

Motion Area Estimated Motion Estimation with Triplet Search Patterns for H.264/AVC

T. Song, and T. Shimamoto

Abstract—In this paper a fast motion estimation method for H.264/AVC named Triplet Search Motion Estimation (TS-ME) is proposed. Similar to some of the traditional fast motion estimation methods and their improved proposals which restrict the search points only to some selected candidates to decrease the computation complexity, proposed algorithm separate the motion search process to several steps but with some new features. First, proposed algorithm try to search the real motion area using proposed triplet patterns instead of some selected search points to avoid dropping into the local minimum. Then, in the localized motion area a novel 3-step motion search algorithm is performed. Proposed search patterns are categorized into three rings on the basis of the distance from the search center. These three rings are adaptively selected by referencing the surrounding motion vectors to early terminate the motion search process. On the other hand, computation reduction for sub pixel motion search is also discussed considering the appearance probability of the sub pixel motion vector. From the simulation results, motion estimation speed improved by a factor of up to 38 when using proposed algorithm than that of the reference software of H.264/AVC with ignorable picture quality loss.

Keywords—Motion estimation, VLSI, image processing, search patterns

I. INTRODUCTION

WITH the rapid growth and improving performance of wireless networks, wireless mobile devices with real time video transmission capability become widely used. H.264/AVC, due to its capability of achieving 2 times higher compression ratio over the current MPEG-4 coding standard, is positioned as the video coding standard of next generation multimedia devices[1]. However, among the several new features which are introduced by H.264/AVC the Motion Estimation (ME) process is highly computationally intensive than traditional algorithms and count for about 80% of the total computation complexity. For the utility of wireless mobile devices, the development of efficient algorithms for the ME of

H.264/AVC is one of the most challenging themes.

ME process exploits temporal correlation between adjacent frames in a video sequence to reduce the data inter-frame redundancy. In the ME, the current frame of video sequence is divided to MacroBlocks (MB) and for each of them a best matched MB in the previous processed frames is searched within the search area. This best matched MB is selected as the MB with lowest coding cost. Then, the differential between the current MB and the best matched MB is coded to achieve high coding efficiency.

Different from the traditional motion estimation algorithms H.264/AVC support seven kinds different block size motion estimation which are from 4x4 pixels to 16x16 pixels block. It also extends the concept of integer pixel motion estimation to half and quarter pixel motion estimation. Furthermore, multiple reference frames are available to search more concise motion vectors. Another Rate Distortion Optimization (RDO) algorithm is adopted by H.264/AVC. It is an exhausted pre-coding loop for each mode of INTRA and INTER type sub-blocks to select the most efficient coding mode. These new techniques highly improved the coding efficiency as well as multiple times coding complexity.

The Full Search (FS) algorithm is the most performing one method for ME since it carries out an exhaustive search of the best match to achieve high image coding efficiency at the expense of high computation cost. Many traditional proposals concerning the mode selection algorithms [2]-[4] or search range management [5]. There are also numbers of fast ME algorithms, such as TSS[6], DS[7], RP[8] were proposed from the view point of search patterns. These traditional works try to search the optimal motion vector from some selected candidates which are approximated in terms of the Sum of Absolute Differential (SAD) and restrict next searching space using this estimated searching point. With these traditional methods, motion search accuracy heavily depends on the SAD of one certain search point. Moreover, to decrease the computation complexity several realistic implementations of H.264/AVC set decreasing search range in the previous reference frames in order of time sequence which induce a high probability to fall into a local minimum.

In this paper a new concept of the motion area is constructed to improve the search accuracy for the first step. In section 2 the in motion area is defined with the discussion of the search patterns. After the discussion of the search patterns the details of proposed algorithm are outlined in section 3. In section 4 the simulation results are shown, and in section 5 the conclusion remarks are addressed.

Manuscript received November 30, 2006.

T. Song is with Computer Systems Engineering, Department of Institute of Technology and Science, Graduate School of Engineering, Tokushima University, Minami-Jyosanijima 2-1, Tokushima City, 770-8506, Japan.
phone: 0081-88-6567484; fax: 0081-88-6567471
e-mail: tiansong@ee.tokushima-u.ac.jp

T. Shimamoto is with Computer Systems Engineering, Department of Institute of Technology and Science, Graduate School of Engineering, Tokushima University, Minami-Jyosanijima 2-1, Tokushima City, 770-8506, Japan.
phone: 0081-88-6567484; fax: 0081-88-6567471
e-mail: simamoto@ee.tokushima-u.ac.jp

II. SEARCH PATTERNS AND MOTION AREA

In this section the concept of motion area and the motion area selection method by proposed patterns are discussed. The motion area is discussed from the viewpoint of the shape of the area, how to describe the shapes with patterns, the pattern distribution and the overlap of these patterns.

A. Shape of the Patterns

Because the rectangle or triangle pattern could be constructed with only 4 and 3 search points totally, small number of search point candidates could construct these patterns for motion search. It is reasonably considered that the rectangle and triangle should be typical basic shapes to describe a motion area. These patterns are depicted in Fig.1.

As shown is Fig.1 well-balanced 25 candidates of the search points in the search area are used to construct these basic shapes. Evaluating the sum of the SAD of all vertices in one pattern the

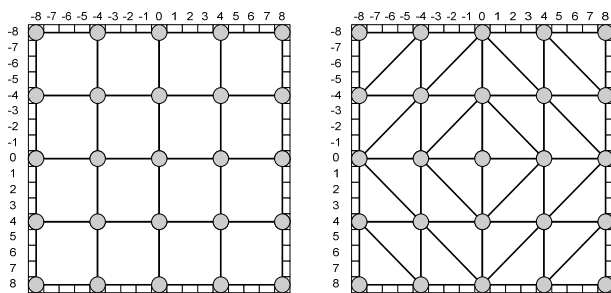
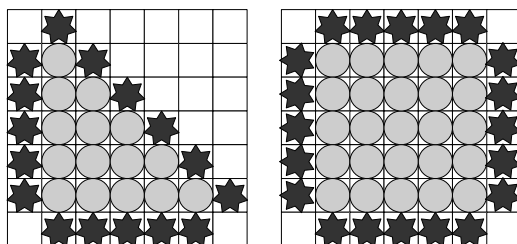


Fig. 1 Rectangle pattern vs. triplet pattern

motion area could be estimated by the pattern with the smallest sum of SAD.

Because the Full Search is positioned as a realistically best solution to find accurate motion vector, we define the motion vector hit rate (MVHR) as a ratio of that the motion vector correctly searched by these patterns to that of using FS method. On the other hand, we have to define the hitting range which shows hit or not. We defined "within fit" and "near fit" as depicted in Fig.2 for the rectangle pattern and the triangle pattern separately. If the motion vector hits the circle points only, define it to "within fit", if hitting the surrounding start points we define it to "near fit".



○ = within the pattern
○ & ★ = near the pattern
Fig. 2 Within fit vs. near fit

Using these patterns showed in Fig.1 some evaluation of the performance for these patterns are evaluated.

TABLE I
MVHR (TRIANGLE PATTERN VS. RECTANGLE PATTERN)

sequence	Shape	MVHR (%)	
		within hit	near hit
Tennis	Rectangle	78.38	89.31
	Triangle	79.32	90.05
Foreman	Rectangle	77.07	88.02
	Triangle	78.50	89.77
Bus	Rectangle	72.76	87.59
	Triangle	74.15	89.22
Football	Rectangle	71.66	86.51
	Triangle	72.63	87.79

The simulation results shown in Table 1 indicate that triangle pattern with near fit always gives a high MVHR than rectangle pattern and within hit does. Therefore, the triangle pattern is considered as an optimal shape to describe a motion area.

B. Pattern Distribution

From these results what are described above triangle patterns show the highest MVHR. Using these results the size of triangle patterns and the distribution of these patterns have to be discussed to detect optimal triangle patterns. Because the motion estimation process for H.264/AVC is sensitive to the search range due to the introducing of the motion vector predictor. Therefore a concept of rings which indicate the distance from the search center is defined in Fig.3.

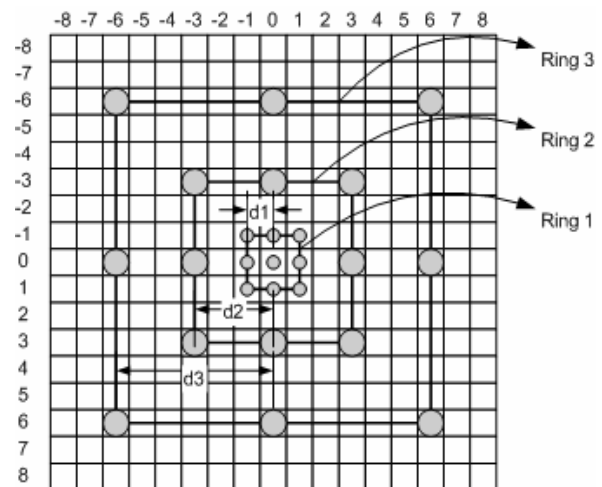


Fig. 3 Definition of rings

As shown in Fig.3 the 25 search point candidate distribute in unequal distance and construct three rings. The d1, d2, d3 indicate the ring size of ring1, ring2, ring3 respectively. Set the ring to linear size with d1=2, d2=2, d3=2, we compare the MVHR with that of logarithmic ring size of d1=1, d2=3, d3=6. From the simulation result shows in Tab. 2 we found that logarithmic ring size can achieve higher MVHR than linear ring size.

TABLE II
MVHR (LINEAR RING SIZE VS. LOGARITHMIC RING SIZE)

sequence	Distribution	MVHR(%)	
		Inter	Near
Tennis	Linear	79.32	90.05
	Logarithmic	89.59	93.82
Foreman	Linear	78.50	89.77
	Logarithmic	88.86	94.33
Bus	Linear	74.15	89.22
	Logarithmic	87.92	91.98
Football	Linear	72.63	87.79
	Logarithmic	85.25	90.42

Several kinds of the ring size combination are simulated. As a result, $d_1=1$, $d_2=3$, $d_3=6$ are optimal ring size to achieve the best MVHR. Therefore, the logarithmic ring size of $d_1=1$, $d_2=3$, $d_3=6$ is adopt in the following discussion and simulations. Using this concept of rings some basic triplet patterns are designed which are shown in Fig.4.

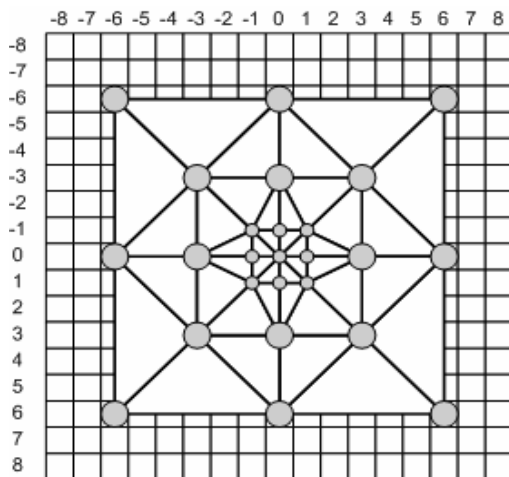


Fig. 4 Triangle patterns

As shown in Fig.4 each of the search point among the 25 candidates together with the adjacent 2 candidates construct 40 triplet patterns. These triplet patterns divide the search range into several areas. Due to the logarithmic ring size the candidate near the center construct small motion area than that of locating far from the center of the search range. This feature ensures that the real motion near the search center have higher probability to be searched with lowest computation complexity.

C. Overlap Patterns

These 40 adopt patterns are defined as basic patterns. Besides these basic patterns there may be some overlap patterns which could achieve higher MVHR. The adjacent patterns are combined and totally eight overlap patterns in Fig.5 which cover the boundary of these basic patterns or across the ring scope are used to evaluate the performance by the MVHR.

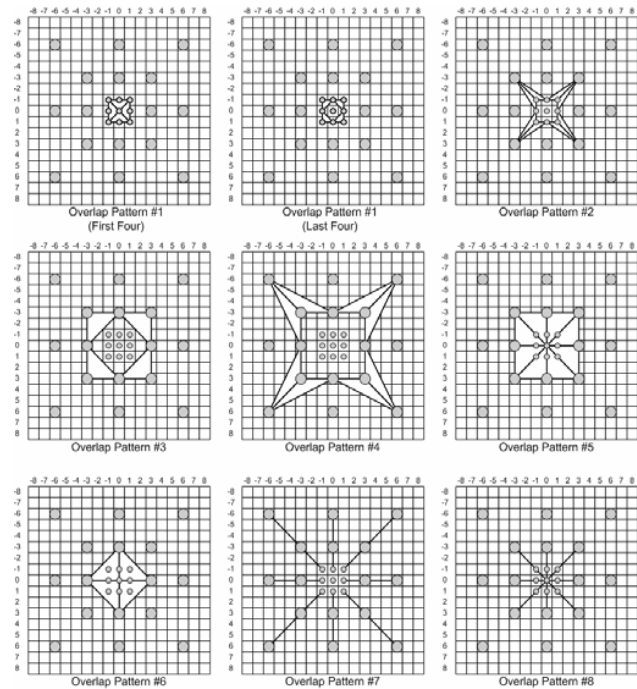


Fig. 5 Overlap patterns

The MVHR of these patterns are simulated using the same video sequence tennis, foreman, bus, and football respectively. Simulation result is shown in Tab.3.

TABLE III
MVHR OF OVERLAP PATTERNS

Overlap patterns	Pattern Number	MVHR (%)			
		Tennis	Foreman	Bus	Football
		inner	near	inner	near
1	48	93.84	94.35	91.99	90.43
2	48	93.88	94.31	91.98	90.53
3	44	93.83	94.35	92.00	90.45
4	48	93.85	94.39	92.01	90.48
5	48	93.93	94.40	92.08	90.50
6	44	93.93	94.41	92.07	90.49
7	48	94.39	94.43	93.17	91.36
8	48	94.11	94.93	92.37	90.77

From these simulation results it is obvious that pattern 5#, 6#, 7# and 8# improved the performance higher but, the improvement by patterns of 1#, 2#, 3# and 4# are slight. To evaluate how many overlap patterns are realistically enough to achieve the best performance another simulation is performed to evaluate the MVHR of the combination of these overlap patterns in addition to the 40 type basic patterns.

From the results showed in Fig. 7 the MVHR is increasing when using some supplemental overlap patterns. However, it does not keep improving performance when increasing overlap patterns numbers are used. On the contrary when using too

much patterns the hit rate will going down. Fig.7 shows that using about 60~70 patterns achieve a peak MVHR.

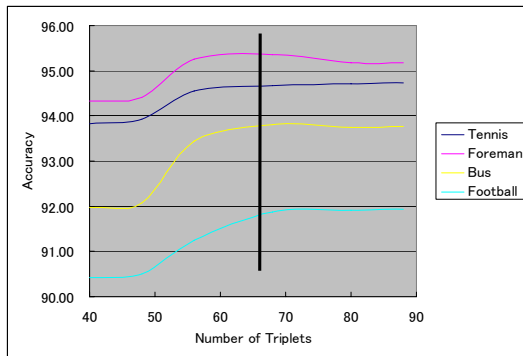


Fig. 6 MVHR with the combination of overlap patterns

From the discussion above, totally 68 patterns which include 40 basic patterns and 24 overlap patterns shown in Fig.7 are selected as the final triplet patterns.

III. PROPOSED ALGORITHM

These finally selected patterns are considered as the optimum patterns to describe the motion area. The areas which have higher MVHR should have high possibility to find the real

motion in this area. Using these triplet patterns in the first search step it is possible to find the real motion with high possibility and to avoid dropping into the local minimum in the first step.

A. 3-step search in the motion area

After decision of the motion area in the first step a localized area is restricted as the search area. This restricted small area limit the search point candidates. Then, in the next two steps the detail search patterns should be designed for each motion area. The search patterns for the step 2 and step 3 are proposed as described in Fig.8.

As shown in Fig.8 the detail search point candidates for step 2 are designed for each pattern of the 40 basic patters and 28 overlap patterns separately. The “●” indicate the vertex of the selected pattern of the motion area and the “△” indicate the search point candidates for step 2.

Fig.9 gives an example of the 3-step TS-ME in which the three “●” stand for a selected triplet pattern, and the triangles describe the selected motion area in the step 1. Then the search points “△” are the search candidates in the step 2. Suppose the “▲” in the motion area is the point with the smallest SAD among these candidates, the surrounding four search point with the mark of “☆” indicate the candidates in the step 3. At last the

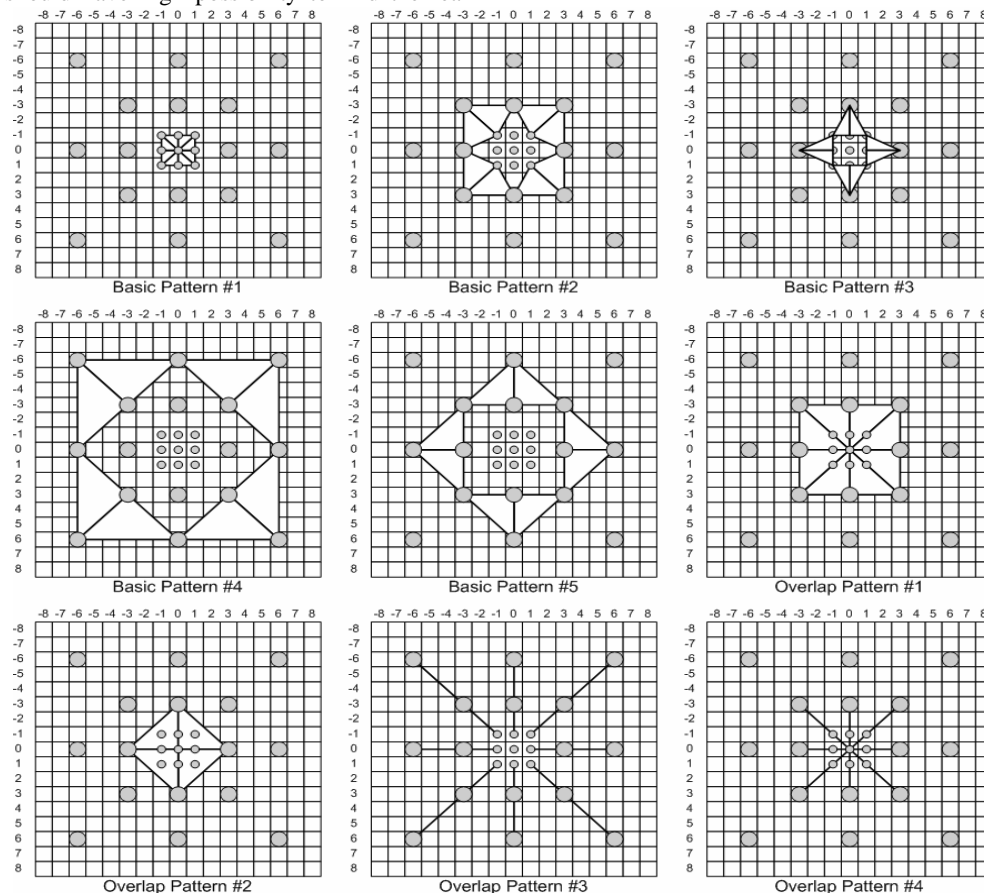


Fig. 7 Proposed triplet patterns

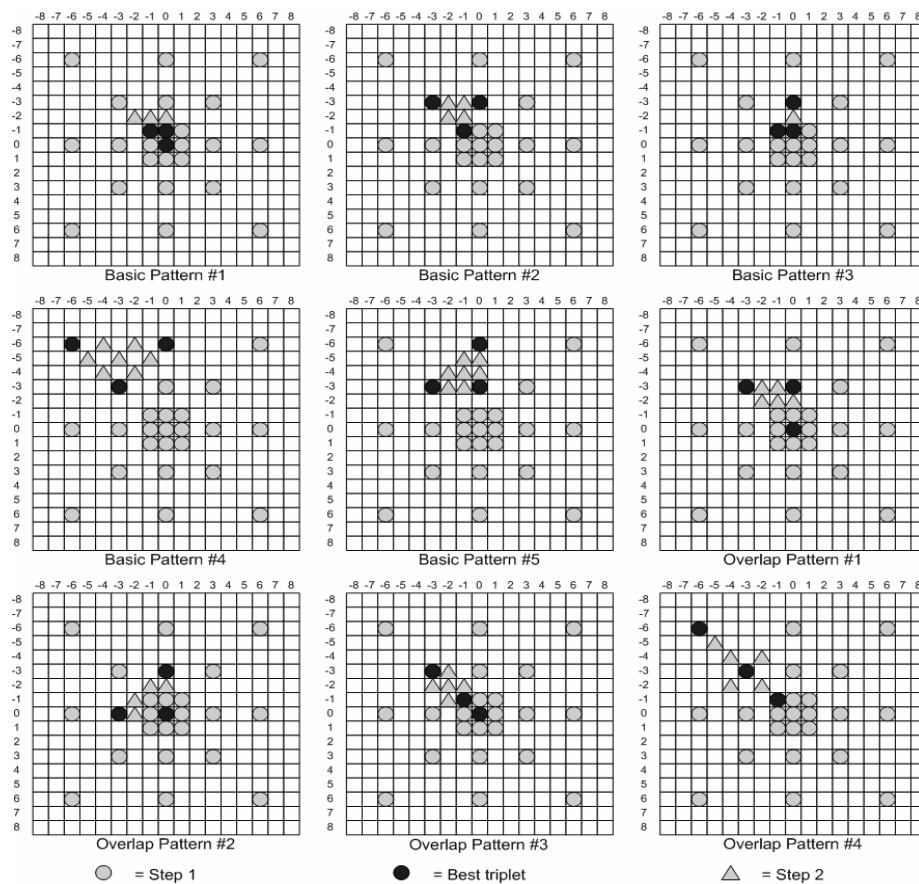


Fig. 8 Search patterns for step 2

best point “★” with the smallest SAD is selected as the integer motion vector. This 3-step process constructs the basic algorithm of proposed algorithm. Another two improvements are also proposed to decrease the computation complexity which will be discussed in the next subsection.

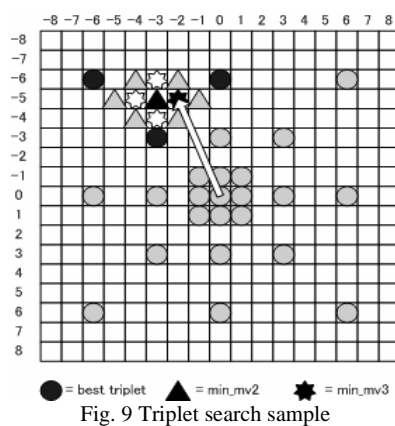


Fig. 9 Triplet search sample

B. Referencing the Surrounding Motion Vectors

Numbers of traditional methods which reference the surrounding motion vectors to reduce the computation

complexity are proposed. However, in H.264, the motion vectors coded in the bitstream define the search center on the basis of the surrounding motion vectors. This feature makes the motion vector more probably to be in the center of searching center or be a small distance from the center. In this paper, utilizing this feature some new rules are proposed to terminate the search process at an early period.

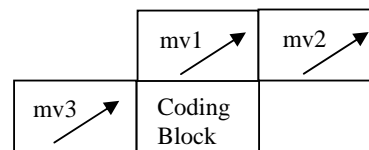


Fig. 10 Reference Motion Vectors

First, pred_diff and mv_predictor are defined as formula (1) and (2) using the surrounding motion vectors which are shown in Fig.10.

$$\text{pred_diff} = \max(\text{mv1}, \text{mv2}, \text{mv3}) - \min(\text{mv1}, \text{mv2}, \text{mv3}) \quad (1)$$

$$\text{mv_predictor} = \text{median}(\text{mv1}, \text{mv2}, \text{mv3}) \quad (2)$$

When the pred_diff is equal to zero the magnitude of motion vectors would be very small and the SAD of the center point

would be very small as well. There is a high probability that the motion vector of current block is (0,0) or the motion vector is within ring1. Therefore, it is enough to only check the triplet patterns in ring1 and choose the best candidate as motion vector. On the contrary when the pred_diff is big we will start search from the ring3. The detail flowchart of the ring selection is described in the Fig.11.

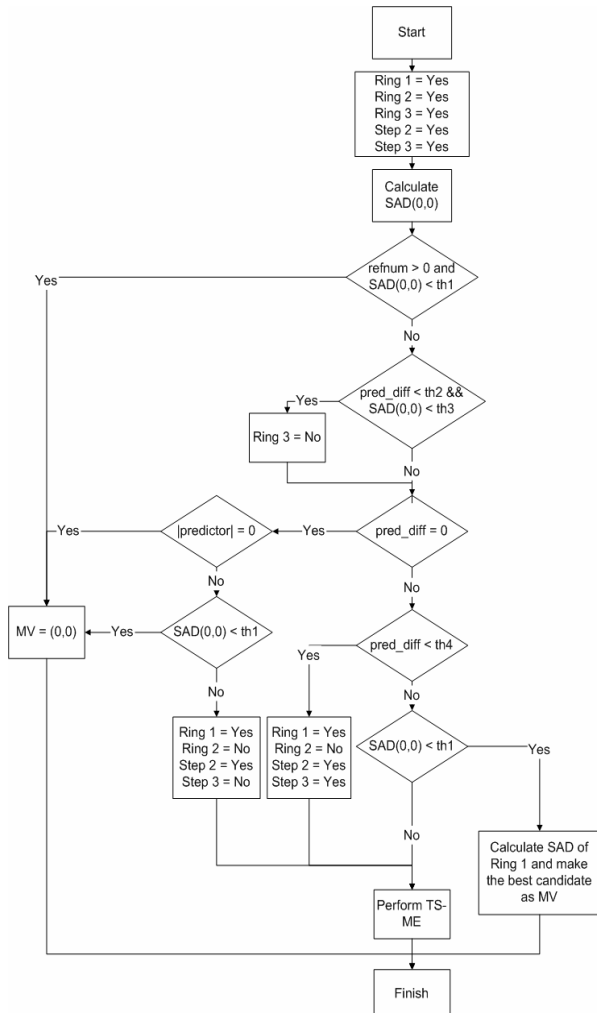


Fig. 11 surrounding motion vector reference rules

As described in Fig.11 to determine whether the difference of magnitude or SAD is small or not, we use several threshold values which are set very small in order to reduce prediction misses.

C. Sub-pixel Motion Estimation

We found a characteristic of the 1/2, 1/4 pixel motion vector results using several image sequences from which it seems not to be uniformly distributed as shown in Fig.12. Around 50%-60% of MVs are located in the center, around 15%-18% of MVs are located in the x-axis direction, while 12%-17% of MVs are located in the y-axis direction. The remaining 11%-14% are located in four corners.

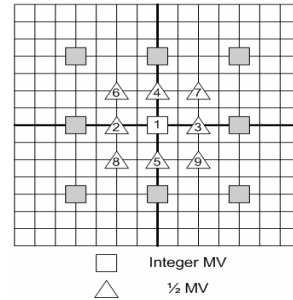


Fig. 12 Search order of 1/2, 1/4 pixel motion vectors

A search order as shown in Fig.12 and searching stop rules shows as follows:

- 1) In case of n is in $[1..5]$: if the $\text{SAD}(n)$ is less than threshold th1 , terminate the process.
- 2) In case of n is in $[6..9]$: if the $\text{SAD}(n)$ is less than threshold th2 and the smallest SAD at that time is $\text{SAD}(0)$, terminate the process.

The n is the search order. The process for 1/4 MV is performed similarly. Because the earlier the termination the greater the risk of miss, it is better to make the rule harder to be passed for small n . With this reason, in case n is $[1..5]$, we make the value of th1 is smaller than the value of th2 to make the rule harder to be passed. Only the destined candidate will pass the rule very early.

IV. SIMULATION RESULT

From the viewpoint of search accuracy, computation cost, and video quality we evaluate the features of the proposed TS-ME method using the reference software JM7.3 [5]. QCIF Image sequence Foreman(100frames), Salesman(100frames), Tennis (250frames), and CIF sequence Bus(150frames), Flower (250frames), Football(125frames) are used and the reference frames is set to 5.

A. Simulation results of the shape patterns

First, we simulate the MVHR compare with other methods to evaluate the motion vector accuracy. The abscissa shows the image sequences, and the ordinate shows the MVHR.

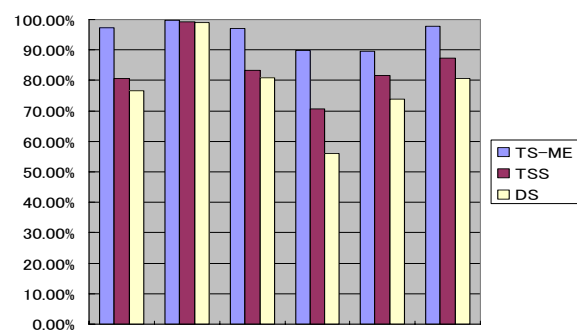


Fig. 13 Motion Vector Hit Rate

From the simulation results show in Fig.13, proposed method shows a high MVHR than TSS and DS at the case of all image sequences. Second, to evaluate the computation cost of TS-ME, the average searching point number of TS-ME, TSS, and DS

method are simulated. The simulation results are shown in Fig.14.

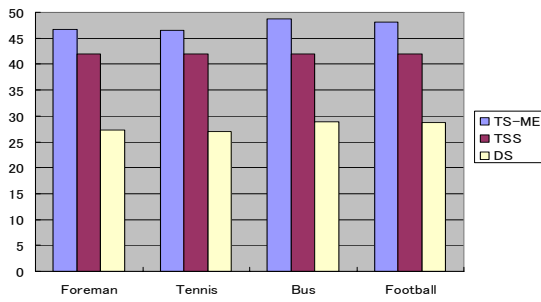


Fig. 14 Average Searching Points

As the Fig.14 shown, compare to TSS, DS method, TS-ME have a little higher computation cost. Last, the comparison of PSNR is performed. The simulation results are shown in Fig.15. The abscissa shows the bit rate, and the ordinate shows the PSNR.

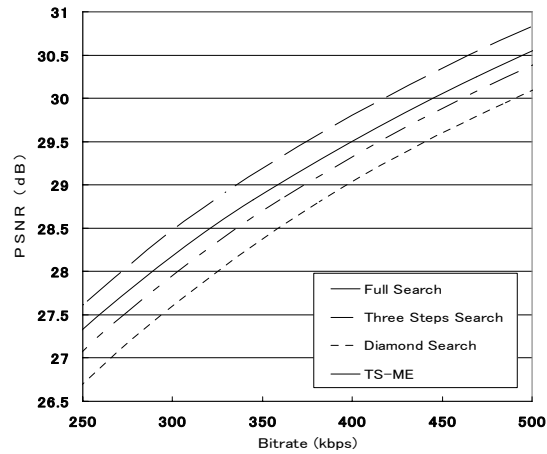


Fig. 15 PSNR-Bus

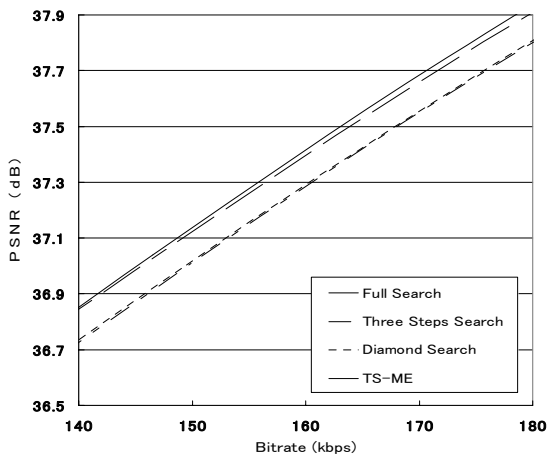


Fig. 15 PSNR-Foreman

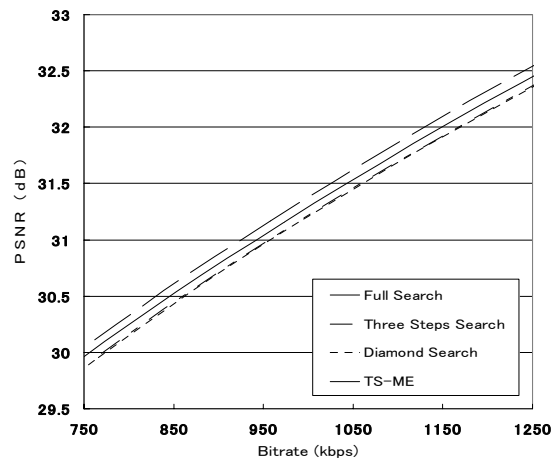


Fig. 15 PSNR-Football

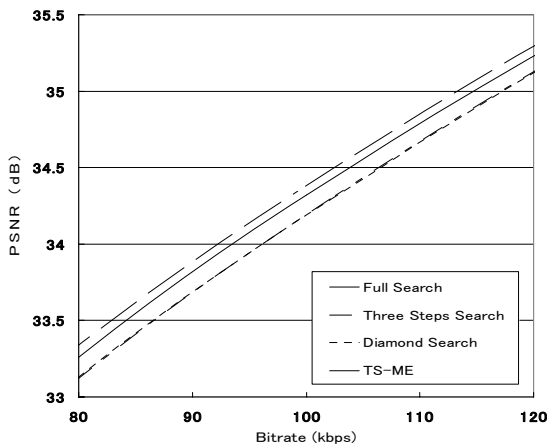


Fig. 15 PSNR-Tennis

The simulation results which are shown here used those image sequences with most drastic motions like Foreman, Bus, Football and Tennis. As the simulation results shown in Fig.15, the proposed TS-ME shows the best performance than the other two traditional methods and shows very close performance to the FS.

B. Simulation results when referencing the surrounding MV

As the simulation results shown in the Fig.14 the computation complexity is still higher than traditional methods. By introducing the method mentioned in section 2 referencing the surrounding motion vectors could achieve further computation complexity decrease. In this sub-section the evaluation when using the method of referenced surrounding motion vectors is proposed. The simulation evaluate from the viewpoint of the MVHR, the computation complexity and the PSNR using the same image sequences. Fig.16 shows the simulation result for MVHR.

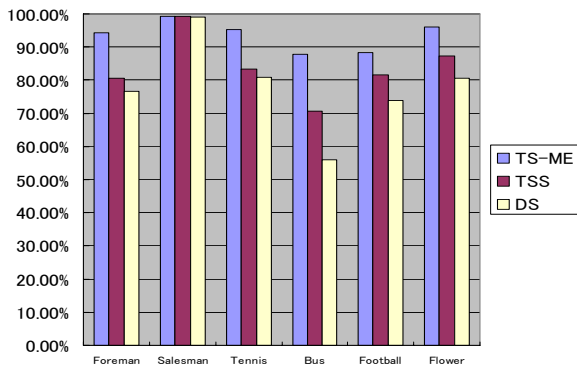


Fig. 16 Motion Vector Hit Rate

From the simulation results it is obvious that referencing the surrounding motion vectors have almost no inference to the MVHR. The average searching point is evaluated in the next simulation and the simulation results are shown in Fig.17.

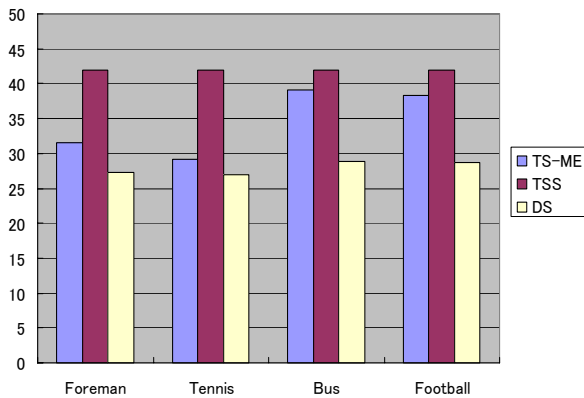


Fig. 17 Average searching points

As the Fig.17 shows the average searching points decrease to almost the same level as the other methods do. Last, the simulation results for the PSNR are shown in Fig.18.

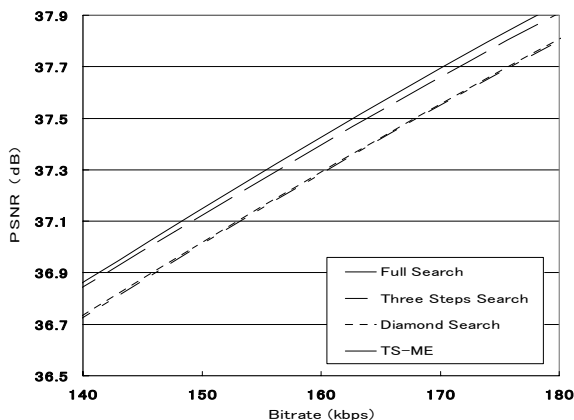


Fig. 18 PSNR-Foreman

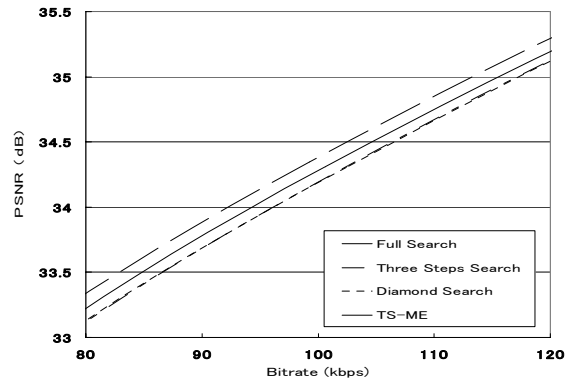


Fig. 18 PSNR-Tennis

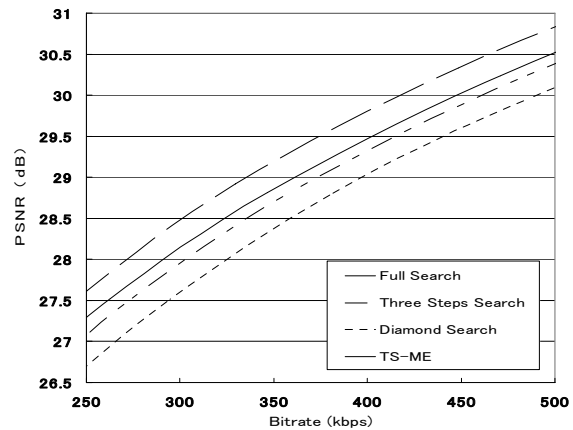


Fig. 18 PSNR-Bus

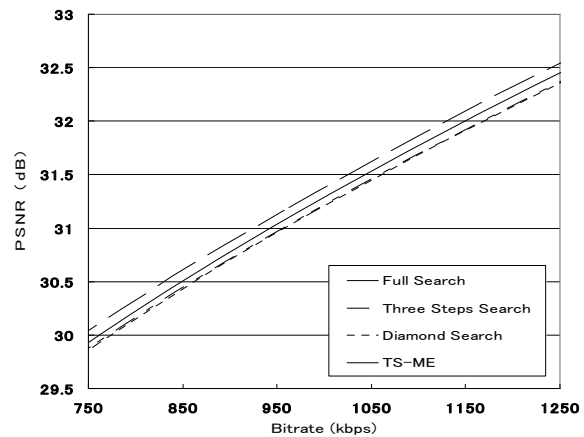


Fig. 18 PSNR-Football

As shown in Fig.18 referencing the surrounding motion vectors have almost no inference to PSNR.

C. Simulation results for Sub-pixel

Another proposed method which is discussed in the previous section together with the proposed method so far is evaluated using the same image sequence and the same viewpoint. The simulation result of the MVHR is shown in the Fig.19.

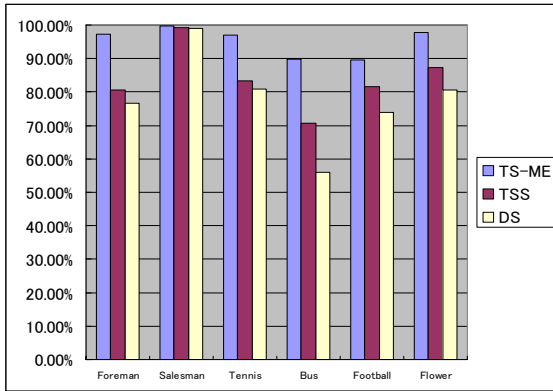


Fig. 19 Motion Vector Hit Rate

As this simulation result shows the MVHR still keep in higher level than other methods with almost no variation compare to the previous simulations. The simulation results for the average search points are shown in Fig.20.

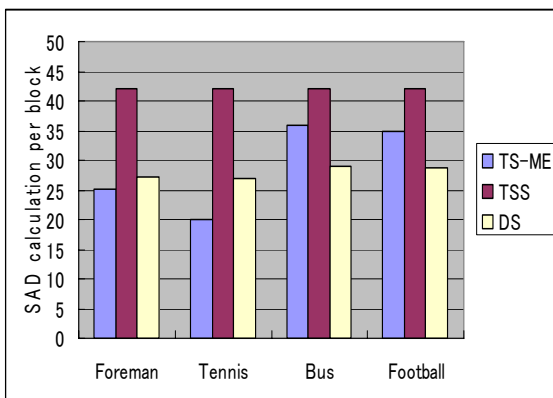


Fig. 20 Average searching points

As shows in Fig.20, TS-ME could accomplish the motion search with an average search point of 29 which is lower than the TSS which has a fixed search point of 39, but almost the same as DS method which has an average search point of 28. PSNR comparing with the FSS, TSS, and DS methods are also evaluated which are shown in Fig.21.

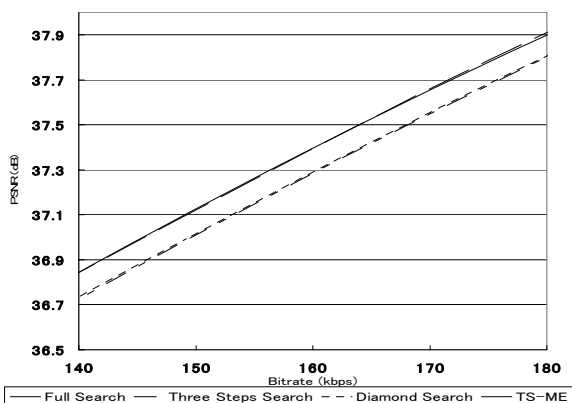


Fig. 21 PSNR-Foreman

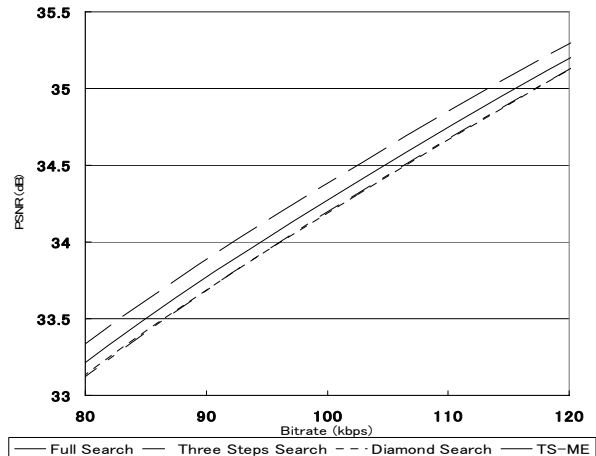


Fig. 21 PSNR-Tennis

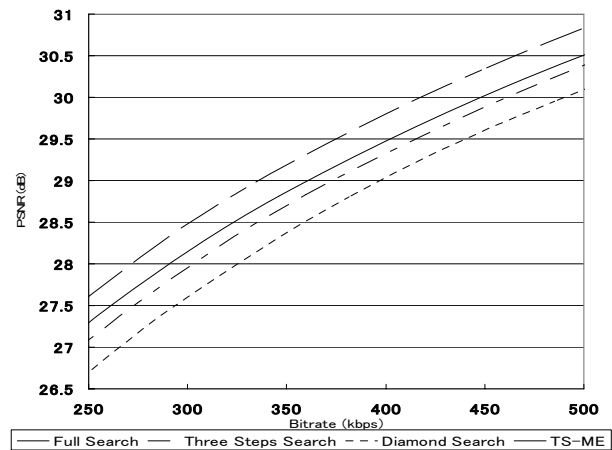


Fig. 21 PSNR-Bus

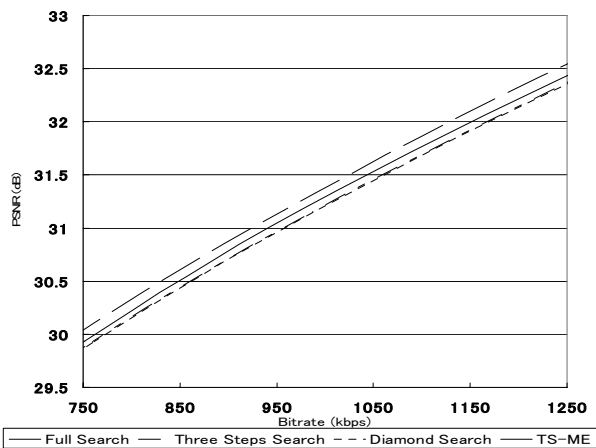


Fig. 21 PSNR-Football

From these simulation results, we found the proposed TS-ME together with proposed improved method could achieve a very close PSNR value to FS as well as a low computation complexity than all other traditional methods.

V. CONCLUSION

In this paper a novel fast motion estimation algorithm is proposed. Due to the characteristic motion estimation process of H.264/AVC, the integer pixel motion estimation make more important role than the traditional algorithms with which could induce dropping into local minimum. Proposed algorithm introduced a new concept of motion area to localize the motion vector in a small area with high accuracy. The concept of the motion area and the proposed triplet patterns are the important features of proposed algorithm which is different from the traditional methods. Another two additional algorithms are also proposed to further decrease the computation complexity. Using these two methods the computational reduction for both integer pixel and sub-pixel motion search are realized while maintaining high picture quality. Based on our performance evaluation, proposed method shows superior quality and computational efficiency over traditional methods. Simulation results show that proposed TS-ME together with the additional improved methods could accomplish the motion estimation 38 times faster than traditional full search method.

There are also other novel proposals which are concerning the other approaches for motion estimation of H.264/AVC. In this paper, proposed algorithm achieved good performance from the viewpoint of search area but did not adopt those proposals into proposed method. Make use of proposed idea together with other excellent method may be able to achieve better performance. Proposed idea is considered not only an efficient method for H.264, but also a suitable method for other motion estimation algorithm.

REFERENCES

- [1] T. Wiegand, G. Sullivan, and A. Luthra, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," Joint Video Team (JVT) of ISO/IEC MPEG&ITU-T VCEG, May 1998.
- [2] Y. L. Xi, C. Y. Hao, Y. Y. Fan and H. Q. Hu, "A fast block-matching algorithm based on adaptive search area and its VLSI architecture for H.264/AVC," in *Journal of Signal Processing: Image Communication*, vol.21, no.8, pp.626-646, sep. 2006.
- [3] Z. B. Chen, J. F. Xu, Y. He and J. L. Zheng, "Fast integer-pel and fractional-pel motion estimation for H.264/AVC," in *Journal of Visual Communication and Image Representation*, vol.17, no.2, pp.264-290, April 2006.
- [4] Z. Zhou, J. Xin and M. T. Sun, "Fast motion estimation and Inter-mode decision for H.264/MPEG-4 AVC encoding," in *Journal of Visual Communication and Image Representation*, vol.17, no.2, pp.243-263, April 2006.
- [5] K. Ogata, K. Saito, T. Song and T. Shimamoto, "Variable search range motion estimation algorithm for H.264/AVC", in *IEICE Society Conference*, Sep. 2006
- [6] R. Li, B. Zeng, and M. L. Liou: "A new three-step search algorithm for block motion estimation," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 4, no. 4, pp. 438-442, Aug. 1994.
- [7] S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. on Image Processing*, vol. 9, no. 2, pp. 287-290, Feb. 2000.
- [8] Y. Nie and K. K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," in *IEEE Trans. on Image Processing*, vol. 11, no. 12, pp. 1442-1449, Dec. 2002.
- [9] <http://bs.hhi.de/~suehring/tml/>

T. Song received his B.E. degree from the Dalian University of Technology, China, in 1995, M.E., Dr.E from Osaka University in 2001, 2004 respectively. From 2004, he joined the faculty of the department of Electric and Electronics

at Tokushima University. He is a member of IPSJ, IEICE, and IEEE. His current research interests include video coding algorithm, VLSI architectures and system design methodology.

T. Shimamoto was born in Tokushima, Japan, on November 22, 1959. He received the B.E. and M.E. degrees in electrical engineering from Tokushima University, and the Dr.E. degree from Osaka University, in 1982, 1984, and 1992, respectively.

He joined Tokushima University in 1984 as an Assistant Professor. Presently he is an Associate Professor of the Department of Electrical and Electronic Engineering, Graduate School of Advanced Technology and Science, Tokushima University. His interests of research include heuristic algorithms for VLSI CAD.