

自动分割及跟踪视频运动对象的一种实现方法

韩军 熊璋 孙文彦 龚声蓉

(北京航空航天大学计算机科学系, 北京 100083)

摘要 随着 MPEG-4 压缩标准的制定, 分割及跟踪视频运动对象的研究显得极其重要. 在 MPEG-4 视频编码标准中, 为了实现基于视频内容的交互功能, 其视频序列的每一帧由视频对象面(VOP)来表示. 为了生成视频对象面, 需要对视频序列中的运动对象进行有效的分割, 并跟踪运动对象随时间的变化, 为此提出并实现了一种用于分割及跟踪视频运动对象的时空联合方法. 该方法首先采用连续帧间差的 4 次统计量假设检验, 确定运动对象的位置, 自动地分离出运动区域与背景区域; 在运动区域内, 采用数学形态学的分水线算法来精确地提取运动对象的轮廓; 最后, 将提取到的运动对象作为模板, 对后续的视频序列, 用 Hausdorff 距离度量, 来跟踪并提取后续帧中运动对象. 实验结果表明, 该方法能有效地分割和跟踪视频运动对象, 且能有效减少计算复杂度, 其调整参数也较少.

关键词 视频分割 图象分割 跟踪运动对象 分水线算法 Hausdorff 距离

中图分类号: TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2001)08-0732-07

A Method for Implementation of Automatic Segmenting and Tracking of Video Moving Objects

HAN Jun, XIONG Zhang, SUN Wen-yan, GONG Sheng-rong

(Beijing University of Aeronautics and Astronautics, Dept. of Computer Science and Engineering, Beijing 100083)

Abstract With emerging of new video coding standard MPEG-4, segmentation and tracking of video moving object is a key research topic in the field of interactive multimedia. In order to support content-based functions in the visual standard of MPEG-4, each frame of video sequences should be represented in terms of video object planes (VOP's). Segmentation and tracking of video moving object is used for guiding the extraction of video object plane from the video sequence. This paper introduces spatio-temporal information to automatically segment and track video moving objects. Firstly, a fourth-orders statistic hypothesis testing in inter-frame difference is used to automatically separate the moving areas from the background in a general video sequence. Secondly, Change areas are extracted as markers of the corresponding moving object. Using the marker, watershed algorithm can locate the precise boundary of moving object. Lastly, using segment moving object as the initial model, moving object can be tracked by Hausdorff distance on a sequence of video, and the model is updated every frame to accommodate for changes in shape. Experimental results demonstrate that this method is effective to segment and track video moving objects. More, the number of parameters updated is the least.

Keywords Video segmentation, Image segmentation, Moving object tracking, Watershed algorithm, Hausdorff distance

0 引言

MPEG-4 压缩标准的制定, 使在互联网上提供

基于内容的交互视频服务成为可能. 分割及跟踪视频运动对象是 MPEG-4 标准能被应用的基础, 其价值主要体现在如下 4 个方面:

(1) 独立的视频运动对象能实现基于视频内容

基金项目: 国防科工委“九五”项目(43301·406), 航空基础科学基金(00F51057), “国家教委跨世纪人才计划”资助

收稿日期: 2000-07-05; 改回日期: 2000-11-29

的存取、粘贴及查询检索等功能；

(2) 分割出的视频运动对象不仅能单独进行压缩编码，而且还能有效提高视频压缩编码效率；

(3) 在现有的视频监控系统中，如能自动检测并跟踪运动对象，将大大提高视频监控系统的智能化程度；

(4) 在未来 Internet 应用中，将实现交互多媒体服务，如虚拟会议系统、虚拟聊天室、交互式视频教学系统等。

由此可见，对视频序列进行分割及跟踪视频运动对象是一个极其重要的研究领域。

一般分割视频运动对象的原则是依据视频序列中运动对象的时空一致性。视频分割通常采用的方法是基于光流场的运动参数估算^[1]，即通过估算运动参数，找出符合运动模型的象素区域，进而合并区域，来构成运动对象，但由于光流估算的可靠性较差，尚需验证光流估算的正确性和精度，为此又提出了贝叶斯分割法。虽然它能同时处理运动估算和分割问题，但这种方法计算较复杂，因此在实用化方面，不宜用于实时视频分割。

鉴于光流估算的视频分割存在上述问题，采用一种新方法来实现视频分割，其实现的原理如图 1 所示。

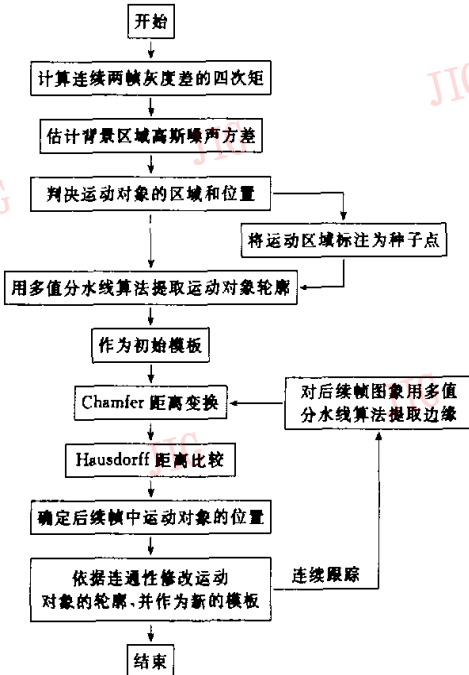


图 1 分割及跟踪视频运动对象原理图

该方法首先依据连续视频图象帧间差 4 次统计量的假设检验，自动地将运动区域与背景分离；然后标注运动的区域成为种子点，使用数学形态学的分水线算法来增长区域，这样即能精确地提取运动对象的轮廓；最后将初始分割的运动对象作为模板，使用 Hausdorff 距离度量方法来跟踪视频运动对象，进而提取运动对象。由于在视频序列中，运动对象的形状会发生旋转、变形等变化，因此依据运动对象轮廓的连通性，来对跟踪到的运动对象边界进行修改，并使其作为新的模板，再跟踪下一帧中的运动对象，直到某一帧中的运动对象从画面中消失。本文所采用的方法全部使用矩阵代数运算，计算效率较高，需调整的参数较少。

1 自动分割运动对象

1.1 自动分离运动区域与背景

视频序列的场景一般分为室内场景和室外场景^[2]。室内场景的背景变化主要是由噪声和亮度的变化引起的；室外场景的背景变化则主要由噪声和图象纹理的慢变化(如云、草、海、或树叶等)所组成。在分离运动区域与背景时，将视频图象中运动的一个或几个对象作为前景，而将静止不变或变化缓慢的区域作为背景。由于视频镜头产生的噪声符合高斯分布，且视频序列中的随机噪声、亮度变化、背景纹理的慢变化等统计量噪声一般也符合高斯特性，因此，如连续帧之间的灰度差为非零值，这可解释为是由噪声的影响或运动对象的变化引起的。一般噪声有高斯特性，而运动对象则有很强的结构性。采用互帧差的 2 次方作为判决时，由于需要考虑运动对象的旋转、局部变形及透视变化，用于实现运动对象与背景分离较复杂，因此，采用互帧差的 4 次统计量判决出运动对象的一般结构。具体地说，将运动区域与背景分离的原理就是在互帧差图象上逐点移动一个小的窗口区域，然后将每个窗口区域内 4 次互帧差和的平均值与视频图象中背景区域估计的高斯噪声的均方差的平均值进行比较，以确定前景区域和背景区域。

设连续帧灰度差图象为 $d(s, t)$ ，移动窗口的大小为 $N_s = 9$ ，计算窗口内的 4 次矩 $\hat{m}_d^{(4)}(x, y)$ ：

$$\hat{m}_d^{(4)} = \frac{1}{N_s} \sum_{(s,t) \in \mathcal{N}(x,y)} (d(s,t) - \hat{m}_d)^4 \quad (1)$$

其中， \hat{m}_d 是 $d(x, y)$ 窗口内的抽样平均，即

$$\hat{m}_d(x, y) = \frac{1}{N} \sum_{r(s, t) \in S} d(s, t) \quad (2)$$

然后再逐像素计算 $\hat{m}_d^{(i)}(x, y)$, 并与一个门限比较(这个门限与噪声变量 σ_{nd}^2 的平方成比例),

若 $\hat{m}_d^{(i)}(x, y) > c(\sigma_{nd}^2)^2$, 则为 H_1

若 $\hat{m}_d^{(i)}(x, y) < c(\sigma_{nd}^2)^2$, 则为 H_0

其中, H_0 表示静止背景, H_1 表示前景或覆盖/显露的背景, 而常量 c 在序列特性中近似为常数。对室内视频序列, 建议 $c=75$, 而在静止背景区域内的噪声均方差 σ_{nd}^2 按下式计算

$$\sigma_{nd}^2 = \frac{1}{N} \sum_{r(s, t) \in S} (d(s, t) - \hat{m}_d)^2 \quad (4)$$

这里, 噪声均方差 σ_{nd}^2 的计算应考虑图象是室外场景, 还是室内场景。对室外场景的背景, 由于是慢的纹理噪声, 因此采用共生矩阵法即可确定噪声门限。



(a) 帧测试原图



(b) 83帧测试原图



(c) 互帧差灰度图象



(d) 运动区域的二值图象

图2 自动分离运动对象的位置和区域图(采用测试序列 tennis)

1.2 运动对象轮廓的提取

使用上述的互帧差4次统计量判决方法虽然可分离出运动的区域与背景, 但提取的运动区域模糊、不完整, 也不能满足 MPEG-4 基于视频内容交互的要求, 而采用数学形态学的分水线算法则能有效、快速地提取运动对象的轮廓。

其中分水线算法^[3]由标记提取和判决两部分组成。标志提取就是将上述分离的运动区域大致分为 16×16 pixel 的均匀区域, 然后将每个小区域的中心点标注为种子点, 再在原视频图象上分割出运动对象的轮廓; 分水线算法的判决过程就是确定分割区域的边界点, 这是一个区域增长过程。其判决准则有两种, 一是采用图象的梯度算子; 二是采用彩色图象的多值均值算子。

本文仅考虑室内噪声的确定。对室内场景静止区域 S' 的选取应远离运动对象区域, 由于运动对象一般占据一帧图象的中央位置, 因此静止区域应在图象4个角选取。在具体实现时, 则是选取互帧差图象4个角的 6×6 pixel 大小的区域, 计算每个区域像素灰度的平均均方差, 再比较4个区域平均均方差, 将其最大值作为背景噪声均方差 σ_{nd} 。经过上述处理后, 图象中还存在一些小的噪声点, 此时可利用形态学的腐蚀算子(选取 3×3 结构元素)来滤出这些小的噪声点。图2是对标准的 tennis 测试序列采用4次互帧差自动分离的运动对象和背景。图2(a), 2(b)是 tennis 测试序列的第81, 83帧图象; 图2(c)是83帧与81帧的互帧差灰度图象, 图2(d)是用上述4次互帧差统计量判决获得的运动区域的二值图象(白色为运动区域点, 即检测到的83帧图象中的运动区域)。

梯度算子一般会产生许多由噪声或量化错误而引起局部小块, 而且使用梯度算子的分水线变换一般的结果是分割过大^[4]。基于彩色图象的多值分水线算法不同于基于梯度的分水线算法, 它不强调空间的一致性, 而是依据空间颜色的相似性。实现这个算法需要分两步计算: 首先, 评估多值表示的标志区域, 然后, 计算相邻像素点与多值表示值之间的差。实际上, 标志的表示被计算成平均彩色值, 距离函数定义为绝对距离。

$$d_i = |r - r_i| + |g - g_i| + |b - b_i| \quad (5)$$

其中 r_i, g_i, b_i 为一相邻像素点的红、绿、蓝分量, r, g, b 是一区域内所有像素红、绿、蓝分量的平均值。

基于运动对象轮廓提取的分水线算法具体实现过程如下:

(1) 首先将分离的运动区域分为 16×16 的小区域, 然后选每个小区域的中心点为种子点, 并将种子点放入一个数组中. 同时设置一个标记数组, 以确定运动区域内每个像素是否被标注.

(2) 做 DO 循环, 即从种子数组中读入一点, 并将这一点的位置 (x, y 坐标) 放入 FIFO 队列中.

(3) 从 FIFO 队列中输出一, 并检查它的 8 个邻域点, 如一个相邻点已被标注, 则证明这一点已属于另一个区域; 如这个相邻点没有被标注, 依据式(5)计算这一点的红、绿、蓝分量与这个区域内红、绿、蓝分量平均值的差, 然后将这个差与门限 δ_1 比较, 实验结果证明: 门限参数 δ_1 选取 30~60 较合理.

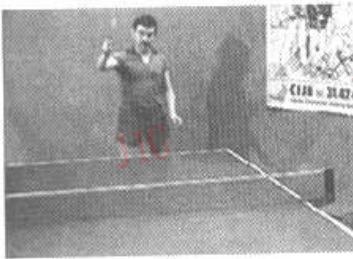
(4) 如上述计算的差小于门限 δ_1 , 则将这个相

邻点设为标记点, 并将其位置放入 FIFO 队列, 使之成为新的种子点. 如大于等于门限 δ_1 , 则将这点标注为边界点.

(5) 检查队列是否为空, 如为非空, 则转入第(3)步循环.

(6) 处理下一种子点直到所有的种子点均处理完, 再进行区域合并处理, 结果生成含边界标记的二值图象.

采用彩色多值分水线算法在运动区域内分割运动对象, 再利用区域合并算法将分割的小区域进行合并, 得到最终的结果如图 3 所示. 这种算法能较精确地提取对象的轮廓, 并且能自动形成闭合的区域, 易于后续的处理.



(a) 83 帧测试原图



(b) 运动对象轮廓

图 3 分割运动对象轮廓图

2 跟踪视频运动对象

2.1 基于 Hausdorff 距离度量的跟踪

在跟踪视频运动对象时, 将分割出的运动对象作为初始模板, 在后续帧中实现自动跟踪运动对象的变化, 并自动地跟踪后续各帧中运动对象轮廓的变化, 直到这一运动对象从画面中消失. 在一般的视频序列中, 传统的运动对象跟踪方法是采用光流运动参数模型, 但它只能跟踪运动对象的轨迹, 而不能很好地跟踪运动对象轮廓的变化, 因为运动对象常会被遮挡, 其形状会发生变形. 所以文本采用 Hausdorff 距离度量的方法来跟踪视频运动对象的变化, 这种方法即使当运动对象被遮挡、其形状发生慢的变化、背景移动时, 也能精确跟踪到运动对象运动后的新变化. Hausdorff 距离度量的跟踪原理如下:

若给定两组有限像素点集合 $A = \{a_1, \dots, a_p\}$ 和 $B = \{b_1, \dots, b_q\}$, 则 Hausdorff 距离定义为

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (6)$$

其中

$$h(A, B) = \max_{a_j \in A} \min_{b_j \in B} \|a_j - b_j\|,$$

$$h(B, A) = \max_{b_j \in B} \min_{a_i \in A} \|b_j - a_i\|$$

式中, $h(A, B)$ 称为从 A 到 B 的有向 Hausdorff 距离. 它的意义是从集合 A 的任意一点 $a \in A$ 到集合 B 中所有点距离中找出最近距离, 因此, 针对集合 A 的所有点, 则生成一个到集合 B 的最近距离集合, 再从这个集合中选择最大值作为 $h(A, B)$ 的值. $h(B, A)$ 的意义同 $h(A, B)$ 相似, 称为反向 Hausdorff 距离. Hausdorff 距离度量会发生这样一种情况, 如果集合 A 的形状与集合 B 的形状是相似的, 但是集合 A 中有一部分点是干扰点或属于背景点, 偏离相似形状区域, 那么计算出的 Hausdorff 距离 $h(A, B)$ 会很大, 因此采用局部 Hausdorff 距离

$$h_i(A, B) = K_{a_i \in A}^n \min_{b_j \in B} \|a_i - b_j\| \quad (7)$$

其中, $K_{a_i \in A}^n$ 是集合 A 的每一点到集合 B 的最近距离集合按从小到大的排序, 由此即将距离集合中第 K 个等级值 (对应于 A 的每个元素) 作为局部 Hausdorff 距离. 从上述分析可见, 局部 Hausdorff

距离对于干扰点或背景点相对不敏感;允许形状的一部分与另一个图象匹配。

在实际计算 Hausdorff 距离时,是将已知模板二值图象与将匹配的未知二值图象转换为距离函数。这种距离函数是将二值图象转换为另一种灰度图象,其中二值图象的“1”,对应于距离图象的“0”;而二值图象的“0”,依据图4的距离变换系数转换为不同的值,在距离图象中离“0”越近,距离值越小,反之会很大。具体的推导见参考文献[5]。

8	5	4	5	8
5	2	1	2	5
4	1	0	1	4
5	2	1	2	5
8	5	4	5	8

图4 距离变换系数

一般地,已知图象 B 的大小小于要匹配的未知图象 A 的大小。设已知模板图象为 B ,其二值图象矩阵为 $B[k,l]$,距离变换为 $D[x,y]$ 。如要匹配的未知二值图象为 A ,其二值图象矩阵为 $A[k,l]$,距离变换为 $D'[x,y]$,则

$$F[x,y] = \max(F_B[x,y], F_A[x,y]) \quad (8)$$

其中

$$F_B[x,y] = \max_{k,l} B[k,l]D'[b_x + x, b_y + y],$$

$$F_A[x,y] = \max_{\substack{(k,l) \\ \begin{matrix} 1 \leq k \leq m+1 \\ 1 \leq l \leq n+1 \end{matrix}}} A[k,l]D[k-x, l-y]$$

而局部 Hausdorff 距离变换为

$$F[x,y] = \max(F_A[x,y], F_B[x,y]) \quad (9)$$

其中

$$F_A[x,y] = K_{k \in B}^h D'[b_x + x, b_y + y]$$

$$F_B[x,y] = L_{a \in A}^h D[a_x - x, a_y - y]$$

$F_B[x,y]$ 是相当于一个矩形区域的二值图象 B 在 A 的距离图象 $D'[x,y]$ 上的逐点移动,在距离图象 D' 的任一位置 (x,y) 上,二值图象 B 的“1”象素点对应于 D' 上相应点的距离值,而二值图象 B 中所有的“1”象素点,构成对应于 D' 距离的集合,从这个距离集合中选出最大值作为这一点的 $F_B[x,y]$ 值。如果在 (x,y) 位置上,图象 B 与图象 A 的一个区域很相似,那么这个最大值应该很小。

反之, $F_A[x,y]$ 是 B 的距离图象 $D[x,y]$ 在二值图象 A 上的移动,在二值图象 A 的任一位置 (x,y) 上,相当于 B 的距离图象 D 覆盖图象 A 的区域内,

图象 A 的所有“1”象素点对应距离图象 D 的所有距离值集合中,选取最大值作为 $F_A[x,y]$ 。如果在 (x,y) 位置上,图象 B 与图象 A 的一个区域很相似,那么这个最大值也应该很小。

最后,将这一点的 $F_B[x,y], F_A[x,y]$ 二者的最大值作为这一点的 Hausdorff 距离 $F[x,y]$ 。由此可见, Hausdorff 距离度量的方法实质上就是从最不相似的形状区域中找出最相似的形状区域,是一种模糊匹配的方法。

2.2 基于 Hausdorff 距离度量的算法实现

实现跟踪运动对象的算法是首先将分割出的运动对象作为初始模板图象,然后使用 Hausdorff 距离度量来找出后续帧中的运动对象,修整其形状作为新模板,这样就能实现连续跟踪,直到运动对象从画面中消失。基于 Hausdorff 距离跟踪视频运动对象的算法如下:

(1)将初始分割的运动对象二值图象作为模板图象 B 。

(2)对后续帧图象,用基于彩色的多值分水线算法分割图象,生成边缘的二值图象,作为 Hausdorff 距离比较的二值图象 A 。

(3)用图4的距离变换系数,将二值图象 A 和 B 变换为距离图象 D' 和 D 。

(4)逐点扫描图象 A ,并对其任一点 (x,y) 进行计算,再对图象 B 中所有“1”的象素点 (k,l) ,查找相应 A 的距离图象中 $(x'+k, y'+l)$ 点的值,放入数组 $Sort$ 中。

(5)计算数组 $Sort$ 中所有值的平均值,作为点 (x', y') 的有向 Hausdorff 距离值,放入数组 F_B 中。

(6)考虑在点 (x', y') 处,图象 B 覆盖图象 A 的区域,对图象 A 在这个区域内所有“1”的象素点 (k_1, l_1) ,查找相应距离图象 B 的 $(k_1 - x', l_1 - y')$ 点的值,放入数组 $Sort_1$ 中。

(7)计算数组 $Sort_1$ 中所有值的平均值,并将其作为点 (x', y') 反向 Hausdorff 距离值,放入数组 F_A 中。

(8)比较在点 (x', y') 上, F_B 和 F_A 的大小,将较大的值放入数组 F 中。

(9)返回到第(4)步,计算图象 A 所有点的 F 值。

(10)查找数组 F 中的最小值,并且需小于一个门限(门限值为2或3)。如满足条件,则该点作为匹配对象的所在位置。

(11)依据图象的连通性,修改图象的轮廓,即跟踪对象轮廓的变化,然后将其作为新的模板来配下

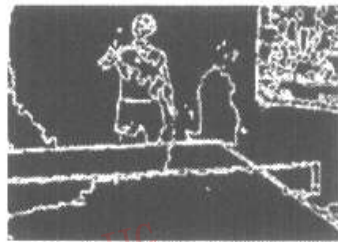
一帧图象中运动对象的变化。

图5是采用 Hausdorff 距离度量跟踪运动对象的实验图。图5(a)是用互帧差4次矩统计量判决和基于彩色的多值分水线算法分割的运动对象,将其作为已知的初始模板二值图象,图5(b)是用基于彩色的多值分水线算法提取 tennis 测试序列 85 帧图象的边缘,图5(c)是用 Hausdorff 距离度量跟踪

到的后续帧(85帧)中运动对象,位置锁定在(83,2)这一点,图5(d)是将跟踪到的运动对象图利用运动对象的连通性进行轮廓修整,并将其作为新的模板图象,跟踪下一帧。实验结果表明,使用 Hausdorff 距离度量来跟踪运动对象是一种极其有效、快速的方法,而且 Hausdorff 距离度量方法在抗图象的噪声干扰、局部遮挡方面显示出有很好的性能。



(a) 初始分割的运动对象



(b) 用分水线算法提取 85 帧原图轮廓



(c) Hausdorff 跟踪运动对象,位置锁定在(83,2)



(d) 经轮廓修正后运动对象

图5 Hausdorff 距离跟踪运动对象图

3 结论

本文从实现视频分割的系统出发,研究了视频分割所涉及的技术问题。由于视频分割是针对任意的视频序列,因此,其关键是在初始时能自动地将前景运动对象和背景分离。本文采用快速的分水线算法精确地分割图象和基于帧间差图象的噪声特性,这两种方法融合来将运动对象与背景分离,再用 Hausdorff 距离度量来跟踪运动对象的变化,实验结果证明,这种方法在抗干扰、运动对象被局部遮挡的情况下是极其有效的。再者算法的实现完全采用代数的运算,因此计算量较少,需调整的参数也较少。

下一步应努力的方向是:从视频分割的实用化、工程化角度来考虑,优化现有的算法,并充分利用计算机硬件处理图象能力和充分研究视频分割的实时性要求。还需从人工智能、图象理解及视频内容的先验知识上考虑,以便将上述方法推广到任意的视频

序列的应用上。

参考文献

- 1 Thomas Meier, King N. Ngan. Automatic segmentation of moving objects for video object plane generation. IEEE Transaction on Circuits and Systems for Video Technology, 1998,8(5):525~538.
- 2 Neri A, Colonese S, Russo G *et al.* Automatic moving object and background separation. Signal Processing, 1998,66(2):219~232.
- 3 Luc Vincent, Pierre Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991,13(6):583~598.
- 4 Philippe Salembier, Montse Pardas. Hierarchical morphological segmentation for image sequence coding. IEEE Transactions on Image Processing, 1994,3(5):639~651.
- 5 Huttenlocher D P, Klanderman G A, Rucklidge W J. Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993,15(9):850~863.



韩 军 1965年生,高级工程师,1990年获西安电子科技大学通信与电子系硕士学位,现为北京航空航天大学博士研究生.曾获航空工业部科技进步一等奖一项.目前主要研究方向为视频处理和视频传输.



孙文彦 1973年生,现为北京航空航天大学计算机科学与工程系博士研究生.主要研究方向为分布式多媒体系统、模式识别.



熊 璋 教授,博士生导师,曾获国家科技进步一等奖、航空工业部科技进步一等奖各一项.目前主要研究方向为多媒体技术和分布式信息处理.发表论文40余篇.



龚声誉 1966年生,副教授,现为在职博士生.研究领域为基于内容的图象查询和分布多媒体信息系统.

投 稿 须 知 (A 版)

1. 为加快审稿速度,我刊要求来稿一律一式三份,其中一份稿件务必写明全部作者真实姓名及简介,单位及部门,中英文摘要(中文摘要需写成200字左右的篇幅,包括研究目的、方法、结果和结论,英文摘要与中文内容一致,字数在200个英文单词),其它两份稿件要求双面(正反面)打印,并略去中英文作者姓名、单位地址,作者简介.
2. 无论稿件录用与否,本刊均将审稿意见寄给作者,不退原稿,请自留底稿.
3. 若方便,来稿时请注明您的E-mail及联系电话,以便今后联系.
4. 本刊暂不接收电子投稿.多谢您的合作!

通信地址:100080 北京 2704 信箱 《中国图象图形学报》编辑部收

《中国图象图形学报》编辑部