM

We use textureless real object like model. Finding correspondence points on those blank surfaces is a hard task even through active methods. This problem is solved by means of two approaches :

- The epipolar geometry that links a point in the first image to the epipolar line on the second image. It is known that the corresponding point lies on the epipolar line so that the area of research is reduced.
- The structured pattern, which is projected, allows to find the point of correspondence on the epipolar line by artificially structuring the object surface.

Although three views are available, we make the choice to compute the fundamental matrices instead of the trifocal tensor.[14] Lens distorsion of the used webcams can not be ignored. The correction of the positions of the points of interest before the calibration, and the estimation of the fundamental matrix is required as detailed in.[12]

3.1. Pattern design

From one point of the pattern, we can estimate the epipolar line for each view of the cameras. The point of correspondence must be on this line, or at least near the line. That is why the point of correspondence is searched on a band centered in this line. The epipolar geometry can not give more information, but the pattern can. The structured pattern is computed so that each element of a cell of five by five pixels is unique.

This pattern is built following this steps :

- The first cell, in the left up corner, is a black cell of 5×5 pixels. A cell of 5×5 pixels is the smallest cell to complete a 1024×768 pattern.
- A column of 5 pixels is selected randomly among all the combinations of possible columns, then it is added to the pattern. If the verification of the uniqueness of each cell of 5×5 pixels is satisfied, another column is added, otherwise the new column is changed.
- A line of 5 pixels is added, with the same method and verification than for the column. The whole line is completed pixel by pixel, randomly for the choice of black or white and the condition of uniqueness is still validated. This step is repeated to complete the pattern.

The figure 4 illustrates these steps. The resulting pattern is depicted in the figure 5.

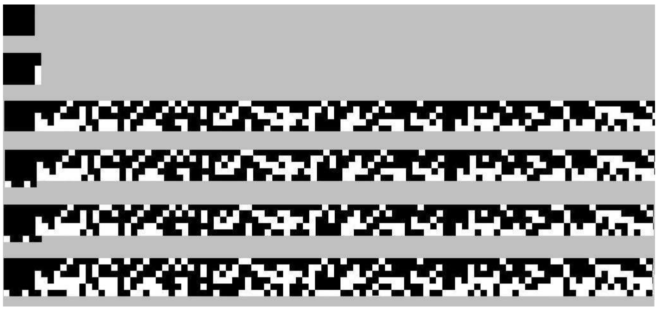

Figure 4. pattern

Note that this pattern is a binary one so that it can be embedded in a texturing light as an imperceptible pattern.[3]

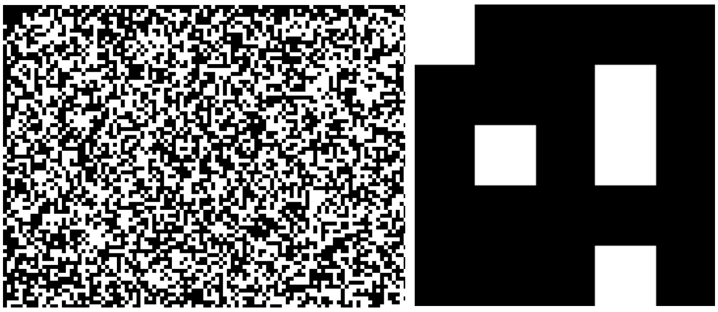

Figure 5. pattern

3.2. Correlation from the pattern : structured light

From one point of the pattern, we can estimate the epipolar line for each view of the cameras. The point of correspondence must be on this line, or at least close to the line. That's why the point of correspondence is searched on a band centered to this line. The epipolar geometry can not give more information, but the pattern can. The structured pattern is computed so that each element of a cell of five by five pixels is unique.

Now we will search this cell on the zone of interest defined by the epipolar geometry. There are many methods to estimate the correlation between the cell of the point of the pattern and those on the zone of interest,[6] based on energy filter. These filters are based by applying a set of masks which define some textural properties of the image. But we will use directly the comparaison between the two cells (see equation 7) of the two images, A and B. Mostly the point for which we find the maximum of correlation must be the point of correspondence.

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{(\sqrt{\sum_m (A_{mn} - \bar{A})^2)(\sum_n (B_{mn} - \bar{B})^2)} \tag{7}$$

where $\bar(A)$ and $\bar(B)$ are the mean of the images A and B.

3.3. The outliers

The outliers are a real problem for our aim, the projection of texture. One outlier can be visually a disease, that's why we have to choose parameters which eleminate all the outlier.

The first parameter is the correlation. We wrote that mostly the point for which we find the maximum of correlation must be the point of correspondence, but there was no criterium of the value of the maximum. We choose a value that eliminate the outlier. The value of the correlation is between 0 and 1, a value of 0.8 appears as good as enough.

The second parameter concerns the geometry epipolar. We only use the fundamental matrices between the projector and the two views, but not between the two views. The distance between a point and its epipolar line is estimated, and due to the first estimation of the epipolar distance, a threshold of one pixel seems good.

4. EXPERIMENTAL RESULTS

We obtain more points of correspondence with this technique (see figure 6), epipolar geometry and structured light, than with a classic detector like harris[7] , for textureless objects. The choice of the two threshold for the value of the correlation and the epipolar distance. We have to find a trade off between identifying the remain outliers and keeping points of correspondence.

The figure 6 shows the 3D reconstruction of the 3D points of the model. Outliers can be identified as points far away from any other, compared to the mean distance between the 3D points.

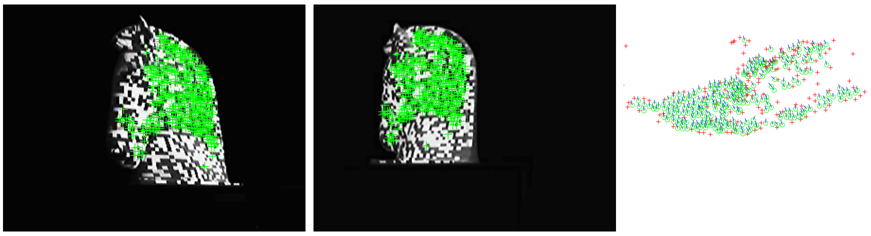

Figure 6. left points of correspondence after the epipolar threeshold,right 3d reconstruction and normal

A second test consists in the projection of different texture on two parallel planes (see fig 7).

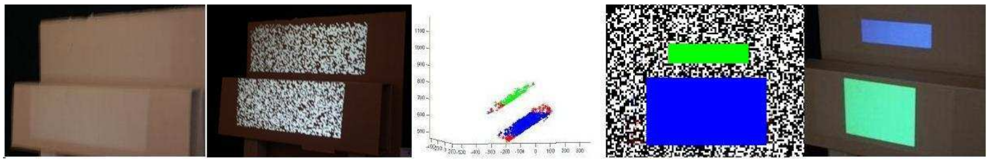

Figure 7. left to right : object, object with structured pattern, reconstruction and identification of the planes, region of correspondence between the 3D and 2D points, projection of the texture on the object

The differents planes are reconstructed and then identified. The correspondence between the 3D and 2D points are known, the region on the pattern corresponding with the planes can be identified and a texture is projected. Two reasons explain the lack of texture on the border. Firstly there are a lack of information because of the occlusion due to the relative position of the projector and the cameras. Secondly at the border of the two planes the cell is cuted and the threshold for correlation eliminates the point as a good one. The border of the blue zone is only cuted by the step of correlation and the green zone by the two reasons.

The figure 8 shows the result of the method on a real model. No prior information on the pose or the shape of the 3D object is necessary.The whole texturation of the object is still forbidden by the deformation of the pattern on the curved surface and by the occlusions .

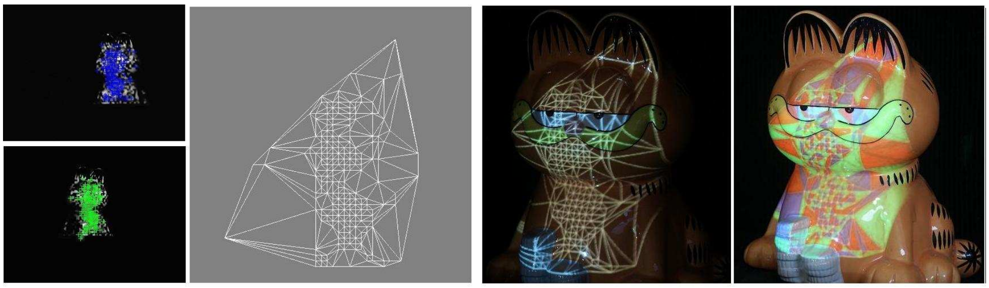

Figure 8. left to right : Points of correspondence on the two views, image of the projector, object with pattern and texture

The next test is on a real object, a model of a vase with the same method, but this time we realize a delaunay triangulation on the 2D points of correspondance of the pattern. Because of the 3D form of the object and the pose of the cameras, only the central part of the model can be textured with the projector. The figure 9 shows the result with different threshold of correlation. Lesser is the threshold, more numberous are the points of correspondance and the number of the outliers.

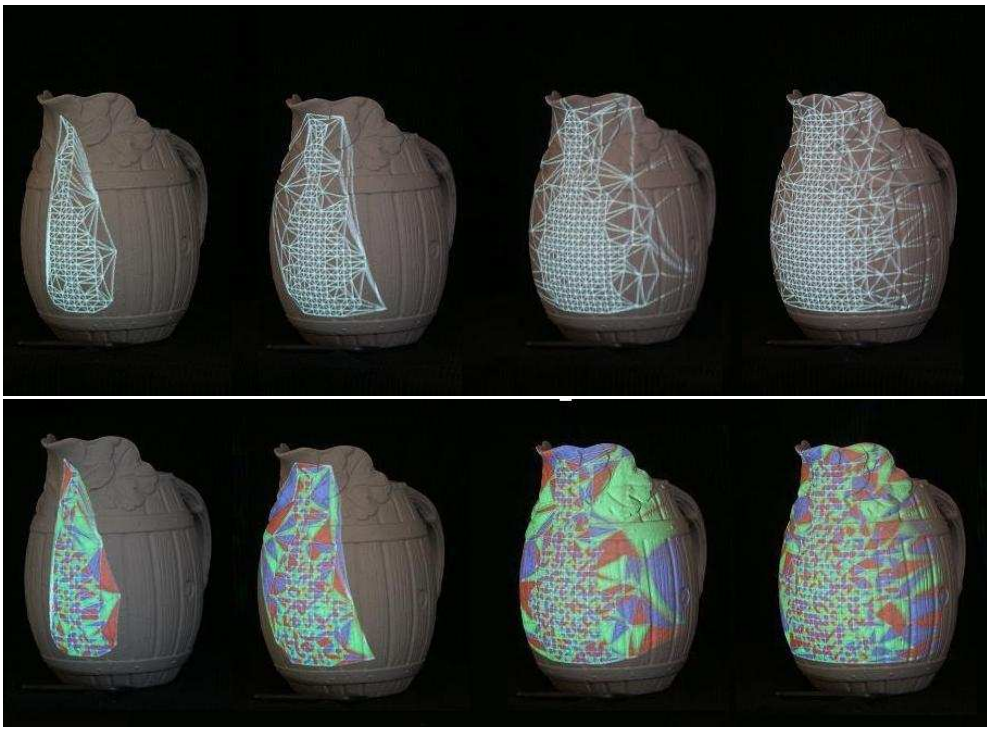

Figure 9. Delaunay triangulation of points of correspondance, with or without color on face

5. CONCLUSIONS

The configuration of our system, especially the respective position of the two cameras, does not allow to use classical algorithm to find points of correspondence between the two images. But a projector is used in the passive way to project texture on an object. The active use of this projector is based on the projection of a structured pattern, in which two same cells can not be found. The zone of search of the corresponding cell on a second image is reduced by the epipolar geometry, because the point of correspondence is on the epipolar line. By restricting the error on the epipolar distance and the correlation of the cells, a large number of points of correspondence is found on textureless object. Finding a large number of points of correspondence, through structured light and epipolar geometry, is described in this article. The few outlier which would not be identified before, because of too low threshold, can be although estimated with an unsual 3D position. The future work will be based on increasing the number of points of correspondence by an adaptive lecture of the unique element of the pattern on the images.

A precise and automatic capture of a cell of the projected structured pattern on a object will capture our attention. It will not only increase the number of points of correspondence, especially on the deformed surfaces on the object, where the pattern is very altered, but also the precision of the position of the points of correspondence, which are on the center of the pattern. It will follow a better estimation of the 3D points after the triangulation and then the estimation of the planes, and finnaly the precision of the projected texture.

REFERENCES

1. Beardsley, P.; Forlines, C.; Raskar, R.; van Baar, J., "Handheld Projectors for Mixing Physical and Digital Textures", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), ISSN: 1063-6919, Vol. 3, pp. 112, June 2005
2. J.P. Bouguet., Camera Calibration Toolbox, http://www.vision.caltech.edu/bouguetj, 1995.
3. Cotting, D.; Naef, M.; Gross, M.; Fuchs, H.,, Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display., Third IEEE and ACM International Symposium on Mixed and Augmented Reality, 2004.
4. Faugeras and Toscani., The Calibration Problem for Stereo, CVPR86, 1986, 15-20.
5. O. Faugeras, Three-Dimensional Computer Vision, MIT Press, Cambridge, 1993.
6. Garcia Campos, Rafael, A proposal to estimate the motion of an underwater vehicle through visual mosaicking, EIA Electrònica, Informàtica i Automàtica, University of girona, Spain, 2001.
7. C. G. Harris and M. Stephens, A combined corner and edge detector, in Proc. 4th Alvey Vision Conference, Manchester, 1988, pp. 147-151.
8. R. Jarvis., Range sensing for compute vision, Advances in Image Communications, Elsevier Sciences Publishers,1993.
9. Ramesh Raskar and Greg Welch and Kok-Lim Low and Deepak Bandyopadhyay, Shader Lamps: Animating Real Objects With Image-Based Illumination, Proceedings of the 12th Eurographics Workshop on Rendering Techniques, 2001,London, UK.
10. Bandyopadhyay, D. and Raskar, R. and Fuchs, H., Dynamic Shader Lamps: Painting on Movable Objects, ISAR'01, 2001 , IEEE Computer Society.
11. J. Salvi, X. Armangué, J. Battle., A comparative review of camera calibrating methods with accuracy evaluation, Pattern Recognition, 1617-1635, 35, 2002.
12. J. Salvi, X. Armangué, J. Pagés. A Survey Addressing the Fundamental Matrix Estimation Problem. IEEE International Conference on Image Processing, Thessaloniki (Greece) October 2001.
13. X. Armangué and J. Salvi, Overall View Regarding Fundamental Matrix Estimation, Image and Vision Computing, volume 21, February 2003, 205-220.
14. Hartley and Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
15. Verlinden, J.C., de Smit, A., Peeters, A.W.J., van Gelderen, M.H.(2003) "Development of a Flexible Augmented Prototyping System ' Journal of WSCG, vol. 11(3), pp. 496-503.
16. K. Matkovic and T. Psik and I. Wagner and D. Gracanin, Dynamic Texturing of Real Objects in an Augmented Reality System, 2005, 245-248, Proceeding of IEEE Virtual Reality 2005, Bonn, Germany.