

视频序列中的镜头渐变识别

钟秋, 潘琪, 姚佩阳

(空军工程大学电讯工程学院, 陕西西安 710077)

摘要: 淡入、淡出和消隐的共同特点是两个镜头之间的切换一个渐变的线性过程。渐变镜头切换识别是视频后台处理和视频数据库检索的关键技术之一。文中提出了自动检测视频序列中渐变镜头切换的数据驱动方法, 首先建立了统一的渐变镜头切换模型, 然后根据帧间差的统计特征, 提出了淡入、淡出和消隐镜头切换识别的算法, 最后给出了算法的实验结果。实验表明该算法具有较好的识别效果。

关键词: 镜头切换识别; 淡入; 淡出; 消隐

中图分类号: TN850.1 文献标识码: A 文章编号: 1009-3516(2005)04-0077-04

目前, 镜头突变的检测方法很多, 特别是利用相邻帧之间颜色直方图差^[1], 在精确度和速度的综合上达到很好的效果, 识别率一般都在90%以上, 但是这些方法一般都没有考虑渐变的情况。由于单纯能识别突变镜头在一个实际的视频检索系统中意义不大, 因此, 目前, 研究者越来越关注能够同时能识别切换和多种渐变的方法。由于淡入、淡出和消隐的构造原理相同, 在这里我们统称为渐变。镜头渐变的识别比镜头突变的识别要复杂得多, 关于镜头渐变检测方法大致可以分为基于像素统计特征、基于边缘和基于数学模型的方法^[2-3], 但都有其局限性。本文提出了一种新的自动识别镜头渐变的方法, 最后给出算法的试验结果和分析, 实验表明该算法具有较好的识别效果。

1 数据驱动方法

1.1 数学模型描述

假设 $g(x, y, k)$ 是某镜头的最后一帧, $E(x, y, t)$ 是视频变换所生成的帧序列, l_s 是该帧序列的长度(帧数), 则

$$E(x, y, t) = g(x, y, k) \times (1 - t/l_s) \quad (1)$$

$E(t)$ 对时间的微分为

$$C(t) = \frac{\delta E / \delta T}{g(x, y, t)} = -\frac{1}{l_s} \quad (2)$$

$C(t)$ 称为一常数图像序列, 该常数值与颜色变换率 $(1/l_s)$ 成正比。这样对渐变的识别问题变成了判断 $C(t)$ 为一常数图像序列的问题。

消隐镜头构成中主要是对图像灰度的特定处理, 因此我们取每帧的平均灰度作为该帧的特征

$$V_r = \frac{1}{N_r} \left(\sum_{(i,j) \in B} v(i, j) \right) \quad (N_r \text{ 为 } B \text{ 中像素的个数}) \quad (3)$$

本文基于渐变镜头的切换, 像素灰度在切换时间段内表现为单调而且通常是线性的变化^[5]。我们的问题可以描述如图1所示, 图中实线段表示镜头的渐变过程, 对消隐镜头切换, 对应第一条线段, 这样对于渐变的实现可以用统一的公式表示。设0表示全黑图像, 则淡出变换是 $S_{out} = g_1, S_{in} = 0$; 淡入是 $S_{out} = 0, S_{in} = g_1$;

收稿日期: 2005-02-28

作者简介: 钟秋(1979-), 男, 辽宁大连人, 硕士生, 主要从事信号与信息处理技术研究
姚佩阳(1968-), 男, 陕西西安人, 教授, 主要从事信号与信息处理技术研究。

而隐现则是淡出和淡入的综合,即 $S_{out} = g_1$ 、 $S_{in} = g_2$ 。若 E_{f_o} 表示由淡出变换生成的帧序列, l_1 表示其帧数; E_{f_i} 表示由淡入变换生成的帧序列, l_2 表示其帧数; E_d 表示由淡入淡出变换生成的帧序列, 则

$$E_{f_i}(x, y, t) = 0 + g_2(x, y)(t/l_2) \tag{4}$$

$$E_{f_o}(x, y, t) = g_1(x, y)((l_1 - t)/l_1) + 0 \tag{5}$$

$$E(x, y, t) = E_{f_o}(x, y, t) |_{(t_1, t_1+l_1)} + E_{f_i}(x, y, t) |_{(t_2, t_2+l_2)} = g_1(x, y)(1 + t \times C(t)) |_{(t_1, t_1+l_1)} + g_2(x, y, t)(-t \times C(t)) |_{(t_2, t_2+l_2)} \tag{6}$$

其中, t_1 和 t_2 分别是 g_1 和 g_2 开始变换的时间, l_1 和 l_2 是 g_1 和 g_2 变换所持续的帧数。淡入和淡出可以用式(4)和式(5)进行识别, 消隐的情况要复杂一些, 因为 g_1 变换(淡出)和 g_2 变换(淡入)的开始时间和变换率不同。其中, x 表示视频帧序列号, y 代表像素的平均灰度。

用 $S(F_i, F_{i+1}, \dots, F_{i+n})$ 表述渐变序列, 对应图 1 中的渐变区间的点可以用以下特征点表示:

$$D(D_i, D_{i+1}, \dots, D_{i+n}), D_i = (i, V_i) \tag{7}$$

其中 i 表示帧号, V_i 如公式(3)所示, 本文中我们采用高斯归一法, 使 V_i 都落在 0 和 1 区间, 于是渐变镜头切换(淡入、淡出和消隐)可以用近似的线段表述。理想状态下, 如果通过阈值 T_n 识别渐变的开始两帧 F_i, F_{i+1} , 则 F_i, F_{i+1} 组成一条特征线如图 2 所示,

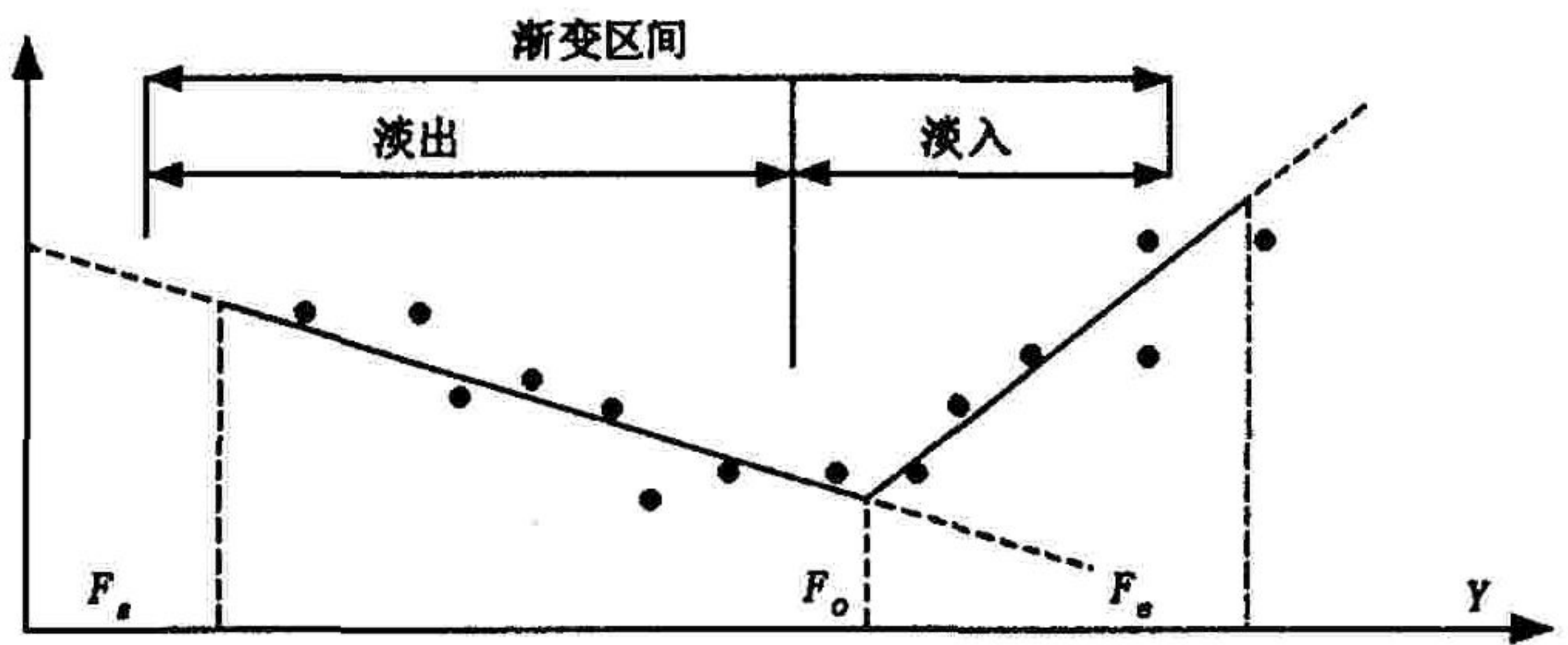


图 1 渐变镜头切换示意图

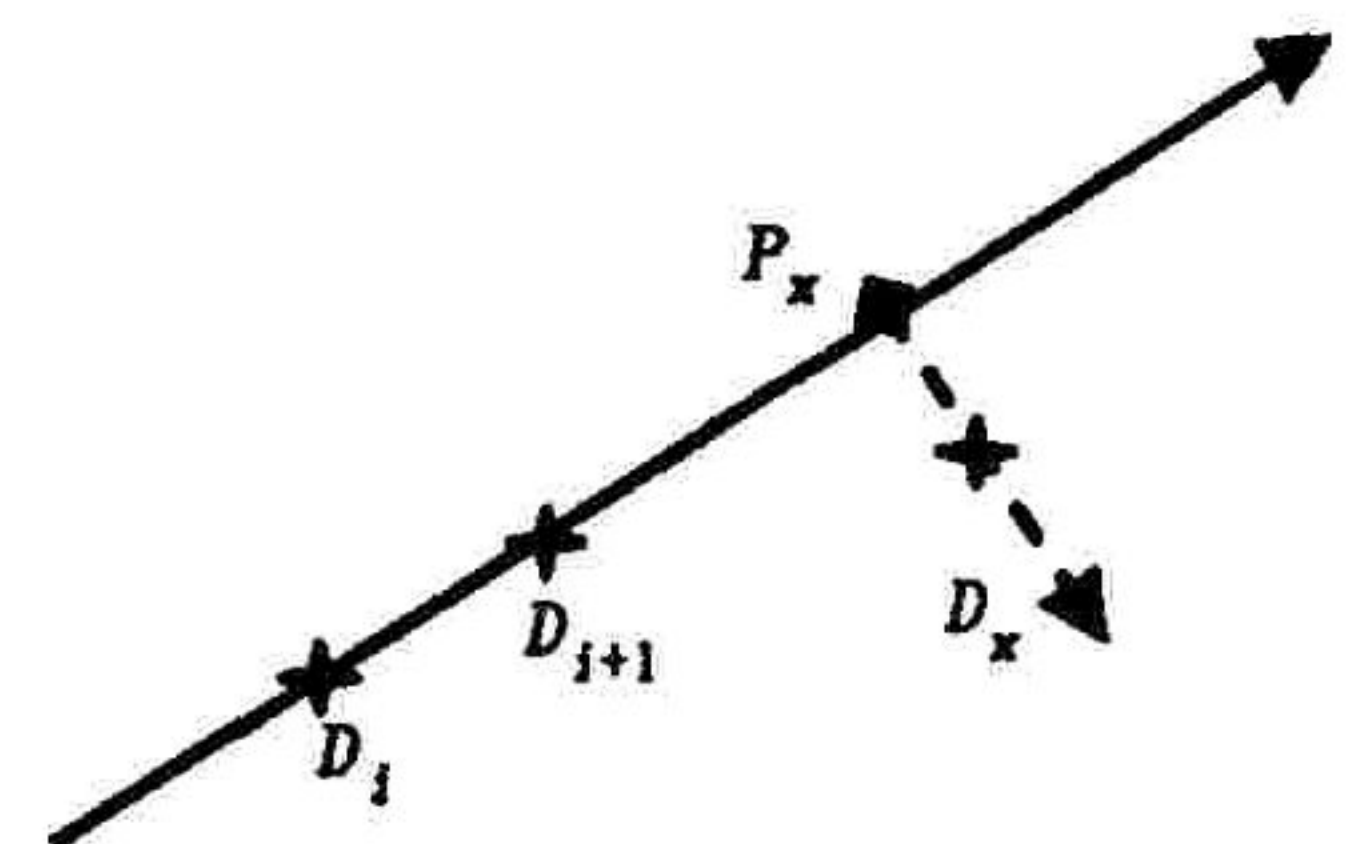
渐变镜头的帧序列都在这条特征线附近, 点 D_x 到特征线 $\overline{D_i D_{i+1}}$ 的距离可定义为 $Dist(f_x, \overline{D_i D_{i+1}})$ 。用 p_x 表示 D_x 在特征线 $\overline{D_i D_{i+1}}$ 上的投影, 则

$$p_x = D_i + \mu(D_{i+1} - D_i) \tag{8}$$

其中

$$\mu = \frac{(D_x - D_i)(D_{i+1} - D_i)}{(D_{i+1} - D_i)(D_{i+1} - D_i)} \tag{9}$$

所以 $Dist(f_x, \overline{D_i D_{i+1}}) = \|D_x - p_x\|$ ($\|\cdot\|$ 表示欧氏距离) $\tag{10}$



设视频中的一帧 D_x 在特征线上对应 f_x (见图 2), 则通过计算该点与特征线的距离 $Dist(f_x, \overline{D_i D_{i+1}})$ 来决定该帧是否属于渐变序列, 对于给定的阈值 δ_{Dist} , 如果 δ_{Dist} 大于或等于 $Dist(f_x, \overline{D_i D_{i+1}})$, D_x 就属于渐变序列, 所以, 阈值 δ_{Dist} 的选取也是要解决的问题。

1.2 结合渐变镜头识别

在实际情况下, 并不象上面讨论的那样, 所有渐变序列中的特征点恰好分布在由 (D_i, D_{i+1}) 组成的特征线 $\overline{D_i D_{i+1}}$ 上, 而是分布在如图 3 所示的特征线 L_i 附近, 如果只是简单的以 (D_i, D_{i+1}) 组成的特征线作为整个渐变序列特征点的参考, 那么后面的特征点与该特征线的距离 $Dist(f_x, \overline{D_i D_{i+1}})$ 会渐变变大, 这样势必会造成很大的识别误差。这里我们采用不断调整特征线的方法来解决这个问题, 首先定义一个集合

$$S^c = \{ D_t | i \leq t \leq i+1 \} \tag{11}$$

集合 S^c 的初始元素为 D_i 和 D_{i+1} , 如图 4 所示, 增加一个特征点 D_{i+2} , S^c 就增加一个元素, 然后再由 S^c 中的 3 个点重新构造特征线 3, 依次在增加一个点, 重新构造特征线 3, 最后使 S^c 中所有的点构造的特征线就是最终所求的特征线 L_i , 每次重新构造特征线时, 采用曲线拟合的最小二乘法。

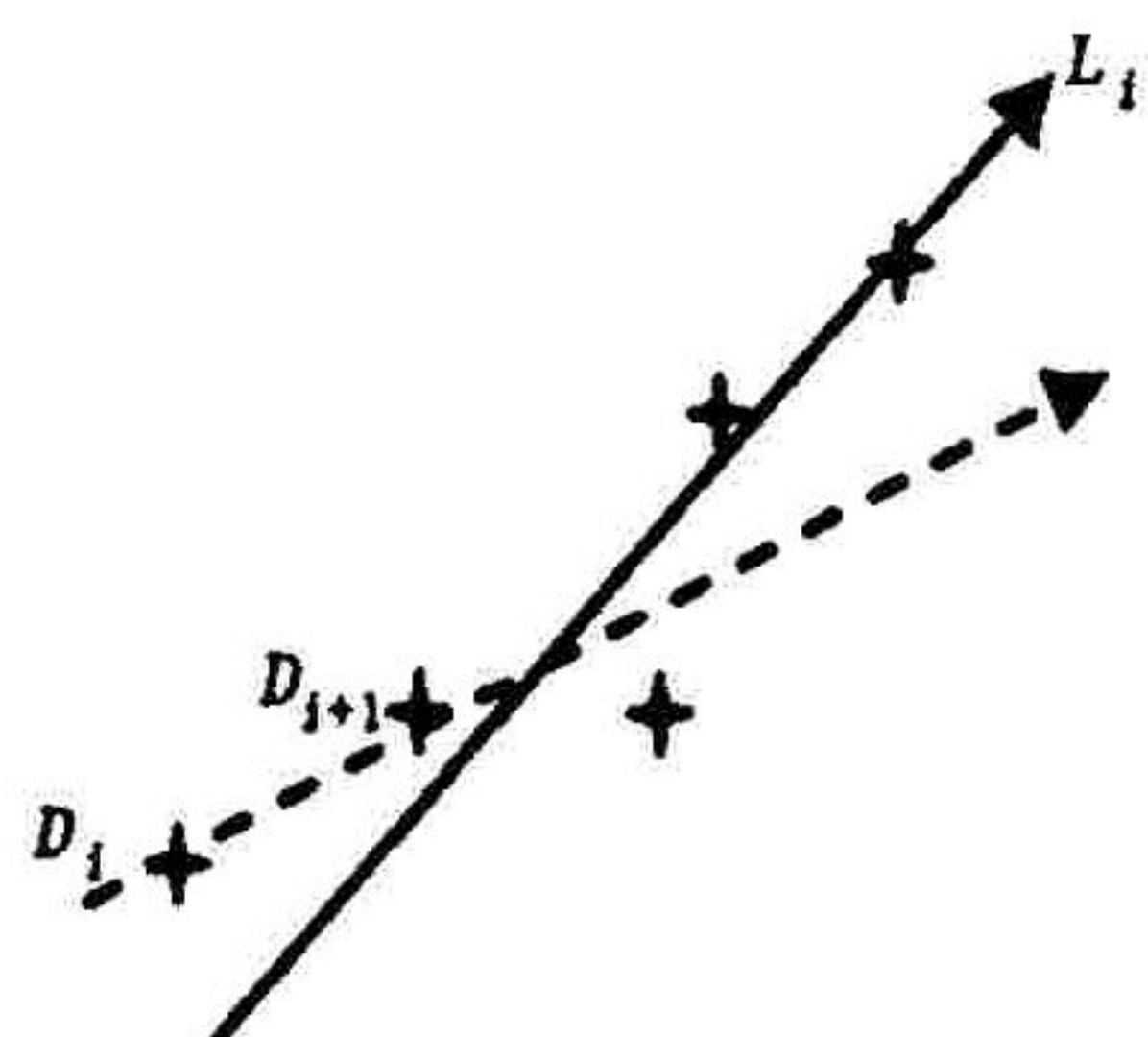


图 3 渐变镜头序列中开始的两个特征点构成的特征线

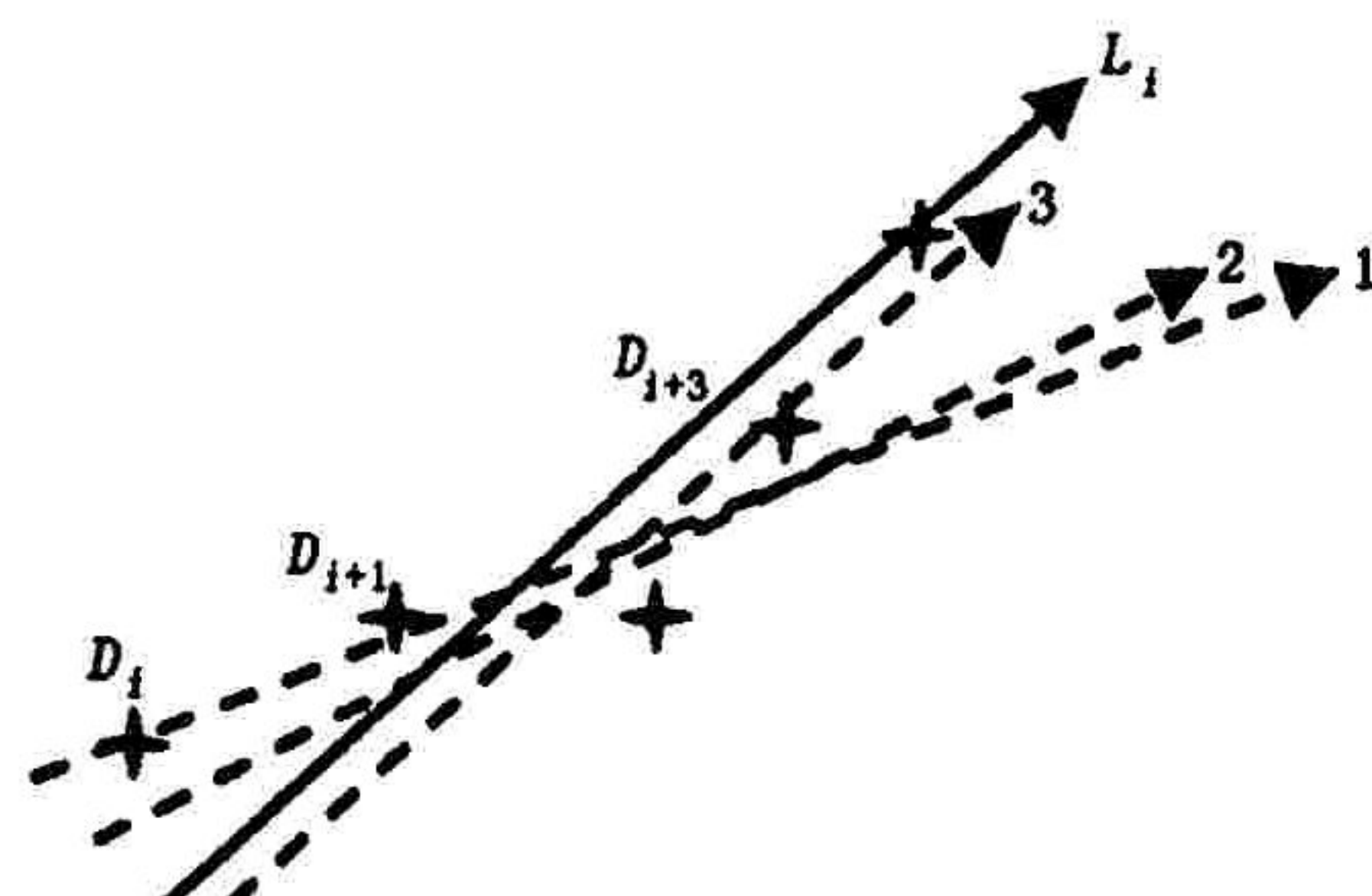


图 4 随着特征点的增加不断调整特征线

对于淡入和淡出,通过比较点到特征线的距离和阈值 δ_{Dist} 来判断镜头切换的结束,对于消隐切换,如式(6)和(7)描述的那样,采用淡出和淡入合成消隐区间。可以这样检测消隐区间,每次在淡出结束后,假设后面出现的视频帧为淡入序列,我们依次取几个特征点,构成一个特征线 L_{temp} ,这几个特征点我们称为“缓冲特征点”,然后计算每个特征点与 L_{temp} 的距离,如果有一个“缓冲特征点”与 L_{temp} 的距离大于阈值 δ_{Dist} ,就认为淡出后面的区间不是淡入,所以不构成消隐区间。图5是我们通过实验,得出“缓冲特征点”个数与淡入、淡出和消隐区间识别率的关系,其中横轴代表“缓冲特征点”的个数,纵轴代表识别率。可以看出,当“缓冲特征点”个数取4时,效果最佳,在我们的算法中缓冲特征点个数取4。通过上面的讨论,下面给出具体镜头渐变识别算法:

输入:视频序列 $Frames = \{Video_1, Video_2, \dots, Video_M\}$

输出:渐变序列 $Graduals = \{Gradual_1, Gradual_2, \dots, Gradual_N\}$

算法:第1步:初始化渐变序列集合 $S^G = 0$,计算视频序列 $Frames$ 中第1帧和第2帧的帧间差 $d_K, K = 2; d_K$ 代表图像的帧间差,是一种计算图像间距离的方法;第2步: $K = K + 1$,计算视频序列 $Frames$ 第 K 帧和 $K - 1$ 帧的帧间差 d_K ,如果 d_K 小于等于阈值 T_n ,则第 K 帧和 $K - 1$ 帧作为渐变序列的初始两帧,构造特征线 L_K 。第3步: $K = K + 1$,选取后续特征点和初始特征点重构特征线,并计算选取的特征点到特征线 L_K 的距离 $Dist(D_K, L_K)$,如果 $Dist(D_K, L_K)$ 大于阈值 δ_{Dist} ,渐变序列结束,把该渐变序列添加到集合 S^G 。否则返回到第2步。这里 T_n 和 δ_{Dist} 是由用户预先设定好的阈值,在我们的实验中 T_n 和 δ_{Dist} 是位于0到1之间的常数, $T_n = 0.125, \delta_{Dist} = 0.152$;第4步:如果视频序列中还有元素,则返回到第2步,否则,退出。

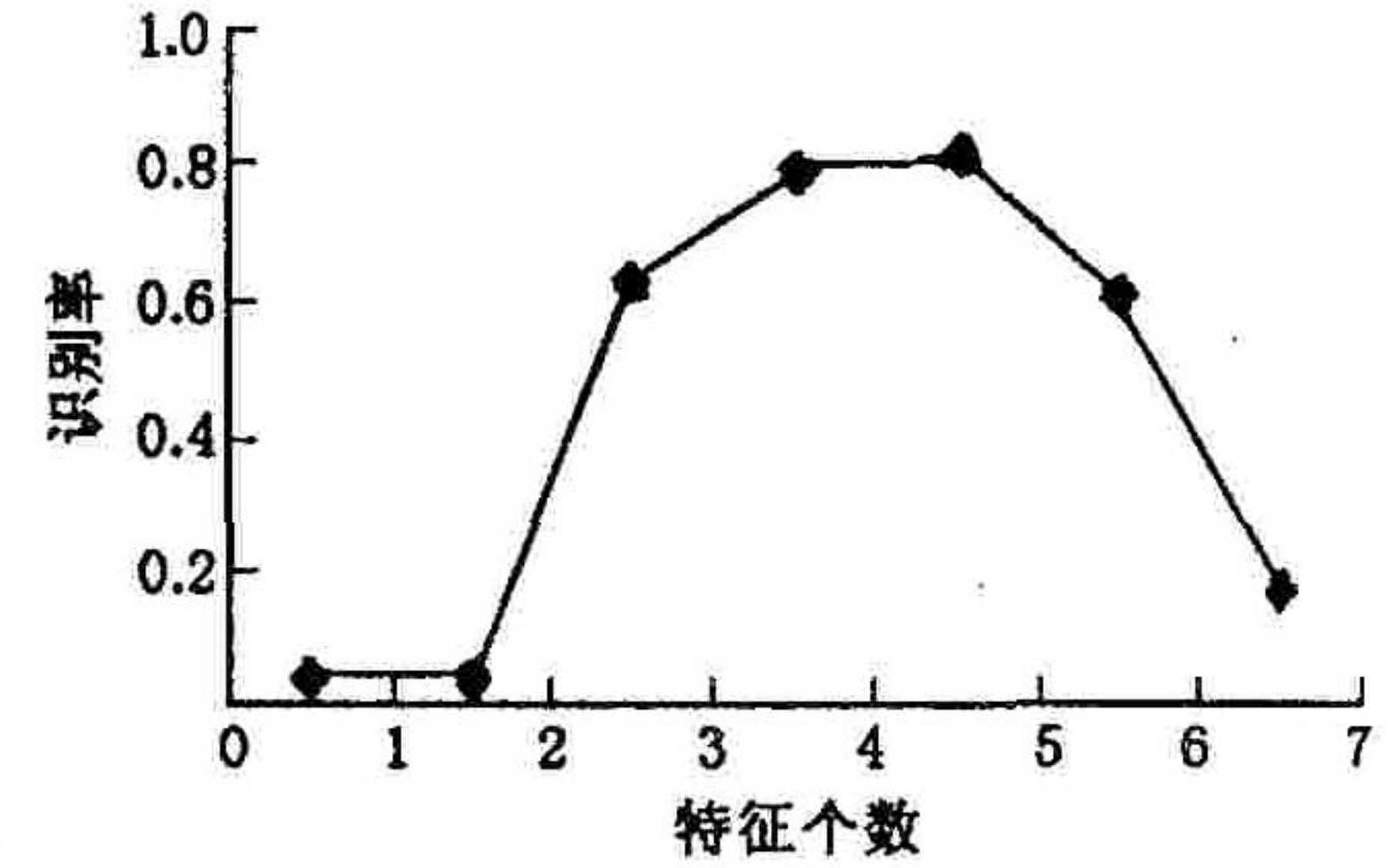


图5 “缓冲特征点”的个数与淡入/淡出和消隐区间识别率的关系

2 实验结果与分析

基于上述的渐变镜头识别算法,我们实现了一个原型系统,在此基础上对一些视频片段进行了实验以测试算法的有效性。实验选了段视频对算法进行了测试。这3段测试视频完全选自实际的视频片断,而且片断2和片断3是两段MTV,它们中实用的镜头渐变切换较多,并且有很强的光照、文字颜色等多种变化干扰。实验中发现,阈值 T_n 对识别率有很大的影响,图6是对4段渐变镜头序列选取不同的 T_n ,对识别结果影响的曲线图,可以看出 T_n 在0.10-0.15之间的识别效果较好,取其平均值 $(0.10 + 0.15)/2$ 作为阈值,所以选择合适的阈值,也是试验结果取得最佳效果的必要条件。通过对视频素材进行测试,得出如表1和表2的实验数据。

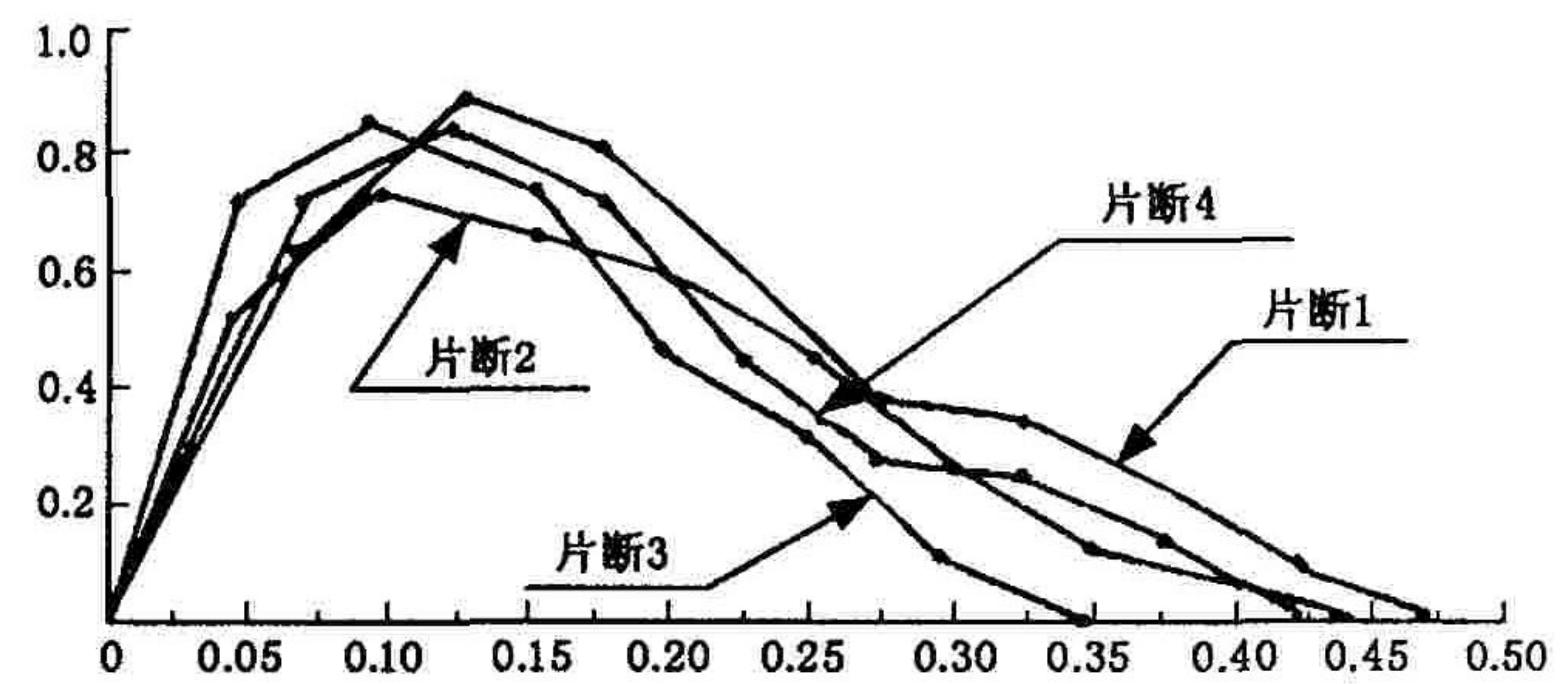


图6 阈值 T_n 对渐变镜头识别的影响

表1 对视频1算法检测出的消隐区间和实际消隐区间

消隐区间	1	2	3	4	5	6	7	8	9
实际	65-105	205-248	381-408	565-600	745-775	789-815	850-881	1 015-1 035	1 062-1 085
检测	70-118	211-249	399-421	577-596	747-778	797-820	857-889	1 015-1 056	1 066-1 082

表2 对视频2算法检测出的淡入/淡出区间和实际淡入/淡出区间

淡入/淡出区间	1	2	3	4	5	6	7
实际	49-58	2 920-2 950	3 371-3 400	4 429-4 450	5 012-5 042	5 760-5 775	6 110-6 140
检测	54-66	2 940-2 954	3 386-3 398	4 420-4 457	5 009-5 040	5 768-5 778	6 115-6 138

由这些数据,可以看出不论是消隐区间还是淡入区间,算法检测的结果总是比实际渐变区间少几个开始渐变帧,另外,在实际渐变结束后,又增加了几帧。一个原因是与阈值 T_n 和 δ_{Dist} 的取值有关,另一原因是镜头变换检测算法的效果目前主要用查全率(recall, R)和准确率(precision, P)两个指标,计算公式如下:

查全率是正确检测数 N_c 与正确检测数及漏检数 N_m 之和的比:
$$R = \frac{N_c}{N_c + N_m} \quad (12)$$

准确率是正确检测数 N_c 与正确检测数及误检数 N_f 之和的比:
$$P = \frac{N_c}{N_c + N_f} \quad (13)$$

对表 1 和表 2 的实验数据进行综合分析,使用公式(13)和(14)来表示查全率和准确率,结果如表 3 所示。可以发现,算法是比较有效的,同时误检率只有 8.3%。数据显示视频 2 比视频 1 和 3 由较高的查全率,因为视频 2 的都是淡入和淡出镜头切换,而视频 1 主要是消隐切换,前面讨论过,消隐切换比较复杂,我们用淡入和淡出来合成(merge),这样在淡出结束后,如图 1 所示,存在一个拐点,所以对拐点的判断正确与否,影响对消隐切换是别的查全率。视频 3 镜头切换包括消隐切、淡入和淡出切换,其查全率在视频 1 和 3 之间,也符合数学规律。

表 3 消隐镜头边界检测算法的检测结果

评价参数	消隐 个数	正确检 测个数	漏检 个数	错检 个数	查全率 (%)	查准率 (%)	漏检率 (%)	误检率 (%)
视频 1	11	9	2	0	81.8	100	18.2	0
视频 2	8	7	2	1	87.5	77.8	12.5	12.5
视频 3	32	27	8	4	84.4	87.1	15.6	12.5
平均值	17	14	4	2	84.6	88.3	15.4	8.3

参考文献:

- [1] Hsu P R, Harashima H. Detecting Scene Change and Activities in Video Database[J]. ICASSP94, 1994, (5): 33 - 36.
- [2] Zhang H J, Kankanhali A, Smoliar S W. Automatic Partition of Animate Video[J]. Tech Rept Institute of System Science National University of Singapore, 1997, (7): 11 - 17.
- [3] Hampapur A, Jain R C, Weymouth T. Production Model Based Digital Video Segmentation[J]. Multimedia Tools and Applications, 1995, (1): 9 - 46.

(门向生)

Gradual Shot Change Recognition in Video Sequences

ZHONG Qiu, PAN Qi, YAO Pei -yang

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

Abstract: There is a common characteristic for fade - in, fade - out and dissolve, and it is a linear gradual process in shot change. It is important to detect the gradual shot change in order to improve the video background process and allow automatic parsing of the video for the purpose of editing and database indexing. In this paper, a data - drive method of gradual shot change recognition, such as fade - in, fade - out and dissolve, is given. A model is built for gradual shot change, and then a recognition algorithm based on frame - difference statistical characteristics is presented, finally, evaluations and experimental results are presented.

Key words: shot change recognition; fade - in; fade - out; dissolve linearity gradual change