MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Fast Visual Detection of Changes in 3D Motion

Antonis A. Argyros*, Manolis I.A. Lourakis*,
Panos E. Trahanias* and Stelios C. Orphanoudakis*
Institute of Computer Science, Foundation for Research and Technology - Hellas
and
Department of Computer Science, University of Crete

Abstract

A method is proposed for the fast detection of objects that maneuver in the visual field of a monocular observer. Such cases are common in natural environments where the 3D motion parameters of certain objects (e.g. animals) change considerably over time. The approach taken conforms with the theory of *purposive vision*, according to which vision algorithms should solve many, specific problems under loose assumptions. The method can effectively answer two important questions: (a) whether the observer has changed his 3D motion parameters, and (b) in case that the observer has constant 3D motion, whether there are any maneuvering objects (objects with non-constant 3D motion parameters) in his visual field. Essentially, the method relies on a pointwise comparison of two normal flow fields which can be robustly computed from three successive frames. Thus, it by-passes the ill-posed problem of optical flow computation. Experimental results demonstrate the effectiveness and robustness of the proposed scheme. Moreover, the computational requirements of the method are extremely low, making it a likely candidate for real-time implementation.

1 Introduction

Most of the research efforts to date in computational vision are influenced by the so called *reconstructionist approach*. Their basic assumption is that the general goal of computer vision is to produce an accurate, quantitative 3D representation of a scene. During the last decade, a new theory of vision has emerged, that of *active* and *purposive vision* [1]. According to the purposive theory, a vision system should be implemented by a set of processes which cooperate for achieving specific goals. Each process is dedicated to understanding certain aspects of the environment and, therefore, the statement and the solution of simpler problems is enabled.

This paper studies one such visual process, namely the visual detection of maneuvering objects. This is related to the problem of independent motion detection by a moving observer. The problem of independent 3D motion detection can be defined as the problem of locating objects that move independently from the observer in his visual field. Recently, this has been approached mainly by assuming some knowledge about the observer's motion. Thomson uses knowledge of certain aspects of egomotion and scene structure [2]; Sharma and Aloimonos [3] assume translational egomotion; Nelson [4] requires some *a priori* knowledge of egomotion parameters and assumes upper bounds on the depth of the scene. In some recent approaches, the problem is tackled without restrictive assumptions regarding the observer or the environment [5, 6], but is formulated for a stereoscopic observer.

Although the general problem of independent 3D motion detection is difficult, we argue that important aspects of it, such as the detection of maneuvering objects, can be solved robustly by simple algorithmic techniques. Based on the principles of purposive vision, this is approached in this paper by employing only an adequate representation of visual motion, rather than trying to fully recover motion information. More precisely, we extract from image sequences the minimum information needed for the detection of maneuvering objects. The developed scheme does not rely on the computation of *optical flow*, rather on the spatiotemporal derivatives of the image intensity function, known as *normal flow*. The latter, although less informative compared to optical flow, can be robustly computed from a sequence of images. Based on the choice of normal flow to represent visual motion information, a method is proposed for the detection of maneuvering objects. The method performs a pointwise comparison of the two normal flow fields that result from three successive image frames; this comparison signals changes in the 3D motion parameters of the observer or of the objects in the field of view.

The problem of detection of maneuvering objects has also been approached in [4] by assuming smooth observer motion and inexact knowledge of the motion field. In our approach the observer's motion is not restricted and, moreover, changes in his 3D motion parameters are signaled. In addition, no assumptions about the objects' motion are imposed,

*Address: PO Box 1385, Heraklion, Crete 711-10, Greece, E-mail {argyros, lourakis, trahania, orphanou}@ics.forth.gr

making it useful in practical applications where the detection of motion changes is desired.

The rest of the paper is organized as follows. Section 2 presents the imaging geometry and the motion representation employed in this work. In section 3, the method for the detection of maneuvering objects is presented. In section 4, experimental results from the application of the maneuvering objects detection algorithm to image sequences are presented. Finally, section 5 presents concluding remarks that summarize the results of this work.

2 Motion representation

Let a coordinate system $OXYZ$ adjusted to the optical center of a camera, such that the OZ axis coincides with the optical axis, as shown in Fig. 1. Under perspective projection, a point $P(X,Y,Z)$

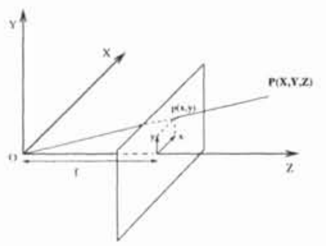

Figure 1: The camera coordinate system.

in 3D space projects on the image plane at point $p(x,y)$. If P is moving relative to $OXYZ$ with translational motion $\vec{t}=(U,V,W)$ and rotational motion $\vec{\omega}=(\alpha,\beta,\gamma)$[1], the equations describing the 2D velocity (u,v) of the image point $p(x,y)$ are written as [7]:

$$\begin{aligned} u &= \frac{(-Uf+xW)}{Z}+\alpha\frac{xy}{f}-\beta\left(\frac{x^2}{f}+f\right)+\gamma y \\ v &= \frac{(-Vf+yW)}{Z}+\alpha\left(\frac{y^2}{f}+f\right)-\beta\frac{xy}{f}-\gamma x \end{aligned} \tag{1}$$

where f is the camera focal length.

Equations (1) describe the 2D *motion field*, which relates the 3D motion of a point with its projected 2D motion on the image plane. The motion field is a purely geometrical concept and is not necessarily identical to the *optical flow* field [8], which describes the apparent motion of brightness patterns observed because of the relative motion between an imaging system and its environment. Verri and Poggio [9] have shown that the motion and optical flow fields are identical in specific cases only. Even in the cases that these two fields are identical, the problem of optical flow estimation is ill-posed [10]. This is often approached using regularization methods, which impose constraints on the solution. Such constraints are related to certain assumptions about the structure of the viewed scene. In practice - especially in the case of independent motion where motion discontinuities exist by definition - these assumptions are quite often violated, resulting in errors in optical flow estimation. For the above reasons, the proposed scheme for the detection of maneuvering objects does not rely on the computation of optical flow, rather on normal flow, i.e. the projection of optical flow along the direction of intensity gradient.

[1] This 3D motion may be due to a motion of the coordinate system (egomotion) and/or independent motion of point P.

Let now the image sequence be modeled as a continuous function $I(x,y,t)$ of two spatial (x,y) and one temporal (t) variables. Assuming that irradiance is conserved between two consecutive frames, the well known *optical flow constraint equation* [11] is formulated as a dot product:

$$(I_x, I_y)\cdot(u,v) = -I_t \tag{2}$$

where, I_x, I_y and I_t are the spatial and temporal partial derivatives of the image intensity function, respectively. Equation (2) gives a constraint for the components u and v of optical flow and enables the computation of the projection of the optical flow along the intensity gradient direction, namely the normal flow.

The normal flow field is not necessarily identical to the normal motion field (the projection of the motion field along the gradient), in the same way that the optical flow is not necessarily identical to the motion field [9]. However, normal flow is a good approximation to normal motion at points where the image gradient magnitude $||\nabla I||$ is large. Such points provide reliable information for motion perception.

Let (n_x, n_y) be the unit vector in the gradient direction. The magnitude u_n of the normal flow vector is given by $u_n = un_x + vn_y$ which, by substitution from eqs. (1), yields:

$$\begin{aligned} u_n &= -n_x f\frac{U}{Z} - n_y f\frac{V}{Z} + (xn_x + yn_y)\frac{W}{Z} \\ &+ \left\{\frac{xy}{f}n_x + \left(\frac{y^2}{f}+f\right)n_y\right\}\alpha \\ &- \left\{\left(\frac{x^2}{f}+f\right)n_x + \frac{xy}{f}n_y\right\}\beta + (yn_x - xn_y)\gamma \end{aligned} \tag{3}$$

3 Detection of maneuvering objects

Suppose that the 3D motion parameters of $P(X,Y,Z)$ remain constant over three frames that are acquired at time instances $t-2$, $t-1$ and t. Suppose also that we compute a normal flow field, from frame $t-1$ to frame t. Then, the normal flow computed at point (x,y) with gradient direction (n_x,n_y), is given by eq. (3). Let us also compute the normal flow from frame $t-1$ to frame $t-2$. Because of the hypothesis of constant 3D motion, point P will again move from time $t-2$ to time $t-1$ with motion parameters (U,V,W) and (α,β,γ) or, equivalently, with parameters $(-U, -V, -W)$ and $(-\alpha, -\beta, -\gamma)$ from time $t-1$ to $t-2$. Therefore,

$$u_n{}^{(t-1)\to(t-2)} = -u_n{}^{(t-1)\to t} \tag{4}$$

Note that since both normal flow fields appearing in eq. (4) are computed with respect to time instant $t-1$, for a given point (x, y), the gradient direction (n_x, n_y) and the depth Z are the same for $u_n{}^{(t-1)\to(t-2)}$ and $u_n{}^{(t-1)\to t}$. Equation (4) provides a simple, yet effective criterion to check whether the 3D motion parameters of a point remain the same over three frames in time. Once the two normal flow fields are computed, then for each point the sum of the normal flow values should be equal to zero. A non-zero value signals a change in the 3D motion parameters of the corresponding point. In practical situations, the sum of normal flow values will not be zero due to errors in the computation of the time derivative. We may, however, require the absolute value of the sum to be small with respect to the sum of the absolute normal flow values, deriving the following criterion:

$$\frac{|u_n{}^{(t-1)\to(t-2)} + u_n{}^{(t-1)\to t}|}{|u_n{}^{(t-1)\to(t-2)}| + |u_n{}^{(t-1)\to t}|} < \delta_{un} \tag{5}$$

where δ_{un} is a threshold controlling the sensitivity to changes in motion in the three frames.

The satisfaction of criterion (5) over subsets of scene points, leads to four interesting cases; they are summarized below, where it is assumed that the majority of the scene points correspond to the static world. Let I_P be the set of image points for which reliable normal flow values have been computed; then:

1. **The criterion holds for all image points in I_P.** This is the case where neither the observer, nor any object(s) changed their motion parameters. Note that the change of motion parameters includes the case of previously static objects that have now started moving.
2. **The criterion holds for the majority of image points in I_P.** This is the case where the motion of the observer remained constant. Points where the criterion does not hold, are points of objects that changed their motion.
3. **The criterion holds for the minority of image points in I_P.** This is a special case where both the observer and the independently moving object(s) changed their motion in exactly the same way, so that no relative change can be detected.
4. **The criterion does not hold for any point in I_P.** The motion of the observer has been changed. It cannot be decided, however, whether some objects have also changed their motion.

Based on whether criterion (5) is satisfied or not at a certain point, a label may be assigned to that point, which describes whether its 3D motion parameters have been changed or not.

It is noted that by employing normal flows, only incomplete information about motion is used. A normal flow value is the projection of an optical flow vector at a certain direction. Infinite many other optical flow vectors may have the same projection in the same direction. Consequently, there are certain changes in the 3D motion parameters of a point that cannot be recovered through summations of normal flow values. However, in a region where 3D motion changed, it is expected that many different gradient directions exist and, therefore, the concentration of points that do not satisfy criterion (5) will be high. This observation, leads to the conclusion that some type of post processing is needed. Such a postprocessing is achieved through a simple majority voting scheme. The label of a point is changed to the label of the majority of the points in a small neighborhood. This allows isolated points to be removed.

4 Experimental results

A set of experiments has been conducted in order to test the performance of the described method. Representative results from these experiments are given in this section. These results refer to the the "coca-cola" sequence which has been appropriately modified with the addition of rotational motion. Fig. 2 shows one frame of this sequence. The

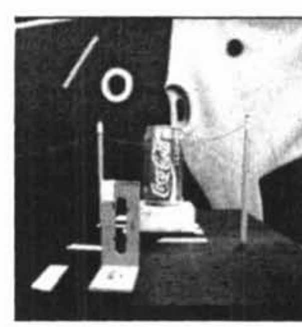

Figure 2: One frame of the "coca-cola" sequence.

camera moves with translational motion approaching the scene. Since there is no independent motion in the scene in view, a rotational motion has been synthetically added in the area of the coca-cola can. More specifically, in the third frame (frame at time t), the coca-cola can has been moved relative to the second frame by adding (synthetically) rotational motion (the observer's egomotion was left unchanged). Rotational motion has been employed because it does not interfere with the scene structure. After smoothing the images, the two normal flow fields were obtained and criterion (5) was computed for all image points with a reliable normal flow value. Figure 3(a) shows a three dimensional plot of the values of criterion (5). x and y dimensions of the plot correspond to the x and y dimensions of the image while the third dimension corresponds to

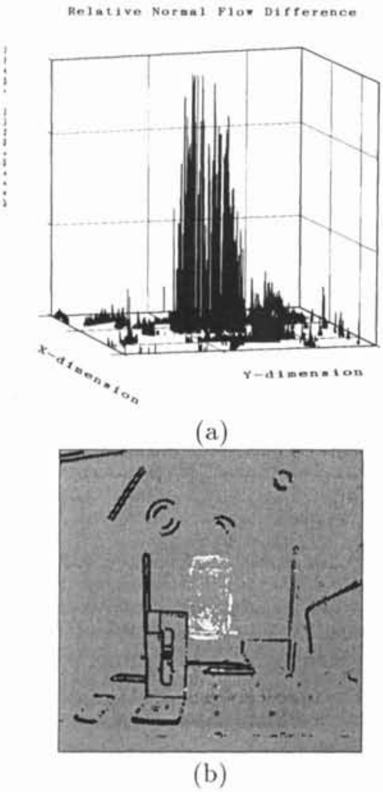

Figure 3: (a) 3D plot of the image points with respect to criterion (5), and (b) characterization of points with respect to the constancy of their 3D motion parameters (see text for explanation).

the values of criterion (5). It is evident that in the points of the coca-cola can where a motion change occurs, criterion (5) gives distinguishably different values than in the other points of the image which move due to the constant egomotion. Figure 3(b) shows the final labeling of the pixels. White pixels correspond to points where the 3D motion has been changed, black pixels correspond to points which kept the same 3D motion parameters and gray pixels correspond to points where no reliable normal flow vectors could be computed.

5 Summary

In this paper, a method for the visual detection of changes in the 3D motion parameters of objects has been described. The method is capable of answering two specific questions regarding a monocular observer and the scene being observed: (a) whether the observer moves with constant 3D motion parameters, and (b) whether some objects are maneuvering within the observer's visual field. Despite the high complexity of the general independent 3D motion detection problem, it has been shown that these two specific questions may be robustly answered by using a simple computational scheme. The method avoids the solution of the structure from motion problem and relies on the comparison of two normal flow fields that are computed from three successive image frames. Relying on normal flow and on its time-reversed computation, enables the method to also avoid the solution of the correspondence problem. The computational requirements of the proposed method are extremely low, facilitating real time implementation.

The experimental results presented serve as an indication of the effectiveness of the method, which answers important questions by employing minimum assumptions about the external world and the observer. Current research is targeted towards integrating the proposed method with other robust visual capabilities, in order to provide synergistic solutions to more complex vision problems.

References

[1] Y. Aloimonos and A. Bandopadhay. Active Vision. In *IEEE 1st International Conference on Computer Vision*, pages 35–54, June 1987.

[2] W.B. Thompson and T.C. Pong. Detecting Moving Objects. *International Journal of Computer Vision*, 4:39–57, 1990.

[3] R. Sharma and Y. Aloimonos. Early Detection of Independent Motion from Active Control of Normal Image Flow Patterns. *IEEE Transactions on SMC*, SMC-26(1):42–53, February 1996.

[4] R.C. Nelson. Qualitative Detection of Motion by a Moving Observer. *International Journal of Computer Vision*, 7(1):33–46, 1991.

[5] A.A. Argyros, M.I.A Lourakis, P.E. Trahanias, and S.C. Orphanoudakis. Independent 3D Motion Detection Through Robust Regression in Depth Layers. In *British Machine Vision Conference (BMVC '96), Edinburgh, UK*, September 9-12 1996.

[6] A.A. Argyros, M.I.A Lourakis, P.E. Trahanias, and S.C. Orphanoudakis. Qualitative Detection of 3D Motion Discontinuities. In *IROS '96, Tokyo, Japan*, November 4-8 1996.

[7] H.C. Longuet-Higgins and K. Prazdny. The Interpretation of a Moving Retinal Image. In *Proceedings of the Royal Society*, pages 385–397. London B, 1980.

[8] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[9] A. Verri and T. Poggio. Motion Field and Optical Flow: Qualitative Properties. *IEEE Transactions on PAMI*, PAMI-11(5):490–498, May 1989.

[10] Y. Aloimonos, I. Weiss, and A. Bandopadhay. Active Vision. *International Journal of Computer Vision*, 2:333–356, 1988.

[11] B.K.P. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.