A Correlation-Relaxation-Labeling Framework for Computing Optical Flow — Template Matching from a New Perspective

Qing X. Wu, *Member, IEEE*

Abstract—**Optical flow estimation is discussed based on a model for time-varying images more general than that implied in the work of Horn and Schunk [21]. The emphasis is on applications where low contrast imagery, non-rigid or evolving object patterns movement, as well as large interframe displacements are encountered. Template matching is identified as having advantages over point correspondence and the gradient-based approach in dealing with such applications. The two fundamental uncertainties in feature matching procedures, whether it is template matching or feature point correspondences, are discussed. Correlation template matching procedures are established based on likelihood measurement. A new method for determining optical flow is developed by combining template matching and relaxation labeling. In this method, a number of candidate displacements for each template and their respective likelihood measures are first determined. Then, relaxation labeling is employed to iteratively update each candidate's likelihood by requiring smoothness within a motion field. Real cloud images taken from meteorological satellites are used to test the usefulness of this method. It is shown in this application that the new method can deal effectively with the uncertainty of multiple peak (*multi-modal*) correlation surfaces encountered in template matching. The results show significant improvement when compared to that of the maximum cross correlation (MCC), which has been operationally used for cloud tracking, and to that of the method of Barnard and Thompson, which estimates displacements based on combining point correspondences with relaxation labeling.**

Index Terms—**Optical flow, dynamic scene analysis, motion estimation, cloud motion fields, template matching, cross-correlation, relaxation labeling.**

I. INTRODUCTION

THIS paper discusses one aspect of low level vision: determining optical flow (OF), the apparent motion from time-varying image sequences [21], [24], [34]. A new approach to computing OF is developed based on combining correlation template matching [28], [43] and relaxation labeling [36], [22].

It has been understood that determining OF, or equivalently, establishing correspondences of identical physical parts between images, is often the first step to acquiring information on 3D scene motion and structure [30], [44], [3]. On the other hand, OF, or image displacement field, is often the desired result itself. For example, cloud motion vectors from meteorological satellite images have been used to investigate atmospheric movement [27], [39], [25]. Video coding of moving picture sequences for efficient transmission and storage is another important direct application of OF estimation [29], [35].

Most previous work in the context of computer vision dealt mainly with scenes containing rigid objects and images from which prominent local features, such as edges, lines, and grey corners, can be easily identified. Images of evolving fluid motion, from which the above local features cannot be unambiguously determined, has not been discussed in depth within the computer vision community. Two predominant examples of evolving motion are the floating clouds and the movement of sea surface temperature patterns seen from meteorological satellite picture sequences [13], [23], [47]. It is worth emphasising the importance of motion tracking in environmental studies. Cloud motion vectors have been operationally derived since the late 1970s [39] and are important input parameters to global numeric weather prediction (NWP) models [25]. In this particular application, however, the correlation matching approach has been the only methodology adopted by operational agencies [1], [2].

For the last four years, we have been researching into developing new techniques for tracking cloud motion and sea surface currents [46], [47], [48], [49], [50], [51]. The new method described in this paper has been shown to produce significantly improved results compared to existing methods based only on correlation matching. This paper attempts to discuss in depth the theoretical background of the new method. In addition, we report new computational results.

In the remainder of this paper we first review existing techniques and discuss their difficulties with regard to the nature of environmental satellite imagery (Section II and Section III). We then relate the likelihood measurement to correlation matching procedures, and formulate the theory behind the new correlation-relaxation-labeling (C-R) method (Section IV and Section V). Some specifics in the implementation of the new procedure are given in Section VI, and in Section VII, results are illustrated and compared with that of two other relevant methods in support of discussions on key matching problems. The last section summarizes the paper and outlines possible future research direction to further improve the method.

Manuscript received July 12, 1994; revised Apr. 10, 1995; recommended for acceptance by A. Singh.

Q. X. Wu is with the Landcare Research Institute of New Zealand, P. O. Box 38491, Wellington Mail Center, New Zealand.
e-mail srghqxw@gih.grace.cri.nz.

IEEECS Log Number P95109.

II. EXISTING METHODS

Most existing techniques for determining OF may be classified into three categories: the gradient-based, the spatiotemporal energy-based, and the feature-based approaches.

A. Gradient-Based Approach

Methods from the gradient-based approach, also known as the differential approach, are based on the following simple image intensity conservation equation [21]:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (1)$$

where $(\delta x, \delta y)$ and δt are the inter-frame displacement and time interval respectively. The first order differential equation of (1) [21],

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (2)$$

can in principle be used to approximate the velocity component in the direction of a spatial gradient of the image intensity. Velocity solutions are typically obtained by combining the conservation equation (2) with a constraint [21], [34], [19], [24] that requires minimum departure from velocity smoothness.

Equations (1) and (2) are derived from two assumptions: Intensity conservation and small inter-frame motion and time interval. These two assumptions lead to two fundamental difficulties which limit the usefulness of the gradient-based approach. The first is that the image intensity is simply not conserved in most real world image sequences not only due to noise and distortions in illumination and 3D imaging geometry, but also due to changes in the scene itself. What is often partially conserved is in fact the structural relation of the intensity function [52], [11]. The second difficulty is that the approach requires the image intensity to be sufficiently smooth so that inter-frame change can be related by the linear relation of (2). This assumption is unrealistic when inter-frame motion is large. To ensure that inter-frame motion is small, i.e., short-ranged [10], [11], temporal sampling rate must be very high. Multi-resolution, coarse-to-fine strategies involving smoothing filtering preprocessing, as described by Glazer [15] and Enkelmann [14], have been used in this approach for increasing the motion range that can be estimated, and were shown to give improved results. These techniques, however, tend to produce oversmoothed flow fields, and it is easy to see that they are successful only under the condition that a large vector obtained at a coarse level is already approximately correct, and only finer adjustments are necessary. This condition corresponds to the *unimodal*, or single peak correlation surface, assumption used in correlation matching procedures, as discussed in Section III. It is noted that more recently Weng et al., [45] proposed a multi-attribute and multi-resolution algorithm, and it appears to have a positive impact on stereo matching, or OF determination in the case of stationary scene and moving camera. No results have been reported for images of a moving scene using this method.

In the case of environmental satellite imagery the above two assumptions are always severely violated. This is because:

1) the dynamic scene changes constantly such that object patterns not only translate and rotate, but also form and deform; and
2) images are formed under severe atmospheric interference in addition to other distortions; and
3) images cannot be acquired at very high temporal rate due to practical constraints, and consequently large inter-frame motion is unavoidable.

B. Spatiotemporal Energy-Based Approach

The more recently developed spatiotemporal energy based approach [5], [18] uses a family of spatiotemporal frequency-tuned Gabor energy filters to extract motion information. The method relies on the use of a large number (7 in [18]) of consecutive images to estimate object motion. The other limitation is that, like the gradient-based approach, it is only suitable for short-ranged motion [5]. While its application to a variety of real world images is to be investigated it appears that it is not suitable for our application for reasons stated at the end of last subsection.

C. Feature-Based Approach

The feature-based approach, contrary to the gradient-based approach, does not suffer, in principle, from the two fundamental difficulties because it tries to establish correspondences of invariant features between time-varying images. Points of high intensity variation, grey level corners [52], [9], [12], [41], as well as blobs [26] have been matched for determining displacement fields. Other features, such as edges and lines [32], [16], [20] have also been used for determining displacements or disparities of stereovision. Very often prominent features such as those mentioned cannot be easily and correctly identified. Images containing mainly non-rigid, low contrast features, such as those of clouds and oceans from meteorological satellites, are typical examples. In this case, segments of image intensity itself, often termed templates, are most frequently used directly as features to be matched in consecutive images [28], [43], [13], [47].

The main difficulties with feature-based approach lie in

1) the uncertainty in determining invariant features due to various image distortions such as those mentioned in Section II.A; and
2) the uncertainty in establishing feature correspondences between any two consecutive images in a sequence. These two uncertainties will be made clear in the next section in substantial detail since this paper concerns mainly the feature-based approach.

III. POINT CORRESPONDENCES AND TEMPLATE MATCHING

A representative method of using prominent points to establish a displacement field is developed by Barnard and Thompson (B-T) [9]. In their approach, a Moravec operator [33] is first used to allocate prominent intensity features points from two images, and then, feature points correspondences are established using a relaxation labeling technique stemming from the work of Rosenfeld et al., [36]. In the context of determining optical flow, the correspondence problem is solved by enforcing the vector field smoothness constraint within the relaxation labeling operation. They reported successful results in estimating displacements or disparities from images both of rigid-body-motion scenes and of stereo pairs. Several other researchers [12], [6], [46], [48] used similar techniques in their respective work.

We have previously applied this approach to estimating ocean surface current using sea surface temperature (SST) images from the advanced very high resolution radiometer (AVHRR) of the NOAA satellite [46], [48]. We have also employed this technique to compute cloud motion vectors from the infrared images of the Japanese GMS satellite (reported in this paper). Although on a large spatial scale the results of B-T show the expected flow patterns (see, for example, Fig. 4b), they invariably exhibit numerous artifacts of two or more neighboring vectors having the same end points. This means that these feature points in the first image of the pair correspond to a same point in the second image. Similar artifacts were shown in the results of Aisbett [6] using the method of B-T (Aisbett did not discuss the artifact). The cause of the error can be identified by observing the types of images used in both cases. In the case of Aisbett, the three pairs of images used were all of low contrast, and there were very few sharp points. In our own work, low contrast infrared images contain basically evolving fluid motion of cloud and ocean surface, sharp features barely exist. The problem is twofold:

- First, the determination of local features within a small neighborhood (5×5 pixels for Moravec operator) from this kind of images will generally be erroneous; and
- Second, only a small number of feature points can be extracted. Consequently, the use of a smoothness constraint in the relaxation labeling operation often unrealistically forces a few neighboring points in the first image to correspond to the same feature point in the second image, and bad matches follow.

This problem exemplifies the first difficulty in the feature-based approach: The uncertainty in detecting invariant features. Apparently this difficulty is not significant for images in which sharp features exist and can be determined using a small neighborhood (SN) operator. We refer to such features as SN features.

On the other hand, we observe that image features are still readily recognizable in low contrast time-varying images. This observation also applies to images of atmospheric and oceanic motion where many features are not only distorted but also gradually change their appearances (deformation) due to fluid dynamic processes. We argue that, in this case, human visual recognition is based on the intensity structure of a relatively large neighborhood (LN) in comparison with SN features. We term these features LN features.

The representation of LN features is a complicated problem. In principle it may be achieved through spline techniques but its usefulness may be limited. One reason is that such representation requires intensive computing power and is subject to severe error when images are distorted. Consequently, as mentioned earlier, a segment of the image intensity itself (sometimes its filtered version) from the image area just enclosing the LN feature is used directly. In this case, motion vectors are always computed through correlation template matching based on (dis-)similarity match measures [37].

In a template matching procedure, for example, in [28], [13], [47], a template from the first image is matched against a larger search area from the second image by shifting its position pixel by pixel over the entire search area. This results in a 2D match measure surface over all possible shifts. Traditionally, a displacement estimate of the template is the shift that either maximizes the match measure if it is a similarity measure, or minimizes it if it is a dissimilarity.

The most commonly used match measure is the cross-correlation coefficient (CCC), which is used in the well known maximum cross-correlation (MCC) method. It is popular because it corresponds to optimal signal-to-noise ratio estimation [37]. The sum of the absolute value of difference (SAVD) and its square, the sum of squared difference (SSD), both dissimilarity measures, have also been used. Their usage is usually justified on the ground that they are easy to implement and use less computing power, especially, when they are used in the fast sequential similarity detection algorithm (SSDA) [8].

The main difficulty with the MCC method is the existence of two uncertainties in searching for the correct match position through peak detection. Ideally, one would expect to find a well defined maximum, i.e., a single sharp peak in a correlation surface. This is the *unimodal* assumption that underlies the MCC method. In practice, however, we often encounter multiply peaked surfaces due to image noise, distortion, as well as limited spatial sampling and quantization resolutions. In other words, correlation surfaces are generally *multimodal*. Even when the unimodal assumption is valid, correlation surfaces may contain single but rather broad peaks, which is the well known *aperture problem* [52]. In both cases, erroneous estimation will occur if one simply selects the maximum. In the case of very low contrast images containing mainly LN features, for example, a template can often be matchable to different parts of a large search area. This can result in a surface that is both multiply and broadly peaked. In passing, we note that these two matching ambiguities equally exist in the uses of dissimilarity measures SAVD and SSD. This problem represents the second difficulty in the feature-based approach: The uncertainty in determining the best match.

Recently, there have been efforts to combine correlation matching with a smoothness constraint for deriving a displacement field [7], [40], [42]. These works are characterized by the following three aspects.

- First, they dealt with mainly rigid object motion.
- Second, a very small template is used, which effectively corresponds to a SN feature representation.
- Third, match measures are used within a global optimisation formulation in conjunction with a smoothness constraint in a fashion similar to the gradient-based approach.

As discussed by Singh [42], the implied assumption in this approach is that the correlation surface is unimodal, i.e., it contains either a well defined maximum, or a well defined minimum. The coarse-to-fine search strategy [7] cannot guarantee the correct match in the presence of multiple maxima or minima, and it essentially avoids some possible match positions. Anandan proved that his formulation [7] is equivalent to the gradient-based approach in the limit that inter-frame time interval and the template size tend to zero. This proof explicitly revealed the relation between correlation matching and the gradient-based method.

In this paper we present a new hybrid technique that overcomes the two main difficulties of the feature-based approach. The method is characterized by three main features:

1) A correlation surface is considered to be generically multi-modal and unimodal is only a special case.
2) Using correlation matching to establish multiple candidate matches for each template. We develop it under a model for time-varying images more general than that defined by equation (1).
3) Using relaxation labeling as an ambiguity reduction process by enforcing the vector field smoothness constraints.

Our relaxation algorithm is probabilistic, and is based on that described by Rosenfeld et al., [36], [37]. Our smoothness constraint is expressed by continuous functions, in contrast to the discrete and binary form in the B-T method. We have previously showed that the former is more suitable for the evolving fluid motion [48] because the discrete and binary function tends to result in unrealistically sharp transitions in flow directions.

IV. Cross-Correlation and Likelihood Estimation

In order to simplify the analysis we deal with strictly 2D problems with the understanding that 2D images are projections of 3D scenes under certain imaging geometry [30], [44], [4].

Equation (1) is a very idealised model for time-varying imagery. Ryan [38] proposed a more general model for the stereo matching problem. The model can be equivalently expressed here as

$$I(x, y, t) = a \cdot I(x + \delta x, y + \delta y, t + \delta t) + m + n \qquad (3)$$

where a is a scale factor introduced to account for the differences in image contrast, and m is a bias term modeling the possible differences in mean intensity level. a and m, depending on the shift (δx, δy) and time in general, can be seen as distortions due to non-uniformity in illumination and transmission channel. We also introduce a noise term n as it is unavoidable in real world imagery. This general model does not seem to have been discussed within the gradient-based approach in the literature.

The MCC method is shown by Ryan [38] to be a maximum likelihood estimator. He used vector analysis in his derivation. We outline the argument in scalar representation and extend it to the multiple candidate matching situation so that relaxation labeling can be used.

Consider the template matching process using digital images based on (3). Without loss of generality we use a square template of $p \times p$ pixels, and rewrite (3) as follows

$$f(x, y) = a \cdot g(x, y; u_s, v_s) + m + n, \qquad 1 \le x, y \le p \qquad (4)$$

where $f(x, y)$ is the template, and $g(x, y; u_s, v_s)$ is assumed to be the (u_s, v_s)-shifted version of $f(x, y)$ found within the corresponding search area in a time-lapsed image. Assuming that the noise n is zero mean Gaussian, independent of image signals, and with a variance δ^2 which is much smaller than the object pattern variation, it is easy to verify that the following relations exist

$$f(x,y) - \bar{f} = a \cdot \left(g(x,y;u_s,v_s) - \bar{g}(u_s,v_s)\right) + n \qquad (5)$$

$$a^2 = \frac{\overline{\left(f(x,y) - \bar{f}\right)^2}}{\overline{\left(g(x,y;u_s,v_s) - \bar{g}(u_s,v_s)\right)^2}} \qquad (6)$$

where $\bar{f}$ and $\bar{g}(u_s,v_s)$ denote the averages estimated over the spatial regions covered by $f(x, y)$ and $g(x, y; u_s, v_s)$. Clearly, $\bar{g}(u_s,v_s)$ depends on the shift (u_s, v_s).

However, the shift (u_s, v_s) is the displacement to be estimated in our matching procedure. Without any a priori knowledge of how the template pattern is moved, it is natural to reason that every subimage $g(x, y; u, v)$ for an arbitrary shift (u, v) from the search area is a possible match for $f(x, y)$. From (5), the probability, therefore the likelihood, that $g(x + u, y + v)$ is a match for $f(x, y)$, given $f(x, y)$, is given by [17]

$$L(u,v) = \Pr\left[g(x,y;u,v) \mid f(x,y)\right] = \frac{1}{\left(\sqrt{2\pi}\delta\right)^{p^2}} \exp\left\{-\frac{1}{2\delta^2}\sum_{y=1}^{p}\sum_{x=1}^{p}\left(\begin{array}{l}\left(f(x,y)-\bar{f}\right) \\ -a\cdot\left(g(x,y;u,v)-\bar{g}(u,v)\right)\end{array}\right)^2\right\} \qquad (7)$$

In the present work, we confine the search area to $0 \le \sqrt{\left(u^2+v^2\right)} \le V_m$, where V_m is the maximum interframe displacement in pixels, and is deemed to be known a priori. This defines a circular region from which the match is to be determined.

Let the summation within the exponential part of (7) be denoted by $i(u, v)$. Replacing a in (7) by (6) and further rearranging the summation we have is the CCC or the normalized covariance between $f(x, y)$ and $g(x, y; u, v)$, ranging from -1 to 1.

$$i(u,v) = 2\left(\rho_{f,g}(u,v) - 1\right) \cdot \sum\sum\left(f(x,y) - \bar{f}\right)^2 \qquad (8)$$

where

$$\rho_{f,g}(u,v) = \frac{\sum_{y=1}^{p}\sum_{x=1}^{p}\left(f(x,y)-\bar{f}\right)\cdot\left(g(x,y;u,v)-\bar{g}(u,v)\right)}{\left(\sum_{y=1}^{p}\sum_{x=1}^{p}\left(f(x,y)-\bar{f}\right)^2\right)^{1/2}\left(\sum_{y=1}^{p}\sum_{x=1}^{p}\left(g(x,y;u,v)-\bar{g}(u,v)\right)^2\right)^{1/2}} \qquad (9)$$

From (7) and (8) we see that the likelihood $L(u, v)$ depends on $\rho_{f,g}(u, v)$ monotonically.

The MCC method determines the displacement estimate by selecting the (u, v) that maximises $\rho_{f,g}(u, v)$, which is equivalent to maximising $L(u, v)$. The MCC is therefore a maximum likelihood (ML) estimator.

CCC is often considered as computationally expensive. An alternative matching method is obtained if one assumes a simpler model by taking $a = 1$ in equations from (4) to (7). Subsequently we have the SSD dissimilarity measure

$$i(u,v) = -\sum\sum\left(\left(f(x,y)-\bar{f}\right)-\left(g(x,y;u,v)-\bar{g}(u,v)\right)\right)^2 \quad (10)$$

A popular ad hoc matching scheme, which is closest to (10) but cannot be justified on second order Gaussian statistics, is the sum of the absolute value of difference (SAVD):

$$i(u,v) = -\sum\sum\left|\left(f(x,y)-\bar{f}\right)-\left(g(x,y;u,v)-\bar{g}(u,v)\right)\right| \quad (11)$$

Both SAVD and SSD can be used in the fast SSDA algorithm for searching a minimum in a correlation surface while CCC cannot be used there because SSDA cannot be used to search for a maximum. Therefore, although SAVD and SSD are derived from a simpler model (a = 1) they are computationally faster than CCC with SAVD being the fastest.

Some researchers [7], [42] working on SAVD and SSD appear to have adopted even simpler forms:

$$i(u,v) = \sum\sum\left|f(x,y)-g(x,y;u,v)\right| \quad (12)$$

and

$$i(u,v) = \sum\sum\left(f(x,y)-g(x,y;u,v)\right)^2 \quad (13)$$

which correspond to the simplest conservation model of (1).

The above discussion shows that every shift (u, v) of a template, within the circular region $0 \le \sqrt{\left(u^2+v^2\right)} \le V_m$, is a candidate for the displacement estimate of the template, and its associated likelihood is $L(u, v)$. At image pixel resolution, we can have up to several thousand candidates within the region. It is neither economical nor necessary to use all of them since many of them have likelihood measures which are so low that they practically have no significance. We therefore use only a small number of candidates which are of the highest likelihood measures within the circular region. We expect that candidates so selected will correspond to correlation peaks and their neighboring positions. Since the likelihood $L(u, v)$ depends on $\rho_{f,g}(u, v)$ monotonically, we use $\rho_{f,g}(u, v)$ to rank the likelihood of each candidate. Fig. 1 illustrates the candidate vectors in relation to the correlation surface. To overcome the main difficulty of the MCC as discussed in the last section, and determine the best match from these candidates, one must use information in addition to image data statistics. This is where the relaxation labeling technique is suitably employed.

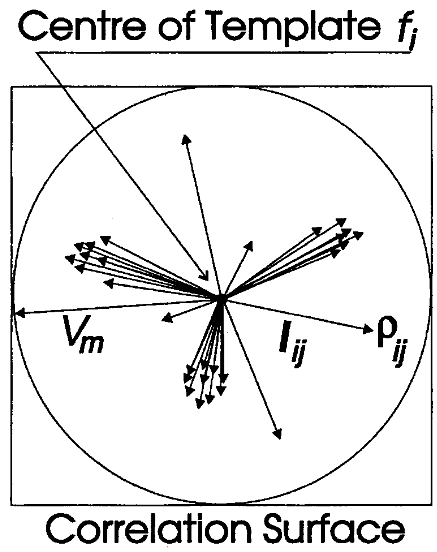


Fig. 1. Initial label assignment. Candidate vectors selected are expected to correspond to correlation peaks and their neighboring positions.

V. Relaxation Labeling

A. Initial Label and Probability Assignment

Assume that N templates f_i, $i = 1, 2, \cdots, N$, are identified in the first of the two images. We note that, in general, neighboring templates may partially overlap. In line with the argument in the last section, we define a candidate vector set for template f_i by $\mathbf{l}_{ij}$, $0 \le j \le M$, where M is the user defined maximum number of candidates selected for each template. $\mathbf{l}_{ij}$ is the shift of f_i to the possible match pattern, now denoted conveniently by g_{ij}, within the search area. Such a candidate vector is conventionally termed a *label* when labeling techniques are used. The correlation coefficient that ranks the likelihood of $\mathbf{l}_{ij}$ can be equally denoted by ρ_{ij}. Note that the number of candidates for a template can be less than M because we require that $\rho_{ij} \ge \rho_0$, a user specified threshold. The label set so defined implicitly includes a label $\mathbf{l}_{i0}$ which does not correspond to any of the candidates selected. We term $\mathbf{l}_{i0}$ the no-match label of f_i accounting for the no solution case, as in [9]. This is necessary because image noise, occlusion, as well as feature pattern distortion can be so excessive that no useful solution should be expected. The initial likelihood of $\mathbf{l}_{i0}$ can be plausibly estimated by either

$$\rho_{i0} = 1 - \max\{\rho_{ij}\}, \qquad 1 \le j \le M \quad (14)$$

or

$$\rho_{i0} = 1 - \frac{1}{M}\sum_{j=1}^{M}\rho_{ij}, \qquad 1 \le j \le M \quad (15)$$

We have used (14) in this work because it implies an assumption that a solution exists if $\max\{\rho_{ij}\} = 1$, which corresponds to the maximum possible likelihood obtainable according to (7) and (9). From a statistical point view $\rho_{ij} = 1$ is a extremely rare occurrence, as it indicates a perfect match.

We showed in the last section that ρ_{ij} indicates the rank of the conditional probability of the label $\mathbf{l}_{ij}$. The probability of $\mathbf{l}_{ij}$ is, in fact, the joint probability, denoted by $Pr[f_i, g_{i,j}]$, that f_i is matchable within the search area and the candidate pattern g_{ij} is a match for f_i. This joint probability relates to the conditional probability by

$$Pr[f_i, g_{i,j}] = Pr[f_i] \cdot Pr[g_{i,j} \mid f_i] \tag{16}$$

where $Pr[f_i]$ denotes the probability of the condition that f_i is matchable. However, (16) shows that all conditional probabilities share the common factor $Pr[f_i]$ so ρ_{ij} also indicates the rank of the joint probability. The use of (14) or (15) can be seen as determining, by cross correlation, the rank of the no-match label's likelihood for f_i.

We now have a set of possible labels, including the no-match one, and their likelihood rankings for each template. Our argument designates that one, and only one, of these labels must be selected as the true outcome. These labels therefore form a complete set of mutually exclusive events of all possibilities. Accordingly, the sum of all probabilities should be 1. Under the circumstance, we estimate the initial probability of each label by normalisation:

$$p_{ij}^0 = \frac{\rho_{ij}}{\sum_r \rho_{ir}}, \qquad 0 \le r \le M \tag{17}$$

Although the estimation is not accurate in terms of (7), it does preserve the likelihood rankings, and is a practical method to initialise these probabilities for the relaxation labeling algorithm.

B. Determine Compatibility Coefficients Using Smoothness Assumption

We now define compatibility coefficients to be used in the consistent labeling operation [36], [37].

Intuitively, or from the knowledge of non-turbulent fluid, one knows that the velocities of two neighboring points constrain each other in that their magnitudes and directions cannot be very different. This constraining relation, as illustrated in Fig. 2, is further strengthened the closer the two points are. The consistency relation, or compatibility, between two neighboring points' movement can therefore be determined from the difference between the two velocities and the distance between the two points. In our case, the two points are the centers of two templates. Since a template represents a feature, we may as well refer to its center as a feature point.

Consider two neighboring features f_i, f_h and their respective labels $\mathbf{l}_{ij}$, $\mathbf{l}_{hk}$ (see Fig. 2). We use $c(i, j; h, k)$, the compatibility coefficient, to describe the consistency relation between $\mathbf{l}_{ij}$ and $\mathbf{l}_{hk}$. We require that $-1.0 \le c(i, j; h, k) \le 1.0$, expressing the degree of consistency or inconsistency in terms of our smoothness consideration.

First consider the differences between the two labels. We define the compatibility coefficient as

$$\alpha(i,j;h,k) = \cos\theta(i,j;h,k) \cdot \left(1 - \frac{\left\| |\mathbf{l}_{ij}| - |\mathbf{l}_{hk}| \right\|}{\max\left(|\mathbf{l}_{ij}|, |\mathbf{l}_{hk}|\right)}\right) \tag{18}$$

where, referring to Fig. 2, $\theta(i, j; h, k)$ is the angle between labels $\mathbf{l}_{ij}$ and $\mathbf{l}_{hk}$, $\max\left(|\mathbf{l}_{ij}|, |\mathbf{l}_{hk}|\right)$ is the length of the longer label. The ratio $\left\| |\mathbf{l}_{ij}| - |\mathbf{l}_{hk}| \right\| / \max\left(|\mathbf{l}_{ij}|, |\mathbf{l}_{hk}|\right)$ is the *relative difference* between the two labels' magnitudes, and it takes values within [0, 1].

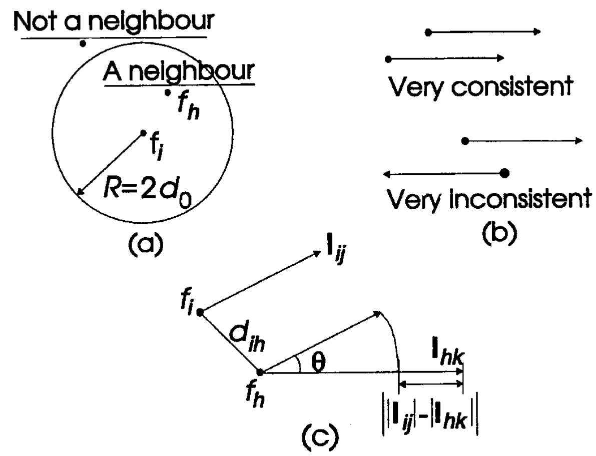


Fig. 2. Consistency check based on the smoothness assumption in a fluid motion field. (a) A feature point f_h is a neighbor of f_i if its distance from f_i is less than or equal to R. (b) Examples of two displacement vectors being very consistent or very inconsistent. (c) The general case in which the consistency strength depends on the differences between displacements' directions and magnitudes, as well as the distance between the two feature points.

Equation (18) can be understood by considering two special situations. Firstly, when $\mathbf{l}_{ij}$ and $\mathbf{l}_{hk}$ are of the same length, the relative difference between their magnitudes is 1, hence

$$\alpha(i, j; h, k) = \cos\theta(i, j; h, k) \tag{19}$$

In this case, the two labels are most compatible, i.e., $\alpha(i, j; h, k) = 1$, if $\theta(i, j; h, k)$ is 0; or most incompatible, i.e., $\alpha(i, j; h, k) = -1$, if $\theta(i, j; h, k)$ is π. Secondly, when the labels have the same direction, then

$$\alpha(i,j;h,k) = 1 - \frac{\left\| |\mathbf{l}_{ij}| - |\mathbf{l}_{hk}| \right\|}{\max\left(|\mathbf{l}_{ij}|, |\mathbf{l}_{hk}|\right)} \tag{20}$$

In this case, the two labels are most compatible, i.e., $\alpha(i, j; h, k) = 1$, if they are of the same length; or most incompatible, i.e., $\alpha(i, j; h, k) = 0$ if the two labels' relative difference is 1. In general, the combined effect from label direction and magnitude differences is that the value of $\alpha(i, j; h, k)$ is within the range [−1, 1]. $\alpha(i, j; h, k) = 0$ indicates that the two labels are independent of each other.

The compatibility relating to the distance d_{ih} between feature points f_i and f_h can be expressed as

$$\beta(i, h) = \exp(-d_{ih}/d_0)$$

where the constant d_0 characterizes the flow. For example, d_0 is small in an area having relatively complicated flow structures such as motion shear and eddy so that neighboring points' movements are loosely related. In general, d_0 is a function of position. Clearly, like some parameters in many real world problems, d_0 cannot be known beforehand and is also application dependent. For simplicity, it is taken as a constant, and is chosen empirically for our cloud tracking application. In order to reduce the amount of computation, we specify a distance $R = 2d_0$ to determine the neighbors of a feature point as illustrated in Fig. 2. Clearly, $\beta(i, h)$ is in the range [0, 1].

From the above discussion, the total compatibility coefficient that meets our requirement can be simply given by

$$c(i,j;h,k) = \alpha(i,j;h,k) \cdot \beta(i,h). \tag{22}$$

It is noted that for the no-match label $\mathbf{l}_{i0}$ we define $c(i, 0; h, k) = 0$. This is because the no-match case is related to the fact that a vector is unable to be defined, and it is reasonable to assume that this fact is independent of the movement of neighboring points. Equally, we have $c(i, j; h, 0) = 0$.

C. Iterative Processing for Updating Probabilities

Our iterative scheme is similar to that given by Rosenfeld and Kak [37]. Within each iteration, say, in the nth iteration, we update label $\mathbf{l}_{i,j}$'s probability, $p_{i,j}^n$, according to $\mathbf{l}_{i,j}$'s consistency relation with the labels of all of neighboring features.

The compatibility coefficient $c(i, j; h, k)$ can be equivalently seen as the *support* of label $\mathbf{l}_{h,k}$ for $\mathbf{l}_{i,j}$. Of course, this support has to be weighted by $p_{h,k}^n$. The total support for $\mathbf{l}_{i,j}$, therefore, is proportional to

$$q_{ij}^n = \sum_{h \neq 1} \sum_k c(i,j;h,k) p_{hk}^n = \sum_{h \neq i} \left(\beta(i,h) \sum_k \alpha(i,j;h,k) p_{hk}^n \right) \tag{23}$$

We define the support function by

$$s_{ij}^n = \frac{1}{C \left| \max_r \left(q_{ir}^n \right) \right|} q_{ij}^n \tag{24}$$

where $\left| \max_r \left(q_{ir}^n \right) \right|$ corresponds to the largest value of the support found within the label set of f_i, and $C \geq 1$ is a constant which controls the speed of convergence as explained in the next paragraph. Clearly, S_{ij}^n is within the range [−1, 1], and it can be -1 or 1 for some values of the j only if $C = 1$.

p_{ij}^n is then updated according to,

$$p_{ij}^{n+1} = \frac{p_{ij}^n \left(1 + S_{ij}^n \right)}{\sum_r p_{ir}^n \left(1 + S_{ir}^n \right)} \tag{25}$$

We can now discuss the meaning of C in more detail based on (25). When $C \gg 1$ in (24), all supports $\left| S_{ij}^n \right|, 0 \leq j \leq M$, will be small so that their probabilities p_{ij}^n get modified slowly, i.e., the system converges slowly according to (24). On the other hand, when C is close to 1, some supports S_{ij}^n approach −1. Consequently their associated probabilities p_{ij}^n are suppressed very quickly in (24). For this reason, we usually do not set $C = 1$ in order to avoid eliminating some label probabilities. It can be seen from (24) that for the no-match label $\mathbf{l}_{i0}$ the support $S_{i0}^n = 0$ so that p_{i0} gets modified only in the normalization procedure of (25).

The iterative process can be summarized as follows. If a label, within the label set of a feature, has relatively more support from neighboring features, its chance of being selected as the feature's displacement will be enhanced. Its probability will be decreased if the label has relatively less support within the label set.

We terminate the iterative process using a threshold on the average absolute probability difference [37] between consecutive iterations, i.e.,

$$\frac{1}{M \times N} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| p_{ij}^n - p_{ij}^{n+1} \right| \leq p_d \tag{26}$$

since it indicates the convergence of the system [37].

When (26) is met, the maximum probability label within the label set of each feature is taken as an estimate of the feature's displacement. If the maximum probability is the no-match label's probability then the displacement of the feature is undetermined.

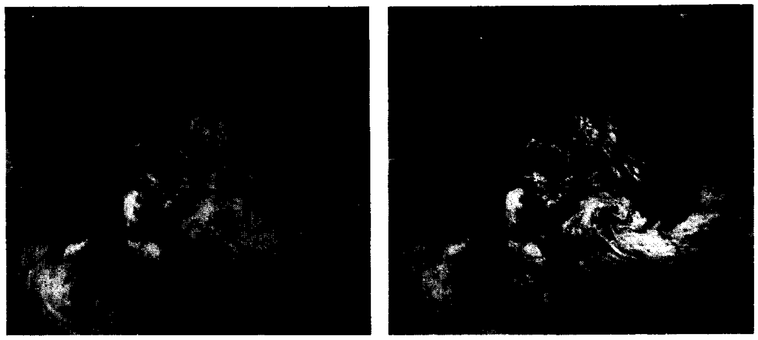
VI. Implementation

In all previous work using the MCC method, for example [27], [13], [23], [47], templates are conventionally centered at regular grid points located evenly across the first of the image pair. When employing the B-T method, feature points are first identified from both images, and then, for computing a similarity measure, a template is centered on a feature point from the first image and another is centered on a feature point from the second image. It is easy to see that the scheme in the B-T method can be modified for use in the MCC method. The difference is that, for the MCC method, we do not need to identify feature points in the second image, and the template is matched against a large search area in the second image for computing a correlation surface. We have used both the traditional scheme and the new schemes within our C-R and the MCC methods to compute displacement fields. In this paper, we use the latter so that we can compare our method with both the MCC and B-T methods. The new scheme has an advantage over the traditional one: featureless regions, which are often not of interest, in an image are mostly excluded because feature points are determined based on computing intensity variation.

Like B-T, we used a 5×5 interest operator for identifying feature points. Our operator [46], [48] is a modified version of the Moravec operator [33] since we found that the Moravec operator tended to generate very sparsely located features points from images containing mainly smooth LN features.

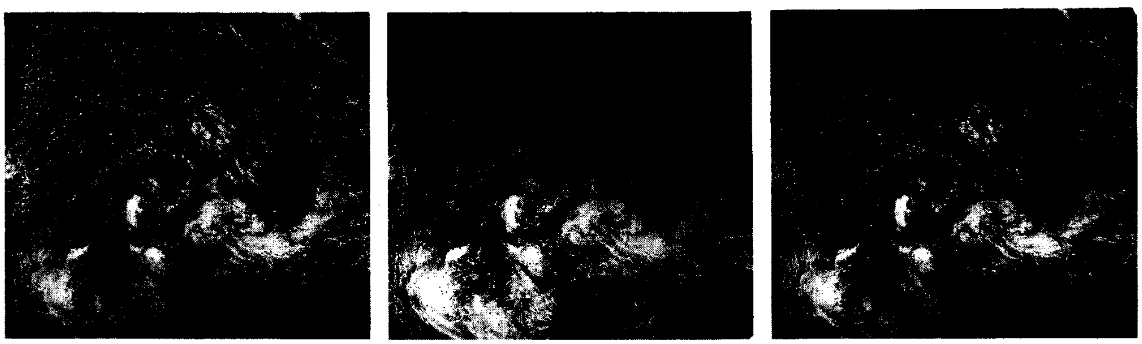
Following Aisbett [6], we used a template size of 12×12 pixels in all of our examples shown here except in one case a template of 32×32 pixels was used for comparison to show the relevance of the size. In all C-R results, the parameters are set as follows: V_m = 15 pixels, $\rho_0 = 0.3$, $M = 20$, $R = 25$ pixels, $C = 1.05$, and p_d = 0.007. We do not discuss in detail the selection of these parameters since they rely on the experience of the user, and are usually application dependent. The same set of parameters are used in either the MCC or the B-T methods.

In all resultant vector fields presented in this paper, a displacement vector is represented by a thin white line, starting from a white dot and ending with a black dot. A white dot indicates the position of a feature point or the center of a template. The length of a vector represents the magnitude of the corresponding pattern displacement.



(a) (b)

Fig. 3. (a) The first and (b) the second of a sequence of five. one hourly GMS images, as described in text.



(a) (b) (c)

Fig. 4. Comparisons between (a) the MCC; (b) the B-T; and (c) the C-R method. Each vector is represented by a white line starting from a white dot and ending on a black dot. All correlation coefficients are thresholded at $\rho_0 = 0.3$ in the MCC and the C-R results. See text for detailed discussion on these results. Note that a small number of feature points near the edges of the image are removed in the MCC and the C-R methods due to their particular handling of the edges.

VII. Results and Discussions

Experiments have been conducted on many GMS infrared cloud motion images sequences, as well as on two NOAA AVHRR SST images. Due to difficulties in obtaining cloud free SST images, we have not applied our algorithm to more of them. In this paper we concentrate on the cloud motion estimation from a sequence of five one-hourly GMS images, and refer readers to our previous publications for results and discussions on sea surface current computation using SST images. We will illustrate the improvement in our new C-R results as compared to the MCC and the B-T results, and discuss the significances in solving template matching problems and in improving existing cloud tracking procedures. Further, since it is well known that the MCC method is very sensitive to the size of the template, we will compare the use of two different sizes and argue that the C-R method is much less sensitive to the template size than the MCC method.

We first compare the three different methods. The five GMS images and the four vector fields estimated from them are of the same nature and quality. It is necessary to show only a sample of the sequence and the estimation results. The first and the second images of the sequence are given in Fig. 3a and Fig. 3b. Correspondingly, displayed in Fig. 4 are the estimated displacement vector fields using the first and the second images. Fig. 4a, Fig. 4b, and Fig. 4c show the results of, respectively, the MCC, the B-T, and the new C-R methods. The three kinds of vector fields are overlaid on the first image.

The full GMS images that we have used are of size 400×576 pixels, of 5km resolution, covering a large area of south pacific around New Zealand. In this work, we have used a subarea of the full images, of 400×460 pixels. We have trimmed off the apparent land mass of Australia on the far left of the full image in order to avoid any possible complication. The entire GMS sequence, from which the five images are taken, illustrates the formation of a low pressure center, and the cloud motion is characterized by a large scale cyclonic spiral around the center. Prior to using motion tracking algorithms, video animation of this sequence were performed by McGregor and Marks [31] to examine the motion characteristics. In most of the image area away from the center of the low, the motion is fairly smooth, especially the large area of shower clouds above and to left of the low center. However, a strong motion shear has been identified between the shower cloud and the cyclonic center. This shear is seen as a band of cloud immediately to the left of the cloud at the center, evident in the MCC and the C-R results shown in Fig. 4a and 4c. Such a motion shear is expected within such a dynamic system by meteorologists [53].

It is seen that the three kinds of results shown in Fig. 4a, (b) and Fig. 4c broadly agree with one another in a sense that they all illustrated the main characteristics of the motion. It is obvious that the resultant vector field from C-R method is far more coherent, especially in the large area of shower clouds, than that of the MCC and the B-T methods. The MCC result is somehow better than that of the B-T method. We can easily notice the dominant artifacts, discussed in Section III, from the B-T results. In comparison, the C-R results do not have these artifacts. In addition, in the B-T results, there are a significant number of white dots for which correspondence points in the second image cannot be found. We argue that this is also due to the uncertainties in determining feature points from these smooth and LN feature dominated images.

We examine again the shower clouds in the MCC results in comparison with that in the new C-R results. Many vectors in the MCC results deviate from the correct flow by large angles, varying, roughly, from 45 to 135 degrees. We examined several candidate sets corresponding to these erroneous MCC results. The positions and initial likelihood values of these candidate vectors indicated the existence of multiple peaks in these correlation surfaces. Wu, et al., [47] illustrated a correlation surface with two major peaks which caused a bad estimation of sea surface current motion. This error of the MCC method is largely corrected by the C-R method, as illustrated by Fig. 4c. This improvement from the MCC results to the C-R results convince us that the C-R method can deal effectively with multi-modal correlation surface. In operational cloud tracking procedures, manual editing of MCC vector fields for improving the vector field coherency has always been a subjective and time consuming final step. This manual editing also tends to cause substantial loss of information because it simply removes many vectors without replacement [1], [2]. One of the current main research objectives in this area is to reduce, and possibly eliminate this manual step [1], [2]. The C-R results therefore demonstrate that the new method can significantly improve cloud tracking procedures in terms of reducing manual editing and preserving information.

We further examine in detail the area of the motion shear in all three kinds of results shown in Fig. 4. It is seen that the B-T vectors fail to clearly illustrate this motion shear while both the C-R and the MCC results do. We notice that one vector in the C-R result is pointing to the spiral center, perpendicular to the direction of motion shear where flow vectors are distinctly in two opposite directions. In the MCC results the corresponding vector, however, is seen to line up with the dominant shower cloud flow. We suspect that this is an artifact of the C-R results due to the requirement of vector field smoothness which is inappropriate in the area of motion shear. The next section discusses possible approaches to reducing errors in cloud tracking using the C-R method.

Next we investigate the sensitivity of the C-R method to template sizes in comparison with the MCC method. We note that this problem is in fact the manifestation of the matching uncertainties, as described in Section III, in digital implementation. The problem affects the MCC estimation in three ways.

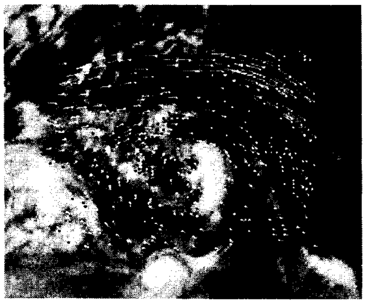
- Firstly, when the template is large compared to an invariant feature pattern, deformation (multiple objects in the case of rigid object motion) is generally severe. Subsequently, motion estimation is erroneous, and at best, it results in an average displacement for the entire template.
- Secondly, when the template is small compared to an invariant feature pattern, the template does not contain enough information to match unambiguously within the search area. This can lead to correlation surfaces which are generally multi-modal and suffer from the aperture problem as well.
- Thirdly, when the template is compatible with the feature pattern but much smaller than the search area, the correlation surface is still likely to be multiply peaked due to, ultimately, the limited quantization and spatial sampling resolutions available to digital images, especially in the case of low contrast and smooth LN features.

We argue that for the C-R method, while large templates will still degrade the estimation results as in the first case of the MCC estimation, the situation is different for relatively smaller templates corresponding to the second and the third cases of the MCC estimation. C-R estimation is inherently an ambiguity reduction process, and it uses additional constraints to determine an estimate from many potential candidates. To demonstrate this property of the C-R method, we compare Fig. 5a and Fig. 5b which are C-R results corresponding to, respectively, a 12×12 pixel template and a 32×32 pixel template.

We choose a 32×32 pixel template because it is used operationally [39] for cloud tracking. The background image in Fig. 5 is the same as that in Fig. 4 but covers only a small area around the cyclonic center. While the quality in terms of coherency is similar in the shower clouds in Fig. 5a and Fig. 5b, the vector field shown in Fig. 5b completely fails to illustrate the expected motion shear. The result confirms that the use of a large template will tend to eliminate details of a motion field. Consequently it is preferable to use relatively small templates for the C-R method. This also leads to a substantial reduction in computation time since the cross correlation matching is the major consumer of computing resource. In processing the GMS images, each vector field is the result of four iterations of relaxation labeling, which accounts for only 10 percent of the total computation time.

Further assessment of the quality of results usually depends on comparison with in situ measurements in respective application areas [1], [2]. This is outside the scope of this paper.

We applied all three algorithms to two NOAA SST images, taken 24 hours apart. The result of using the B-T method can be found in [46]. The comparison between the C-R result and the MCC result was given in [49]. Additional results of using the MCC and the B-T methods can be found in [47], [48]. Our comparison on the performances of these three algorithms when applied to the SST image gives similar conclusions as for the GMS images. These conclusions are only tentative as only two images were used.



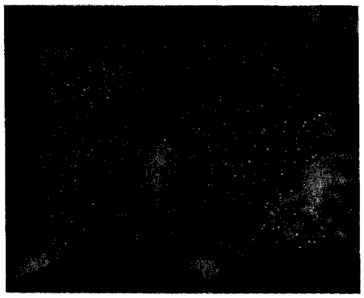


Fig. 5. Comparison between the use of two different template sizes: (a) the result of using a 12×12 pixel template; and (b) the result of using a 32×32 pixel template. (b) fails to illustrate the motion

VIII. SUMMARY AND FUTURE STUDIES

A new C-R method is developed for computing optical flow and its relationship to previous work is shown. The method has been successfully applied to tracking cloud motion from GMS image sequences. Compared to the popularly used MCC method and the B-T method, the C-R method achieves significantly improved results. Although this paper concentrates on its application to images containing mainly smooth LN (large neighborhood) features and evolving pattern motion, the method is equally applicable to images containing sharp SN (small neighborhood) features and rigid object motion. We emphasize that the C-R method's main feature, which is new to all existing methods, is its ability to deal effectively with multimodal correlation surfaces.

One main error source in using the C-R method for cloud tracking is due to the fact that cloud heights (altitudes) have not been preassigned. Clouds at different heights can have different motion, and occlusion does exist. Multi-spectral methods [39] have been used to determine cloud heights. When height information exists, it can be incorporated into the C-R method to separate different cloud layers and remove possible occlusion.

The existing C-R algorithm can be easily extended to processing more than two consecutive images. We are interested in investigating *temporal motion smoothness* within the consistency labeling framework. Candidate vectors for neighboring features in the time domain will be related to one another in terms of the smoothness consistency. Estimates will be chosen from candidates by considering both spatial and temporal motion consistency.

Employing SAVD, SSD, or the exponential part of (7) directly to identify potential vectors through the use of SSDA has the advantage of substantial computational savings. We will investigate the use of SAVD within the C-R framework. While it is understood that SAVD does not perform as well as the CCC matching we are interested in whether the final C-R results are severely affected by the initial candidate selection.

ACKNOWLEDGMENTS

The author wishes to thank his team leader, Dr D. Pairman, for his support during the course of this work, and in particular, his effort in detailed proof reading of this paper.

Funds for this research were provided by the New Zealand Foundation for Research, Science and Technology under Contract No. C09302.

REFERENCES

[1] *Proc. First Int'l Wind Workshop*, Workshop on wind extraction from operational meteorological satellite data. Washington, D.C., Sept. 17-19, 1991, EUMETSAT.

[2] *Proc. Second Int'l Wind Workshop*, Workshop on wind extraction from operational meteorological satellite data. Tokyo, Japan, Dec. 13-15, 1993, EUMETSAT.

[3] G. Adiv, "Determining 3D motion and structure from optical flows generated by several objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, no. 4, pp. 384-401, 1985.

[4] G. Adiv, "Inherent ambiguities in recovering 3D motion and structure from a noisy flow field," *IEEE Trans. Pattern Analysis and Machine Intelligence Intell.*, vol. 11, no. 5, pp. 477-489, 1989.

[5] E. H. Adelson, J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am. A*, vol. 2, no. 2, pp. 284-299, 1985.

[6] J. Aisbett, "Optical flow with an intensity-weighted smoothing," *Trans. Pattern Analysis and Machine Intelligence Intell.*, vol. 11, no. 5, pp. 512-522, 1989.

[7] P. Anandan, "Measuring visual motion from image sequences," PhD Thesis, COINS Department, Univ. of Massachusetts, Amherst, 1987.

[8] D. I. Barnea, H. F. Silverman, "A class of algorithms for fast digital image registration," *Trans. Computers*, vol. 21, pp. 179-186, 1972.

[9] S. T. Barnard, W. B. Thompson, "Disparity analysis of images," *Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 333–340. 1980.

[10] O. J. Braddick, "A short-range process in apparent motion," *Vision Res.*, vol. 14, pp. 519-527, 1974.

[11] O. J. Braddick, "Low-level and high-level processes in apparent motion," *Philo. Trans. Royal Soc. London B*, vol. 290, pp. 137-151, 1980.

[12] L. Dreschler, H. H. Nagel, "Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene," *Comput. Graphics Image Processing*, vol. 20, pp. 199-288, 1982.

[13] W. J. Emery, A. C. Thomas, and M. J. Collins, "An objective method for computing advective surface velocities from sequential infrared images," *J. Geophysical Research*, vol. 91, no. C11, pp. 12865–12878, 1986.

[14] W. Enkelmann, "Investigations of multigrid algorithms for estimation of optical flow fields in image sequences," *Computer Vision Graphics, Image Processing*, vol. 43, no. 2, pp. 150-177, 1988.

[15] F. Glazer, "Hierarchical gradient-based motion detection," *Proc. Image Understanding Workshop*, Los Angeles, CA, 1987, pp. 733-748.

[16] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *Trans. Pattern Analysis and Machine Intelligence Intell.*, vol. 7, no. 1, pp. 17-34, 1985

[17] J. L. Harris, "Resolving power and decision theory," *J. Opt. Soc. Amer.*, vol. 54, pp. 606-611, 1964.

[18] D. J. Heeger, "Model for the extraction of image flow," *J. Opt. Soc. Am. A*, vol. 2, no. 2, pp. 1455-1471, 1987.

[19] E. C. Hildreth, "The measurement of visual motion," Cambridge, MA: MIT Press, 1984.

[20] W. Hoff, N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," *Trans. Pattern Analysis and Machine Intelligence Intell.*, vol. 11, no. 2, pp. 121-136, 1989

[21] B. K. P. Horn, B. G. Schunk, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[22] R. A. Hummel, S. W. Zucker, "On the foundations of relaxation labeling processes," *Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 3, pp 276-287, 1983.

[23] M. Kamachi, "Advective surface velocities derived from sequential images for rotational flow field: limitations and applications of maximum cross-correlation method with rotational registration," *J. Geophysical Research*, vol. 94, no. C12, pp 18227–18233, 1989.

[24] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: An error analysis of gradient-based methods with local optimization," *Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp 229-244, 1987.

[25] G. Kelly, "Satellite observations for global modeling," *Adv. Space Res.* vol. 12, no. 7, pp. (7)263-(7)275, 1992.

[26] R. Kories, G. Zimmermann, "Motion detection in image sequences: an evaluation of feature detectors," in *Proc. 7th Int. Conf. on Pattern Recognition*, Montreal, Canada, 1984, pp. 778-780.

[27] J. A. Leese, C. S. Novak, "An automated technique for obtaining cloud motion from geosynchronous satellite data using cross-correlation," *J. Appl. Meteorology*, vol. 10, pp 118–132, 1971.

[28] J. A. Leese, C. S. Novak, and V. R. Taylor, "The detection cloud pattern motions from geosynchronous satellite image data," *Pattern Recognition*, vol. 2, pp. 279-292, 1970.

[29] J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Computer Graphics Image Processing*, vol. 4, pp. 311-327, 1975.

[30] H. C. Longuet-Higgins, "A computer program for reconstructing a scene from two projection," *Nature*, vol. 293, pp. 133-135, 1981.

[31] J. McGregor, C. J. Marks, "Making the most of GMS imagery," *Proc. 6th New Zealand Image Processing Workshop*, August 29-30, 1991, pp. 61-65.

[32] D. Marr, T. Poggio, "A computational theory of human stereo vision," in *Proc. R. Soc. London*, 1979, pp. 301-324, vol. B204.

[33] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, pp. 584, Cambridge, MA, Aug. 1977.

[34] H. H. Nagel, W. Enkelmann, "An Investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565-593, 1986

[35] A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: Part I," *Bell Syst. Tech. J.*, vol. 58, no. 3, Mar. 1979.

[36] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 6, pp. 420-433, 1976.

[37] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, vol. 2, 2nd ed., New York: Academic Press, 1982.

[38] T. W. Ryan, "The Prediction of Cross-Correlation Accuracy in Digital Stereo-Pair Images". PhD thesis, University of Arizona, 1981.

[39] J. Schmetz, K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and L. Van De Berg, "Operational cloud-motion winds from Meteosat infrared images," *J. Appl. Meteorol.*, vol. 32, no. 7, pp. 1206-1225, Jul. 1993.

[40] G. L. Scott, "Four-line method of locally estimating optic flow," *Image and Vision Computing*, vol. 5, no. 2, 1986.

[41] M. A. Shuh, R. Jain, "Detecting time-varying corners," in *Proc 7th Int. Conf. on Pattern Recognition*, pp. 2-5, Montreal, Canada, 1984.

[42] A. Singh, *Optic Flow Computation, A unified perspective*, IEEE Computer Society Press, 1991.

[43] E. A. Smith, D. R. Phillips, "Automated cloud tracking using precisely aligned digital ATS pictures," *Trans. Computers*, vol. 21, pp. 715-729, 1972.

[44] R. Y. Tsai, T. S. Huang, "Uniqueness and estimation of 3D motion parameters of rigid bodies with curved surfaces," *Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 13-27, 1984.

[45] J. Weng, N. Ahuja, and T. S. Huang, "Matching two perspective views," *Trans. Pattern Analysis and Machine Intelligence Intell.*, vol. 14, no. 8, pp. 806-825, 1992.

[46] Q. X. Wu and D. Pairman, "Computing sea surface velocities from sea surface temperature—A relaxation labeling approach," *Remote Sensing: Global Monitoring for Earth Management*, vol. I, , pp. 157-161, Proc. International Geoscience and Remote Sensing Symposium, Helsinki Univ. of Technology, Espoo, Finland, June 3-6, 1991.

[47] Q. X. Wu, D. Pairman, S. McNeill, and E. J. Barnes, "Computing advective velocities from satellite images of sea surface temperature," *Trans. Geoscience and Remote Sensing*, vol. 30, no. 1, pp. 166-176, 1992.

[48] Q. X. Wu, D. Pairman, E. J. Barnes, "Aspects of applying a relaxation labeling method to computing sea surface velocities from sea surface temperature images," *Proc. First Pacific Ocean Remote Sensing Conference*, pp. 802-807, Okinawa, Japan, August 25–31, 1992.

[49] Q. X. Wu, "A correlation-relaxation algorithm for computing velocity fields from time-lapsed satellite images," in *Proc. 6th Australasian Remote Sensing Conference*, pp. 40-47, Wellington, New Zealand, 2-6, Nov. 1992, vol. 3.

[50] Q. X. Wu, "Computing velocity field from sequential satellite images," in *Satellite Remote Sensing of Oceanic Environment*, Ed. Jones, Sugimori and Stewart, Publisher: Seibutsu Kenkyusha Co. Ltd., 1993, Chap. 2.4, pp. 38-47.

[51] Q. X. Wu, "Computing cloud motion using a correlation relaxation algorithm: Improving estimation by exploiting problem knowledge," *Proc. second Int'l wind workshop*, Tokyo, Japan 13-15 Dec. 1993, EUMETSAT.

[52] S. Ullman, *The interpretation of visual motion*, Cambridge, MA: MIT Press, 1979.

[53] C. Zick and J. McGregor, *Private Communication*, Inst. of Geophysics, Victoria, University of Wellington, 1992.

Qing X. Wu received the BSc degree in physics from Nanjing University, Nanjing, China in 19982. He was a research assistant at Wuhan Institute of Physics, Chinese Academy of Science from 1982 to 1985, working on underwater acoustic imaging. In 1989, he received the PhD in electrical engineering from the University of Canterbury, New Zealand, where he was working on image processing and medical imaging.

Since 1989, Dr. Wu has been a research scientist with the Image Processing and Remote Sensing Group, Physical Sciences Division, Department of Scientific and Industrial Research, New Zealand. The group later became part of the Landcare Research Institute of New Zealand, after a major government science restructuring.

Currently, Dr. Wu has two main research interests. One is motion and dynamic scene analyses and its applications in dynamic meteorology and physical oceanography. His other main current interest is speckle noise modeling and feature detection in synthetic aperture radar images.