Independent 3D Motion Detection Using Residual Parallax Normal Flow Fields

Manolis I.A. Lourakis†‡, Antonis A. Argyros† and Stelios C. Orphanoudakis †‡
{lourakis,argyros,orphanou}@ics.forth.gr

†Institute of Computer Science
Foundation for Research and Technology – Hellas
P.O.Box 1385, Heraklion, 711 10 Crete, Greece

‡Department of Computer Science
University of Crete
P.O.Box 1470, Heraklion, 714 09 Crete, Greece

Abstract

This paper considers a specific problem of visual perception of motion, namely the problem of visual detection of independent 3D motion. Most of the existing techniques for solving this problem rely on restrictive assumptions about the environment, the observer's motion, or both. Moreover, they are based on the computation of a dense optical flow field, which amounts to solving the ill-posed correspondence problem. In this work, independent motion detection is formulated as a problem of robust parameter estimation applied to the visual input acquired by a rigidly moving observer. The proposed method automatically selects a planar surface in the scene and the residual planar parallax normal flow field with respect to the motion of this surface is computed at two successive time instants. The two resulting normal flow fields are then combined in a linear model. The parameters of this model are related to the parameters of self-motion (egomotion) and their robust estimation leads to a segmentation of the scene based on 3D motion. The method avoids a complete solution to the correspondence problem by selectively matching subsets of image points and by employing normal flow fields. Experimental results demonstrate the effectiveness of the proposed method in detecting independent motion in scenes with large depth variations and unrestricted observer motion.

1 Introduction and Previous Work

The visual perception of motion has been the subject of many research efforts due to its fundamental importance for many visually assisted tasks. Independent 3D motion detection (IMD) is an important motion perception capability of a mobile seeing system. The problem of IMD is particularly challenging, since the motion of the observer causes every object in his field of view to appear moving in a manner dependent on its relative motion with regard to the observer and the structure of the viewed scene.

IMD has often been treated as a problem of segmenting the 2D motion field that is computed from a temporal sequence of images. A typical example of this approach appears in [12]. The basic problem of such methods is that they assume scenes where depth variations are small compared to the distance from the observer. However, in real scenes depth variations can be large and, therefore, 2D methods may detect discontinuities that are not only due to motion, but also due to the structure of the scene.

Solutions to the problem of IMD have also been provided using 3D models. The employment of 3D models makes the problem more difficult because extra variables regarding the depths of scene points are introduced. This in turn requires certain assumptions to be made in order to provide additional constraints for the problem. Most of the methods depend on the accurate computation of a dense optical flow field or on the computation of a sparse map of feature correspondences [18]. Other assumptions that are commonly made by existing methods are related to the motion of the observer, to the structure of the viewed scene, or both. In [14] for example, the IMD problem for an observer pursuing restricted translational motion is considered. Adiv [1] performs segmentation by assuming planar surfaces undergoing rigid motion, thus introducing an environmental assumption. Various principles for IMD, when certain aspects of the egomotion or of the scene structure are known, are derived in [16]. However, the practical exploitation of the underlying principles is limited because of the assumptions they are based on and other open implementation issues. Argyros et al [2, 3, 4] have proposed three methods that combine depth and motion information extracted by a binocular observer. Although all three methods avoid any assumptions related to the egomotion or the scene structure and do not require the correspondence problem to be solved, their main disadvantage is that they assume that normal flow can be computed from a pair of stereo images, an assumption that is valid in special cases only.

In order to overcome the limitations of existing methods, this paper proposes a new method for IMD. This method is based on two key observations. The first is that, although an accurate solution to the correspondence problem in the general case is very difficult, the problem can be solved with satisfactory accuracy in special cases, such as those involving corners or points belonging to a planar surface. The second observation is that the *residual parallax field* that remains after the registration of the images of a planar surface in two frames is an epipolar field. The proposed method exploits the information contained in the *normal residual field*, the component of residual motion in the direction of the image gradient. This field is less informative compared to the full residual flow, but can be more accurately computed from a temporal sequence of images. The combination of two such residual normal flow fields allows the elimination of the depth variables from the 3D motion equations, which

*This work was funded in part under the VIRGO research network of the TMR Programme (EC Contract No ERBFMRX-CT96-0049).

in turn leads to the derivation of a model that is linear in the 3D motion parameters. IMD is then handled by applying a robust estimator to solve for the parameters of the linear model. Points that conform to the estimated model are labeled as moving due to the motion of the observer, while points that are characterized as outliers during the estimation process are labeled as independently moving. The proposed method assumes an observer that moves rigidly with unrestricted translational and rotational egomotion. Independent motion can be rigid or non-rigid and no calibration information is necessary.

The rest of this paper is organized as follows. Section 2 presents the input used by the proposed method and issues related to robust regression, which constitutes a basic building block of the proposed approach. Section 3 details a technique for identifying the dominant planar surface in a scene. The estimation of the motion of the dominant plane is outlined in Section 4. Section 5 discusses the decomposition of rigid image motion into the motion of a planar surface and a residual parallax field. The proposed method for IMD is presented in Section 6. Experimental results from the application of the method on real-world image sequences are presented in Section 7 and the paper is concluded in Section 8. A more detailed version of the present paper can be found in [10].

2 Preliminaries

2.1 Visual Motion Representation

Consider a coordinate system $OXYZ$ at the optical center (nodal point) of a pinhole camera, such that the axis OZ coincides with the optical axis. Assuming that the camera is moving rigidly with respect to its 3D static environment with translational motion (U, V, W) and rotational motion (α, β, γ), the equations relating the 2D velocity (u, v) of an image point $p(x, y)$ to the 3D velocity of the projected 3D point $P(X, Y, Z)$ under perspective projection are [6]:

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

Equations (1) describe a 2D motion vector field, which relates the 3D motion of points to their 2D projected motion on the image plane. The motion field is a purely geometrical concept and it is not necessarily identical to the optical flow field [6], which describes the motion of brightness patterns observed because of the relative motion between the imaging system and the viewed scene. Even in the cases that these two fields are identical, the computation of the optical flow field requires some form of smoothness conditions to be satisfied in the neighborhoods of points for a unique solution to exist. Such assumptions, however, are not always satisfied because of depth discontinuities, illumination changes, independent 3D motion, etc.

For the above reason, the proposed IMD method does not rely on the computation of dense optical flow, but rather on the combination of the optical flow of a planar surface and the *normal flow* field for the whole image. As it will be shown in Section 4, once a planar surface in the scene has been identified, the problem of estimating its optical flow is a well-posed problem. On the other hand, the normal flow field is the projection of the optical flow field in the direction of image gradients and can be directly computed from the spatiotemporal derivatives of image intensity. It can be shown that the normal flow field is not necessarily identical to the *normal motion field* (the projection of the motion field along the image gradient), in the same way that the optical flow field is not necessarily identical to the motion field [17]. However, normal flow is a good approximation to normal motion at points where the image gradient has a large magnitude [17]. Normal flow vectors at such points can be used as a robust input to 3D motion perception algorithms.

2.2 Robust Regression

Regression analysis, i.e. the problem of fitting a model to noisy data, is a very important subfield of statistics. The traditional approach to regression analysis employs the least squares (LS) method, which is popular due to its low computational complexity. LS involves the solution of a linear minimization problem, and achieves optimal performance if the underlying noise distribution is Gaussian with zero mean. However, in cases where the noise is not Gaussian, or in the presence of *outliers*, that is observations that deviate considerably from the model representing the rest of the observations, the LS estimator becomes highly unreliable. One criterion for characterizing the tolerance of an estimator with respect to outliers is its *breakdown point*, which may be defined as the smallest amount of outlier contamination that may force the value of the estimate outside an arbitrary range. As an example, LS has a breakdown point of 0%, because a single outlier may have a substantial impact on the estimated parameters.

The *Least Median of Squares* (LMedS) estimator was originally proposed by Rousseeuw [13] and is able to handle data sets containing many outliers. LMedS involves the solution of a nonlinear minimization problem that aims at estimating a set of model parameters that best fit the *majority* of the observations. In contrast, LS tries to estimate a set of model parameters that best fit *all* the observations. Thus, LMedS has a breakdown point of 50%, a characteristic which makes it particularly attractive for the purposes of this work.

3 Dominant Plane Extraction

The traditional approach for identifying planar regions using two images of a scene has been to recover the depth of each point in the field of view and then segment the resulting depth map into planes. This process however, involves computations that are numerically unstable and requires difficult problems, such as point correspondence and camera calibration, to be solved. Fortunately, simpler methods that exploit results from projective geometry have been proposed [11]. A well-known fact is that groups of five corresponding coplanar points give rise to two projective invariants. Another important concept is the *plane homography* (also known as plane projectivity or plane collineation) $\mathbf{H}$, which relates two uncalibrated views of a plane in three dimensions. Each 3D plane Π defines a nonsingular 3×3 matrix $\mathbf{H}$ which relates the image of the plane in two views. More specifically, if $\mathbf{m}$ is the projection in one view of a point belonging to Π and $\mathbf{m}'$ is the corresponding projection in a second view, then $\mathbf{m}' = \mathbf{H}\mathbf{m}$. The plane homography can be estimated

from four pairs of corresponding coplanar points in general position (no three points collinear).

Based on the above results, Sinclair and Blake [15] have proposed an iterative method for identifying coplanar sets of corresponding points. Briefly, this method uses as input a set of matched corners, extracted from a pair of images that have been acquired from considerably different viewpoints in the 3D space. Such an image pair can be captured either by the two cameras of a binocular system, or by the single camera of a monocular system at two instants that are far apart in time. Owing to the significant disparities defined by such images, accurate 3D structure information for the viewed scene can be recovered. Initially, a random sample consisting of five pairs from the set of matched corners is formed. If the selected corners satisfy the plane invariants, they are likely to belong to the same plane. Next, the plane homography corresponding to the selected corners is estimated. To verify that the five selected points lie on the same plane, the estimated plane homography is used to find more coplanar points. For every corner in one image, the plane homography can predict the location of the corresponding corner in the second image. If this location is sufficiently close to the true location of the matching corner, the corner in question is assumed to be coplanar with the corners in the selected sample. If the number of coplanar points identified during this step is above a threshold, the method concludes that a plane has indeed been found. The corresponding plane homography is then re-estimated using the whole set of coplanar points and this set is removed from further consideration. The sampling process iterates until either the number of corners that have not been assigned to a plane drops below a threshold or a predetermined number of iterations is completed.

When the iterative algorithm terminates, a set of planes along with their homographies have been computed. Multiplication of each point in the first view with a homography matrix, warps the second view with respect to the first and registers the image of the corresponding plane in the two views. Change detection between the first and the warped second view can label image points as changing in the two views or not. Points that remain unchanged belong to the plane under consideration. To account for the fact that typical change detection algorithms fail in uniform, textureless areas, a pixel is assumed to belong to a plane when it is labeled as not changing by the change detection algorithm and the magnitude of its gradient is above some threshold. The plane having the largest number of points is declared to be the dominant one. It will become clear in the following sections that the result of change detection does not have to be very accurate, since the part of the proposed method for IMD that makes use of the location of the dominant plane is tolerant to errors.

4 Robust Parametric Estimation of Optical Flow

The problem of estimating 2D image velocity, or optical flow, from image sequences is generally very difficult. This difficulty mainly stems from the fact that transparencies, specular reflections, shadows, occlusions, depth boundaries and independent motions give rise to discontinuities in the optical flow field. This in turn implies that an optical flow field is typically only piecewise smooth [5]. Since the estimation of optical flow involves the combination of constraints arising from an image region, no guarantee is given that the selected region will contain only a single motion. In other words, the primary difficulty of most optical flow estimation techniques is that they lack any information regarding the region of support of a particular motion. This problem is referred to in [5] as the *generalized aperture problem*.

In the case that an image region is known to correspond to a plane in the scene, the optical flow within the region can be accurately modeled as a function of eight parameters and image coordinates [1]. This is known as the *quadratic* model for optical flow, since it contains terms that are of degree two in the image coordinates. At this point, it should be noted that, if the camera is not calibrated, the unknown intrinsic parameters (i.e. focal length and location of principal point) are absorbed in the eight flow parameters. By employing the quadratic model, the estimation of optical flow amounts to the estimation of the eight parameters involved. The combination of the quadratic flow model with the *optical flow constraint equation* [6], permits the derivation of an equation relating the eight planar flow parameters to the spatiotemporal intensity derivatives. This equation is linear in the parameters to be estimated and is overdetermined, since each point of the plane contributes one constraint regarding the eight unknown parameters. To account for errors in the computation of derivatives, violations of the intensity conservation assumption, errors in the determination of the region corresponding to the image of the plane, etc, the LMedS estimator is again employed to give a robust estimate of the parameters satisfying the majority of the constraints. This "robustification" of the optical flow estimation problem has already been suggested by Black and Anandan [5], the major difference being that they employed M-estimators which are less robust compared to LMedS that is employed here.

5 Planar Parallax

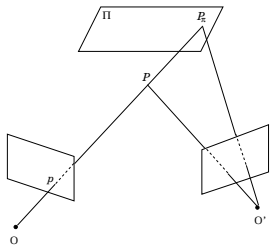

Figure 1: Planar parallax.

Most motion analysis methods express rigid image motion as the sum of two displacement fields, namely a translational and a rotational one. Recently, however, it has been shown that if image motion is expressed in terms of the motion of a parametric surface and a *residual parallax field*, important problems in motion analysis become considerably simpler [9, 8]. In this section, the equations describing the residual field are derived, assuming that the employed parametric surface is a plane.

Let (u, v) be the displacement field between two images $\mathcal{I}_t$ and $\mathcal{I}_{t+dt}$, acquired at time instants t and $t + dt$ respectively. Let also Π be a 3D plane in the viewed scene and let (u_π, v_π) be the 2D motion vector of a single point belonging to Π. As mentioned in Section 4, (u_π, v_π) is defined by a linear model with eight parameters. Warping

$\mathcal{I}_t$ towards $\mathcal{I}_{t+dt}$ according to (u_π, v_π) will register $\mathcal{I}_t$ and $\mathcal{I}_{t+dt}$ over regions of Π, while regions not belonging to Π will be unregistered. According to Eq. (1), the residual flow (u^r, v^r) between the warped $\mathcal{I}_t$ and $\mathcal{I}_{t+dt}$ is [8]:

$$\begin{aligned} u^r &= u - u_\pi &= (xW - Uf)(1/Z - 1/Z_\pi) \\ v^r &= v - v_\pi &= (yW - Vf)(1/Z - 1/Z_\pi) \end{aligned} \quad (2)$$

where $1/Z_\pi$ is the depth of the 3D plane at pixel (x, y). As can be seen from Eq. (2), the residual flow field is purely translational. This is because the rotational component of the motion does not depend on depth and is thus canceled by the warping step. Consequently, all optical flow vectors of the residual flow point towards the FOE[1]. In a similar manner, the *residual normal flow field* between the warped $\mathcal{I}_t$ and $\mathcal{I}_{t+dt}$ is given by:

$$u_{nr} = \{(xW - Uf)n_x + (yW - Vf)n_y\}\,(1/Z - 1/Z_\pi) \quad (3)$$

where (n_x, n_y) is the unit vector in the direction of the intensity gradient.

Figure 1 depicts geometrically the notion of planar parallax. O and O' are the camera focal points in the two views, Π a plane in the viewed scene and P_π, P are two 3D points with P_π belonging to Π. Since P_π and P project to the same image point p in the first view, their corresponding optical flow vectors share the same rotational components.

6 Independent Motion Detection

Consider a rigid observer that is moving with unrestricted egomotion in 3D space. Due to this motion, a reliable normal flow vector can be computed at each point where the image intensity gradient is sufficiently large. 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 $u_n = un_x + vn_y$, and as can be seen by substitution from Eq. (1), each image point introduces one independent depth variable. Thus, n computed normal flow vectors and m 3D motions result in n constraints regarding $n+6m$ unknowns. Evidently, the problem cannot be solved without any additional information on depth.

Let us now suppose that at least one of the surfaces in the scene is planar or can be well approximated by a plane. This assumption is often satisfied in practice, especially in scenes containing man-made objects. Using the technique described in Section 3, the dominant plane in the scene can be extracted. Following this, the parametric model describing the motion of this plane can be estimated as described in Section 4. The residual planar parallax flow can then be computed from Eq. (2). Irani and Anandan [7] have recently described a method for IMD that computes the *relative projective 3D structure* from this residual parallax flow. Their method, however, requires the computation of a dense optical flow field, a difficult problem in its own right. Noting that the residual flow field is translational, another approach to detect independent motion is to locate the FOE and then, similar to [14], label points that violate the epipolar constraint as independently moving. The major drawback of this approach is that it depends critically on the correctness of the estimated FOE. To avoid this problem, the proposed method for IMD does not attempt to estimate the FOE. Instead, it combines the information from two residual normal flow fields computed at consecutive time instants.

Assume that three consecutive images $\mathcal{I}_{t-dt}$, $\mathcal{I}_t$ and $\mathcal{I}_{t+dt}$ are captured at time instants $t-dt$, t and $t+dt$ respectively. Let $\mathcal{I}_0$ be a fourth "distant" image that along with $\mathcal{I}_t$ permits the extraction of the dominant plane. Also, let u_{nr} be the residual normal flow computed by warping $\mathcal{I}_t$ towards $\mathcal{I}_{t+dt}$ using the motion of the dominant plane. Similarly, let u'_{nr} be the residual normal flow computed by warping $\mathcal{I}_t$ towards $\mathcal{I}_{t-dt}$ using the dominant plane. According to Eq. (2), u_{nr} and u'_{nr} are given by:

$$\begin{aligned} u_{nr} &= \{(xW - Uf)n_x + (yW - Vf)n_y\}\,1/Z_\delta \\ u'_{nr} &= \{(xW' - U'f)n_x + (yW' - V'f)n_y\}\,1/Z_\delta \end{aligned} \quad (4)$$

where (U, V, W) and (U', V', W') are the translational velocity vectors for the displacement between t and $t+dt$ and t and $t-dt$ respectively and $1/Z_\delta = 1/Z - 1/Z_\pi$.

Both residual normal flow fields given by Eqs. (4) are defined in the same reference frame, namely $\mathcal{I}_t$. This implies that at each point (x, y) of $\mathcal{I}_t$, having considerable gradient magnitude, two normal flow vectors along the same direction (n_x, n_y) can be computed. Solving the first of Eqs. (4) for $1/Z_\delta$ and substituting into the second results into the following equation

$$\begin{aligned} &W(xn_x + yn_y)u'_{nr} - Ufn_x u'_{nr} - Vfn_y u'_{nr} - \\ &W'(xn_x + yn_y)u_{nr} + U'fn_x u_{nr} + V'fn_y u_{nr} = 0 \end{aligned} \quad (5)$$

in which the terms related to depth have been eliminated. The above equation is linear in the variables $\phi_1 = W$, $\phi_2 = Uf$, $\phi_3 = Vf$, $\phi_4 = W'$, $\phi_5 = U'f$, $\phi_6 = V'f$. These variables involve the 3D motion parameters and the camera focal length. Assuming that the dominant plane is not independently moving, violations of Eq. (5) signal the presence of independently moving objects. LMedS estimation can be applied to a set of observations of the model of Eq. (5) as a means to estimate the parameters ϕ_i, $i = 1, \ldots, 6$. To avoid the trivial solution $\phi_i = 0$, the solutions tried by LMedS are computed with an eigenvector technique that imposes the constraint $||(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6)||^2 = 1$. LMedS will provide estimates $\hat{\phi}_i$ of the parameters ϕ_i and a segmentation of the image points into model inliers and model outliers. Model inliers correspond to image points that move with the dominant 3D motion. The set of outliers is the union of the set of points for which the quantities u_{nr} and/or u'_{nr} have been computed erroneously, with the set of points that move with 3D velocities different from the 3D velocity of the majority of the points. The points in the first set will be few and sparsely distributed over the image plane. This is because only reliable normal flow vectors are considered. The second set contains points that are incompatible with the dominant 3D motion parameters. Thus, in the case of two rigid motions in a scene, the inlier/outlier characterization of points achieved by LMedS is equivalent to a dominant/secondary 3D motion segmentation of the scene. In the case that more than two rigid motions are present in a scene, the correctness of 3D motion segmentation depends on the spatial extent of the 3D motions. If there is one dominant 3D motion[2], the

[1] The FOE is the point $(\frac{fU}{W}, \frac{fV}{W})$ on the image plane, which defines the direction of translation.

[2] Dominant in the sense that at least 50% of the points under consideration move with this motion.

high breakdown point of LMedS ensures that LMedS will handle the situation successfully. Inliers will correspond to the dominant motion (egomotion) and the outliers will coincide with all secondary (independent) motions. Recursive application of LMedS to the set of outliers may further discriminate the remaining motions.

When implementing the method presented in the preceding paragraphs, the residual normal flow can be computed without actually warping the first image towards the second according to the estimated planar flow. Knowledge of the eight parameters defining the planar flow enables the prediction of the normal flow that would result if the dominant plane covered the whole visual field. The residual normal flow can then simply be estimated as the difference between the normal flow computed directly from the pair of input images and the predicted planar normal flow.

6.1 Postprocessing

According to the proposed method for independent motion detection, points are characterized as being independently moving or not based on their conformance to a general rigid 3D model of egomotion. The characterization is made at the point level, without requiring any conditions to hold in the neighborhood of each point. In order to further exploit information regarding independent motion, it is often considered preferable to refer to connected, independently moving areas rather than to isolated points. There are three reasons why the points of a motion segment may not form connected regions. First, the normal flow field is usually a sparse field, because normal flow values are considered unreliable in certain cases (e.g. at points with a small gradient value). Second, there is always the possibility of errors in measurements of normal flow and, therefore, some points may become model inliers (or outliers) because of these errors and not due to their 3D motion parameters. Finally, normal flow is a projection of the optical flow onto a certain direction. Infinitely many other optical flow vectors have the same projection onto this direction. Consequently, a normal flow vector may be compatible with the parameters of two different 3D motions, and therefore a number of point misclassifications may arise.

We overcome the problem of disconnected motion segments by exploiting the fact that, in the above cases, misclassified points are sparsely distributed over the image plane. A simple majority voting scheme is used. At a first step, the number of inliers and outliers is computed in the neighborhood of each image point. The label of this point becomes the label of the majority in its neighborhood. This allows isolated points to be removed. In the resulting map, the label of the outliers is replicated in a small neighborhood in order to group points of the same category into connected regions.

7 Experimental Results

The proposed method has been evaluated experimentally with the aid of several real-world image sequences. During the course of all experiments, quantitative information regarding camera motion and calibration parameters was not available. Due to space limitations, only two of the conducted experiments are reported here.

The first experiment is based on the well known "calendar" image sequence. Frame 2 of this sequence is shown in Fig. 2(a).

In this sequence, the camera is panning in a right to left direction and the viewed scene consists of a planar background and a nonplanar foreground. The background contains a stationary wall and a calendar that is independently moving upwards. The foreground contains three independently moving objects. A pair of spheres is rotating on the left side of the scene, while a ball followed by a toy train are moving in a right to left direction. The dominant plane was extracted using frames 2 and 30.

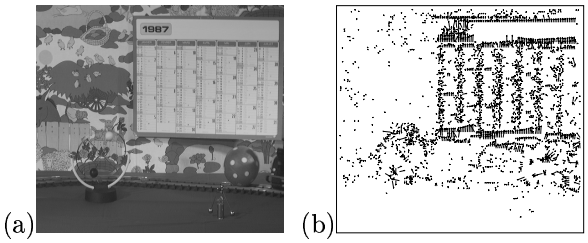

Figure 2: (a) Frame 2 of the "calendar" sequence, (b) residual normal flow field for frames 2-3.

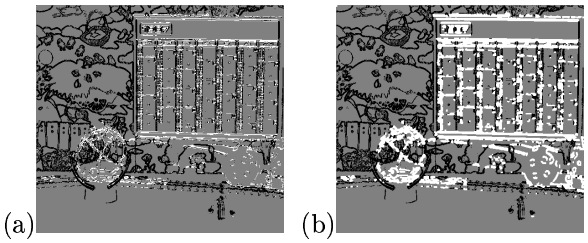

Figure 3: Motion segmentation for the "calendar" sequence (a) before and, (b) after postprocessing.

The pair of residual parallax normal flow fields was computed between frames 2 - 3 and 2 - 1. The residual parallax normal flow for frames 2 - 3 is shown in Figure 2(b). As can be seen from this figure, the residual flow field is zero over the area corresponding to the dominant plane, indicating that the dominant plane has been successfully registered. Figure 3 illustrates the results of motion segmentation on the "calendar" sequence. Figure 3(a) shows the intermediate segmentation results. Black color corresponds to egomotion and white color corresponds to independent motion. Gray color corresponds to points where no decision can be made, due to low image gradient and, therefore, lack of normal flow vectors. It can be verified that the largest concentration of white (i.e. independently moving) points is indeed over the regions of the independently moving objects. Note that independent motion was not detected along the vertical edges of the calendar. This is because the intensity gradient is perpendicular to the direction of motion on these edges, which results in the corresponding normal flow vectors being equal to zero. The elongated areas below the calendar that are marked as independently moving are actually shadows, cast by the calendar and the rotating spheres, that are also moving. Figure 3(b) presents the same result after postprocessing, which eliminates isolated outliers (inliers) in large populations of inliers (outliers) and, in the resulting map, dilates the label of remaining outliers in a small neighborhood. It is clear that after this step, the bodies of the four independently moving objects have been successfully identified as such.

The second experiment concerns the "cars" image se-

quence. Frame 5 of this sequence is shown in Fig. 4.

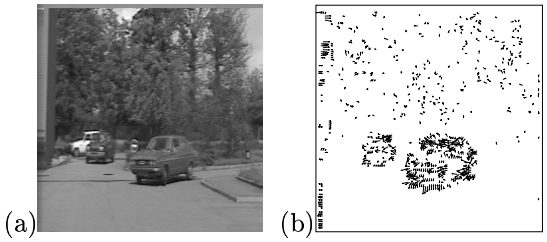

Figure 4: (a) Frame 5 of the "cars" sequence, (b) residual normal flow field for frames 5-6.

In this sequence, the camera is again panning in a right to left direction. The two dark gray cars in the foreground move independently while the white car on the far left is stationary. A few trees in the background form an approximately planar surface. Frames 5 and 20 were used to extract the dominant plane.

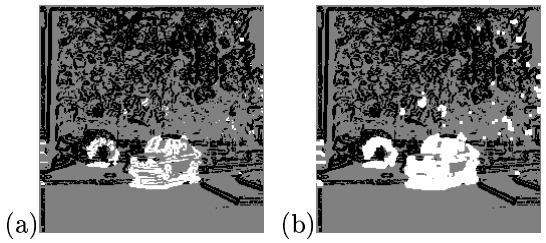

Figure 5: Motion segmentation for the "cars" sequence (a) before and, (b) after postprocessing.

Frames 5 - 6 and 5 - 4 are used to compute the pair of residual parallax normal flow fields. Figure 4(b) shows the residual parallax normal flow computed from frames 5 - 6. The results of motion segmentation on the "cars" sequence before and after postprocessing are illustrated in Figures 5(a) and 5(b) respectively. Black color corresponds to egomotion and white color corresponds to independent motion. Gray color corresponds to points with low intensity gradient, and thus without normal flow vectors. As it can be seen from Fig. 5, the two cars are correctly identified as independently moving. Moreover, the independent motions of small parts of the tree foliage are also detected.

8 Summary and Conclusions

Artificial seeing systems should operate in dynamic environments that consist of both stationary as well as moving objects. The perception of independent 3D motion is crucial because it provides useful information regarding dynamic changes in the environment, indicating areas where attention should be focused and, possibly, maintained. In this paper, independent 3D motion detection was based on a pair of residual parallax normal flow fields that are computed by an observer that moves freely in 3D space. The proposed method employs 3D motion models and is able to perform satisfactorily even in scenes with considerable depth variations. Both rigid and non-rigid independent motion can be detected. Moreover, apart from the requirement for the existence of a planar surface in the viewed scene, no further assumptions regarding the structure of the external world are made. The method avoids a complete solution to the ill-posed correspondence problem by matching only carefully selected sets of image points. To guard against errors caused by false matches, robust estimation techniques are employed. Experimental results from the application of the proposed method on real image sequences were also presented. Future research will address the problem of estimating the 3D egomotion of the observer, based on the motion segmentation results provided by the proposed method.

References

[1] G. Adiv. Determining Three Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects. *IEEE Trans. on PAMI*, 7(4):384–401, July 1985.

[2] A.A. Argyros et al. Independent 3D Motion Detection Through Robust Regression in Depth Layers. In *Proceedings of BMVC '96*, 1996.

[3] A.A. Argyros et al. Qualitative Detection of 3D Motion Discontinuities. In *Proceedings of IROS '96*, 1996.

[4] A.A. Argyros and S.C. Orphanoudakis. Independent 3D Motion Detection Based on Depth Elimination in Normal Flow Fields. In *Proceedings of CVPR '97*, pages 672–677, 1997.

[5] M.J. Black and P. Anandan. The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *CVIU*, 63(1):75–104, 1996.

[6] B.K.P. Horn. *Robot Vision*. MIT Press, 1986.

[7] M. Irani and P. Anandan. A Unified Approach to Moving Object Detection in 2D and 3D Scenes. In *Proceedings of ICPR'96*, pages 712–717, 1996.

[8] M. Irani, B. Rousso, and S. Peleg. Recovery of Ego-Motion Using Region Alignment. *IEEE Trans. on PAMI*, 19(3):268–272, Mar. 1997.

[9] R. Kumar, P. Anandan, and K. Hanna. Direct Recovery of Shape from Multiple Views: A Parallax Based Approach. In *Proceedings of ICPR'94*, pages 685–688, 1994.

[10] M.I.A. Lourakis, A.A. Argyros, and S.C. Orphanoudakis. Independent 3D Motion Detection Using Residual Parallax Normal Flow Fields. Technical Report 206, ICS/FORTH, Aug. 1997.

[11] J.L. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[12] P. Nordlund and T. Uhlin. Closing the Loop: Detection and Pursuit of a Moving Object by a Moving Observer. *IVC*, 14:267–275, 1996.

[13] P.J. Rousseeuw. Least Median of Squares Regression. *J. of Am. Stat. Assoc.*, 79:871–880, 1984.

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

[15] D. Sinclair and A. Blake. Quantitative Planar Region Detection. *IJCV*, 18(1):77–91, Apr. 1996.

[16] W.B. Thompson and T.C. Pong. Detecting Moving Objects. *IJCV*, 4:39–57, 1990.

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

[18] W. Wang and J. H. Duncan. Recovering the Three-Dimensional Motion and Structure of Multiple Moving Objects from Binocular Image Flows. *CVIU*, 63(3):430–440, May 1996.