TOWARD THE VISION BASED SUPERVISION OF MICROFACTORIES THROUGH IMAGES MOSAICING

J. Bert[1], S. Dembélé[1], N. Lefort-Piat[1]
[1] *Laboratoire d'Automatique de Besançon - UMR CNRS 6596 - ENSMM - UFC*
24 rue Alain Savary, 25000 Besançon, France

Abstract The microfactory paradigm means the miniaturisation of manufacturing systems according to the miniaturisation of products. Some benefits are the saving of material, energy and place. A vision based solution to the problem of supervision of microfactories is proposed. It consists in synthetising a high resolution global view of the work field and real time inlay of local image in this background. The result can be used for micromanipulation monitoring, assistance to the operator, alarms and others useful informations displaying.

Keywords: Supervision; microfactories; mosaicing.

1. INTRODUCTION

Since many years a lot of *microproducts* i.e. in the size range between $10\mu m$ and $10mm$ have been developed: micromechanical parts, microopticals parts, microsensors, microactuators, MEMS, MOEMS, ... For the automatic manufacturing of these elements and, their manipulation to get for example assembled products, the main approach have been the use of precision facilities where no constraint of size is considered. That has led to precision production systems which size is usually very large compared to *micrometric* size of parts, for example in semiconductor industry. Those factories take large space, spend a lot of energy and material. Besides that precision production approach, a micro production approach more known as the microfactory approach has been studied. Its purpose is the achievement of small manufacturing and manipulation systems, the philosophy is to match the size of the production facilities with the size of the products. The potential benefits are the saving of material, energy and place, agility and flexi-

bility. Some experimental microfactories are described in the following references: (Tanaka, 2001; Okazaki et al., 2004).

Because of the impossibility or difficulty to use conventional sensors in a microfactory (lack of resolution and place) vision systems are of great interest. They give images of the work field from which a lot of informations can be retrieved. An image can be used for detecting a micropart or an effector, measuring the position of a part, measuring the force applied on a part, measuring defects of a part, verifying the presence of a part in an assembled product, detecting events, displaying alarms, displaying effector position, ... Those informations cover the control, inspection and supervision functions.

The paper deals with the supervision of microfactories, which is still a few studied. The following points are developed next: the characteristic of the vision system in a microfactory, the scene synthesis by mosaicing, proposition and validation of the dynamic mosaicing solution.

2. CHARACTERIZATION OF THE VISION SYSTEM

Supervision is one of the many functions achieved in a production system specially a microproduction one (Breguet et al., 2000). The supervision of a microfactory is a few studied. (Tanaka, 2001) and (Okazaki et al., 2004) simply noticed the presence of three miniature cameras for displaying the image of each machine in the experimental microfactory of Japan AIST. (Kuronita et al., 2001) use a camera to monitor the activity of their swarm robot based drilling system. Actually supervision is more than monitoring i.e. displaying image of the microfactory, it also includes assistance to the operator, detecting events (lost of a microproduct, contact of the effector with the substrat, ...), displaying alarms and informations.

In the paper a vision based supervision paradigm is proposed. The usual vision system used in a microfactory is a camera with an optical (photonic) microscope. The resolution is high (up to $0.25\mu m$ according the law of Raleyght) but the depth of field and the field of view are low, the overall dimensions are important, the images are weakly contrasted (Vikramaditya and Nelson, 1997). Instead of the microscope based vision system, a microfiberscope and a camera also can be used. A fiberscope consists of a bundle of optical fibers for lighting and an other bundle for image transportation, a microoptical set allows the connection with the camera. The system is flexible and not cumbersome (end diameter reaches $0.5mm$). The resolution is less high, the depth of field is high, the field of view is low. That kind of system has been developed

by (Tohyama et al., 2000) for stereoscopic observation of the work field. The microcamera with the appropriate microoptical set (endoscope) also could be used.

The common characteristic of the above vision components is their low field of view, the corresponding image represents a local view. Then the vision system of the microfactory must includes a set of local vision components positioned at adequate places. Finally the vision system must be distributed and if possible modular. We propose the reconstruction of the global view of the microfactory by mosaicing the local images of the vision components.

3. OFF-LINE VIEW SYNTHESIS BY MOSAICING

The mosaicing consists in constructing a large image (the mosaic image) from a set of small images. It virtually increases the field of view of vision systems without lost of resolution and with minimum deformation. It has a lot of applications :

- 360° panoramic image achievement that can be view with the Quicktime®VR software (virtual camera) (Chen, 1995),
- video compression (Irani et al., 1996),
- increasing the field of view (Heckbert, 1989; Kumar et al., 1994; Szeliski, 1994; Potsaid et al., 2003),
- digitization of large printed documents (Zappalá et al., 1997; Pilu and Isgro, 2002; Kumar et al., 2004).

Static mosaicing includes two stages.

- In the first stage the registration of the images is performed, they are aligned in the same reference according to their transformation (camera motion) with this reference (Figure 1).
- The second stage is the blending stage. After been registred the images are fused to form the mosaic image.

There are three approaches of mosaicing according to the method used to recovery the camera motion: the calibrated motion approach, the intensity based approach (named also direct method), the feature based approach. Below, we present these approaches by considering only two images.

3.1 Calibrated motion approach

When the camera motion is perfectly know i.e. the transformation between the images is known, the registration is immediately achieved.

If the motion is a translation or a small rotation the problem is easy, no registration is achieved, the images are strips that are directly aligned. This approach was used by (Rousso et al., 1997; Peleg and Herman, 1997), (Blanc et al., 2001) in satellite image mosaicing and (Potsaid et al., 2003) for increasing work field in biological micromanipulation application. The advantage of this approach is the fact it can be used even if the scene contains no texture.

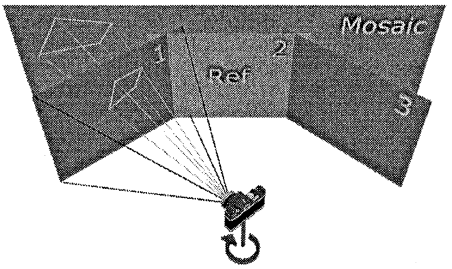

Figure 1. Illustration of images registration: image 1 and 3 are aligned with image 2 (mosaic reference).

3.2 Intensity based approach

The direct approach allows the registration of images with an iterative intensity correlation method. In fact the correlation allows the estimation of the camera motion between the two images. A lot of methods have been developed. (Barnea and Silverman, 1972) used a spatial correlation named L1 Norm. Next, (Kuglin and Hines, 1975) used a phase correlation by FFT that allows the estimation of the translation between two images using the properties of Fourier space. Szeliski and Shum introduced a warp correlation between two images that leads to the colineation matrix (homography matrix in image plan) (Szeliski, 1994; Szeliski and Shum, 1997; Shum and Szeliski, 1997). This colineation matrix integrates translation and rotation of the camera and correspond to a full planar projective motion model (camera pan/tilt):

$$p' \sim Gp \Leftrightarrow \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim \begin{bmatrix} g_0 & g_1 & g_2 \\ g_3 & g_4 & g_5 \\ g_6 & g_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

where $p = (x, y, 1)^T$ and $p' = (x', y', 1)^T$ are respectively pixels in images 1 and 2 represented in homogeneous coordinates (projective coordinates), and $\sim$ indicates equality up to a scale. In translation motion,

only the two parameters g_2 and g_5 are used. To compute the 8 parameters G matrix, an iterative method is used. G is first initialized and then updated according to the following expression $G \leftarrow (I + D)G$ where I and D are respectively the identity and incremental matrixes. Image I_2 is iteratively warped by G until the error with image I_1 is less than a defined value, then the problem becomes the minimization of the error $E(d)$ between I_1 and the warped of I_2 ($\tilde{I}_2$). For that, the error is approximated by the first order Taylor series:

$$E(d) = \sum_i \left[\tilde{I}_2(p_i) - I_1(p_i)\right]^2 \approx \sum_i \left[g_i^T J_i^T d + e_i\right]^2 \qquad (2)$$

where $e_i = \tilde{I}_2(p_i) - I_1(p_i)$ is the intensity error, $g_i^T = \nabla\tilde{I}_2(p_i)$ is the image gradient of $\tilde{I}_2$ at p_i, $d = (d_0, ..., d_8)$ is the incremental motion vector parameter, and $J_i = J_d(p_i) = \frac{\partial p_i}{\partial d}$ is the Jacobian of the resampled point coordinate p_i with respect to d. This least-squared problem, (eq 2), has a simple solution through the *normal equations*:

$$Ad = -b, \quad A = \sum_i J_i g_i g_i^T J_i^T \quad b = \sum_i e_i J_i g_i \qquad (3)$$

A is the *Hessian*, and B is the *accumulated gradient* or *residual*. These equations can be solved using a symmetric positive definite (SPD) solver such as *Cholesky* decomposition. d is solved and G is updated to warp $\tilde{I}_2$ and so on. That 8 parameters projective transformation recovery algorithm works well if initial estimates are close enough to final results. Its contains more free parameters than necessary, it suffers from slow convergence and sometimes gets stuck in local minima. For these reasons, the 3 rotational parameters model is usually preferred. For long images sequences, this approach also suffers from the problem of accumulated misregistration errors. The latter are reduced using a global alignment method next (Szeliski and Shum, 1997; Shum and Szeliski, 1997; Shum and Szeliski, 1998).

3.3 Feature based approach

An alternative solution to estimating the transformation between images by intensity correlation as explained above, consists in the use of invariant feature points. These are points where the intensity changes like corners. The motion estimation follows four stages:

1 find the *interest* points with a corners detector,

2 match points of image 1 with points of image 2,

3 remove outliers i.e. the false matchings,

4 estimate G matrix with at least four pairs of matched points.

The first corners detector algorithm was published by (Moravec, 1977). Today, there are several corners detectors in the literature, but only two are most popular, Susan by (Smith and Brady, 1995) and Harris by (Harris and Stephens, 1988). (Schmid et al., 1998) shows that the Harris detector is the most robust according to illumination changes. This is why, generally Harris detector is often used for features detection. It is based on auto-correlation function since the latter put in light the intensity changes. Actually a bilinear approximation of auto-correlation is used because small shifts are considered. Suppose $[u, v]$ be the shift, the approximation M can be written:

$$M = \sum_{x,y} W(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{4}$$

with $W(x,y)$ a window function (rectangular or Gaussian), I intensity of image, I_x and I_y respectively the gradient along the axes X and Y. The result of the corner detector is a set of points in each image (figure 2).

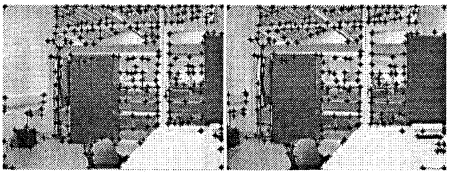

Figure 2. Feature points detected in the images with Harris corner detector. The motion between the two images is a rotation around the camera centre (pan).

The next stage consists in matching the two sets of points. For each point of image 1 a correlation window is defined centered at that point. The latter is used to perform the correlation between that window and the same size region centered at each feature point of image 2. Sum of Squared Difference (SSD) (Smith et al., 1998) correlation is usually used:

$$SSD(p_1,p_2) = \frac{1}{W} \sum_{k=-K}^{K} \sum_{n=-N}^{N} [I_1(x_1+k, y_1+n) - I_2(x_2+k, y_2+n)]^2 \tag{5}$$

If that result is greater than a defined threshold the points are supposed matched each other. The SSD gives some erroneous matchings, then in stage 3 it is necessary to remove the bad matchings. The RANSAC algorithm (Random Sample Consensus) is often use for that purpose and for motion (G) estimation. It is an algorithm for robust models fitting, first introduced by (Fischler and Bolles, 1981). It is robust in the sense it has good tolerance to outliers in the experimental

data. It is capable of interpreting and smoothing data containing a significant percentage of error. The estimation is only correct with a certain probability, since RANSAC is a randomised estimator. The algorithm has been applied to a wide range of model parameters estimation problems in computer vision, such as registration or detection of geometric primitives.

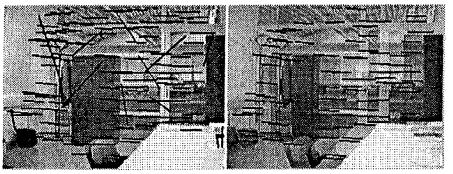

Figure 3. Before applied RANSAC (left), after applied RANSAC (right). Every line indicates the matching between two points

3.4 Blending

A set of images represented in the same reference are obtained after the registration. These images usually contains some overlapping zones i.e. common zones and the problem is to attribute values to the pixels of these zones in order to smooth the transition between the images (i.e. to make invisible the seams). The blending solves this problem. A weighted averaging is usually used: every pixel is weighted according the k^{th} power of its distance to the image boundary (hat filter):

$$f_{res}(p) = \frac{\sum_{i=1}^{N} f_i(p) d_i^k}{\sum_{i=1}^{N} d_i^k} \tag{6}$$

with d the weighting coefficient. After this stage we obtain the image mosaic.

3.5 Comparison

The calibrated motion approach suppose the precise knowledge of the camera motion between the different successive images. It works even the images are not textured, that is not the case for the two others approaches. In fact, they use correlation methods (for motion estimation) that requires textured images. The success of the intensity and feature based approaches is not guaranteed because respectively of problems of convergence and local minima dead end, and of the feature detection algorithm weakness. Finally the choice of a mosaicing approach is application dependent.

4. RESULT WITH THE BENCHMARK

A small local view of the work field is not sufficient to allow the performing of a micromanipulation or microassembly, a global view is required and this is true even if the vision system is distributed or not. So we propose to use mosaicing for syntheting the global view. It can be noticed that (Potsaid et al., 2003) mosaics optical microscope images to get a view of the scene for biological observation and manipulation. Our solution is quite similar to that of (Kourogi et al., 1999), it consists in an off-line construction of the global image by mosaicing and an on-line inlay of dynamical local images in this background. The following benchmark is used to valid that solution: a microendoscope based vision system (the camera is a cylinder of length $20mm$ and diameter $5mm$, the angle of view is 90°, the CCD sensor resolution is 768x576 pixels, a 8 bits frame grabber), an xyz stage (resolution $2.5\mu m$, travel $55mm$, a stand alone control system) (figure 4). The maximal resolution (for the minimal work distance of about $35mm$) is $50\mu m$/pixels. The small products manually manipulated with a brussel gripper are components of watch, an axis, a gear, a support. Matlab, C++ Builder with OpenCV library environments are used to program the application.

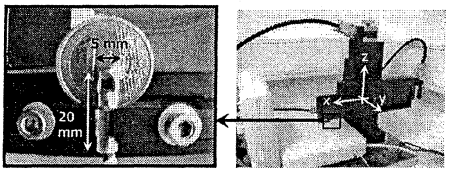

Figure 4. Micro camera compared with a coin of 1 euro (left), xyz stage (right).

4.1 Off-line background construction

Usually, in micromanipulation and microassembly the scenes (work fields) are rarely textured and precise positioning stage are used. So we use the calibrated motion approach.

The positioning stage is equiped with the microcamera and this system is used to scan the work field in order to obtain the set of local images. Calibrated translations are performed. The motion step d_m (metric) corresponds to the image width d_p (pixels), it can be written $d_m = d_p \times S_p$. S_p is the pixel size and depends on the work distance (scene-microcamera) Z_m. We perform partial calibrations of the system by analyzing the image of a $2000\mu m$ diameter circle based template at different work distance. The image of the template is acquired, the diameter of the circle is computed with a resolution of 1/20 pixel using interpolation with B-Spline functions, and the pixel size S_p is computed. By

applying a median least squared method, a robust optimisation method, we find the following function:

$$S_p = 1,277.10^{-3} \times Z_m + 90,532 \tag{7}$$

The images delivered by the vision system are very distorted because of the lens of the microcamera (angle of view of 90°). In order to minimize the distorsion in the final mosaic image we crop a small strip I_c around the centre of each local image (distorsion is always smaller around the image centre) then the actual motion step is $d_m = d_{pc} \times S_p$ where d_{pc} is the width of I_c.
The result of the method is presented figure 6 at the minimum work distance, the image size is 1100 × 1100 pixels for a pixel size of $50\mu m$

4.2 On-line local image inlay

Now, we have to perform the real time inlay of the local view in the above global view. That local view define a dynamic zone in the static background. We do not perform a simple overlay which stays visible the seam between images. We perform the fusion of the two images by a symetric fade mask which part is defined by the following equations (8):

$$\begin{array}{cc} f_1(x,y) = 0 & f_2(x,y) = \frac{1}{\beta-\alpha}x + \frac{\alpha}{\alpha-\beta} \\ f_3(x,y) = \frac{(x-\alpha)(y-\alpha)}{(\beta-\alpha)^2} & f_4(x,y) = 1 \end{array} \tag{8}$$

α and β are respectively start and end slopes, (x,y) is the pixel position. Figure 5 shows the 2D and 3D forms of the mask.

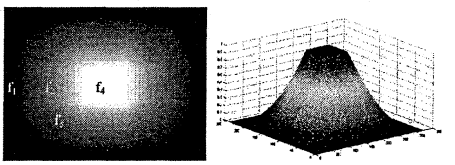

Figure 5. The fade mask in 2D representation with the functions according to equation 8 (left), the fade mask in 3D representation (right)

The local image I_l is multiplied by the mask and the background I_b image is multiplied by the mask complemented:

$$I_{ic}(x,y) = I_l(x,y)f(x,y) + I_b(x_{ic},y_{ic})\,[1 - f(x,y)] \tag{9}$$

I_{ic} represents the final dynamic image, (x_{ic}, y_{ic}) is the centre of the dynamic zone in the final image.

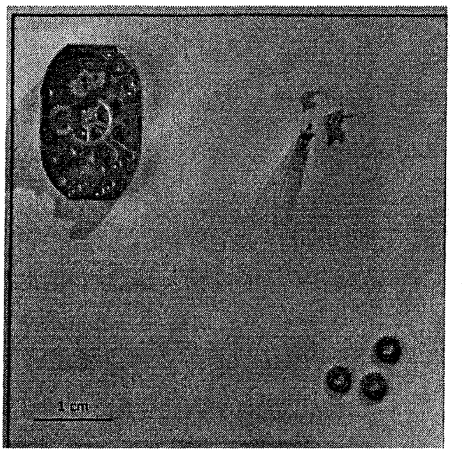

Figure 6. Dynamic mosaic of the benchmark: 1100x1100 pixels, 50μm/pixels, 55mm x 55mm.

Figure 6 shows an image of the final video. In the background we find a watch support (upper left) and gears (lower right). Upper right of that image we can see a gripper manipulating an axis. The user has a global view of the work field with high resolution, he sees the assembly target, products stocks, he also sees the local view showing the manipulation in progress. In addition, the fade mask prevents the gripper to hide the work field, it is visible only in dynamic zone which is updated as often as possible. We obtain a frame rate of 10 Hz, but our code is not optimize and we can easily increase the speed of the process.

5. CONCLUSION

We analyze the vision components used and usable in microfactories. Their main property is the low size of the field of view because of the requirement of high resolution and lower distortion. These local views of the work field are not sufficient to perform the tasks, a global view is required. A solution to perform the latter is static image mosaicing, so we have summarized the three approaches of mosaicing, calibrated motion, intensity based and feature based, and pointed out their advantages and disadvantages. According to the characteristics of micromanipulation (precise positioning stage, non textured work field) we select the calibrated motion approach to reconstruct the global image of the work field with high resolution. That background is dynamically updated by inlay live local images. This dynamic mosaicing was validated with a

benchmark including a xyz positioning stage, a microcamera, a non textured work field containing watch components. The proposed dynamic mosaicing defines a step toward the vision based supervision of microfactories. The mosaic gives visual feedback and could be used to assist the operator, to display alarms and informations about the tasks being performed. The solution is also valid for distributed vision systems and can be combined with visual servoing for the tracking of mobile target every where in the mosaic.

References

Barnea, E. I. and Silverman, H. F. (1972). A class of algorithms for fast digital image registration. In *IEEE Transactions on Computers*, volume C-21, pages 179–186.

Blanc, Philippe, Savaria, Eric, and Oudyi, Farid (2001). Le mosaquage d'images satellitales optiques a haute resolution spatiale. In *18e colloque sur le traitement du signal et des images (GRETSI'01)*, volume 2, pages 251–254, Toulouse, France.

Breguet, Jean-Marc, Schmitt, Carl, and Clavel, Reymond (2000). Micro/nanofactory: Concept and state of the art. In *SPIE Proceedings of the Microrobotics and Microassembly II*, volume 4194, pages 1–12.

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

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

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

Heckbert, P. (1989). Fundamentals of texture mapping and image warping. Master's thesis, The University of California at Berkeley, USA.

Irani, M., Anandan, P., Bergen, J., Kumar, R., and Hsu, S. (1996). Mosaic representations of video sequences and their applications. In *Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application*, volume 8(4).

Kourogi, Masakatsu, Kuratay, Takeshi, Hoshinoz, Jun'ichi, and Muraoka, Yoichi (1999). Real-time image mosaicing from a video sequence. In *IEEE International Conference on Image Processing*.

Kuglin, C. D. and Hines, D. C. (1975). The phase correlation image alignement method. In *Proceeding IEE International Conference on Cybernetic Society*, pages 163–165, New York, USA.

Kumar, G. Hemantha, Shivakumara, P., Guru, D. S., and Nagabhushan, P. (2004). Document image mosaicing: A novel approach. In *SADHANA - Academy Proceedings in Engineering Sciences*, volume 29(3), pages 329–341, India.

Kumar, R., Anandan, P., and Hanna, K. (1994). Shape recovery from multiple views : a parallax based approach. In Publishers, Morgan Kaufmann, editor, *In Image Understanding Workshop*, pages 947–955, Monterey, CA.

Kuronita, Tokuji, Tadokoro, Shigeru, and Aoyama, Hisayuki (2001). Swarm control for automatic drilling operation by multiple micro robots. In *Australian Conference on Robotics & Automation*, pages 7–12, Sydney.

Moravec, H. P. (1977). Towards automatic visual obstacle avoidance. In *Proceeding of the 5th International Joint Conference on Artificial Intelligent*, page 584.

Okazaki, Yuichi, Mishima, Nozomu, and Ashida, Kiwamu (2004). Microfactory - concept, history, and developments. *Journal of Manufacturing Science and Engineering*, 126:837–844.

Peleg, S. and Herman, J. (1997). Panoramic mosaics by manifold projection. In *IEEE Computer Vision and Pattern Recognition*, pages 338–343.

Pilu, M. and Isgro, F. (2002). A fast and reliable planar registration method with applications to document stitching. In *Proceeding of the British Machine Vision Conference*, pages 688–697, Cardiff.

Potsaid, Benjamin, Bellouard, Yves, and Wen, John T. (2003). Scanning optical mosaic scope for dynamic biological observations and manipulation. In *IEEE Intelligent Robotics Systems (IROS)*, Las Vegas, NV, USA.

Rousso, B., Peleg, S., and Finci, I. (1997). Mosaicing with generalized strips. In *DARPA Image Understanding Workshop*, pages 255–260, New Orleans, Louisiana, USA.

Schmid, C., Mohr, R., and Bauckhage, C. (1998). Comparing and evaluating interest points. In *IEEE International Conference on Computer Vision*, pages 230–235.

Shum, Heung-Yeung and Szeliski, Richard (1997). Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research.

Shum, Heung-Yeung and Szeliski, Richard (1998). Construction and refinement of panoramic mosaics with global and local alignment. In *IEEE International Conference on Computer Vision*, pages 953–956, Bombay , India.

Smith, P., Sinclaif, D., Cipolla, R., and Wood, K. (1998). Effective corner matching. In *Proceeding of the British Machine Vision Conference*, volume 2, pages 545–556.

Smith, S.M. and Brady, J.M. (1995). Susan - a new approach to low level image processing. Internal Technical Report TR95SMS1c, Defence Research Agency, Chobham Lane, Chertsey, Surrey, UK.

Szeliski, Richard (1994). Image mosaicing for tele-reality applications. Technical Report CRL 94/2, Digital Equipment Corporation, Cambridge Research Lab.

Szeliski, Richard and Shum, Heung-Yeung (1997). Creating full view panoramic image mosaics and environment maps. In *Computer Graphics (SIGGRAPH'97 Proceedings)*, volume 31, pages 251–258.

Tanaka, Makoto (2001). Development of desktop machining microfactory. In *RIKEN Review*, volume 34, pages 46–49.

Tohyama, Osamu, Maeda, Shigeo, Abe, Kazuhiro, and Murayama, Manabu (2000). Fiber-optic sensors and actuators for environmental recognition devices. *IEEE Trans. Electron.*, E83-C(3):475–480.

Vikramaditya, Barmeshwar and Nelson, Bradley J. (1997). Visually guided microassembly using optical microscopes and active vision techniques. In *IEEE International Conference on Robotics and Automation*, pages 3172–3177, Albuquerque, New Mexico.

Zappalá, Anthony, Gee, Andrew, and Taylor, Michael (1997). Document mosaicing. In *Proceeding of the British Machine Vision Conference*, volume 17, pages 589–595.