基于全局运动信息的视频检索技术

俞天力,章毓晋
(清华大学电子工程系图像图形研究所,北京 100084)

摘 要: 运动信息是描述视频内容的一种重要信息,本文介绍了一个基于全局运动信息的视频检索系统.在该系统中,我们通过对视频数据进行短时全局运动分析,较为精确地提取出了描述全局摄像机运动的双线性运动模型,并以该模型参数为运动特征,以特征点序列顺序匹配以及全局运动矢量距离平方和为基础,构造了一个视频检索方案.实验结果表明,在特定的应用领域,如体育类视频中,我们的检索方案能够实现一定程度的语义内容检索,同时提供了采用其他图像特征所无法实现的检索功能.

关键词: 视频检索; 运动信息; 全局运动; 运动模型
中图分类号: TP391; TN911.7 **文献标识码**: A **文章编号**: 0372-2112 (2001) 12A-1794-05

Video Retrieval Based on the Global Motion Information

YU Tian-li, ZHANG Yu-jin
(Institute of Image and Graphics, Electronic Engineering Department, Tsinghua University, Beijing 100084, China)

Abstract: Motion is an important clue to describe the content of a video sequence. This paper presents a global motion information based video retrieval system. In this system, video sequences have undergone a short-term global motion analysis, which will accurately extract the bilinear motion model of the global camera motion. The extracted model parameters are used as motion features for retrieval, and the retrieval scheme is based on the feature point sequential match technique and the calculation of square sum of global motion vector distance. Experimental results show that in certain application domain, such as sports video, our retrieval scheme could achieve a near-semantic content retrieval, and it provides new functionalities that other image based features cannot achieve.

Key words: video retrieval; motion information; global motion; motion model

1 引言

自 20 世纪 90 年代以来,随着数字技术和互联网的发展,数字视频的产生和传播变得越来越容易,数字电视、多媒体广播、视频会议已经开始逐步走入人们的日常生活中,人们接触到的视频数据以前所未有的速度增长.当今人们面临的问题已不再是视频内容的匮乏,而是面对浩如烟海的视频信息,如何有效地找到自己需要的内容.基于内容的视频检索技术就是为了满足这方面的需求而迅速发展起来的,它通过对视频数据中所包含的视觉内容进行分析和特征提取,使人们可以直接利用计算机搜索符合主观感受的相似内容片段.

一般认为,视频数据所包含的视觉内容包括颜色、纹理、形状和运动的信息[1],其中运动信息是视频区别于图像数据所特有的内容.对于一个视频片段来说,运动信息是反映视频中变化演进的重要信息,要想对视频内容进行全面的刻划,运动信息是必不可少的一个方面.

视频中的运动信息复杂多样,在对视频数据进行运动分析时,通常将摄像机移动形成的运动信息和由场景中物体产生的运动信息分开处理,分别称为全局运动和局部运动.全局运动具有整体性强、计算量小、结果稳定和特征表示方便等特点,而局部运动则相对复杂、计算量大、结果不够稳定且难以用特征完整地刻划.在大多数视频序列中,摄像机的运动总是跟踪着视频中重要人物和事件的运动,因此可以认为全局运动信息在一定程度上反映了视频中的语义内容.本文针对这一特点提出了一种从视频中提取全局运动信息进行检索的技术,从对足球视频的检索结果来看,该技术可以在特定的应用领域,如体育类视频中起到较好的检索作用,提供其他视觉特征所无法实现的检索功能.

目前,大多数基于运动的检索技术都只对全局运动进行粗略的分类处理,如 Chang 等[2]和朱兴全等[3]实现的系统,需要指出的是,这些检索方案比较适用于大粒度的检索,如大型视频剪辑库内的检索,而本论文所提出的检索方案则比较适合于较细粒度的视频检索,如在几段视频数据中定位具有特定相似内容的片段.由于对视频中的全局运动采用了较为精确的模型估计算法,并配以细致的特征匹配方案,使我们能够相当准确地定位出所需要的视频片段的位置.

收稿日期:2001-07-18;修回日期:2001-09-13
基金项目:国家自然科学基金(No. 69672029;No. 60172025)

2 系统概述

图 1 给出了我们提出的基于全局运动信息的视频检索系统框图.

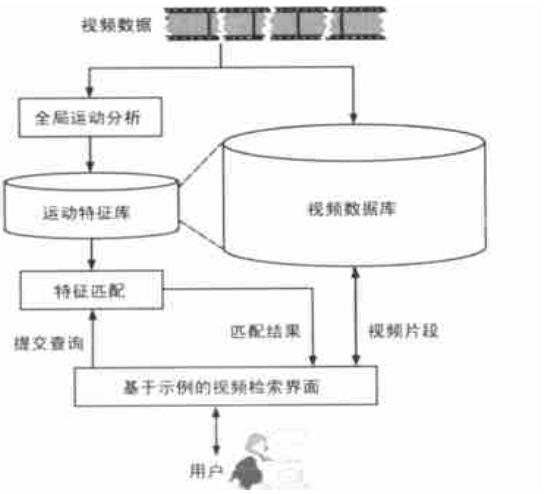

图 1　基于全局运动信息的视频检索系统

在图 1 的系统中,对新加入的视频数据首先进行全局运动分析,提取出其相邻帧的全局运动模型,全局运动模型作为视频序列的运动特征存入特征数据库,用户检索采用示例查询(Query By Example)机制,系统得到用户所提交的示例片段后,根据它的全局运动特征,通过特征匹配模块在特征库中查找具有最相似的视频片段,并根据相似度进行排序,然后按照排序的结果从视频数据库中取出视频片段,返回给用户.

3 短时全局运动分析

所谓短时全局运动分析是指从视频序列中的相邻帧间提取全局运动信息. 提取运动信息实际上是寻找两个视频帧之间对应物体的相对位置变化,而这种对应关系只有在较小的时间间隔的情况下才能够做到较为精确的估计. 因此,首先采用这种短时全局运动分析的方法从视频中提取运动特征. 考虑到全局运动特征是由摄像机引起的整体运动,可以对它建立形式简单的模型.

3.1 全局运动的双线性模型

对于全局运动进行建模既可以从摄像机操作的角度[4],也可以从参数模型的角度[1]出发,由于参数模型具有较好的数学形式,而且计算方便,因此我们的系统中就采用这种模型. 我们用双线性模型来表示由摄像机运动造成的全局运动矢量,如式(1)所示:

$$\begin{cases} u = a_0 xy + a_1 x + a_2 y + a_3 \\ v = a_4 xy + a_5 x + a_6 y + a_7 \end{cases} \tag{1}$$

式中,(x, y)是镜头中某一点的坐标,而(u, v)是该点在 x 和 y 方向的运动矢量.

采用双线性模型的好处是任意四边形在双线性几何变换下仍然是四边形,而原来平行的直线可能变换为不平行的直线[5],这些特性使该模型可以更好地适应摄像机运动时立体景物所造成的透视变换,以更好地描绘全局运动.

3.2 运动矢量的计算

为了估计全局运动的双线性模型参数,必须获得 4 个以上不同位置的运动矢量数据,在本文中,我们用块匹配法来求得这些运动矢量. 块匹配法可以适应较大幅度的运动矢量,而这在全局运动中十分常见,同时,通过选取较大的匹配块尺寸,可以减少由局部物体运动引起的计算偏差,从而为全局运动估计提供较为准确的数据. 图 2 是在一对视频帧上利用 16×16 块匹配全搜索法计算的运动矢量结果.

图 2　由 16×16 块匹配全搜索算法计算出的运动矢量

图 2 是从一个同时带有摄像机扫视和放大的视频中截取的一帧. 从图中可以看出,虽然块匹配法在大多数位置能够给出运动矢量的正确值,但是在图像中的低纹理区域还存在很多随机的误差数据,同时前景物体(球员)的运动,也造成了运动矢量与全局运动不符,面对这些噪声,我们采用下节提出的带异常点剔除的最小二乘估计法来提取出正确的运动模型.

3.3 带异常点剔除的最小二乘估计法

假设运动矢量的随机误差服从高斯分布,那么最优的估计准则是最小均方误差准则,我们可以用最小二乘法来估计出双线性运动模型的 8 个参数. 但是由于最小二乘法对于异常点数据的干扰较为敏感,与摄像机运动模型不符的前景运动物体,以及块匹配在低纹理区域出现的随机错误,都会较大程度地影响最终估计参数的准确性,为了降低这些误差的影响,我们在最小二乘估计的基础上,加入了迭代的异常点剔除的步骤[6],经过改进的估计算法步骤如下:

(1) 将所有块匹配计算的运动矢量加入到估计数据集中.

(2) 用最小二乘法从估计数据集中估计全局运动的双线性模型,使估计的均方误差最小.

(3) 计算所有运动矢量与由第 2 步估计的全局运动模型所恢复出的运动矢量的平均误差值

$$E_{avg} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2} \tag{2}$$

(4) 取 E_{avg}为门限,将块匹配运动矢量数据中所有估计误差低于该门限的数据归为新的估计数据集,其余归为异常数据.

(5) 若新的估计数据集与上次迭代所得到的估计数据集一致,则停止循环,否则更新估计数据集,剔除异常数据,然后转到第 2 步.

该算法对于场景中有较多的运动物体时也有较好的适应性,能够成功地把非全局运动的数据从估计数据集中剔除,从

而获得全局运动模型的较精确的估计. 图 3 是利用该算法对图 2 中的运动矢量进行估计后,根据运动模型恢复出的运动矢量的结果,图中白色线条表示运动矢量的大小和方向,灰色的块表示最后一次迭代后被剔除的数据.

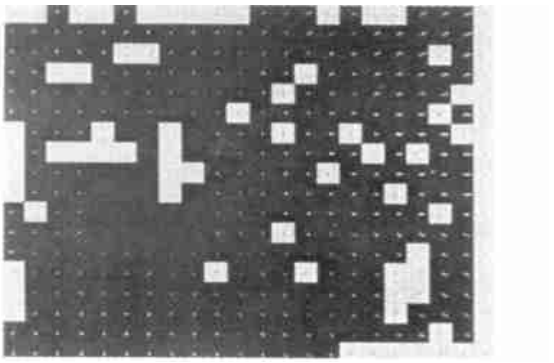

图 3 带异常点剔除的最小二乘估计求出的全局运动模型. (白色直线表示估计出的运动矢量,灰色块表示被剔除的运动矢量数据)

对照图 2 和图 3 可以看出,视频中球员跑动而形成的运动矢量数据都被划为剔除数据,而一些低纹理区域中的错误运动矢量也被成功剔除,由此而估计出的全局运动模型较准确地反映了摄像机的扫视和放大运动.

当然,该算法的前提是前景运动物体和误差运动矢量不能过多,对于一些近镜头,由于前景运动物体占屏幕的比例过大(超过 1/2),即使剔除算法也无法把所有的运动物体区域剔除,这时估计出的全局运动模型会有较大的误差,对于这种情况,只能根据一定的先验知识,对其作特殊的处理.

4 全局运动检索方案

第 3 节中,通过带异常点剔除的最小二乘估计法很好地求出了视频中相邻两帧间的全局运动模型,我们可以将全局运动双线性模型的 8 个参数作为视频的运动特征. 在对视频中每个相邻帧提取全局运动特征的基础上,设计了利用这些全局运动特征进行检索的方案.

4.1 基于特征点序列的顺序匹配检索

由于双线性模型参数描述的是摄像机的短时运动信息(大约几十毫秒),而人们一般理解中的运动则是有一定时间跨度的(一秒或者更长),为了使检索的结果符合人们一般的理解习惯,必须在全局运动检索方案中将短时的运动信息结合起来,为此我们提出了特征点序列的概念.

每个短时全局运动特征可以看作特征空间中的一个点,视频中提取出的全局运动信息可以由一系列点来表示,这一系列点按照相应帧的时间顺序排列,就构成了一个特征点序列,这个特征点序列既包含了每对相邻帧间的全局运动信息,又保留了前后帧的时间顺序关系,能够较好地反映视频中的全局运动变化.

我们所要实现的示例查询,实际上就是在特征库中找出与示例视频的特征点序列距离最近的其他特征点序列. 为了计算两个具有一定长度的特征点序列之间的距离,我们采用如下的特征点序列顺序匹配的方法.

对于两段视频序列 V_1, V_2,假设它们的特征点序列分别为 $\{m_1(i)\}$、$\{m_2(j)\}$,序列长度分别为 L_1、L_2.

如果两个特征点序列的长度相同,即 $L_1 = L_2$,则它们之间的距离可以定义为它们对应的各特征点的距离之和,如式(3):

$$D(V_1, V_2) = \sum_{i=1}^{L_1} d(m_1(i), m_2(i)) \tag{3}$$

在式(3)中 $d(m_1, m_2)$ 是计算两个特征点间距离的函数.

对于两个长度不同的特征点序列,我们必须考虑如何选择它们之间的对应点,假设 $L_2 > L_1$,则可以先在 V_2 中以不同的时间为起点选取长度为 L_1 的特征点序列 V_2',V_1 与 V_2' 之间的距离可以用式(3)计算得到,通过移动 V_2' 的时间起点位置,如图 4 所示,我们可以得到一个以时间起点为变量的函数,此函数中的最小值点,就是特征点顺序匹配的最佳位置点,可以作为两个序列之间的距离度量,用式(4)表示:

$$D_{v_0, v_1}(t) = \min_{t=1}^{L_2 - L_1} \sum_{i=1}^{L_1} d(m_1(i), m_2(i+t)) \tag{4}$$

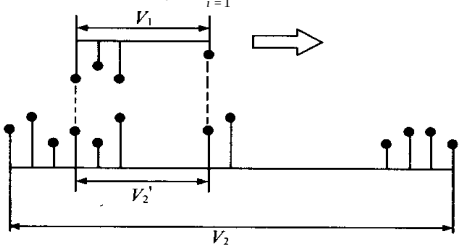

图 4 两个不同长度的视频序列的特征点顺序匹配

在实际的应用中,我们会选取多个最小的局部极值点作为可能的匹配点,返回给用户以做最终的挑选.

4.2 全局运动矢量距离平方和

为了计算全局特征点序列的距离,还必须定义两个全局特征点间的距离函数 $d(m_1, m_2)$. 由于我们的系统着眼于粒度较细的检索应用,希望特征点的距离能够较为准确地反映两个不同的全局运动的差异,在此,通过计算由两个全局运动模型在整个帧画面上恢复出来的运动矢量的距离平方和,即式(5),来确定两个全局特征点间的距离.

$$\begin{aligned} d(m_1, m_2) = & \sum_{x,y} [u_{m1}(x, y) - u_{m2}(x, y)]^2 \\ & + [v_{m1}(x, y) - v_{m2}(x, y)]^2 \end{aligned} \tag{5}$$

假设 m_1, m_2 的模型参数分别为 $a_{10} \cdots a_{17}$ 和 $a_{20} \cdots a_{27}$,则根据式(5)和(1),它们间的距离可以表示成:

$$\begin{aligned} d(m_1, m_2) = & \sum_y \sum_x [(a_{10} - a_{20}) + (a_{11} - a_{21}) x + (a_{12} - a_{22}) y \\ & + (a_{13} - a_{23}) xy]^2 + [(a_{14} - a_{24}) + (a_{15} - a_{25}) x \\ & + (a_{16} - a_{26}) y + (a_{17} - a_{27}) xy]^2 \end{aligned} \tag{6}$$

将式(6)化简并合并,可以得到

$$\begin{aligned} d(m_1, m_2) = & [(a_{13} - a_{23})^2 + (a_{17} - a_{27})^2] \sum_x x^2 \sum_y y^2 \\ & + [(a_{11} - a_{21})^2 + (a_{15} - a_{25})^2] \sum_x x^2 \sum_y 1 \\ & + [(a_{12} - a_{22})^2 + (a_{16} - a_{26})^2] \sum_x 1 \sum_y y^2 \\ & + [(a_{10} - a_{20})^2 + (a_{14} - a_{24})^2] \sum_x 1 \sum_y 1 \end{aligned}$$

$$+2[(a_{13}-a_{23})(a_{11}-a_{21})+(a_{17}-a_{27})(a_{15}-a_{25})]\sum_x x^2\sum_y y$$
$$+2[(a_{13}-a_{23})(a_{12}-a_{22})+(a_{17}-a_{27})(a_{16}-a_{26})]\sum_x x\sum_y y^2$$
$$+2[(a_{13}-a_{23})(a_{10}-a_{20})+(a_{17}-a_{27})(a_{14}-a_{24})]\sum_x x\sum_y y$$
$$+2[(a_{11}-a_{21})(a_{12}-a_{22})+(a_{15}-a_{25})(a_{16}-a_{26})]\sum_x x\sum_y y$$
$$+2[(a_{12}-a_{22})(a_{10}-a_{20})+(a_{16}-a_{26})(a_{14}-a_{24})]\sum_x 1\sum_y y$$
$$+2[(a_{11}-a_{21})(a_{10}-a_{20})+(a_{15}-a_{25})(a_{14}-a_{24})]\sum_x x\sum_y 1 \quad (7)$$

确定了视频帧画面的长 M 和宽 N 之后,有:

$$\sum_x 1=M,\ \sum_y 1=N \quad (8)$$

$$\sum_x x=\frac{M(M+1)}{2},\ \sum_y y=\frac{N(N+1)}{2} \quad (9)$$

$$\sum_x x^2=\frac{M(M+1)(2M+1)}{6},\ \sum_y y^2=\frac{N(N+1)(2N+1)}{6} \quad (10)$$

将式(8)～(10)以及两个运动模型的参数代入到式(7)就可以求出它们之间的距离 $d(m_1,m_2)$ 了.

5　实验结果

采用本文中提出的方案,我们用 Java 语言实现了一个视频全局运动分析和检索的实验系统.实验的数据是一段来自 MPEG-7 测试序列集[7]的 14 分钟的足球比赛视频,在整段视频中,我们用人工观察的方式定位出了 6 个同时具有向左扫视和放大镜头的全局运动的 25 帧视频片段,我们选取这 6 个片段中的一个作为示例进行检索(图 5),图 6 是系统返回的前 10 个检索结果.为了便于比较,我们将由全局运动估计恢复出的运动矢量(白色直线)叠加视频帧上.在处理检索结果时,为了避免相邻视频片段在结果中反复出现,我们采用了相似结果合并的方法,即当一个片段被选为检索结果后,则在其前后同样长度的片段内不再允许出现新的检索结果.

图 5　一个全局运动检索的示例

图 6 中显示的前 10 个返回结果中,R_1、R_2、R_4、R_5、R_8 是其他 5 个正确的结果,分析图 6 中的错检结果,R_3 和 R_7 错检是因为他们都有与查询结果类似的向左扫视的运动,而 R_6、R_9 和 R_{10}是因为球场 3 维立体透视的原因,摄像机的扫视会产生与放大镜头相类似的全局运动矢量.

在通常的足球视频中,摄像机的放大镜头说明该处有较重要的活动发生,比方说突然进攻或射门,而向左扫视表示该活动是由右向左运动的,从正确的检索结果来看,查询示例以及查询结果 R_1、R_2、R_4、R_5 都是射门的镜头,而 $R6$ 是进攻传球的镜头,由此可见,在足球比赛中,摄像机的运动确实表达了一定的视频语义内容.

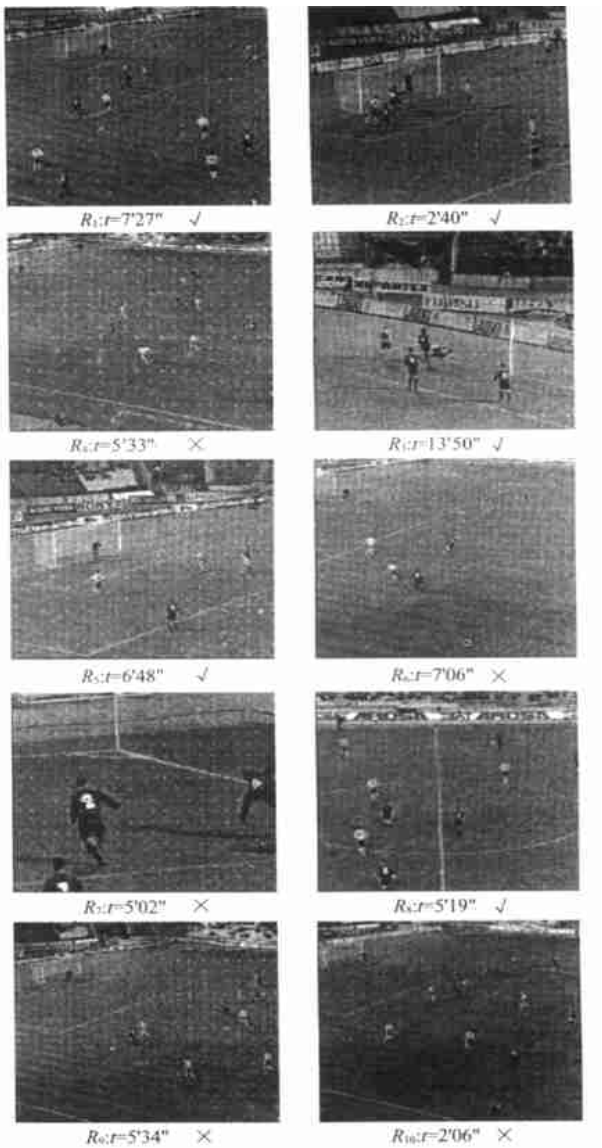

图 6　全局运动示例检索的前 10 个返回结果
(√:正确的结果, X:错检的结果)

表 1　全局运动特征检索结果汇总

	第 N 个正确结果在前 10 个检索结果中的位置				
查询示例	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Z_1	1	2	3	4	——
Z_2	4	6	——	——	——
Z_3	1	2	4	10	——
Z_4	1	2	3	5	——
Z_5	1	2	3	5	8
Z_6	1	3	——	——	——

进一步,将 6 个人工选定的正确结果分别作为示例进行

检索,所返回的前 10 个检索结果中正确结果所处的位置汇总于表 1 中.在这里我们采用在正确结果前 10 个检索结果中所处的位置来衡量检索的效果,这是因为在实际的检索系统中,排在最前面的几个检索结果将对用户的满意程度起到决定性的作用,而且正确的结果越靠前,对用户的效用就越大.

从表 1 的结果来看,多数情况下,全局运动检索算法都可以在前 10 个检索结果里找到大部分的正确结果,当然在对于某些片段的检索结果还不是很好,如 Z_2 和 Z_6 的结果,这是因为拍摄中摄像机的运动并不是规则的,在某些片段中存在较多的晃动,影响了检索结果的准确性.

6 结论

运动信息是视频数据中反映其内容变化演进的特有信息,对于描述视频的内容具有不可替代的作用.本文利用了视频中摄像机的运动通常反映了主要事件和人物的运动情况这一特点,提出了一种利用视频中的全局运动信息进行视频检索的方法.在我们的系统中,短时全局运动信息通过双线性运动模型来描述,采用块匹配和带异常点剔除的最小二乘估计来计算.在以双线性模型参数作为运动特征的基础上,我们提出特征点顺序匹配的方法来检索具有较长时间跨度的视频片段,同时,通过采用在整个帧内的运动矢量距离平方和作为计算特征距离的方法,使我们的检索方案能够较为精确地反映各种全局运动的差异.实验结果表明,在特定的应用领域,如体育类视频中,这种基于全局运动的检索方案能够实现一定程度的语义检索.

基于运动信息的视频检索的优势是,它可以提供使用其他图像特征所无法实现的检索功能,如在体育比赛视频中,场景的颜色和纹理特征通常是相当固定的,用它们作为检索特征就不能很好地区分不同的语义内容,而运动信息就可以弥补这方面的不足.当然,要单靠运动信息对一般的视频数据实现较好的语义检索还是相当困难的,在大多数情况下,运动信息是视频检索的一个重要辅助手段,需要结合其他的图像信息才有可能达到满意的检索效果,这也正是我们今后的研究方向.

参考文献:

[1] Ohm J R, Bunjamin F, Liebsch W, et al. A set of visual feature descriptors and their combination in a low-level description scheme [J]. Signal Processing: Image Communication, 2000, 16: 157 - 179.

[2] Chang S F, Chen W, Meng H J, et al. A fully automated content-based video search engine supporting spatiotemporal queries [J]. IEEE Trans. Circuits and Systems for Video Technology, 1998, 8(5): 602 - 615.

[3] 朱兴全,鲁 翔,薛向阳等. 一个基于运动的视频检索系统[J]. 模式识别与人工智能, 2000, 13(2): 159 - 163.

[4] Jeannin S, Jasinschi R, She A, et al. Motion descriptors for content-based video representation [J]. Signal Processing: Image Communication, 2000, 16: 59 - 85.

[5] Intel Corporation. Intel Image Processing Library Reference Manual [M]. USA: Intel Corporation, 2000. 11 - 16 - 27.

[6] Kim E T, Kim H M. Efficient linear three-dimensional camera motion estimation method with applications to video coding [J]. Opt. Eng. 1998, 37(3): 1065 - 1077.

[7] International Organization for Standardization, Description of MPEG-7 Content Set [S]. ISO/IEC JTC1/SC29/WG11/N2467. Atlantic City: ISO, 1998.

作者简介:

俞天力 男. 1977 年 10 月生于浙江杭州. 1999 年于上海交通大学电子工程系获学士学位. 现在清华大学电子工程系图像图形研究所攻读硕士学位,研究方向主要为视频的处理和分析、基于内容的视频检索等.

章毓晋 男. 1954 年 10 月出生于山西省太原市. 1989 年获比利时列日大学应用科学博士学位,从 1989 年至 1993 年在荷兰德尔夫特大学作博士后及研究工作. 1993 年到清华大学工作,现为图像图形研究所副所长,教授,博士生导师. IEEE 高级会员. 主要研究领域是图像工程(图像处理,图像分析,图像理解及其技术应用),已发表了 160 多篇研究论文,著有《图像分割》等书 4 本.