Synthesizing a virtual imager with a large field of view and a high resolution for micromanipulation.

Julien Bert, Sounkalo Dembélé, Nadine Lefort-Piat

Synthesizing a Virtual Imager with a Large Field of View and a High Resolution for Micromanipulation

Julien Bert, Sounkalo Dembélé and Nadine Lefort-Piat
Laboratoire d'Automatique de Besançon
UMR CNRS 6596 - ENSMM - UFC
25000 Besançon, France
{jbert, sdembele, npiat}@ens2m.fr

***Abstract*— Photon microscope connected with a camera is the usual imager required in micromanipulation applications. That microimager gives high resolution views, but the corresponding field of view are very narrow and do not allow the vision of the entire workfield. The classical solution consists in using multiple views imaging system: a high resolution imager for local view and a low resolution imager for global view. We are developing an alternative solution based on image mosaicing that requires only one microimager. The views from that real microimager are associated in order to achieve a virtual microimager which combines a large field of view with a high resolution.**

I. Introduction

Micromanipulation is the manipulation of parts at the microscale, i.e. in the range from 1 μm to 1 mm, for assembly, sorting or testing. In addition to biomicroparts like cells and pollen seeds, artificial microparts are chemically or mechanically synthetized, or micromachined. Classical examples of the first and second types are respectively grains of powder like drugs or cosmetics, and optomechatronic components like balls, pegs, pins, threads, membranes, lenses, shutters and fibres. In some cases these microparts define final products (MEMS), otherwise they must be assembly to lead to the final products. For that purpose some automated microassembly systems have been developed by [1], [2], [3] and [4]. From those results it can be noticed that a microimaging system is always required, and the most used is the photon microscope connected to a camera. The images and their processing and analysis allow task surveillance, system control or microparts recognition. That microimager gives high resolution views (up to $0.25\mu m$ according to the law of Rayleight), but the corresponding field of view is very narrow and do not allow the vision of the entire work field. To overcome that problem, multiple view imaging systems are used: a microimager for the high resolution and a basic imager for the large field of view [4]. The drawbacks of that solution are its expensiveness and the fact the work field is obstructed.

Image mosaicing is an image based rendering technique that consists in constructing a mosaic image of a scene from set of small overlapping views of that scene. Each small view represents a portion of the scene. It virtually increases the field of view of imaging systems without a lost of resolution and with a minimum of deformation. Some applications are panoramic image synthesis [5], [6], [7], [8], video compression [9], image stabilization [10] and large documents digitization [11], [12], [13], [14]. The solution for the construction of the high resolution and large filed of view imager developed in this paper is based on image mosaicing. A mosaic representing the background of the work field is first constructed off-line from the high resolution real views of an imager, and secondly real time views of that imager are inlaid in-line in this background. The image mosaicing is mentioned in section 2. Section 3 is presents the construction of the background by improving usual stages of mosaicing. The inlaid of real time views are exposed in section 4. Section 5 exposes the results for an microimaging system based on photon microscope.

II. Image Mosaicing

The image based rendering technique of mosaicing consists of a registration stage followed by a blending stage. In the first stage the input images are aligned in the same reference according to their transformation with this reference. One image is selected as the reference, the problem is to find out the transformation between the latter and every image of the input set of images. That transformation corresponds to the motion of the imager between the views. After been registered the images are assembled in a unique view called the mosaic image. Overlapping zones of the input views are mixed in order to make then indistinguishable. According to the method used to recovery the motion of the image, mosaicing techniques are classified into calibrated motion, intensity or feature based approaches.

In calibrated motion approach the motion of the imager is supposed to be known at the beginning or is a translation or a small rotation, then the registration is immediately performed [15], [16], [17], [7]. In intensity based approach the transformation between two images is recovered by an iterative algorithm with all the pixels of the overlapped zones [18]. In feature based approach the transformation is estimated by algorithms with only some pairs of matched feature points in both views [19]. In this paper we do not use calibrated motion approach since we do not know the motion of the imager. We do not use the intensity based approach because of it possible convergence toward a local minimum i.e. the transformation recovery is impossible. We use the Direct Linear Transformation (DLT) [20] algorithm based on matched features in both images.

III. Synthesis of the Background image

The background image of the imager corresponds to a mosaic image. A feature based approach using the DLT algorithm is used. Usual stages are improved in order make them more accurate and robust.

A. Registration

We will explain the registration process by considering only two input images I and I' where I is considered as the reference one. The process is the same for I' and I'', I'' and I''', ... until the last image is reached. The motion of the camera (translation and/or rotation) between I and I' induced a projective transformation between both images. The latter is expressed through a collineation matrix, G. The registration means the rectification of I' according to the reference I i.e. the warping of I' by the inverse of G (Fig. 1). Our feature based approach consists in detecting feature points in both images, their matching, estimation of G by DLT, warping of I'.

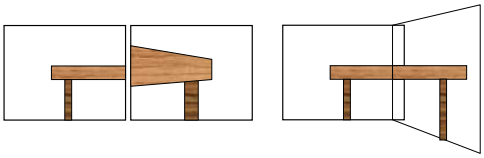

Fig. 1. Example of registration: left, the couple of original images (a table), right, the couple of images with the second rectified by the projective transformation with the reference.

1) Features points detection: In order to find the interest points (features), we use the popular Harris detector [21]. Harris detector is a corner detector which is based on an auto-correlation function since the latter puts in light the intensity changes in an image. By considering a small shift the auto-correlation can be approximated as followed:

$$M = \sum_x \sum_y W(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{1}$$

with I the image, I_x and I_y the derivatives of I along x and y respectively. For every pixel of the image the bilinear approximation (1) is computed and the detector response R can be expressed as followed:

$$R = \det M - k\,(trace\ M)^2 \tag{2}$$

where, $\det M = \lambda_1 \lambda_2$ and $trace\ M = \lambda_1 + \lambda_2$ with λ_1 and λ_2 the eigenvalues of M. The value of k is defined empirically 0.04×10^{-6} in our experiments. According to the detector response and the eigenvalues, it is possible to determine if the point is a corner, an edge or a flat.

We will determine the feature points by the fact the detector response is superior to a specified threshold.

2) Feature Points Matching: This stage consists in matching the two sets of points found by Harris detector, one set for each image. The correspondence begins by defining the correlation window around the feature found in I and the same around every feature of I'. The Zero-mean Normalized Sum of Squared Differences (ZNSSD) criterion, is used to estimate the likelihood:

$$c_{x,y} = -\frac{\sum_{i,j}\left[\left(I(x+i,y+j)-\bar{I}\right)-\left(I'(x'+i,y'+j)-\bar{I'}\right)\right]^2}{\sqrt{\sum_{i,j}\left[I(x+i,y+j)-\bar{I}\right]^2}\sqrt{\sum_{i,j}\left[I'(x'+i,y'+j)-\bar{I'}\right]^2}} \tag{3}$$

where $\bar{I}$ and $\bar{I'}$ are respectively the mean of I and I'. The minimum value of the criterion corresponds to the maximum likelihood between both features. The corresponding features are then matched. Noise and the illumination changes can lead to bad matchings.

3) Removal of outliers and DLT: Robust estimation of the collineation matrix requires to remove bad matchings. The RANSAC algorithm (RANdom SAmple Consensus) is often use for that purpose and G estimation. First introduced by [22] it randomly selects four couples of points (p, p') with p the feature of I and p' the feature of I'. p and p' are linked by:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim \begin{bmatrix} g_1^1 & g_2^1 & g_3^1 \\ g_1^2 & g_2^2 & g_3^2 \\ g_1^3 & g_2^3 & g_3^3 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

where $p = (x, y, 1)^T$, $p' = (x', y', 1)^T$) and $\sim$ equality up to a scale. It can written:

$$\begin{bmatrix} p^{iT} & 0 & -x' p^{iT} \\ 0 & p^{iT} & -y' p^{iT} \end{bmatrix} \begin{bmatrix} g_i^{1T} \\ g_i^{2T} \\ g_i^{3T} \end{bmatrix} \sim 0 \tag{5}$$

That equation is of the form:

$$Ag = 0 \tag{6}$$

For each correspondence (p, p') the DLT computes the matrix A and performs a Singular Value Decomposition (SVD) of it:

$$\mathrm{A} = U \Sigma V^T \tag{7}$$

The collineation matrix G^* for the corresponding subset of points is determined by the last column of V. G^* is used to compute the correspondence p^* of every pixel p of the input subset:

$$p^* = G^* p \tag{8}$$

Then the euclidian distance between the four pixel p' and p^* is computed, if it is superior to a predefined threshold the corresponding pixel p' is removed from the input set. Then RANSAC performs the filtering of the data with several randomly subsets. As soon as the filtering is over DLT is used to recovery the collineation matrix G [20].

4) Image Warping: The matrix G is used to warp the image I'. In order to avoid holes in the result image a backward method is used. It consists in defining an empty image (the color value of every pixel is set to zero). G and bicubic interpolation are used to determine the position of every pixel of the empty image in I', the value of that pixel becomes the value of the pixel in the empty view.

B. Blending

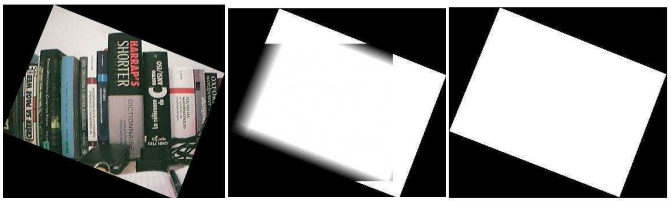

Fig. 2. Three-layer method: I'' image layer, M_b alpha layer and M'_t transfer layer.

Blending consists in fusing the input images, previously rectified as exposed above, in order to obtain only one image, the mosaic. The method used is not a simple overlapping nor an average filtering of common regions since they have been known to make the seams visible. A flexible method of the type exposed in [23] is implemented. But instead of defining a unique layer for all the pixels of the warped image three layers are defined: one for the warped view I'', another for the transparency (the coefficients are the blending coefficients or the alpha coefficients) M_b and the other for the transfer mask M'_t (Fig. 2). The value I_r of every pixel of the mosaic is the pixel value, I the pixel value, I_r is the product I and the complementary coefficients of M_b with the addition of the product of I'' by the coefficient of M_b only for the pixels equal to 1 in the transfer mask:

$$\begin{aligned} I_r(x_r, y_r) = \quad & I(x, y)\,[1 - M_b(x_b, y_b)] \\ & + \{I''(x'', y'')\, M_b(x_b, y_b)\,|M'_t(x'_t, y'_t) = 1\} \end{aligned} \tag{9}$$

Fig. 3. Result of three-layer blending: left, top and bottom the original images, right the mosaic.

Fig. 3 shows an illustration of the process, the left top image is the reference one, the left down image is warped and the right image is the result of the blending of both images. In this mosaic the transition between the two assembled images is invisible, the quality of the mosaic is high. An algorithm is used to automatically crop the image in order to obtain a rectangular mosaic. This algorithm splits in four quadrants the mosaic according to its centroid. In each quadrant, for every pixel of the border, the area of the rectangle between this pixel and the centroid is computed. The pixel for which the are is maximum, is selected as the target points for the bounding box of the cropping zone.

IV. Inlay of views in the mosaic

The final result of blending is the mosaic which is a large view of the scene at the resolution of input images. If the resolution of the latter is high then the resolution of the mosaic is high. The mosaic defines the background image, it is static. In order to get dynamic images, real time views (video) of the scene are inlaid automatically in that background. First the position of the video frame is found in the mosaic using a correlation algorithm.

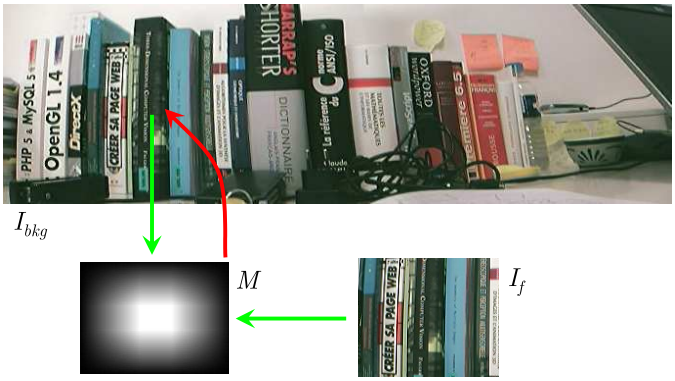

Fig. 4. Method to inlay views in the mosaic.

The algorithm is valid for fixed or mobile imager. Secondly the region of the mosaic of the same size as the video frame is stored and mixed with the video frame:

$$\begin{aligned} I_m(x_m, y_m) = \quad & I_f(x_f, y_f)\, M(x, y) \\ & + I_{bkg}(x_{bkg}, y_{bkg})\,[1 - M(x, y)] \end{aligned} \tag{10}$$

where I_m is the new value of the pixel in the mosaic, I_f the value of the pixel in the frame, I_{bkg} the old value of the pixel and M the alpha layer as explained above. After been mixed, that image is overlaid in the remained mosaic (Fig. 4). The final result a virtual imager combining resolution of the imager (that can be high) with the large field of view.

V. Application

The microimager considered is a miniature videomicroscope from SPI including a 6x objective with a 768×576 pixels camera. The magnification is high then the resolution is also high, but the field of view is only about $3mm^2$ which is

small and does not allow the view of the entire workfield in the majority of micromanipulation applications. For example an assembly means a station where a part is picked up and an other station where the part in placed. The corresponding workfield is very large in comparison of the field of view of the microimager. A xyz stage is used to scan the scene and get the input images of the process. In the first experiment we consider a gear mounted in a watch. We record 25 images with about 50% overlapping rate. We get the mosaic presented Fig. 5 : the size is 1934×1516 pixels with a resolution of $3\mu m$/pixel, the field is about $5.8 \times 4.5mm^2$.

Fig. 5. Mosaicing of a watch gear: 25 images, 1934×1516 pixels, 3 μm/pixel, 26 mm^2 and 7 Mo.

In the second experiment we consider the border of the watch, 36 images with an overlapping rate of 40% are assembled to form the mosaic presented Fig. 6. Its size is 1957×1599 pixels with a resolution of $3\mu m$/pixel. The scanned field is about $5.8 \times 4.7mm^2$. Then, the microimager is positioned somewhere in the work field and its images are projected in that mosaic. The result is a virtual microimager with a high resolution ($3\mu m$/pixel) and large field of view ($5.8 \times 4.7mm^2$). Fig. 6 shows a snapshot from that microimager : a tip is manipulating polystyrene balls of diameter $250\mu m$ over a the border of a watch.

VI. Conclusion

We have developed a way to construct a virtual microimager that combines high resolution with large field of view. First a mosaic is constructed from overlapped images using a feature based approach. The images are registered with robust algorithms and seamless blended. Secondly real time images from the microimager is automatically projected in the mosaic. That kind of microimager will be useful in micromanipulation applications whose require the vision of the entire work field instead only the picking up station or the placing station. Future work will deals with the application of the virtual imaging in real assembly tasks, particularly the assembly of

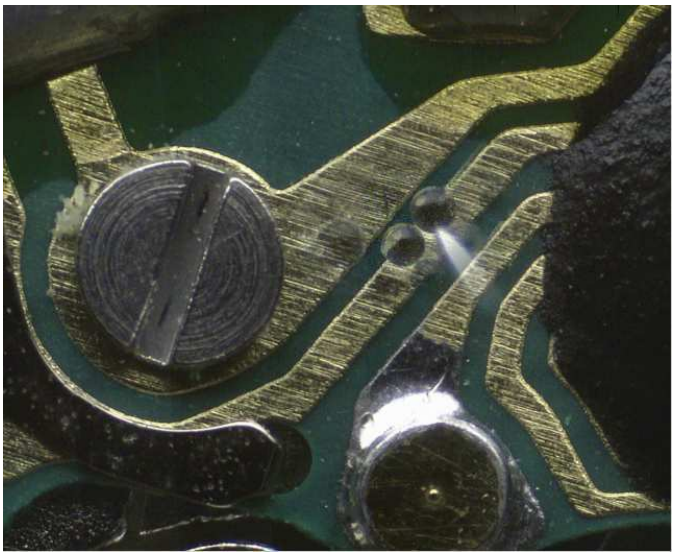

Fig. 6. A snapshot from the virtual microimager: a tip is manipulated polystyrene balls of diameter 250 μm.

silicon parts of $400\mu m \times 400\mu m \times 4\mu m$ under a stereo video microscope with 2x objective and the assembly of silicon parts of $40\mu m \times 40\mu m \times 4\mu m$ under multiple microimaging system with 10x objectives. We will also solve the following problems: the performing of an autofocus in order to avoid defocused images, the projection of images from a mobile microimager, the projection of multiple images.

References

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

[2] A. Matsumoto, T. Akimoto, K. Yoshida, H. Inoue, and K. Kamijo, "Development of mems component assembly machine - application of robotics technology to micromechatronics," in *The International Symposium on Micro-Mechanical Engineering*, December 2003, pp. 83–88.

[3] D. O. Popa and H. E. Stephanou, "Micro and mesoscale robotic assembly," *Journal of Manufacturing Process*, vol. 6, no. 1, pp. 52–71, 2004.

[4] L. Sun, H. Xie, W. Rong, and L. Chen, "Task-reconfigurable system for mems assembly," in *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 2005, pp. 844–849.

[5] S. Chen, "Quicktime vr - an image-based approach to virtual environment navigation," in *Computer Graphics (SIGGRAPH'95 Proceedings)*, August 1995, pp. 29–38.

[6] R. Szeliski, "Image mosaicing for tele-reality applications," Digital Equipment Corporation, Cambridge Research Lab, Tech. Rep. CRL 94/2, May 1994.

[7] B. Potsaid, Y. Bellouard, and J. T. Wen, "Adaptive scanning optical microscope (asom): A multidisciplinary optical microscope design for large field of view and high resolution imaging," *The International Electronic Journal of Optics*, vol. 13, no. 17, pp. 6504–6518, August 2005.

[8] Z. Zhu, G. Xu, E. M. Riseman, and A. R. Hanson, "Fast construction of dynamic and multi-resolution 360 panoramas from video sequences," *Journal of Image and Vision Computing*, vol. 24, pp. 13–26, 2006.

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

[10] C. Guestrin, F. Cozman, and M. G. Simoes, "Industrial apllications of image mosaicing and stabilization," in *International Conference on Knowledge-Based Intelligent Electronic Systems*, Adelaide, Australia, April 1998, pp. 174–183.

[11] A. Zappalá, A. Gee, and M. Taylor, "Document mosaicing," in *Proceeding of the British Machine Vision Conference*, vol. 17, 1997, pp. 589–595.

[12] M. Pilu and F. Isgro, "A fast and reliable planar registration method with applications to document stitching," in *Proceeding of the British Machine Vision Conference*, Cardiff, September 2002, pp. 688–697.

[13] G. H. Kumar, P. Shivakumara, D. S. Guru, and P. Nagabhushan, "Document image mosaicing: A novel approach," in *SADHANA - Academy Proceedings in Engineering Sciences*, vol. 29(3), India, June 2004, pp. 329–341.

[14] P. Shivakumara, G. H. Kumar, D. S. Guru, and P. Nagabhushan, "Sliding window based approach for document image mosaicing," *Journal of Image and Vision Computing*, vol. 24, pp. 94–100, 2006.

[15] B. Rousso, S. Peleg, and I. Finci, "Mosaicing with generalized strips," in *DARPA Image Understanding Workshop*, New Orleans, Louisiana, USA, May 1997, pp. 255–260.

[16] S. Peleg and J. Herman, "Panoramic mosaics by manifold projection," in *IEEE Computer Vision and Pattern Recognition*, June 1997, pp. 338–343.

[17] P. Blanc, E. Savaria, and F. Oudyi, "Le mosaquage d'images satellitales optiques a haute resolution spatiale," in *18e colloque sur le traitement du signal et des images (GRETSI'01)*, vol. 2, Toulouse, France, Septembre 2001, pp. 251–254.

[18] R. Szeliski and H.-Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *Computer Graphics (SIGGRAPH'97 Proceedings)*, vol. 31, 1997, pp. 251–258.

[19] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *IEEE Computer vision and pattern recognition*, 2001.

[20] J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry*. Oxford Press, 1952.

[21] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceeding of 4th Alvey Vision Conference*, 1988, pp. 147–151.

[22] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the Association for Computing Machinery (ACM)*, vol. 24, no. 6, pp. 381–395, 1981.

[23] W. Zhao, "Flexible image blending for image mosaicing with reduced artifacts," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 4, pp. 609–628, June 2006.

[24] R. Inampudi, "Image mosaicing," in *IEEE Geoscience and Remote Sensing Symposium Proceedings*, vol. 5, Seattle, WA, USA, July 1998, pp. 2363–2365.