一种利用人物相似度的视频索引算法

王 鹏[1], 马宇飞[2], 张宏江[2], 杨士强[1]
(1. 清华大学计算机科学与技术系, 北京 100084; 2. 微软亚洲研究院, 北京 100080)

摘 要: 本文面向视频分析和索引技术, 提出一种利用人物相似度进行视频索引的算法. 该算法应用 SVMs 概率输出理论, 将底层特征空间距离映射为语义层人物相似度, 并提出一种新的非监督聚类算法, 修正部分误判, 最终实现对视频节目中人物的自动聚类和索引. 实验结果表明, 该算法实用而高效, 是对现有视频索引算法的有效补充.

关键词: 视频分析和索引; 人物相似度; 机器学习; 非监督聚类
中图分类号: TP37 文献标识码: A 文章编号: 0372-2112 (2004) 06-0968-05

A Novel Video Indexing Approach Based on People-Similarity

WANG Peng[1], MA Yu-fei[2], ZHANG Hong-jiang[2], YANG Shi-qiang[1]
(1. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;
2. Microsoft Research Asia, Beijing 100080, China)

Abstract: This paper presents a novel approach to video indexing. In this approach, we define a people-similarity measure according to both clothing similarity and speaking voice similarity. Such similarity depicts how perceptually similar two people appearing in different scenes are and whether they belong to an identical person. The extended support vector machines are applied to map a serial of low-level feature distances to a perceived people-similarity. To build people-similarity based video indexing, a novel unsupervised clustering algorithm is also proposed, which can more correctly group individual person according to the mutual people-similarities among multiple people pairs.

Key words: video indexing; people-similarity; machine learning; unsupervised clustering algorithm

1 引言

视频分析和索引是有效获取、浏览、检索、查询视频信息的重要技术, 其研究的初级阶段主要集中在视频结构化分析, 包括: 镜头边缘检测、关键帧提取、镜头聚类、场景合成等. 结构化分析能够减少视频的冗余信息, 但很难为用户提供语义信息. 因此, 目前的研究工作主要致力于在语义层对视频进行分析和索引, 可参阅有关文献[1~4]. 在视频分析和索引中, 相似性度量是一项非常重要的研究内容. 利用底层特征(颜色、运动等)相似性很难描述语义层信息[5,6], 这是目前视频语义分析亟需解决的问题. 我们注意到, 用户在观看视频节目时, 对出现重要人物的内容格外关注, 因此建立基于人物的索引信息对视频语义分析具有重要意义, 而以往并没有太多的工作涉及于此. 本文提出一种利用人物相似度进行视频索引的算法, 以满足用户获取高层语义信息的需要.

本文定义的人物相似度基于两种基本的相似度: 衣服相似度和说话声音相似度. 由于目前的人脸识别技术在光照、表情、姿态等方面的不稳定性[7], 本文没有利用人脸相似度来定义人物相似度, 而只用人脸检测算法定位衣服区域. 在实际的视频节目中, 在人脸很小、光线很暗、侧面人脸等情况下人脸检测算法会失效, 但这些情况对我们的工作影响不大, 因为上述情况下重要人物通常不会出现. 因此, 本文提出的算法具有其合理性和实用性. 为了计算感知层的人物相似度, 本文使用支持向量机及其概率输出理论, 将底层特征的距离映射到感知层的相似度. 此外, 提出一种非监督聚类算法, 对人物进行自动组织和聚类. 这里, 有两个主要原因使得以往的聚类算法(例如 K 均值聚类)无法使用: (1) 聚类是在距离空间, 而不是在特征空间进行; (2) 文中定义和计算的人物相似度不满足三角不等式关系.

2 人物相似度的定义和计算

人物活动在大多数视频节目中占主导地位, 其衣着和声音在同一视频场景中基本不变, 为此我们用衣服的相似度和说话人声音的相似度来定义人物相似度. 人物相似度的计算分为两个步骤: (1) 提取底层的衣服和声音特征, 计算特征间的距离; (2) 使用支持向量机[8]及其概率输出理论[9]将底层特征的距离映射到高层的人物相似度. 图 1 给出了计算人物相似度的流程图.

收稿日期: 2003-02-21; 修回日期: 2003-11-13

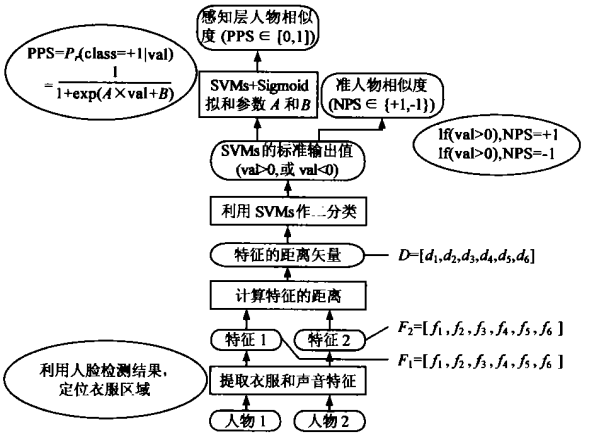

图 1　计算人物相似度的流程图

2.1.1　特征提取

本文共提取和计算 5 个图像特征的距离和 1 个声音特征的距离, 分别用 d_1、d_2、d_3、d_4、d_5、d_6 表示, 其中图像特征具有较高的稳定性和可靠性, 而声音特征容易受背景噪声等干扰.

2.1.1.1　衣服特征　定义人脸以下一个矩形区域为衣服区域, 本文采用一种多角度的人脸检测算法[10]检测人脸的位置和偏转信息. 其中人脸偏转信息用{-3, -2, -1, 0, 1, 2, 3}分别表示从左偏 90 度到右偏 90 度之间共 7 个偏转方向. 定义$((X_1, Y_1), W_1, H_1)$为人脸的矩形区域, 其中(X_1, Y_1)为矩形中心的坐标; W_1, H_1 为矩形的宽和高. 定义$((X_2, Y_2), W_2, H_2)$为衣服矩形, 图 2 给出了人脸和衣服区域的直观描述, 计算衣服区域的公式如下:

$$\begin{cases} X_2 = X_1 + p\#012W_1, \quad p \in \{-3, -2, -1, 0, 1, 2, 3\} \\ Y_2 = Y_1 + H_1 \\ W_2 = 3W_1 \\ H_2 = H_1 \end{cases}$$

在衣服区域, 共提取三类图像特征, 包括: 颜色直方图(Color Histogram)、颜色自相关直方图(Color Auto2Correlogram, CAC)[11]和颜色矩(Color Moment, CM)[12]. 颜色直方图对图像的方向和尺度具有较好的适应性, 在基于内容的图像检索系统中广泛使用. 本文在 YCbCr 颜色空间的三个通道, 分别提取颜色直方图, 每个直方图被量化为 64 级, 用特征向量 f_1、f_2、f_3 表示:

$$f_1 = [H_{Y1}, H_{Y2}, \cdots, H_{Y64}]$$
$$f_2 = [H_{Cb1}, H_{Cb2}, \cdots, H_{Cb64}]$$
$$f_3 = [H_{Cr1}, H_{Cr2}, \cdots, H_{Cr64}]$$

考虑到颜色直方图缺乏对颜色空间分布的描述, 本文还提取颜色自相关直方图 CAC, 以获取颜色在空间的自相关分布特征. 颜色自相关直方图考虑图像中任意两象素点具有相同颜色的概率分布, 定义颜色 C_i(C_i 表示某量化的颜色)在距离 k(k 表示图像中两象素点的 L_1 距离)下的概率分布为:

$$C_{C_i}^{(k)} = P_r[L_1(p_1 - p_2) = k, p_2 \in C_i \mid p_1 \in C_i]$$

实验中, RGB 颜色空间被量化为 4@4@4 级, 设定 k= 1, 3, 5, 7, 提取的 CAC 特征维数为 64@4 级, 用特征向量 f_4 表示:

$$f_4 = [C_1^1, C_2^1, \cdots, C_{64}^1, C_1^3, \cdots, C_{64}^3, C_1^5, \cdots, C_{64}^5, C_1^7, \cdots, C_{64}^7]$$

除颜色自相关特征外, 本文还提取颜色矩 CM 特征, 获取颜色在空间的整体分布特征. 用$(\overline{x_{C_i}}\ \overline{y_{C_i}})$表示颜色 C_i 的所有象素点的分布重心坐标, 其中$\overline{x_{C_i}}$和$\overline{y_{C_i}}$分别被归一化到[0, 1]范围. 实验中, 在量化为 36 级的 HSV 颜色空间提取 CM 特征, 用向量 f_5 表示:

$$f_5 = [\overline{x_1}, \overline{y_1}, \overline{x_2}, \overline{y_2}, \cdots, \overline{x_{36}}, \overline{y_{36}}]$$

对特征向量 f_1、f_2、f_3、f_4, 分别使用线性相关函数 Corr 计算特征的距离, 其中 L, M表示特征向量, L, M表示向量 L, M 的均值, N 表示向量的维数:

$$\mathrm{Corr}(L, M) = \frac{\sum_{i=1}^{N}(M_i - M)(L_i - L)}{\sqrt{\sum_{i=1}^{N}(M_i - M)^2}\sqrt{\sum_{i=1}^{N}(L_i - L)^2}}$$

对特征向量 f_5, 使用 L_1 距离计算其特征的距离:

$$L_1(L, M) = \left(\sum_{i=1}^{N} |L_i - M_i|\right)/N$$

2.1.1.2　声音特征　实验中, 说话人的音频信号分割为长 3 秒的基本单元, 提取的声音特征包括: 线谱对系数(Linear Spectral Pairs, LSP)、Mel 倒谱系数(Mel2Frequency Cepstral Coefficients, MFCCs)、音高(Pitch), 特征维数分别为 10、8 和 2. 使用混合高斯模型(Gaussian Mixture Models, GMMs)对提取的声音特征建模, 实验中使用 GMM232 模型. 假定前 m21 个单元计算的 GMM 模型为 $G_{m-1} \sim N(L_{m-1}, C_{m-1})$, 第 m 个单元 $G_c \sim N(L_c, C_c)$对 G_{m-1} 进行更新得到 $G_m \sim N(L_m, C_m)$, 其中 L_{m-1}, L_c, L_m 是混合高斯模型的均值向量, C_{m-1}, C_c, C_m 是混合高斯模型的协方差矩阵, N_{m-1}, N_c, N_m 是混合高斯模型的特征维数.

$$L_m = \frac{N_{m-1}}{N_{m-1} + N_c} L_{m-1} + \frac{N_{m-1}}{N_{m-1} + N_c} L_c$$
$$C_m = \frac{N_{m-1}}{N_{m-1} + N_c} C_{m-1} + \frac{N_{m-1}}{N_{m-1} + N_c} C_c$$
$$N_m = N_{m-1} + N_c$$

考虑到混合高斯模型的均值向量容易受到背景噪声干扰, 本文只使用协方差矩阵的对角元素描述说话人的声音特征, 用特征向量 f_6 表示:

$$f_6 = [C_1^1, C_1^2, \cdots, C_1^{20}, C_2^1, C_2^2, \cdots, C_2^{20}, \cdots, C_{32}^{20}]$$

使用 K2L 距离计算声音特征的距离:

$$K\text{-}L(L, M) = \mathrm{tr}[(L - M)(M^{-1} - L^{-1})]/2$$

至此, 通过特征提取和特征距离的计算, 得到距离向量 D, 每个分量均被规一化到范围[0, 1]:

图 2　人脸和衣服区域的关系

$$D=[d_1, d_2, d_3, d_4, d_5, d_6]$$

212 人物相似度的计算

为了将底层特征的距离映射到感知层的相似性,本文使用支持向量机(Support Vector Machines, SVMs)的机器学习方法计算人物相似度,计算过程分两个步骤:(1)利用支持向量机的二分类功能,计算离散的人物相似度,记作准人物相似度(Near People Similarity, NPS, NPS({+1, -1});(2)利用支持向量机的概率输出理论[9]计算连续的人物相似度,记作感知层人物相似度(Perceived People Similarity, PPS, PPS([0,1]).

(1)把底层特征距离 D 作为支持向量机的输入向量,使用高斯径向基函数(Gaussian Radial Basis Function, RBF)作为支持向量机的核函数.训练支持向量机的分类模型时,输入数据为$(x_1: y_1), \ldots, (x_n: y_n)$,其中 $x_i \in R^6$,代表输入距离矢量. $y_i \in \{+1, -1\}$,代表输入距离矢量的类别标签: $y_i = +1$ 表示两个人物是同一个人; $y_i = -1$ 表示两个人物是不同的人.假定支持向量机的标准输出值为 val,则当 val> 0 时,NPS= + 1;当 val< 0 时,NPS= - 1,从而计算得到准人物相似度.

利用 SVM 计算准人物相似度存在误判情况.为了纠正误判并对人物进行有效聚类,本文把相似度从离散值拓展为连续值,并在后面提出非监督聚类算法实现人物聚类.

(2)使用 S 形函数拟合支持向量机的标准输出 val 和概率输出 $P_r(\text{class}=+1 \mid \text{val})$间的函数关系[9].通过拟合参数 A 和 B,计算感知层的人物相似度 PPS,该拟合函数的意义为:当 PPS 越趋向 1, val> 0(即 NPS= + 1)的可靠性越大;当 PPS 越趋向 0, val< 0(即 NPS= - 1)的可靠性越大.

$$PPS = P_r(\text{class}=+1 \mid \text{val}) = \frac{1}{1+\exp(A \cdot \text{val} + B)}$$

3 利用人物相似度的视频索引

本文提出的利用人物相似度的视频索引算法的具体步骤如下:(1)利用人脸检测算法定位候选镜头,定义候选镜头具有如下属性:从镜头的大多数帧都能检测到人脸,并且检测到的人脸大小、位置和个数基本不变;(2)从候选镜头的关键帧提取候选人物的衣服特征,在候选镜头内提取候选人物的声音特征.若在视频中检测到 N 个候选人物,则要计算 N·(N - 1)/2 个人物相似度;(3)使用一种新的非监督聚类算法,对出现在不同场景的同一候选人物有效地组织和聚类.

311 非监督的聚类算法

假定 N 个候选人物为$(p_i, i=1, 2, \ldots, N)$,本文将其一一映射为图中 N 个点$(p_i \leftrightarrow n_i, i=1, 2, \ldots, N)$,将 p_i 与 p_j 的人物相似度 PPS 映射为连接 n_i 与 n_j 的边的权值,该映射函数如下,实验中, A= 014, B= 017.

$$F(PPS)=\begin{cases}0, & PPS \in [0, A)\\ PPS, & PPS \in [A, B)\\ 1, & PPS \in [B, 1)\end{cases}$$

图 3 给出映射函数的示意图.对于可靠性较高的区域Ⅰ和Ⅲ的 PPS 值,直接映射为 0 和 1;对于可靠性较低的区域Ⅱ的 PPS 值,利用线性函数代替原有的 PPS 分布.

定义算子 D(n_i, 8),其中 n_i 为图中某节点, 8 为节点集合, D(n_i, 8)表示连接点 n_i 和 8 中各点的边的权值的累加和.聚类过程中,共涉及三类点集: 8(All)、8(C)和 8(~C),分别表示全集、子集和对应于子集的补集,定义 Node(8(C))表示子集 8(C)中的点的个数,聚类算法描述如下:

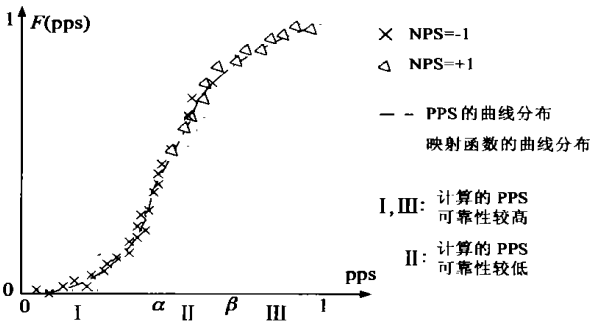

图 3 原 PPS 分布和映射函数 F(PPS)的示意图

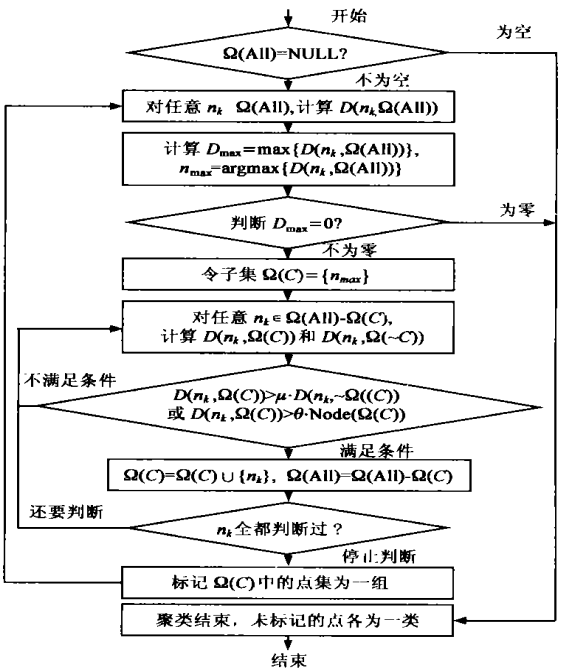

图 4 非监督聚类算法的流程图

如图 5 所示,左半部分描述如何选取子集 8(C)中的初始节点;右半部分描述如何添加子集 8(C)中的后续节点,实验中参数 L 和 H 设为 110 和 015.

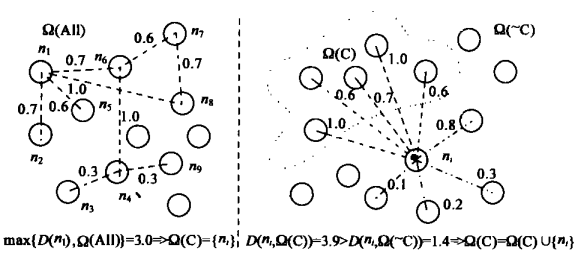

图 5 非监督聚类算法的图示

312　基于人物相似度的视频索引

利用人物相似度进行索引时, 按照人物出现的频率和时间长短排序, 排在前面的人物重要性高; 在一个人物组内, 把不同场景按时间排序, 得到如图 6 所示的树形索引结构, 共有 n 个候选人物组, 每个人物组包括 N_i 个镜头场景. 图的上半部分是对电影 傲慢与偏见0 的一段视频分析后得到的, 该人物是剧中的女主角, 子节点是该人物出现的若干镜头(用关键帧表示).

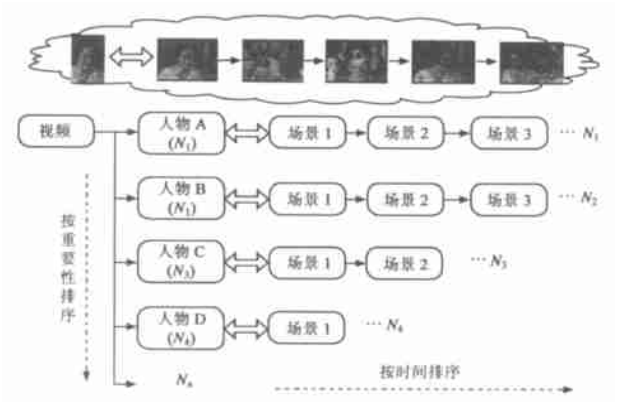

图 6　利用人物相似度的视频索引示例

4　实验结果与分析

411　非特定领域的视频

实验共选取 5 个视频节目, 分别用 V_1, V_2, V_3, V_4 和 V_5 表示, 包括: 2 个新闻节目(N), 1 个家庭录像节目(H), 1 个脱口秀节目(T) 和 1 个电影节目(M), 详细信息如表 1 所示. 手工标注视频数据的人物相似度, 用来衡量算法的准确度, 其中 V_1 的数据用来训练支持向量机的分类模型.

表 1　测试视频的信息

视频	V_1	V_2	V_3	V_4	V_5
类型	N	N	H	T	M
长度(时: 分)	28: 25	30: 27	58: 33	30: 08	14: 16
镜头数目	471	386	472	476	118
候选人物数目	54	56	29	96	43

首先, 利用标注的 V_1 数据训练支持向量机的分类模型, 计算 V_2, V_3, V_4 和 V_5 的人物相似度, 由表 2 可知计算的准人物相似度, 平均准确率达到 96103%(100% ~ 31197%). 经过拓展相似度后使用聚类算法人物分类的平均准确率达到 97126%(100% ~ 21174%). 用计算正确的人物相似度数目(Correctness, Cor.) 和计算的人物相似度总数目(Number of People Pairs, NoPP) 来计算错误率(Recall Error, R_{Err}):

$$R_{Err} = [(NoPP - Cor)/NoPP] @100\%$$

聚类算法能修正一部分 SVMs 计算的准人物相似度的错误, R_{Err} 降低到 21174%.

表 2　聚类前后计算人物相似度的对比

视频	V_2	V_3	V_4	V_5	均值
候选人物	56	29	96	43	)))
NoPP	1540	406	4560	903	)))
Cor. (NPS)	1501	388	4314	871	)))
R_{Err}(NPS) (%)	2. 53	4. 43	5. 39	3. 54	3. 97
Cor. (Cluster)	1531	393	4365	877	)))
R_{Err}(Cluster) (%)	0. 58	3. 20	4. 28	2. 88	2. 74
R_{Err} Reduced(%)	77. 08	27. 77	20. 59	18. 64	36. 02

大多数计算错误的人物相似度是由衣服噪声引起的, 例如胸前举红牌子的人物和穿红色 T 恤衫的人物被计算为同一人物; 此外, 当多个人物同时出现在一个镜头中, 由于现有的说话人分割和识别技术的局限性, 声音特征提取可靠性降低.

表 3　人物聚类的结果

视频	V_2	V_3	V_4	V_5
# H_{GT}	16	5	19	10
# H_{ID}	15	8	22	12
# H_{Cor}	13	3	15	8
# H_{Err}	1	3	3	2
# H_{Mis}	1	2	4	2
Rec. (%)	81. 25	60. 00	78. 95	80. 00
Pre. (%)	86. 67	37. 50	68. 18	66. 67

表 3 给出在 4 个测试视频上得到的人物聚类结果, 符号说明如下: 手标的人物聚类的数目(# H_{GT}), 计算的人物聚类的数目(# H_{ID}), 正确的人物聚类的数目(# H_{Cor}), 错误的人物聚类的数目(# H_{Err}), 漏掉的人物聚类的数目(# H_{Mis}), 准确率(Pre.) 和有效率(Rec.). 从表 3 看出, 家庭录像节目 V_3 的聚类准确率和有效率相对较低, 分析其主要原因是因为非专业拍摄和编辑技术; 对新闻节目 V_2 的聚类效果最好, 因为出现在新闻节目中的人物姿态固定, 噪声较少.

412　新闻视频

新闻视频具有良好的结构性, 人物行为稳定, 不易受噪声干扰, 以往有相当多的工作[14] 检测新闻主持人的镜头. 应用本文提出的人物相似度计算和聚类算法能够对新闻视频中的人物自动聚类, 检测主持人及主持人所在的镜头. 表 4 给出对 4 个新闻节目的人物聚类和主持人检测的实验结果, 每个新闻节目长度约为 30 分钟, 取自于 NBC 晚间新闻报道和 CNN 电视台新闻播报. 使用由视频 V_1 训练 SVMs 的分类模型, 表 4 中主要符号的说明: 候选人物的总数目(Number of People, # NoP); 计算的人物相似度的总数目(Number of People Pairs, # NoPP); 计算正确的人物相似度的数目(聚类算法后的总结果) (Correct People Pairs, # CorrPP); 人物相似度计算的正确率(Precision of computed People Pairs, PrePP); 主持人的数目(Number of Anchor Persons, # NoAP); 正确检测的主持人数目(Correct Anchor Persons, # CorrAP); 主持人检测的准确率(Precision of

detected Anchor Persons, PreAP), 实验表明主持人镜头检测的平均正确率为 91.24%. 本文在检测主持人时, 操作对象是帧中的人物而不是整个帧图像, 所以对主持人的个数和演播室背景没有限制, 对出现多个主持人的镜头能有效定位, 具有很好的鲁棒性和适应性.

表 4 新闻视频的人物聚类和主持人检测结果

新闻视频	N_2	N_3	N_4	N_5	均值
# NoP	54	72	69	61	)))
# NoPP	1431	2556	2346	1830	)))
# CorrPP	1399	2465	2312	1777	)))
PrePP(%)	97.76	96.44	98.55	97.10	97.46
# NoAP	28	36	33	30	)))
# CorAP	25	34	29	28	)))
PreAP(%)	89.29	94.44	87.88	93.33	91.24

5 结语

本文提出一种利用人物相似度进行视频索引的算法, 能够为用户提供一种快速的视频内容获取方式, 是对以往利用底层特征相似度进行视频索引的良好补充. 随着相关技术的提高, 如: 人脸识别、说话人识别, 能够进一步提高计算人物相似度的准确率. 此外, 其他的视频分析技术, 如: 使用基于颜色和运动特征的镜头聚类方法可以对人物相似度的视频索引进行优化, 这也是我们下一步的研究工作.

本文工作为第一作者在微软亚洲研究院做访问学生期间完成.

参考文献:

[1] XU P, XIE L, CHANG S F, DIVAKARAN A, VETRO A, SUN H. Algorithms and systems for segmentation and structure analysis in soccer video[A]. Multimedia and Expo, IEEE International Conference on [C]. Tokyo, Japan: IEEE Computer Society Press, 2001. 721-724.

[2] HUANG Q, LIU Z, GOSERNBER A, GIBBON D, SHARHRARAY B. Automated generation of news content hierarchy by integrating audio, video, and text information[A]. Acoustics, Speech, and Signal Processing, IEEE International Conference on[C]. Arizona, USA: Causal Productions Pty Ltd, 1999, 6. 3025-3028.

[3] NAPHADE M R, HUANG T S. Semantic video indexing using a probabilistic framework[A]. Pattern Recognition, Proceedings 15th International Conference on[C]. Barcelona, Spain: IEEE Computer Society Press, 2000. 3. 79-84.

[4] CHANG S F, CHEN W, SUNDARAM H. Semantic visual templates: linking visual features to semantics[A]. Image Processing, IEEE International Conference on [C]. Illinois, USA: IEEE Computer Society Press, 1998. 3. 531-535.

[5] FABLET R, BOUTHEMY P. Motion-based feature extraction and ascendant hierarchical classification for video indexing and retrieval[A]. Visual Information and Information Systems, Third International Conference[C]. Amsterdam, Netherlands: Springer-Verlag Press, 1999. 221-228.

[6] NGO CH, PONG T C, ZHANG H J. On clustering and retrieval of video shots[A]. Proceedings of the Ninth ACM International Conference on Multimedia[C]. Ottawa, Canada: ACM Press, 2001. 51-60.

[7] CHELLAPA R, WILSON C L, SIROHEY S. Human and machine recognition of faces: a survey[J]. Proceedings of IEEE, 1995, 83(5): 705-740.

[8] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [CP/OL]. http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/, 20012042012001210225.

[9] PLATT J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods[A]. In Advances in Large Margin Classifiers[C]. Cambridge, MA: MIT Press, 2000. 61-74.

[10] LI S Z, ZHU L, ZHANG Z, BLAKE A, ZHANG H J, SHUM H. Statistical learning of multi-view face detection[A]. In European Conference on Computer Vision [C]. Copenhagen, Denmark: Springer-Verlag Press, 2002. 4. 67-81.

[11] HUANG J. Color-spatial image indexing and applications [D]. New York: Department of Computer Science, Cornell University, 1998.

[12] ZHANG L, LIN F, ZHANG B. A CBIR method based on color-spatial feature[A]. IEEE Region 10 Annual International Conference[C]. Korea: IEEE Computer Society Press, 1999. 1. 166-169.

[13] LU L, ZHANG H J. Speaker change detection and tracking in real-time news broadcasting analysis[A]. Proceedings of the 2002 ACM workshops on Multimedia[C]. Juan Les Pins, France: ACM Press, 2002. 602-610.

[14] GAO X, TANG X. Unsupervised video shot segmentation and model-free anchorperson detection for news video story parsing[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2002, 12(9): 765-776.

作者简介:

王 鹏 女, 1979 年 1 月出生于北京, 2001 年毕业于清华大学计算机科学与技术系, 获工学学士学位, 现为清华大学计算机系人机交互与媒体集成研究所博士研究生, 主要从事多媒体信息的基于内容描述、视频分析和索引等.

马宇飞 男, 2000 年毕业于清华大学计算机科学与技术系, 获工学硕士学位, 现为微软亚洲研究院多媒体计算组副研究员, 主要研究方向是基于内容检索技术, 包括基于内容的图像检索、视频内容分析、多媒体数据库、信息检索技术等.

杨士强 男, 1952 年出生于黑龙江省哈尔滨市, 1977 年毕业于清华大学计算机系, 1983 年在清华大学获工学硕士学位, 现为清华大学计算机科学与技术系教授, 博士生导师, 主要研究领域为多媒体信息处理、多媒体网络及人机交互技术等, 在国内外会议和期刊上发表学术论文 50 余篇.