

## 多维矢量矩阵变换域的运动矢量估计

桑爱军\*, 孙敏, 陈绵书, 钟江江, 陈贺新  
(吉林大学通信工程学院, 吉林 长春 130022)

**摘要:**为了实现对视频中运动目标的运动矢量估计,建立了多维矢量矩阵变换域运动估计系统,在变换域对运动视频中运动目标形成的多维能量集中平面进行研究。首先,介绍了多维矢量矩阵理论、变换理论以及运动目标在变换域形成能量集中平面的理论推导;然后,采用平面拟合的方法,求取运动矢量的大小;最后,分析对比了几种方法的效果和迭代速度。实验结果表明,多维矢量矩阵变换域的运动平面拟合方法估计的运动矢量精度达到  $10^{-2}$  pixel,提供了一种变换域运动矢量估计的高精度方法。

**关键词:**视频编码;多维矢量矩阵;运动估计;迭代拟合

中图分类号:TP941.1 文献标识码:A doi:10.3788/CO.20150802.0191

## Motion vector estimation in multi-dimensional vector matrix transform domain

SANG Ai-jun\*, SUN Min, CHEN Mian-shu, ZHONG Jiang-jiang, CHEN He-xin  
(College of Communication Engineering, Jilin University, Changchun 130022, China)  
\* Corresponding author, E-mail: sangaj@jlu.edu.cn

**Abstract:** To realize the motion vector estimation targeted at moving objects in video, the motion estimation system based on transformation domain of multi-dimensional vector matrix is established. The research is then carried out in transformation domain to study on multi-dimensional energy concentration plane formed by moving objects in moving videos. Firstly, the theory of multi-dimensional vector matrix and transformation are introduced. Then the theoretical derivation of energy concentration plane in transformation domain formed by moving objects is presented. The next step is to calculate the motion vector using plane fitting method. Finally, the performance and iteration speed among a few methods are analyzed. Experimental results show that the motion vector accuracy obtained by moving plane fitting in multi-dimensional vector matrix transformation domain can reach  $10^{-2}$  pixel. A new high accuracy motion vector estimation method in transformation domain is provided.

**Key words:** video coding; multi-dimensional vector matrix; motion estimation; iterative fitting

收稿日期:2014-12-11;修订日期:2015-02-18

基金项目:吉林省自然科学基金资助项目(No. 20130101045JC);吉林省科技发展计划国际科技合作资助项目(No. 20130413053GH);国家自然科学基金资助项目(No. 61171078)

## 1 引言

随着视频技术、多媒体通信技术的发展,各种视频信息已普及和深入到日常生产和生活的方方面面。电视会议、网络电视和智能终端等,带动了当今社会的快速发展,同时也为人类社会带来了巨大的便利。随着数字化、信息化进程的不断加深,视频数据向着高分辨率、高采样率、多角度发展,而随之产生的海量数据对视频压缩编码效率提出了更严峻的挑战<sup>[1-3]</sup>。

无论是过去的 H. 264 (AVC)<sup>[4]</sup> 还是刚建立的标准 H. 265 (HEVC)<sup>[5-6]</sup>, 运动估计技术始终是上述视频压缩标准的核心模块, 它是压缩编码的关键技术, 约占视频编码计算总量的 70% 以上, 决定了视频压缩的效果<sup>[7]</sup>。因此寻找一种快速、简便、高效的运动估计算法便成了视频编码领域的重点。传统的运动估计算法, 都是在时空域内采用块匹配的运动估计算法, 这种算法的瓶颈在于精度离散, 块匹配时间长。H. 265 和 H. 264 中亮度运动矢量估计精度都为 0.25 pixel, H. 265 中 0.25 pixel 位置采用七抽头的滤波器进行插值, 半像素位置采用八抽头的滤波器进行插值, 如果想进一步提高其精度, 运算量将会成倍增加。虽然后来有很多文献都提出了新的快速算法, 比如三步搜索、新三步搜索、四步搜索、菱形搜索、六边形搜索等各种方法, 但也只是在不断改进最佳匹配块的搜索算法<sup>[8-9]</sup>, 提高计算速度, 其精度离散这

个问题始终没有解决。

根据全局恒速平移运动所生成的人工视频序列在其变换域表现出的频谱特性可以被利用来进行视频序列运动估计方法的设计。本文将引入多维矢量矩阵的变换体系内, 推导出了视频序列经过多维矢量矩阵的离散余弦变换后, 会在频域内生成特殊的能量痕迹, 即频谱被限制在一个平面上, 并且该平面指示了运动方向。本文提出了基于多维矢量矩阵的离散余弦变换的运动矢量估计, 它不像传统运动估计采用块匹配的方法, 而是提出了一种新的运动估计的方法-迭代拟合。运用该方法求出频谱占据的平面, 然后便可求出对应的运动矢量。

## 2 MVM-DCT 域的变换理论

### 2.1 多维矢量矩阵理论

定义: 如果将多维矩阵的维数分成两组, 分别用 2 个矢量表示, 比如  $\mathbf{M}_{K_1 \times K_2 \times \dots \times K_r}$  表示成  $\mathbf{M}_{(I_1 \times I_2 \times \dots \times I_m) \times (J_1 \times J_2 \times \dots \times J_n)}$ , 记为  $\mathbf{M}_{IJ}$ , 其中  $I, J$  为矢量, 则称多维矩阵  $\mathbf{M}$  为维数按照矢量  $I, J$  划分的多维矢量矩阵, 简称多维矢量矩阵。

### 2.2 2M 维矢量正交变换核矩阵的定义

2M 维矢量 DCT 操作算子又称为 2M 维矢量 DCT 正交变换核矩阵<sup>[10-13]</sup>, 其具体形式如下:

$$\mathbf{C}_{IJ} = (\mathbf{C}_{u_1 u_2 \dots u_M v_1 v_2 \dots v_M})$$

式中  $I = (N_1, N_2, \dots, N_M)$ ,  $J = (N_1, N_2, \dots, N_M)$

$$\mathbf{C}_{u_1 u_2 \dots u_M v_1 v_2 \dots v_M} = \left( \frac{2^M}{N_1 N_2 \dots N_M} \right)^{\frac{1}{2}} C(u_1) C(u_2) \dots C(u_M) \cdot \cos \frac{(2v_1 + 1)u_1 \pi}{2N_1} \cos \frac{(2v_2 + 1)u_2 \pi}{2N_2} \dots \cos \frac{(2v_M + 1)u_M \pi}{2N_M}, \quad (1)$$

式中  $C(u_i) = \begin{cases} \frac{1}{\sqrt{2}} & u_i = 0 \\ 1 & u_i = \text{other} \end{cases}$ ,  $u_i = 0, 1, \dots, N_i - 1$ ;  $v_i = 0, 1, \dots, N_i - 1$ ;  $M, N_i \in N^*$ ;  $i = 1, 2, \dots, M$ 。

$$\mathbf{C}_{v_1 v_2 \dots v_M u_1 u_2 \dots u_M} = \left( \frac{2^M}{N_1 N_2 \dots N_M} \right)^{\frac{1}{2}} C(v_1) C(v_2) \dots C(v_M) \cdot \cos \frac{(2u_1 + 1)v_1 \pi}{2N_1} \cos \frac{(2u_2 + 1)v_2 \pi}{2N_2} \dots \cos \frac{(2u_M + 1)v_M \pi}{2N_M}, \quad (2)$$

式中  $C(v_i) = \begin{cases} \frac{1}{\sqrt{2}} & v_i = 0 \\ 1 & v_i = \text{other} \end{cases}$ ;  $v_i = 0, 1, \dots, N_i - 1$ ;  $u_i = 0, 1, \dots, N_i - 1$ ;  $M, N_i \in N^*$ ,  $i = 1, 2, \dots, M$ 。

下面是一个例子, 表示  $2 \times 2 \times 2 \times 2$  的变换核矩阵。

### 2.3 MVM-DCT 域的能量集中性

根据 FT 与 DCT 的关系 (DCT 是 FT 在实数

$$A_{2 \times 2 \times 2} = \begin{matrix} \begin{matrix} 1 & 1 \\ 1 & -1 \end{matrix} & \begin{matrix} 1 & -1 \\ 1 & -1 \end{matrix} \\ \begin{matrix} 1 & -1 \\ 1 & -1 \end{matrix} & \begin{matrix} 1 & -1 \\ 1 & -1 \end{matrix} \end{matrix} \quad (3)$$

域的对称扩展形式),以及 DCT 的对称性,可以推导出多维矢量矩阵的离散余弦变换同样具有能量频谱折叠特性<sup>[14]</sup>。构造一个宽为  $n_1$ ,高为  $n_2$ ,帧数为  $n_3$ ,每两个连续帧之间的位移为  $(d_1, d_2)^T$  的视频序列

$$u(n_1, n_2, n_3) = u_0(n_1 - d_1 n_3, n_2 - d_2 n_3)$$

将上述公式表示成卷积形式:

$$u_0(n_1 - d_1 n_3, n_2 - d_2 n_3) = [u_0(n_1, n_2)\delta(n_3)] * \delta(n_1 - d_1 n_3, n_2 - d_2 n_3), \quad (4)$$

设  $N_1 = N_2 = N_3 = N$ ,用  $k_1, k_2, k_3$  取代 FT 域中的  $f_1, f_2, f_3$ ,用 MDCT[·]表示输入函数的多维矢量矩阵的离散余弦变换,对式(4)卷积的第一部分应用多维矢量矩阵的离散余弦变换,化简后得:

$$\begin{aligned} \text{MDCT}[u_0(n_1, n_2)\delta(n_3)] &= \left[ \sqrt{\frac{2}{N}} \sqrt{\frac{2}{N}} \mathbf{C}(k_1) \mathbf{C}(k_2) \cdot \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} u(n_1, n_2) \cos \frac{\pi(2n_1 + 1)k_1}{2N} \cdot \cos \frac{\pi(2n_2 + 1)k_2}{2N} \right] \cdot \left[ \sqrt{\frac{2}{N}} \mathbf{C}(k_3) \cdot \sum_{n_3=0}^{N-1} \delta(n_3) \cos \frac{\pi(2n_3 + 1)k_3}{2N} \right], \quad (5) \end{aligned}$$

对式(4)卷积的第二部分应用多维矢量矩阵的离散余弦变换,化简后得:

$$\begin{aligned} \text{MDCT}[\delta(n_1 - d_1 n_3, n_2 - d_2 n_3)] &= \frac{2}{N} \cdot \sqrt{\frac{2}{N}} \mathbf{C}(k_1) \mathbf{C}(k_2) \mathbf{C}(k_3) \cdot \sum_{n_3=0}^{N-1} \left[ \cos \frac{\pi(2d_1 n_3 + 1)k_1}{2N} \cdot \cos \frac{\pi(2d_2 n_3 + 1)k_2}{2N} \cdot \cos \frac{\pi(2n_3 + 1)k_3}{2N} \right]. \quad (6) \end{aligned}$$

以上的式子表明均匀平移图像的频谱是受限于折叠平面  $k_1 d_1 + k_2 d_2 + k_3 = 0$ ,在视频变换后的

系数分布图中,系数是沿轴折叠,并衰减。变换后的系数集中分布在一个折叠平面上,这就为后面的实验提供了理论指导。

### 3 MVM-DCT 域的运动估计

#### 3.1 MVM-DCT 域变换系统

根据上述多维矢量矩阵正交变换的性质,本文设计了全新的 MVM-DCT 域的运动矢量估计算法,对于原始的 YUV 格式视频,先分别对各个分量数据进行分块,得到多维矢量矩阵,然后按照多维矢量矩阵正交变换公式进行变换,得到各个分量的三维变换系数。由于 Y 分量含有主要运动信息,下述的运动估计都是在 Y 分量上进行,U、V 分量算法同理。算法框图如 1 图所示。

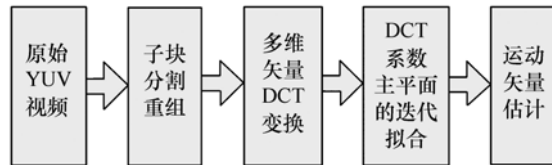


图 1 运动估计的框图

Fig.1 Block diagram of MVM-DCT

#### 3.2 新的运动估计方法——迭代拟合

图 2 为一幅静止图像,定义为  $u_0(n_1, n_2)$ ,在每两个连续的位置  $n_3$  上经过运动矢量  $(d_1, d_2)$  均匀平移得到的图像序列的前 8 帧;图 3 表示对得到的均匀平移图像序列应用多维矢量矩阵的离散

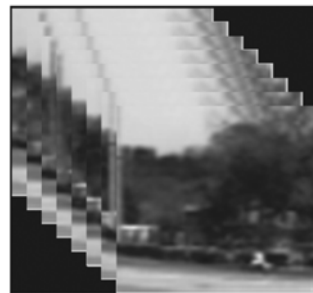


图 2 均匀平移图像序列的前 8 帧

Fig.2 First 8 frames of image sequence formed by uniform translation

余弦变换后系数的分布图。从图 3 中可以看出,变换后的系数能量集中分布在一个折叠平面上。

本文就是通过对折叠平面进行拟合,然后逐步运用迭代的方法,就可以估计出图 2 中静止图像  $u_0(n_1, n_2)$  在每两个连续的位置  $n_3$  上移动的运动矢量  $(d_1, d_2)$  的值,进而来提高编码效率。

视频序列经过多维矢量矩阵的离散余弦变换后,变换后的系数是集中分布在一个折叠平面上的,如图 3 所示。

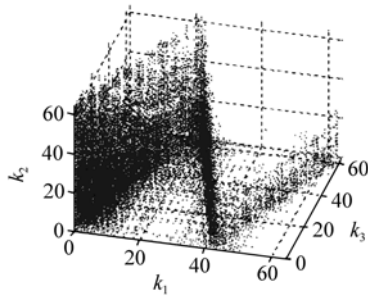


图 3 变换后系数

Fig. 3 Coefficients after transformation

因此本文提出了新的运动估计的方法—

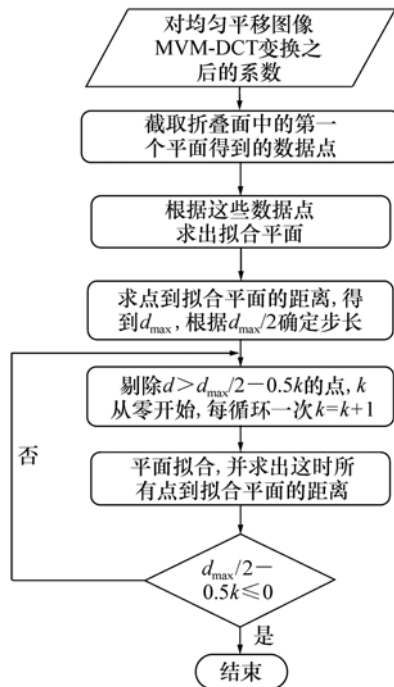


图 4 迭代拟合的流程图

Fig. 4 Iterative fitting flowchart

MVM-DCT 运动矢量估计,其平面迭代拟合的流程图如图 4 所示。

### 4 实验结果

运动矢量  $(3,0)$  和运动矢量  $(2,0)$  的折叠平面如图 5 所示。

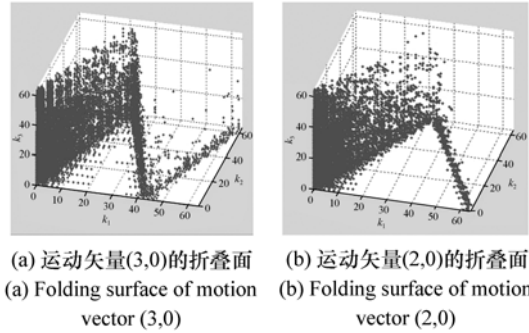


图 5 两种运动矢量的折叠面

Fig. 5 Folding surface of two kinds of motion vector

已知在理想情况下,对一个视频序列进行平移,如果让平移的运动矢量为  $[d_1, d_2]^T = [3, 0]^T$  和  $[d_1, d_2]^T = [2, 0]^T$ ,然后再对视频序列进行多维矢量离散余弦变换,就会得到系数集中分布在折叠平面上,而这个折叠平面就包含了运动信息。本文就是通过仿真实验,对折叠之前的平面进行拟合,然后再通过迭代的方法,进而求出它的运动矢量  $(d_1, d_2)$ 。以运动矢量  $(3,0)$  为例,迭代拟合的方法如下文所述。

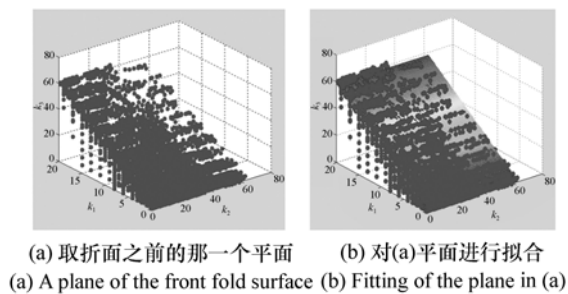


图 6 平面拟合效果图

Fig. 6 Renderings of plane fitting

图 6(a) 是图 5(a) 取折面之前的那一个平面。图 5(a) 的折叠平面包含了运动信息,由于折面之前的那一个平面包含大部分的点数,所以我们只对折面之前的平面进行拟合,然后再运用迭代的方法,就可以求出运动矢量  $(d_1, d_2)$ 。通过对

图 6(a) 进行拟合,得到了(b)图,即拟合平面是:

$$z = -0.0067x + 2.8481y. \quad (6)$$

在图 6(b)中,可以看到,有的点严重偏离了拟合平面,这样会使拟合平面误差比较大。本文则采用了迭代拟合的方法,将严重偏离拟合平面的点剔除,使得到的拟合平面逐渐趋向于一个固定的值。

本文所使用的迭代拟合的方法为:

(1) 首先求出图 6(b)中所有点到式(5)所求拟合平面的距离,将超过某一距离值的点进行剔除,利用剩下的点重新拟合出一个新的平面:

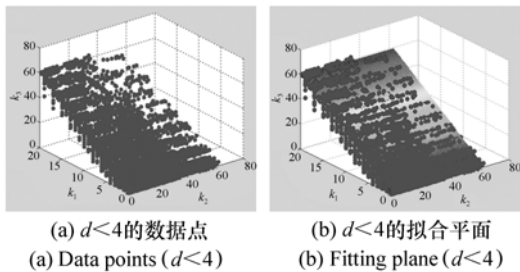


图 7 剔除数据后平面拟合效果图

Fig. 7 Renderings of plane fitting after excluding data

图 7(a)是图 7(b)中所有点到式(6)所求拟合平面的距离小于 4 的点,图 7(b)是将图 6(a)中的所有点重新进行拟合。在图 6(b)中,可以求出所有点到拟合平面的距离,最大的距离是  $d_{\max} = 8$ ,这时取了  $d < d_{\max}/2$  ( $d_{\max}/2 = 4$ ),是因为  $d > 4$  的点比较稀少,所以一开始本文取  $d < 4$ 。可以看到图 7(a)比图 6(a)的左下角少了一些点,也就是能量更集中了。在图 7(b)中,将剩下的所有点再进行拟合,会得到一个新的拟合平面是:

$$z = 0.0001x + 2.9160y. \quad (7)$$

将式(6)与式(7)作比较,可以看到精确度有所提高。

(2) 接着求出图 7(b)中所有点到式(6)所求拟合平面的距离,本文取的步长  $d_1 = 0.5$ ,所以剔除距离  $d > 3.5$  的点,再对剩下的点重新进行拟合,便得到了图 8(a)。然后利用迭代的方法,再继续剔除点到图 8(a)中拟合平面距离  $d > 3$  的,取步长  $d_1 = 0.5$  时,以此类推便分别得到  $d < 2.5, d < 2, d < 1.5, d < 1, d < 0.5$  的拟合平面,部

分拟合结果如图 7 所示。

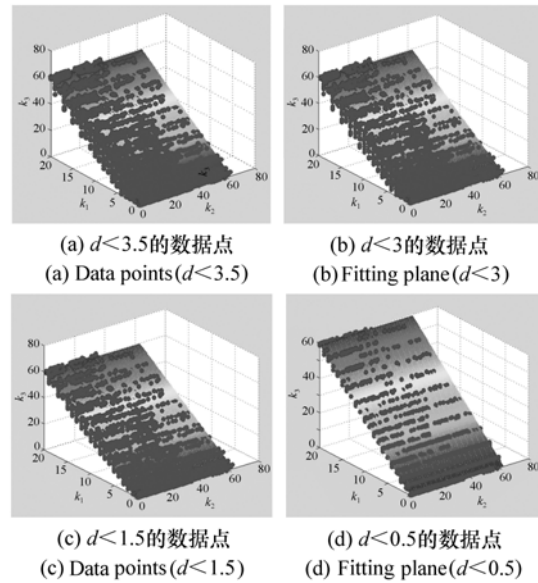


图 8 进一步剔除数据后平面拟合效果图

Fig. 8 Renderings of plane fitting after further excluding data

(3) 为了验证所提出的迭代拟合方法的可行性,本文也对矢量  $(2, 0)$  做了迭代拟合。此次试验理论结果是  $z = 0 \times x + 3 \times y$  和  $z = 0 \times x + 2 \times y$ 。试验结果如图 9 所示,纵坐标是迭代拟合结果误差值,横坐标是迭代拟合的次数,按照上述方法迭代拟合 9 次,直到剔除距离  $d > 0.5$  的点时,拟合平面达到了固定精度。

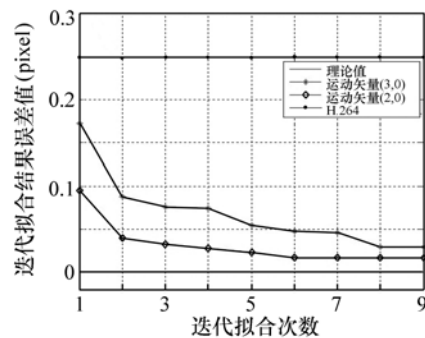


图 9 迭代拟合试验结果

Fig. 9 Iterative fitting test results

由图 9 可知,国际标准的容许误差是 0.25 pixel,本文在第一次迭代,误差就小于国际标准,并且随着迭代次数的增加,逐渐逼近零误差。

差,计算精度远高于国际标准。表1是图9的具体数据。

表1 迭代拟合试验结果  
Tab.1 Iterative fitting test results

精度(pixel)	1	2	3	4	9
运动矢量(2,0)	0.17	0.09	0.08	0.07	0.03
运动矢量(3,0)	0.10	0.04	0.03	0.02	0.02
H.264 国际标准			0.25		

## 5 结论

本文利用理想情况下视频序列 FT 域的频谱特性和离散余弦变换域的特性,获得理想的运动矢量理论值。重点研究了多维矢量矩阵的离散余弦变换域的运动估计,提出了一种新的运动估计

的方法—MVM-DCT 域的运动矢量估计,该运动估计运用了迭代拟合的方法。实验验证了这种方法的正确性和可行性,它不需要进行块匹配,可大大减少计算复杂度,而且由于其连续性,相比 H.265 标准中的 0.25 pixel 精度,采用本文方法获得的精度达到 0.01 pixel,得到了明显提高。

## 参考文献:

- [1] 朱明,杨航,贺柏根,等. 联合梯度预测与导引滤波的图像运动模糊复原[J]. 中国光学,2013,12(6):850-854.  
ZHU M, YANG H, HE B G, *et al.*. Image motion blurring restoration of joint gradient prediction and guided filter[J]. *Chinese Optics*, 2013, 12(6):850-854. (in Chinese)
- [2] 许廷发,李俊涛,张一舟,等. 真彩色传递双波段图像融合[J]. 中国光学,2014,6(7):402-810.  
XU T F, LI J T, ZHANG Y ZH, *et al.*. True color transfer for dual band image fusion[J]. *Chinese Optics*, 2014, 6(7):402-810. (in Chinese)
- [3] 郭立强,朱明. 彩色图像处理的可交换 Clifford 代数方法[J]. 中国光学,2013,12(6):884-891.  
GUO L Q, ZHU M. Commutative Clifford algebra method for color image processing[J]. *Chinese Optics*, 2013, 12(6):884-891. (in Chinese)
- [4] 毕厚杰. 新一代视频压缩编码标准——H.264/AVC[M]. 北京:人民邮电出版社,2005.  
BI H J. *A New Generation of Video Coding Standard——H.264/AVC*[M]. Beijing:People Post Press. 2005. (in Chinese)
- [5] BENJAMIN B, WOO-JIN H, JENS-RAINER O, *et al.*. High Efficiency Video Coding(HEVC) text specification draft 10 [C], JCT-VC-L1003 of JCT-VC Geneva, CH, Jan., 2013.
- [6] 沈燕飞. 高效视频编码[J]. 计算机学报,2013,11(36):2340-2355.  
HEN Y F. High efficiency video coding[J]. *Chinese J. Computers*, 2013, 11(36):2340-2355. (in Chinese)
- [7] 王大伟,严俊,王彬,等. 视频序列的运动估计技术综述[J]. 电视技术,2012(3):20-23.  
WANG D W, YAN J, WANG B, *et al.*. Survey of motion estimation used in video sequence[J]. , 2012(3):20-23. (in Chinese)
- [8] 刘龙. 基于运动矢量时空特性的快速运动估计算法研究[J]. 通信学报,2013,34(1):121-127.  
LIU L. Fast motion estimation based on the special and temporal characteristic[J]. *J. Communications*, 2013, 34(1):121-127. (in Chinese)
- [9] 刘英哲. H.264 中一种基于搜索范围自适应调整运动估计算法[J]. 电子与信息学报,2013,35(6):1382-1387.  
LIU Y ZH. Motion estimation algorithm based adaptive search range adjustment for H.264[J]. *J. Electronics & Information Technology*, 2013, 35(6):1382-1387. (in Chinese)
- [10] 桑爱军,王艇,栾晓利,等. 2M 维矢量余弦整数变换核矩阵[J]. 光学精密工程,2013,7(21):1891-1897.

- SANG A J, WANG T, LUAN X L, *et al.* . [J]. *Opt. Precision Eng.* ,2013,7(21):1891-1897. (in Chinese)
- [11] SANG A J, CHEN M S, CHEN H X, *et al.* . Multi-dimensional vector matrix theory and its application in color image coding[J]. *Imaging Science J.* ,2010,58(3):171-176.
- [12] 桑爱军,崔海廷,王墨林,等. 高维变换域中的系数集中分析及研究[J]. 吉林大学学报(工学版),2013,43(增刊):96-100.
- SANG A J, CUI H T, WANG M L, *et al.* . Coefficient concentration analysis and research in high-dimensional transform domain[J]. *J. Jilin University(Engineering and Technology Edition)* ,2013,43(Supp.):96-100. (in Chinese)
- [13] 桑爱军,吴杨,崔海廷,等. 基于多维矢量矩阵的离散余弦变换快速算法[J]. 光学精密工程,2013(21):799-806.
- SANG A J, WU Y, CUI H T, *et al.* . Fast DCT algorithms based on multi-dimensional vector matrix[J]. *Opt. Precision Eng.* ,2013,3(21):799-806. (in Chinese)
- [14] NIKOLA BOZINOVIC, JANUSZ KONRAD. Motion analysis in 3D DCT domain and its application to video coding, signal processing[J]. *Image Communication* ,2005,6(20):510-528.

#### 作者简介:



桑爱军(1973—),女,山东莱州人,博士,教授,硕士生导师,1994年于华中理工大学获得学士学位,1997年于大连理工大学获得硕士学位,2002年于吉林大学获得博士学位,主要从事数字图像处理等方面的研究。E-mail:sangaj@jlu.edu.cn



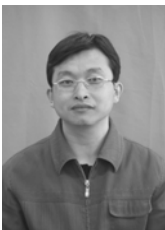
钟江江(1990—),男,湖南郴州人,2012年于长春大学获得学士学位,主要从事H.265(HEVC)运动矢量估计方面的研究。E-mail:zhongjj12@mails.jlu.edu.cn



孙敏(1988—),女,山东菏泽人,2011年于长春工业大学获得学士学位,主要从事多视角视频图像编解码方面的研究。E-mail:sunmin0920@163.com



陈贺新(1949—),男,吉林大安人,博士,教授,博士生导师,1990年于吉林大学获得博士学位,主要从事多维信号处理、图像与视频编解码、视频通信、多媒体数据库技术方面的研究。E-mail:chenhx@jlu.edu.cn



陈绵书(1973—),男,辽宁海城人,博士,副教授,2004年于吉林大学获得博士学位,主要从事图像处理、人工智能方面的研究。E-mail:chenms@jlu.edu.cn