

一种基于皇后模板抽样的全局运动估计方法

樊玉平 田裕鹏

(南京航空航天大学自动化学院,南京 210016)

摘要 在目标跟踪、视频编码等领域都需要对全局运动进行快速准确的估计。针对既有摄像机运动又有前景物体运动的视频图像序列的全局运动估计问题,提出了一种新的全局运动估计算法。该新算法先采用3层金字塔进行多分辨率计算;然后在每层运算中都采用基于块的去外点法来有效去除外点。为了提高运算速度,在金字塔下面两层中还分别采用四皇后、八皇后模板抽取背景像素,只有被抽到的像素才参与迭代计算。通过对比试验可以看到,该算法在不降低精度的情况下,大大提高了估计速度,达到了实时性要求。

关键词 全局运动估计 抽样模板 迭代算法 运动模型

中图法分类号: TN919.81

文献标识码: A

文章编号: 1006-8961(2009)12-2497-06

Global Motion Estimation Algorithm Based on Queen Patterns Subsample

FAN Yu-ping, TIAN Yu-peng

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016)

Abstract Fast and accurate estimation of global motion is necessary in target tracking, video coding and etc. Considering the global motion estimation of video with both camera motion and foreground objects motion, a new global motion estimation algorithm is proposed. In the algorithm a three-level pyramid is utilized. Then outliers are effectively eliminated with the block-based methods in the calculation of each pyramid level. In order to accelerate the algorithm, 4-Queen and 8-Queen patterns are also used to sample background pixels respectively on the intermediate level and the base level, and only the extracted pixels are contributed to the iterative calculation. The experimental results show that the algorithm improved the estimation speed greatly without reducing the accuracy and satisfied the real-time request.

Keywords global motion estimation, subsampling pattern, iterative algorithm, motion model

1 引言

在视频序列中,通常存在着以下两种运动:一种是由于摄像机位置或参数的变化而引起的整个图像的变化;另外一种是由于场景中物体的运动而引起的局部图像的变化。前者称为全局运动,后者则称为局部运动。区分视频序列中的这两种运动并进行相应处理,对于许多视频应用(如目标跟踪、视频稳定和压缩、场景分析等)都具有极为重要的意义。

对于全局运动估计,目前绝大多数算法是根据全局运动估计的基本原理,先利用密度估计的方法来得到最佳的模型参数,然后采用迭代优化的方法来估计这些参数。

在迭代计算中,由于图像背景像素数量很多,且有前景物体的局部运动干扰,致使现有的全局运动估计方法存在以下两个不足:一个是计算量大,很难满足实时性要求;另一个是对前景物体(外点)的去除不彻底,从而影响了参数估计的精度。正是由于这两个方面的不足,使它们的应用受到了很大限制。

因此国内外众多学者都致力于研究如何改进全局运动估计的算法,例如 Yosi Keller 等人提出了多种特征点选取的方法^[1],以减少计算量;Konrad 等人提出了一些去除噪声的方法^[2];贺玉文等人将特征点与去除噪声两方面相结合,提出了一种快速鲁棒的全局运动估计算法^[3],但是计算量依然比较大。本文对全局运动估计进行了深入的研究,提出了一种新的快速全局运动估计算法。该算法对现有的 Levenberg-Marquardt 算法做了相应的改进,不仅提高了全局运动估计的速度,而且还可以保证估计的精确性。

2 现有全局运动估计方法

2.1 摄像机运动模型

根据复杂性不同,图像运动模型有多种描述形式,其中 Affine 模型是一种线性模型,它能精确地描述纯旋转、相机在场景中较小深度变化时的平移和变焦运动。此模型适合绝大多数的室内和室外场景。其表达式为

$$\begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} a_3 \\ a_6 \end{bmatrix} \quad (1)$$

其中, (x_{i-1}, y_{i-1}) 和 (x_i, y_i) 为在时刻 t_{i-1} 和 t_i 时两帧图像的像素坐标, (a_1, a_2, \dots, a_6) 为图像帧间的变换参数,其中 a_3, a_6 与图像的平移运动有关, a_1, a_2, a_4 及 a_5 与图像的缩放、旋转运动有关。本文采用此模型进行全局运动估计的计算。

2.2 传统 Levenberg-Marquardt 算法

Konrad 和 Meyer 等人提出的 Levenberg-Marquardt 迭代法(简称 LM 算法)^[4]是牛顿迭代法的变形,它们都是沿梯度方向逐步下降得到最优化点所在的最低处。传统的 LM 算法是通过最小化两幅图像的亮度残差来求得全局运动参数,并要求能量函数 $E(\alpha)$ 的取值最小^[5]。能量函数 $E(\alpha)$ 的定义如下:

$$E(\alpha) = \frac{1}{n} \sum_{m=1}^n e_m^2 \quad (2)$$

其中,

$$e_m = f_i(x_m, y_m) - f_{i-1}(\tilde{x}_m, \tilde{y}_m)$$

(x_m, y_m) 是当前帧 f_i 中某个像素的坐标, $(\tilde{x}_m, \tilde{y}_m)$ 是参考帧 f_{i-1} 中像素的坐标, f_{i-1} 可以通过某种运动变换到 f_i , 本文选用 6 参数仿射模型进行变换, 如果

f_{i-1} 中的像素位置 $(\tilde{x}_m, \tilde{y}_m)$ 与 f_i 中的像素位置 (x_m, y_m) 满足式(1), 则能量函数 $E(\alpha)$ 可表示为

$$E(\alpha) = \frac{1}{n} \sum_{m=1}^n (f_i(x_m, y_m) - f_{i-1}(a_i^{(1)}x_m + a_i^{(2)}y_m + a_i^{(3)}, a_i^{(4)}x_m + a_i^{(5)}y_m + a_i^{(6)}))^2 \quad (3)$$

如果在第 k 步求得全局参数 $\alpha^{(k)}$, 那么就可以把能量函数 $E(\alpha)$ 在 $\alpha^{(k)}$ 处, 用泰勒公式展开来得到下一步的迭代, 即

$$E(\alpha) = E(\alpha^{(k)}) + d^T(\Delta\alpha) + \frac{1}{2}(\Delta\alpha)^T D(\Delta\alpha) \quad (4)$$

其中, d 是梯度向量, D 是 Hessian 矩阵, 则下一步的迭代公式为

$$\alpha^{(k+1)} = \alpha^{(k)} + \Delta\alpha = \alpha^{(k)} + H^{-1}b \quad (5)$$

其中,

$$H = \frac{1}{2}D = \sum_m \frac{\partial e_m}{\partial \alpha^{(k)}} \frac{\partial e_m}{\partial \alpha^{(k)}} \quad (6)$$

$$b = -\frac{1}{2}d = \sum_m e_m \frac{\partial e_m}{\partial \alpha^{(k)}}$$

传统的 LM 算法的具体步骤如下: 首先利用 3 层金字塔对图像进行多分辨率分解, 然后对每一层进行迭代计算。求取全局运动参数时, 先将利用三步搜索法求得的最上层的平移运动矢量作为迭代初值; 然后利用直方图方法去除外点, 而将剩余的点参与全局运动估计的迭代; 最后把求得的参数映射到下一层, 并用同样的方法求得中间层和最终的全局运动参数。

3 本文对现有 LM 算法的改进

3.1 外点去除

由于传统的 LM 算法采用基于直方图的方法去除的外点较分散, 而局部运动往往都是由前景物体运动造成的, 根据前景运动物体有聚集成块的特性, 贺玉文等人提出了基于块的去局部运动点方法^[3]。这种方法对于局部运动点的去除效果明显好于直方图法, 但文献[3]只是在金字塔的最下层使用了该方法, 上面两层同样使用直方图法。由于前景物体局部运动点的聚集性并不随图像分辨率的降低而改变, 郭丽等人提出在金字塔的每一层都使用基于块的去外点方法^[6], 其去除效果更好, 因此本文也在每一层金字塔上使用基于块的去外点方法。

下面对基于块的去外点方法进行简要介绍,对于金字塔的每一层图像,先经过运动补偿得到残差图像;然后将残差图像分成 $M \times N$ 大小的块,并计算每个块的绝对误差和 (SAD) 值, SAD 的定义为

$$SAD = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |R_{i,j}| \quad (7)$$

其中, $R_{i,j}$ 是坐标为 (i,j) 的块的残差。先将所有块的 SAD 值进行排序,并将 SAD 值大的前 30% 的块作为预选去除块,按照下面的方法确定最后的去除块。去除块时,首先假设块 $B_{(i,j)}$ 是一个预选的去除块,若它的 8 邻域中有多于 4 个预选的去除块,那么 $B_{(i,j)}$ 这个预选的去除块将被去除,并标志为去除块;然后,对于剩下的预选去除块再进行判断,如果它的 8 邻域中有去除块,那么这个预选的去除块也将被去除。所划分块的大小对于去外点所需的计算量和去除效果都有很大的影响,根据经验,对于 288×352 pixels 大小的 CIF 图像, 16×16 大小的块较为合适。

3.2 皇后模板抽样介绍

传统的 LM 算法是利用直方图先去除 10% 的外点,而剩余 90% 的背景像素点都参与迭代计算,因此计算量非常大,此后许多改进的算法被提了出来,其中抽取代表点方法用得最多。抽取代表点方法不仅可以有效地减少计算量,而且还能保证运动估计的精确性,抽取方式有随机抽样、固定模板抽样和根据梯度值抽样 3 种。其中随机抽样和固定模板抽样是依据相邻区域像素的运动具有一致性的特性,通过选取其中的一个像素来代表此区域中所有像素参与全局运动的估计。根据梯度值抽样就是先根据像素点的梯度值找出背景中的特征点,然后仅特征点参与迭代计算。随机抽样最简单,但由于抽取像素的随机性,容易造成抽取结果的不稳定,而且根据梯度值抽样本身需要相当的计算量,这在某种程度上违背了减少计算量的初衷;而固定模板抽样由于计算量小,抽样比率可控,所以用得最多。Hussein Alzoubi 等人提出了多种固定模板^[7],有梅花型模板、四皇后模板、八皇后模板及以上各个模板的两两合成模板。但是文献[7]在特定的计算中只用其中的一个模板对图像进行抽样,没有考虑金字塔各层的分辨率问题。本文在此基础上,提出了一种更合理的抽样方法,即基于皇后模板的抽样方法,由于此方法可根据金字塔各层分辨率的不同而采用不同的模板,从而能更好地兼顾全局运动估计的速度和

精度。

四皇后(4Q)和八皇后(8Q)格局都来源于国际象棋,4Q 格局是在 4×4 的棋盘上无冲突地放置 4 个皇后棋子,在国际象棋中,皇后的移动方式是横竖交叉,所以在任意一个皇后所在位置的水平、竖直和斜 45° 线上都不能有其他皇后棋子,图 1(a)是它的一种摆放格局。同样道理,8Q 是在 8×8 的棋盘上无冲突地放置 8 个皇后棋子,图 1(b)是它的一种摆放格局。

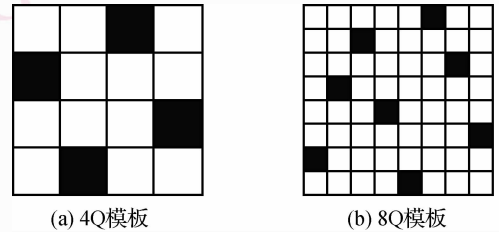


图 1 皇后模板

Fig. 1 Queen patterns

在本文提出的全局运动估计(GME)算法中,对图像金字塔的中间层使用 4Q 抽样模板进行抽样,即首先把背景区域分成 4×4 的子区域,然后对每个子区域采用 4Q 模板进行抽样,因此只有各个皇后所在位置的像素才被抽取出来参与计算。用同样的方法,对最下层图像则采用 8Q 模板进行抽样。在金字塔的最上层,由于分辨率低,相邻像素间的运动不能保证具有一致性,但为了保证计算的准确性,每个背景像素都应该参与计算,因此不对其进行抽样;在中间层,相邻像素间有一定的运动一致性,但是这种一致性并不强,适合小比率的抽样,所以本文采用 4Q 抽样模板来对背景像素进行抽样。由于金字塔最下层的分辨率最高,且相邻像素间具有较强的运动一致性,适合比率较大的抽样,因此本文使用 8Q 抽样模板进行抽样。

由于皇后棋盘的特殊格局,皇后棋子分布基本均匀,因此 4Q 和 8Q 抽样模板保证了抽样的均匀性;而本算法对金字塔中间层和最下层分别采用抽样比率不等的模板,分辨率低时,用抽样比率小的 4Q 模板,分辨率高时,使用抽样比率大的 8Q 模板,这样就保证了抽样的合理性。抽样后,由于参与迭代运算的像素数大大减少,因此使速度明显加快,但是若仅仅为了加快计算速度而一味提高抽样比率却是不可取的,因为在全局运动估计中,估计的准确性

和计算速度同样重要,不能顾此失彼,否则将没有意义。本文算法的抽样比率不高,4Q 模板抽样比率为 4:1,8Q 模板为 8:1,因此它既能提高运算速度,又不影响计算准确性。除了上述优点,与普通的隔行隔列抽样模板相比,皇后模板的优越性还在于它便于与其他的模板组合使用,使估计精度更高。

3.3 本文算法的步骤如下:

- (1)用高斯滤波器滤去图像的随机噪声,并用 $[1/4, 1/2, 1/4]$ 滤波器产生图像金字塔;
- (2)利用三步搜索方法估计金字塔顶层的平移运动,作为迭代初值;
- (3)利用基于块的方法去除外点;
- (4)对于余下的点,利用 LM 算法进行迭代计算,直到循环次数达到 32 次,或平移参数误差小于 0.001 或者其他参数误差小于 0.00001;
- (5)将计算结果映射到下一层,并对相应的参数进行更新;
- (6)利用基于块的方法去除金字塔中间层图像的外点;
- (7)对于背景区域,采用 4Q 抽样模板对其进行均匀抽样;
- (8)重复步骤(4)、(5),求得中间层的全局运动参数;
- (9)在金字塔最下层图像上,根据基于块的去局部运动点方法找出前景区域,并对其进行消除;
- (10)对于背景区域,采用 8Q 抽样模板对其进行均匀抽样;
- (11)重复步骤(4),求得最终的全局运动参数值。

4 实验结果及分析

为验证本文算法(简称 GME_4 算法)的全局运动估计效果,选用 Coastguard, Foreman, Stafen 等多个标准视频测试序列,在 PC 机上用 matlab7.0 进行全局运动估计仿真试验,并与传统的 LM 算法(GME_1)及贺玉文等人提出的算法(GME_2)以及 Hussein Alzoubi 等人所提模板中只选四皇后模板的方法(简称 GME_3_1 算法)和只选八皇后模板的方法(简称 GME_3_2 算法)进行实验对比。

Coastguard 序列主要针对摄像头有平移运动的情况。对 Coastguard 序列第 72 帧(图 2(a))、第 70 帧(图 2(b))进行仿真。表 1 显示各种算法的全局运动估计的参数、时间和补偿后图像的峰值信噪比,在此基础上,再分别用求得的全局运动参数来对当前帧进行运动补偿,并用相同的阈值进行分割来得到二值图(如图 2(c)~图 2(h)所示)。为了说明算法上的优越性,先对本文算法与 GME_1 及 GME_2 算法进行比较分析;为了说明本文提出的抽样模板的合理性,再将本文算法与 GME_3_1 及 GME_3_2 算法进行比较分析。从表 1 中可以看到,本文算法所用时间只是 GME_1 算法的 64.5%,GME_2 算法的 73.1%,而估计精度并不差。从图 2(c)可知,两帧图像之间既有前景的运动,也有背景的运动。图 2(d)~图 2(h)主要都是前景船的轮廓,背景基本上都被去除了,这说明全局运动估计是有效的。图 2(d)中显示还有一些背景没有被去除,而 GME_2 算法和本文算法的全局运动估计效果要好,图

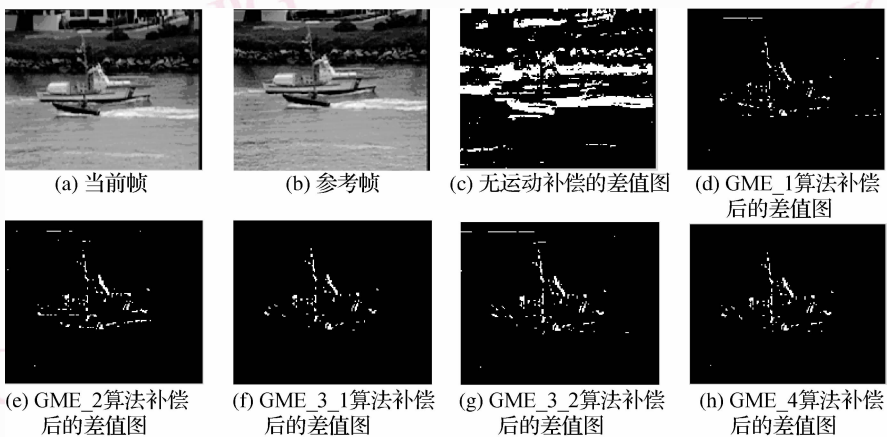


图 2 Coastguard 序列全局运动估计结果

Fig. 2 The GME result of the Coastguard sequence

2(e)和图 2(h)的运动估计效果差不多,但图 2(h)采用的 GME_4 算法的运算时间明显少于图 2(e)采用的 GME_2 算法的运算时间。从表 1 中还可以看到,四皇后模板法用的时间明显多于本文算法和八皇后算法,八皇后算法用的时间最少,但从精度上讲,八

皇后模板法的全局运动估计效果明显不如四皇后模板法和本文算法,从图 3(f)、图 3(g)和图 3(h)中也可以看出。从实际应用要求来讲,只有估计速度和精度二者兼顾才有意义,因此本文算法的应用效果最好,不仅估计时间较快,而且精确性又高。

表 1 Coastguard 序列全局运动估计结果
Tab.1 The GME result of Coastguard sequence

计算方法	参数 a_1	参数 a_2	参数 a_3	参数 a_4	参数 a_5	参数 a_6	信噪比 [dB]	与传统算法相比的花费时间比率
GME_1	0.999 8	-0.008 4	-8.596 0	-0.003 8	1.003 5	-0.518 7	25.54	1
GME_2	0.999 9	-0.005 0	-8.611 6	-0.003 5	1.001 2	-0.490 5	25.60	0.883
GME_3_1	0.999 9	-0.004 7	-8.619 7	-0.002 7	1.000 5	-0.462 1	25.65	0.786
GME_3_2	0.999 9	-0.005 8	-8.613 9	-0.003 5	1.002 0	-0.483 7	25.37	0.610
GME_4	0.999 9	-0.003 8	-8.619 2	-0.003 0	1.000 7	-0.479 9	25.58	0.645

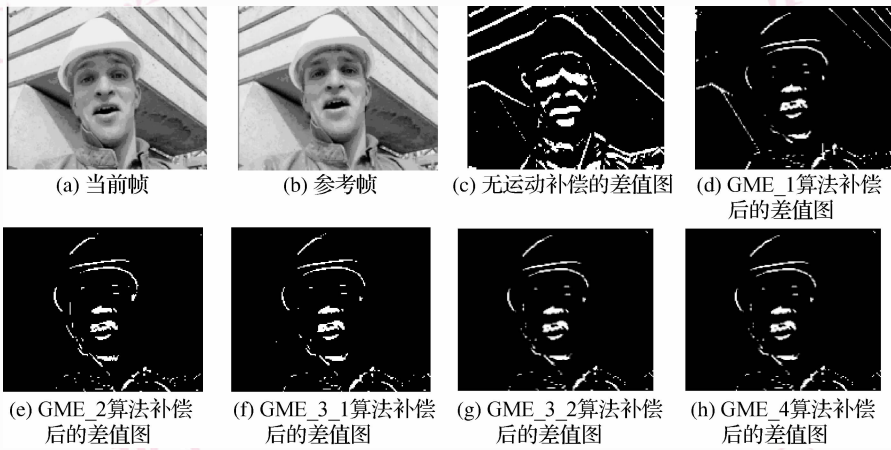


图 3 Foreman 序列全局运动估计结果
Fig.3 The GME result of the Foreman sequence

Foreman 序列主要针对摄像头的旋转运动,采用此序列的第 74 帧、第 72 帧(如图 3(a)和图 3(b)所示)进行仿真试验。各种全局运动算法的运动参数、峰值信噪比及运算时间如表 2 所示。为比较各种算法的分割效果,分别用全局运动参数对当前帧进行运动补偿,并用相同的阈值分割来得到二值图(如图 3(c)~图 3(h)所示)。同样,可通过对本

文算法与 GME_1 及 GME_2 算法进行比较来说明本文算法的优越性,再通过对本文算法与 GME_3_1 及 GME_3_2 算法进行比较来说明本文模板抽样的优越性。从表 2 可以看到,在对 Foreman 序列进行全局运动估计时,在保证准确性的情况下,本文算法所用时间只是 GME_1 算法的 51.7% 和 GME_2 算法的 56.1%。从图 3(c)可以看到,两帧图像之间既有

表 2 Foreman 序列全局运动估计结果
Tab.2 The GME result of Foreman sequence

计算方法	参数 a_1	参数 a_2	参数 a_3	参数 a_4	参数 a_5	参数 a_6	信噪比 [dB]	与传统算法相比的花费时间比率
GME_1	0.993 5	0.007 5	-2.865 8	-0.004 3	0.995 5	-0.347 8	25.18	1
GME_2	0.995 2	0.005 2	-2.860 2	-0.002 9	0.995 5	-0.164 4	25.47	0.921
GME_3_1	0.998 0	0.002 3	-2.854 3	-0.001 7	0.999 1	-0.015 7	25.63	0.785
GME_3_2	0.994 6	0.006 4	-2.862 0	-0.003 3	0.996 0	-0.253 9	22.32	0.490
GME_4	0.996 8	0.004 0	-2.859 8	-0.002 5	0.997 3	-0.064 4	25.50	0.517

前景的运动,也有背景的运动。图 3(d)~图 3(h)主要是前景中的人,而背景则基本上被去除了,这说明全局运动估计是有效的。图 3(d)中显示还有一些背景没有被去除,而 GME_2 算法和本文算法的运动估计效果最好,但本文算法所用的时间却比 GME_2 算法要少。再比较抽样模板间的区别,从表 2 可以看到,四皇后模板法用的时间明显多于本文算法和八皇后算法,八皇后算法用的时间最少,但从图 3(f),图 3(g)和图 3(h)中可以看到,八皇后模板法的估计效果明显不如其他两种算法,考虑到估计时间和精度两个方面,本文所使用的算法最好,不仅运算较快,而且精确性又高。

5 结 论

本文提出了一种新的全局运动估计算法,并将这种全局运动估计算法与现有的多种算法进行了实验比较。该算法是利用基于块的方法去除外点,其去除效果明显好于传统的 LM 算法。迭代算法的主要问题是计算速度太慢,因此提高速度成了关键问题,本文提出的基于皇后模板抽样的全局运动估计方法,在不影响估计精确性的前提下,大大提高了运算速度,基本满足了实时应用的要求。该算法可以用于运动物体跟踪、Sprite 软件编码,还可以用于运动分割,以及实时视频稳定等领域。

参考文献 (References)

- 1 Frederic Dufaux, Konrad Janusz. Efficient, robust and fast global motion estimation for video coding [J]. IEEE Transactions on Image Processing, 2000, 9 (3): 497-501.
- 2 Konrad Janusz, Frederic Dufaux. Digital equipment corporation, improved global motion estimation for N3 [A]. In: Meeting of ISO/IEC/JTC1/SC29/WG11. No. MPEG97/M3096 [C], San Jose, California, USA, 1998.
- 3 He Yu-wen, Zhao Li, Zhong Yu-zhuo, et al. A fast and robust algorithm of global motion estimation [J]. Journal of Software, 2001, 12 (8): 1220-1228. [贺玉文,赵黎,钟玉琢等.快速鲁棒的全局运动估计算法[J].软件学报,2001,12(8):1220-1228.]
- 4 Meyer F G, Bouthemy P. Region-based tracking using affine motion models in long image sequences [J]. Computer Vision and Graphics Image Process: Image Understanding, 1994, 60 (2): 119-140.
- 5 He Yu-wen, Zhong Yu-zhuo, Yang Shi-qiang. A fast global motion compensation coding method [J]. Chinese of Journal Electronics, 2001, 29 (2): 175-177. [贺玉文,钟玉琢,杨士强.一种快速全局运动补偿编码方法[J].电子学报,2001,29(2):175-177.]
- 6 Guo Li, Gong Sheng-rong, Cui Zhi-ming. A fast global motion estimation algorithm [J]. Journal of Image and Graphics, 2007, 12 (12): 2119-2126. [郭丽,龚声蓉,崔志明.一种快速全局运动估计算法[J].中国图象图形学报,2007,12(12):2119-2126.]
- 7 Alzoubi Hussein, Pan W. David. Efficient global motion estimation using fixed and random subsampling patterns [A]. In: Proceedings of the International Conference on Image Processing (IPIC) [C], San Antonio, Texas, USA, 2007, 1: 477-480.