Robust Reweighted MAP Motion Estimation

Dong-Gyu Sim, *Student Member, IEEE,* and Rae-Hong Park, *Member, IEEE*

Abstract—This paper proposes a motion estimation algorithm that is robust to motion discontinuity and noise. The proposed algorithm is constructed by embedding the least median squares (LMedS) of robust statistics into the maximum a posteriori (MAP) estimator. Difficulties in accurate estimation of the motion field arise from the smoothness constraint and the sensitivity to noise. To cope robustly with these problems, a median operator and the concept of reweighted least squares (RLS) are applied to the MAP motion estimator, resulting in the reweighted robust MAP (RRMAP). The proposed RRMAP motion estimation algorithm is also generalized for multiple image frame cases. Computer simulation with various synthetic image sequences shows that the proposed algorithm reduces errors, compared to three existing robust motion estimation algorithms that are based on M-estimation, total least squares (TLS), and Hough transform. It is also observed that the proposed algorithm is statistically efficient and robust to additive Gaussian noise and impulse noise. Furthermore, the proposed algorithm yields reasonable performance for real image sequences.

Index Terms—Motion estimation, regularization, MAP estimation, robust statistics, LMedS.

1 Introduction

In early vision problems, a solution may not exist at all nor be unique. Even if a unique solution exists, it may be sensitive to noise. Because some early vision problems are ill-posed, they are sensitive to noise, and their solutions might be incorrect if the actual observations do not fit the model assumed [1]. To implement a robust computer vision algorithm, a stable system effectively coping with noise and distortions in the input is required [2], [3]. In spite of the usefulness of the motion field, conventional motion estimation methods suffer from such problems [4], [5]. This paper proposes a motion estimation algorithm robust to motion discontinuity and noise.

Because of ill-posedness, it is difficult to find a solution in all cases [6]. Generally, such a problem has been solved by using additional constraints [7], [8], [9], [10]. To obtain a solution having a physical meaning, conventional approaches based on regularization or maximum a posteriori MAP scheme [11] were presented. Although these two approaches were derived from different points of view, they used a similar framework for data fusion. Regularization obtains a smooth solution approximating observation data sets. But discontinuities in a solution and the corresponding observations occur in real applications, resulting in poor results. To overcome this problem, a variety of constraints have been proposed in the regularization [1], [6], [11], [12], [13], [14], [15], [16]. Also recently, a MAP scheme generalizing the regularization has been widely used for many applications, however its performance is still degraded if actual observations do not fit to an assumed model or contain outliers that are anomalous data far away from the assumed error distribution. Thus, several conventional algorithms have been proposed by replacing each part of the MAP estimator by a robust estimator commonly used in robust statistics [12], [17]. On the other hand, least squares (LS) algorithms based on maximum likelihood (ML) were also employed in computer vision problems [18]. While these algorithms are optimal for Gaussian noise, their performance is severely deteriorated by a few outliers, which is common in practical computer vision applications. To cope with this problem, M-estimators based on ML were applied to some computer vision problems [19], [20], [21], [22]. Also, a robust least median squares (LMedS) algorithm having a high breakdown point was successfully applied to visual reconstruction problems [23] and moton estimation [24], where the breakdown point denotes a performance measure representing the percentage of outliers that an estimator can deal with.

In this paper, we present a robust motion estimation algorithm based on the reweighted robust MAP (RRMAP). It is constructed by applying a median operator and reweighted least squares (RLS) in robust statistics to the MAP motion estimation technique. To explain the concept of the proposed algorithm, we first briefly present regularization and MAP estimation methods. The proposed algorithm is introduced to accurately estimate the smooth and discontinuous motion field by combining the LMedS method as local optimization and the MAP approach as global optimization.

The rest of the paper is structured as follows. In Section 2, a number of estimation methods and existing robust motion estimation algorithms are discussed. In Section 3, the proposed RRMAP motion estimation method is presented. In Section 4, experimental results and discussions are presented and finally conclusions are given in Section 5.

• *The authors are with the Department of Electronic Engineering, Sogang University, C.P.O. Box 1142, Seoul 100-611, Korea.*
E-mail: rhpark@ccs.songag.ac.kr.

Manuscript received 31 July 1996; revised 23 Dec. 1997. Recommended for acceptance by R. Szeliski.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106142.

2 Various Estimation Methods and Existing Robust Motion Estimation Algorithms

In this section, we briefly describe the robust regression that is fundamental to the proposed algorithm. Also we briefly present regularization, MAP estimation, and existing motion estimation methods based on robust statistics.

2.1 Robust Regression

Regression determines a set of parameters of a model that effectively fits observation data sets. In general, a linear model given by $y_i = \sum_{k=1}^{p} x_{ik}\theta_k + e_i, 1 \le i \le n$, has been widely used, where both y_i and x_{ik} represent observation data, and e_i denotes the error in observation data. The observations x_{i1}, x_{i2}, ..., x_{ip} are denoted by a column vector $\vec{x}_i$, and similarly observations y_1, y_2, ..., y_n are represented by a column vector $\vec{y}$. Parameters θ_ks should be estimated. An estimated parameter vector is given by a column vector $\vec{\theta} = \left[\theta_1\, \theta_2 \cdots \theta_p\right]^T$, where p denotes the number of parameters, and $\hat{y}_i = \sum_{k=1}^{p} x_{ik}\theta_k$ represents the prediction of y_i. In most LS methods, the estimated parameter vector $\vec{\theta}^*$ is determined by

$$\vec{\theta}^* = \arg\min_{\vec{\theta}} \sum_{i=1}^{n} r_i^2 \tag{1}$$

where $r_i = y_i - \hat{y}_i$ represents the estimation error assumed as zero-mean Gaussian. The parameter vector $\vec{\theta}^*$ is estimated by minimizing the summation of squared errors. But this method has a drawback that a single outlier included in observation data can severely deteriorate the estimation result.

Robust statistics introduces the finite sample breakdown point ϵ defined by

$$\epsilon(T, \vec{y}) = \min\left\{\frac{m}{n}; \text{bias}(m; T, \vec{y}) \text{ is } \infty\right\}$$

where $\text{bias}(m; T, \vec{y})$ is represented by $\sup\left\|T(\vec{y}') - T(\vec{y})\right\|$, $T(\vec{y})$ equals $\vec{\theta}$, and n denotes the number of observations [22]. The vector $\vec{y}$ represents an observation vector, and $\vec{y}'$ denotes an observation vector generated by deleting m outliers from $\vec{y}$. The breakdown point is defined by the minimum outlier ratio of m and n when the estimation error with m outliers in n observations is infinite. Any estimator with the breakdown point equal to ϵ can reliably estimate parameters if the outlier ratio is less than ϵ. Note that the breakdown point of an LS estimator is $\frac{1}{n}$, i.e., the performance may be severely degraded by a single outlier in n observations. To overcome this drawback, M-estimator was proposed based on ML estimation, by replacing a square cost function in (1) by a convex symmetric cost function ρ to discriminate inliers from outliers. M-estimator is effective when outliers are included in observation y_i. The breakdown point of M-estimator is defined by $\frac{1}{p+1}$. Thus, as p increases, the breakdown point decreases. If p is equal to two in the case of translational motion estimation, the breakdown point becomes at most 33 percent. Furthermore, only one outlier in observation x_{ik} works as a leverage point, in which the breakdown point decreases to zero [20], [22], [25].

Other robust estimators have been proposed in robust statistics. LMedS estimator [25] is one of the robust estimators having a high breakdown point. Its breakdown point is equal to $([n/2] - p + 2)/n$, where p represents the number of unknown parameters and n is the number of observations, with $[z]$ denoting the largest integer less than or equal to z. Its breakdown point can be as large as 50 percent. Characteristics of LMedS are explained from theoretical and experimental points of view [25]. LMedS is derived by replacing a sample sum in (1) by a median operator. Because LMedS has a high breakdown point, it yields good results in discontinuous regions that contain many outliers [23], compared with other estimators. As a modified version, the least trimmed squares (LTS) scheme was proposed to increase the breakdown point and the RLS scheme was presented as a two-stage algorithm to improve the statistical efficiency, where the statistical efficiency represents the ratio between the lowest achievable variance and actual variance produced by the given method depending on the noise distribution. For instance, the mean estimator has an asymptotic statistical efficiency of one whereas median estimator's statistical efficiency is $2/\pi = 0.637$ for Gaussian noise distribution [23]. While the LTS yields good results in homogeneous regions because of the good statistical efficiency, M-estimator does not produce good results. Higher breakdown point and better statistical efficiency of LTS have been theoretically justified in robust statistics [25]. Also recently R-, L-, and total least squares (TLS) estimators in robust statistics have been proposed for various computer vision applications [26], [27].

2.2 Regularization and MAP Estimation

Various low-level vision problems can be solved by regularization, a generalized tool that obtains a smoothed solution by approximating given observations. But this approach yields an oversmoothed solution at discontinuous regions. To alleviate this artifact, many approaches have been presented [4], [5], [6]. On the other hand, MAP estimators using a Markov random field (MRF) with line process were proposed for various applications [12], [17].

Regularization converts ill-posed or ill-conditioned problems into well-posed ones by constraining the solution with a priori assumption. The energy function $J(\Theta, \alpha)$ in Tikhonov's regularization is defined by [1], [6]

$$J(\Theta, \alpha) = \alpha\mathbf{S} + \mathbf{D} = \alpha\mathbf{S} + \sum_{\vec{y}_i \in \mathbf{Y}} \left\|A_i\vec{\theta}_i - \vec{y}_i\right\|^2$$

where $\mathbf{S}$ and $\mathbf{D}$ represent smoothness and data terms, respectively, and Θ denotes the parameter field to be estimated. Note that the smoothness term can be the norm of the first or second derivative of the parameters. A_i is determined depending on the given application and may be observations. The smoothness of a solution depends on α and can be adaptively determined according to a priori knowledge. This energy function defines the error as a linear combination of smoothness of a solution and the sum of the squared differences between observation $\vec{y}_i$ and estimated value $A_i\vec{\theta}_i$. By minimizing this energy function, the solution $\Theta^* = \arg\min_{\Theta} J(\Theta, \alpha)$ is obtained. Recently, the motion estimation method [24] was proposed based on LMedS. This algorithm calculates the optic flow by solving the over-determined linear equations within local image patches.

2.3 Existing Robust Motion Estimation Methods

To reduce the sensitivity of motion estimation to discontinuities and noise, regularization and MAP based on M-estimator were proposed recently [12], [17], [28], [29], [30], [31], [32]. Among them, motion estimation methods based on global optimization were derived by replacing the stabilizing term of the regularization or a priori probability distribution of the MAP estimator by Huber's min-max function or the Lorentzian. Also an algorithm embedding M-estimation into both stabilizing and motion constraint terms was proposed to preserve discontinuities [28], [29]. But discontinuities can be preserved when the initial solution is almost exact. Its proper scale cannot be easily estimated because the initial solution is not known for most motion estimation problems. Also the M-estimation method can not effectively cope with leverage data [25], resulting in poor performance in many cases. As a local regression approach, the motion estimation method based on the TLS [26] was proposed without the constraint of motion continuity. However this algorithm requires a large amount of memory, for example, memory for ten input image frames and 30 filtered frames. Also the estimated motion field is smoothed at motion discontinuities because of the large window size employed. In Bober and Kittler's M-estimation method based on the Hough transform [33], the breakdown point decreases in proportion to the number of parameters and an exhaustive search results in high computational complexity because of the large Hough space. So the method using a variant of the multiresolution and gradient search on discrete grid was proposed to reduce the computation time. They also proposed the robust motion estimation algorithm based on Hough transform and multiresolution MRF (MMRF) for the global interpretation of the local estimates [34].

It is important to have a robust motion estimation method with a high breakdown point because the percentage of outliers is high in many regions such as boundary and corner regions. In boundary regions, there exists approximately 50 percent inliers in practice. Furthermore, there may be much less inliers in some regions containing corner points. Also a single outlier in terms of gray level may yield three outlier observations (E_x, E_y, and E_t), where E_x, E_y, and E_t represent partial derivatives of gray-level E with respect to x, y, and t, respectively. The motion estimation problem is very unstable, especially, at boundary and noisy regions. So a robust motion estimation method with a high breakdown point is required to cope with this problem. We propose an effective hybrid algorithm, a robust MAP motion estimation algorithm, by combining MAP as a global optimization method and robust regression as a local optimization technique. The proposed algorithm is also extended to multiframe motion estimation cases and can also be applied to other applications, such as surface reconstruction or stereo vision. On the other hand, a number of methods to model the motion in multilayers have been proposed [7], [9], [10], [16], [18]. These algorithms yielded good results by computing the multiple models of motions in multiple layers.

3 Proposed RRMAP Motion Estimation Method

In this section, the RRMAP motion estimation method is proposed. At first, we describe it with a pair of image frames, then extend it for multiple image frame cases.

3.1 RRMAP Motion Estimation With Two Image Frames

The energy function proposed by Horn and Schunck [35] for motion estimation is used in the global optimization method that consists of a motion constraint equation and a smoothness constraint. Their energy function $J(\Theta, \alpha)$ is defined by

$$J(\Theta, \alpha) = \sum_{s \in S} \left(\left(E_{s,x} u_s + E_{s,y} v_s + E_{s,t} \right)^2 + \alpha \left(u_{s,x}^2 + v_{s,x}^2 + u_{s,y}^2 + v_{s,y}^2 \right) \right)$$

where S, a set of sites in the $N \times M$ image space, is denoted as $\{s = (i, j) \mid 0 \le i < N, 0 \le j < M\}$. The motion field Θ is represented by $\left\{ \vec{\theta}_s = (\theta_1, \theta_2) = (u_s, v_s) \middle| s \in S \right\}$, assuming the motion vector at a pixel is 2D translational. The constant α is a regularization term, E_s represents the gray level at site s, and $E_{s,x}$, $E_{s,y}$, and $E_{s,t}$ are partial derivatives of E_s with respect to x, y, and t, respectively. By minimizing $J(\Theta, \alpha)$, a smooth motion field is estimated because of the smoothness constraint. A large number of algorithms based on Horn and Shunck's algorithm have been presented [28], [36], however, they all tried to preserve discontinuities in the gray-level and motion field.

We apply the concept of LMedS to motion estimation. Because of ill-posedness, motion estimation cannot be solved by simple regression alone, so we embed the robust operator into the MAP estimator using a priori knowledge such as smoothness constraints. Because the proposed algorithm can discriminate outliers from inliers by using the robust estimator, it is effective at discontinuous regions. We define an a posteriori likelihood probability distribution, $P(\Theta \mid E^{cur}, E^{prev})$, given a degraded image sequence, as

$$P\left(\Theta \middle| E^{cur}, E^{prev}\right) = \frac{P\left(E^{cur} \middle| \Theta, E^{prev}\right) P\left(\Theta \middle| E^{prev}\right)}{P\left(E^{cur} \middle| E^{prev}\right)} \tag{2}$$

where observation image frames E^{cur} and E^{prev} represent the gray level of the current and previous images, respectively. Assuming that the prediction error follows Gaussian probability distribution, the likelihood probability distribution $P(E^{cur} \mid \Theta, E^{prev})$ for ML-type motion estimation is defined by

$$P\left(E^{cur} \middle| \Theta, E^{prev}\right) = \frac{1}{Z_{ML}} \exp\left(-\sum_{s \in S} \sum_{n \in N_s \cup s} \left(E_n^{cur} - E_{n-\vec{\theta}_s}^{prev} \right)^2 \right)$$

where N_s represents the neighborhood of any site s and Z_{ML} denotes a normalizing constant. Note that $P(E^{cur} \mid \Theta, E^{prev})$ depends on the sum of the squared differences between the observed current image E_n^{cur} at site n and an image $E_{n-\vec{\theta}_s}^{prev}$ predicted by the motion vector $\vec{\theta}_s$ at neighborhood sets N_s and s. Because the translational motion within N_s is considered in this paper, we obtain the ML-type conditional distribution:

$$P\left(E^{cur}\middle|\Theta, E^{prev}\right) = \frac{1}{Z_{ML}}\exp\left(-\sum_{s\in S}\sum_{n\in N_s\cup s}\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2\right) \quad (3)$$

by neglecting higher-order terms in the Taylor series expansion and by assuming small $\vec{\theta}_s$. Also, an a priori distribution of the motion field given E^{prev} is defined by

$$P\left(\Theta\middle|E^{prev}\right) = \frac{1}{Z_{AP}}\exp\left(-\sum_{s\in S}\sum_{n\in N_s}\alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right) \quad (4)$$

where Z_{AP} is a normalizing constant and the constant $c_{s,n}^{E^{prev}}$ denoted by

$$c_{s,n}^{E^{prev}} = \frac{1}{1+\left(E_s^{prev}-E_{n-\hat{\theta}_s}^{prev}\right)^2}$$

controls smoothness. Thus, an a priori distribution constrains the available motion field space with respect to changing motion and gray level. By putting (3) and (4) into (2), we can express $P(\Theta \mid E^{cur}, E^{prev})$ as

$$P\left(\Theta\middle|E^{cur}, E^{prev}\right) = \frac{1}{Z_T}\exp\left(-\sum_{s\in S}\sum_{n\in N_s\cup s}\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2\right) \exp\left(-\sum_{s\in S}\sum_{n\in N_s}\alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right)\Bigg/P\left(E^{cur}\middle|E^{prev}\right) \quad (5)$$

where $Z_T = Z_{ML}Z_{AP}$ represents an overall normalizing constant. Note that (5) is formulated by considering the local and global optimization, where MAP estimator corresponds to a global optimization method if the second summations in the exponential terms are removed. The second summation term corresponds to LS regression estimator of a local optimization method. Equation (5) corresponds to MAP estimator as a global optimization method consisting of local regression estimators at every site s in the image space. By simplification of (5), $P(\Theta \mid E^{cur}, E^{prev})$ is rewritten as

$$P\left(\Theta\middle|E^{cur}, E^{prev}\right) = \frac{\prod_{s\in S}\left(\exp\left(-\sum_{n\in N_s\cup s}\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2-\sum_{n\in N_s}\alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right)\right)}{P\left(E^{cur}\middle|E^{prev}\right)}.$$

Next, as in the derivation of the LMedS [25], if we replace the summation of errors by the median of errors, the proposed a posteriori probability distribution is given by

$$P\left(\Theta\middle|E^{cur}, E^{prev}\right) = \frac{\prod_{s}\left(\exp\left(-\mathrm{med}_{n\in N_s\cup s}\left\{\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2+\alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right\}\right)\right)}{P\left(E^{cur}\middle|E^{prev}\right)}$$

The above equation signifies that LS as local regression is replaced by LMedS. Maximization of the a posteriori probability is converted into a minimization problem. Because $P\left(E^{cur}\middle|E^{prev}\right)$ is independent of the estimated parameters, the energy function $J(\Theta, \alpha)$ can be rewritten as

$$\begin{aligned} J(\Theta,\alpha) &= \sum_{s\in S}\mathrm{med}_{n\in N_s\cup s}\left\{\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2 + \alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right\} \\ &= \sum_{s\in S}\mathrm{med}_{n\in N_s\cup s} r_{s,n} = \sum_{s\in S} J_s\left(\vec{\theta}_s,\alpha\right) \end{aligned}$$

where J_s represents the local energy function at s, $r_{s,n}$ signifies the regularization error at n of s, and $\vec{\theta}_s$ denotes the motion vector $\left(u_s, v_s\right)$ at s. In this energy function, if the a priori knowledge term is removed, it can be considered as a cost function for the ML-type LMedS motion estimator. By adopting an a priori distribution to convert the ill-posed nature of motion estimation into well-posed one, we can obtain the continuous motion field. Also, by employing the median operator, the proposed algorithm is robust to motion discontinuities and noise.

The energy function of the proposed algorithm is not convex because of the median operation and local interaction between neighboring pixels. A noniterative optimization solution can not be derived, thus, a deterministic algorithm is employed by assuming a possible solution at local minimum. With the assumption of local interaction, the motion field is iteratively estimated by minimizing J_s, which is defined by

$$J_s\left(\vec{\theta}_s,\alpha\right) = \mathrm{med}_{n\in N_s\cup s}\left\{\left(E_{n,x}u_s + E_{n,y}v_s + E_{n,t}\right)^2 + \alpha c_{s,n}^{E^{prev}}\left(\left(u_s-u_n\right)^2+\left(v_s-v_n\right)^2\right)\right\}.$$

A relaxation equation for local optimization is defined by

$$\left(u_s^{(k)}, v_s^{(k)}\right) = \arg\min_{u_s^{(k)}, v_s^{(k)}}\mathrm{med}_{n\in N_s\cup s}\left\{\left(E_{n,x}u_s^{(k)} + E_{n,y}v_s^{(k)} + E_{n,t}\right)^2 + \alpha c_{s,n}^{E^{prev}}\left(\left(u_s^{(k)}-u_n^{(k-1)}\right)^2+\left(v_s^{(k)}-v_n^{(k-1)}\right)^2\right)\right\} \quad (6)$$

where the superscript (k) represents an iteration index, $1 \le k \le K$. It is not possible to get a closed form solution for (6), so we obtain a solution by the same sampling scheme as used in LMedS estimation [25]. This sampling scheme at first selects n observation sets each of which consists of p observations, where p denotes the number of parameters to be estimated. With the parameters estimated for each observation set, errors are calculated for all observations. Then the parameters giving the minimum median error is selected.

In the proposed motion estimation, the set of observations $\left(E_x, E_y, E_t, u, v\right)$ is considered, where observations E_x and E_y are obtained from forward, backward, and central differences, and observation E_t at each set of eight neigh-

boring sites. u and v are individual components of the estimated motion vector in neighborhood region of s. With these observation sets, we can construct a large number of combinatorial samples. Because the computation time is proportional to the number of samples and motion estimation is iteratively calculated, it is necessary to reduce the number of samples. Considering the consistency of locality, nine observation sets

$$\{(E_x, E_y, E_t, u, v)\} =$$
$$\{ (E_{x,(b)}, E_{y,(b)}, E_{t,(1)}, u_{(1)}, v_{(1)}), (E_{x,(c)}, E_{y,(b)}, E_{t,(2)}, u_{(2)}, v_{(2)}),$$
$$(E_{x,(f)}, E_{y,(b)}, E_{t,(3)}, u_{(3)}, v_{(3)}), (E_{x,(f)}, E_{y,(c)}, E_{t,(4)}, u_{(4)}, v_{(4)}),$$
$$(E_{x,(f)}, E_{y,(f)}, E_{t,(5)}, u_{(5)}, v_{(5)}), (E_{x,(c)}, E_{y,(f)}, E_{t,(6)}, u_{(6)}, v_{(6)}),$$
$$(E_{x,(b)}, E_{y,(f)}, E_{t,(7)}, u_{(7)}, v_{(7)}), (E_{x,(b)}, E_{y,(c)}, E_{t,(8)}, u_{(8)}, v_{(8)}),$$
$$(E_{x,(c)}, E_{y,(c)}, E_{t,(mean)}, u_{mean}, v_{mean})\}$$

are selected as observation samples at each site in the proposed algorithm, where the subscripts (b), (c), and (f) represent backward, central, and forward, respectively, and the subscript (n), $1 \le n \le 8$, denotes the nth neighboring site around s, numbered clockwise from the top left pixel. The subscript "mean" represents the mean value of observations at the eight neighboring sites. With these observation samples, the residual errors are computed, and the parameters minimizing the median error are determined as the solution of the robust MAP based on the LMedS. In the proposed algorithm, the uniform sampling scheme is used. Note that the probability giving a good estimate with a small number of samples for LMedS is close to one. Because the order of samples is arbitrary for each patch, the uniform sampling yields the results similar to those of the random sampling. Furthermore, the proposed system can reliably estimate a reasonable solution by using the reweighted LS.

To improve the statistical efficiency, we apply a RLS scheme [25] to the proposed robust MAP, resulting in a reweighted robust MAP (RRMAP). With the parameters estimated by a sampling scheme, an initial scale [25] is computed by

$$scale_s^{(k)} = 1.4826(1+\beta)\sqrt{J_s(\vec{\theta}_s, \alpha)}$$

where β represents a term compensating for the finite sample effect. We consider β as a parameter because it cannot be mathematically determined because of the a priori distribution of the motion field. Weights are determined as

$$w_{s,n}^{(k)} = \begin{cases} 1, & \left|\dfrac{r_{s,n}}{\text{scale}_s^{(k)}}\right| \le 2.5 \\ 0, & \text{otherwise} \end{cases}$$

based on the statistical criterion that distinguishes inliers from outliers. With the selected weights, the final motion vector is estimated as

$$\left(u_s^{*(k)}, v_s^{*(k)}\right) = \arg\min_{u_s, v_s} \sum_{n \in N_s \cup s} w_{s,n}^{(k)} r_{s,n}. \quad (7)$$

Our algorithm can be summarized as follows:

for all k {
  for all pixel s {
    for all samples {
      compute $\left(u_s^{(k)}, v_s^{(k)}\right)$ by (6)
      compute error with $\left(u_s^{(k)}, v_s^{(k)}\right)$ for all samples
      store the median value of errors }
    select $\left(u_s^{(k)}, v_s^{(k)}\right)$ that minimizes the stored median value
    compute $\left(u_s^{*(k)}, v_s^{*(k)}\right)$ by (7) } }

Compared with Horn and Schunck's algorithm, the proposed algorithm requires additional computation of motion vector estimation, the computational complexity of which depends on the number of samples.

High computational complexity is required for pixelwise motion estimation of an entire image. The sampling scheme also requires large computational resources to guarantee convergence. To reduce the computation time, a coarse-to-fine control scheme is adopted, with three-level hierarchy in both the image space and motion field. A 4×4 Gaussian filter, sampler, and interpolator are used for either downsampling or upsampling by a factor of two in horizontal and vertical directions. At first, three-level image pyramids of a pair of frames are constructed by using the Gaussian filter and sampler. Next, optical flow is computed from coarse level to fine level, in which optical flow estimated in higher level is interpolated in the next lower level with the upsampling rate equal to two in both directions.

3.2 RRMAP Motion Estimation With Multiple Image Frames

Most conventional robust motion estimation methods use a pair of images. A recursive robust optical flow estimation algorithm was proposed by Black [31]. It computes the current optical flow by prediction with optical flow computed in the previous frame, based on the temporal continuity of optical flow. We extend the proposed robust motion estimation algorithm to multiple-image frame cases. For multiple image frames, Weber and Malik adopted a TLS regression scheme, assuming that the motion field is continuous during short time intervals [26]. Thus, their algorithm is not robust to scene changes. Whereas, the proposed algorithm detects outliers in motion vectors and copes well with scene changes in an image sequence, resulting in an algorithm robust to scene changes. Also it is efficient for image sequences having a smooth motion. Note that it is difficult to apply the proposed RRMAP algorithm directly to multiple image frames because of the high computational complexity, and of motion discontinuities or scene changes in successive image frames. The RRMAP algorithm is extended to multiple frame cases by separating inliers from outliers using the motion field estimated between two successive image frames. That is, the proposed motion estimation algorithm with multiple frames can be considered as a two-stage algorithm. At first, the motion field estimated from two successive image frames is calculated by (7). Next, the multiframe weight at pixel s is determined by

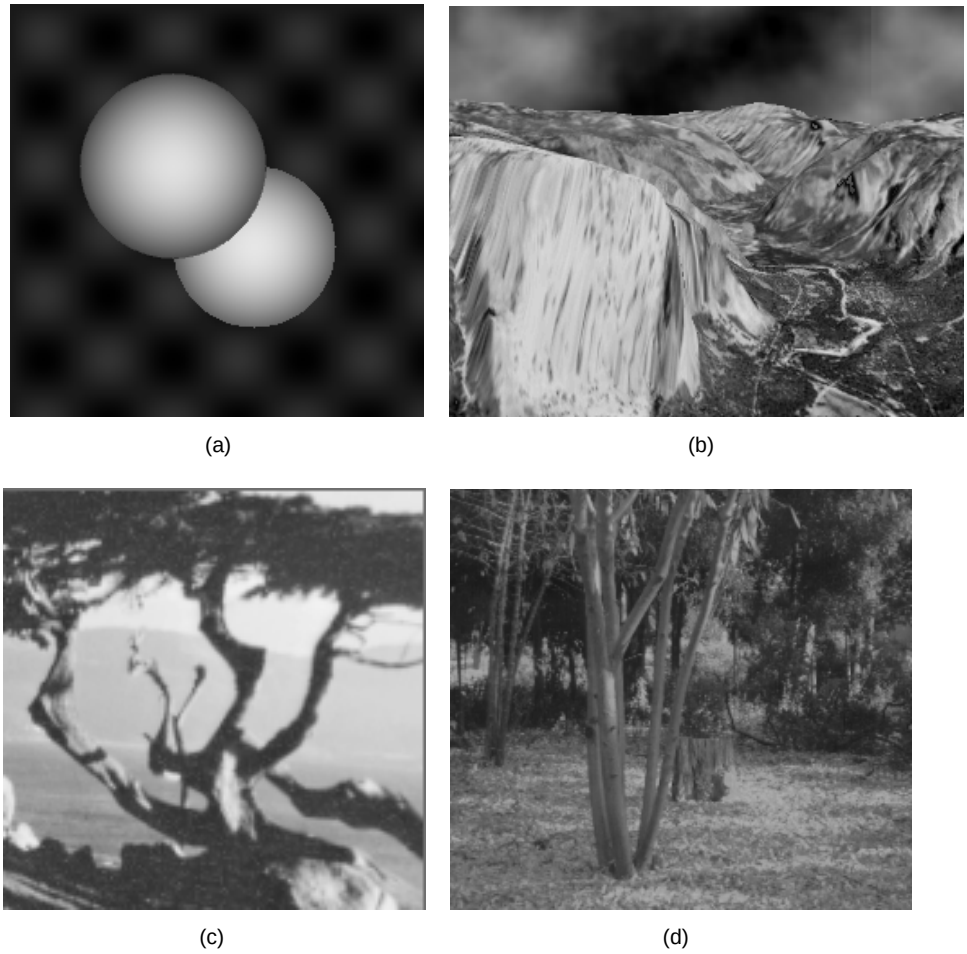


Fig. 1. Synthetic and real image sequences used in experiments. (a) Synthetic Circle sequence (256 × 256). (b) Synthetic Yosemite sequence (316 × 252). (c) Synthetic Tree sequence (150 × 150). (d) Real SRI Tree sequence (256 × 233).

$$w_{s,n}^{l} = \begin{cases} 1, & \left| \frac{r_{s,n+l\vec{\theta}_s}^{l}}{\text{scale}_s^{(K)}} \right| \leq 2.5 \\ 0, & \text{otherwise} \end{cases}$$

according to the estimated residuals, where the superscript l represents a sequential ordering of the residual field that is computed between a pair of image frames, the subscript $n + l\vec{\theta}_s$ denotes a motion trajectory in terms of l based on the first-order prediction at the neighboring pixel n of s, and K is the maximum number of iterations. With these weights, the motion vector at s is calculated by

$$\left(u_s^*, v_s^*\right) = \arg\min_{u_s, v_s} \sum_{|l| \leq F} \sum_{n \in N_s \cup s} w_{s,n}^{l} \left(r_{s,n+l\vec{\theta}_s}^{l} \right)$$

in multiple-image frames, where $2F + 2$ image frames are used and $l = 0$ represents the two center frames among the $2F + 2$ image frames. Because of the reweighting scheme in multiple image frame cases, the proposed algorithm shows robustness to abrupt scene changes.

4 Experimental Results and Discussions

In this section, we show the effectiveness of the proposed algorithm by computer simulations with several synthetic and real image sequences. With the synthetic image sequences, the performance of the proposed algorithm is compared with that of three existing robust methods in terms of two error measures: angle average error (a.e.) and angle standard deviation (s.d.) [37], where the angle error represents $\arccos\left(\vec{\theta}_t \cdot \vec{\theta}_e\right)$ between the true 3D unit velocity $\vec{\theta}_t$ and the estimated 3D unit velocity $\vec{\theta}_e$. Note that the 3D unit velocity $\vec{\theta}_t$ is defined by $\vec{\theta}_t = \frac{1}{\sqrt{u_t^2 + v_t^2 + 1}}\left(u_t, v_t, 1\right)^T$, where u_t and v_t represent the components of the true motion vector. Similarly, $\vec{\theta}_e$ is defined. Existing methods compared in experiments include Black and Anandan's algorithm [32] based on M-estimation, Weber and Malik's algorithm [26] based on TLS, and Bober and Kittler's algorithm [33] based on the robust Hough algorithm. Also the performance of these algorithms against additive Gaussian noise and impulse noise is compared.

Fig. 1 shows synthetic and real image sequences used in the experiments. Fig. 1a (256 × 256 Circle sequence consisting of 10 frames) shows the previous image synthesized from the current image using the specified motion. For an upper circle, the motion field is specified by $u(x, y) = R\sin\left(\pi \frac{y-100}{180}\right)$ and $v(x, y) = R\sin\left(\pi \frac{x-100}{180}\right)$, where R

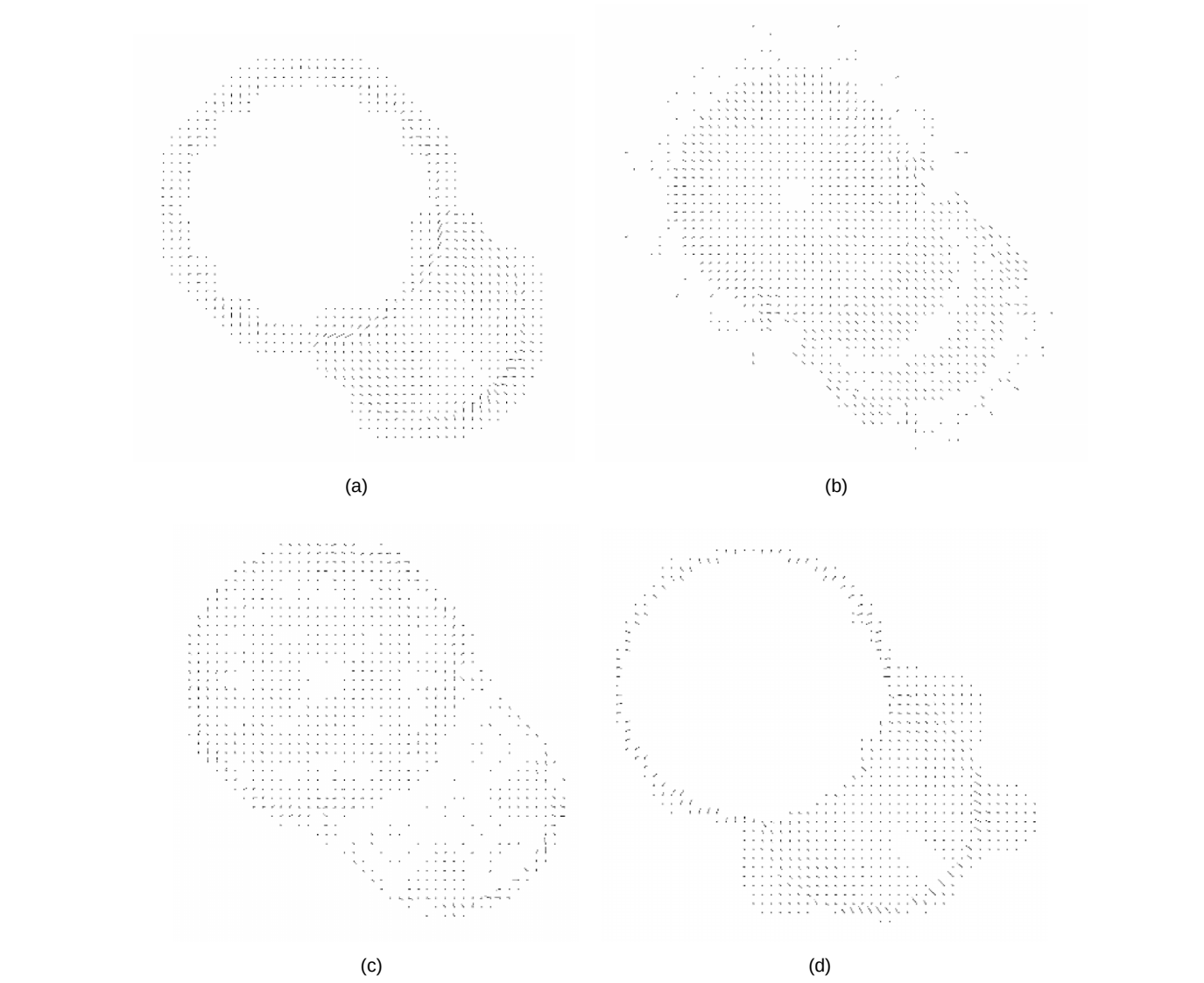


Fig. 2. Comparison of motion estimation errors for the synthetic Circle sequence (R = 3). (a) Black and Anandan's algorithm. (b) Weber and Malik's algorithm. (c) Bober and Kittler's algorithm. (d) Proposed algorithm.

denotes a zooming parameter and $(x, y) = (0, 0)$ corresponds to top left position of an image. For a lower right circle, the motion parameter is specified by $(dx, dy, d\omega) = (R, R, 0)$, where dx and dy denote translation parameters along the x and y axes, respectively, and $d\omega$ signifies the clockwise rotation angle. Fig. 1b (316 × 252 Yosemite sequence consisting of 15 frames) and Fig. 1c (150 × 150 Tree sequence consisting of 30 frames) illustrate realistic synthetic images. Images from synthetic image sequences are shown in Figs. 1a, 1b, and 1c, and they are employed to compare the quantitative performance of each method, whereas the real image sequence shown in Fig. 1d (256 × 233 SRI Tree sequence consisting of 20 frames) is used to compare the qualitative performance. We synthesize the test sequence in Fig. 1a and other three test sequences are obtained from ftp site (ftp.csd.uwo.ca/pub/vision) [37].

For performance comparison of each algorithm, Fig. 2 and Fig. 4 show motion estimation errors for the synthetic data, subsampled by a factor of four in both directions, whereas Fig. 6 illustrates horizontal motions estimated for the real sequence. Motion estimation error is defined by the vector difference between the estimated motion vector and the true one that is known a priori.

The parameter values for each method are experimentally determined. Small changes in these parameter values do not affect the final results much. In Black and Anandan's algorithm, the Lorentzian function is used as error norm. Also three-level pyramid using a six-stage continuation method [32] is used, in which the scale σ_D for the data term varies linearly from $18/\sqrt{2}$ to $5/\sqrt{2}$ and the scale σ_S for the smoothness term changes linearly from $3/\sqrt{2}$ to $0.03/\sqrt{2}$. We use the weight $\lambda_D = 5$ for the data term and the weight $\lambda_S = 1$ for the smoothness term in simultaneous over relaxation (SOR). Also the number of iterations at each level of the pyramid is set to 100. In Weber and Malik method, the number of scales is set to five and the number of bands per scale is set to six, where 30 spatiotemporal bandpass filters are used. The initial scale of the bandpass filter is set to 0.5

and the ratio between adjacent scales is set to $\sqrt{2}$. Also arithmetic progression is employed as a scale progression scheme, a spatial ratio of filters is set to 1.5, and the border size to nine. In Bober and Kittler's algorithm, a four-parameter model is used in a 15×15 block. The motion vector is calculated pixelwise based on the search method and subpixelwise based on the gradient descent method. Because Weber and Malik's algorithm cannot estimate the motion vectors at border regions of the image, these regions are excluded in computing error measures. Motion vectors at pixels not defined are interpolated to fairly compare the performance of different algorithms, and the performance of each algorithm is evaluated in terms of two error measures. In the proposed algorithm, we use $\alpha = 10$, $\beta = 0$, and $K = 200$. All algorithms except for Weber and Malik's algorithm use a pair of image frames. Any prefiltering scheme is not employed to fairly compare the performance of algorithms.

The computation time of algorithms based on the relaxation scheme depends on the number of iterations. Computer simulation is done with MIPS R-4000 workstation. For the Yosemite sequence with parameters mentioned above, Black and Anandan's algorithm takes about 16 minutes. Bober and Kittler's algorithm takes about 28 minutes with an exhaustive search whereas the proposed algorithm takes 70 minutes. Also the noniterative method, Weber and Malik's algorithm takes 87 minutes. Most of existing and proposed algorithms are based on robust statistics, requiring high computational complexity. So the reduction of the computation time is to be investigated.

For the Circle sequence, all algorithms except for Weber and Malik's algorithm estimate the motion field between the fifth and sixth frames whereas Weber and Malik's algorithm requires 10 frames with the fifth and sixth frames in the middle of a total of 10 frames. For the Yosemite and Tree sequences, the motion fields between 10th and 11th frames are estimated. Also the motion fields between the 10th and 11th frames are estimated for the SRI Tree sequence.

4.1 Motion Estimation Results for Synthetic Image Sequences

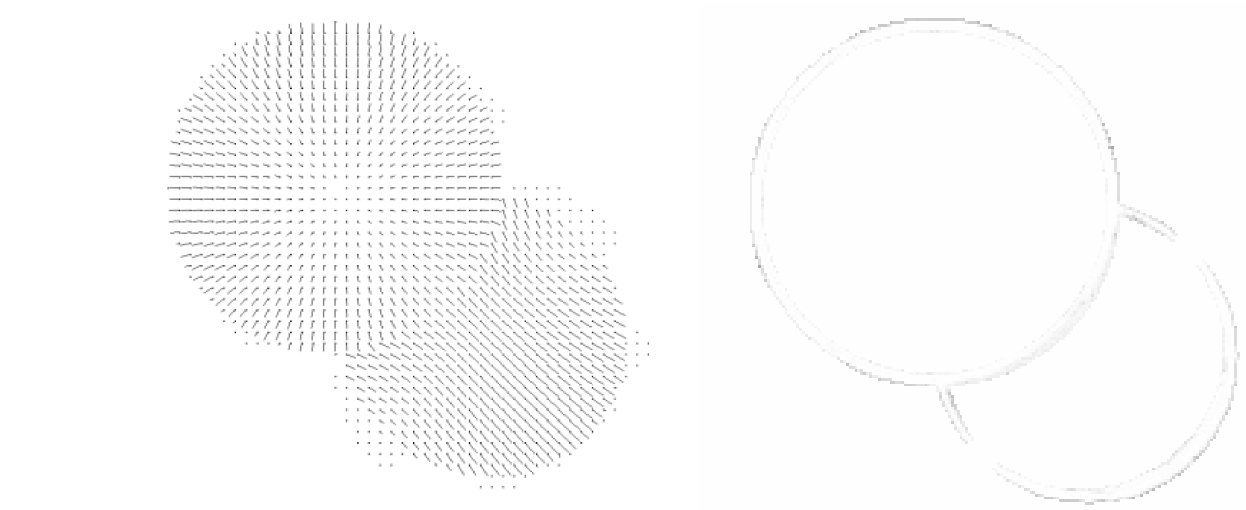
Table 1 shows comparison of motion estimation results for the Circle sequence with varying motion parameter R. Because the test sequence contains abrupt discontinuities of gray level and motion, accurate motion estimation is difficult. Table 1 shows that the proposed algorithm yields better performance than the existing methods simulated, in terms of quantitative error measures. The proposed algorithm adopts the robust MAP based on the LMedS scheme, thus it can estimate accurately not only discontinuous but also smooth motion field. Black and Anandan's algorithm based on M-estimation is degenerated by leverage data, where the breakdown point is at most 33 percent. Weber and Malik's algorithm can estimate continuous motions because of regression with several image frames, however, discontinuous motion field is smoothed and is severely deteriorated if a large window is employed for regression. The proposed algorithm based on the median operator has a high breakdown point, thus its performance is not affected much by leverage data.

TABLE 1
COMPARISON OF MOTION ESTIMATION RESULTS WITH VARYING R (SYNTHETIC CIRCLE SEQUENCE)

R	Methods	Error measures	
		Angle a.e.	Angle s.d.
1	Black and Anandan	3.213	7.411
	Weber and Malik	5.413	9.431
	Bober and Kittler	3.156	7.437
	Proposed	2.411	3.963
2	Black and Anandan	5.816	8.129
	Weber and Malik	6.192	11.431
	Bober and Kittler	4.159	8.248
	Proposed	2.594	5.194
3	Black and Anandan	6.247	9.823
	Weber and Malik	7.714	12.121
	Bober and Kittler	4.815	9.379
	Proposed	2.843	6.241
4	Black and Anandan	6.796	10.134
	Weber and Malik	8.127	12.989
	Bober and Kittler	5.018	9.948
	Proposed	3.011	6.924
5	Black and Anandan	7.771	11.414
	Weber and Malik	8.961	13.414
	Bober and Kittler	6.043	10.694
	Proposed	3.381	7.513

Fig. 2 shows the error of the motion field estimated by each method, for the synthetic Circle sequence with $R = 3$. Because the proposed algorithm can preserve motion discontinuities by embedding LMedS having higher breakdown point into MAP, it yields smaller error near occluded regions than the existing algorithms. The proposed algorithm estimates the motion field based on hierarchical RRMAP, thus it also yields smaller error than other robust estimation algorithms in homogeneous regions as shown in Fig. 3a. The proposed algorithm is based on a reweighted scheme and it has higher statistical efficiency than other existing ones, thus in homogeneous regions, it yields good results similar to Horn and Schunck's [35]. Fig. 3b shows the sum of weights $\sum_{n \in N_s \cup s} w_{s,n}^{(K)}$ of the last iteration at s by the reweighted scheme for the Circle sequence, because it is difficult to show all weight values at neighbors of s. Black regions, motion boundaries, represent pixels containing outliers, while white regions do not contain outliers.

Table 2 shows the performance comparison of the motion estimation results for the Yosemite and Tree sequences. The Yosemite sequence is a realistic synthetic sequence generated by translation of a camera in depth. The clouds in this sequence is produced by the fractal brownian motion. It has large divergent flow vectors and shows the discontinuous motion field between mountains and clouds. Because the motion for the cloud region is stochastic, it is difficult to estimate the motion field. There is no true motion field in this area, thus the error in this area is ignored in the numerical comparison. As shown in Table 2, the proposed algorithm shows satisfactory performance for large zoom. Fig. 4 shows comparison of motion estimation errors for the synthetic Yosemite sequence, in which the proposed algorithm yields smaller error field than the existing ones tested.



(a) (b)

Fig. 3. Estimated motion field and sum of weights for the synthetic Circle sequence (R = 3). (a) Estimated motion field. (b) Sum of weights, $\sum_{n \in N_s \cup s} w_{s,n}^{(K)}$, by the RLS scheme.

In Table 2, results of the motion estimation for the two synthetic Tree image sequences that are generated by translational and zooming motions are also listed, where those sequences satisfy the smoothness constraint of the motion field. The proposed algorithm shows the performance comparable to that of the existing algorithms. The proposed algorithm shows better performance than existing ones in homogeneous regions because it is based on local and global optimization techniques with a large number of observations.

4.2 Motion Estimation Results for Noisy Image Sequences

We compare the performance of existing and proposed algorithms with noisy image sequences. Table 3 shows estimation results for the noisy Circle sequence (R = 3, first to 10th frames) contaminated by additive Gaussian noise and impulse noise. We estimate the motion field for additive Gaussian noisy images with varying standard deviation of the Gaussian noise: 3, 5, 7, and 9. The change rate of the performance with varying noise level is used, where the change rate in precent is defined by

$$|\text{error with noise} - \text{noise-free error}| \,/\, |\text{noise-free error}| \times 100.$$

The proposed algorithm is based on a reweighted scheme, thus, it yields good performance against Gaussian noise.

We also simulate motion estimation methods for impulse noise cases with varying the noise probability, where the noise probability represents the ratio of the number of noisy pixels to the total number of pixels. A noisy pixel has a value of zero or 255 with equal probability. In experiments, all computer simulations are performed for each motion estimation algorithm without preprocessing such as median filtering. Because motion estimation uses derivatives of image brightness, the motion estimation results are severely deteriorated by a few outliers generated by noise sources. It is difficult to accurately estimate motion vectors for noisy image sequences because motion estimation at a pixel is done with the small number of samples in a small finite-size local window, which is also true for other robust methods. Its performance is deteriorated if the number of outliers increases. But the proposed algorithm is robust against impulse noise compared with other algorithms because of the employment of LMedS. While the absolute error of the proposed algorithm is smaller than that of existing ones, the change rate is larger. As the number of outliers increases, the number of inlier samples decreases and the absolute error generated by the proposed algorithm is also small, which results in the large change rate for the proposed algorithm.

TABLE 2
COMPARISON OF MOTION ESTIMATION RESULTS
(SYNTHETIC YOSEMITE AND TREE SEQUENCES)

Sequences	Methods	Error measures	
		Angle a.e.	Angle s.d.
Yosemite sequence	Black and Anandan	4.492	4.325
	Weber and Malik	10.547	17.349
	Bober and Kittler	14.140	19.316
	Proposed	4.127	4.026
Tree sequence (Translation)	Black and Anandan	0.481	0.387
	Weber and Malik	0.513	0.377
	Bober and Kittler	0.358	0.254
	Proposed	0.365	0.233
Tree sequence (Zooming)	Black and Anandan	3.484	3.891
	Weber and Malik	4.011	3.117
	Bober and Kittler	3.719	4.442
	Proposed	3.621	3.075

4.3 Motion Estimation Results for Multiple Image Frames

For the Yosemite sequence, motion estimation error of the proposed RRMAP algorithm for multiple image frames is (angle a.e., angle s.d.) = (4.01, 3.93). We use 10 frames (sixth

TABLE 3
COMPARISON OF MOTION ESTIMATION RESULTS FOR ADDITIVE GAUSSIAN NOISE AND IMPULSE NOISE (SYNTHETIC CIRCLE SEQUENCE, $R = 3$)

Noise type (Noise level)	Noise level	Methods	Error measures: Angle a.e. (change rate)	Error measures: Angle s.d. (change rate)
Gaussian (s.d.)	3	Black and Anandan	6.286 (1.1%)	9.900 (0.8%)
		Weber and Malik	7.991 (3.6%)	13.491 (11.3%)
		Bober and Kittler	4.961 (3.0%)	9.843 (4.9%)
		Proposed	2.891 (1.7%)	6.374 (2.1%)
	5	Black and Anandan	6.827 (9.5%)	10.110 (3.1%)
		Weber and Malik	8.413 (9.1%)	14.318 (18.1%)
		Bober and Kittler	5.716 (18.7%)	11.024 (17.5%)
		Proposed	3.027 (6.5%)	7.496 (20.1%)
	7	Black and Anandan	7.687 (23.5%)	10.915 (11.4%)
		Weber and Malik	9.317 (20.8%)	14.943 (23.3%)
		Bober and Kittler	6.374 (32.4%)	12.306 (31.2%)
		Proposed	3.647 (28.3%)	8.486 (36.0%)
	9	Black and Anandan	8.683 (39.1%)	12.819 (30.5%)
		Weber and Malik	9.915 (28.5%)	15.782 (30.2%)
		Bober and Kittler	7.201 (49.6%)	13.435 (43.2%)
		Proposed	4.484 (57.7%)	10.176 (63.1%)
Impulse (Percentage)	3	Black and Anandan	7.094 (14.3%)	11.317 (15.8%)
		Weber and Malik	8.794 (14.0%)	15.236 (25.7%)
		Bober and Kittler	5.746 (19.3%)	11.468 (22.4%)
		Proposed	3.597 (26.5%)	8.017 (28.5%)
	5	Black and Anandan	7.867 (26.7%)	12.417 (27.1%)
		Weber and Malik	9.541 (23.7%)	16.713 (37.9%)
		Bober and Kittler	6.923 (43.8%)	13.368 (42.5%)
		Proposed	4.193 (47.5%)	9.817 (57.3%)
	7	Black and Anandan	9.557 (53.0%)	13.918 (42.8%)
		Weber and Malik	11.003 (42.6%)	18.015 (48.6%)
		Bober and Kittler	8.246 (71.3%)	15.204 (62.1%)
		Proposed	5.514 (94.0%)	10.968 (64.1%)
	9	Black and Anandan	11.536 (85.7%)	16.748 (71.5%)
		Weber and Malik	13.110 (70.0%)	19.516 (61.0%)
		Bober and Kittler	10.146 (110.7%)	17.277 (84.2%)
		Proposed	6.814 (139.7%)	13.473 (115.9%)

to 15th frames) and estimate the motion field between the 10th and 11th frames. Comparing the result for multiple frame cases with that for two frame cases, we observe that the performance of the proposed algorithm using multiple frames is better than that using only two frames.

Next, we simulate with an image sequence having a scene change, which is constructed by concatenating the Circle and Claire sequences. We construct the test image sequence consisting of ten frames by combining the Circle sequence ($R = 3$, first to sixth frames) and the Claire sequence (seventh to 10th frames) as shown in Fig. 5, where both sequences are 256×256 and the scene change occurs between the sixth and seventh frames. The motion field is calculated between the fifth and sixth frames. We compare the performance of Weber and Malik's algorithm and that of the proposed algorithm using multiple-image frames. Estimation errors of Weber and Malik's algorithm and the proposed algorithm are (18.316, 20.314) and (2.781, 5.907), respectively. Because Weber and Malik's algorithm is based on TLS and uses a large spatio-temporal filter, observations are oversmoothed along the temporal axis, resulting in poor performance for the test sequence having a scene change. But the proposed algorithm is robust to scene changes because of the RRMAP estimation along the temporal axis. The multiframe weights $w_{s,n}^{l}$ at the cross mark in the fifth frame ($s = n = (160, 160)$) are {1, 1, 1, 1, 1, 0, 0, 0, 0}. That is, the pairs of frames from $l = -4$ to $l = 0$ are considered as inliers, while other frames are considered as outliers. As a result, Claire images are excluded in estimating the motion field between the fifth and sixth frames of the Circle sequence. For smooth motion along the temporal axis, because a lot of observations contaminated by Gaussian noise are detected as inliers, the proposed algorithm using multiple image frames yields better performance than that using a pair of image frames. Temporal robust motion estimation yields the statistical efficiency and robustness for estimation of smooth and discontinuous motions along the temporal axis.

4.4 Motion Estimation Results for Real Image Sequences

We also compare motion estimation results for the real image sequence. Since the real motion field is not known a priori, the performance of each algorithm is subjectively compared. Fig. 6 shows the horizontal flow of the motion field estimated by each method for the SRI Tree sequence.

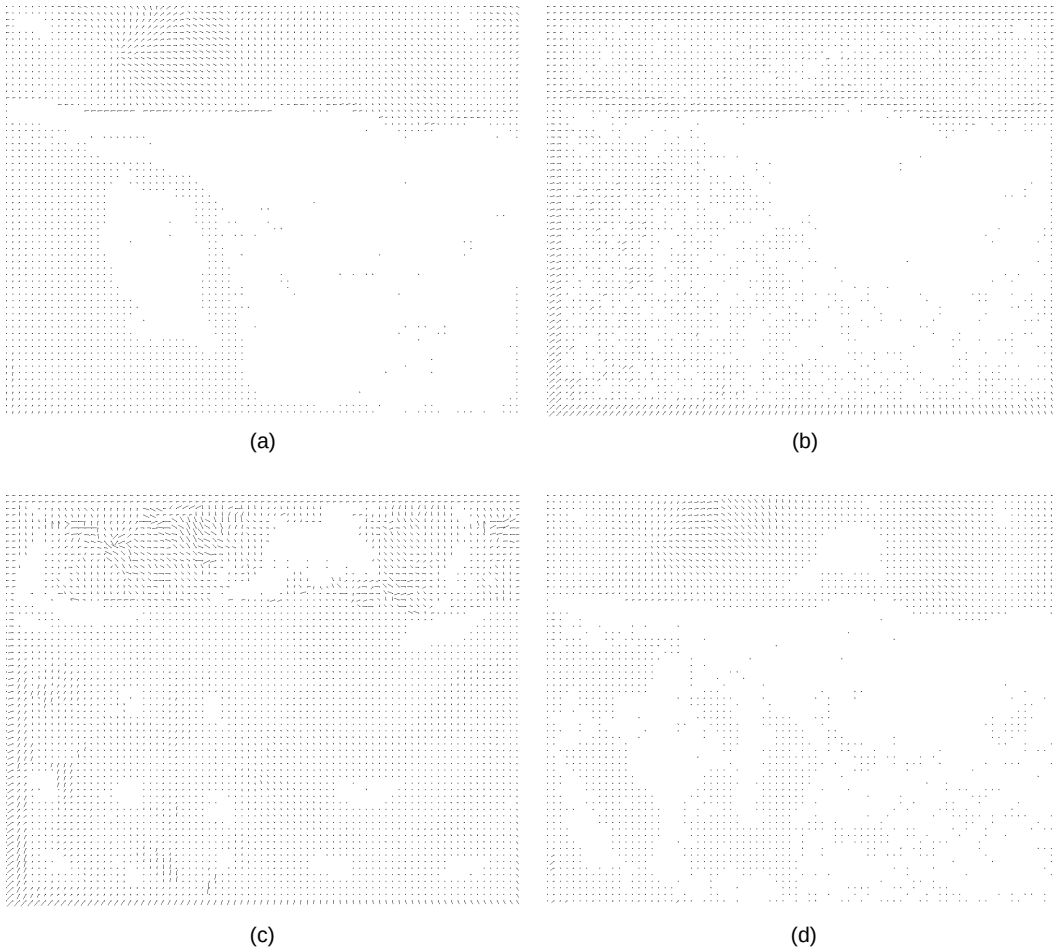


Fig. 4. Comparison of motion estimation errors for the synthetic Yosemite sequence. (a) Black and Anandan's algorithm. (b) Weber and Malik's algorithm. (c) Bober and Kittler's algorithm. (d) Proposed algorithm.

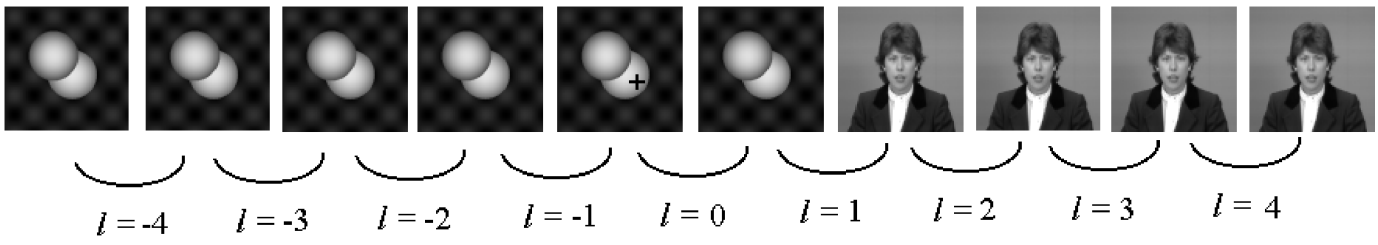


Fig. 5. Frame construction for motion estimation with multiple-image frames.

The amount of horizonal flow u_s is represented by a gray-level image, in which gray level is expressed as $u_s \times 50 + 20$ for easy comparison. The proposed algorithm preserves motion discontinuities between trees and background better than the existing ones. Weber and Malik's and Bober and Kittler's algorithm are sensitive to noise and discontinuities. Black and Anandan's algorithm is comparable to the proposed algorithm, however their algorithm yields the disperse motion field in several regions.

5 Conclusion

In this paper, we propose the RRMAP motion estimation algorithm that is robust to motion discontinuity and noise, by embedding a robust median operator into a MAP estimator. It is effective for both smooth and discontinuous motions. Also the proposed RRMAP is extended to multiple image frame cases having scene changes. Computer simulation with various synthetic sequences shows that the proposed algorithm yields better performance, compared to existing ones tested. Also the proposed algorithm is robust to Gaussian noise and impulse noise, and it yields the more natural motion field for real image sequences.

Further research will focus on application of the proposed algorithm to various computer vision paradigms such as 3D motion estimation and surface reconstruction.

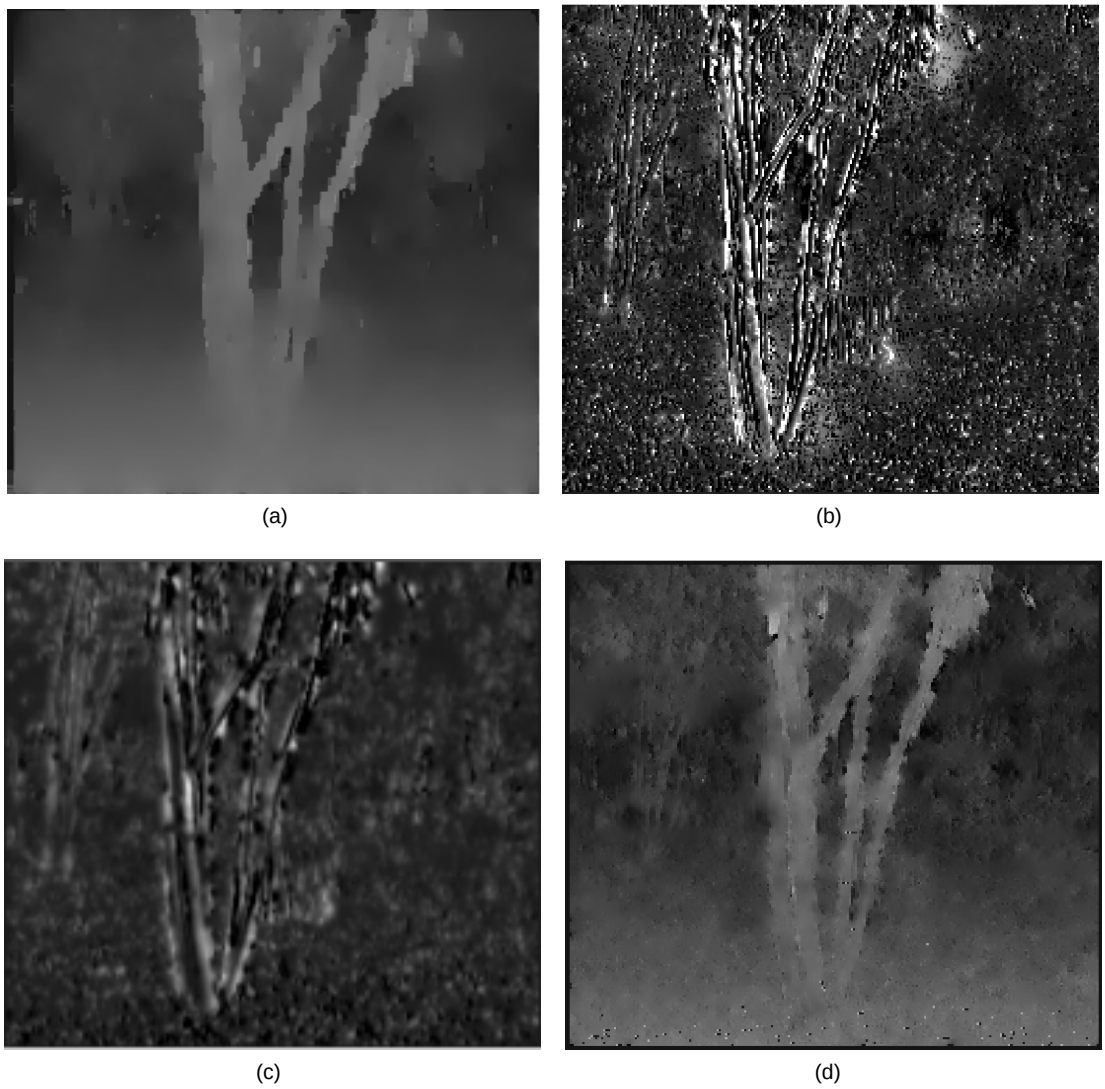


Fig. 6. Comparison of the detected horizontal motion in terms of gray level for the real SRI Tree sequence. (a) Black and Anandan's algorithm. (b) Weber and Malik's algorithm. (c) Bober and Kittler's algorithm. (d) Proposed algorithm.

ACKNOWLEDGMENTS

We would like to thank J. Weber for providing the source code [26]. We obtained the code for Black and Anandan algorithm [28] from website http://www.parc.xerox.com/black. Also the authors thank the anonymous referees for many valuable comments.

REFERENCES

[1] M. Bertero, T.A. Poggio, and V. Torre, "Ill-Posed Problems in Early Vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869-889, Aug. 1988.

[2] S.A. Kassam and H.V. Poor, "Robust Techniques for Signal Processing: A Survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433-481, Mar. 1985.

[3] X. Zhuang, T. Wang, and P. Zhang, "A Highly Robust Estimator Through Partially Likelihood Function Modeling and Its Application in Computer Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, pp. 19-35, Jan. 1992.

[4] J.K. Aggarwal and N. Nandhakumar, "On the Computation of Motion From Sequences of Images—A Review," *Proc. IEEE*, vol. 76, no. 8, pp. 917-935, Aug. 1988.

[5] J.F. Vega-Riveros and K. Jabbour, "Review of Motion Analysis Techniques," *IEE Proc*. pt. 1, vol. 136, no. 6, pp. 397-404, Dec. 1989.

[6] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *J. Am. Statistical Assoc*., vol. 82, no. 397, pp. 76-89, Mar. 1987.

[7] A. Jepson and M.J. Black, "Mixture Models for Optical Flow Computation," *Proc. Computer Vision Pattern Recognition*, pp. 760-761, New York, June 1993.

[8] E. Szeliski and H.Y. Shum, "Motion Estimation With Quadtree Splines," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 757-763, Boston, June 1995.

[9] Y. Weiss and E.H. Adelson, "A Unified Mixture Framework for Motion Segmentation: Incorporating Spatial Coherence and Estimating the Number of Models," *Proc. Computer Vision Pattern Recognition*, pp. 321-326, San Francisco, June 1996.

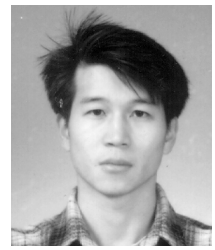
[10] T. Darrell and A. Pentland, "Cooperative Robust Estimation Using Layers of Support," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 474-487, May 1995.

[11] J.J. Clark and A.L. Yuille, *Data Fusion for Sensory Information Processing Systems*. Norwell, Mass.: Kluwer Academic Publishers, 1990.

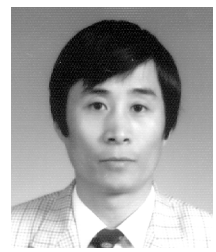
[12] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation," *IEEE Trans. Image Processing*, vol. 2, no. 3, pp. 296-310, July 1993.

[13] D.G. Sim and R.H. Park, "A Two-Stage Algorithm for Motion-Discontinuity Preserving Optical Flow Estimation," *Computer Vision and Image Understanding*, vol. 65, no. 1, pp. 19-37, Jan. 1997.

[14] D.G. Sim and R.H. Park, "Anisotropic Hierarchical Motion Estimation Method Based on Decomposition of the Functional Domain," *J. Visual Comm. Image Representation*, vol. 7, no. 3, pp. 259-272, Sept. 1996.
[15] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *Int'l J. Computer Vision*, vol. 2, no. 3, pp. 283-319, Sept. 1989.
[16] S.X. Ju, M.J. Black, and A.D. Jepson, "Skin and Bones: Multi-Layer, Locally Affine, Optical Flow and Regularization With Transparency," *Proc. Computer Vision Pattern Recognition*, pp. 307-314, San Francisco, June 1996.
[17] R.L. Stevenson, B.E. Schmitz, and E.J. Delp, "Discontinuity Preserving Regularization of Inverse Visual Problems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 24, no. 3, pp. 455-469, Mar. 1994.
[18] S. Ayer and H.S. Sawhney, "Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 777-784, Boston, June 1995.
[19] X. Yu, T.D. Bui, and A. Krzyzak, "Robust Estimation for Range Image Segmentation and Reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 530-538, May 1994.
[20] R.L. Kashyap and K.B. Eom, "Robust Image Modeling Techniques With an Image Restoration Application," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 8, pp. 1,313-1,325, Aug. 1988.
[21] A.J. Koive and C.W. Kim, "Robust Image Modeling for Classification of Surface Defects on Wood Boards," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1,659-1,666, Nov. 1989.
[22] I. Pitas and A.N. Venetsanopoulos, *Nonlinear Digital Filters*. Norwell, Mass.: Kluwer Academic Publishers, 1990.
[23] P. Meer, D. Mintz, D.Y. Kim, and A. Rosenfeld, "Robust Regression Methods for Computer Vision: A Review," *Int'l J. Computer Vision*, vol. 6, no. 1, pp. 59-70, Apr. 1991.
[24] A. Bab-Hadiashar and D. Suter, "Optic Flow Calculation Using Robust Statistics," *Proc. Computer Vision Patern Recognition*, pp. 988-993, San Juan, Puerto Rico, June 1997.
[25] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, 1987.
[26] J. Weber and J. Malik, "Robust Computation of Optical Flow in a Multi-Scale Differential Framework," *Int'l J. Computer Vision*, vol. 14, no. 1, pp. 67-81, Jan. 1995.
[27] S.S. Sinha and B.G. Schunck, "A Two-Stage Algorithm for Discontinuity-Preserving Surface Reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, pp. 36-55, Jan. 1992.
[28] M.J. Black and P. Anandan, "A Framework for the Robust Estimation of Optical Flow," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 231-236, Berlin, Germany, May 1993.
[29] M.J. Black and P. Anandan, "Robust Dynamic Motion Estimation Over Time," *Proc. Computer Vision Pattern Recognition*, pp. 296-302, Lahaina, Maui, Hawaii, June 1991.
[30] M.J. Black and A. Rangarajan, "The Outlier Process: Unifying Line Processes and Robust Statistics," *Proc. Computer Vision Pattern Recognition*, pp. 15-22, Seattle, Wash., June 1994.
[31] M.J. Black, "Recursive Non-Linear Estimation of Discontinuous Flow Fields," *Proc. European Conf. Computer Vision*, vol. 800, pp. 138-145, Stockholm, Sweden, May 1994.
[32] M.J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75-104, Jan. 1996.
[33] M. Bober and J. Kittler, "Robust Motion Analysis," *Proc. Computer Vision Pattern Recognition*, pp. 947-952, Seattle, Wash., June 1994.
[34] M. Bober and J. Kittler, "On Combining the Hough Transform and Multiresolution MRFs for the Robust Analysis of Complex Motion," *Proc. Second Asian Conf. Computer Vision*, pp. 301-305, Singapore, Dec. 1995.
[35] B.K.P. Horn and B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, Aug. 1981.
[36] C. Schnörr, "Computation of Discontinuous Optical Flow by Domain Decomposition and Shape Optimization," *Int'l J. Computer Vision*, vol. 8, no. 2, pp. 153-165, Aug. 1992.
[37] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Performance of Optical Flow Techniques," *Int'l J. Computer Vision*, vol. 12, no. 1, pp. 43-77, Feb. 1994.



Dong-Gyu Sim (S'94) received the BS and MS degrees in electronic engineering from Sogang University, Seoul, Korea, in 1993 and 1995, respectively. He is currently working toward his PhD at the same university. His current research interests are computer vision and pattern recognition.



Rae-Hong Park (S'76-M'84) received the BS and MS degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the MS and PhD degrees in electrical engineering from Stanford University, Stanford, California, in 1981 and 1984, respectively.

He joined the faculty of the Department of Electronic Engineering at Sogang University, Seoul, Korea, in 1984, where he is currently a professor. In 1990, he spent his sabbatical year at the Computer vision Laboratory of the Center for Automation Research, University of Maryland, College Park, Maryland, as a visiting associate professor. His current research interests are computer vision, video communication, and pattern recognition.