A trifocal transfer based virtual microscope for robotic manipulation of MEMS components.

Sounkalo Dembélé, Julien Bert, Brahim Tamadazte, Nadine Piat

A trifocal transfer based virtual microscope for robotic manipulation of MEMS components

Sounkalo Dembélé*(a) Julien Bert (b)
Brahim Tamadazte (a)
Nadine Piat (a)
(a) FEMTO-ST Institute UMR CNRS 6596 / UFC / ENSMM
24, rue Savary, 25000 Besançon, France
+33(0)381 40 27 98, firstname.lastname@femto-st.fr
(b) LaTIM INSERM U650
5, avenue Foch 29609 Brest cedex, France
+33(0)298 01 81 05, julien.bert@univ-brest.fr

June 9, 2010

Abstract

The paper deals with the problem of imaging at the microscale. The trifocal transfer based novel view synthesis approach is developed and applied to the images from two photon microscopes mounted in a stereoscopic configuration and observing vertically the work scene. The final result is a lateral virtual microscope working up to 6 frames per second with a resolution up to that of the real microscopes. Visual feedback, accurate measurements and control have been performed with, showing it ability to be used for robotic manipulation of MEMS parts.
Keywords: Novel view synthesis, trifocal tensor, photon microscope, microassembly, micromanipulation, MEMS.

Nomenclature

$\mathbf{A}$	two-D homography mapping each point of ψ to a point of ψ'
a_i	i^{th} column of $\mathbf{A}$ $(i = 1, 2, 3)$
a^i_j	ij^{th} entry of $\mathbf{A}$ $(i, j = 1, 2, 3)$
$\mathbf{B}$	two-D homography mapping each point of ψ to a point of ψ''
b_i	i^{th} column of $\mathbf{B}$ $(i = 1, 2, 3)$
b^i_j	ij^{th} entry of $\mathbf{B}$ $(i, j = 1, 2, 3)$
$\mathbf{D}$	rotation matrix between ψ' and ψ''

*corresponding author: sounkalo.dembele@femto-st.fr

d_i	i^{th} column of $\mathbf{D}$ $(i = 1, 2, 3)$
d^i_j	ij^{th} entry of $\mathbf{D}$ $(i, j = 1, 2, 3)$
d	working distance of the microscope [mm]
$e(N)$	error between the desired and current positions [pixel]
$\mathbf{F}$	fundamental matrix of the stereo microscope
$\mathbf{F}^i_j$	ij^{th} entry of $\mathbf{F}$ $(i, j = 1, 2, 3)$
f	focal length of the microscope [μm]
$\mathbf{K}$	intrinsic matrix of the microscope
k	scale factor of the microscope [pixel/μm]
O	centre of the right microscope
O'	centre of the left microscope
O''	centre of the virtual microscope
$P = (X, Y, Z, W)^\top$	point of the 3-space
$p = (x, y, 1)^\top = (p^1, p^2, p^3)^\top$	image of P in the view ψ
$p' = (x', y', 1)^\top = (p'^1, p'^2, p'^3)^\top$	image of P in the view ψ'
$p'' = (x'', y'', 1)^\top = (p''^1, p''^2, p''^3)^\top$	image of P in the view ψ''
$\mathcal{R}_i$	frame of the i^{th} object $(i = 1, 2, 3, 4)$
$\mathcal{T}$	tensor of the three-view system (ψ,ψ',ψ'')
$\hat{\mathcal{T}}$	tensor of the stereoview system (ψ,ψ')
$\mathbf{T}_i$	i^{th} matrix of $\mathcal{T}$ $(i = 1, 2, 3)$
$\mathcal{T}_i^{jk}$	ijk^{th} entry of $\mathcal{T}$ $(i, j, k = 1, 2, 3)$
$\mathbf{t}$	translation vector between ψ' and ψ''
t^i	i^{th} entry of $\mathbf{t}$ $(i = 1, 2, 3)$
$u(N)$	value of the control [μm]
$v = (v^1, v^2, v^3)^\top$	epipole of the left microscope
$v' = (v'^1, v'^2, v'^3)^\top$	epipole of the right microscope
$v'' = (v''^1, v''^2, v''^3)^\top$	epipole of the virtual microscope
$[v']_\times$	skew-symmetric matrix corresponding to v'
x_0	X-coordinate of the principal point of the microscope [pixel]
y_0	Y-coordinate of the principal point of the microscope [pixel]
z^a	initial position of the gripper [pixel]
z^b	final position of the gripper [pixel]
z^d	desired position of the gripper [pixel]
$\hat{z}$	estimation of the current position of the gripper [pixel]
α	angle of rotation between the frames of ψ' and ψ'' [$^\circ$]
Δz	displacement of the gripper [pixel]
γ	magnification of the microscope
π	plane of the 3-space
ψ	image from the right microscope
ψ'	image from the left microscope
ψ''	image from the virtual microscope
i, j, k, n, q, r, s	indexes
N	index of the sampling step

1 Introduction

1.1 Related work

Novel View Synthesis (NVS) is a part of computer vision introduced by Chen and Williams in 1993 (Chen and Williams, 1993). It deals with the obtaining of a maximum of views of an environment from a minimum of data, for example the construction of side views of an object from two frontal views. Two classes of methods in NVS may be considered: the model-based rendering (MBR) and the image-based rendering (IBR). In MBR, virtual environments are created from mathematical models, a typical example is 3D characters synthesis in movies and video games by modeler softwares. In IBR, a set of images of the scene is used to construct the novel views.

According to the knowledge about scene geometry Shum and Kang (Shum and Kang, 2000) propose the following classification for IBR: rendering with no geometry, rendering with explicit geometry and rendering with implicit geometry.
Rendering with no geometry concerns the fulfilling of a mosaic image from a set of local views (Chen, 1995); (Szeliski and Shum, 1997); (Bhosle et al., 2002); (Trakaa and Tziritasa, 2003).
Rendering with explicit geometry is close to MBR: its purpose is the reconstruction of 3D view from real views of the scene. It requires weak as well as strong calibrations of the system of views and is computationally expensive (Saito et al., 2003); (Li and Hartley, 2006); (Geys and Gool, 2007).
By opposition to above cases, rendering with implicit geometry requires only weak calibration. Three usual techniques of that type are the line-of-sight, the epipolar transfer and the trifocal transfer. The line-of-sight approach is based on ray-tracing, it main drawback is the high number of input images: at least ten images are required to obtain the synthetic view (Irani et al., 2002); (Connor and Reid, 2002); (Cooke et al., 2006). The epipolar transfer approach introduced by Faugeras and Robert is based on epipolar geometry where the epipolar constraint defines the point-line duality in the pair of reference views. It may be use to create a virtual view from two real views: each point of the virtual view is the intersection between the line of a point in the left view and the line of a point in the right view (Faugeras and Robert, 1993); (Connor and Reid, 2002). The trifocal transfer approach developed by Avidan and Shashua is based on the trifocal constraint between three views: with two reference images and a tensor, the points of the references are transferred into the novel view (Avidan and Shashua, 1997).

In addition to multimedia their usual domain of application (Chen and Williams, 1993); (Pollard et al., 2000); (Fehn et al., 2001); (Saito et al., 2003); (Trakaa and Tziritasa, 2003), virtual views generated by NVS may be of great interest in robotic micromanipulation which deals with the handling of objects in the range from 1 μm to 1 mm.

The main purpose of micromanipulation is assembly. In addition to biological objects like cells and pollen seeds, artificial objects are chemically or mechanically manufactured. As examples one may quote grains of powder like drugs or cosmetics, optomechatronic parts like balls, pegs, pins, threads, membranes, lenses, shutters and fibres.
Rarely these microparts define the final products, they usually must be assembled into 3D MEMS (Micro Electro Mechanical Systems) (Tsuchiya et al., 1999); (Yang et al., 2003); (Dechev et al., 2004); (Pascual, 2005); (Xie et al., 2007); (Sieber et al., 2008); (Tamadazte et al., 2009) and MOEMS (Micro Opto Electro Mechanical Systems) (Hoffmann and Voges, 2001); (Aoki et al., 2003); (Popa and Stephanou, 2004); (Pascual, 2005). For that purpose some robotic setup have been developed: (Sato et al., 1995), (Kasaya et al., 1999), (Nelson et al., 1999), (Ralis et al., 2000), (Yang et al., 2003), (Matsumoto et al., 2003), (Popa and Stephanou, 2004), (Shacklock and Sun, 2005), (Kim et al., 2006), (Xie et al., 2007), (Cvetanovic et al., 2008), (Cvetanovic et al., 2008), (Probst et al., 2009), (Fatikow et al., 2009), etc. In addition to robotic and gripping systems these setup always include at least one microscope based imaging system whose images enable either the tracking and recognition of objects or the control of the systems.

(Weinstein et al., 2004), (Potsaid et al., 2005) and (Bert et al., 2006) use an image mosaicing NVS to create virtual views that can be used in the manipulation of MEMS or bioMEMS parts. Several high resolution local images of the work scene are dynamically stitched to give a mosaic image representing a global view of the scene. The solution by Potsaid et al. is particularly interesting: a device is embedded in their system enabling the scanning of the scene in order to acquire the local images. The mosaic image and the local images have the same resolution but the size of the former is larger. It defines a virtual microscope with a high resolution and a large field-of-view, overcoming by the fact the resolution - field-of-view tradeoff of the microscope.
The present paper also addresses the use of NVS to create virtual views for the manipulation of MEMS parts, but instead of image mosaicing it uses an approach based on trifocal transfer.

1.2 Contribution

Two photon microscopes mounted in a stereoscopic configuration is considered. The angle between them is about 30°. Then the system gives regularly two vertical overlapped views of the scene. A location corresponding to a lateral viewpoint that gives the maximum of information about the geometry of the scene is defined: an angle of about 80° from the vertical axis (figure 1). Then the matched points of every pair of input images are transferred into a novel view positionned in the determined location. Only the points belonging to the edges are transferred since that solution defines the best tradeoff between the quality of information of the images and the rate of synthesis of the views.
The final result is a virtual microscope delivering lateral views of the workscene

from which depth information can be retrieved. It resolution is controllable and is up to that of the input microscopes. It is required to control a gripper and to avoid its collision with the part to pick up.

This is the first time trifocal transfer is studied and applied in real time to images from a photon stereo microscope. One may notice that it has been developed and applied until now to static images from conventional lens camera systems for multimedia purpose. The result is conclusive, the obtained images are rich in information that is easily exploitable.

The main advantage of that solution over the use of real microscope is the fact it sets free the work scene. Indeed the analysis of the manipulation setup found in the literature shows very cluttered work space with multiple degree-of-freedom (DOF) robotic system, grippers, multiple view imaging system and multiple lighting system. By setting free the workscene from imaging and lighting systems, more robotic degree-of-freedom may be added to improve the manipulation capability of the setup.

Since it is based on stereoscopy where the pair of images embedded depth information, that approach is close to pure 3D reconstruction i.e. 3D reconstruction without the stage of rendering. They share the same stage, the computation of correspondence between the points of both input images, and have equivalent computation times. But if one considers that stage of rendering, 3D reconstruction becomes longer since trifocal transfer has no intermediate stage, it renders directly 2D views from 2D views.

1.3 Contents

The paper is organized as followed. Section 2 states the setup used to perform the experiments. Section 3 summarizes the trifocal geometry i.e. the geometry of three views which is the basis of the NVS approach. Section 4 describes the synthesis of novel view by the trifocal geometry embedded in the trifocal tensor. Section 5 presents the way the stereomicroscope is calibrated and its geometry is recovered. Sections 6 describes the synthesis of views of some micromanipulation scenes, and the performing of the positioning of a gripper with respect to an object using of a vision based control with the virtual microscope.

2 Experimental setup

The setup includes two TIMM microscopes (from SPI) positioned with an angle of about 30 $^\circ$ between each other and observing vertically the work scene. Every microscope magnification, γ, varies between 0.0001$\times$ and 4$\times$. The working distance, d, varies from 10 mm to 70 mm. The microscopes are connected to an acquisition system that delivers 600 $\times$ 540 pixels with a frame rate up to 25. A xz positioning stage (from OWIS) that can support many types of gripper

according to the experiment to achieve completes the setup (figure 2). A PC (Pentium (R) D, CPU 2.80 GHz, 2 GB of RAM) is used to run the system and, C++ and Matlab (R) are used as programming languages.

3 Multiple view geometry

Trifocal transfer underlied the concepts of epipolar and trifocal geometries which will be briefly reminded below. A detailled analysis is stated in the excellent book from Hartley and Zisserman (Hartley and Zisserman, 2006). According to (Zhou and Nelson, 1999) and (Tamadazte et al., 2009) the photon microscope can be modelled by a linear projective model as any standard lens camera.

3.1 Epipolar geometry

The epipolar geometry is the projective geometry between two views. It is independent of scene structure, and only depends on the cameras internal parameters and relative pose. The fundamental matrix $\mathbf{F}$, a 3×3 array of rank 2, encapsulates this intrinsic geometry.

Consider two cameras of centres O and O' observing a scene. Let ψ and ψ' be the images obtained, respectively. The line (OO') defines the baseline. It intersection with ψ and ψ' defines the epipole v of the first camera and the epipole v' of the second camera, respectively.
Consider a point $P = (X, Y, Z, W)^\top$ in 3-space belonging to the plane $\pi = (0, 0, 0, 0)^\top$ not coplanar with O and O'. It is imaged as $p = (x, y, 1)^\top$ in ψ and $p' = (x', y', 1)^\top$ in ψ' (figure 3).

There is a 2D homography $\mathbf{A}$ induced by π mapping each point of ψ to a point of ψ':

$$p' = \mathbf{A}p \tag{1}$$

Then, one may write:

$$p = [\mathbf{I}|0]P \tag{2}$$

$$p' = [\mathbf{A}|v']P \tag{3}$$

The epipolar constraint linking p and p' is written:

$$p'^\top \mathbf{F} p = 0 \tag{4}$$

$\mathbf{F}$ is the fundamental matrix that may be written as:

$$\mathbf{F} = [v']_\times \mathbf{A} \tag{5}$$

with $[v']_\times$ the skew-symmetric matrix corresponding to v':
if $v' = (v'^1, v'^2, v'^3)^\top$ then $[v']_\times = \begin{pmatrix} 0 & -v'^3 & v'^2 \\ v'^3 & 0 & -v'^1 \\ -v'^2 & v'^1 & 0 \end{pmatrix}$.

As one can see **F** is independent of scene structure, it only depends on the cameras internal parameters and relative pose embedded in **A** and v': it represents the geometry of the two views. It is a 3×3 array of rank 2 whose estimation is also known as weak calibration.

3.2 Trifocal geometry

The trifocal geometry deals with the projective geometry between three views. The trifocal tensor $\mathcal{T}$, a 3×3×3 array, plays an analogous role in three views to that played by the fundamental matrix **F** in two views. It encapsulates all the projective geometric relations between three views that are independent of scene structure, and only depends on the cameras internal parameters and relative poses.

Let a third camera of center O'' be added to above stereo system, and let ψ'' be the corresponding image. The epipole v'' in ψ'' of the first camera can be defined as above. Let $p'' = (x'', y'', 1)^\top$ be the image of P in ψ'' (figure 3).

The following equations are added to equations 1 and 3 respectively:

$$p'' = \mathbf{B}p \tag{6}$$

$$p'' = [\mathbf{B}|v'']P \tag{7}$$

where **B** is the homography mapping ψ to ψ''.

The trifocal constraint between p, p' and p'' may be written:

$$[p']_\times (\sum_{i=1}^{3} p^i \mathbf{T}_i)[p'']_\times = 0_{3\times 3} \tag{8}$$

with:

$$\mathbf{T}_i = a_i v''^\top - v' b_i^\top \tag{9}$$

The terms a_i and b_i are respectively the i^{th} column of **A** and **B**.
The set of three 3×3 matrices $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ constitutes the trifocal tensor.
Equations 8 and 9 are stated in matrix notation, however tensor notation may be preferred.
An image point is represented by a homogeneous column 3-vector, i.e. $\mathbf{x} = (x^1, x^2, x^3)^\top$. The ij^{th} entry of a matrix **M** is denoted by m^i_j, superscript index i being the contravariant (row) index, subscript index j being the covariant (column) index. The convention states that indices repeated in the contravariant and covariant positions imply summation over the range (1,2,3) of the index. For instance, the equation $\mathbf{x}' = \mathbf{M}\mathbf{x}$ is equivalent to $x'^i = \sum_{j=1}^{3} m^i_j x^j$, which may be written $x'^i = m^i_j x^j$.
Then the constraint and the tensor become:

$$p^i (p'^n \epsilon_{njr})(p''^q \epsilon_{qks}) \mathcal{T}_i^{jk} = 0_{3\times 3} \tag{10}$$

$$\mathcal{T}_i^{jk} = a_i^j v''^k - v'^j b_i^k \tag{11}$$

with:

$$\epsilon_{rst} = \begin{cases} 0 & \text{unless } r, s \text{ and } t \text{ are distinct,} \\ +1 & \text{if } rst \text{ is an even permutation of 123,} \\ -1 & \text{if } rst \text{ is an odd permutation of 123.} \end{cases}$$

The tensor $\mathcal{T}$ represents the geometry of the three views, it is a 3×3×3 array, and thus has 27 entries, however, it has only 18 degrees of freedom.
The relation 10 actually leads to four independent relations or trilinearities.

4 Novel view synthesis using the trifocal tensor

4.1 The principle

The fundamental matrix is embedded in the tensor since one can consider the pair (ψ,ψ') as the triplet (ψ,ψ', ψ') and then rewrites equation 11:

$$\hat{\mathcal{T}}_i^{jk} = a_i^j v'^k - v'^j a_i^k = \epsilon_{rjk} \mathbf{F}_r^i \tag{12}$$

Then, the tensor can be used for the transfer of points between three images: given two matched points (p,p') in the pair (ψ,ψ') and a tensor $\mathcal{T}$, it is possible to determine the location of the corresponding point p'' in the third view ψ'' without reference to image content by resolving equation 10:

$$p''^k = p^i (p'^n \epsilon_{njr}) \mathcal{T}_i^{jk} \tag{13}$$

That approach enables the use of the whole experience on epipolar geometry, particularly the estimation of the fundamental matrix by points matching. However the interesting property of tensor transfer over others solutions, notably epipolar transfer, is the fact it continues to work even if the three camera centres are collinear. It also have been shown by Barrett et al. (Barrett et al., 1995) that in challenging situations, the tensor approach performs the best.

The transfer of points leads to the synthesis of novel views as stated by Avidan and Shashua (Avidan and Shashua, 1998) if the texture is also transferred: the texture of p'' is determined by a function of the textures of p and p'. To simplify the user definition of the tensor they propose to define the location of the novel view ψ'' with respect to the view ψ' as a rotation matrix $\mathbf{D}$ and a translation vector $\mathbf{t}$:

$$p''^i = a_j^i p'^j \quad (p'' = \mathbf{A} p') \tag{14}$$

$$\mathbf{t}^k = v''^k - v''^n d_n^k = v'' - v' \tag{15}$$

$$b_i^k = d_n^k a_i^n \quad (\mathbf{B} = \mathbf{D}\mathbf{A}) \tag{16}$$

Those equations leads to:

$$\mathcal{T}_i^{jk} = d_n^k \hat{\mathcal{T}}_i^{jn} + \mathbf{t}^k a_i^j \tag{17}$$

4.2 Application

The idea developed in the paper is based onto the use of the pair of images from the stereomicroscope, positionned over the scene, to repeatedly reconstruct a side view, at an angle α with respect to the right camera and at the working distance d from the work scene as if there is a real camera (Figure 1). Let $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, $\mathcal{R}_4$ and $\mathcal{R}_s$ be the frames of the left, right and novel microscope, an object and the scene, respectively. The calibration frame $\mathcal{R}_4$ is chosen in order to have D and t be defined as the rotation matrix of angle α and the translation vector d in $\mathcal{R}_4$, respectively.
Since the novel view is used mainly to track the gripper and the part to pick up in order to estimate their vertical relative position, only the edges of both objects are sufficient which avoids the transfer of every point and the texture as done usually in image rendering.

The process starts by:

- the calibration of the stereo microscope enabling the computation of the intrinsic matrix $\mathbf{K}$ (supposed the same for both microscopes) and the rotation matrices and translation vectors between the different frames,
- the recovery of the geometry of the stereomicroscope, i.e. the computation of the fundamental matrix $\mathbf{F}$.

If these operations are usual for standard lens camera, they require specific developments in the case of the microscope lens camera which are described below.

The main stage consists for every frame of the left and right microscopes to match the points of the edges and transfer them in the novel camera using equation 13.

5 Preprocessing

Microscope is characterized by a high optical magnification, a weak depth-of-field and a weak field-of-view. Then, the calibration as well as weak and strong becomes more difficult to implement than usual.

5.1 Calibration

As described by (Tamadazte et al., 2009) every microscope can be modelled according to the multiple scale paradigm because of its modifiable magnification. The standard projective model $\mathbf{K}$ is modified to explicitly includes the magnification value γ:

$$\mathbf{K} = \mathbf{K}(\gamma) = \begin{pmatrix} f(\gamma)k(\gamma) & 0 & x_0 \\ 0 & f(\gamma)k(\gamma) & y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{18}$$

The parameters f, k, x_0, y_0 are the focal length, the scale factor, the X and Y coordinates of the principal point, respectively. Actually the following simplifications are considered:

- the scale factors along X and Y axes are equal to k,
- the distorsions are neglected,
- the image axes are orthogonal.

The calibration involves two stages: the establishment of the function mapping each value of γ into a value of the scale factor k followed by the estimation of $\mathbf{K}$ for a given value of γ.
Because of the properties of the microscope (narrow depth-of-field, weak field-of-view) a planar virtual pattern is achieved from the tracking of a MEMS part. Feature points from that pattern is used in the calibration algorithm (refer to (Tamadazte et al., 2009) for more information). The following results were obtained:

- $\gamma = 4$,
- $k = 0.1$ pixel/μm,
- $f = 2000$ μm,
- $x_0 = 300$ pixels,
- $y_0 = 220$ pixels.

5.2 Geometry recovery

The recovery of epipolar geometry requires a sample that should satisfy some constraints. It should be textured in order to allow the use of a feature point detector. The texture must not be repetitive because of the ambiguity during the matching. The feature points must be at several depths for a reliable estimation. The images of the 400 μm $\times$ 400 μm $\times$ 100 μm (length,width, thickness, respectively) MEMS part respect these constraints (figure 4).

The recovery includes the following stages (Bert et al., 2006).

- Detection of invariant points is performed by an Harris detector modified to impose *a priori* the number of points in order to ensure enough points for the matching.
- Matching of points is performed by a ZNSSD (Zero-mean Normalized Sum of Squared Differences) correlation where the correlation window is sampled in order to improve the accuracy by interpolation, combined with a closer neighbor algorithm.
  Since views are close and are at the same level Harris detector (Harris and

Stephens, 1988) is enough.
Because of the short depth-of-field that prevents the obtaining of a focused view of the whole pattern, the images are acquired at several levels. The invariant points of the pair of the same level are matched, that is a way to improve the accuracy of the matching.

- RANSAC (RANdom SAmple Consensus) algorithm (Fischler and Bolles, 1981) is used to robustly compute the homography **A** and the fundamental **F** matrices. As expressed above the computation of **A** and **F** allows that of $\hat{\mathcal{T}}$.

6 Results

6.1 First experiment

A light hue gripper and part (500 μm diameter gear) are imaged. They are in plastic. The gripper is positioned 1 mm over the gear and an image is recorded from the left and right microscopes (figure 5). A Sobel edge detector is applied to both images and the points obtained are matched with an accuracy of 1/2 pixel i.e. an interpolation factor of 2. Then the points are transferred in the lateral view with the angle α varying from 0° to 80°.
Figure 6 shows the views obtained. The pixels blobs surrounded do not correspond to any object, they are the effects of false matchings. The result shows that the interpolation factor of 2 is not sufficient. To overcome that problem the interpolation factor is increased: the value of 20 i.e. an matching accuracy of 1/20 pixel gives a very good result and then is used in next experiments.

The views obtained are rich in geometric information, notably depth information: the gear is underneath the gripper and then cannot be picked up.

6.2 Second experiment

A dark hue gripper and part (400 μm $\times$ 400 μm $\times$ 100 μm corresponding to length, width and thickness, respectively) are imaged (figure 7). They are in silicon as usual for MEMS components. The part is exactly between the gripper fingers so it can be picked up. Left and right images are recorded and a Canny edge detector is applied to them. As explained above the edge points are matched with an accuracy of 1/20 pixel leading to no false matching.
Figure 8 shows the views obtained for the angle α varying from 0° to 80°. Depth information is correct since one can see clearly that the part is between the fingers and can be picked up.

6.3 Third experiment

The silicon gripper and part are still considered with the virtual microscope at the angle α of 80°. A third TIMM microscope is positioned in the location

of the virtual microscope and the images are compared. For that purpose the gripper is displaced over the part from a position z^a to a position z^b and some measurements are made. Figure 11 shows the lateral virtual image with pixel blobs corresponding to the gripper and the part. The following results are obtained:

- the controller of the positioning stage indicates 424 μm,
- the distance of 39 pixels is measured in the virtual image,
- the distance of 42 pixels is obtained in the real image.

Those measurements lead to the same scale factor for both microscopes, real as well as virtual: 0.1 pixel/μm.

The rate of synthesis is also evaluated: it is possible to synthesis up to 6 frames (600×540 pixels) per second. This result combined with the measurement capability leads to the performing of a vision based control in the virtual views.

6.4 Fourth experiment

The setup includes the two-TIMM stereomicroscope and the xz positioning stage where only the z DOF is controlled (figure 2). A third TIMM is still positionned laterally but does not intervene in the experiment.

Now, the task to achieve is the vertical positioning of the gripper. From every input pair of frames, the virtual image is generated and the Z position of the gripper is estimated in the latter. In a control viewpoint, this becomes the regulation toward zero of the error between the desired, z^d, and current, $\hat{z}$, gripper positions. At every sampling step N the control $u(N) = \frac{e(N)}{N}$ is applied to the positioning stage supporting the gripper (figure 12). Usually that control law leads to a polynomial decrease of the error.

Figure 13 shows initial and final images of that experiment, and the corresponding real images. Figure 14 shows the evolution of the positioning error with respect to the number of iterations. It decreases towards zero and stabilizes at $0\pm3\mu m$.

7 Conclusion

The paper develops and applies the novel view synthesis of the type trifocal transfer to the photon microscope imaging system which is widely used in the setup dedicated to 3D MEMS assembly. The matched points of every two images from two microscopes mounted in a stereoscopic configuration are transferred directly into a novel view positioned laterally. The transfer is limited to the

edges since the targetted application is assembly of MEMS devices where the rendering is less important than in usual NVS applications. It is based on the trilinear tensor of the three views computed from accurate image feature detection and robust matching.

The rate of synthesis reaches 6 frames (600×540 pixels) per second leading to the obtaining of an absolute virtual microscope. The resolution of the latter is up to that of input ones. Experiments involving two types of gripper, a 500 μm diameter gear and a 400 μm $\times$ 400 μm $\times$ 100 μm part have shown the relevance of the concepts.

Trifocal transfer is usually used in multimedia applications, that paper shows that microassembly is another possible application. The view obtained are rich in information and set free the work space that are usually cluttered in assembly setup. In a more general point of view, microscopy can be considered as a promising application since stereomicroscope is being used more and more.

Later the results obtained in this paper will be extended to a compact stereomicroscope of the type MZ16A from Leica. In the latter two optical paths deflected 10 $^\circ$ with respect each other lead to a left and a right cameras of 1024 $\times$ 772 pixels and a rate of 12 fps. The zoom and thus the magnification, and the focus are motorized and may be controlled by a computer. The work distance is approximately 112 mm. That microscope is an element of a setup enabling the robotic assembly of MEMS parts which also includes a 5-DOF (degrees of freedom) robotic system and a 4-DOF gripper (Figure 15). The tasks to perform require the tracking of the gripper fingers as well as the parts to pick up in order to control their relative position and to avoid any collision that may lead to damage of the system.

References

Aoki, K., H. T. Miyazaki, H. Hirayama, K. Inoshita, T. Baba, K. Sakoda, N. Shinya, and Y. Aoyagi (2003). Microassembly of semiconductor three-dimensional photonic crystals. *Nature Materials 2*(2), 117–121.

Avidan, S. and A. Shashua (1997). Novel synthesis in tensor space. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June*, pp. 1034–1040.

Avidan, S. and A. Shashua (1998). Novel view synthesis by cascading trilinear tensors. *IEEE Transactions on Visualization and Computer Graphics 4*(4), 293–306.

Barrett, E., P. Payton, and G. Gheen (1995, July). Robust algebraic invariant methods with applications in geometry and imaging. In T. F. Schenk

(Ed.), *Proceedings of SPIE Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes*, Volume 2572, pp. 30–42.

Bert, J., S. Dembélé, and N. Lefort-Piat (2006). Synthesizing a virtual imager with a large field of view and a high resolution for micromanipulation. In *5th International Workshop on MicroFactories, IWMF'2006, october 25-27, 2006, Besanon, France*, pp. 230–238.

Bhosle, U., S. Chaudhuri, and S. D. Roy (2002). A fast method for image mosaicing using geometric hashing. *IETE-Journal of Research 48*(3-4), 317–324.

Chen, S. (1995). Quicktime vr: an image-based approach to virtual environment navigation. In *Computer Graphics (SIGGRAPH'95 Proceedings)*, pp. 29–38.

Chen, S. E. and L. Williams (1993). View interpolation for image synthesis. In *Computer Graphics (SIGGRAPH'93 Proceedings)*, pp. 279–288.

Connor, K. and I. Reid (2002). Novel view specification and synthesis. In *Proceedings of British Machine Vision Conference, BMVC*, pp. 243–252.

Cooke, E., P. Kauff, and T. Sikora (2006). Multi-view synthesis: A novel view creation approach for free viewpoint video. *Signal Processing : Image Communication 21*, 476–492.

Cvetanovic, A., A. Cvetanovic, M. Soucek, D. Andrijasevic, and W. Brenner (2008). Micro assembly in a sem chamber and the solution for collision prevention. *Microsystems Technology 14*, 835–839.

Dechev, N., W. L. Cleghorn, and J. K. Mills (2004). Microassembly of 3-d microstructures using a compliant, passive microgripper. *Journal of Microelectromechanical Systems 13*(2), 176– 189.

Fatikow, S., C. Dahmen, T. Wortmann, and R. Tunnel (2009). Visual feedback methods for nanohandling automation. *International Journal of Information Acquisition 6*(3), 159–169.

Faugeras, O. and L. Robert (1993). What can two images tell us about a third one ? Technical Report 2018, INRIA, France.

Fehn, C., P. Kauff, O. Schreer, and R. Schafer (2001). Interactive virtual view video for immersive tv applications. In *Proceedings of International Broadcast Conference*, pp. 14–18.

Fischler, M. A. and R. C. Bolles (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*(6), 381–395.

Geys, I. and L. V. Gool (2007). View synthesis by the parallel use of gpu and cpu. *Image and Vision Computing 25*, pp. 1154–1164.

Harris, C. and M. Stephens (1988). A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151.

Hartley, R. and A. Zisserman (2006). *Multiple view geometry in computer vision* (2nd edition ed.). Cambridge, United Kingdom: Cambridge University Press.

Hoffmann, M. and E. Voges (2001). New silicon-based fibre assemblies for applications in integrated optics and optical mems. *Applied Physics B 73*, 629–633.

Irani, M., T. Hassner, and P. Anandan (2002). What does the scene look like from a scene point ? In *Proceedings of European Conference on Computer Vision (ECCV), Copenhagen (2002)*, pp. 883–897.

Kasaya, T., H. Miyazaki, S. Saito, and T. Sato (1999). Micro object handling under sem by vision-based automatic control. In *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, pp. 2189–2196.

Kim, B., H. Kang, D.-H. Kim, and J.-O. Park (2006). A flexible microassembly system based on hybrid manipulations scheme for manufacturing photonics components. *The International Journal of Advanced Manufacturing Technology 28*(3-4), 379–386.

Li, H. and R. Hartley (2006). Inverse tensor transfer with applications to novel view synthesis and multi-baseline stereo. *Signal Processing : Image Communication 21*, 724–738.

Matsumoto, A., T. Akimoto, K. Yoshida, H. Inoue, and K. Kamijo (2003). Development of mems component assembly machine - application of robotics technology to micromechatronics. In *The International Symposium on Micro-Mechanical Engineering, December 1-3*, pp. 83–88.

Nelson, B. J., S. Ralis, Y. Zhou, and B. Vikramaditya (1999). *Force and vision feedback for robotic manipulation of the microworld*, Volume 250/2000 of *Lecture Notes in Control and Information Sciences*, Chapter Haptics In Experimental Robotics VI, pp. 433–442. Springer Berlin, Heidelberg.

Pascual, D. N. (2005). Fabrication and assembly of 3d mems devices. *Solid State Technology*, 38–42.

Pollard, S., M. Pilu, S. Hayes, and A. Lorusso (2000). View synthesis by trinocular edge matching and transfer. *Image and Vision Computing 18*, pp. 749–757.

Popa, D. O. and H. E. Stephanou (2004). Micro and mesoscale robotic assembly. *Journal of Manufacturing Process 6*(1), 52–71.

Potsaid, B., Y. Bellouard, and J. T. Wen (2005). Adaptive scanning optical microscope (asom): a multidisciplinary optical microscope design for large field of view and high resolution imaging. *Optics Express 13*(17), 6504–6518.

Probst, M., C. Hrzeler, R. Borer, and B. J. Nelson (2009). A microassembly system for the flexible assembly of hybrid robotic mems devices. *International Journal of Optomechatronics 3*(2), 69–90.

Ralis, S., B. Vikramaditya, and B. J. Nelson (2000). Micropositioning of a weakly calibrated microassembly system using coarse-to-fine visual servoing strategies. *IEEE Transactions on Electronics Packaging Manufacturing 23*(2), 123–131.

Saito, H., S. Baba, and T. Kanade (2003). Appearance-based virtual view generation from multicamera videos captured in the 3-d room. *IEEE Transactions on Multimedia 5*(3), 303–316.

Sato, T., T. Kameya, H. Miyazaki, and Y. Hatamura (1995). Hand-eye system in nano manipulation world. In *Proceedings of the 1995 IEEE International Conference on Robotics and Automation, ICRA 1995, May 21-27, Nagoya, Japan*, pp. 59–66.

Shacklock, A. and W. Sun (2005). Integrating microscope and perspective views. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, Barcelona, Spain, 18-22 April*, pp. 454–459.

Shum, H. and S. B. Kang (2000). A review of image-based rendering techniques. In *Proceedings of the Visual Communications and Image Processing*, pp. 2–13.

Sieber, A., P. Valdastri, K. Houston, A. Menciassi, and P. Dario (2008). Flip chip microassembly of a silicon triaxial force sensor on flexible substrates. *Sensors and Actuators A 142*(1), 421–428.

Szeliski, R. and H.-Y. Shum (1997). Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques,SIGGRAPH'97*, Volume 31, pp. 251–258.

Tamadazte, B., S. Dembélé, G. Fortier, and N. L. Fort-Piat (2009). Robotic micromanipulation for microassembly: Modelling by sequencial function chart and achievement by multiple scale visual servoings. *Journal of Micro-Nano Mechatronics 5*(1-2), 1–14.

Tamadazte, B., S. Dembélé, and N. Piat (2009). A multiscale calibration of a photon video microscope for visual servo control: Application to mems micromanipulation and microassembly. *Sensors & Transducers Journal (Special issue in Robotics and Sensors Environments) 5*, 37–52.

Tamadazte, B., N. L. Fort-Piat, E. Marchand, and S. Dembélé (2009). Microassembly of complex and solid 3d mems by 3d vision-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'09, Saint-Louis, Missouri: United States*, pp. 3284–3289.

Trakaa, M. and G. Tziritasa (2003). Panoramic view construction. *Signal Processing: Image Communication 18*, 465–481.

Tsuchiya, K., A.Murakami, G. Fortmann, M. Nakao, and Y. Hatamura (1999). Micro assembly and micro bonding in nano manufacturing world. In *SPIE Proceedings*, Volume 3834, pp. 132–140.

Weinstein, R. S., M. R. Descour, C. Liang, G. Barker, K. M. Scott, L. Richter, E. A. Krupinski, A. K. Bhattacharyya, J. R. Davis, A. R. Graham, M. Rennels, and W. C. Russum (2004). An array microscope for ultrarapid virtual slide processing and telepathology. design, fabrication and validation study. *Human Pathology 35*(11), 1303–1314.

Xie, H., W. Rong, and L. Sun (2007). A flexible experimental system for complex microassembly under microscale force and vision-based control. *International Journal of Optomechatronics 1*(1), 81–102.

Yang, G., J. A. Gaines, and B. J. Nelson (2003). A surpervisory wafer-level 3d microassembly system for hybrid mems fabrication. *Journal of Intelligent and Robotic Systems 37*(1), 43–68.

Zhou, Y. and B. J. Nelson (1999). Calibration of a parametric model of an optical microscope. *Optical Engeenering 38*(12), 1989–1995.

8 Figures and tables

List of Figures

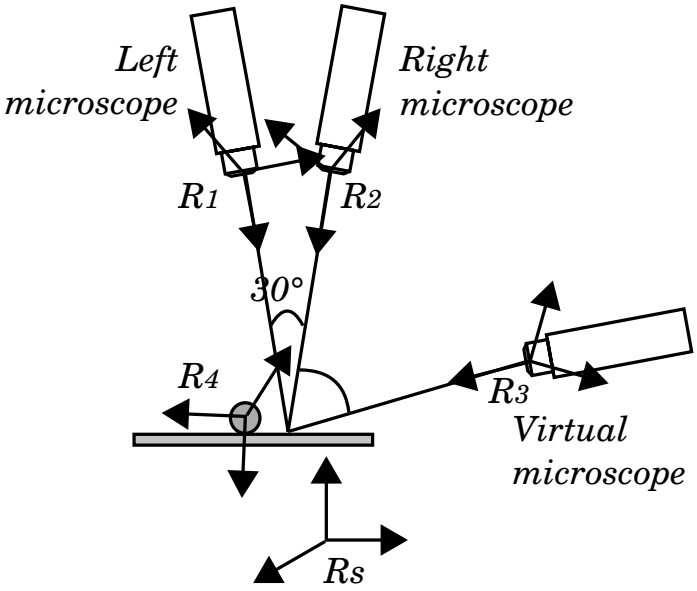


Figure 1: Configuration of the 3-view imaging system. $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, $\mathcal{R}_4$ and $\mathcal{R}_s$ are the frames of the left, right and novel microscope, an object and the scene, respectively.

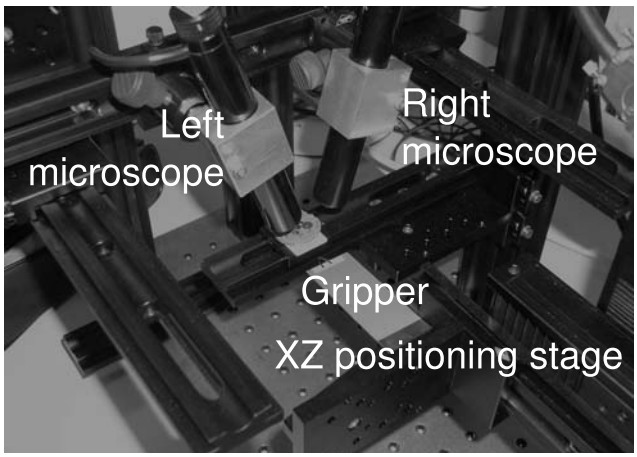


Figure 2: Experimental setup.

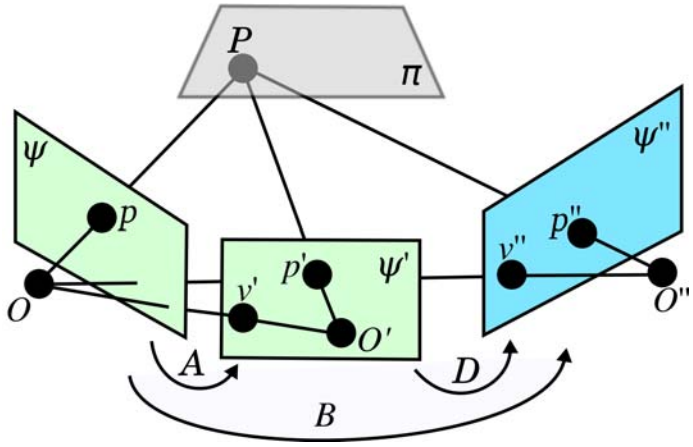


Figure 3: Trifocal geometry.

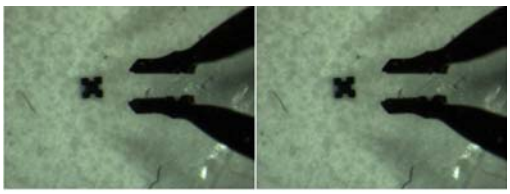


Figure 4: Left and right images used to recovery geometry: the 400 μm $\times$ 400 μm $\times$ 100 μm part is underneath the gripper.

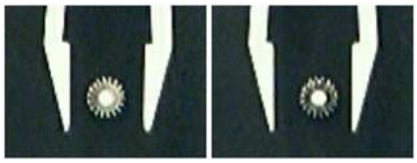


Figure 5: Input left and right views respectively: the gripper is over the gear.

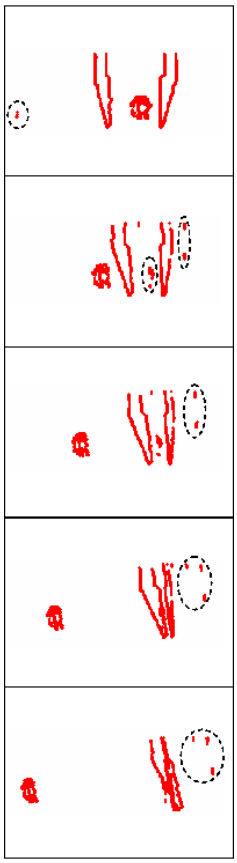


Figure 6: Novel views with the angle α varying from 0° to 80°: the gripper is over the gear. The surrounded elements correspond to errors of matching.

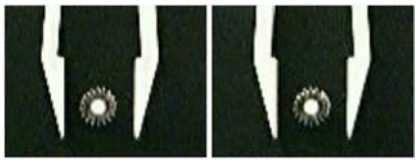


Figure 7: Input left and right views respectively: the gear is within the gripper.

Figure 8: Novel views with the angle α varying from 0° to 80°: the gear is within the gripper.

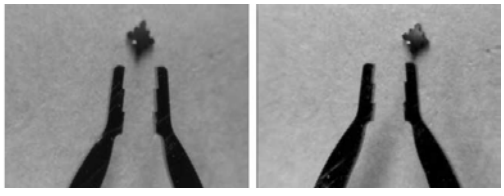


Figure 9: Input left and right views respectively: the gripper is over the part.

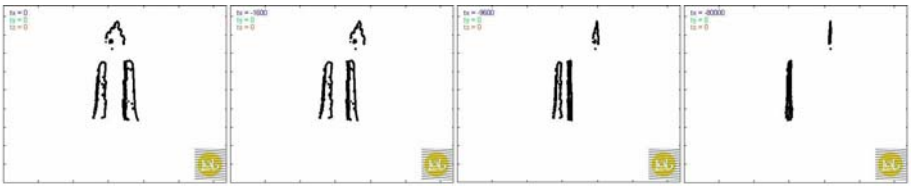


Figure 10: Novel views with the angle α varying from 0° to 80°: the gripper is over the part.

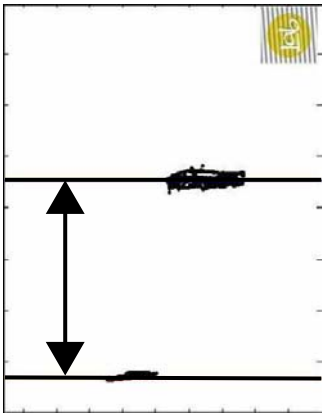


Figure 11: Scale factor estimation for the virtual microscope.

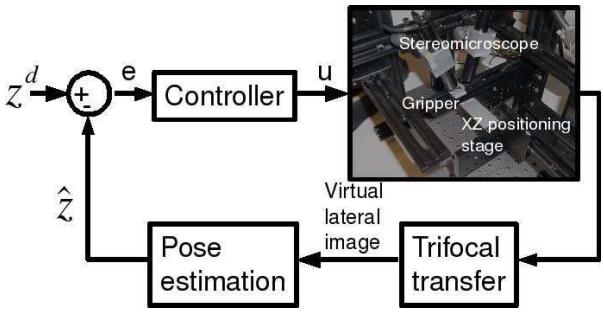


Figure 12: Vision based control scheme with the virtual images.

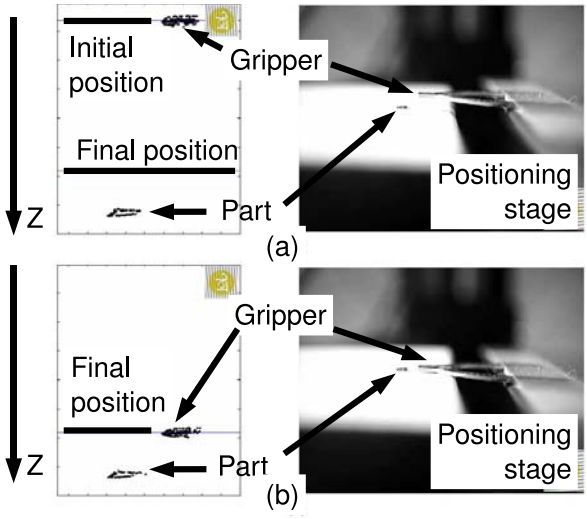
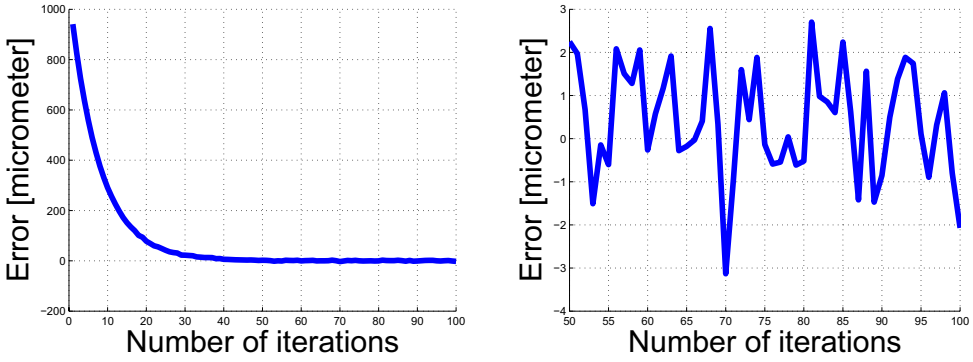


Figure 13: Positioning of the gripper with respect to the part: (a) virtual and real initial views, respectively (b) virtual and real final views, respectively.



(a) Error with 100 iterations.

(b) Zooming into the last 50 iterations.

Figure 14: Positioning of the gripper with respect to the part: evolution of error with respect to the number of iterations.

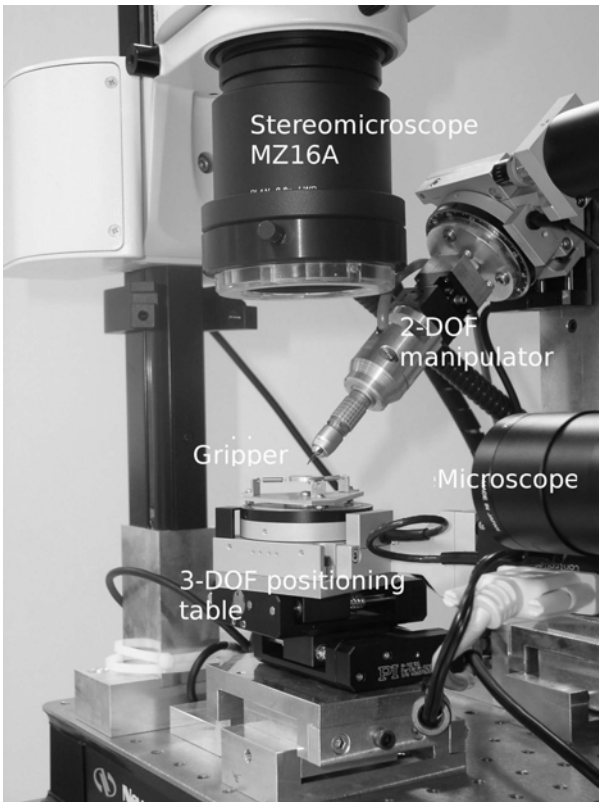


Figure 15: Assembly setup of FEMTO-ST Institute.