A New Predictive Search Area Approach for Fast Block Motion Estimation

Kuo-Liang Chung, *Senior Member, IEEE*, and Lung-Chun Chang

***Abstract*—According to the observation on the distribution of motion differentials among the motion vector of any block and those of its four neighboring blocks from six real video sequences, this paper presents a new predictive search area approach for fast block motion estimation. Employing our proposed simple predictive search area approach into the full search (FS) algorithm, our improved FS algorithm leads to 93.83% average execution-time improvement ratio, but only has a small estimation accuracy degradation. We also investigate the advantages of computation and estimation accuracy of our improved FS algorithm when compared to the edge-based search algorithm currently published by Chan and Siu; experimental results reveal that our improved FS algorithm has 74.33% average execution-time improvement ratio and has a higher estimation accuracy. Finally, we further compare the performance among our improved FS algorithm, the three-step search algorithm, and the block-based gradient descent search algorithm.**

***Index Terms*—Block matching, estimation accuracy, full search algorithm, motion estimation, search area determination, search window.**

I. Introduction

MOTION estimation plays an important role in video coding [13], [11], [9]. The block matching algorithm is widely used in motion estimation. In the block matching algorithm, the encoder first divides the current image frame into many fixed-size blocks, and the motion vector for each current block, say B_c, with size $b \times b$, where $b = 2^m$, is determined by finding the best matching reference block in the reference image frame according to the defined matching criterion. For example, b is selected as 16. Commonly, for the current block to find the best matching block, a search window of size $(2W+1) \times (2W+1)$ is used to confine the search range in the reference image frame. For example, W is selected as 16. Throughout this paper, the matching criterion used is the accumulated absolute difference (AAD) which is defined by

$$\mathrm{AAD}(v_x, v_y) = \sum_{x=1}^{b} \sum_{y=1}^{b} |B_c(x, y) - B_r(x+v_x, y+v_y)| \quad (1)$$

for $-W \leq v_x, v_y \leq W$ where $B_c(x, y)$ denotes the gray level of the pixel at position (x, y) of the current block in the current image frame and $B_r(x+v_x, y+v_y)$ denotes the gray level of the pixel at position $(x+v_x, y+v_y)$ of the reference block in the reference image frame. By (1), the best matching reference block within the search window is the found block with the minimal AAD (v_x, v_y).

The full search (FS) algorithm is the most simple block matching algorithm for finding the best matching block among all the possible search positions in the search window. Although the FS algorithm can indeed obtain the global optimal result, however it is much time-consuming. Therefore, many efficient block matching algorithms have been developed, for example, the one-dimensional search algorithm [2], the two-step search algorithm [3], the cross-search algorithm [4], the two-dimensional logarithmic search algorithm [6], the three-step search (TSS) algorithm [7], the block-based gradient descent search (BBGDS) algorithm [8], the four-step search algorithm [10], the conjugate directional search algorithm [12], and the edge-assisted search (EAS) algorithm [1]. Among these developed block matching algorithms, many different kinds of techniques have been proposed to reduce the computational time as much as possible, and make the estimation accuracy degradation as least as possible.

In this paper, according to the observation on the statistical distribution of motion differentials among the motion vector of any block and those of its four neighboring blocks from six real video sequences, we present a new predictive search area approach for fast block motion estimation. Employing our proposed predictive search area approach into the full search (FS) algorithm, it leads to 93.83% average execution-time improvement ratio, but only has a small estimation accuracy degradation. We next investigate the advantages of computation and estimation accuracy of our proposed improved FS algorithm when compared to the EAS algorithm currently published by Chan and Siu [1]; experimental results reveal that our improved FS algorithm has 74.33% average execution-time improvement ratio and has a higher estimation accuracy. Furthermore, we compare the performance among our improved FS algorithm, the TSS algorithm, and the BBGDS algorithm.

The remainder of this paper is organized as follows. Our proposed predictive search area approach is presented in Section II. Section III presents our improved FS algorithm. Experimental results are demonstrated in Section IV. Some concluding remarks are addressed in Section V.

II. New Predictive Search Area Approach

In this section, we want to present the predictive search area approach which will be used in our proposed improved FS algorithm. We first give an observation on the motion differentials, among the motion vector of any block and those of its four

Manuscript received March 6, 2002; revised February 6, 2003. This work was supported by the National Science Council, R.O.C., under contracts NSC90-2213-E-011-056 and NSC91-2213-E-011-028. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bruno Carpentieri.

K.-L. Chung is with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10672, Taiwan, R.O.C. (e-mail: klchung@cs.ntust.edu.tw).

L.-C. Chang is with the Department of Information Management, National Taiwan University of Science and Technology, Taipei 10672, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TIP.2003.812756

neighboring blocks. Here, all the motion vectors used to evaluate the statistical distribution of motion differentials are obtained by using the FS algorithm in advance from six real video sequences, namely, the calendar train with 30 image frames, the football with 30 image frames, the garden with 30 image frames, the salesman with 21 image frames, the Susie with 30 image frames, and the table tennis with 30 image frames. In fact, these six video sequences can be viewed as the training videos. According to our observation on the statistical motion differentials, a predictive search area is thus determined using our proposed search area determination strategy where the determined search area is within the search window. Since the determined search area is much smaller than the predefined search window of size $(2W+1)\times(2W+1)$, the computation cost required in our improved FS algorithm can be drastically reduced when compared to the FS algorithm.

A. Observation

Since in video sequence, there are about thirty image frames being taken in a second, two consecutive image frames are possibly very similar. We now investigate the distribution of motion-vector differentials between any block, say B, and its four neighboring blocks, say B_1, B_2, B_3, and B_4, as shown in Fig. 1.

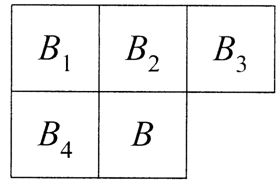


Fig. 1. Any block and its four neighboring blocks.

Let the displacement variable D be defined by

$$D=\min_{1\le i\le 4}\{\max\{|v_x-mv_{i,x}|,\,|v_y-mv_{i,y}|\}\} \tag{2}$$

where (v_x, v_y) denotes the motion vector of the block $B(x, y)$ and mv_i $(=(mv_{i,x}, mv_{i,y}))$, $1\le i\le 4$, denotes the motion vector of the neighboring block B_i. For example, let the motion vector of the block $B(x, y)$ be (2,3) and let the motion vectors of its four neighboring blocks be (3,3), (4,3), (−2,3), and (3,4), respectively, then by (2), it yields to $D=\min\{\max\{1, 0\}, \max\{2, 0\}, \max\{4, 0\}, \max\{1, 1\}\}=1$. From this example, the obtained displacement $D=1$ means that the displacement of motion-vector differentials between the block $B(x, y)$ and the nearest neighboring block is one pixel.

For the jth image frame in the video sequence l, $1\le l\le 6$, let $\mathrm{Pr}_{j,l}(D=d)$ denote the probability when $D=d$. The size of the search window used in our experiment is 33 × 33, so we have the constraint $0\le d\le 16$. Except the first image frame in the lth video sequence, for the remaining image frames in the same video sequence, the average probability of $D=d$ is defined by

$$Prob_l(D=d)=\frac{1}{n_l-1}\sum_{j=2}^{n_l}\mathrm{Pr}_{j,l}(D=d) \tag{3}$$

where n_l denotes the number of image frames in the video sequence l. The first column of Table I denotes the values of d's. By (3), the ith column, $2\le i\le 7$, of Table I lists the average probability of D for the $(i-1)$th video sequence. The eighth column denotes the average probability of D for the six video sequences where the related average probability is defined by

$$Prob_{nl}(D=d)=\frac{1}{6}\sum_{l=0}^{5}\frac{1}{n_l-1}\sum_{j=2}^{n_l}\mathrm{Pr}_{j,l}(D=d).$$

For example, when $D=1$, we have $Prob_{nl}(D=1)=0.7838(=(0.8919+0.5884+0.7962+0.7972+0.8621+0.7669)/6)$.

The ninth column of Table I denotes the accumulated probability of $D=d$ for the six video sequences where the related accumulated probability is defined by

$$Prob_{nl}(D\le d)=\frac{1}{6}\sum_{l=0}^{5}\frac{1}{n_l-1}\sum_{j=2}^{n_l}\sum_{D=0}^{d}\mathrm{Pr}_{j,l}(D=d).$$

For example, when $D=2$, we have $Prob_{nl}(D\le 2)=94.24\%$ $(=0.7838+0.1348+0.0238)$; when $D=3$, we have $Prob_{nl}(D\le 3)=95.54\%$ $(=0.7838+0.1348+0.0238+0.0130)$.

From the ninth column of Table I, we have the following observation.

Observation 1: The average accumulated probability is 94.24% for $D=2$ (95.54% for $D=3$).

In fact, the accumulated probability of $D=d$ can be calculated in an adaptive way. That is, once we have obtained the motion vector of the current block $B_c(x, y)$ by using (1), the motion vector of $B_c(x, y)$ associated with the four motion vectors of the four neighboring blocks (see B_1, B_2, B_3, and B_4 in Fig. 1) can be used to obtain the new $Prob_{nl}(D\le d)$ although some extra time is needed.

B. Search Area Determination

From Observation 1, due to the high probability, e.g., 94.24%, $D=2$ is a good choice to confine the search area for the current block B_c to find the best matching block in the reference image frame. Of course, $D=3$ is also another choice to confine the search area for the current block B_c. In Section IV, both experiments for $D=2$ and $D=3$ will be carried out to demonstrate the good performance of our proposed improved FS algorithm for block motion estimation. For exposition, we only consider the case $D=2$ although our proposed predictive search area approach is applicable for $D=3$.

For the current block B_c in the current image frame, according to Observation 1 and the four known motion vectors of B_cs four neighboring blocks, namely mv_1, mv_2, mv_3, and mv_4, the determined search area of the current block B_c, which is used to confine the search area in the reference image frame, can be represented by

$$S=S_1\cup S_2\cup S_3\cup S_4$$

where S_i, $1\le i\le 4$, denotes the subsearch area with size $(2D+1)\times(2D+1)=5\times 5$ for mv_i [see Fig. 2(a)] where from Observation 1, we have $D=2$. The union of the four 5 × 5 subsearch areas is depicted in Fig. 2(b) which is used as the search area in the reference image frame for the current block B_c in the current image frame.

From Observation 1 and the determined search area S, for the current block B_c in the current image frame, it has about 94.24% probability that the best matching reference block in

TABLE I
$Prob_l(D = d)$, $Prob_{nl}(D = d)$

D	calendar	football	garden	salesman	Susie	tennis	$Prob_{nl}(D = d)$	$Prob_{nl}(D \leq d)$
0	0.8919	0.5884	0.7962	0.7972	0.8621	0.7669	0.7838	0.7838
1	0.0648	0.1839	0.1361	0.1803	0.1289	0.1152	0.1348	0.9186
2	0.0162	0.0660	0.0272	0.0084	0.0051	0.0197	0.0238	0.9424
3	0.0065	0.0358	0.0015	0.0044	0.0005	0.0163	0.0130	0.9554
4	0.0042	0.0257	0.0077	0.0030	0.0001	0.0100	0.0084	0.9638
5	0.0021	0.0170	0.0028	0.0010	0.0000	0.0083	0.0052	0.9690
6	0.0021	0.0147	0.0017	0.0007	0.0001	0.0088	0.0047	0.9737
7	0.0018	0.0113	0.0013	0.0003	0.0001	0.0082	0.0038	0.9775
8	0.0009	0.0092	0.0018	0.0001	0.0000	0.0039	0.0027	0.9802
9	0.0009	0.0054	0.0013	0.0001	0.0000	0.0036	0.0019	0.9821
10	0.0008	0.0060	0.0005	0.0003	0.0000	0.0041	0.0019	0.9840
11	0.0008	0.0071	0.0009	0.0001	0.0001	0.0036	0.0021	0.9861
12	0.0006	0.0055	0.0014	0.0001	0.0001	0.0048	0.0021	0.9882
13	0.0007	0.0035	0.0006	0.0003	0.0000	0.0058	0.0018	0.9900
14	0.0007	0.0039	0.0003	0.0003	0.0000	0.0046	0.0016	0.9916
15	0.0012	0.0051	0.0015	0.0003	0.0001	0.0048	0.0022	0.9938
16	0.0002	0.0055	0.0011	0.0003	0.0000	0.0053	0.0021	0.9959

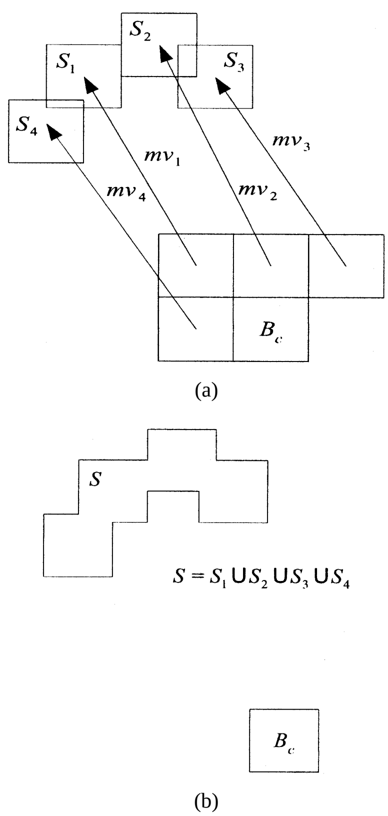


Fig. 2. Four subsearch areas and the determined search area. (a) The four 5 × 5 subsearch areas, S_1, S_2, S_3, and S_4. (b) The determined search area S.

the reference image frame is within the determined search area S since two consecutive image frames are quite similar. From Table I, it is known that the larger the value of D is, the higher the probability is when the best matching reference block is inside the determined search area S. On the other hand, the higher the value of $Prob_{nl}(D \leq d)$ is, the higher the estimation accuracy and the computational cost are.

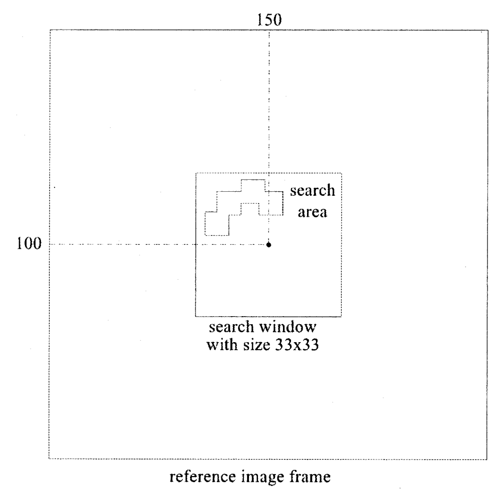


Fig. 3. Search window and the determined search area in the reference image frame.

III. Proposed Improved FS Algorithm

Based on the above predictive search area approach, our improved FS algorithm is outlined in this paragraph. Given a video sequence as the input, for the $(j-1)$th reference image frame and the jth current image frame, $j \geq 2$, our improved FS algorithm for motion estimation is the same as the conventional FS algorithm, but instead of finding the best matching reference block within the search window for the current block B_c, the current block B_c in the current image frame finds the best matching reference block in the reference image frame within the determined predictive search area described in Section II.

In Fig. 2, suppose the upper-left corner of the current block B_c in the current image frame is located at position (150 100). As shown in Fig. 3, taking (150 100) as the center of the search window in the reference image frame, the determined search area in the reference image frame for the current block B_c, i.e.,

TABLE II
PSNR COMPARISON

PSNR	FS	EAS	OURS(D=2)	OURS(D=3)	TSS	BBGDS
Calendar	31.6685	31.6400	31.6400	31.6478	31.3085	31.6224
Football	26.0563	25.5200	25.6779	25.7536	25.4400	25.2757
Garden	27.0244	26.3300	26.9123	26.9209	26.0011	26.8254
Salesman	34.4373	34.3200	34.4197	34.4308	34.1611	34.3770
Susie	38.5160	38.5000	38.5155	38.5156	37.9631	38.5022
Tennis	28.3772	27.5200	27.9365	28.0544	26.7966	27.1818
average PSNR	31.0133	30.6383	30.8503	30.8872	30.2784	30.6308

TABLE III
EXECUTION-TIME COMPARISON

Time	FS	EAS	OURS(D=2)	OURS(D=3)	TSS	BBGDS
Calendar	3.1660	1.1283	0.1615	0.2750	0.0912	0.0655
Football	3.1624	1.3128	0.2635	0.4226	0.0923	0.0957
Garden	3.1532	2.0025	0.1894	0.3067	0.0924	0.0878
Salesman	3.1660	0.5111	0.1956	0.2857	0.0906	0.0564
Susie	3.1853	0.4378	0.1704	0.2765	0.0929	0.0611
Tennis	3.3500	0.6176	0.2034	0.3064	0.0923	0.0715
average time	3.1971	1.0077	0.1973	0.3122	0.0920	0.073

S, is inside the search window. Instead of performing the block matching [(1)] within the search window, based on our proposed improved FS algorithm, we only perform the block matching within the determined search area S. Thus for finding the best matching reference block, the computation improvement ratio (CIR) of our proposed improved FS algorithm within the determined search area over the conventional FS algorithm within the search window is given by

$$CIR = \frac{Area(SW) - Area(S)}{Area(SW)}$$

where $Area(SW)$ and $Area(S)$ denote the area of the search window (SW) and the area of the determined search area S, respectively.

Because of the area of S_i, $1 \le i \le 4$, is equal to 25 $(=(2\times 2+1)\times(2\times 2+1))$ for $D=2$, the area of the determined search area S is at most 100 $(=4\times 25)$. Suppose the search window is of size 33×33 and $D=2$, then we have the following result.

Proposition 1: For performing block matching of the current block in the current image frame, the computation improvement ratio (CIR) of our proposed improved FS algorithm using the predictive search area approach over the FS algorithm using the conventional search window approach is at least 90.82% $(=(1089-100)/1089)$, where $1089 = 33 \times 33$.

In general, suppose the search window is of size $(2W+1)\times(2W+1)$ and the displacement is an arbitrary D, we immediately have the following result.

Corollary 1: For block matching of the current block in the current image frame, the CIR of our proposed improved FS algorithm over the conventional FS algorithm is at least $(4W^2 + 4W - 16D^2 - 16D - 3)/(4W^2 + 4W + 1)$.

From Corollary 1, suppose the search window is still of size 33×33 but $D = 3$, the CIR is at least 82%. Previously, we have known that the larger the value of D is, the higher the probability is when the best matching reference block is inside the determined search area S. That is, the higher the value of $Prob_{nl}(D \le d)$ is, the higher the estimation accuracy and the computational cost are. However, it is a tradeoff among the value of D, the computational cost, and the estimation accuracy.

IV. EXPERIMENTAL RESULTS

In this section, some experiments are carried out to compare the performance among the FS algorithm, the EAS algorithm [1], our proposed improved FS algorithm (OURS) for $D = 2$ and $D = 3$, the TSS algorithm [7], and the BBGDS algorithm [8]. All the concerning block matching algorithms are implemented using Borland C++ builder and the Pentium III 600-based PC.

For convenience, each of the six training video sequences mentioned above is used in each block matching algorithm for evaluating the performance such as the PSNR (Peak Signal-to-Noise Ratio), the execution-time requirement in terms of seconds, and the execution-time improvement ratio. Each image frame in the video sequence is of size 352×256. The block size and the size of the search window are selected as 16×16 and 33×33, respectively. Here, the PSNR is defined by

$$\text{PSNR} = 10\log_{10}\left(\frac{255^2}{\text{MSE}}\right)$$

where MSE is the mean square error between the estimated image frame and the original image frame. In fact, the PSNR is also a guideline for the estimation accuracy.

Tables II–IV show the performance comparison in terms of the PSNR, the execution-time requirement, and the execution-time improvement ratio. From Table II, the FS algorithm has the highest PSNR, i.e., the highest estimation accuracy. Our proposed improved FS algorithm for $d = 2$ has the second PSNR; our proposed improved FS algorithm for $d = 3$ has

TABLE IV
EXECUTION-TIME IMPROVEMENT RATIO COMPARISON

FS	EAS
93.8259%	74.3344 %

the third PSNR, and the EAS has the fourth PSNR. In the EAS block matching algorithm, before performing the edge detection, a smoothing operation is applied on the current image frame in order to filter out noises. Then, Sobel operator [5] is used in edge detection and the value 40 is selected as the threshold to separate whether the testing pixel is an edge pixel or a nonedge pixel; the value 16 is selected as the threshold to separate whether the current block is an edge block or a nonedge block. In addition, in the EAS algorithm, $\alpha = 0.5$ and $\beta = 2.0$ are used to help the determination of the staring points.

From Table III, the BBGDS algorithm and the TSS algorithm are the fastest and the second one, respectively; our proposed improved FS algorithm for $D = 2$ ($D = 3$) is the third (fourth) one. From Table III, the average execution-time required in the FS algorithm is 3.1971 seconds; the average execution-time required in the EAS algorithm is 1.0077 seconds, and the average execution-time required in our proposed improved FS algorithm is 0.1973 seconds. Table IV demonstrates that the average execution-time improvement ratio of our proposed improved FS algorithm over the FS algorithm is 93.83% which nearly confirms the theoretical analysis in Proposition 1. The average execution-time improvement ratio of our proposed improved FS algorithm over the EAS algorithm is 74.33%.

V. CONCLUSION

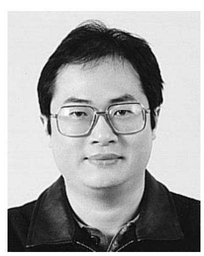
This paper has presented the new predictive search area determination strategy which has also been plugged into the FS algorithm. Not only our proposed simple predictive search area approach can reduce the execution-time required in the motion estimation, but it also can retain a good estimation accuracy. As can be seen from the experimental results, our proposed improved FS algorithm has least execution-time requirement when compared to that of the FS algorithm and the EAS algorithm. Considering the PSNR measure, our proposed improved FS algorithm has the best estimation accuracy in each video sequence when compared to that of the TSS algorithm, the BBGDS algorithm, and the EAS algorithm.

ACKNOWLEDGMENT

The authors appreciate the three anonymous referees, the Associate Editor Prof. B. Carpentieri, and S. W. Yu for their valuable comments that lead to the improved version of this paper.

REFERENCES

[1] Y. L. Chan and W. C. Siu, "An efficient search strategy for block motion estimation using image features," *IEEE Trans. Image Processing*, vol. 10, pp. 1223–1238, Aug. 2001.
[2] M. J. Chen, L. G. Chen, and T. D. Chiueh, "One-dimensional full search motion estimation algorithm for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 504–509, Oct. 1994.
[3] F. H. Cheng and S. N. Sun, "New fast and efficient two-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 977–983, Oct. 1999.
[4] M. Ghanbari, "The cross-search algorithm for motion estimation," *IEEE Trans. Commun.*, vol. 38, pp. 950–953, July 1990.
[5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.
[6] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. 29, pp. 1799–1808, Dec. 1981.
[7] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," in *Proc. Nat. Telecommunications Conf.*, Nov./Dec. 1981, pp. G.5.3.1–G.5.3.5.
[8] L. K. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 419–422, Aug. 1996.
[9] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG Video Compression Standard*. New York: Chapman & Hall, 1997.
[10] L. M. Po and W. C. Ma, "A novel four-step algorithm for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 313–317, June 1996.
[11] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[12] R. Srinivasan and K. R. Rao, "Predictive coding based on efficient motion estimation," *IEEE Trans. Commun.*, vol. 33, pp. 888–896, Aug. 1985.
[13] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.



Kuo-Liang Chung (SM'01) received the B.S., M.S., and Ph.D. degrees in computer science and information engineering from National Taiwan University in 1982, 1984, and 1990, respectively.

From 1984 to 1986, he was a soldier. From 1986 to 1987, he was a Research Assistant with the Institute of Information Science, Academic Sinica, Taiwan. He is now a Professor with the Department of Computer Science and Information Engineering at National Taiwan University of Science and Technology. His research interests include image compression, image processing, video compression, coding theory, and algorithms.

Dr. Chung received the Distinguished Professor Award from the Chinese Institute of Engineers in May 2001.

Lung-Chun Chang received the B.S. degree in mathematics from Tunghai University, Taichung, Taiwan, R.O.C., in 1995, the M.S. degree from the Graduate School of Engineering Technology, National Taiwan Institute of Technology, Taipei, Taiwan, in 1997, and the Ph.D. degree in the Department of Information Management of the National Taiwan University of Science and Technology, Taipei, in 2002.

His research interests include image processing and image security.