Trifocal transfer based novel view synthesis for micromanipulation.

Julien Bert, Sounkalo Dembélé, Nadine Lefort-Piat

Trifocal Transfer Based Novel View Synthesis for Micromanipulation

Julien Bert, Sounkalo Dembélé, and Nadine Lefort-Piat

Laboratoire d'Automatique de Besançon
UMR CNRS 6596 - ENSMM - UFC
25000 Besançon, France
Email: {jbert, sdembele, npiat}@ens2m.fr

Abstract. In trifocal transfer based novel view synthesis, matched pixels of both input views are projected in the novel view. The angle of view of this latest is usually narrow, i.e. the novel view is very close to input ones. In this paper we improve the method to get a large angle of view. A simplex approach is used to compute the model of the virtual views pose. This model allows the computation of the novel view at any desired angle of view. We also show that those results are very useful in micromanipulation tasks where transfer of edges is enough instead of the entire pixels of input views.

1 Introduction

Novel View Synthesis (NVS) is a part of computer vision introduced by [1]. It deals with the obtaining of a maximum of views of an environment from a minimum of real data on it. For example, a lateral view of an object can be synthesized with only two real top views. There are two classes of methods in NVS: the model-based rendering and the image-based rendering.

In model-based rendering (MBR), virtual environments are created from mathematical models. A typical example is 3D characters synthesis in movies and video games by modeler softwares. In image-based rendering (IBR), a set of real images of the scene is used to build a novel view. According to the knowledge about scene geometry, [2] proposes the following classification: rendering with no geometry, rendering with explicit geometry and rendering with implicit geometry. Rendering with no geometry i.e. no calibration is used to create a mosaic from a set of local views that leads to a novel global view [3], [4]. Rendering with explicit geometry i.e. with strong calibration is close to MBR. Its purpose is the reconstruction of a 3D view from real views of the scene [5]. This technique needs a strong calibration and is computationally expensive. Rendering with implicit geometry only needs a weak calibration. Ref. [6], presents three techniques of NVS of this type: the line of sight, the epipolar transfer and the trifocal transfer. The line of sight approach is based on ray-tracing [7]. Its drawback is the fact that at least ten images are required to obtain a synthetic view. The epipolar transfer approach is introduced by [8], it is based on epipolar geometry where

the epipolar constraint defines the point-line duality in pair of images: one point in the left view corresponds to a line in the right view. This concept is used to create a virtual view from two real views. Each point of the virtual view is the intersection of the lines of the points from the real views. The trifocal transfer approach first proposed by [9] is based on the trifocal constraint between three views. The later is defined by a tensor. With two real images and a tensor, all the points of the images are transferred into a novel view. The control of that corresponding virtual view pose is performed through a transformation matrix.

This virtual view can be very useful in micromanipulation which concerns the manipulation of parts at the microscale, i.e. in the range from 1 μm to 1 mm. The main applications of micromanipulation are assembly, sorting and testing of microparts. In addition to biomicroparts like cells and pollen seeds, artificial microparts are chemically or mechanically synthetized, or micromachined. Classical examples of the first and second types are respectively grains of powder like drugs or cosmetics, and optomechatronic components like balls, pegs, pins, threads, membranes, lenses, shutters and fibres. In some cases these microparts define final products (MEMS), otherwise they must be assembly to lead to the final products. For that purpose some automated microassembly systems have been developed by [10], [11], [12] and [13]. From those results it can be noticed that a microimaging system is always required, and the most used is a photon microscope connected to a camera. The images and their processing and analysis allow the task surveillance, system control or microparts recognition. The field-of-view of the microscope is very narrow that leads to the use of multiple views imaging : global view (usually at the top), left and right lateral views. The second reason of multiple views use is the fact that top view only allows the access to the xy position of the microgripper. Lateral view is required to get the z position. The third reason is the occurrence of components occlusions during assembly, the microgripper can hide the microparts to pick. However multiple views imaging has a drawback, the microimaging component cannot be positioned anywhere, so some views are not accessible. Sometimes, it is also useful to set free the work field. A view from a virtual imaging system using a novel view synthesis method can overcome that problem.

In this paper we use a trifocal approach without explicit 3D data to synthesize a virtual view that can be very useful in micromanipulation. In Image Based Rendering (IBR) literature, the novel view is close to real views, but in this paper we extrapolate the angle of view up to 85°. Section 2 summarizes the trifocal geometry and describes it use to generate a novel view. Section 3 presents a new method to automatically obtain the angle of view wanted. Section 4 presents experimental results.

2   Trifocal Transfer

The trifocal transfer is the method of IBR with implicit geometry. It only requires a weak calibration which implies the estimation of the fundamental matrix.

Trifocal transfer is based on the geometry of three views, named the trifocal geometry.
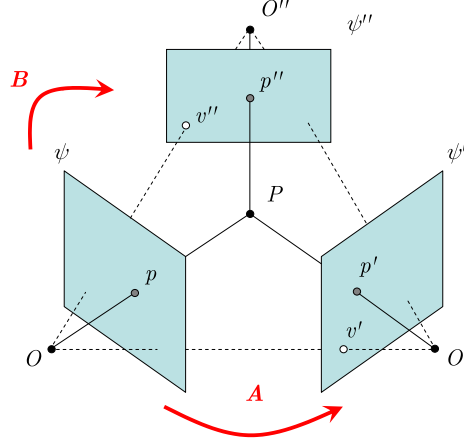
2.1   Trifocal Geometry and Trilinear Tensor

Fig. 1. Trifocal geometry.

Trifocal geometry is the extension of epipolar geometry to three views. Let us consider three views of $\mathcal{P}^2$ ψ, ψ' and ψ'' (Fig. 1). A point $P \in \mathcal{P}^3$ is projected onto the point $p = (x, y, 1)^T$ in ψ, $p' = (x', y', 1)^T$ in ψ' and $p'' = (x'', y'', 1)^T$ in ψ''. Let us note:

– A and B the collineation matrixes corresponding respectively to the projective transformations $\psi \rightarrow \psi'$ and $\psi \rightarrow \psi''$,
– v' and v'' the epipoles i.e. the projection of the optic center O on respectively ψ' and ψ''.

The trilinearity defines the constraint between three views [14]: p, p' and p'' are linked by the same projected point P. The epipolar geometry of (ψ, ψ') and (ψ, ψ'') allow to write the following equations:

$$\begin{aligned} p' &\cong Ap + \delta v' \\ p'' &\cong Bp + \delta v'' \end{aligned} \qquad (1)$$

where δ is the relative affine structure of P. The coefficient δ is independent of ψ', i.e., is invariant according to the choice of the second view [15]. Then δ can be isolated from both (1) to obtain a set of equalities. From those equalities trilinear equations linking p, p' and p'' can be recovered: four linearly independent equations with 27 distinct coefficients are obtained. Each of these is an element of the trilinear tensor $\mathcal{T}_i^{jk}$ $i, j, k \in [1, 3]$:
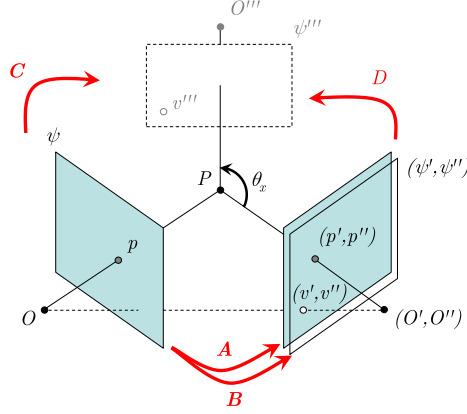
Fig. 2. The principe of novel view synthesis from two input views.

$$\mathcal{T}_i^{jk} = v'^j b_i^k - v''^k a_i^j \qquad (2)$$

where a_i^j and b_i^k are the elements of the collineation matrix A and B with $i, j, k \in [1, 3]$ (i is the index of the column, j is the index of the rows and k is the index of the layer). The trilinear tensor $\mathcal{T}_i^{jk}$ is a $3 \times 3 \times 3$ array of the 27 trilinear coefficients. Then the four trilinear equations, can be written with the tensor:

$$
\begin{aligned}
x'' \mathcal{T}_i^{13} p^i - x'' x' \mathcal{T}_i^{33} p^i + x' \mathcal{T}_i^{31} p^i - \mathcal{T}_i^{11} p^i = 0 \\
y'' \mathcal{T}_i^{13} p^i - y'' x' \mathcal{T}_i^{33} p^i + x' \mathcal{T}_i^{32} p^i - \mathcal{T}_i^{12} p^i = 0 \\
x'' \mathcal{T}_i^{23} p^i - x'' y' \mathcal{T}_i^{33} p^i + y' \mathcal{T}_i^{31} p^i - \mathcal{T}_i^{21} p^i = 0 \\
y'' \mathcal{T}_i^{23} p^i - y'' y' \mathcal{T}_i^{33} p^i + y' \mathcal{T}_i^{32} p^i - \mathcal{T}_i^{22} p^i = 0
\end{aligned}
\qquad (3)
$$

2.2   Novel View Synthesis by Trilinear Tensor

The first application of trilinear tensor to NVS is reported in [9] where three real views are used to compute the trifocal tensor and the virtual view: three real views lead to a virtual view. Later the authors proposed in [16] a more subtle approach that consists in merging two of the three input views: as a result, a novel view (virtual) is obtained from two real views.

Let us consider three views ψ, ψ' and ψ''. As explained above the trilinear tensor $\mathcal{T}(\psi, \psi', \psi'')$ can be calculated by (2). Now suppose ψ'' is merged with ψ' (Fig. 2). That means the collineation matrixes A ($\psi \rightarrow \psi'$) and B ($\psi \rightarrow \psi''$) and the epipoles v' and v'' are identical. Thus (2) becomes:

$$\mathcal{T}_i^{jk} = v'^j a_i^k - v'^k a_i^j \qquad (4)$$

This latest defines what is called the seed tensor. Let us suppose the view ψ'' becomes the view ψ''' by a collineation matrix D, then a collineation ma-
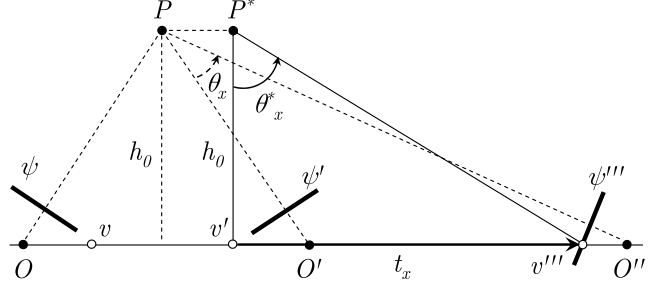
Fig. 3. The angle of view of ψ''' according to the translation t_x.

trix C links ψ and ψ'''. As the same, the seed tensor $\mathcal{T}(\psi, \psi', \psi'')$ changes to $\mathcal{G}(\psi, \psi', \psi''')$. As $C = DB$ the new tensor $\mathcal{G}$ can be computed from (2) and (4):

$$\mathcal{G}_i^{jk} = d_l^k \mathcal{T}_i^{jl} + t^k a_i^j \tag{5}$$

where $t^k = d_l^k v''^k - v'''^k$ is the element of the translation vector t that changes $v'' \to v'''$ and d_l^k is the element of collineation matrix of D, with $i, j, k, l \in [1, 3]$. Every point p''' of ψ''' can be calculated by (3):

$$x''' = \frac{x'\mathcal{G}_i^{31}p^i - \mathcal{G}_i^{11}p^i}{x'\mathcal{G}_i^{33}p^i - \mathcal{G}_i^{13}p^i} \qquad y''' = \frac{x'\mathcal{G}_i^{32}p^i - \mathcal{G}_i^{12}p^i}{x'\mathcal{G}_i^{33}p^i - \mathcal{G}_i^{13}p^i} \tag{6}$$

The process requires the weak calibration of the imagers and the points correspondence between both input views.

2.3 Computation of the Angle of View

We have exposed above how to synthesize a virtual view using trilinear tensor approach. In practice, the view ψ''' does not exist. It synthesis from ψ and ψ' requires the computation of the tensor element $\mathcal{G}_i^{jk}$ which is a function of two sets of parameters d_l^k and t^k. In order to simplify the synthesis D will be set equal to the identity matrix $I_{3\times 3}$. The translation vector is a function of the angle of view, $t = f(\theta)$, defined by the angle between the lines $[PO']$ and $[PO''']$ in the trifocal plane. We only use the component t_x and suppose it depends on θ_x. Thus the problem is to find the value of t_x for a given value of θ_x.

The value of θ_x is approximated by θ_x^*, the angle between the lines $[P^*v']$ and $[P^*v''']$ where $[PP^*]$ is parallel to $[OO']$ and $[P^*v']$is perpendicular to $[OO']$ (Fig. 3). Then t_x can be written:

$$t_x = h_0 \tan \theta_x^* \tag{7}$$

In order to estimate h_0 we create a dummy segment in a plane of $\mathcal{P}^3$ parallel to ψ' which we project in the view ψ'''. At $\theta_x^* = 0$ and then $t_x = 0$, the length of the pattern in ψ''' is L_0, and at $\theta_x^* \neq 0$ then $t_x \neq 0$ the length become $L(t_x)$. We can write:

Fig. 4. Our stereo images.

$$L_0 \cos \theta_x^* = L(t_x) \tag{8}$$

That equation can be solved using a Nelder-Mead simplex method [17]. That optimization method compares the values of the objective function with zero and does not require the use of any derivatives. A simplex in $\mathbb{R}^n$ is a set of $n+1$ points that do not lie in a hyperplane. For example a triangle is a simplex of 2 dimensions. In the Nelder-Mead method, the simplex can vary in shape from iteration to iteration following reflect, expand, contract and shrink. The simplex finds the minimal of (8) according to t_x. As soon as t_x and θ_x^* are known, h_0 can be computed according to (7). Thus the model of displacement is entirely defined and can be used for the computation of the tensor and then the novel view.

3   Experimental Results

The principles exposed above are used to compute a novel view, a lateral one, from two front views (Fig. 4).

3.1   Making the Dummy Pattern

This process requires the calculation of the disparity interval i.e. the displacement between every (left and right) couple of points. A Canny [18] is used to compute the edges and the matching is achieved by Zhang method [19] with a Sum of Squared Differences (SSD) correlation criterion. Usually the correspondent of the point p of the left image in the right one (p') is searched along the epipolar line. The edges transfer is enough for our experiment since our application is the surveillance and control of micromanipulation task. Complete images with texture and color are not necessary. At the end the correspondence of every point of the edges is achieved and the edge-map disparity is computed (Fig. 5). As exposed above the dummy pattern is inlayed in the virtual view and allows the computation of the displacement vector of the view. All the points of the pattern must be on the same plane i.e. at the same layer. For the later we choose the one for which the number of points is maximal.
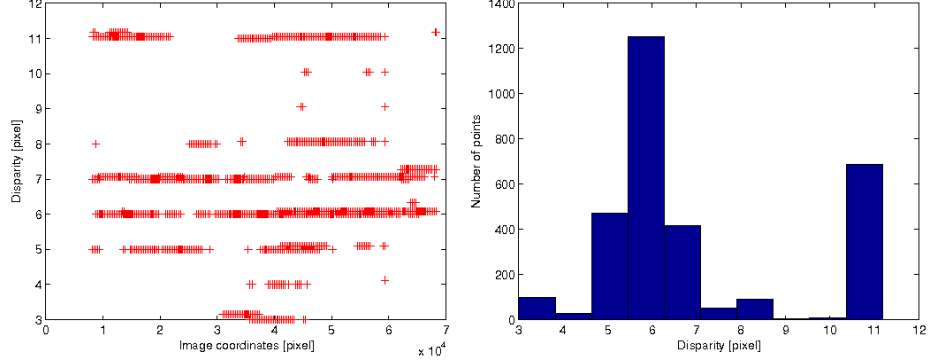
Fig. 5. Left, the layer representation of the disparity edge map between images. Right, distribution of disparity.
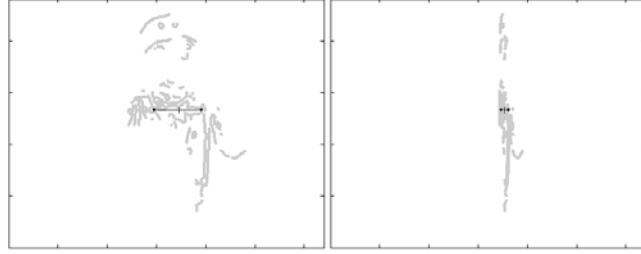

Fig. 6. Left, the master layer with the dummy pattern for $t_x = 0$. Right, the same with $t_x = 900000$.

Let us note Ω the set of points. Since the novel view displacement is reduced to θ_x (see above) Ω can be divided perpendicularly to x axis into left Ω_L and right Ω_R sets of points. The centroid of the three sets are computed and they coordinates are used to define the following points: (x_L, y), (x, y) and (x_R, y) where x, x_L and x_R correspond respectively to the x coordinate of the centroid of Ω, Ω_L and Ω_R and y corresponds to the y coordinate of Ω. Those points define the dummy pattern (Fig. 6).

3.2   Computing the View

It is impossible to compute the view for $\theta_x = 90^o$, in is this case t_x trends toward infinity according to (7). So, for the lateral view we choose a maximum angle of view of 85°. For an arbitrary angle of view of $\theta_x^* = 70^o$, the simplex method leads to a t_x of 255943.

Figure 7 shows the length of the pattern versus t_x and the angle of view (model and result) versus t_x. According to (7), the value of h_0 is 68580. Finally, for the view at 85°, the displacement vector is $t_x = 68580 \tan(\theta_x)$. Where the
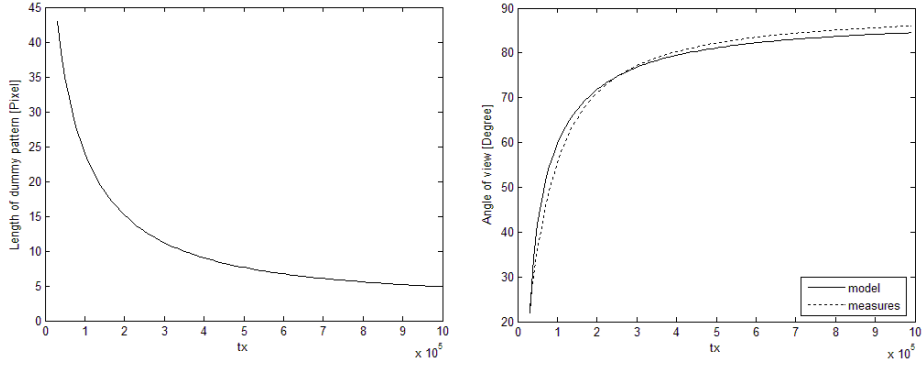
Fig. 7. Left, the size of the dummy pattern according to t_x. Right, comparison between model and experimental measures of the angle of view according to t_x.
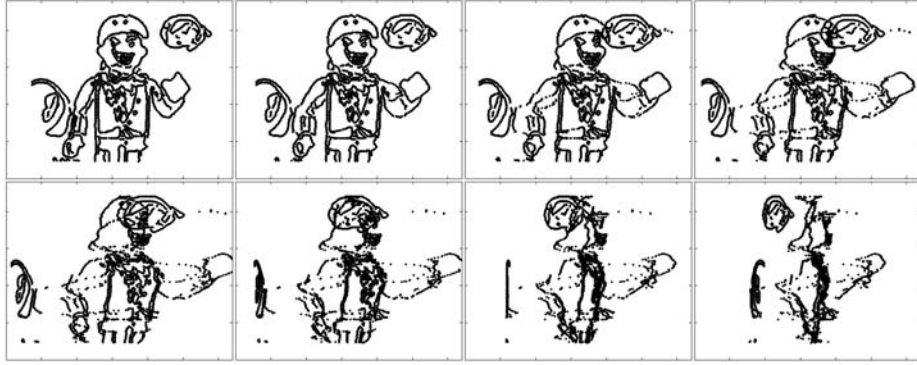

Fig. 8. Lateral views where the angle of view increases from 0° to 85°.

value of t_x is known, the lateral view is computed in real time with the two input views using (6).

But, that is not sufficient to ensure the quality of the view. The points are also rectified by minimizing their shift and maintaining the center of the pattern at the center of the view. Figure 8 shows eight lateral views from 0° to 85° angle of view.

3.3   Application to Micromanipulation

We apply above principles to a microassembly scene: the picking up of a microgear by a microgripper. The lateral view, at 85° angle of view, allows the access to the z position of the gripper according to the microgear. Two cases are considered: the part is not between the tips and the part is between the tips (Fig. 9).

Fig. 9. Top, stereo top images and lateral image of the scene with the microgear outside the gripper tips. Bottom, the same scene with the microgear inside the gripper tips.
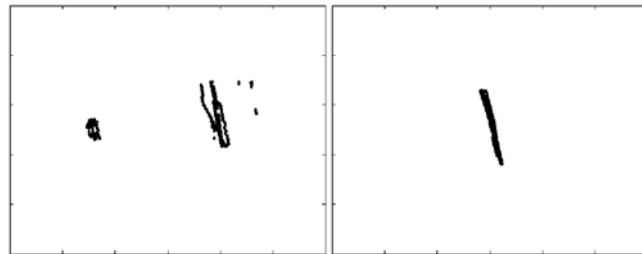

Fig. 10. The views at 85° according to views of Fig. 9. Left corresponds to top views, right to bottom views.

Figure 10 shows the views at an angle view of 85° for the two configurations. In spite the weak number of layers, it is possible to evaluate the distance between the gripper and the microgear. Then it is possible to know if the gear can be picked up or not.

4   Conclusion

We summarized trifocal geometry and explained how it allows the expression of trifocal constraints through trilinear tensor. This latest is required to transfer matched pixels in both input views into the virtual one. We quickly computed from the measure of the length of a dummy segment in a novel view and the Nelder-Mead simplex method, the model of the novel views pose. That model allows the computation of novel views with very large angle of view.

We applied that approach to images from a micromanipulation scene and showed that the obtained image of edges is enough to ensure the surveillance of the task.

Future work will deal with the deepening of the modelisation of the views pose and it application to synthesize virtual imagers in micromanipulation. The great merit of that idea is it will set free the work field.

References

1. Chen, S.E., Williams, L.: View interpolation for image synthesis. In: Computer Graphics (SIGGRAPH'93 Proceedings). (1993) 279–288
2. Shum, H.Y., Kang, S.B.: A review of image-based rendering techniques. In: SPIE Proceedings of the Visual Communications and Image Processing. (2000) 2–13
3. Chen, S.: Quicktime vr - an image-based approach to virtual environment navigation. In: Computer Graphics (SIGGRAPH'95 Proceedings). (1995) 29–38
4. Szeliski, R., Shum, H.Y.: Creating full view panoramic image mosaics and environment maps. In: Computer Graphics (SIGGRAPH'97 Proceedings). Volume 31. (1997) 251–258
5. Saito, H., Baba, S., Kanade, T.: Appearance-based virtual view generation from multicamera videos in the the 3-d room. In: IEEE Transactions on Multimedia. Volume 5. (2003) 303–316
6. Connor, K., Reid, I.: Novel view specification and synthesis. In: Proceeding of the British Machine Vision Conference, Cardiff, England (2002) 243–252
7. Irani, M., Hassner, T., Anandan, P.: What does the scene look like from a scene point? In: Proceedings of European Conference on Computer Vision (ECCV), Copenhagen (2002) 883–897
8. Faugeras, O., Robert, L.: What can two images tell us about a third one? Technical report 2018, INRIA (1993)
9. Avidan, S., Shashua, A.: Novel view synthesis in tensor space. In: IEEE Computer Vision and Pattern Recognition. (1997) 1034–1040
10. Yang, G., Gaines, J.A., Nelson, B.J.: A surpervisory wafer-level 3d microassembly system for hybrid mems fabrication. Journal of Intelligent and Robotic Systems **37** (2003) 43–68
11. Matsumoto, A., Akimoto, T., Yoshida, K., Inoue, H., Kamijo, K.: Development of mems component assembly machine - application of robotics technology to micromechatronics. In: The International Symposium on Micro-Mechanical Engineering. (2003) 83–88
12. Popa, D.O., Stephanou, H.E.: Micro and mesoscale robotic assembly. Journal of Manufacturing Process **6** (2004) 52–71
13. Sun, L., Xie, H., Rong, W., Chen, L.: Task-reconfigurable system for mems assembly. In: IEEE International Conference on Robotics and Automation, Barcelona, Spain (2005) 844–849
14. Shashua, A.: Algebraic functions for recognition. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 17. (1994) 779–789
15. Shashua, A., Navad, N.: Relative affine structure: Theory and application to 3d reconstruction from perspective views. In: IEEE Computer Vision and Pattern Recognition (CVPR). (1994) 483–489
16. Avidan, S., Shashua, A.: Novel view synthesis by cascading trilinear tensors. In: IEEE Transactions on Visualization and Computer Graphics (TVCG). Volume 4. (1998) 293–306
17. Nelder, J.A., Mead, R.: A simplex method for function minimization. Computer Journal **7** (1965) 308–313
18. Canny, J.F.: A computational approach to edge detection. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 8. (1986) 679–698
19. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.T.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical report 2273, INRIA (1995)