Robust Regression for the Detection of Independent 3D Motion by a Binocular Observer

Antonis A. Argyros, 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

e-mail: {argyros, trahania, orphanou}@ics.forth.gr

Abstract

A method is proposed for the visual detection of objects that move independently of the observer in a 3D dynamic environment. Many of the existing techniques for solving this problem are based on 2D motion models, which is equivalent to assuming that all the objects in a scene are at a constant depth from the observer. Although such methods perform well if this assumption holds, they may give erroneous results when applied to scenes with large depth variations. Additionally, many of the existing techniques rely on the computation of optical flow, which amounts to solving the ill-posed correspondence problem. In this paper, independent 3D motion detection is formulated using 3D models and is approached as a problem of robust regression applied to visual input acquired by a binocular, rigidly moving observer. Similar analysis is applied both to the stereoscopic data taken by a non-calibrated stereoscopic system and to the motion data obtained from successive frames in time. Least Median of Squares (LMedS) estimation is applied to stereoscopic data to produce maps of image regions characterized by a dominant depth. LMedS is also applied to the motion data that are related to the points at the dominant depth, to segment the latter with respect to 3D motion. In contrast to the methods that rely on 2D models, the proposed method performs accurately, even in the case of scenes with large depth variations. Both stereo and motion processing is based on the normal flow field which can be accurately computed from the spatiotemporal derivatives of the image intensity function. Although parts of the proposed scheme have non-trivial computational requirements, computations can be expedited by various ways which are discussed in detail. This is also demonstrated by an on-board implementation of the method on a mobile robotic platform. The method has been evaluated using synthetic as well as real data. Sample results show the effectiveness and robustness of the proposed scheme.

1 Introduction

Because of the egomotion of an observer equipped with visual sensors, his visual field appears to be moving in a manner that depends on the observer's 3D motion parameters and the structure of the scene in view. In case that certain objects move independently, the 3D velocity of the observer relative to all points in its environment is not the same. The problem of independent 3D motion detection is defined as the problem of locating such objects, if they exist in a scene.

The ability to detect independent 3D motion is very important for an observer that interacts with a dynamic environment. It is known [1] that independent motion detection is one of the basic visual competences of most of the biological organisms possessing the sense of vision. More complex behaviors like obstacle avoidance, visual target tracking and surveillance seem to be based on the perceptual information provided by such mechanisms. The ability to detect independent motion is also very crucial for robots that should navigate autonomously in dynamic environments.

The importance of visual motion understanding in general, and of independent 3D motion detection[1] in particular, has been recognized for years and a lot of work has been done along this research direction. A lot of this work depends on the accurate computation of the optical flow field [2, 3]. Moreover, in many cases 2D motion models (affine or quadratic) have been employed [2, 4]. Independent motion detection is usually achieved by detecting the discontinuities (in the 2D model) and reporting them as 3D motion discontinuities. This has been approached using a variety of techniques. In [5] robust statistics, and more specifically M-estimators are used to distinguish the dominant 2D motion from the secondary 2D motions. A similar idea is exploited by Ayer et al in [6], where two other robust estimators, namely Least Median of Squares and Least Trimmed Squares are used to discriminate the dominant from the secondary 2D motions. Bouthemy and Francois [7] present a motion segmentation method relying on 2D affine motion models and a statistical regularization approach.

Most of these methods produce good results in certain classes of scenes. However,

---

[1]In the remainder of this paper and, unless it is explicitly stated otherwise, we will refer to the problem of independent motion detection for the case of a moving observer.

the methods that are based on the computation of optical flow tend to eliminate the effect they are trying to detect, because optical flow computation is based on the assumption of a smooth motion field which, by definition, does not hold for scenes with independently moving objects. Another drawback stems from the 2D motion models they employ. The projection of 3D motion on the 2D image plane depends on certain characteristics of the observer (e.g. focal length of the optical system), on the parameters of the relative 3D motion between an object and the observer and, on the depth of the scene. Therefore, discontinuities in the computed 2D motion field are not only due to 3D motion discontinuities (i.e. independently moving objects), but also due to depth discontinuities. For this reason, methods that employ 2D models of motion perform well only in scenes where depth variations are small compared to the distance from the observer. In cases of scenes with large depth variations, these methods fail to provide reliable results.

In an effort to overcome the inherent problems of employing 2D models of motion, the more accurate 3D models have been employed. By employing 3D models the problem becomes much more complicated because extra variables are introduced regarding the depth of the scene points. Thus, certain assumptions are made in order to provide additional constraints. Common assumptions of existing methods are related to the motion of the observer, the structure of the scene in view, or both. Moreover, in the majority of reported works, it is assumed that either a dense optical flow field or a sparse set of point correspondences can be computed from a set of images. Jain [3] 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 cannot have rotational components), knowledge of the direction of translation is required. Adiv [8] performs segmentation by assuming planar surfaces undergoing rigid motion, thus introducing an environmental assumption. Thompson and Pong [9] derive various principles for detecting independent motion when certain aspects of the egomotion or of the scene structure are known. Although it is an inspiring work, the practical exploitation of the underlying principles is limited because of the assumptions they are based on and open implementation issues.

Some alternative methods that follow the active and purposive vision paradigm [10,

11, 12], do not rely directly on the estimation of the parameters of a motion model, but rather on qualitative geometrical properties of the normal flow field. Thus, the solution of the structure from motion problem is by-passed. A method that falls in this context has been proposed by Sharma and Aloimonos [13]. However, as in the case of [3], restricted (translational) egomotion is assumed. Nelson [14] presents two such methods for independent motion detection. 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. Despite their theoretical elegance, these methods suffer in that they perform independent motion detection by assuming either known or restricted egomotion.

In this paper, the problem of independent motion detection is formulated using 3D motion models and is approached as a robust regression problem. Robust regression has also been employed in the past in the problem of motion segmentation. Ayer et al [6] have proposed a method which uses robust regression to identify independently moving objects. However, since no information on scene structure is used, the method works well only when the scene in view forms a *frontoparallel plane.*

The capability that has been developed (henceforth referred to as *IMDRR*[2]) makes use of the *Least Median of Squares (LMedS)* estimation technique [15]. LMedS is initially applied to the stereoscopic data that are acquired by the binocular observer, to separate the set of image points in two subsets: stereo inliers and stereo outliers. The stereo inliers correspond to the dominant scene depth, while the stereo outliers correspond to the rest of the scene points. At a second stage, LMedS is applied to the motion data that correspond to the points of the dominant depth. This second application of LMedS results in a segmentation of the points at the dominant depth into motion inliers and motion outliers. Motion inliers correspond to the dominant motion (egomotion), whereas motion outliers can only be due to independent motion. The exploitation of stereo information relaxes the constraint of a strictly frontoparallel scene that other methods assume (e.g. [6]).

---

[2]*IMDRR* is an acronym for *Independent Motion Detection based on Robust Regression.*

The motion analysis component of *IMDRR* is based on the spatiotemporal derivatives of the image intensity function (normal flow). The ill-posed problem of point correspondence is also avoided for the stereo case. The stereo configuration is treated as the hypothetical motion that would map the position of the left camera to the position of the right camera. Again, normal flow is computed between the two frames of the stereo pair.

Regarding the egomotion of the imaging system, it is assumed to be a rigid motion. This is a valid assumption because, although an observer may move in a non-rigid fashion, all biological eyes and artificial cameras are essentially rigid. The motion of the independently moving objects, however, is not constrained to be rigid. Since the method does not rely on the estimation of motion parameters for independently moving objects, no particular assumptions are actually made regarding their motion.

The rest of the paper is organized as follows. Section 2 describes the geometry of the imaging system, the input used by *IMDRR* and the 3D motion models employed. Section 3 is a brief introduction to robust regression and the LMedS estimation technique, for the sake of self-completeness of the paper. Section 4 describes *IMDRR* in detail and section 5 focuses on the implementation and performance issues of *IMDRR*. In Section 6, experimental results from the application of *IMDRR* to synthetic data and real world image sequences are presented and discussed. Finally, section 7 concludes the paper and gives directions for future research work.

2 The visual input

Before proceeding with the description of *IMDRR*, the geometry of the imaging system, the input to the motion and stereo processing modules and the 3D motion models employed are presented. Moreover, the choice of the normal flow field in all computations is justified.

2.1 The coordinate system

Consider a coordinate system $OXYZ$ centered in the optical center (nodal point) of a camera, such that the axis OZ coincides with the optical axis. Suppose 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.

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 given by [16]:

$$u = \frac{(-Uf + xW)}{Z} + \alpha \frac{xy}{f} - \beta \left(\frac{x^2}{f} + f \right) + \gamma y \tag{1a}$$

$$v = \frac{(-Vf + yW)}{Z} + \alpha \left(\frac{y^2}{f} + f \right) - \beta \frac{xy}{f} - \gamma x, \tag{1b}$$

where f represents the focal length of the imaging system.

2.2 Motion field - optical flow field

Equations (1) describe the 2D motion vector field, which relates the 3D motion of a point with its 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 [17], which describes the apparent motion of brightness patterns observed because of the relative motion between an imaging system and its environment. Verri and Poggio [18] have shown that the motion and optical flow fields are identical in specific cases only.

Even in the case that these two fields are identical, the computation of the optical flow field is an ill-posed problem [19], since special conditions (such as smoothness) should be satisfied for a unique solution to exist. Such conditions are not satisfied in practice, especially in the case of independent motion, where by definition motion discontinuities do exist. Furthermore, although progress on optical flow computation has been made [20, 21, 22], it is well known that the estimation of 3D motion is very sensitive to the presense of noise in the optical flow field [23, 24, 25].

For the above reasons, the proposed scheme for independent motion detection does not rely on the computation of the optical flow field, but rather on the normal flow field,

i.e. the projection of optical flow on the direction of intensity gradient. The normal flow field has also been used in the past for both egomotion estimation [26, 27, 28] and independent motion detection [14, 13].

2.3 Normal flow field - normal motion field

Assuming that a sequence of images can be modeled as a continuous function $I(x, y, t)$ of two spatial (x, y) and one temporal (t) variables, and that irradiance is conserved between two consecutive frames, it is possible to write the well known optical flow constraint equation [29] at each image point:

$$I_x u + I_y v + I_t = 0 \tag{2}$$

where (u, v) is the optical flow vector at point (x, y) and I_x, I_y, I_t are the partial derivatives of function I with respect to variables x, y and t. Equivalently, eq (2) can be written in the form of a dot product

$$(I_x, I_y) \cdot (u, v) = -I_t, \tag{3}$$

Equation (3) constitutes the mathematical expression of the *aperture problem*. Optical flow cannot be recovered based on local information only. What can be computed is the projection of optical flow along the direction of image spatial gradients, also known as normal flow. Figure 2 gives a schematic view of the aperture problem.

According to eq (3), the magnitude u_{nm} of the normal flow is equal to

$$u_{nm} = -\frac{I_t}{\sqrt{I_x^2 + I_y^2}} \tag{4}$$

The normal flow field is not necessarily identical to the normal motion field (the projection of the motion field on the direction of the image spatial gradient), in the same way that the optical flow is not necessarily identical to the motion field. In [18], it is shown that the difference between the algebraic values of a normal flow vector and a normal motion vector is equal to $\frac{1}{||\nabla I||}\frac{dI}{dt}$, where $||\nabla I||$ is the magnitude of the intensity gradient and $\frac{dI}{dt}$ is the time derivative of the intensity function. Therefore, the normal flow field is a good approximation of the normal motion field in points where

the intensity gradient has a large value. Normal flow vectors at such points can be used as a robust input to 3D motion analysis.

2.4   Normal flow field due to motion

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

$$u_{nm} \quad = \quad un_x + vn_y \qquad (5)$$

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

$$u_{nm} \quad = \quad -n_x f \frac{U}{Z} - n_y f \frac{V}{Z} + (xn_x + yn_y) \frac{W}{Z} +$$
$$+ \quad \left\{ \frac{xy}{f} n_x \left(\frac{y^2}{f} + f \right) n_y \right\} \alpha - \left\{ \left(\frac{x^2}{f} + f \right) n_x + \frac{xy}{f} n_y \right\} \beta + (yn_x - xn_y)\gamma \qquad (6)$$

Equation (6) highlights some of the inherent difficulties of independent motion detection. Each image point (in fact, each point at which the intensity gradient has a significant magnitude) provides one constraint on the 3D motion parameters. In case that only the observer is moving, the above equation holds for each point and for one specific unknown set of 3D egomotion parameters (U_E, V_E, W_E), $(\alpha_E, \beta_E, \gamma_E)$. In the case of independent motion, there is at least one more set of motion parameters (U_I, V_I, W_I), $(\alpha_I, \beta_I, \gamma_I)$ that is valid for some of the image points. Furthermore, if no assumption is made regarding the depth Z, each point provides at least one independent depth variable. Evidently, the problem cannot be solved if no additional depth information is available. Equation (6) also shows why the problems of egomotion estimation and independent motion detection can be considered as chicken-and-egg problems. If egomotion is known, then it can be compensated and, therefore, independent motion detection is greatly facilitated. On the other hand, if independently moving objects have not been previously identified, each image point may or may not be part of an independent motion; egomotion estimation would then be confused by the wrong evidence provided by the points belonging to an independently moving object. For this reason, a lot of techniques for estimating egomotion [26, 27, 28, 30] assume the absence of independently moving objects.

2.5 Normal flow field due to stereo

Consider the typical stereo configuration of a fixating pair of cameras, as shown in Fig. 3(a). This type of configuration may easily be achieved by existing mechanical heads [31] and is very common in biological organisms, including humans.

The images acquired simultaneously by a stereoscopic system encapsulate information relevant to depth, that manifests itself in the form of *disparities* defined by the displacements of points between images. Since these images are acquired simultaneously, there is no dynamic change in the world that can be recorded by them. It can easily be observed that a stereo image pair is identical to the sequence that would result from a hypothetical (ego)motion that brings the one camera to the position of the other. The hypothetical motion that transforms the position of one camera to the other is simpler than the one described by the general motion model of eq (1). Fig. 3(b) shows the motion that maps the position of the left camera, to the position of the right camera. Evidently, there is no rotation around the X and Z axes, and no translation along the Y axis. Consequently, if $(U_s, 0, W_s)$ and $(0, \beta_s, 0)$ are the translational and rotational parameters of the hypothetical motion, then at each point, a normal flow value u_{ns} due to stereo may be computed, which is equal to

$$u_{ns} = -n_x f \frac{U_s}{Z} + (xn_x + yn_y) \frac{W_s}{Z} - \left\{ \left(\frac{x^2}{f} + f \right) n_x + \frac{xy}{f} n_y \right\} \beta_s \qquad (7)$$

In practical situations, the computation of normal flow from a pair of stereo images needs further consideration. The computation of normal flow is based on the optical flow constraint equation, which 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. We may observe, however, that since normal flow is the projection of optical flow in a certain direction, the value of normal flow at a certain point is bounded by the value of optical flow. Thus, we may select suitable stereo configurations for which the maximum obtained optical flow (and consequently the maximum normal flow) does not exceed a certain upper bound. Additionally, the normal flow magnitude is expressed in pixels. By appropriately selecting the spatial image resolution for the above computations, the requirements posed on the

9

stereo configuration can be drastically relaxed at the cost of a coarser estimation of the normal flow field.

3   Robust regression

Regression analysis (fitting a model to noisy data) is a very important statistical tool. In the general case of a linear model [32], given by the expression

$$y_i \;=\; x_{i1}\theta_1 + \ldots + x_{ip}\theta_p + e_i, \tag{8}$$

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 in each of the observations. In classical applications of regression, e_i is assumed to be normally distributed with zero mean and unknown standard deviation. Let $\hat{\theta}$ be the vector of estimated parameters $\hat{\theta}_1, \ldots, \hat{\theta}_p$. Given these estimates, predictions can be made for the observations:

$$\hat{y}_i \;=\; x_{i1}\hat{\theta}_1 + \ldots + x_{ip}\hat{\theta}_p \tag{9}$$

Thus, a residual between the observation and the value predicted by the model may be defined as:

$$r_i \;=\; y_i - \hat{y}_i \tag{10}$$

Traditionally, $\hat{\theta}$ is estimated by the Least Squares (LS) method. LS involves the solution to a minimization problem, namely:

$$Minimize \sum_{i=1}^{n} r_i{}^2 \tag{11}$$

and achieves optimal results if the underlying noise distribution is Gaussian. However, in cases where the noise is not Gaussian, the LS estimator becomes unreliable. The LS estimator becomes highly unreliable also in the presence of outliers, that is observations that deviate considerably from the model representing the rest of the observations. One criterion for measuring the tolerance of an estimator with respect to outliers is its *breakdown point*, which is 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.

In order to be able to handle data sets containing large portions of outliers, a variety of robust estimation techniques have been proposed. Many of them have been used in computer vision [33, 34, 35, 36, 32]. Meer et al [36] provide an excellent review of the use of robust regression methods in computer vision.

Probably, the most popular robust estimators are the M-estimators [37, 32]. M-estimators are based on the idea of replacing the sum of the squared residuals r_i^2 by another function of the residuals. The aim of these functions is to protect the estimate from strongly outlying observations. However, M-estimators have two major drawbacks. First, it can be shown that although they behave better than Least Squares in practical situations, their breakdown point is equal to $1/n$ [32], where n is the number of observations. This becomes equal to zero as n increases[3]. Second, it can be shown that they require a reliable initial estimate of the model parameters because otherwise, they can be trapped in local minima.

In an effort to provide robust estimators with a higher breakdown point, Rousseeuw and Leroy [32] introduced the so-called S-estimators which are defined by minimizing a robust measure of the scatter of the residuals. The *Least Median of Squares (LMedS)* is an S-estimator which is described in detail in the next section. LMedS has a breakdown point of 50%. It can be demonstrated that 50% is the highest possible breakdown point of an estimator, because for larger outlier contaminations it is impossible to distinguish the "good" from the "bad" data. Recently, a new robust regression method, namely MINPRAN, has been proposed [34] which reports a breakdown point that is higher than 50%. However, MINPRAN makes extra assumptions regarding the distribution of the outliers. More specifically, it assumes a random distribution of the outliers and tries to group data according to a linear model so that the probability of randomness of the grouped data is minimized. The concept of MINPRAN is very interesting, although its computational complexity poses restrictions on its practical exploitation.

---

[3]Note that the Least Squares method is in fact a trivial case of an M-estimator.

3.1 Least Median of Squares (LMedS)

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

$$Minimize\{median_{i=1,...,n} r_i^2\} \tag{12}$$

Qualitatively, LMedS tries to estimate a set of model parameters that best fit the *majority* of the observations. The presence of some outliers in a set of observations will not influence LMedS estimation, as long as the majority of the data fit into the particular model. More formally, LMedS has a breakdown point of 50%.

Once LMedS has been applied to a set of observations, a standard deviation estimate may be derived:

$$\hat{\sigma} = C\sqrt{median\ r_i^2} \tag{13}$$

where C is an application-dependent constant. Rousseeuw and Leroy [32] suggest a value of

$$C = 1.4826\left(1 + \frac{5}{n-p}\right) \tag{14}$$

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

$$w_i = \begin{cases} 1, & \text{if } \frac{|r_i|}{\hat{\sigma}} \leq THR \\ 0, & \text{if } \frac{|r_i|}{\hat{\sigma}} > THR \end{cases} \tag{15}$$

All points with weight $w_i = 1$ correspond to model inliers, while points with weight $w_i = 0$ correspond to outliers. The threshold THR controls the sensitivity to outliers. Typically, a value of 2.5 is used. This value reflects the fact that in the case of a Gaussian distribution, very few residuals should be larger than $2.5\hat{\sigma}$. Note that the criterion according to which the labels of inlier and outlier are assigned to data is itself robust, since it involves calculations over the median of residuals. Moreover, the method adjusts automatically to the noise levels of the observations. The better the estimated model fits to the observations, the smaller the median residual is and, therefore, the finer the outlier detection becomes.

The computational requirements of LMedS are reported as high [32, 36]. This is because LMedS involves a non linear minimization problem which must be solved by a search in the space of possible estimates generated by the data. However, in section 5, modifications are proposed, which improve the computational performance of LMedS.

4   Independent motion detection

The proposed method for independent 3D motion detection makes a twofold use of the LMedS estimation technique. First, LMedS is applied to the stereo normal flow field, according to eq (7). This results in a characterization of the image points as stereo inliers (i.e. points that belong to a dominant depth) or as stereo outliers (i.e. points that deviate from the dominant depth). More formally, let S denote the set of image points where a reliable normal flow vector due to stereo has been computed. Then, the above application of LMedS partitions S into two subsets, the stereo inliers S_I and the stereo outliers S_O. Obviously,

$$S = S_I \bigcup S_O$$
$$S_I \bigcap S_O = \emptyset$$

LMedS is subsequently applied to the motion normal flow field in the points of S_I, according to the model of eq (6). In general, this relation forms a linear model, when depth Z and the motion parameters (U, V, W) and (α, β, γ) are constant for all image points. In terms of LMedS estimation, the outliers of the linear model will be points for which either Z deviates from a dominant depth, or points whose 3D motion parameters are different from the dominant motion. For the purpose of independent motion detection, we are interested in the second class of points. Since the latter application of LMedS is restricted to S_I (points at dominant depth), the outliers of the model can only be due to independent motion.

13

4.1 The *IMDRR* algorithm

The complete 3D independent motion detection capability can now be described; a block diagram of it is shown in Fig 4.

At time t, a pair of images L_t and R_t is acquired by the stereo configuration. Both images are smoothed producing SL_t and SR_t. The normal flow field NF_s due to the stereo configuration is computed. To do so, the spatiotemporal derivatives of the image intensity function are computed. Without loss of generality, we compute the temporal derivatives assuming a left-to-right image transition. The image intensity gradient is then thresholded; all points for which the magnitude of intensity gradient is lower than a threshold, are removed from further consideration. NF_s is fed into the LMedS module that makes use of the model described by eq (7). The output of the LMedS module is a partition of the scene points into two classes: stereo inliers (S_I) and stereo outliers (S_O). Stereo inliers correspond to the dominant depth of the scene. Stereo outliers correspond to either noisy observations (that is, points where a large error was introduced in the computation of normal flow) or points with depth deviating sufficiently from the dominant scene depth. Note that U_s, W_s, and β_s are the same in eq (7) for all image points.

In a way completely analogous to that of stereo, the motion normal flow field NF_m is computed from smoothed images SR_{t-1} and SR_t. By using the stereo inliers set S_I, we mask out the points of NF_m that correspond to the stereo outliers. The remaining normal flow values are fed to the second LMedS module that makes use of the model given by eq (6). The output of this second LMedS module is a partition of the points at the dominant depth layer (S_I) into motion inliers and motion outliers. Motion inliers correspond to the motion with the largest spatial support. It is reasonable to assume that this motion corresponds to egomotion. Motion outliers correspond to independent motion. In other words, S_I is further partitioned in this step into two subsets: (a) S_I^E, that is image points that move according to the observer's egomotion and, (b) S_I^{IM}, that is image points that move independently of the observer.

The above partition of S_I may result in isolated points for a number of reasons. The normal flow field is usually a sparse field, because normal flow values below a

certain threshold are rejected. In some cases, points may become members of the inlier or outlier classes due to errors in the computation of normal flow and not because of depth variations or motion discontinuities. Finally, normal flow is just a projection of the optical flow onto a certain direction. Infinitely many other optical flow vectors have the same projection onto this direction and, consequently, a number of point misclassifications may arise. We overcome the problem of isolated points 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.

Points that are finally classified as independently moving, may belong to an independent rigid or nonrigid motion. No model is actually assumed by *IMDRR* for the independent motion. The rigid motion model is assumed only for egomotion.

5   *IMDRR* implementation

The implementation of *IMDRR* follows closely the block diagram of Fig 4. In this section an attempt is made to elaborate on implementation issues pertaining to *IMDRR* and also to demonstrate its amenability to parallel implementation.

The first step in *IMDRR* is image smoothing, which is implemented by the convolution of the input images with a 5×5 Gaussian kernel with standard deviation $\sigma_G = 1.4$. Image smoothing is a characteristic example of low level, data parallel algorithms [38]. The parallelization of such algorithms has been extensively studied [39], and efficient implementations have been developed for both SIMD [40] and MIMD [41] parallel architectures.

The computation of normal flow (either stereo normal flow or motion normal flow), also belongs to the class of low level, data parallel algorithms. Normal flow computation

can be decomposed into three independent stages which correspond to the computation of the two spatial and the one temporal derivative of the image intensity function. Each of these stages involves some local mask correlation, which has the same general computational characteristics as image smoothing. In particular, the image gradient is computed by convolving the image with the Sobel operators [42]. The time derivative is computed by substructing averaged (in 3×3 windows) intensities from the successive frames. The spatial and temporal derivatives are combined through eq (4) to give the normal flow values (for both stereo and motion normal flow).

LMedS estimation, is the most computationally intensive part of the independent motion detection scheme. LMedS minimization cannot be reduced to a closed form solution, but must be solved by a search in the space of possible estimates generated by the data. Let p denote the number of parameters to be estimated; then there are $O(n^p)$ possible p-tuples. Because this search space may become prohibitively large, in practical situations, a Monte-Carlo type of speedup technique is employed [36], in which a certain probability of error is tolerated. If e is the fraction of outliers in the data, then the probability Q that at least one out of m p-tuples has only uncorrupted observations is equal to:

$$Q = 1 - [1 - (1 - e)^p]^m \qquad (16)$$

Thus, the solution of eq (16) for m, gives a lower bound for the number of p-tuples that should be tried. Note that eq (16) is independent of n. Each of the m trials, requires the selection of candidate parameter values and the computation of the squared residuals between the observations and the predictions of the model. It is noted that we are interested in obtaining an estimate for m for $p = 3$ and $p = 6$. These cases correspond to the stereo model of eq (7) (three parameter model) and the motion model of eq (6) (six parameter model), respectively. Figures 5(a) and 5(b) show a 3D plot of the number of required iterations m (for $p = 3$ and $p = 6$, respectively) as a function of the confidence level Q and the outlier ratio e. As can be observed, the number of iterations increases with respect to the confidence level Q and the outlier ratio e. In most practical cases, the spatial extent of independently moving objects is small, compared to the static background. This is equivalent to confining the outlier ratio in a range $[0..a]$, with

16

$a < 0.5$. Referring to Figs 5(a),(b) we observe that in this case the number of required iterations is considerably reduced.

These figures hold for the case that at each iteration, p observations are randomly selected, a linear system is solved and the median of the squared residuals is tested against the minimum of the squared residuals computed so far. Alternatively, candidate solutions can be formed by the results of least squares parameter estimation in rectangles of random dimensions and locations over the input image. In this case, all points with reliable normal flow values in the random rectangle contribute to the least squares solution. Both approaches were tested. With the first approach (linear system solution) a number of computationally cheap iterations are required. The second approach (least squares solution) requires fewer, but computationally more expensive iterations. Experimental results demonstrated that the overall computational performance is better for the second approach because of the significant reduction in the number of iterations required. Experimentation has revealed that in this second case, a selection of $m = 20$ for $p = 3$ and $m = 150$ for $p = 6$ works satisfactorily for real world scenes.

Another algorithmic improvement is achieved through avoiding the use of sorting to compute the median in each of the m iterations. Instead, we use an algorithm that selects the kth largest number out of n numbers, originally suggested in [43]. 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 $O(mn)$.

Furthermore, the performance of LMedS can be greatly improved by exploiting parallel processing techniques. Each of the m iterations does not depend on the outcome of another iteration. All candidate solutions could be evaluated in parallel. An extra stage for comparing the partial results is needed, which amounts to the problem of finding in parallel the minimum in a set of numbers. Going one step further, in each of the m iterations, the computation of the residual for each observation does not depend on the computation of the residual for another observation. Therefore, the n residuals can be computed in parallel.

The motion segmentation of the dominant depth layer is post-processed towards enhancing its usability. The post processing is again a local operation for which efficient parallel algorithms may be developed.

One interesting computational characteristic of the overall method for independent motion stems from the computational nature of LMedS. Through the various iterations over possible parameter sets, LMedS keeps the best solution found thus far (the one that minimizes the median of the residuals). Therefore, the whole algorithm can be viewed as an *any time algorithm* (i.e. an algorithm that continuously improves a given solution to the problem). This characteristic is very important because the execution time of the algorithm can be appropriately adjusted, taking into account the resource limitations of the system and the hard real time constraints that should be met.

Additional computational savings can be gained by considering a stereo configuration that does not change over time. In such a case, after few time instances, a robust estimate of β_s can be achieved. Given this estimate, the 3-parameter model of eq (7) can be reduced to a simpler model for the phase of estimating the dominant depth layer, which reduces the computational requirements for the stereo processing.

6   Experimental results

The proposed method for independent motion detection has been tested using both synthetic data and real world image sequences. Sample results from these experiments are reported in the following sections.

6.1   Simulation results

In order to facilitate the experimental evaluation of the proposed method, a simulation environment has been built. This environment enables the creation of synthetic normal flow fields (both stereo and motion) for a certain scene. The simulator can be used to define the focal length of the hypothesized cameras, the parameters of the stereo configuration, the noise level of the resulting normal flow fields and the number of rigidly moving regions. For each rigidly moving region, one can define its location and

dimensions on the image (rectangular regions are assumed), the mean and the variance of the Gaussian distribution that models the depth of the points in that region, the density of the normal flow field in the region, and its 3D motion parameters. Given the above parameters, a very large variety of scenarios can be effectively simulated. The output of a simulation is one possible normal flow field that can be due to the scene structure and the motion parameters. At each image point, the simulator assumes a random gradient direction which is selected from a uniform distribution in the range $[0, .., 2\pi)$.

A set of experiments has been carried out in order to evaluate the performance of *IMDRR*, as a function of the noise in the motion and stereo normal flow fields. The proposed method has also been compared to a motion segmentation method that employs a 2D motion model. This method has very close resemblances to the one proposed by Ayer et al [6], and tries to estimate the parameters of a 2D affine model by employing robust regression on a motion normal flow field.

In order to evaluate the performance of *IMDRR*, a synthetic motion and stereo normal flow field has been constructed by using the simulation environment. The synthetic flow fields refer to 256×256 images. A focal length of 600 pixels has been assumed for both cameras of the simulated stereoscopic observer. The cameras have been arranged in a parallel stereo configuration, with a 7cm baseline. The normal flow values in 50% of the points have been rejected, simulating the rejection of normal flows due to small image gradient. The simulated scene contains three areas of interest. The layout of the scene[4] can be seen in Fig 6. The green and the blue regions correspond to areas of the static environment, but differ in their average depth. The green area is located at approximately 6m from the observer, while the blue area is at approximately 3m from the observer. The red region corresponds to an independently moving object, which is at the same depth with the green area (distant background). Thus, an independent motion detection algorithm should produce a common label for the points of the green and blue regions (egomotion) and another label for the points of the red region (independent mo-

---

[4]Note that the simulation does not create synthetic images, but synthetic normal flow fields. Thus, Fig 6 is given to illustrate the layout of the regions in the hypothesized scene and it is not relevant to the image intensities.

tion). The independently moving object covers 23% of the total area of the scene and the close-to-the-observer object covers 25% of the scene. Note also that the independently moving object covers the 30% of the scene points that are at the dominant depth. The observer has been assumed to perform a complex translational and rotational motion with parameters $(U_e, V_e, W_e) = (60.0, 60.0, 6.0)$ and $(\alpha_e, \beta_e, \gamma_e) = (0.001, 0.0, 0.0001)$, while the relative motion between the observer and the independently moving object is $(U_i, V_i, W_i) = (4.0, 40.0, 80.0)$ and $(\alpha_i, \beta_i, \gamma_i) = (0.002, 0.0002, 0.0001)$.

Various simulations were performed, each with different noise added to both motion and stereo normal flow fields. In all cases, Gaussian noise with zero mean has been hypothesized. The standard deviation σ_n was set relative to the average magnitude $\overline{M}_{NF}$ of the normal flow vectors in the whole scene. In different runs, σ_n varied from 0.0 (noise-free case) to $0.48 \cdot \overline{M}_{NF}$ (highly contaminated data). In each experiment, both the motion and the stereo normal flow fields were affected by the same type of Gaussian noise.

Figures 7(a) to 7(i) illustrate the results of the 2D motion segmentation method for the above noise distributions. Each of the images of Fig 7, is a map with dimensions equal to the dimensions of the input images. Each point in this map corresponds to a point in the image and takes one of three possible labels: Black, white and dark gray, corresponding to egomotion, independent motion and points with rejected normal flows, respectively. The results of Fig 7 are characteristic of the inherent weakness of the 2D motion models when applied to scenes with large depth variations. More precisely, the method recognizes the object that is close to the observer as independently moving, regardless of the level of noise. Moreover, when the noise becomes high, the method fails to capture the real independent motion. This is because the apparent motion of the independently moving object is (in terms of the affine model) more similar to the apparent motion of the distant background than it is to the static foreground object.

The results of the application of the *IMDRR* method to the same data set are shown in Figs 8(a) through 8(i). As can be observed, the correct 3D motion characteristics of the scene are captured by this method. The method recognizes the independent motion of the distant object and fails only when the noise of the normal flow fields (for both stereo and motion) becomes extremely high ($\sigma_n \geq 0.48 \cdot \overline{M}_{NF}$). Note that

20

the points of the foreground object have been correctly recognized as such after the application of LMedS to the stereoscopic data. Therefore, they were rejected from further consideration; this is indicated in Fig 8 by the gray color that shades these points.

6.2 Experiments with real world sequences

IMDRR has been implemented on TALOS, the mobile robotic platform available at the Computer Vision and Robotics Laboratory (CVRL) of FORTH. TALOS includes:

- A mobile robotic platform (equipped with a 486 and a PENTIUM processors running Linux, Radio Link communications, sonar, infrared, and tactile sensors).

- A binocular, active vision head (independent control of pan, tilt, left and right vergence).

Figure 9 shows a picture of TALOS. The system is configured so that the PENTIUM processor is responsible for vision processing and control of the head, while the 486 processor controls the motion of the robot as well as the sonar, infrared and tactile sensors.

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 for the intrinsic camera parameters, the stereo configuration parameters and the observer's egomotion were unknown. However, this does not affect the outcome of the method, since LMedS differentiates between model inliers and outliers without using any prior knowledge of the actual model parameter values.

As a testbed for evaluating the performance of *IMDRR*, the "cart" sequence (512 × 384 images) has been employed. One frame (the left from the stereo pair) of this sequence is shown in Fig 10(a). In this image sequence, the observer performs a translational motion with U and W components as well as with a rotational β component[5].

---

[5]It is noted that in this, as well as in all other experiments held, the indicated egomotion parameters are the dominant ones; since no calibration has been applied, egomotion may also have minor translational and rotational components with respect to the axes that are not mentioned explicitly.

The horizontal translation is the motion that dominates. The field of view consists of 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.

Figures 10(b),(c) illustrate the motion segmentation results of the 2D method for the "cart" sequence. In Fig 10(b), the same coding of colors is employed as in previous results; gray color corresponds to points where normal flow values have been rejected, black color corresponds to the model inliers and white color corresponds to the model outliers, i.e. points where independent motion has been detected. Figure 10(c) shows the results of postprocessing of Fig 10(b), in which independently moving objects appear as in the original image while the rest of the scene is masked-out. As can be verified, the method is able to detect the independent motion that is present in the distant background. However, it also detects the toy car and the table as independently moving, although they belong to the static foreground. This is because the 2D method does not exploit any information regarding depth. It actually assumes that all points in the scene are at an equal distance from the observer. Thus, the apparent motion of the toy-car which is due to its relatively small distance from the observer, is interpreted as independent motion.

Figure 11 shows the results of *IMDRR* when applied to this image sequence. The estimation of the dominant depth resulted in the depth map of Fig 11(a). In this figure, gray color corresponds to points where stereo normal flow values have been rejected, black color corresponds to depth inliers (dominant depth) and white color corresponds to stereo outliers. Robust regression within the dominant depth layer gives rise to the 3D motion segmentation map that is illustrated in Fig 11(b). Finally, the postprocessed results of Fig 11(b) are presented in Fig 11(c). This result demonstrates that *IMDRR* is capable of detecting the independent motion that is present in the scene, without misinterpreting the apparent motion due to depth variation as independent motion.

The results presented in Fig 11 have been obtained after off-line processing of the

acquired images. However, since *IMDRR* is intended for on-board use on the robotic platform TALOS, a set of experiments has also been conducted to test the on-line performance of the method. In all these experiments the on-board PENTIUM processor has been used; due to the limited processing power and for reducing the overhead for image acquisition, processing and dumping of the results on disk, the image size has been kept small, namely 144×106. Moreover, the number m of iterations of LMedS for motion segmentation (refer to section 5 and Fig 5(b)) has been kept relatively small, namely $m = 100$. However, this did not affect the motion segmentation results since the spatial extent of the motion was rather small.

A sample result from these experiments is presented in Figs 12 and 13 for demonstration purposes. Fig 12 shows twelve frames of a sequence that correspond to the left frames of the stereo pairs. These frames are not consecutive in time but show intermediate snapshots of the whole sequence. As can be observed, in the scene in view there is a man who is initially sitting on a chair (right of the scene). The man then stands up, moves to the cart, takes it to the leftmost part of the room and then returns to his initial position. Meanwhile, TALOS is moving with W and U components (i.e. approaches the scene and also moves to the left). Figure 13 shows the motion segmentation (after post processing) that has been achieved by *IMDRR*. This figure demonstrates clearly the correct segmentation that has been achieved. It is also worth noting that the small table in the foreground (bottom-left of the scene) has not been misinterpreted as moving, although it is placed at substantially different depth from the rest of the scene.

7 Conclusions

In this paper, the *IMDRR* method for independent 3D motion detection has been described. Robust regression and, more specifically, LMedS constitutes an essential part of the method. Independent motion detection is achieved by the combination of motion information with stereoscopic information acquired by an uncalibrated stereo 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 experimental results presented serve as an indication of the robustness of *IMDRR*, as well as of its capability to handle multiple and possibly non-rigid motions. *IMDRR* can also be used as a preprocessing stage to other modules that are able to accurately estimate the parameters of one rigid motion [26].

From a computational performance point of view, although LMedS estimation has non-trivial computational requirements, *IMDRR* can be speeded-up by various ways that were discussed in section 5. Moreover, by appropriately adjusting the image size and the computational requirements of LMedS, an on-board implementation of *IMDRR* has been demonstrated.

The motivation behind *IMDRR* is to provide robust 3D motion segmentation by employing the minimum possible assumptions about the external world and the observer. As a consequence, information representation remains close to the data, and *IMDRR* may become a robust building block to other visual capabilities of an observer navigating in a three dimensional dynamic environment. Under such a perspective, the basic weakness of *IMDRR* is that, at least theoretically, 50% of the scene points should constitute a frontoparallel plane. Current research [44] is targeted towards alternative exploitations of stereo information in order to remove this constraining assumption.

Acknowledgements

REFERENCES

[1] A. Horridge. The Evolution of Visual Processing and the Construction of Seeing Systems. In *Proc. of the Royal Society, London B 230*, pages 279–292, 1987.

[2] J.Y.A. Wang and E.H. Adelson. Representing Moving Images with Layers. *IEEE Transactions on Image Processing*, 3(5):625–638, September 1994.

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

[4] M. Irani, B. Rousso, and S. Peleg. Computing Occluding and Transparent Motions. *International Journal of Computer Vision*, 12(1):5–16, 1994.

[5] Y. Huang, K. Palaniappan, X. Zhuang, and J.E. Cavanaugh. Optic Flow Field Segmentation and Motion Estimation Using a Robust Genetic Partitioning Algorithm. *IEEE Transactions on PAMI*, 17(12):1177–1190, December 1995.

[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] 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.

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

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

[10] R. Bajcsy. Active Perception. *Proceedings of the IEEE*, 76(8):996–1005, August 1988.

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

[12] D. Ballard and C. Brown. Principles of Animate Vision. *Computer Vision, Graphics and Image Processing*, 56(1):3–21, 1992.

[13] 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.

25

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

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

[16] 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.

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

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

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

[20] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[21] P. Anandan. A Computational Framework and an Algorithm for the Measurement of Visual Motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[22] A. Del Bimbo, P. Nesi, and J.L.C. Sanz. Optical Flow Computation Using Extended Constraints. *IEEE Transactions on Image Proc.*, 5(5):720–739, May 1996.

[23] R.Y. Tsai and T.S. Huang. Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. *IEEE Transactions on PAMI*, PAMI-6(1):13–26, January 1984.

[24] M.E. Spetsakis and Y. Aloimonos. Optimal Motion Estimation. In *IEEE Workshop on Visual Motion*, pages 229–237, 1989.

[25] G. Adiv. Inherent Ambiguities in Recovering 3D Motion and Structure from a Noisy Flow Field. *IEEE Transactions on PAMI*, PAMI-11(5):477–489, May 1989.

[26] C. Fermuller. *Basic Visual Capabilities*. PhD Dissertation, Center for Automation Research, University of Maryland, 1993.

[27] Y. Aloimonos and Z. Duric. Estimating the Heading Direction using Normal Flow. *International Journal of Computer Vision*, 13(1):33–56, 1994.

[28] D. Sinclair, A. Blake, and D. Murray. Robust Estimation of Egomotion from Normal Flow. *International Journal of Computer Vision*, 13(1):57–69, 1994.

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

[30] W. Burger and B. Bhanu. Estimating 3-D Egomotion from Perspective Image Sequences. *IEEE Trans. on Image Processing*, 12(11):1040–1058, Nov. 1990.

[31] K. Pahlavan and J.O. Eklundh. A Head-Eye System - Analysis and Design. *CVGIP: Image Understanding, Special Issue on Purposive, Qualitative Active Vision*, 56:41–56, 1992.

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

[33] 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.

[34] C.V. Stewart. MINPRAN: A New Robust Estimator for Computer Vision. *IEEE Transactions on PAMI*, 17(10):925–938, 1995.

[35] R. Kumat and A.R. Hanson. Robust Methods for Estimating Pose and a Sensitivity Analysis. *CVGIP:Image Understanding*, 1994.

[36] 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.

[37] P.J. Huber. *Robust Statistics*. John Wiley and Sons Inc., New York, 1981.

[38] C.C. Weems. Architectural Requirements of Image Understanding with Respect to Parallel Processing. *Proceedings of the IEEE*, 79(4):537–547, April 1991.

27

[39] Q.F. Stout. Mapping Vision Algorithms to Parallel Architectures. *Proceedings of the IEEE*, 76(8):982–995, August 1988.

[40] R. Deriche. Fast Algorithms for Low Level Vision. *IEEE Transactions on PAMI*, 12(1):78–87, January 1990.

[41] P.J. Hatcher et al. Data Parallel Programming on MIMD Computers. *IEEE Transactions on Parallel and Distributed Systems*, 3(2):377–383, July 1991.

[42] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*, chapter 7, pages 418–420. Addison-Wesley, third edition, 1992.

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

[44] 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.
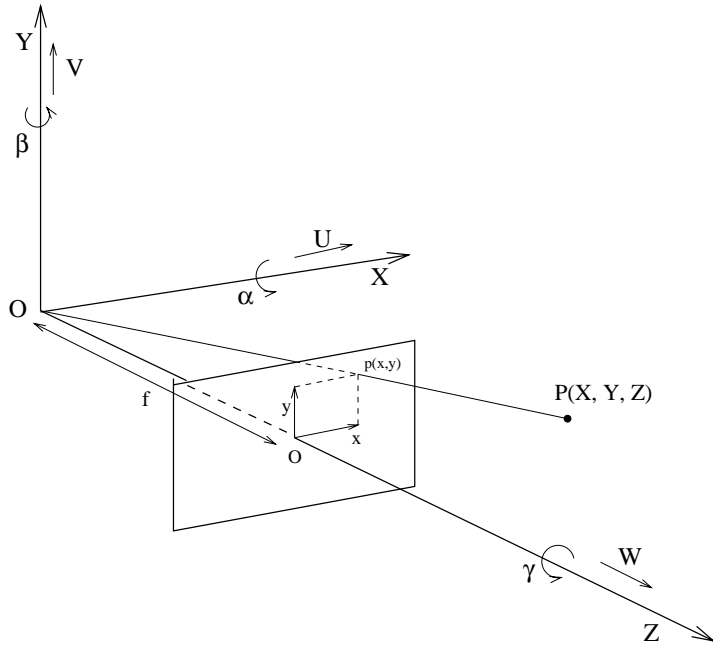
Figure 1: The camera coordinate system.
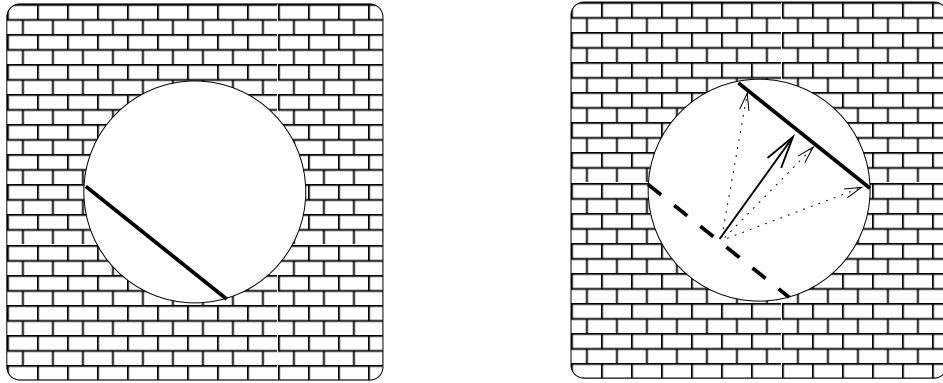
Figure 2: A schematic view of the aperture problem. The solid line in the left image has moved to a new position in the right image. Based on the information that is visible through the aperture, it is not possible to decide which of the dashed vectors corresponds to the motion vector of the line. However, whatever the motion vector may be, its projection to the direction perpendicular to the line is unique and is represented by the solid vector.
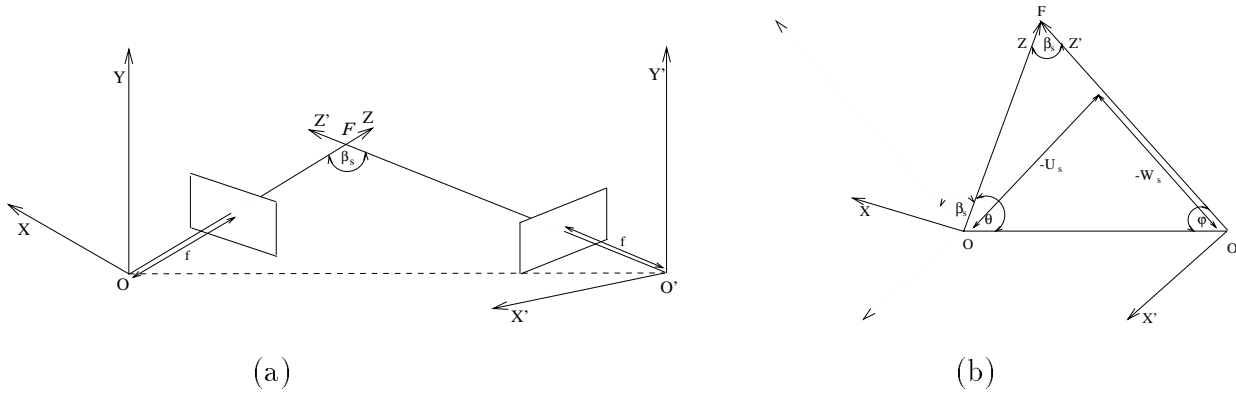
(a)   (b)

Figure 3: (a) The geometry of a fixating stereo configuration, (b) the parameters of the motion that transforms the position of the left camera to the position of the right camera.
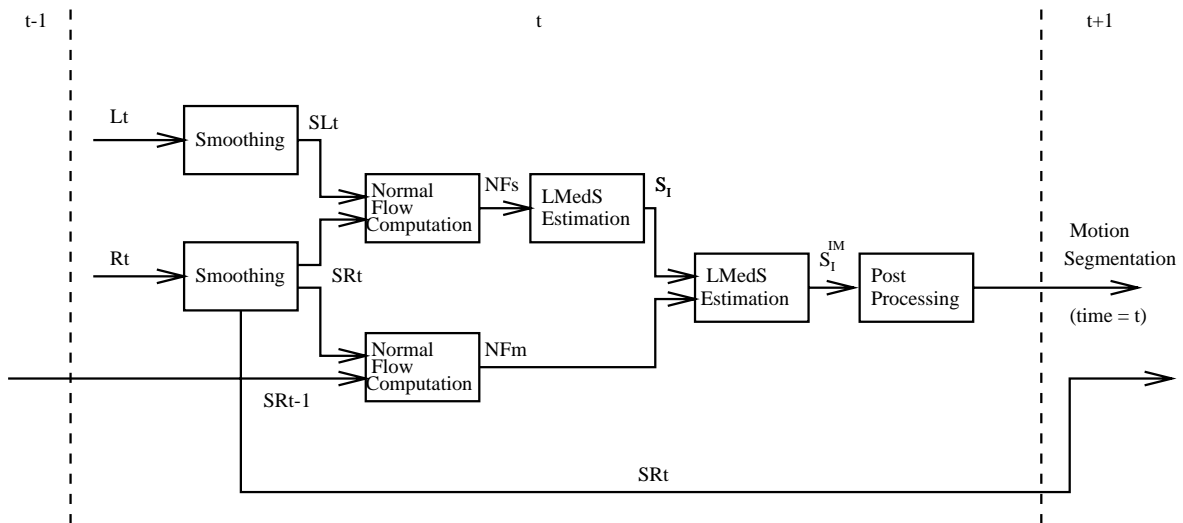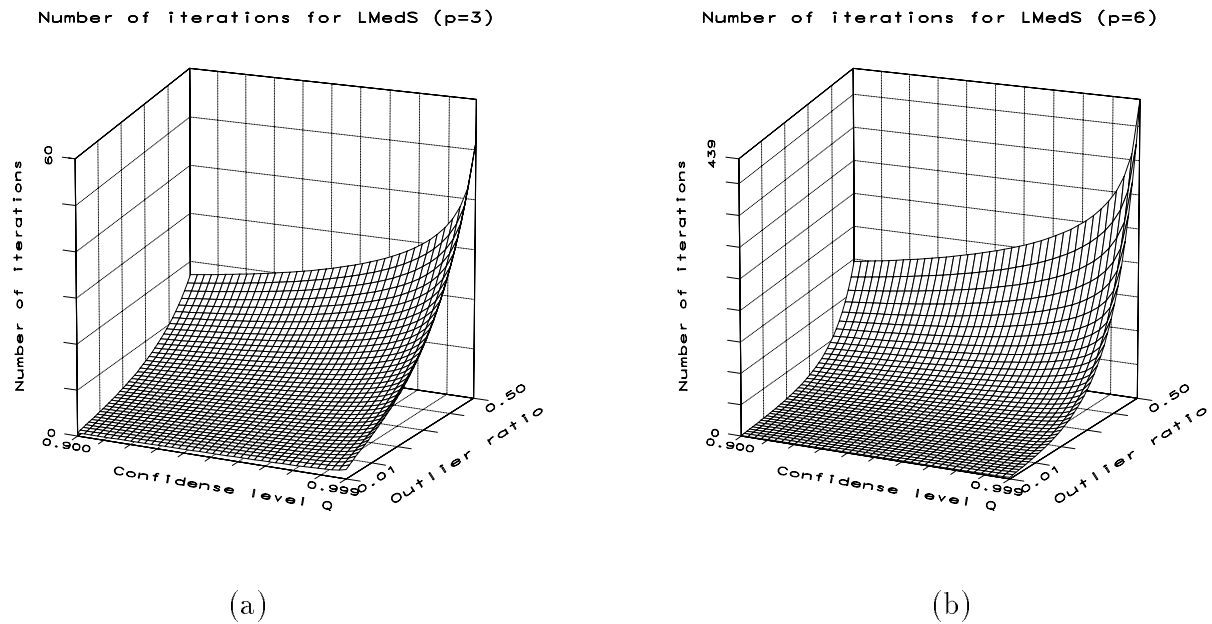

Figure 4: Block diagram of *IMDRR*.

(a)   (b)

Figure 5: Number of iterations m for LMedS estimation (eq (16)), as a function of Q and e. The number of model parameters is kept equal to (a) p = 3, and (b) p = 6.
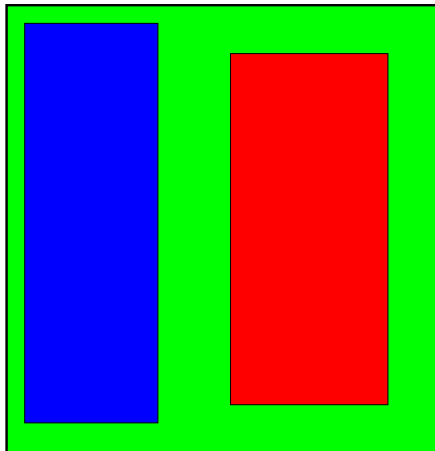

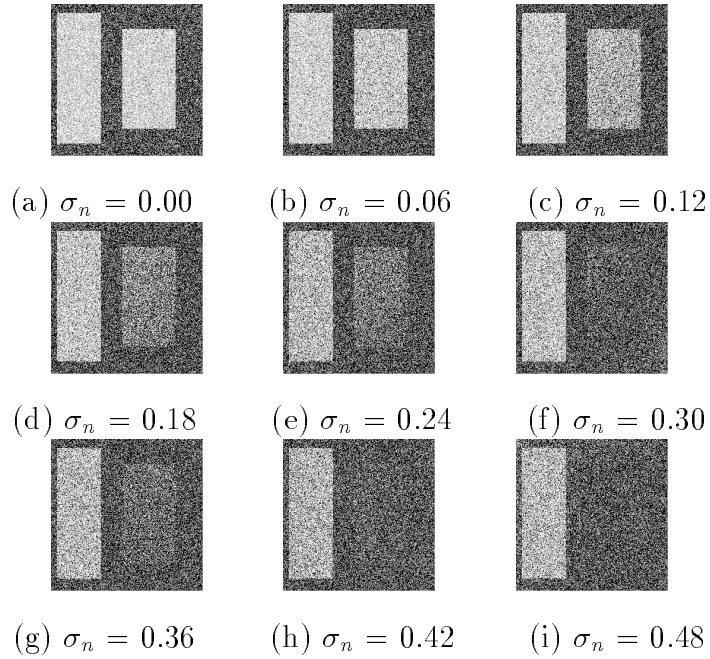Figure 6: The layout of the scene used for the evaluation of *IMDRR* under different levels of noise.

(a) $\sigma_n = 0.00$ (b) $\sigma_n = 0.06$ (c) $\sigma_n = 0.12$

(d) $\sigma_n = 0.18$ (e) $\sigma_n = 0.24$ (f) $\sigma_n = 0.30$

(g) $\sigma_n = 0.36$ (h) $\sigma_n = 0.42$ (i) $\sigma_n = 0.48$

Figure 7: Results of 2D independent motion detection for different levels of noise.

(a) $\sigma_n = 0.00$ (b) $\sigma_n = 0.06$ (c) $\sigma_n = 0.12$

(d) $\sigma_n = 0.18$ (e) $\sigma_n = 0.24$ (f) $\sigma_n = 0.30$

(g) $\sigma_n = 0.36$ (h) $\sigma_n = 0.42$ (i) $\sigma_n = 0.48$
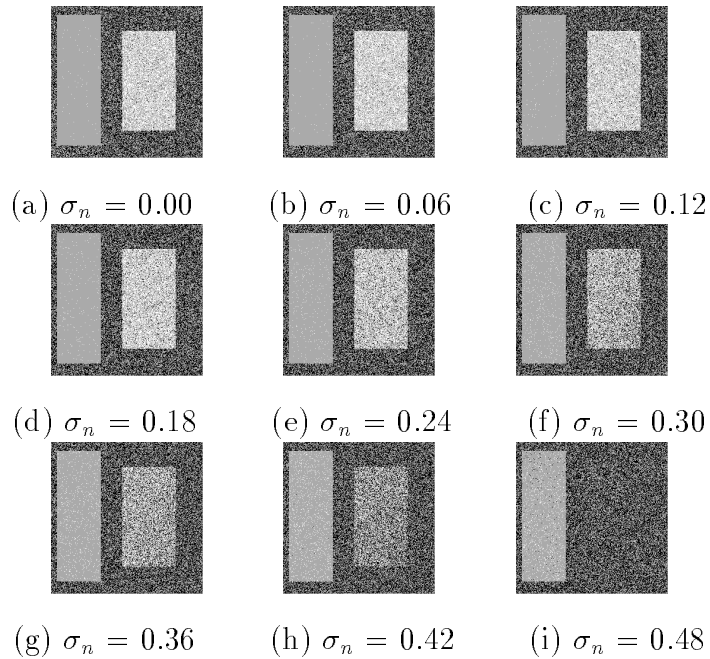
Figure 8: Results of *IMDRR* for different levels of noise.

Figure 9: TALOS, the multisensor mobile robot of ICS FORTH.

(a)   (b)   (c)
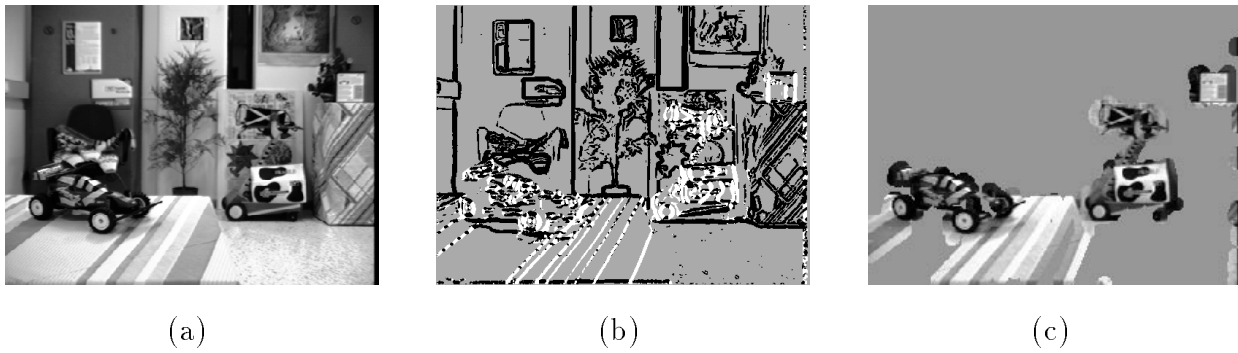
Figure 10: (a) One frame of the "cart" sequence, (b),(c) 2D motion segmentation results; (b) before and, (c) after postprocessing.
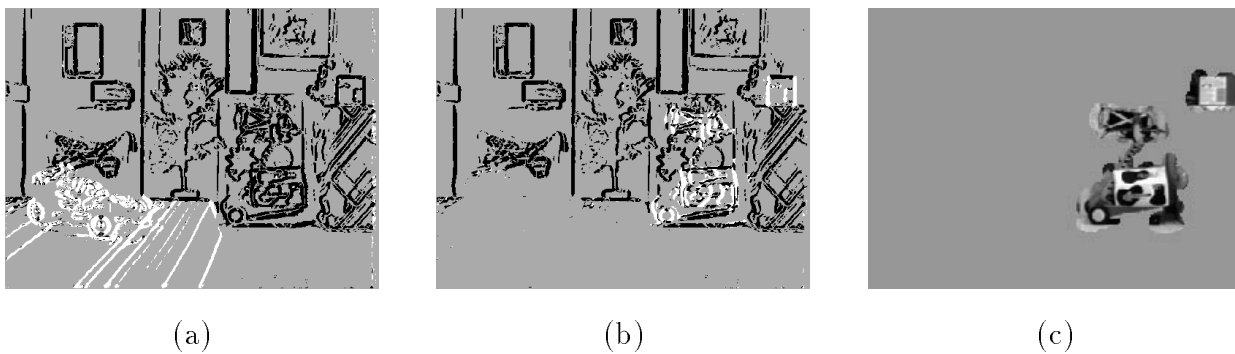

(a)   (b)   (c)

Figure 11: *IMDRR* results for the "cart" sequence; (a) Dominant depth estimation, (b) the outliers of motion segmentation within the dominant (distant) depth and, (c) motion segmentation after postprocessing.

Figure 12: Twelve images from the "moving man" sequence

Figure 13: Results of *IMDRR* for the "moving man" sequence