MOTION RECOVERY BASED ON FEATURE EXTRACTION FROM IMAGE SEQUENCES

C. Lazar, E. Casandra and A. Lungu,

"Gh. Asachi" Technical University of Iasi
Dept. of Automatic Control and Industrial Informatics
Blvd. Mangeron 53A, 700050 Iasi, Romania
E-mail: clazar@ac.tuiasi.ro

Abstract: The paper concentrates on detection of motion from 2D image sequences and the image analysis used to extract feature points. It is considered a still camera with constant background and a moving object. Object moving across the background will generate changes of its feature points. Detection of these points is used to recover the trajectory of the moving object. The developed procedure for motion recovery was tested to detect the moving of a robot end-effector and real-time experimental results demonstrate the efficiency of the method.

Keywords: Motion recovery, Interesting point, Interest operator, Motion trajectory, Motion field

1. INTRODUCTION

Motion in image sequences captured by a video camera is produced by objects movements in a 3D scene and by camera motion. Since the camera parameters regarding its 3D motion (rotation, translation) or focal length are usually known, only object motion needs to be recovered. The 3D object motion produces 2D motion on the image plane through a suitable projection system. 2D motion, also called apparent motion or optical flow, is recovered from the information of image sequences and it is used in various applications such as image processing and compression or computer vision. Optical flow consists of the computation of the displacement of each pixel between frames. This yields a vector map representing the motion of 3D scene points that is called the motion field.

Changes in an image sequence give features for detecting objects that are moving and for computing their trajectories. To compute motion trajectories, three elements are used (Stiller and Konrad, 1999): a motion model, an estimation criterion and search strategy to find the motion parameters that optimize the chosen criterion. The search strategy has to achieve a trade-off between the optimization performance and computation load. From this reason, simplified search approaches are frequently used including matching and gradient based methods. However, due to the accuracy and robustness aspects, the motion estimation from general image sequences remains an extremely difficult problem.

Estimating the apparent motion from image sequences is used in many applications including robotic motion control, object tracking, autonomous navigation and automatic image sequence analysis (Horn, 1986; Cedras and Shah, 1995; Zucchelli *et al.*, 2002; Davis and Taylor, 2002; Nicolescu and Medioni, 2003). The motion vector field was used to estimate the three-dimensional motion of the imaging system, referred to as ego-motion (Tian et al., 1996), or employing structure from motion (Zucchelli *et al.*, 2002) to determine three-dimensional properties of rigid object. Other applications category is video processing and compression where motion estimation is an important component of the video compression algorithms (Keller and Averbuch, 2003). Also, in medical imaging, a sort of unimodal registration, useful for diagnostic medical conditions, is similar to global motion estimation (Maintz and Viergever, 1998).

This paper presents an algorithm for motion recovery based on future extraction from image sequences. An image processing technique is applied to extract interesting points from the first image of the sequence. After that, corresponding points are computed for the following images of the sequence. The corresponding points are used to compute motion vectors and, based on them, the motion trajectory is recovered. The developed algorithm for motion recovery was tested using a workspace which consists of a moving robot end-effector, a fixed object and a still camera. Real time experimental results are given.

2. MOTION RECOVERY

Our aim is to find an important feature of the motion, the trajectory, and to use it in tracking a robot end-effector in a grasping application. For this reason, a motion recovery algorithm based on feature extraction from image sequences was developed.

2.1 Motion estimation

An image acquisition system generates measurements of the image intensity function $f(\mathbf{x},t)$, which represents the light emanating from a point $\mathbf{X} = (X,Y,Z)^T$ on an object moving in 3D space. At time t, the position of $\mathbf{X}$ is considered:

$$\mathbf{X}(t) = \left(x(t), y(t), z(t)\right)^T \in \mathbb{R}^3 \qquad (1)$$

expressed in camera coordinates. Using (1) ($\mathbf{X}(t)$, t) defines a 3D curve over time called world motion trajectory. For any two time instants t_0 and t_1, the world motion trajectory provides a 3D displacement in position:

$$\mathbf{D}(\mathbf{X}) = \mathbf{X}(t_1) - \mathbf{X}(t_0) \qquad (2)$$

The image acquisition system projects the 3D point $\mathbf{X}$ onto a 2D image plane with image coordinates:

$$\mathbf{x}(t) = (x(t), y(t))^T \in \Lambda \qquad (3)$$

where Λ is the sampling grid. This projection transforms the world motion trajectory in a 2D motion trajectory ($\mathbf{x}(t)$,t), only if in the time interval the associated point is visible in image (Dubois and Konrad, 1993).

Starting from (2), the 3D displacement leads to the 2D vector displacement:

$$\mathbf{d}(\mathbf{x}) = \mathbf{x}(t_1) - \mathbf{x}(t_0), \qquad (4)$$

as depicted in Fig.1. The motion vector in the image represents displacements of the images of moving 3D points.

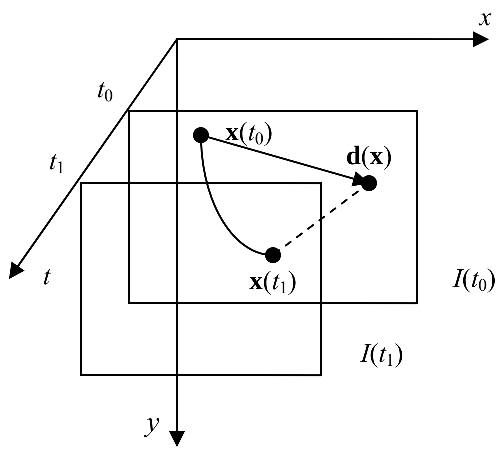

Fig.1 Motion vector $\mathbf{d}(\mathbf{x})$

Each motion vector is formed, as in Fig.1, with its tail at an imaged 3D point at time t_0 and its head at the image of the same 3D point imaged at time t_1.

Alternatively, each motion vector might correspond to an instantaneous velocity estimated at time moments t_0 and t_1. The velocity vectors field is defined as:

$$\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x},t_0), v_2(\mathbf{x},t_1))^T \qquad (5)$$

In this case the image intensity functions are modeled as temporally evolving according to:

$$f\left(\mathbf{x},t\right) = f\left(x - v_1\left(\mathbf{x},t\right), y - v_2\left(\mathbf{x},t\right), 0\right) \qquad (6)$$

Equation (6) specifies that the intensity function for a given region remains the same even the location of the region moves as a function of time and it is known as the intensity conservation assumption (Horn, 1986).

The objective of motion estimation is to compute the vector field based on measurements of the image sequence $f(\mathbf{x},t)$. There are two methods for estimating 2D motion: motion correspondence and optical flow. Motion correspondence matches interesting image features that can be tracked through time. Optical flow consists of the computation of the displacement of each pixel between frames. This gives a vector map of flows in the image and individual or regional flows may be analyzed/tracked. The first approach leads to a sparse motion field and using the pixel based method a dense motion field is obtained.

The sparse motion field is based on matching and can be computed by identifying pairs of interesting points that correspond to two images taken at times t_0 and, respectively, t_1. Estimating image flow at all points of an image leads to a dense motion field using, e.g., differential methods based on computing the velocity from spatio-temporal derivatives of image intensity.

2.2 Deriving motion vectors from interesting points

Taking into account the necessity of a reduced computation load for the real time application of tracking the robot end-effector, a sparse motion field for motion analysis was chosen. In order to compute the sparse motion field, an interest operator (Shapiro and Stockman, 2001) was developed and implemented. The operator determines the variance v_i of the image intensity function in the horizontal, vertical and diagonal directions of a neighborhood centered in the pixel $\mathbf{x}$. Considering the 3x3 neighborhood depicted in Fig. 2, the variances on i directions are computed as:

$$\begin{aligned}
v_1(x,y) &\triangleq \max\Big(\big|f(x+1,y)-f(x,y)\big|, \\
&\qquad \big|f(x-1,y)-f(x,y)\big|\Big) \\
v_2(x,y) &\triangleq \max\Big(\big|f(x,y-1)-f(x,y)\big|, \\
&\qquad \big|f(x,y+1)-f(x,y)\big|\Big) \\
v_3(x,y) &\triangleq \max\Big(\big|f(x-1,y+1)-f(x,y)\big|, \\
&\qquad \big|f(x+1,y-1)-f(x,y)\big|\Big) \\
v_4(x,y) &\triangleq \max\Big(\big|f(x+1,y+1)-f(x,y)\big|, \\
&\qquad \big|f(x-1,y-1)-f(x,y)\big|\Big)
\end{aligned} \tag{7}$$

The pixel location (x,y) is declared an interesting point only if the minimum of the four variances given by (7) exceeds a threshold α.

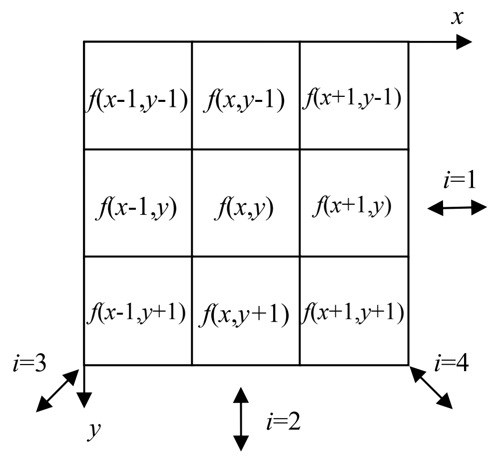

Fig.2 $F(x,y)$- the 3x3 neighborhood of $f(x,y)$

The locations of interesting points constitute a map $m(x,y)$, which is defined as:

$$m(x,y) = \begin{cases} 1, & (x,y) \in I_f \\ 0, & \textit{otherwise}, \end{cases} \tag{8}$$

where

$$I_f \triangleq \left\{(x,y);\ \min_i V_i(x,y) \geq \alpha\right\}. \tag{9}$$

Typically, the threshold α is selected so a few pixels are declared interesting point.

The interest operator might be implemented using the masks from Fig. 3 which permits the computation of differences from (7) via convolution with the image window from Fig. 2.

0	0	0
1	-1	1
0	0	0

i=1

0	1	0
0	-1	0
0	1	0

i=2

0	0	1
0	-1	0
1	0	0

i=3

1	0	0
0	-1	0
0	0	1

i=4

Fig. 3 Masks of interest operator

After finding the interesting points $\mathbf{x}(t_0)$ identified on the image $I(t_0)$ taken at time t_0, corresponding points $\mathbf{x}(t_1)$ must be extracted from the image $I(t_1)$ taken at time t_1.

A way of determining $\mathbf{x}(t_1)$ is based on using the cross-correlation method under the assumption that the amount of movement is limited. Thus, the interesting point $\mathbf{x}(t_0)$ found in the image $I(t_0)$ is considered now in a mxn small search window $W(x,y)$, with $m,n > 3$ having the center (x,y). For the interesting point $\mathbf{x}(t_0)$. a mxn window of image $I(t_1)$ is searched for finding the best match to the small neighborhood $F(x,y)$ of $\mathbf{x}(t_0)$. The center of the best correlated neighborhood in $I(t_1)$ is considered to be the corresponding point $\mathbf{x}(t_1)$.

With interesting points $\mathbf{x}(t_0)$ and $\mathbf{x}(t_1)$, the motion vector can be constructed, as in Fig. 1, having the tail $\mathbf{x}(t_0)$ and the had $\mathbf{x}(t_1)$. In Fig. 4 is depicted the process of cross-correlation which leads to the finding of the best correlated 3x3 neighborhood having as center the interest point $\mathbf{x}(t_1)$. Starting from 3x3 neighborhood $F(x,y)$ from image $I(t_0)$, considering the mxn search window $W(x,y)$ and using the cross-correlation of window $W(x,y)$) from $I(t_1)$ with the neighborhood $F(x,y)$, defined as:

$$\begin{aligned} g(x,y) &= W(x,y) \otimes F(x,y) = \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} W(x+i,y+j) F(i,j) \end{aligned} \tag{10}$$

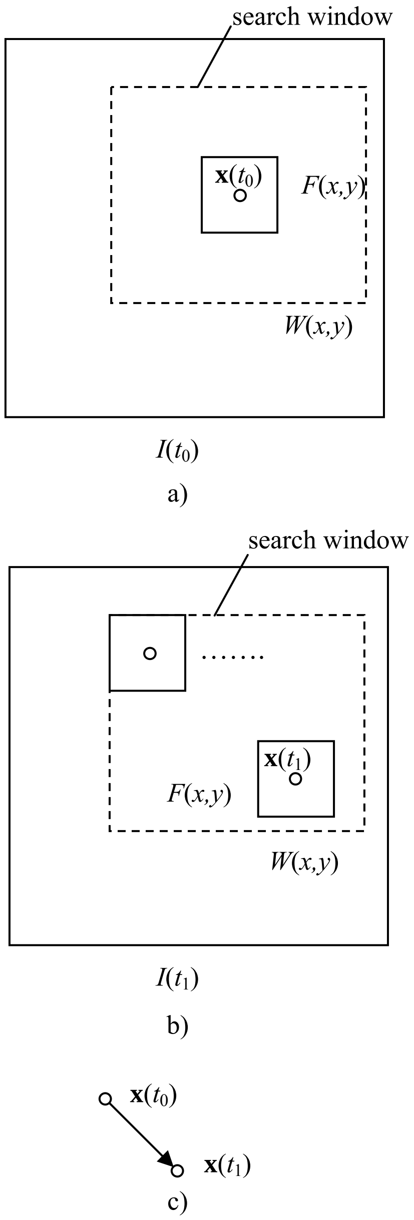

Fig. 4 Process of finding motion vector:
a) neighbourhood $F(x,y)$; b) the correlating neighbourhood; c) motion vector.)

the best match is obtained having as center the corresponding interest point $\mathbf{x}(t_1)$ of $\mathbf{x}(t_0)$.

Aiming at obtaining a high search speed it is necessary to choose a small search window $W(x,y)$, but in relation with the velocity of the object.

The identified pairs of points $\mathbf{x}(t_0)$ and $\mathbf{x}(t_1)$ that correspond in two images taken at times t_0 and respectively t_1, are employed to compute a sparse motion field. This field permits motion recovery from image sequences.

The proposed algorithms for motion recovery can be summarized as follows:

Step 1. Take a sequence of monocular and monochromatic images as input;

Step 2. Extract the interesting points from the first image using the interest operator defined by relations (7)-(9) or using the convolution of the image with the masks from Fig.3;

Step 3. Calculate the corresponding point in the following images of the sequence based on interesting points and using the cross-correlation defined by (10);

Step 4. Use point correspondences and compute the sparse motion field based on motion vectors;

Step 5. Display the recovered motion trajectory.

3. EXPERIMENTAL RESULTS

The motion recovery algorithm has been tested with a workspace consisting of a moving robot end-effector, a fixed rectangular object and a still camera located up the scene. The sequence utilized in the motion recovery testing and depicted in Fig. 5 has 5 images captured at time moments t_0,..,t_4 representing the 5 successive position from a grasping application.

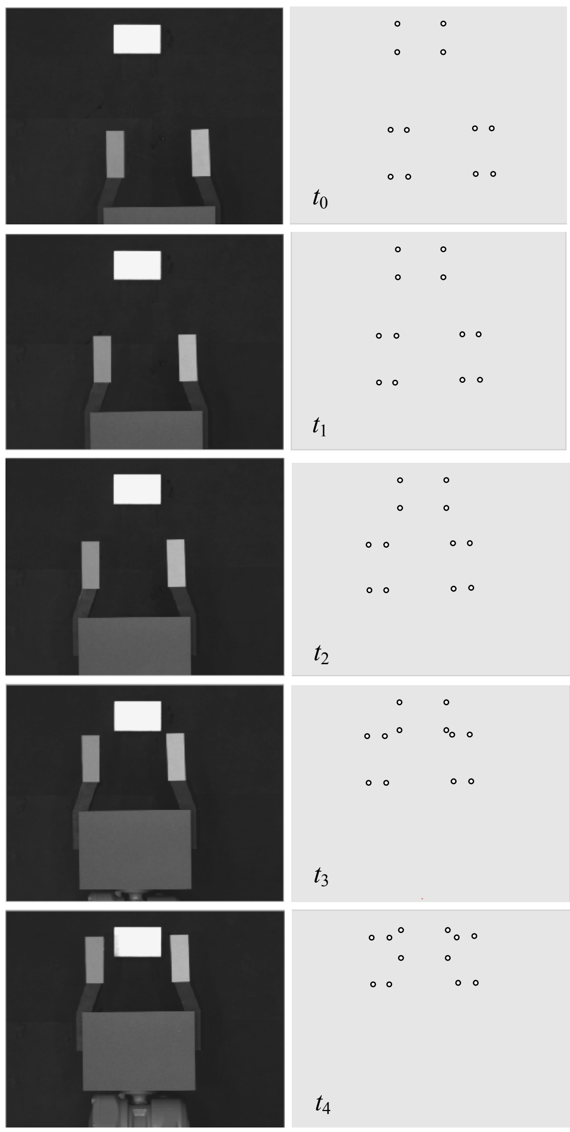

Fig. 5 The 5 positions of end-effector and the corresponding interesting points

For image acquisition the camera of OptiMaster system was used and the motion recovery algorithm was implemented on a Matlab environment running on a Pentium-IV 1 GHz with 512 Mb of system memory. An image preprocessing algorithm was used to transform the color image acquired with OptiMaster camera in a filtered monochrome image.

The considered moving object was the end-effector of an IRB 2400 robot which had bright surfaces. In order to eliminate the disturbances introduced by reflections during the moving, mat surfaces were added with different grey levels. For each position, both grey level workspace images and the corresponding interesting points are shown in Fig. 5. The interesting points were computed for the fixed object and for the fingers of the end-effector.

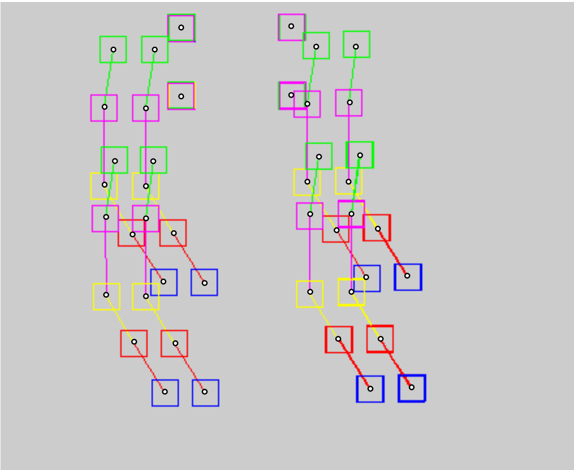

Fig.6. Motion trajectory

Using the motion recovery algorithm from Section 2, the obtained motion trajectory is shown in Fig. 6.

4. CONCLUSION

An algorithm for motion recovery based on interesting points extraction from image sequences was presented. Experiments show that reconstructed results are encouraging while some improvements are needed, the algorithm being depending on the object shapes. Future work includes the usage of this algorithm for complex object tracking by visual servoing based on 2D image motion.

REFERENCES

Cedras, C., and M. Shah (1995). Motion based recognition: A survey, *IEEE Proceedings Image and Vision Computing*, **13**, pp. 129-155.

Davis, J. and S. Taylor (2002). Analysis and recognition of working movements, *Proceedings of 16th International Conference on Pattern Recognition*, **1**.

Dubois, E. and J. Conrad (1993). Estimation of 2D motion fields from image sequences with application to motion-compensated processing, In *Motion Analysis and Image Sequence* (Eds. M. Sezan and R. Lagendijk), Kluwer Academic Publishers, Ch. 3, pp. 53-87.

Horn, B.K. (1986). *Robot vision*, Cambridge:MIT Press.

Keller, Y. and A. Averbuch (2003). Fast gradient methods based on global motion estimation for video compression, *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, pp. 300-309.

Maintz, A. and M. Viergever (1999). A survey of medical image registration, *Medical Image Analysis*, **2**, pp. 1-36.

Nicolescu, M. and G. Medioni (2003). Motion segmentation with accurate boundaries – a tensor voting approach, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*

Shapiro, L. and G. Stockaman (2001). Computer Vision, Prentice Hall.

Stiller, C. and J. Konrad (1999). Estimating motion in image sequences – A tutorial on modeling and computation of 2D motion, *IEEE Signal Processing Magazine,*July*,* pp. 70-91.

Tian, T., C. Tomasi and D. Heeger (1996). Comparison of approaches to ego-motion computation, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.315-320.

Zucchelli, M., J. Santos-Victor and H. I. Christensen (2002). Constraint structure and motion estimation from optical flow, *Proceedings of 16th International Conference on Pattern Recognition,* **1**.