VISUAL TRACKING BASED ON ACCUMULATED DIFFERENCES [1]

Manuel Vargas and **Francisco R. Rubio**

Dpto. Ingeniería de Sistemas y Automática
Escuela Superior de Ingenieros
Avda. Reina Mercedes s/n, 41012-Sevilla (Spain)
Tel: 34-5-4556855, Fax: 34-5-4556849, E-mail: vargas@cartuja.us.es

Abstract. This article presents an algorithm for tracking an object using a robot provided with a camera in the final effector. Simple methods are studied for estimating the optical flow based on accumulated differences which permit tracking in real time. The study is first realized by simulation and then the best results are tried experimentally.

Keywords. Robotics, Image processing, Visual motion, Optical flow, Tracking.

1. INTRODUCTION

This article is centered on the analysis of the tracking problem linking a vision system and an articulated arm by the adaptation of a camera in the final effector of the arm (this is termed *eye-in-hand configuration*). The following problem is posed: "Observing an object on the visible scene, compensate its displacements in such a way that it always occupies the same position in the image" (preferably at the center).

The information about the movement achieved from a sequence of images can be characterized by the so called *optical flow* (Ballard D.H., 1982). The optical flow is a consequence of the relative movement between the camera and the objects on the scene. Several techniques for estimating the optical flow have been developed (Martin W.N., 1988), (Fu K.S., 1986), (Papanikolopoulos N.P., 1993), notable amongst them being the techniques based on the *intensity function gradient*. Such techniques are based on the so called *gradient constraint equation*, which relates at every pixel, the space-time gradient to the velocity vector associated to that pixel. There are also some studies combining stereoscopy with the optical flow detection (Allen P.K., 1993). In their original form, all these algorithms are very restricted. On one hand, they assume that any change of the intensity function at each pixel is only due to the movement. On the other hand, they require images without discontinuities in space and time.

Many mono-camera visual tracking algorithms try to detect and follow well-known objects, which have some visual features, the position of each is accurately known relative to an object coordinate system (Hashimoto, 1993).

The tracking process establishes strong impositions in real time because an immediate response to the displacements of the object is needed. This article presents a method based on the accumulated difference technique, which allows for a rapid response using not too powerful hardware. In this method the interest is centered in detecting and tracking an object without a previously defined shape or structure. We are just focused to objects which are distinguible from the background and which are moving (since the purpose is to track moving objects, the motion itself will be used as a differenciating property).

Hence, our purpose is to track arbitrary shaped moving

[1] The authors would like to thank to CICYT for supporting this work under grant TAP 95-0370.

spots in two dimensions.

2. OPTICAL FLOW

The first approximation for creating a difference image (Fu K.S., 1986) is based on the criterion expressed in the equation,

$$ADI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } |I(x,y,t_i) - I(x,y,t_j)| > \theta; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right\} \quad (1)$$

The new image, called ADI (*Absolute Difference Image*) is the result of a comparison between two *frames* (grabbed images) successive in time, the first one taken at t_i and the second one at t_j. This difference between the intensity of the pixel before and afterwards is considered significant or negligible, depending on the threshold value used (θ).

Thus in the ADI image those pixels of the differences resulting from the movement will appear as ones, and also those due to noises which cause a change of intensity superior to the θ threshold. The small areas of 1's which appear can be due to noise or to real but insignificant differences, are eliminated by techniques of *erosion* and posterior *dilation* of the image objects; this will also regularize the shape of said objects (as said before, the exact object's shape is not of interest).

The set of pixels of the mobile object which take up new positions in the image which previously belonged to the background, will be called the *leading edge* of the object. The set of pixels of the mobile which leave positions which previously were occupied by the object will be called the *leaving edge* of the object.

The ADI image will show both the object's leading edge and leaving edge as 1's. However it is often more interesting to obtain in a difference image only one of these edges. This leads us to two new types of difference images. The way of calculating both is given by the equations,

$$PDI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } I(x,y,t_i) - I(x,y,t_j) > \theta; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right\}$$

$$NDI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } -(I(x,y,t_i) - I(x,y,t_j)) > \theta; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right\} \quad (2)$$

This formulation does not ensure that one edge or the other is obtained independently. This is only true when the range of intensities of the mobile is higher than the one of the background (in this case PDI provides the leaving edge and NDI the leading edge), or when the intensity range of the mobile is lower than the one of the background (PDI gives the leading edge and NDI the leaving one). However this situation deteriorates if the range of the object's intensities is higher than the background's in some areas and lower in others.

2.1 *Alternative Formulation of the Difference Method*

Let's suppose the intensity range of the mobile is known: $r = [i_{min} \cdots i_{max}]$ (even if this range is not known from the beginning, it can be estimated taking advantage of the motion of the object of interest, see (Vargas, 1997)).

In order to obtain an image with the leading edge and separately another with the leaving edge the following formulation can be used:

$$PDI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } I(x,y,t_i) \in r \text{ and } I(x,y,t_j) \notin r; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right\}$$

$$NDI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } I(x,y,t_i) \notin r \text{ and } I(x,y,t_j) \in r; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right\} \quad (3)$$

$$ADI_{ij}(x,y) = \left\{ \begin{array}{l} 1 \text{ if } PDI_{ij}(x,y) = 1 \text{ or } NDI_{ij}(x,y) = 1; \ (t_j > t_i) \\ 0 \text{ otherwise;} \end{array} \right.$$

whatever the relationship between the intensities of the background and those of the mobile might be.

In this case it is unimportant that in some regions the intensities of the mobile are above and in other regions below those of the background. The only problem which may arise is if there are intensities of the background within the r range, because in this case it is not possible to distinguish what is background and what is object; these regions will be called *interference regions*. Actually, these regions have no effect during tracking so long as the mobile does not pass over any of them.

Up to now, the difference images have been obtained using only two consecutive images. However, what is usually used is the so called *accumulated difference method.*

In this method a first image (I_o) is taken as a reference, R; and the following n frames are all compared to R and are accumulated on top of the same resultant image. Next, the new reference to be used for the following n frames is taken, and so on. Thus, the equations will now be like:

$$NDI_j(x,y) = \left\{ \begin{array}{l} 1 \text{ if } R(x,y) \notin r \text{ and } I(x,y,t_j) \in r; \\ 0 \text{ otherwise;} \end{array} \right\} \quad (4)$$

The fact that these simple differences are accumulated can be expressed by the equation,

$$NADI_n(x,y) = \sum_{j=1}^{n} NDI_j(x,y) \quad (5)$$

Thus the PADI (*Positive Accumulated Difference Image*), NADI (*Negative Accumulated Difference Image*) and AADI (*Absolute Accumulated Difference Image*). images arise. The PADI indicates, for each pixel on which the object was in the reference frame, the number of frames (from the reference) in which the object has been absent from this pixel. The NADI indicates the number

of frames (from the reference) in which this pixel, which initially was not occupied by the object, has been occupied by said object.

Figure 1 illustrates these concepts with an example. It shows a mobile being displaced at the rate of 1 pixel/frame to the right.

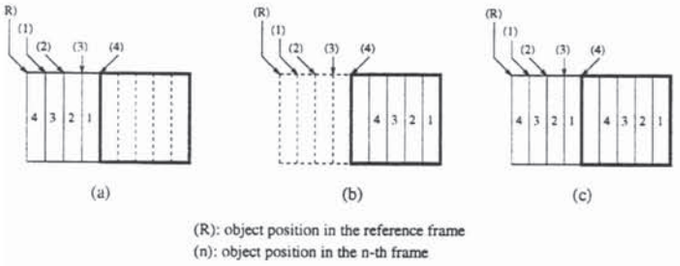

Fig. 1. (a) PADI differences. (b) NADI differences. (c) AADI differences.

This information about the number of frames in which there was no object (for the PADI) or there was it (for the NADI) is translated into the number of frames elapsed since the object left (for the PADI) or reached (for the NADI) said pixel, (this second interpretation of the accumulated differences allow, as will be seen, the velocity information to be obtained). However, this interpretation is not valid, for example, if the mobile is very small in relation to the number of frames which are accumulated.

Let's suppose, for instance, the object in *Figure 2*, paying attention to the pixel (x, y) showed, and calculate the NADI which would be generated over this pixel. In case (a), in which the object moves to the right at the rate of one pixel per frame, the object begins to be over (x, y) when frame 3 is taken, and in all the following frames the point will continue to be occupied by the object; therefore, starting from frame 3 the value accumulated over the pixel is increased by 1 for each frame taken. At the end $NADI(x, y) = 5$ is obtained; and this coincides with the number of frames grabbed since the pixel was occupied.

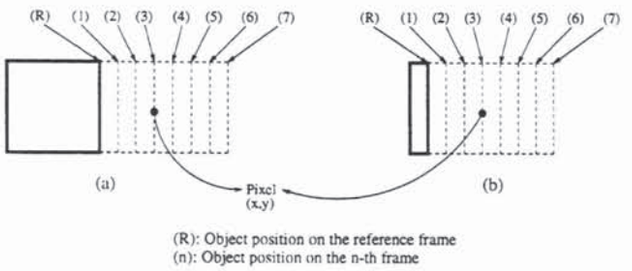

Fig. 2. Object moving at 1 pixel/frame towards the right. (a) Wide object. (b) Thin object.

Let's now look at case (b) of *Figure 2* in which the object is specially narrow. The pixel is only occupied by the mobile during frames 3 and 4, therefore the value finally accumulated in $NADI(x, y)$ is 2. However, at the end, the number of frames elapsed since the pixel was occupied is 5 (the same as in case (a)).

This inconvenient is almost overcome in the velocity extraction process, given that the difference between the values accumulated in a pixel and its neighbors and not the absolute values are considered, as will be shown in the next section.

It is convenient to point out a significant difference between PADI and NADI; It is obvious that both stop growing when the object stops, but the PADI also stops growing when the object completely quits the area it was occupying on the reference image. This point can be used, for example, to extract static images from images on which, from the beginning, there are mobile objects.

2.2 *Optical Flow Starting from Accumulated Differences*

The optical flow field (the field of velocities at each pixel) can be obtained from an image of accumulated differences. We are going to present this deduction starting from the NADI.

The image containing the NADI will be called D for short throughout this deduction. If, as said before, it can be assumed that the differences accumulated at every pixel can be taken as the number of frames elapsed since the object occupied the pixel then, $D(x+1, y) - D(x, y)$, represents the number of difference frames from when the object occupied the position (x, y) until it occupied $(x + 1, y)$. Given that a discrete approximation to the derivative is a difference quotient, the following equation can be written,

$$D_x = \frac{\partial D}{\partial x} \approx \frac{D(x+1,y) - D(x,y)}{(x+1) - x} = D(x+1,y) - D(x,y)$$
$$D_y = \frac{\partial D}{\partial y} \approx \frac{D(x,y+1) - D(x,y)}{(y+1) - y} = D(x,y+1) - D(x,y) \quad (6)$$

The horizontal component of the velocity v_x is the number of pixels passed over horizontally per frame elapsed (the time unit is the frame). We can define the v_y component in the same way.

If D_x is the number of frames elapsed for the object to advance one pixel (assuming that it moves at a constant speed), and v_y is the number of pixels passed over per frame elapsed, one is the inverse of the other. However, there is one more detail, if the object moves to the right the NADI decreases in this direction, therefore D_x is negative; however, the velocity is positive in that direction. According to this, the relationship between the components of the velocity and of the difference image gradient is,

$$v_x = -\frac{1}{D_x} \; ; \quad v_y = -\frac{1}{D_y} \tag{7}$$

In order to make this gradient calculation more robust, the pixel's 8-neighbours will be taken into account, using the Sobel masks. In this way, an approximation for the gradient is shown in the following equation,

$$D_x = -\frac{Sobel_x}{8} \; ; \quad D_y = -\frac{Sobel_y}{8} \tag{8}$$

2.3 *Estimation of the Centroid of the Mobile*

Another method making use of accumulated difference images will be shown. It does not try to estimate the optical flow at each pixel, but the centroid of the mobile at each instant. It implements a very intuitive idea to solve the tracking problem.

The *centroid* of the NADI region produced by the mobile will be calculated, and this centroid is assumed as an estimation of the real object's centroid at each instant. This method uses accumulated difference images although the accumulated values at each pixel themselves are not of interest, but the extension and location of the accumulated difference region.

This is an approximation which can be inaccurate under some conditions. In *Figure 3* three cases are presented.

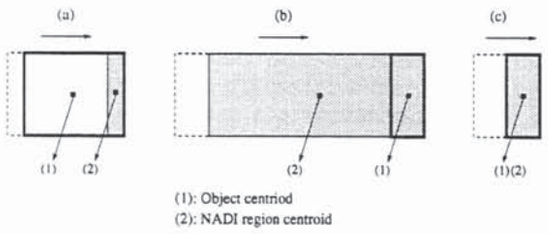

Fig. 3. Centroid of the NADI region produced by the mobile.

(a) In this case the area of the NADI region produced by the mobile object is small relative to the object area; this causes the centroid of the NADI region to be far from the real centroid of the mobile. The greater the difference between the mentioned areas the greater the error.

(b) This is another extreme case in which the NADI region is much wider than the mobile itself. This case is worse than the previous one, here the discrepancy between the position of the NADI region centroid and the real object centroid can be very large.

(c) This is the best case, here the estimation is very precise. The area of the mobile and the area of the NADI region are quite similar. In this case the mobile occupies almost exactly the NADI region and because of this both centroids practically coincide.

As can be seen the precision of the estimate is strongly dependent on the quotient *NADI area / mobile area.* The most favorable case is when this quotient is close to one. Actually, the situation in case *(a)* is not incorrect from the tracking point of view, because the robot is directed in front of the object.

2.4 *Restrictions of Difference Based Methods*

The methods based on differences have several restrictions such as:

- The range of intensities of the mobile must not change too much throughout the process. This means that there cannot be significant illumination differences along the path followed by the mobile.
- If there is a significant area of interference the algorithms continue working well, provided that the mobile does not pass over those regions which interfere.
- The velocity of the mobile should be as little variable as possible, at least during each accumulation cycle, for the optical flow estimation method (*Section 2.2*).
- The method of estimating optical flow is affected by irregularity in the objects' shape, while the centroid estimation method (*Section 2.3*) is indifferent to this aspect.

3. PROPOSED METHOD OF ACCUMULATED DIFFERENCES

The proposed method consists of estimating the centroid of the mobile using accumulated differences (NADI). Starting from this, the absolute position of the mobile in the image (more precisely, the position of its centroid) can be estimated. Provided that the aim is to keep the object centered, the displacement which should be applied to the camera is given by the difference between the position of the centroid and the coordinates of the image center.

The displacement vector thus obtained is transformed into the universal reference system, and the robot is ordered to displace the camera according to the resulting vector. The general structure is shown in *Figure 4*.

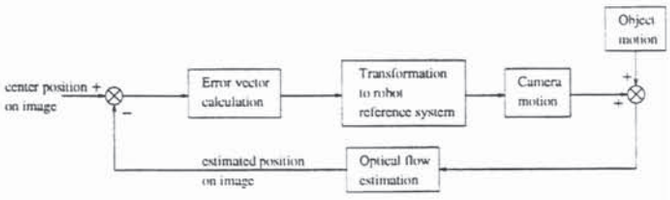

Fig. 4. Block diagram of the tracking process.

In the tracking process an adjustment of the ratio between pixels and distance in the real world (let's call this ratio the *scale factor*) is previously required. If this factor is not accurately known, or if the depth of the object trayectory is not constant, relative to the camera plane of motion, it can be estimated and adapted on line.

It is simpler to use the *look-and-move* approximation. That is, the robot remains motionless while the secuence of n consecutive images (the first of which is the reference frame in equation 4) is being grabbed. Otherwise, motion information is generated due to the camera movement, in addition to the object motion. This undesired information affects to the interference regions too, and must be removed making neccesary a very precise calibration of the scale factor.

The main advantage of this method in relation to the use of simple binarization is that it can cope with nonperfectly structured scenes. That is, there can be some background areas which have the same intensity range as the object of interest.

4. SIMULATIONS AND EXPERIMENTAL TESTS

In order to test the proposed method, in the first place simulations using simple synthetic images have been carried out. The point $(0,0)$, the origin of the graphs which are presented, is the point which occupies the center of the camera at the moment when tracking begins.

Firstly the behavior in the ideal case is analysed. The ideal conditions are given by:

- Mobile object of regular shape (rectangular).
- Constant velocity of the mobile. (at exactly 1 pixel per frame towards the right and down: $(1,-1)$).
- The scale factor used by the algorithm has its real value. In following examples the effect of using an inexact scale factor is shown (so, a deficient calibration of this parameter will be simulated).

In all the simulated examples, the starting points of the trajectories are the same: the object centroid at $(-16,10)$, and the camera center at $(0,0)$. The trajectory of the mobile is represented by a continuous line and the center of the camera by a dashed one.

Figure 5 shows the trajectories from a simulation for the given conditions. *Figure 6* shows the y-coordinates of the mobile and camera center, corresponding to those trajectories.

Figure 7 shows the response when a sudden change in the mobile trajectory at frame number 30 is given. The mobile changes from a displacement in a south-east direction to a north direction. The same trajectory is shown in *Figure 8* but using scale-factor automatic adjustment.

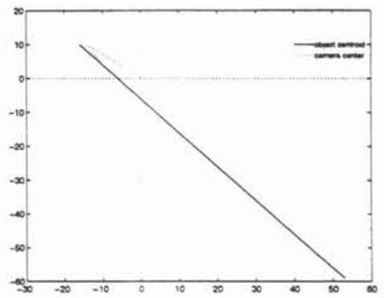

Fig. 5. Trajectories using the proposed method under ideal conditions.

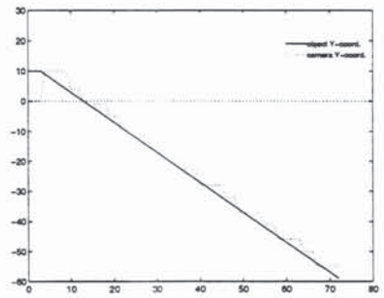

Fig. 6. Y-coordinates using the proposed method under ideal conditions.

It can be seen how the changes in the trajectory affect this method.

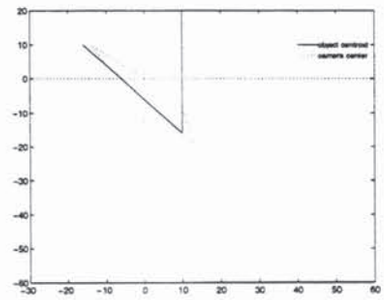

Fig. 7. Trajectories under ideal conditions, in the presence of a sudden change in the trajectory.

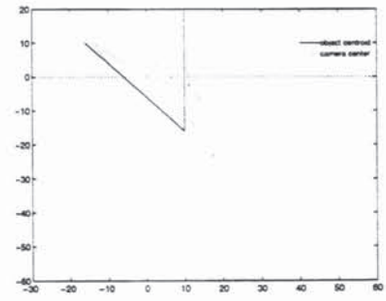

Fig. 8. Trajectories using the scale-factor adjustment method.

Figure 9 presents the trajectories when using the algorithm without scale-factor correction mechanism, and using a scale factor twice its real value.

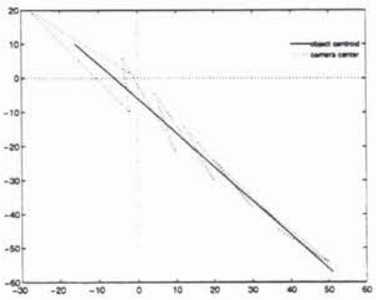

Fig. 9. Trajectories in the case of a system ill-calibrated, without scale-factor correction mechanism.

Figure 10 shows a comparison of the horizontal coordinates, when automatic correction of the scale factor is made and when it is not made. It can be seen that correcting the scale factor improves the tracking when the system is not well-calibrated.

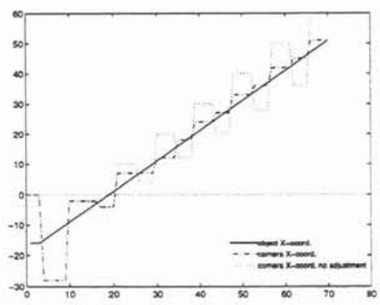

Fig. 10. Comparison of the X-coordinates with and without adjustment of the scale factor.

In view of the simulation results, a real test has been made using the proposed method with constant scale factor and calibrating the system before the test. The components used were:

- A *PUMA 560* robot.
- A *CCD camera* attached to the end-effector of the robot.
- A personal *486 computer* with:
- A Matrox board model *Image-1280*, for image processing.

In the next figures, continuous lines represent the camera motion, and dotted lines the object motion. *Figure 11* shows the trajectories and *Figure 12* shows the respective y-coordinates.

It can be seen how the obtained result is quite good, taken into account that the object was moving at about 80 pixels/second (most of the time the velocity was constant).

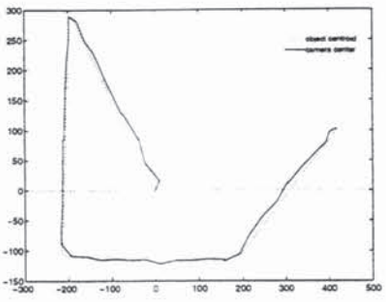

Fig. 11. Trajectories for a real test of the tracking algorithm.

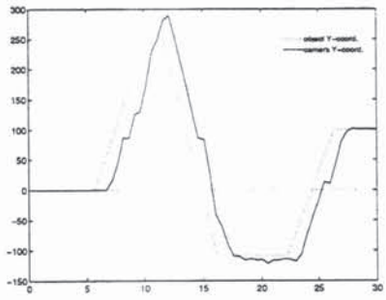

Fig. 12. Y-coordinates corresponding to the above trajectories.

5. CONCLUSIONS

In this article a method to estimate the optical flow based on the accumulated difference technique has been presented. The proposed method has been tested by simulation and experimentation to track an object in real time, using a PUMA 560 robot with a camera in its final effector.

6. REFERENCES

Allen P.K., A. Timcenko, B. Yoshimi (1993). Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Trans. on Robotics and Automation* **Vol.9**, pp.152-165.

Ballard D.H., C.M. Brown (1982). *Computer Vision.* Prentice-Hall. Englewood Cliffs, N.J.

Fu K.S., R.C. González, C.S.G. Lee (1986). *Robotics: Control, Sensing, Vision and Intelligence.* McGraw-Hill.

Hashimoto (1993). *Visual Servoing.* World Scientific.

Martin W.N., J.K. Aggarwal (1988). *Motion Understanding.* KAP (Kluwer Academic Publishers).

Papanikolopoulos N.P., P.K. Khosla (1993). Adaptive robotic visual tracking: Theory and experiments. *IEEE Transactions on Automatic Control* **Vol.38**, pp.429-445.

Vargas, M. (1997). Binarización optima de imágenes basada en histograma. *Internal Report,GAR 1997/02.*