

PAPER

Fast, Robust Block Motion Estimation Using Simulated Annealing*

Mon-Chau SHIE[†], Wen-Hsien FANG^{††}, Kuo-Jui HUNG^{††}, and Feipei LAI[†], *Nonmembers*

SUMMARY This paper presents a simulated annealing (SA)-based algorithm for fast and robust block motion estimation. To reduce computational complexity, the existing fast search algorithms move iteratively toward the winning point based only on a finite set of checking points in every stage. Despite the efficiency of these algorithms, the search process is easily trapped into local minima, especially for high activity image sequences. To overcome this difficulty, the new algorithm uses two sets of checking points in every search stage and invokes the SA to choose the appropriate one. The employment of the SA provides the search a mechanism of being able to move out of local minima so that the new algorithm is less susceptible to such a dilemma. In addition, two schemes are employed to further enhance the performance of the algorithm. First, a set of initial checking points which exploit high correlations among the motion vectors of the temporally and spatially adjacent blocks are used. Second, an alternating search strategy is addressed to visit more points without increasing computations. Simulation results show that the new algorithm offers superior performance with lower computational complexity compared to previous works in various scenarios.

key words: blocking motion estimation, simulated annealing, video coding

1. Introduction

Motion estimation underlines the core of motion compensated predictive coding of image sequences. The block matching algorithms (BMA), which render efficient implementations, have in particular received a great amount of attention and have been adopted by various video compression standards such as the H.261, MPEG 1-2, etc. [1].

The most straightforward BMA is the full search (FS) algorithm, which searches exhaustively over all allowable displaced points in the reference frame to locate the best match. The enormous amount of computations involved, however, has hindered it from practical implementations. To mitigate this, various attempts have been made to reduce the number of search points while, at the same time, without serious degradation of the reconstructed image quality. One of

the most widespread BMA is the simple yet effective three-step search (TSS) algorithm [2], which iteratively checks the winning point as well as the surrounding eight points with a diminishing window size. Several variants of the TSS such as the new three-step search (NTSS) algorithm [3], the four-step search (FSS) algorithm [4], and the diamond search (DS) algorithm [5] were also addressed, aiming at locating the motion vector more precisely with reduced computations. All of these fast search algorithms, however, are based on the implicit assumption that the block distortion measure (BDM) monotonically increases around the global minimum [2]. The search thus moves iteratively toward the point which achieves the minimum BDM in every stage. Since this underlying assumption does not hold in many practical situations, the search process is therefore easily trapped into local minima if the point chosen in any stage is not the one that leads to the global minimum. Consequently, despite their efficiency, the search process is vulnerable to local minima, especially for high activity image sequences which contain lots of them.

To overcome this difficulty, several robust BMA were advocated to refrain the search from being easily trapped into local minima. For example, Jan et al. [6] determined the motion vector based on the concept of "motion flow," which employs multiple candidates in the search process and thus is more robust against local minima. Tang et al. [7] decomposed the template matching into pixel and subpixel levels and considered a robust and accurate two-stage approach for motion estimation at the price of substantially more computations.

In this paper we propose a new algorithm by utilizing the SA technique together with the local characteristics of the image sequences. As discussed above, we can note that the main problem of the current fast search algorithms is that they are "greedy" in the search process by only pursuing the winning point in every stage. Since only a limited amount of points are used in every stage in order to reduce the computational complexity, these algorithms are inevitably susceptible to the local minima dilemma. In light of this, the proposed algorithm uses two sets of checking points in every search stage and invokes the SA in the decision of the appropriate one. The SA provides the search a mechanism of being able to jump out of local minima so that the new algorithm is more robust against such

Manuscript received April 28, 1999.

Manuscript revised September 2, 1999.

[†]The authors are with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan.

^{††}The authors are with the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.

*This research is supported by National Science Council of R.O.C. under contracts NSC 87-2213-E-011-014 and 88-2218-E-011-007.

undesired situations. In addition, two schemes are employed to further enhance the performance of the algorithm. First, a set of initial checking points which exploit high correlations among the motion vectors of the temporally and spatially adjacent blocks are used. Second, an alternating search strategy is addressed to visit more points without increasing computations. As a consequence of these techniques, the new algorithm can accurately locate the motion vectors with lower computational overhead as compared with previous works.

This paper is organized as follows. Section 2 addresses the proposed SA-based algorithm. Some experimental results are furnished in Sect.3 to justify the new algorithm. Section 4 provides a concluding remark to summarize the whole paper.

2. Simulated Annealing-Based Algorithm for Motion Estimation

In this section, we describe an SA-based algorithm for fast and robust motion estimation. Before addressing the proposed algorithm in detail, we briefly review the SA.

2.1 Simulated Annealing

The SA is a well-known effective technique for solving *iterative optimization* problems and has been successfully applied to various contexts of applications such as computer-aided circuit design and image processing [8]. In such problems, the traditional iterative search process is easily trapped into local minima on the way to find the global minimum. The employment of the SA allows a possibility of escaping from local minima by providing a search a mechanism of uphill move according to a stochastic decision rule. More specifically, the SA is constituted by the following two rules in the transition of “states” [8]:

1. A change of state which induces a reduction of the associated energy is always allowed.
2. If a change of state induces an increase of the associated energy, the change is then accepted with the probability

$$P = e^{\frac{-(E_2 - E_1)}{kT}} \quad (1)$$

where E_1 and E_2 are the associated energy of the present and next states, respectively, T is a control parameter (called “temperature”), and k is the Boltzmann constant.

In light of these two rules, we can note that the scheme always takes a downhill step (reduction of energy) in a transition of states, and sometimes an uphill step (increase of energy) with probability P , which decreases as the energy difference increases. To illustrate this more explicitly, suppose that we intend to iteratively find the solution of x to achieve the smallest associated energy as shown in Fig. 1. Then, if the current

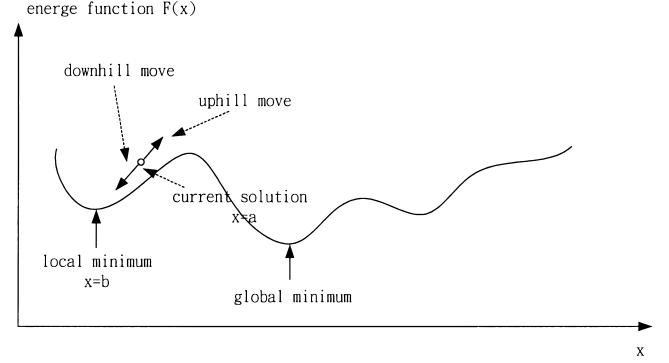


Fig. 1 The iteratively search strategy of the SA.

solution is $x = a$, the traditional algorithms choose the next search point only at the left-hand side of a (downhill move). This one-sided search will eventually arrive at the solution $x = b$, which is only a local minimum. In contrast, the SA, in addition to these points, possesses a possibility of choosing the next search point at the right-hand side of a (uphill move) with a probability decided by (1). It is this uphill climbing technique which enables the search process to be more robust against local minima. Also, the temperature T in general decreases as iterations proceed to reflect the fact that the search is less likely to be trapped into local minima as it gets closer to the global minimum. A simple mechanism to attain this is to reduce T by a fixed ratio in each iteration.

2.2 The SA-Based Approach

Since the fast search block motion estimation algorithm is an iterative two-dimensional search process, to follow we consider a new algorithm which incorporates the aforementioned SA technique.

To apply the SA in this problem, the state will denote the checking point pattern and the energy associated with the state will correspond to the minimum BDM based on these checking points [9]. In every search stage, we consider two checking point patterns, the current one and the next one, and invoke the SA in deciding the appropriate one. More specifically, suppose that E_1 and E_2 stand for the minimum BDM based on the current and the next checking point patterns, respectively. If $E_1 > E_2$, then the search will proceed based on the next checking point pattern. Otherwise, the search will either jump to the next checking point pattern or stay in the current one to further the search with a probability decided by (1). Also, since the search is carried out for every block in the frame, we suggest a normalized initial T as $T = \alpha(E_1 - E_2)$ be used so that the same probability measure (1) can be used for every block.

Furthermore, two strategies are employed to further enhance the performance of the proposed algo-

rithm. In view of the fact that most image sequences only involve gentle movements, there exists high correlations among the motion vectors of the temporally and spatially adjacent blocks. The first scheme is then to appropriately choose the initial checking points by exploiting these intimate relationships. Similar ideas have also been reported in previous literature [10]. Here, we extend their ideas by using not only the motion vector of the temporally adjacent block but also those of the spatially adjacent blocks and their adjacent neighboring points. More specifically, we adopt the starting candidate points which include C_0^t , C_{i-1}^t , C_{j-1}^t , and C_0^{t-1} as shown in Fig. 2, where C_0^t denotes the checking point pattern that includes the center of the search window of the present block and its eight neighboring points, C_{i-1}^t and C_{j-1}^t denote the checking point patterns based on the motion vectors of the two spatially adjacent blocks, respectively, and C_0^{t-1} that of the corresponding block of the previous frame. In light of the center-biased characteristic of the image sequences as addressed above, most of these checking points are in common and therefore do not incur too many computations. In addition, in order to reduce the computational complexity while still visiting enough checking points, the second scheme is to use alternating checking point patterns as shown in Fig. 3 between two adjacent blocks.

The proposed algorithm begins with the determination of the energy of the initial state, E_1 , as the minimum BDM based on the aforementioned starting checking points. Also, we consider the energy of the next state, E_2 , as

$$E_2 = \text{median}\{b_{i-1}^t, b_{j-1}^t, b_0^{t-1}\} \quad (2)$$

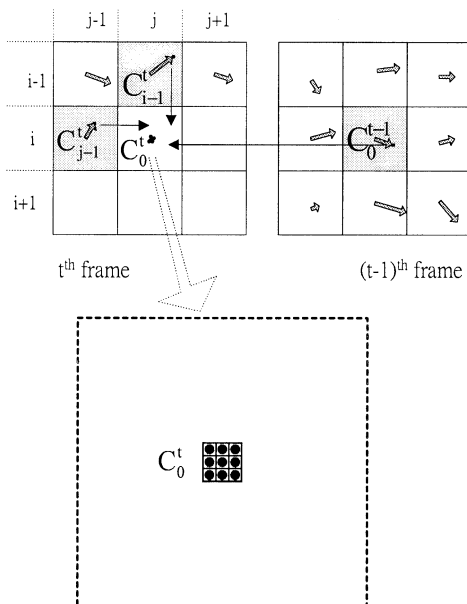


Fig. 2 The starting checking point patterns C_0^t , C_{i-1}^t , C_{j-1}^t , and C_0^{t-1} , where the short arrows denote the corresponding motion vectors.

where $\text{median}\{\dots\}$ stands for the median value of the variables inside the bracket, b_{i-1}^t , b_{j-1}^t , and b_0^{t-1} denote the minimum BDM of the two spatially and one temporally adjacent blocks based on their respective optimum motion vectors, and have already been determined.

If $E_2 < E_1$, we expect that the present block involves fast movement and the next search pattern should be closer to the boundary of the search window. As such, we use a larger search pattern C_d , where d denotes the corresponding search size, and determines the new E_2 based on the minimum BDM of these new checking points. If the new E_2 is still smaller than E_1 , we set this E_2 as the new E_1 and begins a new search with the center at the winning point with a search pattern of shrinking size, $C_{\frac{d}{2}}$, and repeat the above steps. On the other hand, if $E_2 > E_1$, we use the SA to determine whether we should proceed the same steps as the above or we should go directly to the final fine search. The above procedures repeat until the minimum search window is attained. The final fine search pattern, C_{end} , is to check the optimum motion vector determined so far along with its eight neighboring points to more accurately locate the best match.

Based on the discussion above, the overall procedures for the proposed algorithm can be summarized as:

Initializations

E_1 = minimum BDM based on the checking point patterns C_0^t , C_{i-1}^t , C_{j-1}^t , and C_0^{t-1} , E_2 = medium $\{b_{i-1}^t, b_{j-1}^t, b_0^{t-1}\}$, search window size = $(2w+1) \times (2w+1)$, $l=1$, and $T = \alpha(E_1 - E_2)$.

Begin of iterations

1. Do **SA**($E_1, E_2, T, \text{ACCEPT}$). If $\text{ACCEPT} = 0$, go to Step 2; otherwise, $E_1 \leftarrow \min\{E_1, E_2\}$, $d = \frac{w+1}{2^l}$, E_2 = minimum BDM based on the checking point

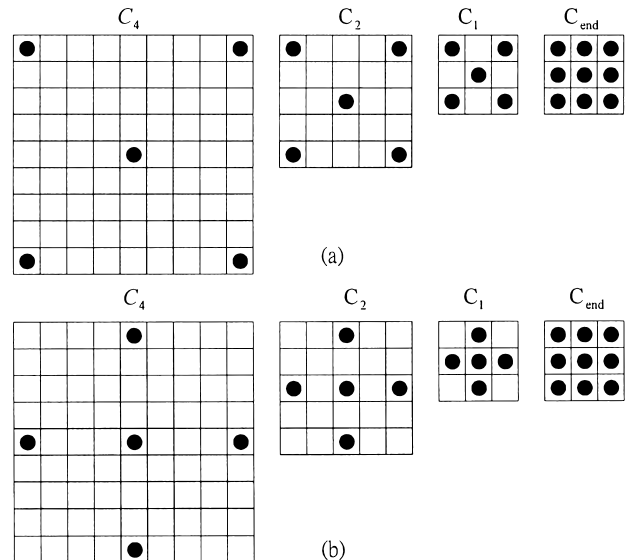


Fig. 3 $\{C_d\}$ and C_{end} for alternating checking point patterns (a) and (b) with $w = 7$.

pattern \mathcal{C}_d with the center being the winning point in the previous step. If $d > 1$, $T \leftarrow \beta T$ (β is a reduction ratio), $l \leftarrow l + 1$, repeat Step 1.

2. Perform the final fine checking point pattern \mathcal{C}_{end} .

End of iterations

where

Procedure **SA**($E_1, E_2, T, \text{ACCEPT}$)

If $E_2 < E_1$, $\text{ACCEPT}=0$; Otherwise, $\text{ACCEPT}=1$ with probability P as given in (1).

The checking point patterns will alternate between patterns (a) and (b), as shown in Fig. 3 with $w = 7$. An example to illustrate the above algorithm is shown in

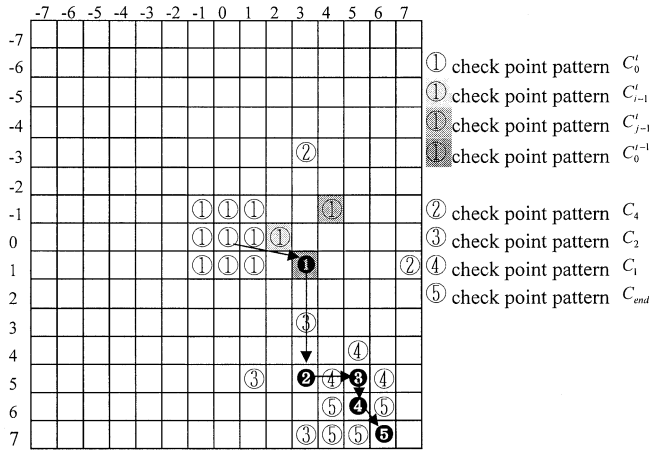


Fig. 4 An example to demonstrate the proposed algorithm.

Fig. 4.

3. Simulation Results and Discussion

Some simulations are conducted in this section to assess the proposed algorithm. Two sets of image sequences are considered. The first set, including the “Claire,” “Miss America,” and “Salesman” image sequences, mainly consists of low activity blocks; whereas the second set, including the “Flower,” “Football,” and “Table tennis” image sequences, consists of lots of high activity blocks. The computationally efficient mean absolute difference (MAD) given by

$$MAD_d(\mathbf{x}) = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{i} \in \mathcal{B}} |f_k(\mathbf{x} + \mathbf{i}) - f_{k-1}(\mathbf{x} + \mathbf{i} + \mathbf{d})| \quad (3)$$

is used as the BDM, where $f_k(\mathbf{i})$ denotes the density of pixel at $\mathbf{i} = (i_1, i_2)$ of the k th frame, \mathcal{B} is an $M \times N$ block, and $|\mathcal{B}|$ denotes the number of points in \mathcal{B} . For comparison, six existing algorithms, including the FS, dependent FS (referred to as the DFS) [10], TSS, NTSS, FSS, and DS algorithms, along with the proposed one (with the parameters $\alpha = 0.7$, $\beta = 0.8$, and $k = 1$), are carried out. Two standard criteria: the average mean squares error per pixel (MSE/pixel) and the average search points per block are utilized to evaluate the effectiveness of these algorithms. The former criterion is to measure the reconstructed image quality, whereas

Table 1 Comparison of MSE/pixel for various algorithms based on the first 90 frames of the test image sequences.

	Clair	Miss America	Salesman	Flower	Football	Table Tennis
FS	9.14	10.11	27.60	277.00	384.88	184.86
DFS	9.31	10.37	29.28	309.21	560.40	257.84
TSS	9.35	10.57	28.29	320.33	416.43	240.07
NTSS	9.31	10.24	27.92	285.01	412.54	217.46
FSS	9.31	10.50	28.16	299.69	428.89	213.28
DS	9.29	10.26	28.13	287.03	433.32	205.94
Proposed	9.26	10.22	27.77	279.80	411.00	196.35

Table 2 Comparison of average number of search points per block for various algorithms based on the first 90 frames of the test image sequences.

	Clair	Miss America	Salesman	Flower	Football	Table Tennis
FS	204.28	204.28	204.28	202.05	202.05	202.05
DFS	24.11	24.58	24.95	25.01	24.07	24.76
TSS	23.28	23.44	23.23	23.25	23.09	23.32
NTSS	20.28	21.78	16.85	21.58	20.56	21.57
FSS	17.59	18.83	16.24	18.90	18.04	19.03
DS	14.99	16.60	12.92	17.02	16.06	16.87
Proposed	14.20	16.53	11.70	15.81	14.53	15.97

Table 3 Comparison of the average number of performing (1) per block based on the first 90 frames of the test image sequences.

	Clair	Miss America	Salesman	Flower	Football	Table Tennis
average number of performing (1) per block	0.99	1.11	1.09	1.12	1.10	1.11

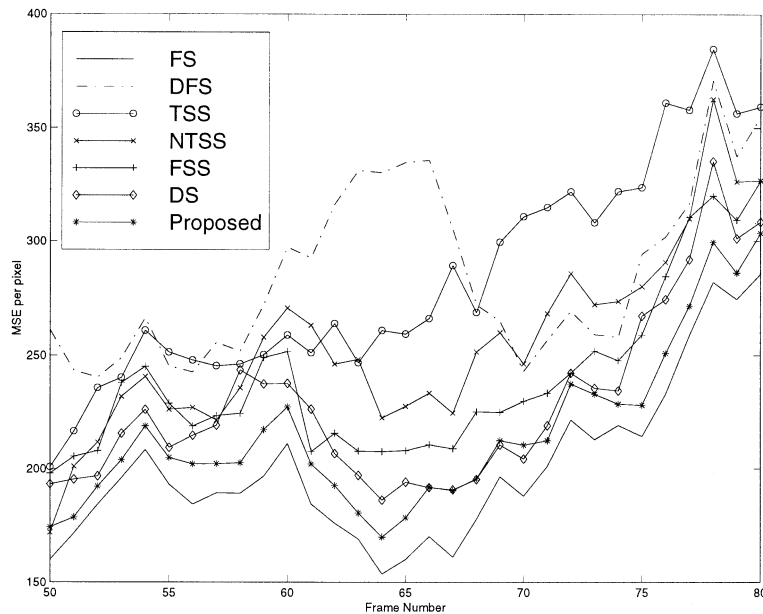


Fig. 5 Comparison of MSE/pixel vs. frame number for various algorithms based on the Table Tennis sequence.

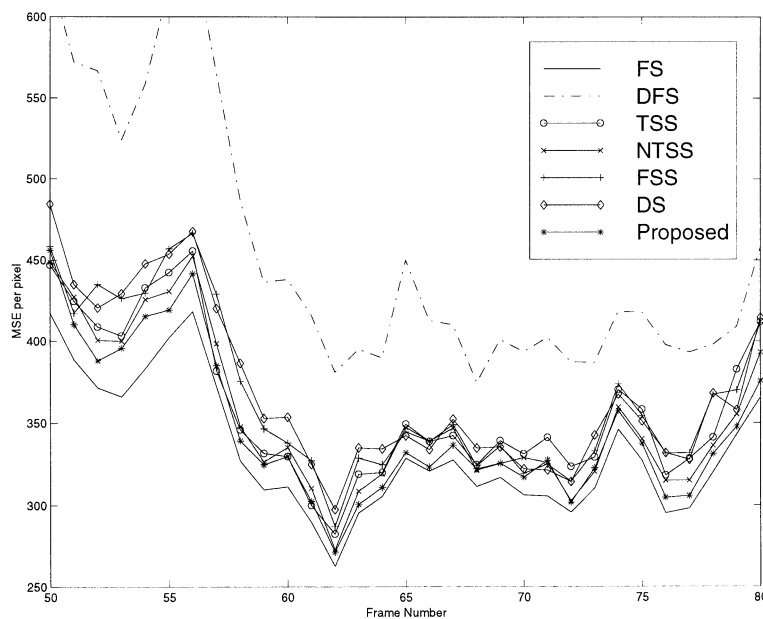


Fig. 6 Comparison of MSE/pixel vs. frame number for various algorithms based on the Football sequence.

the latter the computational complexity.

The resulting average MSE/pixel and the average search points per block for the first 90 frames of the test image sequences using these algorithms based on the H.261 with $M = N = 16$ and $w = 7$ are listed in Tables 1 and 2, respectively. The DFS algorithm employs the temporal information for the initial estimate of the motion vector and thus, as suggested in [10], uses a smaller search window 5×5 to carry out the full search. Table 3 lists the average number of car-

rying out (1) per block for the test image sequences. We can find from Table 3 that all of these values are about one for all of the image sequences and thus the overhead induced is negligible. As a vivid illustration, the MSE/pixel vs. frame number for the test image sequences “Table tennis” and “Football” are also shown in Figs. 5 and 6, respectively (for clarity, only the results from the 50th to the 80th frames are given).

From Table 1, we can note that the proposed algorithm outperforms the others by providing a smaller

MSE/pixel (except the FS), whereas we can note from Table 2 that the new algorithm also calls for the lowest computational complexity. The DFS algorithm addressed in [10] can indeed reduce the computations required by the original FS algorithm. It, however, can not provide quite satisfactory performance for high activity image sequences such as the "Football" sequence because the motion vectors of the adjacent frames are not that highly correlated in these scenarios. The small MSE's of the proposed algorithm are due to the SA technique as well as the full exploitation of the temporal and spatial relationships among the motion fields. The incorporation of the SA scheme also explains that the proposed algorithm works in particular well for high activity image sequences, in which the previous fast algorithms are easily trapped into local minima in the search process.

4. Conclusions

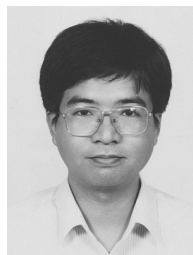
In this paper, we describe a new fast block-based motion estimation algorithm which is also robust against local minima. This new algorithm uses high correlations among the motion fields of the spatially and temporally adjacent blocks, the SA in the transition of search point patterns, and the alternating search technique for adjacent blocks, to attain superior performance with lower computations compared to previous works. As such, it offers an appealing alternative to block motion estimation in view of the performance it can provide and the computational complexity it calls for.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments which have enhanced the readability and quality of this paper.

References

- [1] A.M. Tekalp, Digital Video Processing, Prentice-Hall, Englewood Cliffs, NJ, 1995.
- [2] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," Proc. NTC 81, pp.C9.6.1-9.6.5, New Orleans, LA, Nov./Dec. 1981.
- [3] R. Li, B. Zeng, and M.L. Liou, "A new three-step search algorithm for block motion estimation," IEEE Trans. Circuits Syst. for Video Technol., vol.4, pp.438-442, Aug. 1994.
- [4] L.-M. Po and W.-C. Ma, "A novel four-step search algorithm for fast block motion estimation," IEEE Trans. Circuits Syst. for Video Technol., vol.6, pp.313-317, June 1996.
- [5] J.-Y. Tham, S. Ranganath, M. Ranganath, and A. Kassim, "A novel unrestricted center-biased diamond search algorithm for block motion estimation," IEEE Trans. Circuits Syst. for Video Technol., vol.8, pp.369-377, Aug. 1998.
- [6] J.-S. Jan, W.-H. Fang, and M.-Y. Yu, "An adaptive flow-based dynamic search algorithm for block motion estimation," Proc. IEEE Int'l Symp. Circuits and Sys., pp.2092-2095, Hong Kong, 1997.
- [7] C.-Y. Tang, Y.-P. Hung, and Z. Chen, "Robust two-stage approach for image motion estimation," Electronics Letters, vol.34, pp.1091-1093, May 1998.
- [8] P.J.M. Van Laarhoven and E.H.L. Aarts, Simulated Annealing: Theory and Applications, Kluwer Academic Publishers, Dordrecht, Holland, 1987.
- [9] M.-C. Shie, W.-H. Fang, K.-J. Hung, and F. Lai, "Fast block motion estimation using adaptive simulated annealing," Proc. IEEE Asia-Pacific Symp. Circuits and Sys., pp.607-610, Thailand, 1998.
- [10] A. Puri, H.-M. Hang, and D.L. Schilling, "Motion-compensated transform coding based on block motion-tracking algorithm," Proc. Int. Conf. Commun., pp.136-140, 1987.



Mon-Chau Shie is a Ph.D. candidate in the Department of Electrical Engineering of the National Taiwan University. He is also a lecturer in the Department of Electronic Engineering of National Taiwan University of Science and Technology. His research interests include computer video and image processing, operating system and microprocessor system design. He received an M.S.E.E. from the National Taiwan University and a B.S.E.E.

from the National Taiwan University.

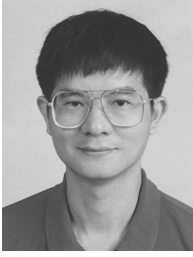


Wen-Hsien Fang was born in Taipei, Taiwan, Republic of China, in 1961. He received his B.S. degree from the National Taiwan University in 1983, and his M.S.E. degree and Ph.D. degree from the University of Michigan, Ann Arbor, in 1988 and 1991, respectively, in electrical engineering and computer science. From 1988 to 1991, he was a Research Assistant at the University of Michigan. In fall 1991, he joined the faculty of National Taiwan

University of Science and Technology, where he currently holds a position as an Associate Professor in the Department of Electronic Engineering. His research interests include fast algorithms for signal processing and their VLSI hardware implementations, statistical signal processing, wireless communication, and video coding.



Kuo-Jui Hung was born in Taiwan, Republic of China, in 1971. He received the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University of Science and Technology, respectively. He is now with the ASUS Comp., designing the PC motherboards and notebooks.



Feipei Lai received a B.S.E.E. degree from National Taiwan University in 1980, and M.S. and Ph.D. degrees in computer science from the University of Illinois at Urbana-Champaign in 1984 and 1987, respectively. He is a professor in the Department of Electrical Engineering and in the Department of Computer Science and Information Engineering at National Taiwan University. He was a visiting professor in the Department of Computer Sci-

ence and Engineering at the University of Minnesota, Minneapolis, U.S.A. He was also a guest Professor at University of Dortmund, German and a visiting senior computer system engineer in the Center for Supercomputing Research and Development at the University of Illinois at Urbana-Champaign. Dr. Lai holds four Taiwan patents and two USA patents currently. He served as a consultant at ERSO, ITRI during 1988 and at Faraday Technology Corp. from 8/94 to 7/95. His current research interests are high performance microprocessor chip design, computer architecture, optimizing compiler, VLSI design. Prof. Lai is one of the founders of the Institute of Information & Computing Machinery. He is also a member of Phi Kappa Phi, Phi Tau Phi, ACM, The Chinese Institute of Engineers, and The Institute of Electronics, Information and Communication Engineering. He received Acer awards five times in 1989, 1991, 1992, 1993 and 1995 and The Taiwan Fuji Xerox Research award in 1991. Dr. Lai is a Senior member of IEEE and included in "Who's Who in Science and Engineering" and "Who's Who in the World."