Estimation of Complex Multimodal Motion: An Approach based on Robust Statistics and Hough Transform.

Miroslaw Bober and Josef Kittler
Department of Electronic and Electrical Engineering,
University of Surrey, Guildford GU2 5XH, United Kingdom

Abstract

An application of Robust Statistics in a Hough Transform based motion estimation approach is presented. The algorithm is developed and experiments are performed, proving its superior performance in terms of estimate accuracy, convergence, robustness and better segmentation. Comparative results with standard methods are also included.

1 Introduction

A substantial body of work in computer vision has been concerned with motion analysis. The diverse range of applications includes autonomous navigation, the recovery of 3D structure of the scene, data compression and object tracking. Each of the prospective applications requires a robust approach that can reliably extract motion information from real sequences. This means that an algorithm has to cope with multiple objects moving with independent complex motions across a non-stationary background. In addition the possible range of displacements may be large, and the noise level significant. Yet the majority of the studied algorithms make unrealistic assumptions, limiting the number of objects in the scene, or assuming coherent translational pixel displacements. The purpose of this paper is to present an approach in which the application of Robust Statistics and the Hough Transform technique results in a robust approach capable of handling complex real sequences.

The problem of extracting motion information (i.e. motion estimation and segmentation) can be viewed from the following perspective. Each pixel on the image plane undergoes a certain 2D displacement that is a projection of some 3D motion. If a 2D correspondence is known (i.e. the correspondence between pixels in the reference and consecutive frames), the displacement vector can be easily computed. Such a correspondence can be established using the assumption that pixel grey-level is preserved in time. The extraction of motion information involves usually two main steps (though they may be performed in reverse order or in parallel): i) finding the displacement for each pixel and ii) grouping pixels moving with coherent motion into objects.

The first step corresponds to recovery of so-called optic flow while in the second step the flow is segmented into regions moving with coherent motions. The optic flow recovery is an ill-posed problem. No matter what method is applied, it may

BMVC 1993 doi:10.5244/C.7.24

fail for some pixels because of one or more of the following reasons. Firstly the assumptions about the preservation of pixel gray-levels may be violated due to noise, changes of illumination, reflections, or shadows. Secondly there may be more than one solution because of the aperture problem in non-textured regions. Finally, in case of occluded or uncovered background a pixel may not be present in one of the frames. In all these examples an additional constraint is necessary in order to regularise the problem.

Most of the approaches that explicitly recover optic flow use the notion of smoothness as a regularisation constraint [8]. In consequence, flow is smoothed out and significant errors are introduced along the boundaries between differently moving regions. Subsequent segmentation is then difficult or even impossible to perform. Some approaches try to inhibit the smoothness constraint across the motion boundaries, although this is difficult in practice because motion segmentation is seldom known a priori. For example Nagel proposed 'oriented smoothness constraint' [12]. Other researchers use Markov Random Fields incorporating line process and include additional information relevant to segmentation (e.g. intensity edges) to estimate translational [7] or complex [4] motions. These techniques tend to be slow in convergence and are sensitive to the selection of the model parameters.

The local smoothness constraint can be replaced by a global model of the optic flow. Bergen et al.[2] described a general framework introducing a family of global and local motion models that constrain the overall structure of the optic flow. They report a number of experiments in which a complex flow structure was successfully recovered. However the method may fail if multiple moving objects are present in the scene. Researchers have also employed the Hough Transform techniques for the interpretation of optical flow fields [1][11] and for parallel estimation and segmentation of complex motions [14]. For example, Adiv [1] firstly partitions flow into connected segments of coherent motion, which are later grouped into objects. A multipass technique is used to cope with multiple objects. The process is however slow (two explicit 3D Hough transforms are computed) and errors in optic flow (e.g. oversmoothing) propagate on segmentation. In [14] estimation and segmentation are performed in parallel, thus constraining each other to give better results. Since the Hough Transform is formulated as an optimisation problem and computed only implicitly the method is fast. However our experimentation proved that the method may fail if the scene contains many moving objects.

In this paper we describe an approach which uses robust statistics and Hough transform techniques. The combination of these techniques and a family of advanced global motion models will be shown to result in an algorithm that is superior in terms of accuracy and robustness. The paper is organised as follows. In the next section we outline the mathematical formulation of the model and develop the algorithm. Section 3 describes several experiments on real and simulated data, and provides some comments on the performance. Finally, conclusions are drawn in Section 4.

2 The Robust Hough Transform for Motion Extraction

This section provides an outline of our Robust Hough Transform (RHT) approach for motion analysis. We begin with a short discussion on the selection of optimal motion model. Then we present the principles of Hough Transform based motion

estimation and show how this idea is implemented by various researchers. We show that the existing implementations may sporadically fail to converge and that the motion estimate may be biased. We apply the Robust Statistics theory to address these drawbacks and propose an algorithm with dramatically improved performance. Finally we explain important aspects of the implementation such as the hierarchical approach, multiresolution search and the segmentation strategy.

2.1 Motion Model

It was mentioned above that motion estimation requires a regularisation constraint and consequently all algorithms make some additional assumptions about the structure of the motion. The estimate can be constrained either locally (e.g. smoothness) or globally within a larger region. Estimates based on large regions are more accurate and robust, provided that the motion within the region is uniform and that the motion model is flexible enough to describe it. However complex models require large regions to extract motion parameters accurately and are more computationally demanding. Therefore the selection of the motion model, which is a tradeoff between complexity and flexibility, depends on the application. For example, the algorithm estimating the velocity of the objects (cars) from the side view may employ a purely translational model with good results. On the other hand such restriction cannot be used for unknown scenes. In our approach we employ an affine motion model in which pixel positions in the reference frame $p(x, y)$ and consecutive frame $p'(x', y')$ are related by the following equation:

$$p' = T_{\vec{a}}(p) = \quad p + (a_1 x + a_2 y + a_3, a_4 x + a_5 y + a_6) \tag{1}$$

where vector $\vec{a} = (a_1 ... a_6)$ represents the model parameters. The above model is capable of handling translation, rotation, change of scale and shear. We believe that it gives an optimal balance between complexity and performance for most real sequences. However our approach need not be restricted to the above model, and therefore we use a general notation $T_{\vec{a}}(p)$ for motion transformation. In the next subsection we show how the Hough Transform can be tailored to efficiently and robustly solve the motion estimation and segmentation problem.

2.2 Hough Transform and motion analysis

The Hough Transform was originally proposed for the detection of parametric curves, e.g. lines or ellipses. Subsequently it was applied to a large range of machine vision problems including motion extraction. The HT is effectively a method segmenting feature points into groups satisfying some parametric constraint. It can be also considered as an estimation procedure in which parameter estimates are defined through the extrema of a function. For a comprehensive review of the HT refer to [10].

Let us assume that the pixel intensity is preserved, i.e.:

$$I_0(p) = I_1(p'); \tag{2}$$

where $I_0(p)$ and $I_1(p')$ are the grey-level intensities at pixel location p and p' in the reference and consecutive frame respectively. Pixel positions p and p' are constrained by the motion model $T_{\vec{a}}$. The displaced pixel difference is defined as:

$$\epsilon(\vec{a}, p) = I_0(p) - I_1(T_{\vec{a}}(p)) \tag{3}$$

In the standard Hough Transform a pixel p votes for a particular motion parameter vector $\vec{a}$ if it satisfies the condition: $|\epsilon(\vec{a}, p)| < T$ where T is a predetermined threshold. The points in the parameter space that collect large number of votes indicate motion of individual objects. We employ the 'HT by optimisation' approach [14],[15] in which the support h from pixel p is defined by a kernel function $\rho(\cdot)$:

$$h(\vec{a}, p) = \rho(\epsilon(\vec{a}, p)) \tag{4}$$

The total amount of support $H(\Re, \vec{a})$ received by the motion vector $\vec{a}$ from the region $\Re$ can be expressed as:

$$H(\Re, \vec{a}) = \sum_{p \in \Re} \rho(\epsilon(p, \vec{a})) \tag{5}$$

The motion parameter vector can be estimated by finding a minimum in function H. This requires an iterative minimisation procedure such as *steepest descent* or *conjugate gradient* methods. We employed the *steepest descent* method [15]. Partial derivatives of the support function H in the parameter space can be expressed in terms of the spatial gradients of the image intensity functions I_0 and I_1:

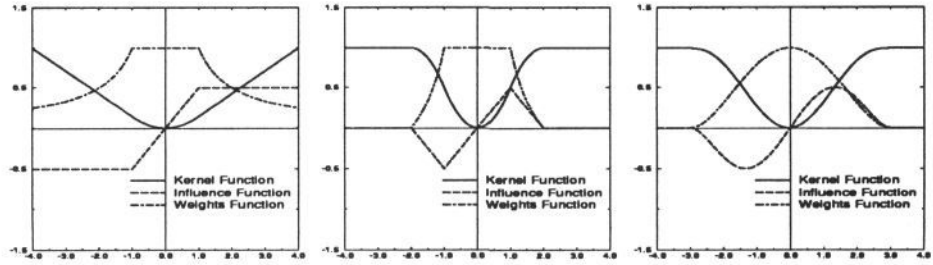
$$\frac{\partial H}{\partial a_i} = \sum_{i \in \Re} -\frac{\partial \rho(\epsilon)}{\partial \epsilon} \frac{\partial I_1(x', y')}{\partial a_i} \tag{6}$$

$$\frac{\partial I_1(x', y')}{\partial a_i} = \frac{\partial I_1(x', y')}{\partial x'} \frac{\partial x'}{\partial a_i} + \frac{\partial I_1(x', y')}{\partial y'} \frac{\partial y'}{\partial a_i} \tag{7}$$

where $i \in \{1, 2, 3, 4, 5, 6\}$.

When methods like *steepest descent* are used to solve optimisation problems, two issues become of a primary concern: the initial (starting) point and the convexity of the minimised function in the region between the starting point and global minimum. It can be shown that the support function H (eq. 5) is a well behaved function in the vicinity of the optimal motion vector 0 provided that the Taylor expansion is valid within the region. If long range motion is present, the aliasing of high spatial frequencies may cause the failure of the algorithm. We resolve this problem using hierarchical estimation [13], which is explained later in this section.

Many algorithms employ the above principle either explicitly [1], [14], [15] or implicitly [2]. However experiments prove that they tend to perform poorly or fail to converge on sequences containing few moving objects. Investigation of the reason underlying this sporadic failure of the algorithms revealed that the minima of the support function H are sometimes displaced from the position corresponding to true motion (Figure 3). The effect is amplified when the size of the objects is comparable and when a quadratic function is used as a kernel. Such behaviour can be explained on the grounds of the estimation theory. When quadratic error function is used, the minimisation of support function (eq. 5) effectively corresponds to the Least Square (LS) or mean estimator. The LS estimator has a number of limitations, the most important being its sensitivity to outliers. When objects are of comparable size, outliers may constitute half of the pixels. The use of absolute value as error norm produces the *median* estimator which is known to be more robust than the *mean* one. This fact is confirmed by experiments; however in some cases it was not robust enough. To overcome these problems we propose a robust estimator employing a redescending kernel, and experimentally demonstrate its dramatically improved properties.



(a) Huber Minimax (b) Hampel (1,2,1) (c) Tukey Biweight

Figure 1: Robust error norms and corresponding Influence and Weighting functions

2.3 Robust statistics

Robust Statistics is a branch of statistics that investigates the sensivity of statistical procedures to the violation of the underlying assumptions. The technical term 'robust' was first introduced by Box [3] in the fifties. In the sixties Tukey summarised earlier work in the field and demonstrated the drastic non-robustness of the least-square estimators. The solid mathematical foundations were introduced by Huber [9] and Hampel [5]. Here we define basic terms used in robust statistics theory and summarise the most important properties. More details can be found in textbooks (e.g by Hampel at al. [6]).

Two important terms concerning the performance of the estimator should be explained here: *efficiency* and *robustness*. *Efficiency* refers to the ability of a procedure to provide optimal estimates from data that fulfills the underlying assumptions (e.g assumptions made during the design of the method). *Robustness* reflects insensitivity to the assumptions being violated. The Least Square Estimator is efficient but non-robust. In this method the quadratic error term weights heavily the contributions to the 'optimal' solution from the data points which have a large residual errors (e.g. outliers).

Huber's class of 'M-estimators' are a generalisation of the Maximum Likelihood estimators. Given a set of N data samples $\{d_i, z(d_i)\}$ and a function constraining the structure of the data $z = f(\vec{a}, d)$ we estimate the optimal parameter vector $\vec{a}$ that minimises an error metric $H(\vec{a})$. The error metric H is usually the sum of the error norm (e.g. kernel) $\rho(\cdot)$ of the residual errors $\epsilon(d_i) = z(d_i) - f(\vec{a}, d_i)$:

$$H(\vec{a}) = \sum_{i=1}^{N} \rho(\epsilon(d_i)/s) \tag{8}$$

where s is a scale estimate. The robustness of the estimator is increased by modelling the kernel function $\rho(\epsilon(d_i)/s)$ so that the influence of the outliers (for which values of $\epsilon(d_i)/s$ are significant) is scaled down. Figure 1 illustrates examples of such designs, namely the *Huber Minimax* (a), *Hampel (1,1,2)* (b) and *Tukey Biweight* (c) error norms. Hampel [6] introduced the influence functions (IF) $\phi(\epsilon) = \partial\rho/\partial\epsilon$ as a convenient tool for analysing the behaviour of the variety of robust estimators. The influence functions for the kernels mentioned above are not scale-invariant and the spread of the data distribution has to be computed. A Median Absolute Deviation (MAD) robust scale estimator can be used for that

purpose [6]:

$$s(\vec{a}) \quad = \quad 1.4826 median\left(|h(d_i) - median(h(d_i))|\right) \tag{9}$$

The scale is based on the median of the absolute errors between the data points and the initial parameter estimate. The coefficient 1.4826 derives from the assumption that the model error terms are normally distributed random variables. Its value is equal to the ratio of the standard deviation to the median of absolute deviations from the mean of a Gaussian distribution.

2.4 Details of the implementation

In this section we explain important details of the implementation, namely the hierarchical strategy, multiresolution in parameters space and the segmentation algorithm.

The hierarchical approach has been used by various researchers. The key idea is to start estimation at the low (coarse) resolution in the image pyramid and then refine the estimate on the subsequent, finer levels. Such strategy improves the computational efficiency but, even more importantly, helps in convergence. This is particularly important if large displacements are involved. In this case the effect of aliasing of high spatial frequency components of the intensity pattern may prevent the convergence. In our algorithm the hierarchical approach is combined with a multipass strategy. Once a moving object is detected at the low resolution the estimate and motion segmentation are passed to the next level in the hierarchy. In several iterative steps the final motion estimate is recovered and all pixels conforming with this motion are marked as 'segmented'. The above procedure is repeated for pixels not labelled as 'segmented' until the majority of pixels have a motion parameter vector assigned to them.

The minimisation of the function H is performed on a discrete grid rather then in continuous space. This approach has two advantages. Firstly, it facilitates the detection of the local minima without the need for the computation of the support function values after each step. Secondly, computations are simplified because the step size need not be computed. In order to make the search process more efficient the parameter space resolution has several levels. Minimisation starts at a coarse grid and explores subsequently higher resolutions to obtain the desired accuracy.

The final motion segmentation takes place when motion vectors of all objects present in the scene are recovered. For each motion vector a_k we use the corresponding displaced frame difference DFD_k (smoothed with a Gaussian-shaped kernel) to compute a 'likelihood' image L_k:

$$L_k(p) = exp\{-0.5[DFD_k(x,y)/\sigma_k]^2\}; \tag{10}$$

where σ_k is the estimate of noise (which may depend on object) and may be computed from the robust estimate of the scale. An additional region $k = 0$ corresponds to unknown motion parameters (e.g. occluded or uncovered background) and has a constant likelihood value assigned to all pixels.

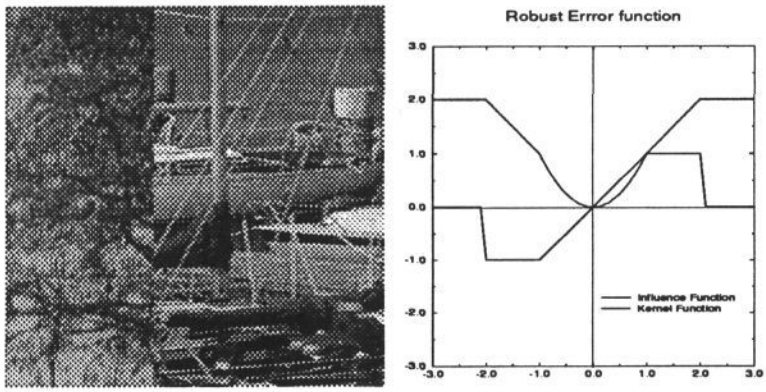
The assumption is made that each pixel belongs to one of the regions and likelihood functions L_k are normalised so that the following equation holds for each pixel p_i:

$$\sum_{k \in \{0,..,n\}} L_k(p_i) \quad = \quad 1 \tag{11}$$

Finally each pixel p_i is assigned to the region k with the highest probability value.

3 Results

In this section we present some experimental results. The first experiment shows how multiple moving objects can result in displaced minima in the Hough function causing the estimate to be biased. We also show that the problem can be corrected by introducing a redescending error norm. In the sequence two objects of a similar size are moving with purely translational motion. Translational motion was selected for the sake of easy visualisation. The comparable size of the objects (the actual ratio of the object areas is 1:1.3) simulates the worst-case scenario when the number of outliers is close to 50%. Objects are referred to as the left and right one. Figure 2(a) presents the reference frame. Three kernels were used during estimation: the Quadratic, Absolute and Hampel (1-2-2). Figure 2(b) shows the Kernel and Influence function for the latest norm.



(a) 'Gen2' Reference Image (b) Hampel (1,2,2) error norm

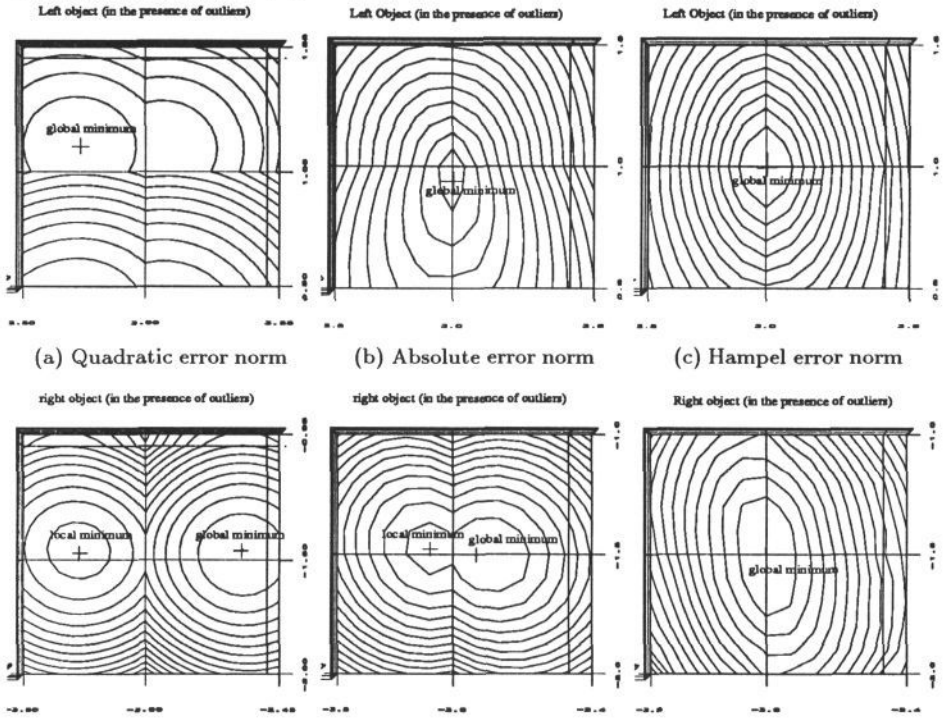
Figure 2: Reference image from Gen2 sequence and Hampel (1,2,2) error norm

The real motion parameters and the estimate based on each error norm is presented in Table 1. The estimates based on Quadratic and Absolute error norm are strongly biased by the presence of outliers. In the multipass procedure a biased estimate of the motion of one object propagates on segmentation (e.g. not all pixels belonging to the object are segmented out) and the algorithm may fail to recover the remaining objects. Indeed, such behaviour was observed during the experiment (marked as (*) in the table above). The Hampel (1-2-2) error function proved sufficiently robust to recover exact motion parameters even in this worst-case experiment.

		True	Quadratic	Absolute	Hampel (1-2-2)
left	Vx	3.00	2.68	3.00	3.0
object	Vy	1.00	1.30	0.96	1.00
right	Vx	-3.00	-2.89	-2.86	-3.01
object	Vy	-1.50	-1.47*	-1.50	-1.50

Table 1: **A comparison of the estimate accuracy for different kernels.**

Figures 3(a-f) depict the contour plots of the Hough function obtained with the above mentioned error norms. The first three plots focus around the minimum related to the left object, with the next three corresponding to the right one. It can be readily seen that for the robust Hampel (1-2-2) kernel ((c),(f)) the Hough space exhibits a nicely-shaped unbiased minimum. The application of non-robust error norms, not only bias the estimate but also result in local minima for the left object (Panels (d) and (e)).



(a) Quadratic error norm (b) Absolute error norm (c) Hampel error norm

(d) Quadratic error norm (e) Absolute error norm (f) Hampel error norm

Figure 3: Contour plots of the Hough space for different error norms

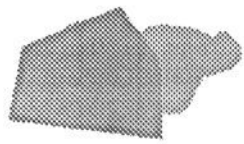
For our next example we consider a sequence with two objects moving past non-stationary background (Figure 4 (a)). The motion was artificialy generated so that the exact motion parameters are known (Table 2). The background was subject to a change of scale (with the focus of expansion $FOE = (120, 120)$ and change of scale $z = 0.98$) and translation (translation vector $T_v = (2.8, 3.2)$). The first object was rotated (angle of rotatation $\Theta = 10$ degrees and center of rotation $R_c = (150, 114)$) and translated ($T_v = (0.4, 1.2)$). The second object was scaled ($FOE = (132, 150)$, $z = 0.92$ and translated ($T_v = (0.25, 0.33)$). An affine motion model was used. The motion segmentation was satisfactorily recovered (4(c)) and motion parameters accurately estimated for each object (Table 2). This example demonstrates that the presented algorithm can successfully cope with complex multiple motions.

Finally we present an example of motion estimation from a real sequence depicting a view from the bridge on a A3 highway (Figure 5(a)). The sequence was taken using a hand-held camera, and therfore the background is not stationary.

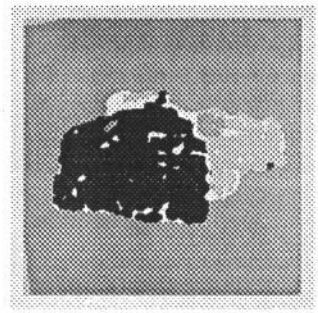
Object		Motion parameters					
		a1	a2	a3	a4	a5	a6
Background Region	True Motion	-0.020	0.000	5.19	0.0	-0.020	5.59
	Estimate	-0.020	0.000	5.01	0.0	-0.020	5.42
First Region	True Motion	-0.015	-0.173	22.47	0.173	-0.0151	-23.11
	Estimate	-0.060	-0.155	16.91	0.155	-0.060	-19.36
Second Region	True Motion	-0.080	-0.0	10.80	0.0	-0.080	12.32
	Estimate	-0.080	-0.003	9.47	-0.003	-0.080	11.95

Table 2: Ground truth and estimate of motion parameters

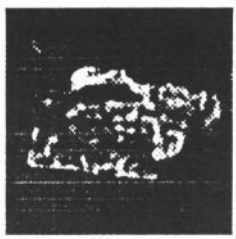
(a) Reference Image



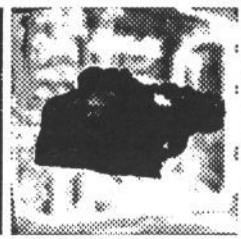
(b) True Segmentation



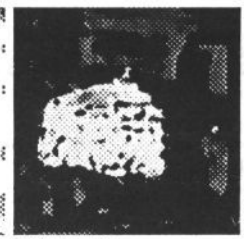
(c) Motion Segmentation



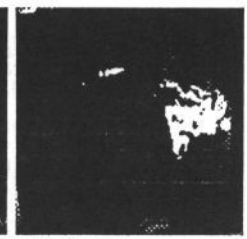
(d) Confidence map 'unknown' region



(e) Confidence map Background

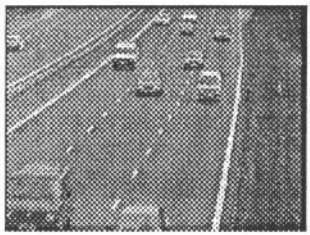


(f) Confidence map Region 1

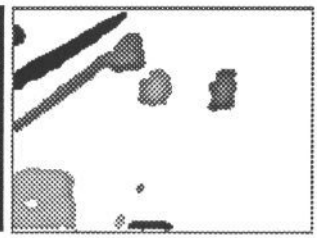


(g) Confidence map Region 2

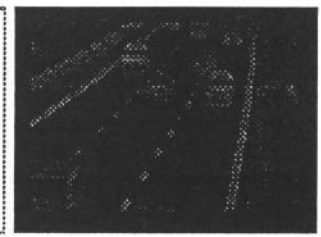
Figure 4: 'Two objects' sequence



(a) Reference Image



(b) Motion Segmentation



(c) TFD Image

Figure 5: A highway sequence

The motion of the cars is a combination of translation, zoom and the motion of the camera. Three motions were detected - the background, the first line of cars, and the second plane of cars (Fig. 5(b)). The DFD image is shown on Figure 5(c).

4 Conclusions

An approach to motion estimation and segmentation with dramatically improved robustness and accuracy has been presented. This significant improvement in performance is achieved by the use of robust redescending kernels and Hough Transform. Multiple moving objects on the non-stationary background are not a problem, since multipass strategy is used. The algorithm is fast because of the multiresolution in the parameter space. Moreover, the multiresolution in image space gives the algorithm the ability to cope with complex motions even when large displacements are involved. A family of motion models can be used, depending on the perspective applications. Finally, a new segmentation strategy based on probabilistic theory is proposed. Experimental results on generated and real sequences are presented. Our current research concentrates on optimisation of the algorithm and real time implementation on the 'DATA CUBE' parallel machine.

References

[1] Adiv G. *Determining three-dimensional motion and structure from optical flow generated by several moving objects*, IEEE Trans. PAMI-7, pp. 384-401, 1985.

[2] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. *Hierarchical model-based motion estimation*, ECCV1992, Italy, May 19-22, Springer-Verlag 1992, pp. 237-252

[3] Box G. E. *Non-Normality and tests on variance.* Biometrika 40, pp.318-335 1953.

[4] Bober M. Kittler J. *General Motion Estimation and Segmentation*, V Simposium Nacional de Reconocimiento de Formas y Analisis de Imagenes, Valencia, Spain, September 1992.

[5] Hampel F. R. *Contributions to the theory of robust estimation*' Ph. D. Thesis, Univ. of California, Berkeley, 1968.

[6] Hampel F. R., Ronchetti E., Rousseeuw P. and Stahel W. A. *Robust Statistics: The Approach based on Influence Functions*, John Wiley and Sons, 1986.

[7] Heitz F. and Bouthemy P. *Multimodal Motion Estimation and Segmentation Using Markov Random Fields*, 10th ICPR, Atlantic City, June 1990.

[8] Horn B.K.P. and Schunck B.G. *Determining optical flow*, Artificial Intelligence, Vol. 17, 1981, pp. 185-203.

[9] P J. Huber *Robust estimation of a location parameter*, Annals of Mathematical Statistics, 35:73-101, 1964.

[10] J. Illingworth and J. Kittler *A Survey of the Hough Transform*, CVGIP, 44, pp. 87-116. 1988

[11] Jayaramamurthy S. N. and Jain R. *An approach to the segmentation of textured dynamic scenes*, CVGIP, 21, pp.239-261, 1983.

[12] Nagel H. H. and Enkelmann W. *An investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields From Image Sequences*, IEEE Trans PAMI, Vol. 8, 1986, pp.565-593.

[13] Terzopoulos D. *Image Analysis Using Multigrid Relaxation Methods*, IEEE Trans. PAMI, vol. PAMI-8, No.2, March 1986 pp 129-139.

[14] Wu F. *General Motion Estimation and Segmentation*, PhD thesis, Dep. of Electronic and Electrical Eng., University of Surrey, Nov. 1990.

[15] Bober M. Kittler J. *A Hough Transform based Hierarchical Algorithm for Motion Segmentation and Estimation*, The 4th International Workshop on Time-Varying Image Processing and Moving Object Recognition, Firenze, June 1993. Elsevier.