International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.4 (2014), pp.93-102
http://dx.doi.org/10.14257/ijmue.2014.9.4.10

A Fast Adaptive Block-matching Motion Estimation Algorithm

Youwei Yuan[1], Weilei Xu[1],Yong Li[1] and Lamei Yan*

[1]School of Computer Science and Technology, Hangzhou Dianzi University,Hangzhou, Zhejiang, P.R China, 310018
** School of Media and Design, Hangzhou Dianzi University,Hangzhou, Zhejiang, P.R China, 310018*

** y.yw@163.com*

Abstract

In this work, we have developed a new bio-inspired neural network algorithms for block-based motion estimation. The main goal is to bridge the gap between algorithmic and biological vision by suggesting a bio-inspired motion estimation model based on neural network. We simplify the matching criterion for the block matching algorithm to reduce the hardware complexity and a number of input ports which maintaining the good quality. This paper implements the optimized algorithm in the reference model of H.264 compiled by VC6.0, and chooses six typical video sequences for simulation. The results show that our algorithm can reduce the average search points up to 82% to the full search black-matching algorithm. The optimized algorithm has reduced the motion estimation by 13.792% compared with UMHexagonS, and it gets better optimization to video testing sequences with low complexity.

Keywords: *Bio-inspired, neural networks, Motion Estimation (ME), Peak signal-to-noise ratio (PSNR)*

1. Introduction

Motion estimation is one of the key features for many vision and robotic application. Several methods have been proposed in the literature, but the H.264/AVC is the most applicable method used to estimate motion. Basically, H.264/AVC has applied advanced block-size macroblock modes, sub-pixel motion estimation(ME), multiple reference frames, deblocking filter, integer discrete cosine transform (DCT), and efficient entropy coding techniques (Luthra *et al.*, 2003). Also, it is the most recent video coding standard that meets the need for high quality video at lower bit rates. However, the H.264/AVC requires greater power consumption due to high complexity coding (Wiegand *et al.*, 2003). Usually, the motion estimation consumes up to 80% of the total encoding time. Thus, it is an issue on how to speed up motion estimation in H.264/AVC in order to meet real-time and low power consumptions. Some scholars have proposed different methods to optimize the complexity of the H.264 coding process in video quality (Sullivan and Wiegand, 2005, Sarwer and Wu, 2009, Maas and Sontag, 2000).

Bio-inspired neual network is a kind of processiong system which can contain a huge number of highly interconnected processing neurons. Bio-inspired neurons mimic the functional motion behavior of a real biological neuron. Generally speaking, all these neurones remove a spike whenever their intemal potential exceeds some predefined threshold. These bio-inspired neurons are more generalized than classical neural networks. Moreover the

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2014 SERSC

information in these bio-inspired networks can be coded in the phase and these codings make bio-inspired neurons mathematically much more powerful than classical neural networks (Jannson et *al.*, 2009).

Tomasz Jannson *et al.*, presented a novel approach to bio-inspired neural networks which follow the infrastructure of the biological brain neurons, except that biological synapses are also replaced by virtual ones based on cellular telephony modeling (Battiato and Rundo, 2009). In (Escobar *et al*., 2009) the inspired neural networks based core is able to afford a method to extract the DNA gene expression information from DNA images.

This paper proposes a new approach using bio-inspired neural networks in order to reduce block-based motion estimation with low complexity. The proposed approach able to perform a preliminary segmentation of block-based image getting foreground. Next, we present an image compression method which can guarantee the lossless encoding of the pixel data with lower intrinsic entropy than original 16-bit images. The results show that the proposed approach reduces the motion up to 13.7% and search point up to 82%.

The rest of the paper is organized as follows: Section 2 provides the proposed bio-inspired neural networks structure, the connections, and the neurons functionalities. Section 3 moves through the algorithm principle and experimental results are shown in section 4. Finally section 5 addresses a set of main conclusion and future works.

2. Bio-inspired neural networks

2.1. The Structure

A multi-layer neural network can be established by assembling basic two layer neural network (Zheng *et al.*, 2011, Zhu and Ma , 2000, Zhu *et al*., 2011). The frame of bio-inspired neural network is shown in Figure 1 that consists of one input layer, one output layer, and multiple hidden layer. From Figure 1, the solid lines represent the bottom-up process and dash lines describe top-down process. The input layer is called data layer with *I* neurons. The output layer is called the feature layer with *N* neurons. Neurons of different layer are connected. The input layer receives the input image and generates a few hypotheses at the output layer via the bottom-up pathway, the output layer then produces refined motion estimation image through the top-down pathway.
The input layer is decided as the following equations:

$$\frac{dx_i}{dt} = b_i, \theta_i = 1, \rho_i = 0.2\sin 2\pi t \tag{1}$$

where x_i, θ_i are internal potential and the threshold respectively, and ρ_i represents the relaxation level of neural networks *I*, and b_i is the binary input .

For the output layer ,we have the following equations:

$$y(t) = \sum_{b_j=1}\sum_{b_l=1} y_i(t - d_l(i, j)) \tag{2}$$

$$y_i(t) = \sum_{n=n_0}^{\infty} \delta(t-n) \tag{3}$$

where $d_l(i,j)$ *is the lateral distance* and n_0 represent positive integer. The expression for the output of the final summing node now becomes:

Copyright ⓒ 2014 SERSC

$$y(t) = \sum_{n=n_0}^{\infty}\left[\sum_{b_j=1}\sum_{b_i=1}\delta(g - d_l(i, j) - n\right] \tag{4}$$

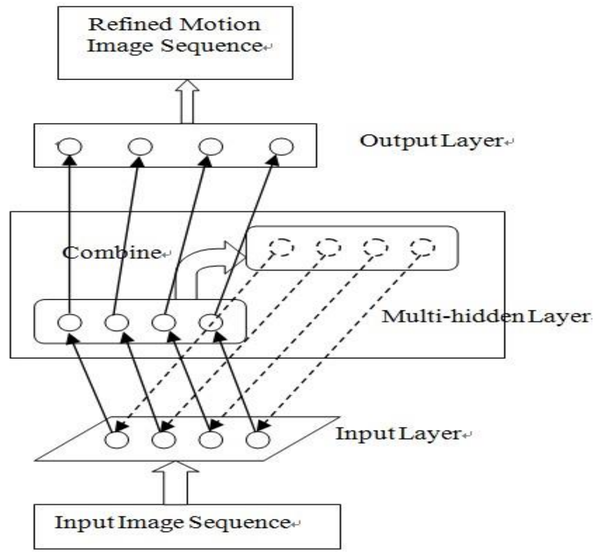


Figure 1. The proposed bio-inspired neural network structure

2.2. The transform-domain transcoding

Figure 2 shows our transform-domain transcoder architecture using bio-inspired neural network. VC-1 uses a variation of the discrete cosine transform to convert blocks of samples into a transform domain to reduce the complexity coding. H.264 encoder works on both macroblocks and motion-compensation and the video is formed by a series of picture frames. Each picture frame is an image which can be splitted down into blocks and the block sizes can be different in H.264. The encoder also behaves intra-coding for the macroblocks of a picture (Zheng *et al.*, 2011). The incoming video stream has been decoded through variable length decoding followed by inverse quantization as done in (Jannson *et al.*, 2009). The H.264 employs a hierarchical transform structure and DC coefficients of neighboring 4x4 transforms for luma and chroma signals are divided into 4x4 blocks. The blocking artifacts have been removed by the deblocking filter owing to the block based encoding pattern. The transform applied after intra-prediction or inter-prediction is on blocks; the transform coefficients then undergo quantization. The de-blocking filter works on it to get rid of the artifacts and to achieve the macroblocks for the subsequent input using bio-inspired neural networks (Zhu *et al*., 2011, Yan *et al.*, 2006).

3. The Algorithm

This paper proposes an efficient adaptive search range algorithm based on bio-inspired neural networks. An approximate functions SAD(X, Y) is used to obtain low-resolution video frames for both the current and the reference frames. The function is similar to the one used in (Lim *et al.*, 2003, Yan and Liu, 2010).

The new technique composes four distinct steps:

(1) The first is a preprocessing phase through which:

The median value called MVP and COL_MV in the dynamic part:

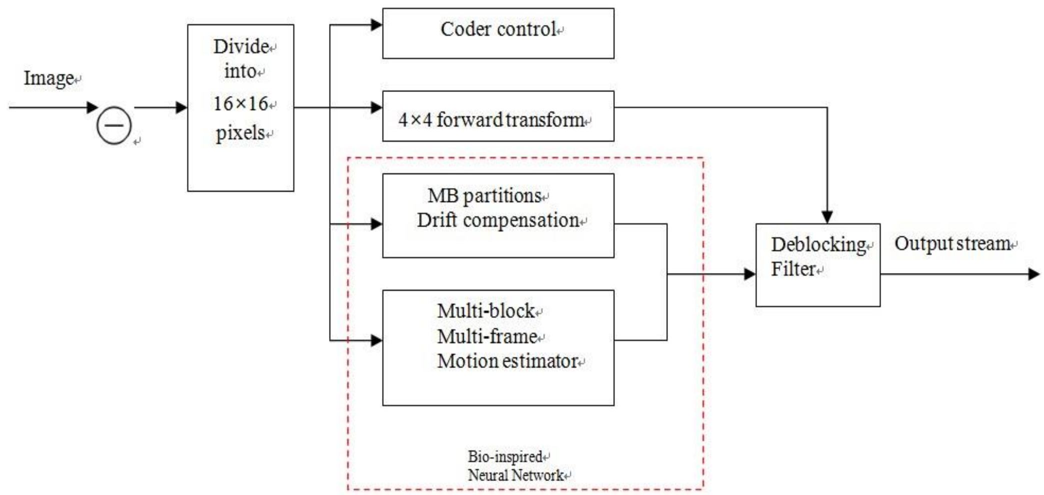


Figure 2. The structure of Transform-Domain Transcoding

$$\text{dynamic-part}=\max\{|MVP\_X-COL\_MV\_X|,|MV-COL\_MV\_Y|\} \tag{5}$$

where MVP represents the median value of the current block's neighboring blocks, and COL_MV is the block's motion vector in the previous frame. A vector followed by "_X(Y)" stands for its X(Y)-axis projected length based on the observation *R* defined as follows:

$$R=\begin{cases}0 & if \quad E\le 2\\ 1 & otherwise\end{cases} \tag{6}$$

$$\text{and} \quad E=SAD/(M*N) \tag{7}$$

where *E* represents the average pixel error, SAD is the cost of the search center (0,0) and M*N is the size of the current block.

$$SAD_{pred\_MD}=\min(SAD_{x\_median},SAD_{y\_median}) \tag{8}$$

Where SAD_{x_median} represents the abscissa of motion vector in corresponding block A,B and C:

$$x\_median = Median(MV_A(x),MV_B(x),MV_C(x)) \tag{9}$$

$$MCOST_{\min}=\arg[\min_{mi} J(m_i,\lambda_{MOTION})], s.t. mi\in S \tag{10}$$

The search of asymmetric cross-block is below:

$$\Omega_1=\{m=(m_x,m_y)^T \mid m=(cm_x\pm 2i,cm_y)^T, i=0,1,2,...,\frac{W}{2}; m=(cm_x,cm_y\pm 2j)^T, j=1,2,...,\frac{W}{4}\}$$

$$cm=m_{\min}$$

 Copyright ⓒ 2014 SERSC

Where W is the size of search area, the search step includes 3 steps:

(a) $m_{\min 2} = \arg[\min_{m_i} J(m_i, \lambda_{MOTION})], s.t. m_i \in \Omega_1$

(b) $m_{\min} = \arg[\min J(m_{\min}, \lambda_{MOTION}), J(m_{\min 2}, \lambda_{MOTION})]$

(c) Early-Termination

(2) The second is expanded search of multi-layer gird of Hexagon with 16points:

$$\Omega_{16-HP} = \{m = (x, y)^T \mid m = (\pm 4, \pm 2)^T, (\pm 4, \pm 1)^T, (\pm 4, 0)^T, (\pm 2, \pm 3)^T, (0, \pm 4)^T\} \quad (11)$$

for (k=1; k<W/4; k++)

{

$\prod_k = \{m = (m_x, m_y) \mid m_x = cm_x + kgx', m_y = cm_y + kgy', (x', y') \in \Omega_{16-HP}\}$,

then：

$$m_{\min k} = \arg[\min_{m_i} J(m_i, \lambda_{MOTION})], s.t. m_i \in \prod{}_k$$

$$m_{\min} = \arg[\min(J(m_{\min}, \lambda_{MOTION}), J(m_{\min k}, \lambda_{MOTION}))]$$

Early-Termination

}

Where the character m is equal to 2^{b1}, b1 represents the number of bits of pixel resolution in the low resolution frames.

(3) Search for small triangle:

Definite the assembly of candidate movement vectors

$$\Omega_2 = \{m = (m_x, m_y)^T \mid m_x - cm_x \mid \le 2, \mid m_y - cm_y \mid \ge 2\}, cm = m_{\min} \quad (12)$$

The search step is following:

$$m_{\min 3} = \arg[\min_{m_i} J(m_i, \lambda_{MOTION})], s.t. m_i \in \Omega_2$$

$$m_{\min} = \arg[\min(J(m_{\min}, \lambda_{MOTION}), J(m_{\min 3}, \lambda_{MOTION}))]$$

Early-Termination

(4) The third is expanded Hexagon search algorithm which can be divided into 2 steps:

Step 1: The definition of Hexagon search model:

$$\Omega_3 = \{m = (m_x, m_y)^T \mid m = (cm_x \pm 2, cm_y)^T, (cm_x \pm 1, cm_y \pm 2)^T\}, cm = m_{\min}$$

Then：

$$m_{\min 4} = \arg[\min_{m_i} J(m_i, \lambda_{MOTION})], s.t. m_i \in \Omega_3$$

$$m_{\min} = \arg[\min(J(m_{\min}, \lambda_{MOTION}), J(m_{\min 4}, \lambda_{MOTION}))]$$

If $m_{\min} = cm_2$, jump to Step 2; otherwise jump to Step 1.

Copyright ⓒ 2014 SERSC

Step 2：Search Algorithm for Diamond：

$$\Omega_4 = \{m = (m_x, m_y)^T \mid m = (cm_x \pm 1, cm_y)^T, (cm_x, cm_y \pm 1)^T\}, cm = m_{\min}$$

$$m = \arg[\min Cost(m, \lambda_{MOTION})], m \in \Omega_4$$

If $m_{\min} = cm$,

then End

4. Simulation Results

The useful objective metric peak signal-to-noise ratio (PSNR) is the ratio between the maximum value of a signal (255 for 8-bit video) and the quantization noise. PSNR can determine the visual quality of the proposed generalized quadtree motion compensation technique and the two-stage global motion compensation technique. For any codec, PSNR is expected to increase at higher bit rates to less noisy and aggressive compression (Botella *et al*., 2010).

In our test, we use JM 10.1 reference software and different types of video sequences with QCIF format and the reference model of H.264 compiled by VC6.0.The test conditions are as follows: all video sequences are tested under the windows XP operating system, and image's YUV=4:2:0, some parameters' setting such as: Input File="video sequence needed to test", FramesToBeEncoded=100, FrameRate=30.0, SourceWidth=176, Sourceheight=144, UseHadamard=1, SearchRange=16, UseFME=1, NumberReferenceFrames=5, QP=28, and others parameters are default setting.

In our experiment, three layered neural networks which include input layer, hidden layer and output layer fully connected network are used. As image contain structure information, we then divide the input image space into a number of sub-spaces. The grid of neurons with 40 rows by 200 columns ((40×200) is chosen for the specific image segmentation implementation. The input is normalized and the initial weighting vectors of the neurons are set equal to the coordinates of the points extracted from the input image. After the process of adaptation of the bio-inspired neural network, the weighting vectors of the input neurons will have values identical to the appropriate points.

The test vector of training phase is performed by addressing the network with the coordinates of selecting points sampled image randomly. The neuron with weight vector closest to signal will be selected to adjust its weight vector and its neighboring neurons as well as modify their weight vectors. The neighboring neurons are confined to a window of 3×3 neurons throughout the network training.

The test consequence is shown in Table 1 where the proposed method reduces both encoding time and ME time while maintain relatively stable PSNR and bit rate.

Figure 3 shows cat's face tracking experiment under a dynamic environment. It shows that our suggested method can always track the cat's face accurately.

As shown in Figure 4, our algorithm can reduce both the total encoding time and motion estimation time as compared to UMHexagonS. Specially, the proposed algorithm behaves well on mobile and coastguard video sequence, which are the typical of big moving sequences.

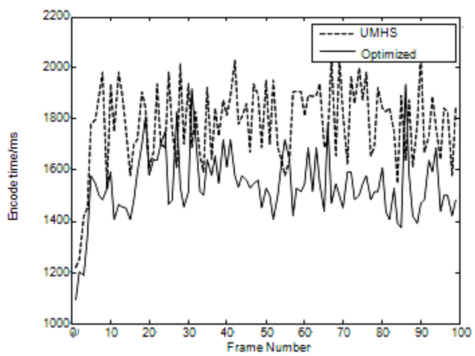
The simulations show that our method is nearly the same as that obtained through operating directly on the sub-bands. It keeps a similar PSNR results on average except some single frames and also works well on mobile and coastguard video sequence, which are the typical of big moving sequences.

 Copyright ⓒ 2014 SERSC

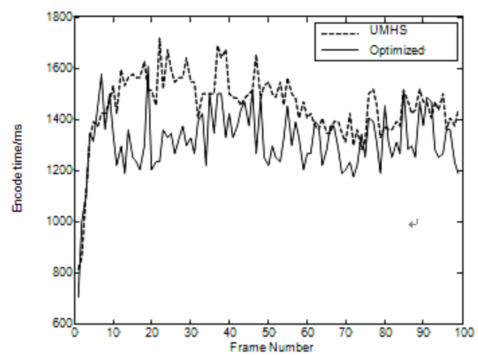
Table 1. Test consequences

Video sequence	*UMHexagonS*				*Our optimized algorithm*			
	PSNR (dB)	*bit rate (kbit/s)*	*Enc.T (s)*	*ME.T (s)*	*PSNR (dB)*	*bit rate (kbit/s)*	*Enc.T (s)*	*ME.T (s)*
Mobile	33.09	265.72	357.411	159.033	33.09	266.43	345.361	151.382
Coastguard	34.28	169.42	291.847	154.666	34.28	169.18	289.468	150.267
Silent	36.13	66.63	267.367	125.538	36.11	66.63	246.084	113.511
Container	36.42	31.15	233.467	102.005	36.42	31.03	230.766	100.938
Highway	37.64	53.26	222.382	110.021	37.65	53.58	218.271	108.393
Foreman	37.64	53.26	219.587	108.370	37.62	54.05	214.701	105.315

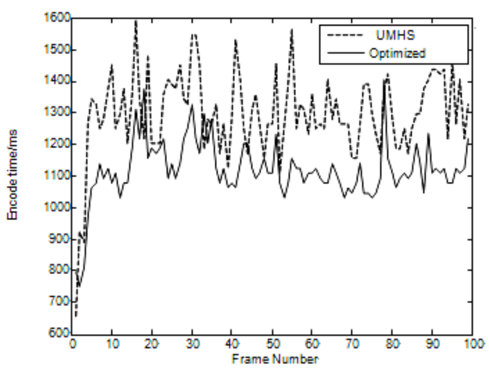
Figure 3. Cat's face tracking experiment



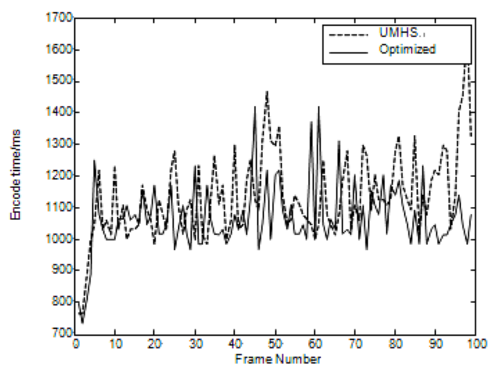
（a）Mobile_qcif.yuv Sequence



（b）Coastguard_qcif.yuv Sequence

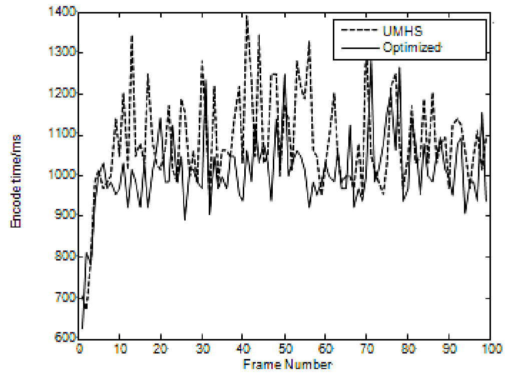


（c）Silent_qcif.yuv Sequence

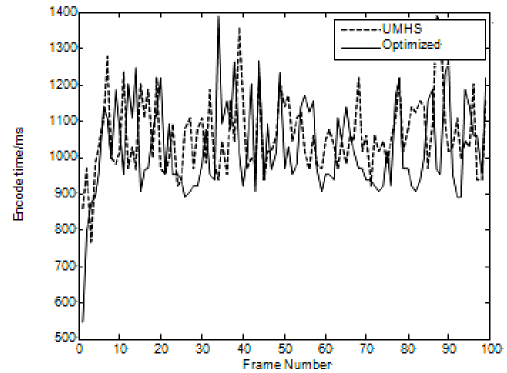


（d）Container_qcif.yuv Sequence

Copyright ⓒ 2014 SERSC



(e) Highway_qcif.yuv Sequence



(f) Foreman_qcif.yuv Sequence

Figure 4. Various Test Sequence

5. Conclusions and future work

As a whole, in this paper we proposed a new bio-inspired neural network algorithm for block-based motion estimation. Our contribution is to bridge the gap between algorithmic and biological vision by suggesting a bio-inspired motion estimation model based on neural network. The paper gets further research of the kernel algorithm named UMHexagonS adopted in H.264, and proposes the optimization algorithms according to the existing deficiencies, which contain new search pattern based on distribution of motion vector, adaptive search range method and spiral subset matching strategy. Another advantage of the proposed methodology is that it can be very easily made compatible to a standard H.264 encoder/decoder.

Experimental results indicate that our proposed method reduces the size of the search window dynamically and enhances encoding efficiency but has few changes in the reconstructed image quality and bit rate. It is recommended that future research involves optimal implementation of the spectral extension of the model.

Acknowledgments

The work was supported by NSFC (Grant No. 61272032) and also supported by the nature science foundation of Zhejiang province (No. Y6090312).

References

[1] A. Luthra, G. J. Sullivan and T. Wiegand, "Introduction to the special issue on the H.264/AVC video coding standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, **(2003)**, pp. 557-559.

[2] T. Wiegand, G. J. Sullivan and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. On Circuits and System for Video Technology, vol. 13, no. 7, **(2003)**, pp. 560-576.

[3] G. J. Sullivan and T. Wiegand, "Video compression—from concepts to the H.264/AVC standard", Proceeding of the IEEE, vol. 1, **(2005)**, pp. 18–31.

[4] M. G. Sarwer and Q. M. Wu, "Adaptive Search Area Selection of Variable Block-Size Motion Estimation of H.264/AVC Video Coding Standard", IEEE International Symposium on Multimedia, vol. 11, **(2009)**, pp. 100-105.

[5] W. Maas and E. D. Sontag, "Neural systems as nonlinear Alters", Neural Computaion, vol. 12, no. 8, **(2000)**, pp. 1743-1772.

[6] T. Jannson, T. Forrester and K. Degrood, "Wireless synapses in bio-inspired neural networks", Proc. of SPIE 09, vol. 7347, **(2009)**, pp. 1-13.

Copyright ⓒ 2014 SERSC

[7] S. Battiato and F. Rundo, “A Bio-inspired CNN with re-indexing engine for lossless DNA microarray compression and segmentation”, ICIP09, **(2009)**, pp. 1737-1740.
[8] M. J. Escobar, G. S. Masson, T. Vieville and P. Kornprobst, “Action recognition using a bio-inspired feedforward spiking network”, International Journal of Computer Vision, vol. 82, no. 3, **(2009)**, pp. 284.
[9] Y. Zheng, Y. Meng and Y. Jin, “Object recognition using a bio-inspired neuron model with bottom-up and top-down pathways”, Neurocomputing, vol. 74, no. 17, **(2011)**, pp. 3158-3169
[10] S. Zhu and K. Ma, “New Diamond Search Algorithm for Fast Block Matching Motion Estimation”, IEEE Transcations on Image Processing, vol. 9, no. 2, **(2000)**, pp. 287-290.
[11] C. Zhu, X. Lin and L. P. Chau, “Enhanced Hexagonal Search for Fast Block Motion Estimation”, IEEE Trans on Circuits and Systems for Video Technology, vol. 14, no. 10, **(2004)**, pp. 1210-1214.
[12] L. M. Yan, X. H. Zeng and M. M. Deris, “A Dynamic and Efficient Method for Three-Dimensional Adaptive Meshing”, Chinese Journal of Electronics, vol. 6, no. 3, **(2006)**, pp. 257-260.
[13] Y. C. Lim, K. Y. Min and J. W. Chong, “A Pentagonal Fast Block Matching Algorithm for Motion Estimation Using Adaptive Search Range”, IEEE International Conference on Acoustics, Speech, and Signal Proceedings(ICASSP’03), **(2003)**, pp. 669-672.
[14] L. M. Yan and B. Liu, “Optimal control of switching systems with impulsive effects”, International Journal of Advances on differential equations, vol. 4, **(2010)**, pp. 1-14.
[15] G. Botella, A. García, M. Rodriguez, E. Ros, U. Meyer-Bäse and M. C. Molina, “Robust bioinspired architecture for optical-flow computation”, IEEE Transactions onVery Large Scale Integration (VLSI) Systems, vol. 18, no. 4, **(2010)**, pp. 616–629.

Authors

Youwei Yuan

He was born in 1966, received his doctor degree in computer science from Wuhan University of Technology, China, in 2007. He is currently a professor of computer science, Hangzhou Dianzi University (Hangzhou, China). His research interests include artificial intelligence, image process and distributed parallel processing, He has published over 50 technical papers in prestigious journals and conferences, 20 papers been indexed by SCI, EI and ISTP.

Weilei Xu

He was born in 1990, is a postgraduate student of computer science, Hangzhou Dianzi University (Hangzhou, China).His research interests include computer vision and image processing.

Lamei Yan

She was born in 1968, is currently a full professor of Hangzhou Dianzi University (Hangzhou, China). Her research interests include three-dimensional reconstruction, artificial intelligence and computer vision.

Copyright ⓒ 2014 SERSC