



HAL
open science

Estimating the 3D orientation of a microgripper by processing the focus data from the images delivered by a videomicroscope.

Guillaume Fortier, Brahim Tamadazte, Sounkalo Dembélé, Nadine Le Fort-Piat

► **To cite this version:**

Guillaume Fortier, Brahim Tamadazte, Sounkalo Dembélé, Nadine Le Fort-Piat. Estimating the 3D orientation of a microgripper by processing the focus data from the images delivered by a videomicroscope.. IEEE International Workshop on Robotic and Sensors Environments, ROSE'08., Oct 2008, Ottawa, Canada. pp.58-63, <10.1109/ROSE.2008.4669181>. <hal-00338546>

HAL Id: hal-00338546

<https://hal.science/hal-00338546v1>

Submitted on 13 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Estimating the 3D orientation of a microgripper by processing the focus data from the images delivered by a videomicroscope

Guillaume Fortier, Brahim Tamadazte, Soukalo Dembélé and Nadine Le Fort-Piat

FEMTO-ST Institute, UMR CNRS 6174 - UFC / ENSMM / UTBM,
Automatic Control and Micro-Mechatronic Systems Department,
24 rue Alain Savary, 25000 Besançon, France

Phone: +33 (0)3 81 40 27 97, Fax: +33 (0)3 81 40 28 09 ,
E-mail: {gfortier, btamadazte, sdembele, npiat}@ens2m.fr

Abstract – Visual servoing is a key technique to automate microassembly tasks since videomicroscope is widely used to recover information about the scene. The parts as well as the gripper should be tracked in the images which do not deliver a complete focused view of the objects because of the weakness of the depth-of-field. The paper investigates the problem of computing the 3D orientation of a rigid body by analysing the focus data extracted from the images of a videomicroscope. The solution is based on the shape-from-focus approach. But the modelling stage is simplified, instead of performing a 3D modelling, a 1D modelling is achieved which enables the estimation of the 3D orientation of the object. The concept is applied with success to the tips (70 μm width and 1500 μm length) of a gripper imaged by a 10 \times videomicroscope.

Keywords – Shape-from-focus, model construction, orientation estimation, visual tracking, microscopy, microassembly

I. INTRODUCTION

Visual servoing appears to be an incontrovertible control approach for the performing of microassembly tasks since the image is the main information source of microassembly scenes. Furthermore visual servo is precise and robust to the errors of modelling of the robotic system as well as the imaging system. Those images usually come from high magnification videomicroscopes characterized by a weak depth-of-field ([20]). Microassembly by means of visual servo requires fast (real-time if possible) tracking of the objects (sub-millimeter or micrometric parts) in the images. Tracking is a typical purpose of machine vision, it corresponds to the repeated locations of instances of a particular object, or class of objects, in successive frames of a video sequence ([2]). The information delivered can be the main motion in the image (a geometrical transformation of the type affine, euclidean or projective, ...) as well as the pose (2D or 3D) of the target in a reference frame. The use of some knowledge about the target, usually its description or model, can notably improve the process.

The plentiful approaches of visual tracking can be classified as exposed below ([2], [12]). The first class includes the approaches where no model of the target is used : optic flow or spatio-temporal approaches. The seeking result is obtained from the computation of spatial and temporal derivatives of images ([24]). The second class concerns the approaches where an implicit model of the target is considered : feature-based approaches. Indeed, knowledge is not absent, it is implicit since the presence of features (points, segments of line, ...) is the preliminary condition. The core of that class is the matching of above features between at least two images ([11], [7]). The third class contains the approaches where a 2D photometric model (or template of the type region) is used : appearance based approaches. That template is matched onto the current image of the sequence in order to minimize a correlation criterium ([2], [8], [1]). The fourth class also includes approaches where a 2D model of the target is used, but the latter is of the type geometric : active (or deformable) contours approaches ([2]). The fifth and last class concerns the approaches using a 3D model (of the type geometric) : 3D (CAD) model-based approaches. Images of the scene are compared to that derived from the model ([17], [6], [4], [21], [10], [5]). That class is very suitable for industrial applications where a CAD model of the handled part is often available.

Methods have been developed for images from the standard lens imaging system (standard camera, indoor or outdoor scenes), however, some of them are applicable to the microscope imaging (video microscope for indoor scenes like microassembly scenes) as shown by Yesin and Nelson ([21]) where the work of Cippola and Drummond is extended ([6]). For the videomicroscope, some specific tracking approaches have been developed. Zhang et al. ([23]), Sandoz et al. ([16], [15]), Brufau et al. ([3]), Kim et al. ([9]) process fringe patterns to recover the pose with very high accuracy (the position or motion is derived from phase shift of the fringes).

This paper investigates the tracking of rigid bodies in the images delivered by a high magnification videomicroscope,

more precisely a solution is proposed for the estimation of the 3D orientation of a two-finger microgripper. Because of the high optical magnification, the depth-of-field of the used videomicroscope becomes very weak, then it is not possible to get a complete focused image of any 3D object, and then to use above tracking approaches. Some stack images of the scene are acquired (by scanning the scene), but instead of achieving a 3D modelling (reconstruction) of objects ([13], [14], [19], [22]) only the medium lines (1D modelling) are computed enabling the estimation of the orientation in the videomicroscope frame.

Section 2 states the modelling of an object from the analysis of focus in the images, section 3 presents the computation of the orientation and section 4 presents the results with a $10\times$ videomicroscope and a two-finger microgripper.

II. MODELLING BY SHAPE FROM FOCUS

In microscopy, because of the limitation of the depth-of-field, it is not often possible to completely perceive a 3D object. The focus must be continuously adjusted to view the region of interest (ROI). To compensate this drawback, the depth-from-focus method can be used to get a full 3D representation of the object. That method consists in computing the focused area of every image of the scene image sequence. Each image is acquired at a different focus (corresponding to a depth), from the same point of view. It can be noticed that a videomicroscope with a motorized focus enables obtaining equidistant focal planes. Next, the focused areas are stacked up according to their position to give the 3D reconstruction of the scene. The different stages of the approach are exposed below.

A. Specification of a binary mask

The first stage of the shape-from-focus is to identify, in the set of images, the object surfaces to be reconstructed. A mask must be accurately found to eliminate all the pixels that do not correspond to the object. This mask is generated automatically by thresholding all the set of images with a common threshold. In a first time, this common threshold is defined manually. To eliminate the false detections because of the blur on the edges of the object, the areas of interest of all the images are blended and only the pixels which are present in all the thresholded images are kept to define the mask.

B. Maximization of the variance

The computation of the focus of an image corresponds to the measure of the sharpness of that image. A lot of focus estimators can be found in the literature ([18]): based on variance, standard deviation, gradient, Haar wavelets, The variance based estimator establishes a trade-off between rapidity and accuracy, and then is used in that paper.

In the shape-from-focus, the variance based estimator is used as followed. For the same pixel (i,j) of each image (N) of the sequence, the local variance in a $n\times n$ neighborhood is computed (Fig.1):

$$V_{(x,y)}(N) = \frac{1}{n^2} \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} (I_{(i,j)} - \overline{I_{(x,y)}})^2 \quad (1)$$

with:

- $V_{(x,y)}(N)$: the value of the variance at the pixel (x,y) of the image N , with $N = \{0, 1, \dots, \text{Number of Images}\}$
- n : the size of the filter
- k : depends on the neighborhood, equals to $(n-1)/2$
- $I_{(i,j)}$: gray value of the pixel (i,j)
- $\overline{I_{(x,y)}}$: mean gray value in the $n\times n$ neighborhood of the pixel (x,y)

The depth of each 3D point (X,Y,Z) is computed by maximizing the variance. X and Y correspond to the coordinates of the pixel (i,j) . Z is equal to the value of the focus for the image N^* with the highest variance $V_{(x,y)}(N^*)$. With a motorized-controlled focus, the focus value can be obtained directly and accurately. So, a sequence of 3D points (X,Y,Z) can be computed representing the 3D reconstruction of the scene. As the value of Z is calculated by the sharpness in a neighborhood, to improve the robustness of the algorithm, the object to be reconstructed must be textured.

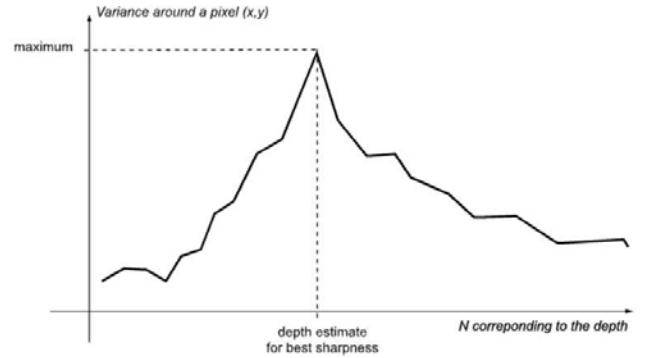


Fig.1 Variance at the same pixel (x,y) in a sequence of images

C. Filtering of points

The result of the shape-from-focus is a set of 3D points that is a 3D representation of the object. But some errors coming from the quality of the image, the accuracy of the mask and the calculation of the variance, cause false detections. Moreover, the size of the acquisition step also influences point detection. A high step leads to sharp distinct regions, but a weak step leads to common sharp regions on several images. So, to avoid that case, the step must be higher than the depth-of-field of the videomicroscope. Finally, to get relevant informations, some filters are applied on the sequence of points. Due to the limited

depth of field, the outliers are mainly positioned at the boundary of the point cloud, where the blur is more important. The point cloud defines the sharpness area. In this area, the result of maximum variance is well-defined. According to the accuracy at the boundary of the binary mask, some pixels are exploited but do not belong to the object. The idea of the spatial filtering is to take off the points far from the mean distribution. Firstly a factor analysis is performed onto each stack of points in order to extract the two main axes (axis 1 orthogonal to axis 2) of the data (Fig.2). They are defined respectively by the first and second eigenvectors of the covariance matrix of the data. Those axes associated with the barycentre become the frame for the spatial distribution filtering.

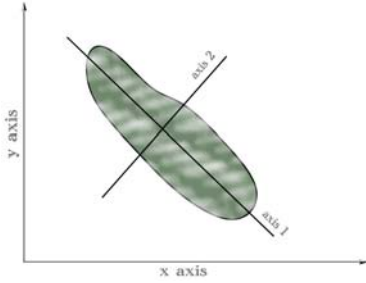


Fig.2 The two factorial axis of set of points

Secondly, the spatial distribution along every factorial axis is smoothed by an appropriate filter (for example, a locally weighted scatter plot smooth using linear least squares fitting and a second-degree polynomial). Thirdly, the standard deviation (σ) and the maximum of points (\max) along axis 1 (respectively along axis 2) of the smoothed data are computed. The points of the following interval are kept (Fig. 3) :

$$\max \pm i * \sigma \quad (2)$$

with i the width of the filter.

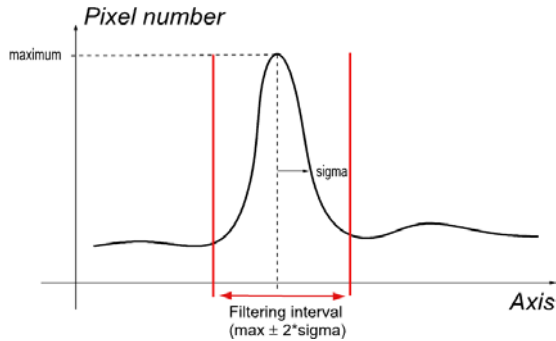


Fig. 3 Filtering interval of a spatial distribution

Fourthly, the filtered points are reprojected in the original frame.

III. ESTIMATION OF OBJECT ORIENTATION

In the classical shape-from-focus approach, the filtered planes are stacked up to achieve the 3D modelling of the object. But, in the case of orientation estimation that process is simplified, a 1D modelling of the object is performed. Each plane is replaced by the barycentre of the points, and finally the medium line of those points is calculated by a regression method.

Let $A(X_A, Y_A, Z_A)$ and $B(X_B, Y_B, Z_B)$ be respectively the first and last points of the segment of line, the elevation (El) and azimuth (Az) angles of that segment are computed as followed (Fig. 4) :

$$El = \arctan \left(\frac{z * FocusStep}{\sqrt{x^2 + y^2}} \right) \quad (3)$$

$$Az = \arccos \left(\frac{x}{\sqrt{x^2 + y^2}} \right) \quad (4)$$

with :

- $x = X_B - X_A$
- $y = Y_B - Y_A$
- $z = Z_B - Z_A$

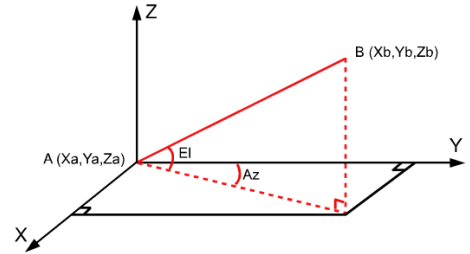


Fig.4 Representation of Azimuth and Elevation angles of a line

The medium line represents the 1D modelling of the object considered and the elevation and azimuth angles represent the orientation of the object in the reference frame. Actually the videomicroscope is supposed positioned perpendicularly to the microassembly scene and then XY plane is parallel to the image plane.

IV. EXPERIMENTS

Above concepts have been applied to an active two-finger microgripper imaged by a videomicroscope. The whole set-up (Fig.5) including a robotic system, a gripping system (the gripper) and an imaging system (the videomicroscope), is positioned in a clean-room enabling the control of the temperature and humidity. That set-up allows the robotic assembly

of micrometric parts ($40\ \mu\text{m} \times 40\ \mu\text{m} \times 10\ \mu\text{m}$) to get 3D compound products. Those tasks require the tracking of the fingers in order to control the gripper (for the aligning of fingers and their closing) and the robot (for the displacement of the part). Because of the small size of the part, the non-aligning of the fingers cause most of the time the failure of gripping.

Fig.5 Experimental micromanipulation station

A. Robotic System

The robotic system includes actuators, sensors and control units mechanically grouped to ensure teleoperated and partially-automated manipulation. The motions of the micromanipulation substrate, which supports the microparts, are obtained by two linear stages (xy) and one rotating stage (θ). The active gripper is mounted on an arm actuated by a linear stage for the vertical motion (z) and two rotating stages for pitch (β) and roll (γ) rotations.

B. Imaging System

In micromanipulation, micro objects are viewed through at least one videomicroscope which is a combination of a high resolution objective with a camera. The lens focuses photons onto the image sensor of the camera. The system used in the experiments includes an inversed microscope LEICA DM IRBE, the magnification is $10\times$ leading to a depth-of-field of about $7\ \mu\text{m}$. The CDD sensor is a LEICA DFC 320 with a frame rate of about 11 fps for a resolution of 1044×772 pixels, that provides a field-of-view of about $968 \times 716\ \mu\text{m}^2$.

C. Gripping System

The concept is tested on a gripper in order to get its two fingers orientations. A special manufacturing process was defined and performed, enabling the achievement of textured fingers compatible with the shape-from-focus. The silicon finger-tips glued on the piezoelectric fingers are made in SOI (Silicon On Insulator) wafers. Two silicon layers of 12 and $400\ \mu\text{m}$ thickness are separated by a thin buried oxide layer of $1\ \mu\text{m}$ thickness. The tips are etched by DRIE (Deep Reaction Ion Etching) process and then the oxide layer remains on their back side. An incomplete RIE (Reaction Ion Etching) process is used to etch 500 nm of this oxide layer. After these chemical and mechanical attacks, this surface is enough textured to ensure a clear

delimitation of focused areas.

The size of each tip is $70\ \mu\text{m}$ width and $1500\ \mu\text{m}$ length (Fig.5).

D. A sequence of images of the gripper

The videomicroscope observes the gripper through a transparent sample-holder fixed on the $xy\theta$ table. A multifocus image sequence is acquired with a step of $10\ \mu\text{m}$ (which is higher than the depth-of-field) of the arm (Fig.6). Each tip of the gripper is considered separately that gives two sequences of points.

Fig.6 The 2th, 6th, 10th and 17th images of the multifocus set with a $10\ \mu\text{m}$ step. The viewed length is $130\ \mu\text{m}$

E. Filtering

Fig.7 shows an example of points for one plane before filtering. A wide area of right detections (inliers) and some false detections (outliers) are present. The filtering enables to keep only inliers.

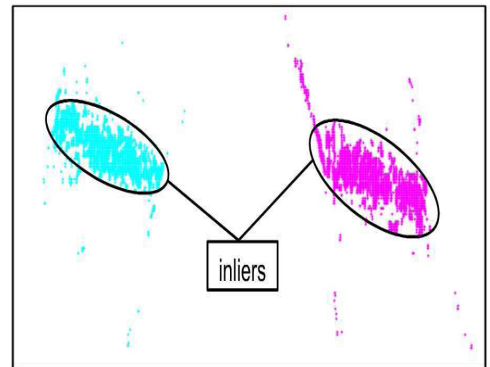


Fig.7 A set of detected points before filtering

Fig.8 shows the factorial axes of a set of points.

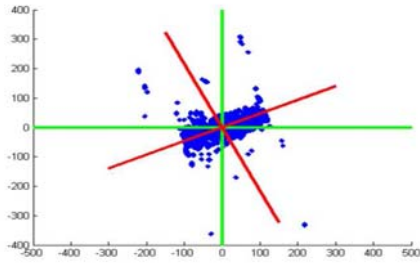


Fig.8 The factorial axes of the data

Fig.9 shows the spatial distribution of the points represented in Fig.8 :

- the spatial distribution along x axis (A) and the same distribution along the first factorial axis (B),
- the spatial distribution along y axis (C) and the same distribution along the second factorial axis (D),

As expected the data are better distributed and the curves less noisy in the factorial frame.

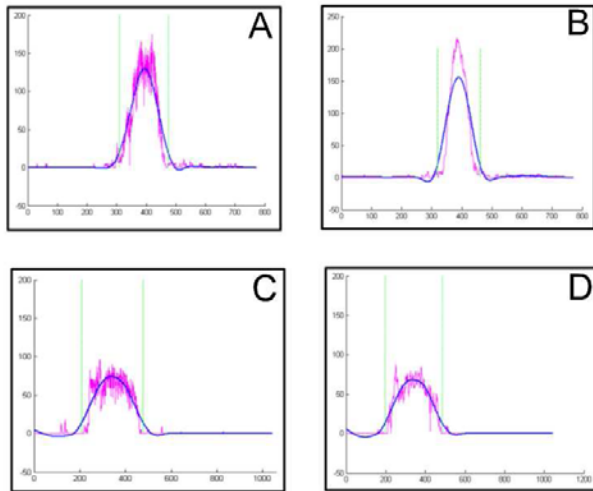


Fig.9 Spatial distribution of points

The result of the overall filtering is shown in Fig.10 indicating the relevance of the adopted approach.

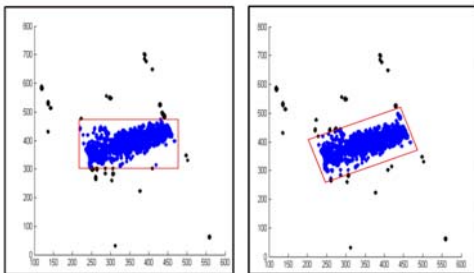


Fig.10 The result of the filtering onto the gripper data. The points outside the rectangles are taken off. The data are in the normal frame (left) and in the factorial frame (right).

F. Model and orientation of the gripper

Fig.11 shows the 3D reconstruction of the surface of the tips of the silicon gripper using the previously described algorithm with a 5×5 patch for the variance. It looks like a staircase as expected, the size of each stair corresponds to the focus step.

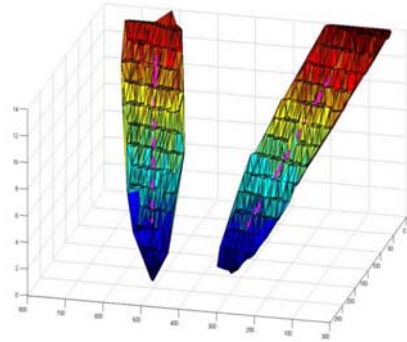


Fig.11 3D model of the gripper tips. The viewed length is $130 \mu m$.

Fig.12 shows the lines (1D models) modelling the tips and the corresponding orientations.

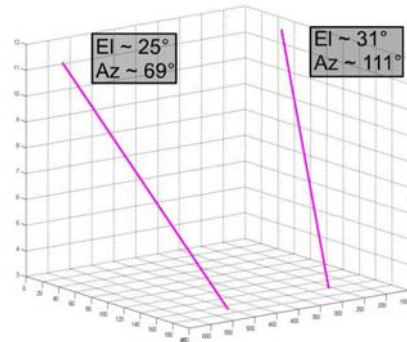


Fig.12 1D model of the gripper tips. The viewed length is $130 \mu m$.

V. CONCLUSION

The paper investigates the problem of localizing a rigid body in the images of a videomicroscope. More precisely the orientation of an active gripper in a microassembly scene is achieved. (The solution developed is based on the shape-from-focus approach. Instead of constructing the 3D model from the data, the 1D model is achieved.) The stages involve the calibrated acquisition of some stack images of the scene, the computation of the focused points by the variance based focus measurement, filtering of the data (binary mask, smoothing, spatial distribution filtering).

The concept is applied to an active microgripper (the silicon tip size is $70 \mu m \times 1500 \mu m$) imaged by a $10 \times$

videomicroscope positioned vertically to the scene in that way the image plane is practically parallel to the scene. The accurate vertical motions of the gripper are obtained by a robotic arm. Relevant modelling and orientation of the two tips have been obtained.

The future work will concern the deeply evaluation of the approach in terms of accuracy and robustness. The real-time aspect also will be considered, by limiting for example the number of stack images. Later, the approach will be used to perform the visual control of the gripper as well as the robot, to achieve the assembly of 3D compound products. The size of every part is $40\ \mu\text{m} \times 40\ \mu\text{m} \times 10\ \mu\text{m}$. In those tasks because of the smallest of the part, a non-aligning of the gripper tips can cause the failure of the gripping. The real-time estimation of the tips orientation enables to align them.

ACKNOWLEDGEMENTS

This work is conducted with financial support from the projet PRONOMIA (ANR n5-BLAN-0325-01) funded by ANR

- [1] S. Benhimane and E. Malis, *Homography-based 2d visual tracking and servoing*, The International Journal of Robotics Research **Vol. 26, No. 7** (2007), pp. 661–676.
- [2] A. Blake, *Visual tracking : a very short research roadmap*, Electronics Letters, Sensors and Actuators A **Vol. 42 (5)** (2006), 254–256.
- [3] J. Brufau, M. Puig-Vidal, J. Lopez-Sanchez, J. Samitier, W. Driesen, J6M. Breguet, J. Gao, T. Velten, J. Seyfried, R. Estana, and H. Woern, *Micron : Small autonomous robot for cell manipulation applications*, Proceedings of ICRA 2005, Barcelona, Spain, 2005.
- [4] Fabrizio Caccavale, Vincenzo Lippiello, Bruno Siciliano, and Luigi Villani, *Real-time tracking of 3d objects*, ch. Advances in control of articulated and mobile robots, pp. 125–151, Springer, 2004.
- [5] Andrew I. Comport, Member, Eric Marchand, Muriel Pressigout, and François Chaumette, *Real-time markerless tracking for augmented reality: The virtual visual servoing framework*, IEEE Transactions on Visualization and Computer Graphics **Vol. 12 (4)** (2006), 615–628.
- [6] Tom Drummond and Roberto Cipolla, *Real-time visual tracking of complex structures*, IEEE Transactions on Pattern Analysis and Machine Intelligence **Vol. 24 (7)** (2002), 932–946.
- [7] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2006.
- [8] Cheng-Ming Huang, David Liu, and Li-Chen Fu, *Visual tracking in cluttered environments using the visual probabilistic data association filter*, IEEE Transactions on Robotics **Vol. 22 (6)** (2006), 1292–1297.
- [9] Jung H. Kim, Shih-Kang Kuo, , and Chia-Hsiang Menq, *An ultraprecision six-axis visual servo-control system*, IEEE Transactions on Robotics **Vol. 21 (5)** (2005), 985–993.
- [10] Vincenzo Lippiello, *Real-time visual tracking based on bsp-tree representations of object boundary*, Robotica **Vol. 23** (2005), 36575.
- [11] Ezio Malis and François Chaumette, *2 1/2 visual servoing with respect to unknown objects through a new estimation scheme of camera displacement*, International Journal of Computer Vision **Vol. 37, N 1** (2000), 79–97.
- [12] Eric Marchand and François Chaumette, *Feature tracking for visual servoing purposes*, Robotics and Autonomous Systems **Vol. 52 (1)** (2005).
- [13] Shree K. Nayar, *Shape from focus*, Tech. Report CMU-RI-TR-89-27, Department of Electrical and Computer Engineering The Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania 15213, November 1989.
- [14] Markus Niederost, Jana Niederost, and Jiri Scucka, *Automatic 3d reconstruction and visualization of microscopic objects from a monoscopic multifocus image sequence*, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV-5/W10, 2003.
- [15] Patrick Sandoz, Vincent Bonnans, and Tijani Gharbi, *High accuracy position and orientation measurement of extended two-dimensional surfaces by phase sensitive vision method*, Applied Optics **Vol. 41 (26)** (2002), 5503–5511.
- [16] Patrick Sandoz, Jean Christophe Ravassard, Soukalo Dembele, and Andre Janex, *Phase-sensitive vision technique for high accuracy position measurement of moving targets*, IEEE Transactions on instrumentation and measurement **Vol. 49 (4)** (2000), 867–872.
- [17] Jay Stavnitzyk and David Capson, *Multiple camera model-based 3d visual servo*, IEEE Transactions on Robotics and Automation **Vol. 16 (6)** (2000), 732–739.
- [18] Yu Sun, Stephan Duthaler, and Bradley J. Nelson, *Autofocusing algorithm selection in computer microscopy*, Proceedings of 2005 IEEE/RSJ IROS, 2005.
- [19] J. Wedekind, *Focus set based reconstruction of micro-objects*, International IEEE Conference on Mechatronics & Robotics (MechRob'04), Aachen, Germany, 13-15 September 2004, 2004.
- [20] Ge Yang, James A. Gaines, and Bradley J. Nelson, *Optomechatronic design of microassembly systems for manufacturing hybrid microsystems*, IEEE Transactions on Industrial Electronics **Vol. 52 (4)** (2005), 1013–1023.
- [21] Kemal B. Yesin and Bradley J. Nelson, *A cad model based tracking system for visually guided microassembly*, Robotica **23** (2005), 409–418.
- [22] Thierry Zamofing and Heinz Hugli, *Multiresolution reliability scheme for range image filtering*, SPIE5606, 2004.
- [23] Xiaoyu Zhang and Richard J. Mammone, *Range estimation from two-dimensional images using phase modulation*, Optical Engineering **Vol. 33 (12)** (1994), 4079–4092.
- [24] Shaohua Zhou, Rama Chellappa, and Baback Moghaddam, *Visual tracking and recognition using appearance-adaptive models in particle filters*, IEEE Transactions of Image Processing **Vol. 13 (11)** (2004).