

基于 GA 的压缩域中全局运动估计及 在字幕遮挡区域恢复中的应用

钱学明,刘贵忠

(西安交通大学电子与信息学院信息与通讯工程系,陕西西安 710049)

摘 要: 文章提出了一种直接使用压缩域中的运动矢量进行全局运动估计的方法,并用遗传算法优化输入运动矢量与全局运动参数所产生运动矢量的平均匹配误差. 最终的实验结果表明本文的方法能够很好的估计出全局运动的参数,并且提出了一种运用全局/局部运动信息进行视频中字幕遮挡区域的恢复的方法. 实验结果表明该方法取得了较好的视觉效果.

关键词: 全局运动估计; 字幕; 运动矢量; 遗传算法; 字幕遮挡区域恢复

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112 (2006) 10-1920-05

GA-Based Global Motion Estimation for Compressed Videos and its Application in Text Occluded Region Recovery

QIAN Xue-ming, LIU Gui-zhong

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

Abstract: A Genetic Algorithm (GA) based Global Motion Estimation (GME) method is proposed for compressed videos. Using motion vectors in compressed videos directly for GME, this approach utilizes GA to optimize the fitting errors between the input motion vectors and the ones generated from the estimated global motion parameters. Experimental results on different type of global motions show that the proposed genetic algorithm based global motion estimation method is satisfying. Moreover, a text-occluded region recovery approach is presented using the estimated global motion parameters, based on the Global/Local Motion (GM/LM) information for text occluded region in the video sequences. Experimental results on different videos show that this method achieves better visual results.

Key words: global motion estimation; video text; motion vector; genetic algorithm; text occluded region recovery

1 引言

视频中的全局运动在计算机视觉、视频编码、基于语义的视频分析与检索中有广泛应用. 全局运动是由摄像机的运动引起,这种运动通常可以用二维图像的参数变换来表达^[1]. 估计这种参数的方法称为全局运动估计. 全局运动估计的目的是对当前帧的运动用一个参数模型来表达.

全局运动估计(GME)方法分为基于象素域和压缩域的方法. 在象素域中常用基于最小二乘的迭代方法来搜索最优解,但是计算量非常大. 为了提高 GME 的速度,通常采用多分辨分析^[2]或者运用象素子集的方法^[3]. 如 MPEG-4 中使用由粗到精的 GME 方法^[2],对原始分辨率下的图像进行下采样,在最低分辨率图像中进行平移分量的估计,然后在其它分辨率下逐渐更新参数.

压缩视频中的运动矢量是基于匹配误差最小的准则得出的. 为了提高基于运动矢量信息的 GME 精度,需要剔除匹配误差较大以及平滑区域内宏块的运动矢量,然后采用基于梯度下降的搜索方法进行全局运动参数优化^[4]. 在文献[5]中,提出了一种非迭代的压缩域中 GME 方法,该方法是通过一种固定的空间采样模式来选取 4 个运动矢量以组成一个运动矢

量组,并用奇异值分解的方法来求解每个运动矢量组所对应的全局运动参数. 最后使用直方图聚类的加权平均方法来获得的全局运动参数. 本文采用遗传算法进行压缩域中全局运动参数估计,针对 8 参数的投影透视模型,对 8 个参数进行编码,然后用遗传算法搜索全局运动参数的最优解.

基于运动矢量的全局运动估计已应用于基于全局运动的错误消除等方面^[4,5]. 在本文中使用了全局运动/局部运动信息进行视频中字幕遮挡区域的恢复. 在文献[6]中采用一种时空信息相结合的字幕区域恢复方法. 在恢复的过程中把整个字幕区域看成运动一致的区域,没有对字幕区域的运动进行区分. 本文则充分利用字幕遮挡邻域的全局/局部运动信息来进行遮挡区域的恢复.

2 基于 GA 的全局运动估计

遗传算法(GA)是一种基于自然选择原理和自然遗传机制的寻优搜索算法,它模拟自然界中生命进化过程中的优胜劣汰适者生存的机制,实现特定目标的优化^[7].

2.1 适应度函数与个体选择

压缩域中的背景区域的运动矢量和全局运动具有相似性和一致性. 度量其一致性的一个准则是用输入运动矢量与由

全局运动参数所产生的运动矢量之间的平均匹配误差 $MME^{[4]}$,为了和现有基于 MV 的全局运动参数估计方法^[4,5]进行比较,本文仍选用平均匹配误差函数作为遗传算法的目标函数,其表达式为

$$MME = \frac{\sum_{i=1}^N w_i (ex_i^2 + ey_i^2)}{\sum_{i=1}^N w_i} \quad (1)$$

$$\begin{cases} ex_i = MVx_i - x_i + x_i \\ ey_i = MVy_i - y_i + y_i \end{cases} \quad (2)$$

其中 N 表示图像中宏块数目; w_i 表示第 i 个宏块的权值; (MVx_i, MVy_i) , (x_i, y_i) 和 (x_i, y_i) 分别表示其输入运动矢量,该宏块在当前帧和参考帧中的坐标位置, ex_i, ey_i 为其在水平、垂直方向上的匹配误差. 对于那些匹配误差较大或者平滑区域内的宏块设置其权值为 0, 其余的宏块的权值设置为 1, 以消除 outliers 对最终全局运动参数估计的影响^[2,4].

针对 8 参数的投影透视模型,用遗传算法来优化该模型的参数. 8 个参数所对应的二进制编码形式来表示一个全局运动参数所对应的染色体. 设第 k 条染色体,其所对应的全局运动参数为 $m_k = [m_{k,0}, m_{k,1}, \dots, m_{k,7}]$,从式(1)可知, MME 的大小体现该个体所提供的解的优劣. 在进化过程中希望解较好的个体有较大的概率选入下一代,相应的适应度函数的定义为

$$Fitness(k) = \frac{1}{MME(k) + 1} \quad (3)$$

其中 $MME(k)$ 表示第 k 条染色体的平均匹配误差.

相应地个体的选择是按照个体适应度比例来进行的^[7],对于某个染色体 k ,其被选中的概率 $P(k)$ 为

$$P(k) = \frac{Fitness(k)}{\sum_{k=1}^{Pop\ size} Fitness(k)} \quad (4)$$

其中 $Pop. Size$ 为染色体数目.

2.2 初始种群与算法终止条件

假设全局运动的参数模型为 8 参数的投影透视模型,并且假设每个参数的取值范围为 $m_j \in [-10, 15]$, ($0 \leq j \leq 7$). 在这里,对每个参数采用 20 位进行编码,即每条染色体的长度位 160 位. 设初始种群数目为 $Pop. size = 40$,交叉概率 $P_c = 0.85$,变异概率 $P_m = 0.05$.

对于初始种群的产生,可以用随机生成的方法或者按照先验知识产生. 本文充分利用了运动矢量场和全局运动的关系,来产生初始种群,以限定算法的搜索范围,加快算法的收敛速度. 初始种群的产生方法如下:

- (1) 随机选取 $Pop. size/4$ 个运动矢量,并构造相应的平移运动所对应的染色体形式;
- (2) 依次为几何变换、仿射变换、投影变换运动随机选取 $Pop. size/4$ 个运动矢量组(每个运动矢量组中运动矢量个数分别为 2,3,4),并用奇异值分解的方法来求解各个运动矢量组所对应的全局运动参数^[5].

在算法运行过程中如果相邻两代中所产生的对应的最佳个体的平均匹配误差的变化小于设定阈值 MME_{TH} ,或者运行的次数大于设定的最大运行次数 $MAXGEN$,则算法停止. 文中设定 $MME_{TH} = 10^{-4}$, $MAXGEN = 200$.

2.3 算法流程

本文基于遗传算法的 GME 算法流程如下

- step 1 按照 2.2 节中的方法产生 $Pop. size$ 个初始种群 Pop_0 , 计算其相应的适应度函数;
- step 2 初始化当前进化代数 $gen = 1$;
- step 3 基于适应度函数 $Fitness(k)$ ($1 \leq k \leq Pop. size$) 的个体选择;
- step 4 交叉、变异以产生新的种群 Pop_{gen} ;
- step 5 计算 Pop_{gen} 中每个个体的适应度函数;
- step 6 如果 $gen > MAXGEN$ 或者 $|BMME(gen) - BMME(gen - 1)| > MME_{TH}$, (其中 $BMME(gen)$ 是第 gen 代最佳个体所对应的 MME) 转至 step 8, 否则转至 step 7;
- step 7 重复步骤 step 3 到 step 6;
- step 8 算法结束, 输出最佳个体所对应的全局运动参数.

3 实验结果及其分析

为了评价本文基于遗传算法的全局运动估计方法(简称为 GA),用如下的 6 个全局运动来进行说明.

GM. Trans: $m = [1, 0, 10, 0.0001, 0, 1, -5.9927, 0, 0]$;

GM. Zoom: $m = [0.3542, 0, 0, 0, 0.3542, 0, 0, 0, 0]$;

GM. Geom: $m = [0.95, 0, 10, 0.4238, 0, 0.95, 5.7927, 0, 0]$;

GM. Affine: $m = [0.9964, -0.0249, 1, 0.0981, 0, 0.00856, 0, 0.9457, -7.2, 0, 0]$;

GM. Persp: $m = [0.9964, -0.0249, 6, 0.0981, 0, 0.0249, 0, 0.9964, 2.5109, -2.7e-5, 1.9e-5]$;

GM. Other: $m = [1, 0, 4, 0.4154, 0, 1, 0, -1.263e-4, 0]$.

我们采用所估计出的全局运动 \hat{m} 与原始参数 m 的信噪比(SNR)^[4]作为比较准则来评价全局运动参数估计的性能,其具体表达式如式(5).

$$SNR(m, \hat{m}) = 10 \times \log_{10} \left[\frac{\sum_{i=1}^N ((MVx_i(m))^2 + (MVy_i(m))^2)}{\sum_{i=1}^N (MVx_i^2 + MVy_i^2)} \right] \quad (5)$$

其中 $MVx_i = MVx_i(m) - MVx_i(\hat{m})$, $MVy_i = MVy_i(m) - MVy_i(\hat{m})$, $(MVx_i(m), MVy_i(m))$, $(MVx_i(\hat{m}), MVy_i(\hat{m}))$ 表示对第 i 个宏块用全局运动参数 m 和 \hat{m} 所产生的运动矢量.

3.1 算法收敛性能分析

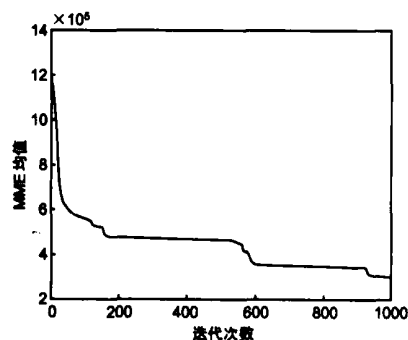


图 1 MME 均值与迭代次数的关系

图 1 给出了全局运动 GM. Affine 在随机产生的种群条件下初始种群大小 $Pop. size = 60$, 运行 20 次平均的最佳个体的 MME 均值和迭代次数的关系. 从中我们可以看出在随机产生种群下, 算法的收敛速度非常慢, 在运行 1000 代的情况下, MME 仍然大于 10^5 .

图 2 (a) 和 (b) 中分别给出了 6 种全局运动在初始种群大小 $Pop. size$, 100 次运行下的每代最佳个体所对应的 MME 均

值和 SNR 均值与迭代次数的关系. 比较图 1 和图 2, 我们可以看出, 基于约束条件所产生的种群要比随机产生的种群, 收敛速度有显著提高. 这也充分说明, 初始种群选择的必要性. 图 2 (c) 和 (d) 分表给出了 GM. Affine 和 GM. Other 初始种群数目为 80, 60 和 40 三种情况下, 100 次运行的最佳个体所对应的 MME 均值、SNR 均值与迭代次数之间的关系. 从中可以看出, 种群个体数每增加 20, 相应的平均 SNR 大约有 2dB 的提高.

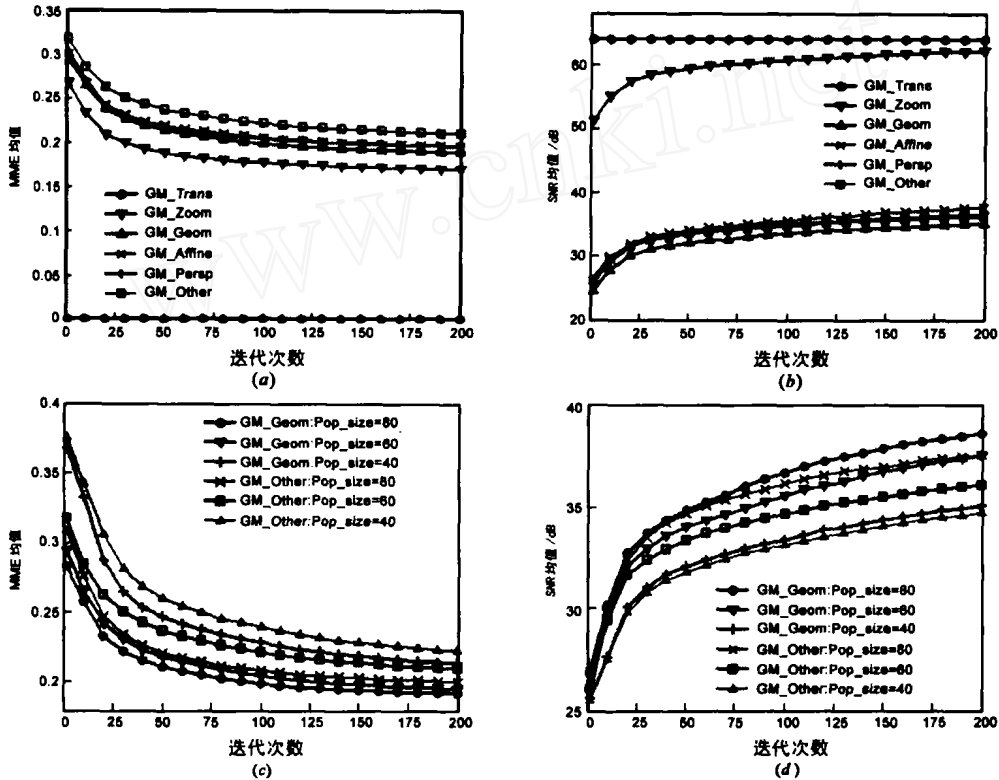


图 2 (a) MME 均值与迭代次数关系; (b) SNR 均值与迭代次数关系; (c) 全局运动 GM_Geom 和 GM_Other 在不同种群数目下 MME 均值与迭代次数关系; (d) 全局运动 GM_Geom 和 GM_Other 在不同种群数目下 SNR 均值与迭代次数关系。

表 1 基于运动信息的全局运动算法估计性能比较

方法	SVD		Newton		GA	
	估计出的全局运动参数	SNR	估计出的全局运动参数	SNR	估计出的全局运动参数	SNR
GM. Trans	$m = [1, 0, 9, 0, 0, 1, -6, 0, 0, 0]$	64.07	$m = [1, 0, 9, 0, 0, 1, -6, 0, 0, 0]$	64.07	$m = [1, 0, 9, 0, 0, 1, -6, 0, 0, 0]$	64.07
GM. Zoom	$m = [0.352273, -0.00, -0.026, 0, 0, 0.354167, 0, 0, 0, -0.0]$	52.20	$m = [0.3570, 0.002, -0.3887, 0.0008, 0.3564, -0.1114, 0, 0]$	61.59	$m = [0.3508, 0.0001, 0.1857, -0.0013, 0.3522, 0.2739, 0, 0]$	62.26
GM. Geom	$m = [0.948864, 0, 10, 4.28571, 0, 0.950231, 5.88889, 0, 0]$	28.72	$m = [0.9508, -0.0001, 10, 2105, 0.0001, 0.9508, 5.687, 0, 0]$	33.62	$m = [0.9530, 0.0008, 10.0995, 0.0005, 0.9511, 5.7844, 0, 0]$	34.98
GM. Affine	$m = [0.9933, -0.02417, 1.2087, 0.0856, 0.9442, -7.3092, 0, 0]$	16.89	$m = [0.9967, -0.0249, 1.0252, 0.0857, 0.9462, -7.287, 0, 0]$	45.23	$m = [0.9947, -0.0249, 1.2168, 0.0859, 0.9453, -7.3301, 0, 0]$	37.53
GM. Persp	$m = [0.99527, -0.0272, 6.1910, 0.025, 0.99671, 2.63947, -2.6e-5, -1.2e-6]$	19.16	$m = [0.9988, -0.025, 5.9526, 0.0263, 0.9977, 2.2946, -1.94e-5, -1.77e-5]$	36.54	$m = [0.9987, -0.0231, 5.7115, 0.0252, 0.9992, 2.347, -2.236e-5, -2.606e-5]$	36.38
GM. Other	$m = [1.0010, 0, 4.3810, 8.8947e-4, 1.0007, -0.1417, -1.1011e-4, 0]$	41.68	$m = [1.0051, -0.0, 4.002, 0.0014, 1.0022, -0.3340, -0.0001, 0]$	33.55	$m = [0.9969, -0.002, 4.6811, -8e-4, 0.9957, 0.2971, -1.186e-4, 0]$	36.07

3.2 算法性能比较

表 1 中给出了本文方法 GA,文献[4,5]中的全局运动估计方法(分别记为 SVD,Newton)的估计效果比较.从中我们可以看出基于迭代的方法所估计出的全局运动参数较非迭代的方法精度要高,并且本文 GA 算法的 SNR 相对于 Newton 性能有所提高.在表 2 中给出了不同运动估计算法在 PIV 2.8G PC 上的运行时间比较,从中可以看出 GA 的计算复杂度要高于 SVD 和 Newton.但是对于视频序列中的全局运动一般会持续几百帧或者更长,这些帧的全局运动具有一定的连续性.在这种情况下,初始种群的产生方法是从前一帧最后一代的种群中按照适应度函数选取 $Pop. size$ 个个体并且用当前帧的运动矢量信息参数另外的 $Pop. size/4$ 条染色体,对于图 3 所用的几个例子序列的平均每帧全局运动的平均时间为 27.54ms.这相对于单帧的 GME 而言,计算速度约有 10 倍的提高.

表 2 几种全局运动的不同方法的估计时间比较(ms)

	SVD	Newton	GA
GM. Trans	3.078	3.274	10.612
GM. Zoom	3.726	4.466	68.310
GM. Geom	3.638	6.483	257.625
GM. Affine	4.067	9.859	268.113
GM. Persp	3.683	9.066	275.074
GM. Other	3.855	8.448	276.867

4 全局运动估计在字幕遮挡区域恢复中的应用

首先,对视频中的字幕采用文献[8]中的方法进行检测、定位.其优点是在字幕检测的同时可以判断字幕是否跨越镜头,并视频中字幕的起始帧以及字幕在像素域中的位置 TR.记字幕所在的宏块区域及其周围 8 邻域的非字幕宏块区域为 TMBR 和 OTMBR.本文所提出的基于全局/局部运动信息的全局运动估计方法的恢复方法是通过 OTMBR 的运动信息对 TMBR 的运动信息进行估计与恢复.对 OTMBR 中的每个宏块,计算其全局/局部运动差异.在这里用压缩视频中的运动矢量同用全局运动参数所生产的运动矢量的差异来描述.假设宏块 $j \in OTMBR$,运动差异的幅值 EMV_j 为

$$EMV_j = \sqrt{ex_j^2 + ey_j^2} \quad (6)$$

在理想情况下,背景区域以及那些相对于背景静止的对象物区域,其局部运动和全局运动具有一定的一致性,也即对于这些宏块有 $EMV_j = 0$.因此,在对字幕遮挡区域进行恢复时,可以按照全局运动的背景区域来进行处理.然而,对于那些存在显著运动的对象物所在的宏块区域,则有 $EMV_j > 0$,可以通过全局/局部运动信息的差异来进行恢复.具体的实现方法如下:

- (1) 确定字幕所在的宏块区域 TMBR 以及其周围 8 邻域的非字幕宏块区域 OTMBR;
- (2) 对于 OTMBR 中的每个宏块按照式子 (6) 来计算其全局运动和局部运动差异信息;并对 OTMBR 中的每个宏块进行标记为有显著运动差异的和没有显著差异两种;

$$SalientMB(j) = \begin{cases} 1, & \text{if } EMV_j > EMV_{th} \text{ and } j \in OTMBR \\ 0, & \text{else} \end{cases} \quad (7)$$

其中 EMV_{th} 为一个判断显著性差异的阈值,在本文中我们设定 $EMV_{th} = 2$.

- (3) 设字幕区域宏块 $s \in TMBR$,其运动差异信息可以通过其 8 邻域中非字幕宏块运动差异信息 $(ex_{(s,l)}, ey_{(s,l)})$, $(l = 1 \dots 8)$ 的加权平均来获得

$$\begin{cases} EMVTx(s) = \frac{\sum_{l=1}^8 ex_{(s,l)} \times SalientMB(l)}{NonTextMBNum(s)} \\ EMVTy(s) = \frac{\sum_{l=1}^8 ey_{(s,l)} \times SalientMB(l)}{NonTextMBNum(s)} \end{cases} \quad (8)$$

其中 $NonTextMBNum(s)$ 表示该字幕宏块 s 的 8 邻域中非字幕宏块的数目.在这个宏块中的字幕区域的每个象素 $(x, y) \in s \subseteq TMBR$ 的全局/局部运动差异的估计量为 $(EMVTx(s), EMVTy(s))$.

- (4) 判断字幕的起始帧之前或者终止帧之后是否为一个完整的非字幕遮挡帧,然后可以分为如下的两种情况进行恢复.

(a) 如果有一个完整的非字幕遮挡帧,则可以用式 (8) 所得出的全局/局部运动差异信息进行恢复.设第 t 帧中字幕遮挡区域的恢复结果为 $T_{TR}(x, y, t)$,相应的恢复方法如下

$$\begin{cases} T_{TR}(x, y, t) = I_{temporal}(x, y, t \pm 1) \\ \begin{cases} x = \frac{m_0 x + m_1 y + m_2}{m_6 x + m_7 y + 1} + EMVTx(x, y) \\ y = \frac{m_3 x + m_4 y + m_5}{m_6 x + m_5 y + 1} + EMVTy(x, y) \end{cases} \end{cases} \quad (10)$$

其中 $(x, y, t \pm 1) \in TR(t \pm 1)$ 表示用全局/局部运动补偿后所计算出的坐标位置, $I_{temporal}(g)$ 表示用时间冗余信息恢复的参考图像.

- (b) 如果字幕的出现帧之前或者消失帧之后没有一个完整的背景帧,对于字幕所在的区域先进行基于全局运动信息的恢复,以得出部分字幕遮挡区域,对于剩下的部分则采用基于空间信息的恢复方法.

$$T_{TR}(x, y, t) = I_{temporal}(x, y, t \pm 1) + I_{spatial}(x, y, t) \quad (11)$$

其中 $I_{spatial}(x, y, t)$ 表示用空域冗余信息恢复的参考图像^[6].

通过上面的方法,可以构造出一个完整的非字幕遮挡的参考帧,其可以为视频中后续帧字幕遮挡区域的恢复提供参考.通过该帧相关的全局运动参数和已经确定的背景,我们可以对连续帧中的字幕所遮挡的区域进行恢复.如图 3 所示给出了 3 幅字幕遮挡区域的恢复效果图.可以看出字幕遮挡的对象物和背景区域都得到很好的恢复.

5 结论

本文提出了一种基于遗传算法的压缩域中的全局运动估计方法,该方法充分使用压缩域中的运动矢量信息来进行全局运动参数的估计.本文通过运动矢量组产生初始种群,有效地解决了随机产生种群条件下算法收敛速度慢的问题.对于



图 3 基于全局/局部运动信息的字幕遮挡区域恢复效果

不同类型的全局运动,最终的实验结果表明本文方法的有效性.另外,本文提出了一种使用全局/局部运动信息进行字幕遮挡区域恢复的方法,取得了较好的恢复效果.

参考文献:

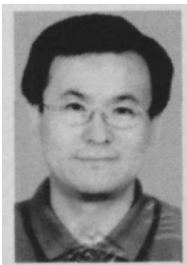
- [1] C Stiller, J Konrad. Estimating motion in image sequences, a tutorial on modeling and computation of 2D motion [J]. IEEE Signal Process Mag, 1999, 16(7): 70 - 91.
- [2] ISO/IEC JTC1/SC29/WG11-2001. MPEG-4 Video Verification Model version 18.0[S].
- [3] Y Keller, A Averbuch. Fast gradient methods based on global motion estimation for video compression[J]. IEEE Trans Circuits Syst Video Technol, 2003, 13(4): 300 - 309.
- [4] Y P Su, M T Sun, et al. Global motion estimation from coarsely sampled motion vector field and the applications [J]. IEEE Trans Circuits Syst Video Technol, 2005, 15(2): 232 - 242.
- [5] Y P Su, M T Sun. A noniterative motion vector based global motion estimation algorithm [A]. International Conference on Multimedia and Expo[C]. Taipei, Taiwan, 2004, 1: 703 - 706.
- [6] C W Lee, K Jung, et al. Automatic text detection and removal in video sequences [J]. Pattern Recognition Letters, 2003, 24(15): 2607 - 2623.
- [7] J H Holland. Adaptation in natural and artificial systems [M]. Ann Arbor: University of Michigan Press, 1975: 30-58.
- [8] X O Tang, X B Gao, et al. A spatial-temporal approach for video caption detection and recognition [J]. IEEE Trans Neural Networks, 2002, 13(4): 961 - 971.

作者简介:



钱学明 男, 1975 年 1 月出生于安徽潜山, 现为西安交通大学电信学院博士研究生, 感兴趣的研究方向为图像处理、视频编解码、基于语义的视频分析、摘要和检索.

E-mail: qxm@mailst.xjtu.edu.cn



刘贵忠 男, 1962 年 9 月出生于陕西延安, 1982 年 9 月获西安交通大学学士学位; 1985 年 9 月获硕士学位; 1989 年 9 月获荷兰埃因霍温大学数学与信息科学系博士学位, 主要从事小波分析时频分析等非平稳信号分析方法及自适应处理方法、图象视频多媒体压缩编码及传输控制、信号采集与分析系统、遥感图象压缩分析处理及高分辨地震勘探信号处理等方面的研究工作.

E-mail: liugz@mail.xjtu.edu.cn