In Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'96), pp. 1630-1637, Osaka, Japan, Nov. 4-8, 1996.

Qualitative Detection of 3D Motion Discontinuities

Antonis A. Argyros, Manolis I.A. Lourakis, Panos E. Trahanias and Stelios C. Orphanoudakis

Institute of Computer Science
Foundation for Research and Technology - Hellas
PO Box 1385, Heraklion, Crete 711-10, Greece
and
Department of Computer Science
University of Crete
PO Box 1470, Heraklion, Crete 714-09, Greece

Abstract

This paper presents a method for the detection of objects that move independently of the observer in a 3D dynamic scene. Independent motion detection is achieved through processing of stereoscopic image sequences acquired by a binocular, rigidly moving observer. A weak assumption is made about the observer's motion (egomotion), namely that the direction of the translational and rotational components of egomotion are constant in small image patches. This assumption facilitates the extraction of qualitative information about depth from motion, while additional qualitative depth information is independently computed from image stereo pairs acquired by the binocular vision system. Robust regression in the form of Least Median of Squares estimation is applied within each image patch to test for consistency between the depth functions computed from motion and stereo. Possible inconsistencies signal the presence of independently moving objects. In contrast to other existing approaches for independent motion detection, which are based on the ill-posed problem of optical flow computation, the proposed method relies on normal flow fields for both stereo and motion processing. By exploiting local constraints of qualitative nature, the problem of independent motion detection is approached directly, without relying on a solution to the general structure from motion problem. Experimental results indicate that the proposed method is both effective and robust.

1 Introduction

Most of the primitive survival tasks of biological organisms are based on the perception of motion: a moving object is likely to be either prey to be caught, or an enemy to be avoided. Thus, the perception of motion is crucial for many behaviors that an autonomous biological or man-made system should exhibit in real world, dynamic environments. Autonomous robotic systems should move relative to their environment in order to accomplish their goals. Due to their *egomotion* in 3D space, the visual field appears to be moving in a specific manner, which depends on their 3D motion parameters and the structure of the viewed scene. The problem of independent 3D motion detection can be defined as the problem of locating the objects that move independently of an observer in his field of view.

The importance of independent 3D motion detection has been recognized for years and a lot of work has been done along this research direction. Most of this work depends on the accurate computation of the optical flow field. Moreover, certain limiting assumptions are usually made about the observer's motion. Jain [1] has considered the problem of independent 3D motion detection by an observer pursuing translational motion. In addition to imposing constraints on egomotion (the observer's motion should not have rotational components), knowledge of the direction of translation is required. Thompson and Pong [2] derive various principles for detecting independent motion when certain aspects of the egomotion or of the scene structure are known. However, the practical exploitation of these principles is made difficult by the limiting assumptions they are based on and other open practical issues. Bouthemy and Francois [3] view motion segmentation as a problem of statistical regularization using Markov Random Field models. Sharma and Aloimonos [4] have proposed a method which uses the spatiotemporal derivatives of the image intensity function (*normal flow field*), rather than optical flow. However, as in the case of [1], translational egomotion is hypothesized. Nelson [5] presents two methods for independent motion detection which are also based on the normal flow field. The first of these methods requires *a priori* knowledge of egomotion parameters and assumes upper bounds on the depth of the scene. The second method detects abrupt changes of independent motion rather than independent motion itself.

In this paper, independent 3D motion detection is formulated as a robust regression problem. Robust regression has also been employed in the past for motion segmentation. Ayer et al [6] has proposed a method which uses robust regression to identify independently moving objects. However, since no infor-

mation on scene structure is used, the applicability of this method is limited to scenes forming a *frontoparallel plane*.

The method proposed in this paper detects 3D motion discontinuities using qualitative depth information. Specifically, two qualitative functions of depth are computed patchwise, one from motion and one from stereo. If, for a certain image patch, the depth functions from stereo and motion are consistent, it turns out that only one rigid 3D motion is present in that patch. The depth functions within an image patch are considered consistent if there is a linear relation that can map one to the other. Inconsistencies in the compared depth functions can only be due to multiple (rigid or non-rigid) motions. The *Least Median of Squares (LMedS)* estimation technique is used to determine whether the two depth functions computed from stereo and motion are consistent or not.

In contrast to other approaches for motion segmentation that use optical flow [1, 2], the proposed method is based on normal flow. Moreover, the ill-posed correspondence problem is avoided for the case of stereo, which is treated as the hypothetical motion that would map the position of the left camera to the position of the right camera.

The rest of the paper is organized as follows. Section 2 describes the geometry of the imaging system, the input used by the independent motion detection method and gives a brief introduction to the Least Median of Squares estimator, which constitutes a basic building block for the independent motion detection method. The method itself is described in detail in section 3. In section 4, experimental results are presented and discussed. Finally, section 5 summarizes the proposed method, experimental results and conclusions.

2 Preliminaries

2.1 Visual motion representation

Let a camera-centered coordinate system $OXYZ$ be positioned at the nodal point of the camera, with the OZ axis coinciding with the optical axis. Assume that the camera is moving rigidly with respect to its 3D static environment with translational motion $\vec{t} = (U, V, W)$ and rotational motion $\vec{\omega} = (\alpha, \beta, \gamma)$, as shown in Fig. 1. Under perspective projection, the relation between the 2D velocity (u, v) of an image point $p(x, y)$ and the 3D velocity of the projected 3D point $P(X, Y, Z)$ are [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 focal length of the imaging system. What eqs. (1) describe is the 2D *motion field*, which relates the 3D motion of a point to 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

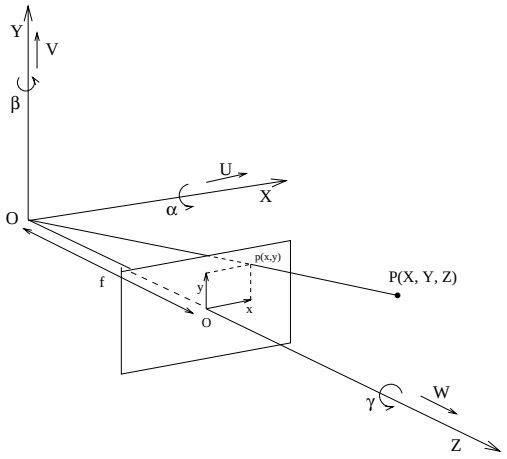


Figure 1: The camera coordinate system.

apparent motion of brightness patterns observed because of the relative motion between an imaging system and its environment. Even in the cases that these two fields are identical, the problem of optical flow estimation is ill-posed [9]. However, the projection of optical flow on the direction of the image intensity gradient, i.e. the normal flow, can be robustly computed using the well known *optical flow constraint equation*, originally developed by Horn and Schunk [10]:

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

where I_x, I_y and I_t are the spatial and temporal partial derivatives of the image intensity function respectively, and "$\cdot$" denotes the dot product. The normal flow is a good approximation to the normal motion field (the projection of the motion field along the image spatial gradient direction) at points with large spatial gradient magnitude [11]. Normal flow vectors at such points can be used as reliable input to 3D motion analysis. Therefore, the proposed method for independent 3D motion detection makes use of normal flow fields.

Let (n_x, n_y) be the unit vector in the gradient direction. The magnitude u_{nm} of the normal flow vector is given by

$$u_{nm} = un_x + vn_y \tag{3}$$

which, by substitution from eq. (1), yields

$$\begin{aligned} u_{nm} &= -n_x f\frac{U}{Z} - n_y f\frac{V}{Z} + (xn_x + yn_y)\frac{W}{Z} \\ &+ \left\{\frac{xy}{f}n_x + \left(\frac{y^2}{f} + f\right)n_y\right\}\alpha \\ &- \left\{\left(\frac{x^2}{f} + f\right)n_x + \frac{xy}{f}n_y\right\}\beta + (yn_x - xn_y)\gamma \end{aligned} \tag{4}$$

Equation (4) contains seven unknowns, a six-tuple $(U, V, W, \alpha, \beta, \gamma)$ of motion parameters and a depth

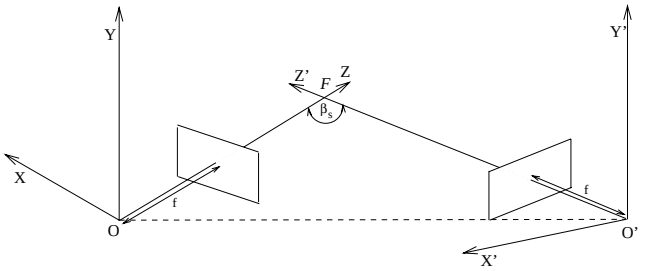


Figure 2: The geometry of a fixating stereo configuration.

variable Z. If only the observer is moving, the same set of 3D egomotion parameters $(U_E, V_E, W_E, \alpha_E, \beta_E, \gamma_E)$ holds true for all points. In the case, however, of independent motion, there is at least one additional set of motion parameters $(U_I, V_I, W_I, \alpha_I, \beta_I, \gamma_I)$ which is valid for some of the image points. Furthermore, if no assumptions regarding the depth Z are made, each point introduces an extra independent depth variable. Evidently, the system of equations that result by writing eq. (4) for all image points - in fact, for all points where a reliable normal flow value can be computed - cannot be solved if no additional information regarding depth is available.

2.1.1 Stereo configuration

Consider a typical stereo configuration of a fixating pair of cameras, as shown in Fig. 2. A pair of images captured with such a configuration encapsulates information relevant to depth, that manifests itself in the form of *disparities* defined by the displacements of points between images. Seen from the vantage point of camera motion, a stereo image pair can be considered as the sequence that would result from a hypothetical (ego)motion that brings one camera to the position of the other. This fact enables the analysis of a stereo pair employing motion analysis techniques. Specifically, two translational motions, U_s and W_s along the X and Y axes and a rotational motion β_s around the Y axis, suffice to describe the hypothetical motion. A normal flow value u_{ns} due to stereo can be computed at each point, which is equal to

$$\begin{aligned} u_{ns} &= -n_x f \frac{U_s}{Z} + (x n_x + y n_y) \frac{W_s}{Z} \\ &- \left\{ \left(\frac{x^2}{f} + f \right) n_x + \frac{xy}{f} n_y \right\} \beta_s \end{aligned} \quad (5)$$

In practice, the optical flow constraint equation on which the computation of normal flow is based, does not hold if the two images differ too much. Moreover, normal flow is computed from discrete images through spatial and temporal differentiation with small masks. In the case that 5×5 masks are used, every normal flow that is more than 3 - 4 pixels is not reliable. However, in areas close to the fixation point, normal flow can be computed robustly. In addition, a proper selection of image resolution can be made, so that a reliable estimate of normal flow can be obtained. The use of lower spatial resolutions facilitates the computation of normal flow due to stereoscopic configurations, at the cost of computing coarser depth information.

2.2 Robust regression

Regression analysis (fitting a model to noisy data) is a very important statistical tool. In the general case of a linear model [12], given by the relation

$$y_i = x_{i1}\theta_1 + \ldots + x_{ip}\theta_p + e_i, \quad (6)$$

the problem is to estimate the parameters θ_k, $k = 1, \ldots, p$, from the observations y_i, $i = 1, \ldots, n$, and the explanatory variables x_{ik}. The term e_i represents the error present in each of the observations. Let $\hat{\theta}$ be the vector of estimated parameters $\hat{\theta_1}, \ldots, \hat{\theta_p}$. Given these estimations, predictions can be made for the observations:

$$\hat{y_i} = x_{i1}\hat{\theta_1} + \ldots + x_{ip}\hat{\theta_p} \quad (7)$$

Thus, a residual between the observation and the value predicted by the model may be defined as $r_i = y_i - \hat{y_i}$. Traditionally, $\hat{\theta}$ is estimated by the popular least squares (LS) method. However, the LS estimator becomes highly unreliable in the presence of outliers, that is observations that deviate considerably from the model describing the rest of the observations. Robust regression methods [12] have been proposed in order to cope with such cases. The main characteristic of robust estimators is their high breakdown point, which is defined as the smallest amount of outlier contamination that may force the value of the estimate outside an arbitrary range. A variety of robust estimation techniques have been used in computer vision. Some of them have been developed within the vision field (eg. [13]), while others have been borrowed from statistics (eg. [14]).

The LMedS method, which was proposed by Rousseeuw [15], involves the solution of a non-linear minimization problem, namely:

$$Minimize\{median_{i=1,\ldots,n} {r_i}^2\} \quad (8)$$

Qualitatively, LMedS tries to find a set of model parameters which best fits the *majority* of the observations. Once LMedS has been applied to a set of observations, a standard deviation estimate may be derived. Rousseeuw and Leroy [12] suggest a value of

$$\hat{\sigma} = 1.4826 \left(1 + \frac{5}{n-p} \right) \sqrt{median\ {r_i}^2} \quad (9)$$

Based on the standard deviation estimate, a weight may be assigned to each observation

$$w_i = \begin{cases} 1, & \text{if } \frac{|r_i|}{\hat{\sigma}} \leq 2.5 \\ 0, & \text{if } \frac{|r_i|}{\hat{\sigma}} > 2.5 \end{cases} \quad (10)$$

All points with weight equal to 1 correspond to model inliers, while points with weight 0 correspond to outliers. The threshold 2.5 reflects the fact that in the case of a Gaussian distribution, very few residuals should be larger than $2.5\hat{\sigma}$.

3 Qualitative Detection of 3D Motion Discontinuities

The proposed method takes a qualitative approach to the problem of independent motion detection, in the sense that solving for the motion or stereo configuration parameters is totally avoided. This is in fact an issue of central importance in the purposive theory of vision [16] according to which problems should be solved directly, by using suitable, specific representations instead of relying on general representations that are difficult to extract accurately. For the specific problem of independent motion detection, this methodological principle favors the use of the normal flow fields and encourages a direct solution to the problem, i.e. one that would not be based on the estimation of 3D motion parameters.

The current method approaches independent motion detection as a problem of pattern matching and, more specifically, as a problem of line fitting. The quantities compared are functions of depth, computed from a temporal sequence of image stereo pairs. One qualitative depth function is computed due to the stereo configuration and another depth function is computed due to motion. Both functions are defined and computed in local image patches. It is shown that these two functions should have a linear relationship in the case that there is only one 3D motion in the image patch under consideration. The violation of this linear relationship in an image patch is attributed to the existence of more than one 3D motions. Consequently, in the general case, boundaries of independently moving objects are detected.

3.1 Method description

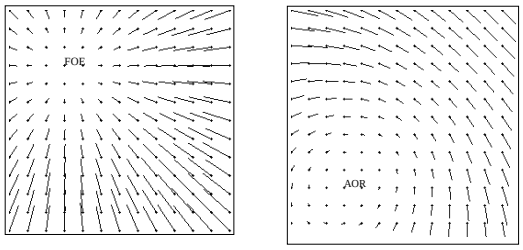
3.1.1 Qualitative depth information in image patches due to motion

The proposed method relies on the assumption that within a small image patch, the translational and rotational components of the egomotion can be accurately approximated with constant vectors. Based on this assumption, eq (1) can be written as

$$\begin{aligned} u &= \frac{u_T}{Z} + u_R \\ v &= \frac{v_T}{Z} + v_R \end{aligned} \tag{11}$$

where (u_T, v_T) and (u_R, v_R) are the constant translational and rotational components of motion for that image patch. This assumption, differs conceptually from the assumption of patchwise constant optical flow. The constancy of the translational and rotational components of egomotion depends only on the observer's 3D motion parameters and poses restrictions on the body of the observer. On the contrary, optical flow constancy implies constancy of depth, which is an environmental assumption.

The hypothesis for constant translational and rotational components of motion is valid in image areas which are far from the *Focus Of Expansion* (FOE) and the *Axis Of Rotation* (AOR). As demonstrated in Fig. 3(a), in the case of pure translational motion, optical flow vectors emanate from the FOE, whose coordinates are $(\frac{Uf}{W}, \frac{Vf}{W})$. In a small image patch far away from the FOE, translational optical flow vectors can be regarded as parallel. The farther the FOE from the image patch, the more accurate the approximation. The approximation becomes perfectly accurate if the W component of translational motion is zero, which is equivalent to the FOE being at infinity. Thus, if the translational component of 3D motion along the Z axis is small relative to the 3D motion along the X and Y axes, the approximation holds for all patches that can be defined on the field of view.



(a) (b)

Figure 3: An example of (a) a translational and, (b) rotational optical flow field.

Similar arguments hold for the case of rotational motion. As can be verified from Fig. 3(b), the rotational optical flow vectors can be approximated by parallel vectors in a small patch far from the AOR, defined as $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$. The hypothesis of a constant rotational component is even more realistic in practical situations, compared to the assumption regarding the translational component, because in the case of egomotion the rotation γ around the Z-axis is usually close to zero and, therefore, the AOR is far outside the field of view.
By substituting u and v of eq (11) in eq (3) we obtain:

$$u^M = \frac{(u_T n_x + v_T n_y)}{Z} + u_R n_x + v_R n_y \tag{12}$$

Assuming that $n_x \neq 0$ and dividing eq (12) by n_x, we obtain:

$$\frac{u^M}{n_x} = \frac{(u_T + n v_T)}{Z} + u_R + n v_R$$

where $n = \frac{n_y}{n_x}$, is the direction of the image gradient. Thus, for a given normal flow direction n in a certain image patch, we can get a depth function of the form:

$$g(Z) = \frac{K}{Z} + L \tag{13}$$

where

$$\begin{aligned} K &= u_T + n v_T \\ L &= u_R + n v_R \end{aligned} \tag{14}$$

are unknown, constant quantities. Equation (13) constitutes the first of the two functions, the comparison of which leads to conclusions about independent motion.

3.1.2 Qualitative depth information in image patches due to stereo

In a way similar to the case of motion, the translational and rotational components of the stereo equivalent motion can be considered constant in image patches. As can easily be shown using geometrical arguments, the translational part of the stereo equivalent motion for a fixating stereo configuration has only a U_s and a W_s component. In all practical situations, U_s is one or two orders of magnitude greater than W_s. Therefore, the FOE for the stereo equivalent motion is far from the image center. The hypothesis for a constant rotational component is also realistic. The rotational component is due to a rotation β_s around the Y-axis, which produces an almost horizontal flow field. By denoting the translational and rotational components of the stereo equivalent motion with (u_{Ts}, v_{Ts}) and (u_{Rs}, v_{Rs}), respectively, we can derive a depth function, corresponding to each patch, of the form:

$$h(Z) = \frac{A}{Z} + B, \tag{15}$$

where

$$\begin{aligned} A &= u_{Ts} + n v_{Ts} \\ B &= u_{Rs} + n v_{Rs} \end{aligned} \tag{16}$$

are again unknown, constant quantities.

3.1.3 Comparison of depth functions

Suppose that for an image point p_i which corresponds to a scene point with depth Z_i, the values $h(Z_i)$ and $g(Z_i)$ of functions h and g are computed. By solving eq (15) for Z and substituting in (13) we obtain:

$$g(Z) = \frac{K}{A} h(Z) + \left(L - \frac{KB}{A} \right) \tag{17}$$

Let $s = \frac{K}{A}$ and $t = L - \frac{KB}{A}$. Then the above equation can be written as:

$$g(Z) = s \cdot h(Z) + t \tag{18}$$

Equation (18) states that the functions of depth $h(Z)$ and $g(Z)$, due to stereo and motion, respectively, have a linear relation for all points with the same gradient direction within an image patch. The scaling parameter s and the shift parameter t depend on the motion parameters and the stereo configuration parameters. Since the quantities A, B, K and L remain constant in an image patch, the same is true for s and t. Therefore, the linear relation of eq (18) is valid for all points with the same gradient direction within an image patch.

Consider now an image patch, in which there are points that correspond to more than one rigid 3D motions. If there are two such motions, eq (13) will hold for some parameters K_1 and L_1, corresponding to points of one motion, and for some other parameters K_2 and L_2, corresponding to points of the other motion. Equation (18) will not hold for the same parameters s and t for all points in an image patch. The detection of such situations signals the presence of more than one rigid motions and, therefore, the presence of independent motion.

A method for comparison treats the problem as a problem of line fitting. Each point in a patch provides one equation of the form of eq (18). $f(Z)$ and $g(Z)$ are computable quantities, and s and t are unknown parameters. In fact, the set of equations (18) for all points in an image patch form an overdetermined set of equations over variables s and t, that can be solved with the LMedS robust estimator. In the presence of two rigid motions, robust regression will estimate the parameters for the majority of the points (dominant motion within that patch). Model inliers will correspond to the points of the dominant motion, while model outliers will correspond to the points of the secondary motion. The absence or the presence of outliers signals one or more 3D motions, respectively. Note that if there are two rigid motions, then the high breakdown point of LMedS suffices to handle correctly the segmentation of the scene. In case that there are more than two rigid 3D motions within a patch, and none is dominant (in the sense that 50% of the total number of points corresponds to that motion) the method will fail to estimate a correct set of parameters s and t. However, it is very likely that whatever the estimated parameters are, outliers will exist and, therefore, discontinuities will be detected.

3.2 Ambiguities in independent 3D motion detection

According to the previously described method, if no set of parameters s and t can be found such that eq (18) holds for all points of a certain gradient direction in a patch, then there is a 3D motion discontinuity in that patch. However, the reverse is not always true: If eq (18) holds for all image points of a certain gradient direction within a patch, then it is not certain that no 3D motion discontinuity exists. Assume that in a certain tile, the depth function acquired due to stereo is given by the relation $h(Z) = \frac{A}{Z} + B$. Suppose also that there are two different 3D motions m_1 and m_2, each of which gives a function of depth $g_1(Z) = \frac{K_1}{Z} + L_1$ and $g_2(Z) = \frac{K_2}{Z} + L_2$, respectively. If, (see eq (17)) the relations

$$\begin{aligned} s &= \frac{K_1}{A} = \frac{K_2}{A} \\ t &= L_1 - \frac{K_1 B}{A} = L_2 - \frac{K_2 B}{A} \end{aligned} \tag{19}$$

hold simultaneously, then although m_1 and m_2 are different motions, eq (18) holds for all points of an

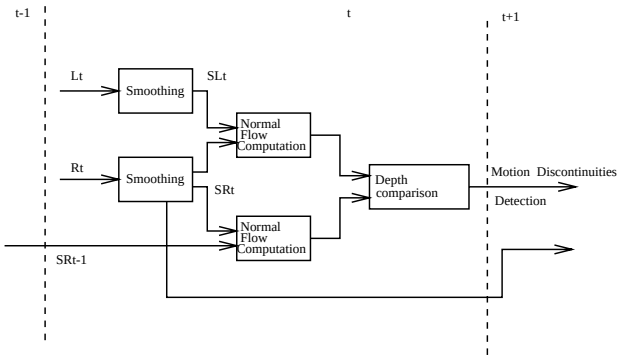


Figure 4: A schematic presentation of the method for 3D motion discontinuities detection.

image patch. Equations (20) after substitution from eqs. (14), yield:

$$(u_{T1} - u_{T2}) \quad = \quad n(v_{T2} - v_{T1}) \tag{20}$$

and

$$(u_{R1} - u_{R2}) + n(v_{R1} - v_{R2}) =$$
$$= \frac{B}{A}\{(u_{T1} - u_{T2}) - n(v_{T2} - v_{T1})\} \tag{21}$$

For both equations to hold, the gradient direction and the motions m_1 and m_2 should be such that:

$$n = \frac{u_{T1} - u_{T2}}{v_{T2} - v_{T1}} = \frac{u_{R1} - u_{R2}}{v_{R2} - v_{R1}}$$

This is a rather restricted case, because it requires the selected gradient direction to have a special relation with both the translational and the rotational components of both motions. An easy way to avoid these ambiguous situations is to test if eq (18) holds for more than one gradient directions.

3.3 Implementation and performance issues

Figure 4 summarizes schematically the method for the detection of discontinuities of 3D motion. At time t, a pair of images L_t and R_t is acquired by the stereo configuration. Both images are smoothed. Image smoothing is implemented by the convolution of the input images with a 5×5 Gaussian kernel of standard deviation $\sigma_G = 1.4$. Two normal flow fields (from stereo and motion) are computed from the smoothed images SR_t, SL_t and SR_t and SR_{t-1}. The image is then partitioned into tiles. For each tile, a histogram of normal flow directions is computed. The dominant normal flow directions are determined and for these directions, the functions h and g are computed. Functions h and g are then compared by LMedS estimation. Independent motion detection is reported in an image patch if, for at least one gradient direction the depth functions due to motion and stereo do not have a linear relationship.

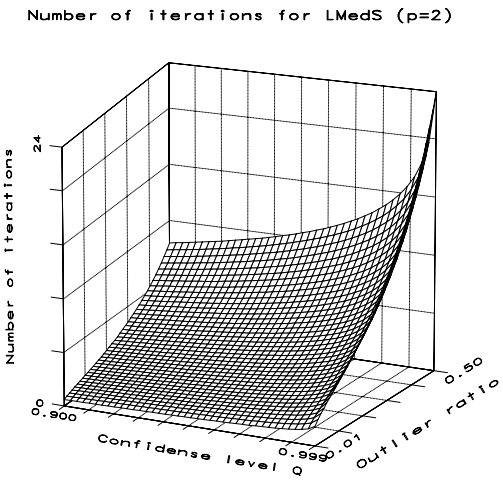


Figure 5: Number of iterations m for LMedS estimation (eq (22)), as a function of Q and e. The number of model parameters is kept equal to p = 2.

LMedS estimation is the most computationally intensive part of the independent motion detection scheme. The robustness of LMedS is offered at some additional computational cost. LMedS minimization cannot be reduced to a least-squares based solution, but must be solved by a search in the space of possible estimates generated by the data. If p is the number of parameters to be estimated, there are $O(n^p)$ possible p-tuples. Because this search space may become too large, in practical situations a certain probability of error is tolerated, which enables the use of a Monte Carlo type of speed-up technique. If e is the fraction of data contaminated by outliers, then the probability Q that at least one out of m p-tuples has only uncorrupted observations is equal to:

$$Q \quad = \quad 1 - [1 - (1 - e)^p]^m \tag{22}$$

Thus, solving eq. (22) for m, gives rise to a lower bound for the number of p-tuples that should be considered. Note that eq. (22) is independent of n. The performance of LMedS estimation depends on the number m of the required iterations. The number of iterations depends on the number of parameters p to be estimated. In this case, $p = 2$. Figure 5 shows a 3D plot of the number of required iterations m as a function of the the confidence level Q to be reached and the outlier ratio e. Each of the m iterations, requires the selection of candidate parameter values and the computation of the squared residuals between the observations and the predictions of the model. The selection of candidate parameter values can be made in various ways. The straightforward one is to randomly select a number of points equal to the number of parameters and solve a linear system of equations. Alternatively, candidate solutions can be formed by the results of least squares parameter estimation in re-

gions of the input normal flow field. Both approaches were tested. For the linear system approach, normal flow vectors were randomly selected over the whole image plane. For the least squares approach, rectangles of random dimensions and locations were selected; all points within such rectangles contributed to the least squares solution. Experimental results demonstrated that the least squares approach gives better results (smaller median) compared to the linear system approach, for the same number of candidate solution trials. However, the least squares approach incurs an extra computational overhead associated with it.

In our implementation, we do not make use of sorting to compute the median in each of the m repetitions. Instead we use an algorithm that selects the kth largest number out of n numbers, originally suggested in [17]. This algorithm has a time complexity of $O(n)$, rather than the $O(n \log n)$ complexity of the best serial sorting algorithm. Thus, the overall computational complexity of LMedS becomes $O(mn)$.

4 Experiments with image sequences

The independent motion detection method has been tested using image sequences that were acquired by specialized equipment available at the Computer Vision and Robotics Laboratory (CVRL) of ICS-FORTH.

Several experiments have been conducted to test the proposed independent motion detection method. It should be stressed that during the course of all the experiments the exact values of the camera focal length and image origin were unknown. Therefore, the results should also be interpreted on the basis of the fact that the proposed method does not require camera calibration.

As a testbed for evaluating the performance of the proposed method two main image sequences have been employed, namely the "toy-car" sequence and the "cart" sequence. One frame of the "toy-car" sequence is shown in Fig. 6. The scene captured consists of a background wall covered with paper and a toy-car together with some other objects of similar size in the foreground. The toy-car and the various objects are closer to the observer, compared to the background wall which is farther (at a larger depth). The observer performs translational motion with a right to left direction. Apart from the toy-car, the rest of the scene is stationary. The toy-car is moving across the scene in a left to right direction, with unrestricted 3D motion.

The proposed method has been applied to this image sequence. The 3D motion segmentation results are presented in Fig 7. In this image, gray color corresponds to points where normal flow has been rejected due to low image gradient. Black color has been assigned to all the points in tiles where the comparison of the depth functions due to motion and due to stereo resulted in ony one rigid motion. Finally, white color has been assigned to all the points in tiles where the comparison of the depth functions revealed more than one rigid motions. The results demonstrate that correct discrimination of the independent motion of the toy-car has been achieved.

Figure 6: One frame from the "toy-car" sequence.

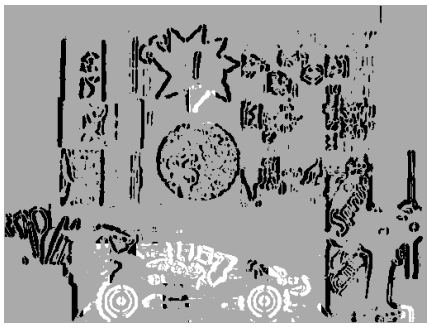


Figure 7: Motion discontinuities detection for the "toy-car" image sequence.

The method has also been tested in the "cart" sequence. One frame of this sequence is shown in Fig. 8. In this set of images, the observer performs a translational motion with U and W components as well as with a rotational β component. The horizontal translation is the motion that dominates. The field of view contains a distant background and a close to the observer foreground. The background contains two independently moving objects: A cart that translates in the opposite direction of the observer (middle of the scene) and a small box (to the right of the scene) that translates at the same direction with the observer, but with different velocity. The foreground of the scene contains a table on which there is a toy car. Both objects are stationary relative to the static environment.

The results of the motion segmentation of the "cart" sequence are shown in Fig. 9. It can be seen that the method has correctly identified most of the tiles where more than one 3D motions are present. Although the results are complete for the case of the moving box, the method fails to detect the elongated (upper) part of the cart because there are not enough normal flow vectors (in terms of the requirements of the method) to support the existence of independent motion.

5 Summary

A novel, qualitative method for independent 3D motion detection has been presented. Independent motion detection has been achieved by exploiting

Figure 8: One frame from the "cart" sequence.

Figure 9: Motion discontinuities detection for the "cart" image sequence.

depth information that can be computed by a rigidly moving binocular vision system. Instead of using optical flow, which amounts to solving the ill-posed correspondence problem, the normal flow field is used in both the stereo and motion domains. The processing of both the stereo and motion pairs yield qualitative information about the scene structure within image patches. Lack of consistency between the evidence provided by stereo and motion leads to conclusions about the number of rigidly moving regions within an image patch. The experimental results obtained demonstrate the robustness and effectiveness of the approach.

Acknowledgements: The authors would like to thank Prof. Yiannis Aloimonos, Prof. George Tziritas and Dr. Cornelia Fermüller for helpful discussions on the problem of independent motion detection.

References

[1] R.C. Jain. Segmentation of Frame Sequences Obtained by a Moving Observer. *IEEE Transactions on PAMI*, PAMI-7(5):624–629, September 1984.

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

[3] P. Bouthemy and E. Francois. Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence. *International Journal of Computer Vision*, 10(2):157–182, 1993.

[4] 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.

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

[6] S. Ayer, P. Schroeter, and J. Bigun. Segmentation of Moving Objects by Robust Motion Parameter Estimation over Multiple Frames. In *European Conference on Computer Vision*, 1994.

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

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

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

[12] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley and Sons Inc., New York, 1987.

[13] K.L. Boyer, M.J. Mirza, and G. Ganguly. The Robust Sequential Estimator: A General Approach and its Application to Surface Organization in Range Data. *IEEE Transactions on PAMI*, PAMI-16:987–1001, 1994.

[14] P. Meer, A. Mintz, and A. Rosenfeld. Robust Regression Methods for Computer Vision: A Review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[15] P.J. Rousseeuw. Least Median of Squares Regression. *Journal of American Statistics Assocciation*, 79:871–880, 1984.

[16] Y. Aloimonos. Purposive and Qualitative Active Vision. In *Proceedings DARPA Image Understanding Works.*, pages 816–828, 1990.

[17] R. Sedgewick. *Algorithms*. Addison-Wesley, Reading, MA, 1988.