MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

1—1 Enhanced Vision for the Teleoperator of an All-Terrain Fast Mobile Robot

Christian Tavanti *
DSB/DED/ROBotique
GIAT Industries

Jean Devars †
LIS - Laboratoire Instrumentation et Systemes
University of Paris 6

Abstract

This paper is concerned with the enhancement of the visual feedback for the teleoperator of an all-terrain fast mobile robot. Indeed, remote control trials have shown inferior mobility performances (as speed, accuracy) in relation to direct driving of the same vehicles. Our goal of improving these performances for an operational application needs notably an better visual perception.

This article deals with the first step of this goal. Studies showed notably that the teleoperator needs a large field of view. We decide to take advantage of the rough movement of the camera induced by the robot motion on unstructured ground. We use this environment sweep of the camera to dynamically build an image mosaic which is a way for extending the operator field of view using one camera. The image mosaic is made of the successive video images aligned together to display an coherent and stable environment in spite of the camera motion. The correct alignement is realized using an iterative multiresolution method and points of interest matching.

1 Introduction

We are interesting in improving performances of remote control of an heavy (over 5 t.), fast (over 30 km/h) mobile robot operationg in all-terrain and unknown environment for its use in operational context. Indeed, a decrease in mobility performance in terms of accuracy and speed, as well as an increasing workload for the operator were observed in remote control in regard to direct driving.

This article presents initial works realized to improve teleoperator's vision on the environment where the robot moves.

The approach is notably based on the identified importance of a large field of view (width but height too) and the need to have three-dimensional informations on the environment [5]. Furthermore, we have several other constraints:

- to reduce the data rate involved between the robot and the control-and-command station,
- to reduce costs by using shelf equipment in exploiting at the best their resources,
- the robot moves in an all-terrain unstructured environment, which induces a rough movement of the camera. The result of this visual feedback to the operator can be a loss of spatial orientation and motion sickness, wich deteriorates performance even more.

According to the foregoing, this first approach consists in exploiting one camera and taking advantage of the camera movement induced by the robot motion on all-terrain ground. The camera field of view which scans the environment will be used to extend the field of view displayed to the teleoperator by dynamically compiling an image mosaic. This latter will consist of all the images broadcasted by the robot which are arranged in a coincident way so as to display a coherent environment.

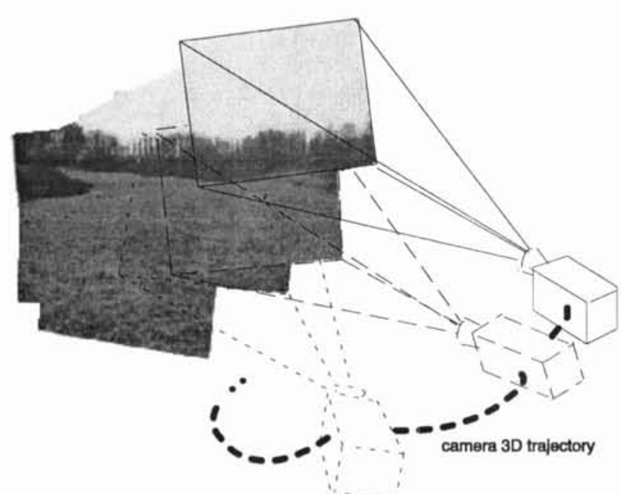


Figure 1. Building of an image mosaic allowing an enlarged field of view by using robot and subsequent video camera changing attitude.

In this paper, after this introduction, we present the method used for the image mosaic building. For this purpose, we use points of interest, which are characterised by a two-dimensional variation on the grey scale surface (high curvature). The matching of

* Adress: 13, route de la Miniere 78022 Versailles Cedex - France
† Adress: Case Courrier 164; 4, place Jussieu 75252 Paris - France
E-mail: devars@ccr.jussieu.fr

these features computed between successive images allows to determine transforms between images in a robust way. The method for matching is based on correlation. However, in order to make this method more robust facing the transforms between images, we use a multiresolution approach: the transform refined in each level is used to compute correlation using neighbourhood shapes determined by the transform.

2 Mosaic building

In order to build the mosaic which will display a wider and more stable view of the environment, positioning of images to each other is required. The context assesses various constraints i. e.: unstructured all-terrain environment, object occultation between two images, no possibility of modelling an a priori unknown environment, simplifying computations (for real time), using caracteristic data within the image (to minimize the error rate caused by less caracteristic informations within the image). On account of these constraints, we chose to do registration with points of interest that are caracteristic of environment areas rich in information (corners, high texture variation areas).

Points of interest use

To detect these points of interest, we use the Harris operator [2] for its simplicity and therefore low processing time.

Let I(x,y) be the image luminance at pixel (x,y) :

$$M = G(x,y,\sigma) \otimes \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix}$$

G gaussian function with standard deviation σ,
I_x and I_y the x and y derivatives of I,

$$R = \det(M) - k \,.\, \mathrm{trace}^2(M) \qquad k = 0.04$$

(x,y) is a point of interest when $R > 0$.

This work is a first step and then we restrict ourself to consider that the field of view extension from the successive images comes to a simple image registration. The parallax problem caused by the three-dimensional structure of the environment is not yet tackled. The registration is made by estimating in a robust way (least median of squares method [6]) the transform between two successive images by matching the points of interest. In this simplified case, the transforms used are affine (valid for straight line moves and for far objects [1]) and projective (more general than the other: e.g. exact for a camera rotated around its optical center [4]) models.

$$\min_{k} \left(\underset{i}{\mathrm{Med}_k} \left(r_i^2 \right) \right)$$

Least Median of Squares (LMedS): r_i is the résidual error of the i-th pair of matched points, that is to say the distance between a point in one image and the matched one in the other image, adjusted by the calculated transform (k index)

In order to position one image to another, one points of interest set has been determined in each one. For a correct positioning, the intersection between the subsets of each image points of interest located in the area shared by both images must be large. It means that a majority of detected feature points within the common region must be the same. The Harris operator enables this result as it was shown by tests with different images in the works of Schmid [3] and Zoghlami [6]; this operator is quite insensitive to Euclidean transforms and to a slight change of the point of view (scale factor, viewing axis, viewing point change). This faculty is called "the repeatability of the points of interest" by Schmid and "the operator stability" by Zoghlami.

The video images are inserted in the mosaic by combining the transform between images. But with this method, there is an accumulation of small errors (roundness, accuracy) after every image insertion. The result may be a bad positioning of the last image with respect to older images. To keep a coherent environment display with the mosaic, we refine the position of the last image directly in relation to the mosaic (and not in relation with the previous image).

Image alignement

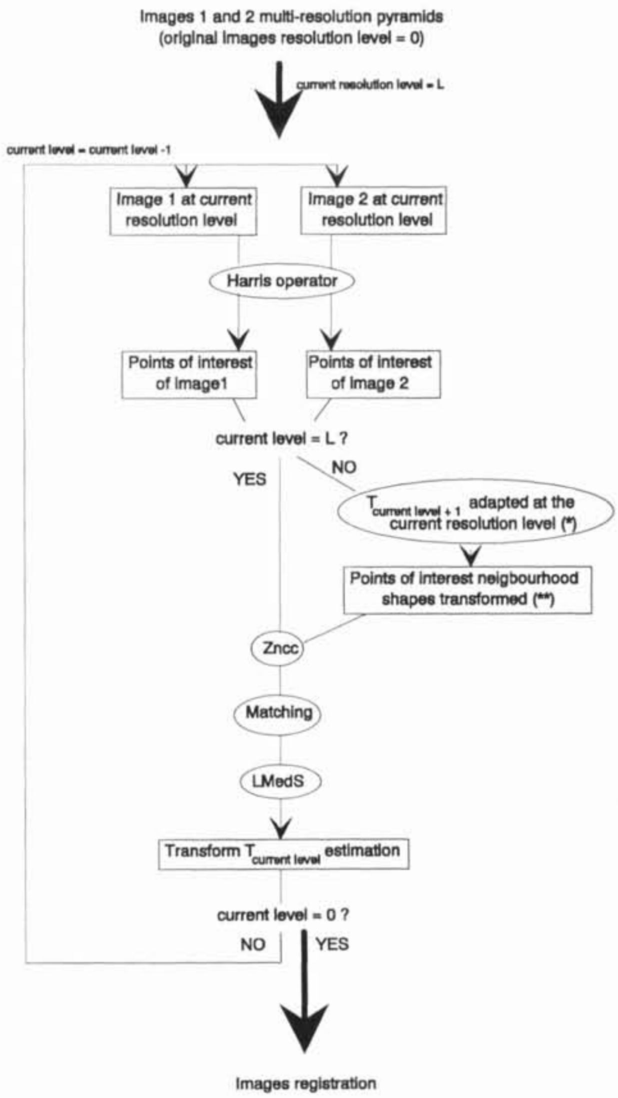
To match these points of interest, we use the information richness in the neighbourhood of these points by using the zncc correlation measure, which leads to good results an is insensitive to luminance uniform variations.

Let (x_1, y_1) and (x_2, y_2) be points of image 1 and image 2, (2m+1,2n+1) is the size of correlation window

$$S((x_1,y_1),(x_2,y_2),m,n) = \frac{\sum_{i=-m,j=-n}^{m,n} \bar{I}_1(x_1+i,y_1+j)\,\bar{I}_2(x_2+i,y_2+j)}{\sqrt{\sum_{i=-m,j=-n}^{m,n} \left(\bar{I}_1(x_1+i,y_1+j)\right)^2}\sqrt{\sum_{i=-m,j=-n}^{m,n} \left(\bar{I}_2(x_2+i,y_2+j)\right)^2}}$$

with $\bar{I}_k(x_k+i,y_k+j) = I_k(x_k+i,y_k+j) - \overline{I_k}(x_k,y_k)$

However, the correlation measures have a lack of robustness in presence of transforms between images, except for translation. In order to reduce the sensitivity to these transforms, deal with larger camera movements and to speed up the registration process, we use a multi-resolution analysis.



(*) For example, if the scale factor between two successive images is 2 for each image dimension and $T_{\text{current level}}$ = (Rotation, Translation) then $T_{\text{current level}}$ adapted is (Rotation, 2*Translation).

(**) See next figure

Figure 2. Multiresolution image alignment process

Indeed, the transform performed at the previous resolution level is used to warp the neighbourhoods of the points of interest. The correlation measure is then more efficient in cases where the traditional approach fails.

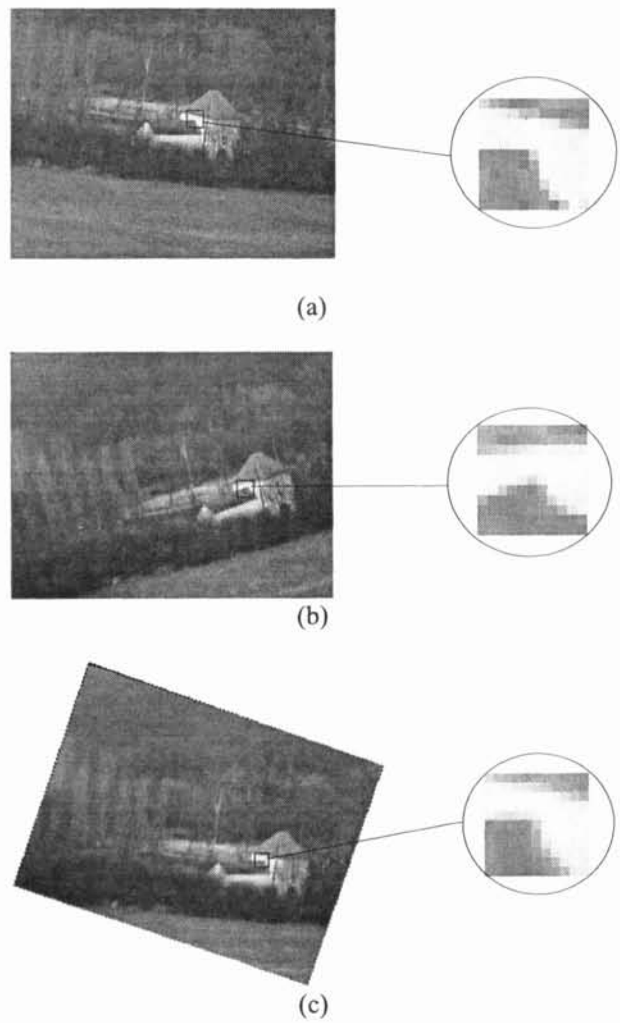


Figure 3. Use of the transforms (a rotation in this example) computed for the lowest resolution levels to warp the neighbourhood areas of the points to be correlated. (a) shows one image and a neighbourhood area of one of its points of interest. (b) shows the next image (to be position adjusted to the previous one) and the neighbourhood of the corresponding point of interest. (c) shows the neighbourhood of the same point of interest (in the second image) transformed as explained in the text, in order to obtain an efficient correlation with the neighbourhood of (a).

5 Conclusion

Our work is pursuing to consider the parallax and then to better manage the mosaic evolution according to the robot motion in the environment. More, our effort will also turn to the 3D computation. That will enrich the operator's informations for a best understanding and positioning in the environment. We do not forget real time implementation for the operational application of these processings.

References

[1] P.J. Burt, P. Anandan, " Image stabilization by registration to a reference mosaic ", ARPA Image Understanding Workshop, 1994

[2] C. Harris, M. Stephens, " A combined corner and edge detector ", 4th Alvey Vision Conference, 1988

[3] C. Schmid, " Appariement d'images par invariants locaux de niveaux de gris ", PhD Thesis, Institut National Polytechnique de Grenoble, France, 1996

[4] R. Szeliski, " Image mosaicing for tele-reality applications ", 2nd IEEE Workshop on Applications of Computer Vision, 1994

[5] C. Tavanti, J. Morillon, J. Devars, "Vision and teleoperation for outdoor rapid mobile robotics", Accepted, to appear in 4th Japan-France/2nd Asia-Europe Congress on Mechatronics, 1998

[6] Z. Zhang, " Parameter estimation techniques : a tutorial with application to conic fitting ", INRIA Research Report n°2676, 1995

[7] I. Zoghlami, O. Faugeras, R. Deriche, " Using geometric corners to build a 2D mosaic from a set of images ", IEEE Conference on Computer Vision and Pattern Recognition, 1997

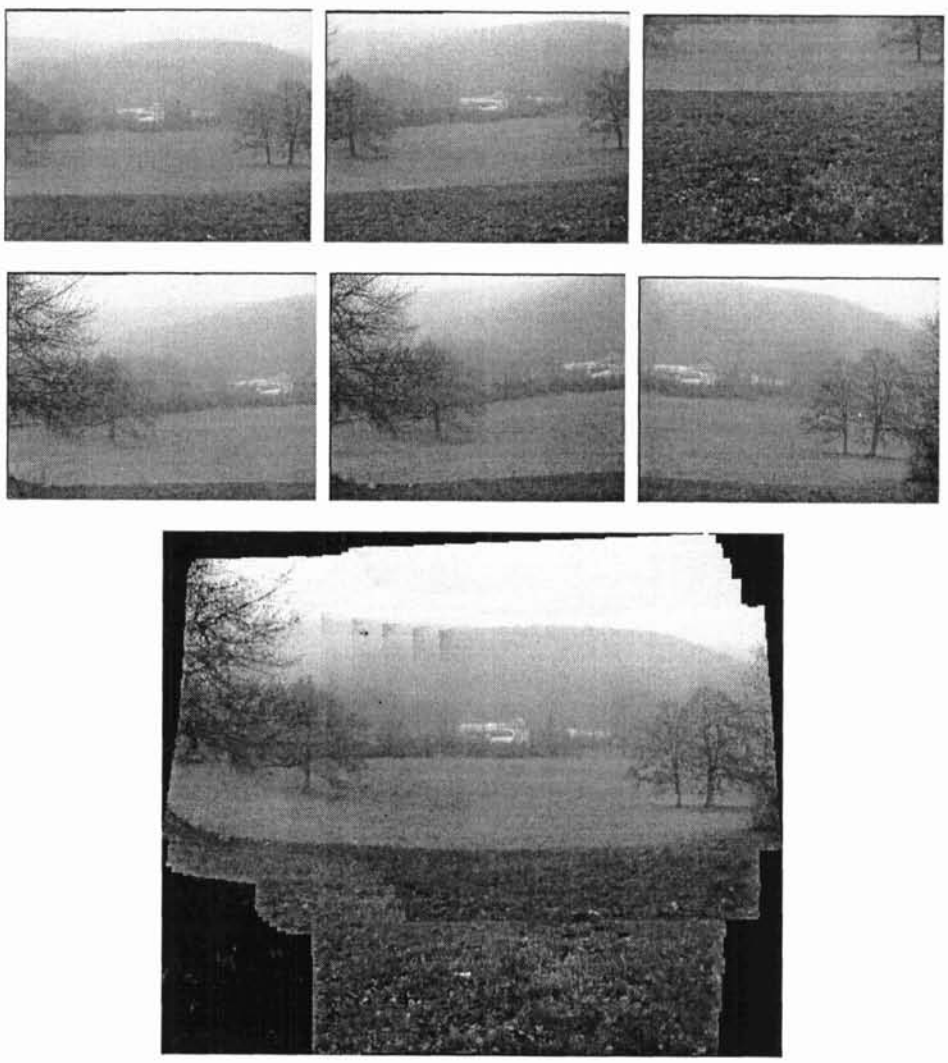


Figure 4. Some images of a sequence as it should be displayed on a monitor and the corresponding mosaic (by using affine transforms between images) wich extend the field of view and improve the situational and orientation awareness. The sequence shows a camera movement and a zoom factor.