PROCEEDINGS OF SCIENCE

High Precision of Global Motion Estimation in Multi-dimensional Transform Domain

Yang Yu[1]
College of Communication Engineering,Jilin University ,Changchun, 130022, China
*E-mail:*yy13@mails.jlu.edu.cn

Xinyu Cui
College of Communication Engineering,Jilin University ,Changchun, 130022, China
E-mail: 522022658@qq.com

Aijun Sang[2]
College of Communication Engineering, Jilin University ,Changchun, 130022, China
*E-mail:*sangaj@jlu.edu.cn

Mianshu Chen
College of Communication Engineering, Jilin University, Changchun, 130022, China
*E-mail:*1257870326@qq.com

Jiangjiang Zhong
College of Communication Engineering, Jilin University, Changchun, 130022, China
E-mail: 451642948@qq.com

Based on the multi-dimensional vector matrix discrete cosine transform (MVM-DCT), this paper presents a global and accurate motion estimation in multi-dimensional transform domain in order to solve the problem that the precision of the block matching motion estimation algorithm is discrete and the block matching is a local optimum .The algorithm breaks the traditional spatial and temporal limits, and adopts the dimension reduction and the linear iterative fitting with dynamic intelligent windows. The algorithm ensures the accuracy continuity and global optimality of the motion estimation. Experiment results verify the correctness of the theory. The accuracy can reach the magnitude of 10^{-4} and the speed of the convergence of linear iterative fitting is very fast.

CENet2015
12-13 September 2015
Shanghai, China

[1] Speaker
[2] Correspongding Author

1. Introduction

With the rapid development of internet and mobile network, multimedia services, especially the video service, account for a growing proportion of the network data. Although the network bandwidth is increasing, it still can’t meet human’s pursuit for high quality video. The popularity of high-definition television and the invention of 4k ultra high-definition television make the emergence of a higher compression efficiency standard to be imminent. The video technology experts all over the world have worked together to develop a video standard called H.265 from the beginning of 2010, which saved a half of rate than H.264 because of its advanced compression technology such as flexible division of data[1], more sophisticated intra frame prediction, sub-pixel interpolation based on DCT and SAO[2], etc.; however, the motion estimation section of video coding standards including H.265 is still using the traditional block-matching motion estimation algorithm. The bottleneck of block-matching motion estimation algorithm is of discrete accuracy, the luminance motion vector estimation accuracy is of quarter-pixel in H.264 and H.265. H.265 adopts the seven-tap filter to interpolate at the position of quarter-pixel and the eight-tap filter to interpolate at half-pixel. If we want to further improve the accuracy, the computation complexity will grow exponentially. Although there are a lot of papers which have proposed new fast algorithms, they just only improved the search algorithms of the best matching block with the discrete accuracy yet not solved. Nikola Boinovié and Janusz Konrad proposed a new algorithm by using the three-dimensional discrete cosine transform for multi-frame image and estimated the motion vector by detecting the plane that the coefficients occupied. Although the new algorithm solved the discrete accuracy, it had large amount of calculations[3].

In order to solve the traditional temporal and the spatial motion vector estimation discrete accuracy as well as the large computation of detecting plane, this paper proposes the global motion vector estimation in multi-dimensional transform domain which based on MVM-DCT. Firstly, the video data will be multi-dimensionally blocked and reorganized, next MVM-DCT; then the dimension reduction and the dynamic intelligent windows linear iterative fitting method is used to calculate the slope of the line; finally, we discuss the accuracy and continuity of the motion vector estimation. This algorithm based on the motion vector estimation can effectively solve the estimation discrete accuracy and greatly improve the detection speed by reducing the dimension to translate the plane detection into the linear detection.

The structure of this paper is shown as follows: firstly, we introduced the motion estimation in the current advanced algorithms; secondly, the theoretical basis such as the multi-dimensional vector matrix (MVM) is introduced; thirdly, our algorithm is introduced with detailed description; finally, the experiment results and analysis are given.

2 Multi-dimensional Vector Matrix DCT Transform Theory

2.1 Multi-dimensional Vector Matrix

Definition: if the dimension of the multi-dimensional matrix is divided into two groups, each of which has two vectors for representation, such as $M_{K1\times K2\times\cdots\times Kr}$, expressed as $M_{(I1\times I2\times\cdots\times Im)\times(J1\times J2\times\cdots\times Jn)}$, denoted by $\mathbf{M}_{\mathbf{IJ}}$. Obviously a multi-dimensional matrix can be expressed as a variety of multi-dimensional vector matrices, but a multi-dimensional vector matrix corresponds to only one multi-dimensional matrix.

2.2 Definition of 2M-dimensional Vector Orthogonal Transformation Kernel Matrix

The orthogonal transform and inverse transform formulas of 2M-dimensional vector matrices are:

$$F_{IJ} = C_{II} f_{IJ} C_{JJ}^{T} \tag{2.1}$$

$$f_{IJ} = C_{II}^{T} F_{IJ} C_{JJ} \tag{2.2}$$

2M-dimensional vector DCT operator is also known as 2M-dimensional vector orthogonal transform DCT kernel matrix with the specific form shown as:

$$C_{IJ} = (c_{u_1u_2\cdots u_M v_1v_2\cdots v_M}) \tag{2.3}$$

Where $\mathbf{I} = (N_1, N_2, \cdots, N_M)$, $\mathbf{J} = (N_1, N_2, \cdots, N_M)$.

$$c_{u_1u_2\cdots u_M v_1v_2\cdots v_M} = \left(\frac{2^M}{N_1N_2\cdots N_M}\right)^{\frac{1}{2}} c(u_1)c(u_2)\cdots c(u_M) \cos\frac{(2v_1+1)u_1\pi}{2N_1}\cos\frac{(2v_2+1)u_2\pi}{2N_2}\cdots\cos\frac{(2v_M+1)u_M\pi}{2N_M} \tag{2.4}$$

Where $c(u_i) = \begin{cases}\frac{1}{\sqrt{2}} & u_i=0\\ 1 & u_i=other\end{cases}$, $u_i = 0,1\cdots N_i - 1$, $v_i = 0,1\cdots N_i - 1$ $M, N_i \in N^*$ $i = 1,2,\cdots M$

$$c_{v_1v_2\cdots v_M u_1u_2\cdots u_M} = \left(\frac{2^M}{N_1N_2\cdots N_M}\right)^{\frac{1}{2}} c(v_1)c(v_2)\cdots c(v_M) \cos\frac{(2u_1+1)v_1\pi}{2N_1}\cos\frac{(2u_2+1)v_2\pi}{2N_2}\cdots\cos\frac{(2u_M+1)v_M\pi}{2N_M} \tag{2.5}$$

Where $c(v_i) = \begin{cases}\frac{1}{\sqrt{2}} & v_i=0\\ 1 & v_i=other\end{cases}$, $u_i = 0,1\cdots N_i - 1$, $v_i = 0,1\cdots N_i - 1$ $M, N_i \in N^*$ $i = 1,2,\cdots M$

3. Global Motion Vector Estimation Theory

Suppose $u_0(n_1, n_2)$ is the brightness function of spatial coordinate (n_1, n_2). Similarly, $u(n_1, n_2, n_3)$ is the brightness function of the time varying image, where n_3 refers to the time coordinate. When all spatial coordinate of brightness function of uniform translation have the constant speed $[d_1, d_2]$, they will generate a time varying image function:

$$u(n_1, n_2, n_3) = u_0(n_1 - d_1 \bullet n_3, n_2 - d_2 \bullet n_3) \tag{3.1}$$

The image sequence by Equation (2.5) will be subjected to MVM-DCT. According to the symmetry of DCT, the spectrum is limited to the folded plane of $k_1d_1 + k_2d_2 + k_3 = 0$.

$$u[k_1, k_2, k_3] = \sqrt{(2-\delta[k_1])(2-\delta[k_2])(2-\delta[k_3])}/\sqrt{\mathrm{N}^3} * 0.25 \; * \; ((P_1 + P_2)(u_0[k_1/2, k_2/2]) + (P_3 + P_4)(u_0[k_1/2, -k_2/2])) \tag{3.2}$$

Where $0 \le k_1, k_2, k_3 \le N-1$ and P_1 P_2 P_3 P_4 are:

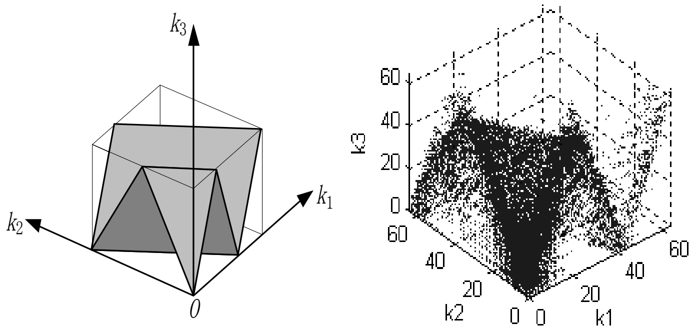
$$P_1 = \delta[(k_1d_1 + k_2d_2 + k_3)/2] * \cos(\phi_0[k_1/2, k_2/2] - \pi(k_1 + k_2 + k_3)/2\mathrm{N}) \tag{3.3}$$

$$P_2 = \delta[(k_1d_1 + k_2d_2 - k_3)/2] * \cos(\phi_0[k_1/2, k_2/2] - \pi(k_1 + k_2 - k_3)/2\mathrm{N}) \tag{3.4}$$

$$P_3 = \delta[(k_1d_1 - k_2d_2 - k_3)/2] * \cos(\phi_0[k_1/2, -k_2/2] - \pi(k_1 - k_2 - k_3)/2\mathrm{N}) \tag{3.5}$$

$$P_4 = \delta[(k_1d_1 - k_2d_2 + k_3)/2] * \cos(\phi_0[k_1/2, -k_2/2] - \pi(k_1 - k_2 + k_3)/2\mathrm{N}) \tag{3.6}$$

From Equation (3.1), the transformed coefficients are concentrated in a folded plane. The folded plane consists of four planes including P_1 P_2 P_3 P_4, which are perpendicular to the direction of motion vector. The plane is always perpendicular and mirror to the prior plane connected to it; thus the motion vector estimation by detecting either of the planes has the same effect. An experiment is carried out herein: $u_0[n_1,n_2]$ is a still an image, translated between each two successive locations of n_3, then generating an image sequence and applying multi-dimensional vector matrix DCT. Fig. 1 shows the converted coefficient distribution of the sequence, (a) is the distribution of an ideal condition, and (b) is the distribution of real condition.



(a) Distribution of Ideal Condition (b) Distribution of Real Condition

Figure1 : Coefficients Distribution of 64*64*64 Image Sequence

4. Experiment Results and Ananlysis

The special energy footprint in frequency domain of MVM-DCT coefficients has been proved in reference[4], that is to say, the data of large coefficients are concentrated around a folded plane. As the plane is always passing the origin of the frequency domain and is orthogonal with the direction of motion vector, so by detecting and locating the plane, we can determine four possible conditions of the best motion vector $[\pm d_1, \pm d_2]$.

As the transformed coefficients are concentrated around a folded plane of $k_1 d_1 + k_2 d_2 + k_3 = 0$, a new method is proposed in this paper. In the method, firstly we extract the data of $k_1 = 0$, so $k_1 d_1 + k_2 d_2 + k_3 = 0$ is converted to $k_2 d_2 + k_3 = 0$. The whole scheme is showed in Fig. 2, and the part of linear iterative fitting is in Fig. 3.

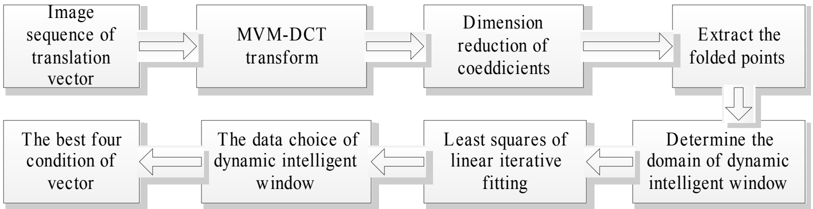


Image sequence of translation vector
MVM-DCT transform
Dimension reduction of coeddicients
Extract the folded points
Determine the domain of dynamic intelligent window
Least squares of linear iterative fitting
The data choice of dynamic intelligent window
The best four condition of vector

Figure 2 : Whole Scheme of Multi-dimensional Transform Domain Motion Vector

In order to adapt to the current HDTV and UHDTV, the biggest coding blockings in H.265 is 64*64 which was 16*16 in H.264; thus we use 64*64*64 image sequence as the video source in the experiment[5-6]. A bmp gray image will be translated based on $[d_1, d_2]$, then MVM-DCT is used to the translated image sequence. Here is the example in respect of the translated vector [3, 2]:

Step 1: a race.bmp gray image was translated 63 times based on vector [3, 2], we got a 64 frames image sequence, then every frame was sub-blocked into 64*64 blockings, reconstructed the pixel of every corresponding frame, finally we got the 64*64*64 image sequence.

Step 2: the MVM-DCT was used to the reconstructed image sequence.

Step 3: dimension reduction of coefficients. As shown in Fig. 1 (b), according to the 3d distribution of transformed coefficients, we extracted the two sections of $k_1 = 0$ and $k_2 = 0$ respectively, results were in Fig. 4.

Step 4: determination of window domain. We distinguished several folded points and determined the window domain by the folded points because details of the movement corresponding to the high frequency were the most obvious while the object was moving, so we usually chose the last section or the last second section data for the linear fitting.

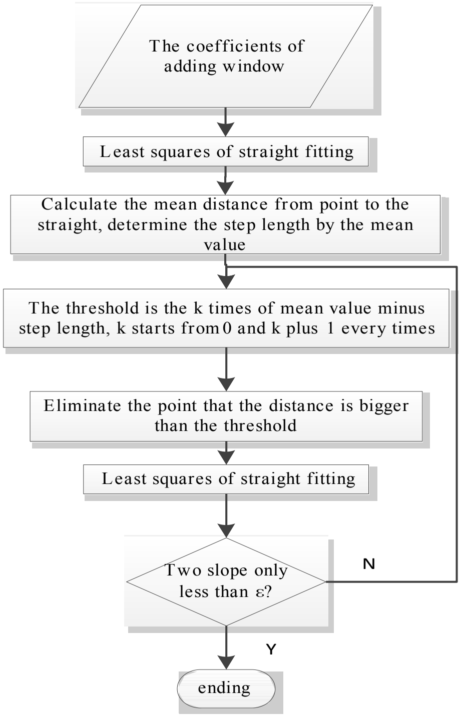
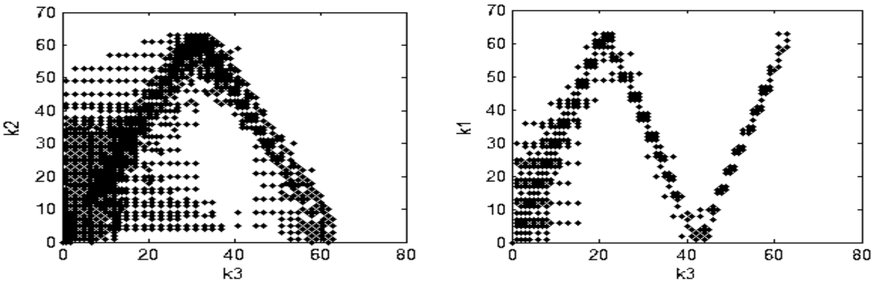

The coefficients of adding window
Least squares of straight fitting
Calculate the mean distance from point to the straight, determine the step length by the mean value
The threshold is the k times of mean value minus step length, k starts from 0 and k plus 1 every times
Eliminate the point that the distance is bigger than the threshold
Least squares of straight fitting
Two slope only less than ε?
N
Y
ending

Figure 3 : Scheme of Linear Iterative Fitting

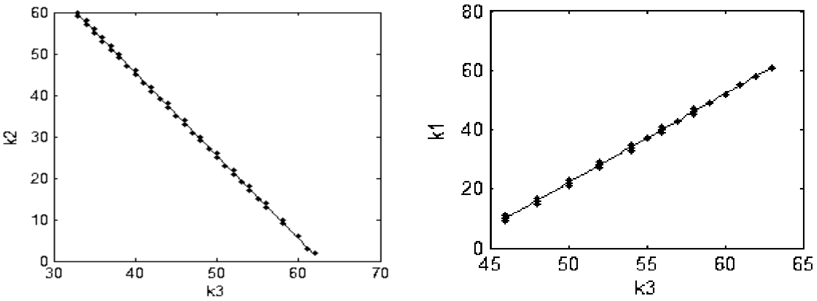


(a) data of $k_1 = 0$ plane

(b) data of $k_2 = 0$ plane

Figure 4: Extracted Data of Two Planes

Step 5: the window least squares linear iterative fitting. According to the original data in the window, we fit a straight line, then computed the mean value of straight line distance of all points as the threshold, decreased it with certain step length and iterative fitting until it converged to a straight line so that the slope of the line was the absolute value of motion vector, so did the iterative fitting to the points in another plane. Fig. 5 was the final fitting straight line which $k_1 = 0$ and $k_2 = 0$ in translated vector [3,2] and the absolute value of motion vector was listed in Table 1.



(a) plane fitting line when $k_2 = 0$ (b) plane fitting line when $k_1 = 0$

Figure 5: Final Fitting Line of Two Planes

In order to estimate the accuracy of motion vector estimation and the performance of dynamic intelligent window in this paper, lots of experiments has been done. After calculating d_1 and d_2, according to the above conclusion, the motion vector should be one of $[\pm d_1, \pm d_2]$; then we take the absolute value of the experiment results in order to prove high accuracy of this method. Based on the above method, the image sequence which translated **.bmp gray image was the video source. Comparison of the theoretical valueand the experimental value was shown in Table 1 and times of the least square linear iterative fitting were listed in Table 2.

Translated vector	[1,2]	[1,3]	[2,2]	[3,2]
race(pixel)	[0.9998,2.0000]	[1.0000,3.0002]	[2.0000,2.0000]	[3.0000,1.9994]
lena(pixel)	[1.0000,2.0000]	[0.9999,2.9998]	[1.9995,2.0001]	[3.0005,2.0000]

Table1 : Results of Dynamic Intelligent Window Linear Fitting Motion Vector Estimation

Translated vector	[1,2]	[1,3]	[2,2]	[3,2]
race	[5,9]	[8,8]	[8,4]	[7,6]
lena	[7,4]	[6,6]	[7,8]	[6,5]

Table2 : Results of Dynamic Intelligent Window Linear Fitting Motion Vector Estimation

According to Table 1, the motion vector estimation precision was continuous. As the dynamic intelligent window was dynamically determining the domain by the folded points and capturing the data after determination of the window domain, the precision of error can be reduced to 10^{-4} without changing the data. According to Table 2, we can get the conclusion that the rate of convergence of dynamic intelligent window linear iterative fitting was very fast; in addition, the motion vector estimation as proposed in this paper utilized the 2^m image information to perform the motion vector estimation, which made this method more precious and improved the vector prediction speed at the same time.

5. Conclusion

In order to solve the motion vector precision discrete of the traditional spatial domain and the computational complexity of finding the local optimal matching block, the high precision of global motion estimation in multi-dimensional transform domain has been put forward on the basis of the special energy plane of MVM-DCT transformed coefficients so as to estimate the motion vector, reduce the dimension in the original plane detection, and use the dynamic intelligent window and the least squares iterative fitting. Experiment results not only show the motion vector estimation precision is continuous but also the precision of error can be reduced to 10^{-4} while this method is more precious than the traditional motion vector estimation algorithm.

The method primarily aims at providing a new direction for motion vector estimation. Although the research is at the preliminary stage and the effect is good, there is still much for improvement. In the following study, the motion vector as calculated by this algorithm will be used to localize the energy concentrated plane and the corresponding coefficient scanning, quantification, entropy coding to achieve more efficient video compression coding.

References

[1] Y. F. Shen. *High efficiency video coding*[J].Chinese journal of computers, 11(36),2340-2355(2013)

[2] G. J. Sullivan, J. R. Ohm, W. J. Han,T.Wiegand. *Overview of the High Efficiency Video Coding(HEVC) Standard* [J].IEEE Transaction on Circuits and Systems for Video Technolgy, 22(12),1649-1668(2012)

[3] S. Mu. *Research on Multi-view Coding Based on Multidimensional Vector Matrix*[D].Jilin:Jilin University(2012) ()

[4] A. V. Paramkusam,V. S. K. Reddy, *Two-layer motion estimation algorithm for video coding*[J]. *Electronics Letters*,50(4),()276-278(2014)(2014)

[5] A. J. Sang, M. S. Chen, H. X. Chen. *Multi-dimensional vector matrix theory and its application in color image coding*[J].Imaging Science Journal,58(3),171-176(2010)

[6] N. Boinović, J. Konrad. *Motion analysis in 3D-DCT domain and its application to video coding*[J]. Signal Processing: Image Communication.20(6),510-528(2005)