

基于灰度和纹理特征的广告视频序列匹配算法

陈伟, 张宪民

(上海交通大学图像处理与模式识别研究所, 上海 200240)

摘要: 针对基于内容的视频序列检索匹配提出一种综合颜色灰度和纹理特征对分块图像进行描述的方法, 根据颜色灰度特征引入灰阶顺序码的概念。结合 Harr 小波变换与 Gabor 变换提出一种新的纹理提取方案。针对广告视频序列检索匹配的应用提出逐级过滤检索匹配策略, 明显减小了计算复杂度, 提高了检索效率。

关键词: 图像检索; 广告视频序列匹配; 分块; Gabor 变换

Advertising Video Sequences Matching Algorithm Based on Grey and Texture Feature

CHEN Wei, ZHANG Xian-min

(Image Processing & Pattern Recognition Institute, Shanghai Jiaotong University, Shanghai 200240)

【Abstract】 For content-based video sequence retrieval and matching, this paper brings forward a description method to block-image combined by grey and texture feature, while proposes the conception of grey-gap-order to describe grey feature. It unites Harr wavelet and Gabor transform, and puts forward a new method to distill texture feature. In addition, it brings forward a stepped-sift scheme for advertising-video retrieval and matching, which remarkably decreases complexity of computing and boosts the retrieval efficiency.

【Key words】 image retrieval; advertising video sequences matching; block; Gabor transform

1 概述

基于内容的图像检索匹配是利用图像的颜色、纹理、形状和空间关系等底层特征作为比对信息进行检索。选择特征的原则是维度尽量紧凑, 能表达图像内容, 具有足够的辨识度。由于图像信息的复杂性, 单个特征不能有效准确地表达图像内容, 因此本文提出一种综合利用多种特征的快速有效的实时视频序列匹配算法。

在广告视频序列检索匹配这一实际应用中, 需从电视节目视频流中检索出与基准广告视频匹配的视频段。这在检索匹配原理上与其他图像检索领域类似。可以将基准广告视频当作图像数据库, 而电视节目视频流里的每一帧图像相当于查询帧(简称 Q 帧), 要在基准广告视频中检索出与之匹配的基准帧(简称 I 帧), 一段匹配的 Q 帧也就组成为匹配视频段。

2 基准视频的预处理

镜头(Shot)是视频的结构单元, 本质上是一个内容相似度较高的连续帧段; 关键帧(Key Frame)是从中提取出的能代表该镜头的一帧图像, 视频内容用少量具有代表性的关键帧代替, 可以大大减少与 Q 帧的检索匹配运算量。

2.1 镜头划分与关键帧

检测镜头的方法主要有像素差值法^[1]、区域边界检测法^[2]、直方图法^[3]。前2种方法虽然能够较精确地判断镜头变化, 但存在运算量大、无统一特征的缺点。而检索匹配算法切分镜头并不仅是为了精确切分镜头本身, 而是为提取出代表整个镜头的相似性特征, 用于与 Q 帧作匹配运算。图像帧的全局灰度直方图是用于镜头划分的一种计算简单、有效、抗噪声能力强的特征, 本文算法以其作为划分镜头的特征。

每个镜头以其第1帧为关键帧, 镜头的划分以其他帧与关键帧直方图的相似度是否大于门限阈值 δ_{shot} 为评判标准。

相似度可通过计算相关系数得到, 对于2个 n 维向量 $X = [x_0, x_1, \dots, x_{n-1}]$, $Y = [y_0, y_1, \dots, y_{n-1}]$, 常用的计算方法有余弦相关、Pearson 积矩相关、归一化直方图相交法等。本文选取归一化直方图相交法:

$$S(X, Y) = \frac{\sum_{i=0}^{n-1} \min(x_i, y_i)}{\sum_{i=0}^{n-1} x_i} \quad (1)$$

2.2 特征的聚类

为了进一步细化 Q 帧的检索空间, 采用聚类方法对各关键帧直方图特征进行聚类。 Q 帧的匹配检索首先是在聚类中进行的, 然后细分到相似关键帧的帧段中做更详细的匹配运算, 以较大程度地减少检索空间。图1是对镜头划分与聚类的层次结构描述。

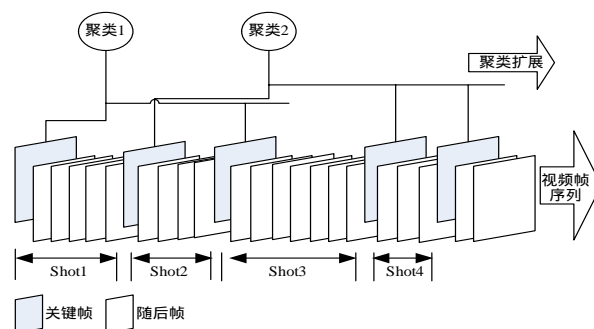


图1 镜头划分与聚类的层次结构

3 分块图像的特征描述

图像检索匹配的主要工作是排除非相似图, 由此提出去

作者简介: 陈伟(1983-), 男, 硕士研究生, 主研方向: 图像处理, 视频检索; 张宪民, 副教授

收稿日期: 2007-12-15 **E-mail:** chenvey@sjtu.edu.cn

伪计算量、求真计算量的概念。去伪计算量是判断 2 幅图不相似所需的计算量，应尽量小，以使大多数非相似图能以较小的代价过滤掉；而求真计算量则是判断 2 幅图相似所需的计算量，应相对较大。

从图像提取的全局特征虽具有整体概括性强的优点，但往往计算复杂、维度较高、去伪计算量大。分块匹配就是将 2 幅图像的相似性衡量变成对应分块图像特征的循序比较，又因为分块本身就带有空间布局信息，所以能够避免全局特征在这方面的不足，且对较小的分块图像提取特征可灵活应用多种策略。综合颜色灰度和纹理特征分别在概括内容和描述细节方面的优势，可以得到一种对分块图像较全面的描述方式。

3.1 颜色灰度特征

经实验统计发现，虽然局部域内颜色灰度值可能受噪声影响发生有限变化，但其阶差顺序关系较为稳定，不易受单调线性变换影响，具有较好的鲁棒性。

如图 2 所示，将每个分块图像 $Block_i$ 平均划分为 4 个子块 S_1, S_2, S_3, S_4 。各子块平均灰度值分别为 M_1, M_2, M_3, M_4 ，按由小到大排序，得到子块标号排列 (l_1, l_2, l_3, l_4) ， $l_k = 1, 2, 3, 4$ 。将其编码为 $L = L(l_1, l_2, l_3, l_4)$ ，称之为灰阶顺序码，由此反映出分块内的灰度阶差顺序关系。其取值模式共有 $4! = 24$ 种，维数较低，如 $L(1, 2, 3, 4)$ 可编码为 1， $L(2, 1, 4, 3)$ 可编码为 7。

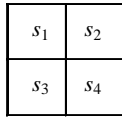


图 2 分块图像的子块划分

提取整个分块图像灰度均值 $\bar{m} = (m_1 + m_2 + m_3 + m_4)/4$ 作为另一个灰度特征，以弥补灰阶顺序码不能反映灰度值的不足。由此得到分块图像的颜色灰度特征 $g = (\bar{m}, L)$ 。对于 Q 帧和 I 帧的对应分块 $Block_i^Q, Block_i^I$ ，其颜色灰度特征矢量 g_i^Q, g_i^I 的相似度定义如下：

$$S(g_i^Q, g_i^I) = w_m \cdot \left(1 - \frac{|\bar{m}_i^Q - \bar{m}_i^I|}{C_{N-1}}\right) + w_L \cdot \text{sign}(L_i^Q, L_i^I) \quad (2)$$

其中，

C_{N-1} 是 N 级灰度级的最大灰度值；

$$\text{sign}(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

\bar{m}_i^Q, L_i^Q 和 \bar{m}_i^I, L_i^I 分别是分块图像 $Block_i^Q$ 和 $Block_i^I$ 的颜色灰度特征；

w_m 和 w_L 分别是分块灰度均值和灰阶顺序码的权重系数 (且 $w_m + w_L = 1$)。

3.2 Gabor 变换与纹理特征

图像纹理反映的是图像的一种局部结构化性质，纹理的定义和量化主要有 2 种方法^[4]：结构方法和统计方法。前者是将图像中非常具有结构规律的特征加以分析，后者则是对图像中色彩强度的空间分布信息进行统计。由于大多实际图像并没有明显的结构化规律，因此使用最广泛的纹理提取方法，对图像变换后的系数提取统计信息来得到纹理特征。

在比较多种小波变换后，文献[5]认为 Gabor 变换最符合人类的视觉特性，其对纹理的描述性能优于基于小波变换的方法。取二维实数 Gabor 小波函数为^[6-7]

$$\psi(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \cdot \exp(2\pi\omega x) \quad (3)$$

其中， ω 为调制频率，对该函数进行膨胀和旋转以得到一系列自相似函数：

$$\psi_{mn}(x, y) = a^{-m} \psi(\tilde{x}, \tilde{y}) \quad (4)$$

其中， $\tilde{x} = a^{-m}(x \cos \theta + y \sin \theta)$ ； $\tilde{y} = a^{-m}(-x \sin \theta + y \cos \theta)$ 。这里， a^{-m} ($a > 1$) 为尺度因子， θ 为旋转角度。由此，图像 $I(x, y)$ 的二维 Gabor 变换离散化形式定义为

$$G_{mn}(x, y) = \sum_s \sum_t I(x-s, y-t) \psi_{mn}^*(s, t) \quad (5)$$

其中， $\psi_{mn}^*(s, t)$ 是自相似函数 $\psi_{mn}(s, t)$ 的共轭复数。对图像的 Gabor 变换可当作在多尺度和方向上的边缘检测，在一个给定区域中，这些微观特性的统计可用来反映其纹理特征。

3.3 分块图像灰度与纹理特征的提取

由式(5)可看出，Gabor 变换的计算量是与卷积相当的。出于时效性考虑，不宜进行过多层次和方向的 Gabor 变换。所以本文对分块图像只取 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 4 个角度方向上的纹理特征。

为进一步简化计算，本文提出一种 Harr 小波变换与 Gabor 变换相结合的小波纹理提取方法。二维 Harr 变换后 LL 子图保留了原图大部分的信息，其灰度和纹理信息是可靠的，所以在 Harr 变换时，只需计算 LL 部分的 Harr 系数，且只对 LL 子图进行 4 个方向上的 Gabor 变换。需要指出的是，统计分块图像的颜色灰度特征 $g = (\bar{m}, L)$ 也是通过 LL 子图进行的。

变换后可得到 4 个系数矩阵 $G_0(x, y), G_1(x, y), G_2(x, y), G_3(x, y)$ ，求各系数矩阵幅度和：

$$E_k = \sum_x \sum_y |G_k(x, y)|, \quad k = 0, 1, 2, 3 \quad (6)$$

以系数幅度的均值 μ_k 和标准方差 σ_k 表示图像的纹理特征：

$$\mu_k = \frac{E_k}{N_k}, \quad \sigma_k = \frac{\sum_x \sum_y (|G_k(x, y)| - \mu_k)^2}{N_k} \quad (7)$$

其中， N_k 为系数矩阵 G_k 中的系数总个数；以 μ_k, σ_k 构成的图像纹理特征矢量为 $f = (\mu_0, \mu_1, \mu_2, \mu_3, \sigma_0, \sigma_1, \sigma_2, \sigma_3)$ ； Q 帧和 I 帧的对应分块 $Block_i^Q$ 和 $Block_i^I$ 的纹理特征矢量 f_i^Q, f_i^I 的相似度定义如下：

$$S(f_i^Q, f_i^I) = \frac{1}{4} \left[w_\mu \sum_{k=0}^3 \frac{\min(\mu_k^Q, \mu_k^I)}{\max(\mu_k^Q, \mu_k^I)} + w_\sigma \sum_{k=0}^3 \frac{\min(\sigma_k^Q, \sigma_k^I)}{\max(\sigma_k^Q, \sigma_k^I)} \right] \quad (8)$$

其中， μ_k^Q, σ_k^Q 和 μ_k^I, σ_k^I 分别是分块图像 $Block_i^Q$ 和 $Block_i^I$ 的纹理特征，综合颜色灰度特征 $g = (\bar{m}, P)$ ，定义分块图像 $Block_i^Q$ 和 $Block_i^I$ 的总相似度：

$$S(Block_i^Q, Block_i^I) = w_c S(g_i^Q, g_i^I) + w_t S(f_i^Q, f_i^I) \quad (9)$$

4 逐级过滤检索匹配算法

为减少去伪计算量的冗余，提出逐级过滤策略。将检索匹配过程分成粗检索和精检索 2 个阶段。前者采用全局直方图在聚类中筛选检索空间图像集，后者先采用颜色灰度特征过滤检索空间，然后综合颜色灰度和纹理特征得到 Q 帧与各 I 帧的整体相似系数，最后选出相似系数最大的 I 帧作为 Q 帧的最佳匹配帧。

电视节目视频流以每秒 25 帧的速度不断向 Q 帧队列输入图像帧，需要从中检索出与基准视频匹配的视频段。首先检索出匹配的首帧，即首个与基准视频中 I 帧匹配的 Q 帧，然后检测随后的帧是否匹配。

4.1 首帧检索

视频匹配中首帧的检索类似于通常意义上的图像检索，即在图像数据库(这里为基准视频)中找出与当前 Q 帧匹配的 I 帧。首帧检索需要经历粗检索和精检索 2 个阶段。

(1)粗检索阶段

1)从 Q 帧队列中取出图像帧,作为当前 Q 帧,转步骤 2)。

2)计算当前 Q 帧与基准视频各聚类的相似度,选取大于阈值 $\delta_{cluster}$ 的前 3 个最大值所对应的聚类作为筛选空间,转步骤 3)。

3)计算 Q 帧与筛选空间中所有关键帧的相似度,若大于阈值 δ_{short} , 将该关键帧所属镜头的所有帧加入检索空间图像集,进入精检索阶段;若检索空间为空,则说明没有与 Q 帧相似的关键帧,回到步骤 1)取下一 Q 帧。

(2)精检索阶段

4)从检索空间中取出 I 帧作为当前 I 帧,转步骤 5)。若检索空间为空,说明对 Q 帧的检索完毕,转步骤 7)。

5)计算 Q 帧与当前 I 帧的颜色灰度特征相似度。逐个计算 Q 帧分块的颜色灰度特征,并将其与 I 帧对应块作比较,若相似度小于阈值 T_c , 则说明该颜色灰度特征不匹配,累计不匹配颜色灰度特征个数 Num_{dif} , 若 Num_{dif} 超过阈值 $a \cdot Num_{block}$, 说明 2 个图不相似(a 为一设定的比例系数, Num_{block} 为分块总数),回到步骤 4); 否则说明 Q 帧与 I 帧的颜色灰度特征基本相符,图像相似,待进一步通过综合纹理特征计算总相似度,转步骤 6)。

6)计算各分块的纹理特征,综合前面得到的颜色灰度特征相似度,按式(9)计算各分块图像的相似度,再统计 Q 帧与 I 帧的总相似度:

$$S(Q, I) = \frac{1}{Num_{block}} \sum_{i=0}^{Num_{block}-1} S(Block_i^Q, Block_i^I) \quad (10)$$

将该相似 I 帧加入 Q 帧的相似帧队列。继续步骤 4)直到检索空间为空,转步骤 7)。

7)从 Q 帧的相似帧队列中取相似系数最大的 I 帧作为 Q 帧的最佳匹配帧。

4.2 随后帧的检索

在检索首帧后,检测随后的帧是否与对应位置的 I 帧相似,该方法较简单,不需要筛选检索空间,而是直接与相应位置的 I 帧做从第 5)步开始的相似度计算。若与该 I 帧不相似,则在其位置前后 $[-5, +5]$ 范围内寻找相似帧;若在此范围还是为无相似帧,则记录丢帧数,若丢帧数大于阈值 T_{lost} , 则认为一段视频序列检索匹配完毕。

5 实验结果

算法用 Visual C++实现,实验平台为:CPU Intel Celeron 2.68 GHz, Memory Kingston 1 GB DDR400。实验过程分为对基准视频的预处理(包括镜头划分、关键帧提取和各帧特征的计算,产生聚类并构成特征库)及电视节目视频流图像帧在基准视频特征库中的检索匹配过程。

选取的基准广告视频时长 14 s 960 ms, 帧数为 374, 图像帧大小为 720×576 , 分块大小为 72×72 。本文算法综合颜色灰度和纹理特征描述分块图像,为验证有效性,与以直方图作为分块图像特征的算法进行了同等条件的对比实验。2 种方法的结果统计数据见表 1。其中,综合匹配率指综合特

征和帧数的匹配程度得出的匹配率;非相似帧平均费时指判断电视节目流中的某 Q 帧不是相似帧的平均花费时间,它反映了检索匹配系统排除非相似帧的性能,应当尽量小;首帧检索匹配计算费时指检索到首个匹配帧时,从粗检索到精检索找出最佳匹配帧这个过程花费的时间;随后帧匹配计算费时指首帧之后其他帧匹配计算的平均费时,因省去了首帧的粗检索部分,所以费时少些;平均每帧费时指整个检索匹配过程对 Q 帧(包括匹配与不匹配)的平均处理时间,是图像检索系统整体工作性能的反映。实时视频序列检索匹配应当控制在 40 ms 左右。

表 1 对比实验统计结果

	匹配帧数	综合匹配率(%)	非相似帧平均费时/ms	首帧检索匹配计算费时/ms	随后帧匹配计算费时/ms	平均每帧费时/ms
本文算法	374	93.11	18	67	46	28
分块直方图法	362	86.78	15	32	18	17

由表 1 可以看出,本文算法的处理速度与视频流每 40 s 一帧的速度在同一个数量级上,且非相似帧平均费时与分块直方图法相近,具有较好的时效性。分块直方图法虽然计算简便,速度更快,但其仅使用直方图特征,对内部细节信息的忽略使其在可靠性上存在较大隐患,因此,在整体性能评价上低于本文算法。

6 结束语

与分块直方图法的对比实验表明,本文提出的综合颜色灰度和纹理特征的新方法在保证时效性的前提下,其对视觉特性有较为全面的描述,使得检测结果更为可靠。该方法还可应用到更广义的图像检索领域。然而,基于内容的图像视频检索匹配还有待进一步的研究,如还需在 Gabor 变换描述纹理特征上进行更多的探索,以使特征系数能更好地代表图像的纹理特性。

参考文献

- [1] Xiong Wei, Lee Chung-Mong. Net Comparison: A Fast and Effective Method for Classifying Image Sequences[C]//Proceedings of SPIE Conf. on Storage and Retrieval for Image and Video Databases III. San Jose, USA: [s. n.], 1995: 318-328.
- [2] Zabih R, Miller J. A Feature-based Algorithm for Detecting and Classifying Scene Breaks[C]//Proceedings of the 4th ACM International Conference on Multimedia. San Francisco, USA: [s. n.], 1995: 189-200.
- [3] Zhang Hongjiang, Low C Y, Smoliar S W. Automatic Partition of Full-motion Video[J]. Multimedia Systems, 1993, 1(1): 10-28.
- [4] 李向阳, 庄越挺, 潘云鹤. 基于内容的图像检索技术与系统[J]. 计算机研究与发展, 2001, 38(2): 46-52.
- [5] Manjunath B S. Texture Features for Browsing and Retrieval of Large Image Data[J]. IEEE Transactions on PAMI, 1996, 18(8): 837-842.
- [6] Qiu Sigang, Feichtinger H G. Discrete Gabor Structures and Optimal Representations[J]. IEEE Trans. on Signal Processing, 1995, 43(10): 2258-2268.
- [7] 李弼程, 彭天强, 彭波. 智能图像处理技术[M]. 北京: 电子工业出版社, 2004-07.