基于组合相似性的视频检索

邓 丽 金立左 费树岷
(东南大学自动控制系 南京 210096)

摘 要：该文研究基于镜头的视频检索问题，提出了一种新的基于组合相似性的镜头相似性度量方法。首先把镜头看成由帧序列组成的一个组合，镜头的相似性通过帧组合的相似性来度量。其次通过用一个非线性映射，把帧组合所在的空间映射到一个高维空间，在这个空间中，假设帧组合服从正态分布，利用核方法，抽取出关键帧序列，并计算出两个正态分布之间的概率距离，这个距离表明了帧组合的相似程度，从而得到两个镜头之间的相似性。最后将这种方法应用于基于镜头的视频检索中，实验表明在相同条件下，基于该方法的检索效果明显优于传统的欧式距离和直方图交方法。

关键词：镜头相似性；组合相似性；核方法；概率距离；视频检索
中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2007)05-1023-04

Ensemble Similarity-Blased Video Retrieval

Deng Li Jin Li-zuo Fei Shu-min
(*Department of Automatic Control Engineering*, *Southeast University*, *Nanjing* 210096, *China*)

Abstract: In this paper, a novel method is proposed to determine the similarity between shots. Firstly, a shot is treated as an ensemble that consists of a sequence of video frames. Shot similarity can be measured by ensemble similarity. Secondly, the original space is mapped to a high dimension space by a nonlinear mapping. In this space, distribution of the ensemble can be assumed as a normal distribution. Finally, by kernel method, the probability distance is computed directly. This distance is equivalent to the ensemble similarity. So, the shot similarity is also obtained. Experimental results show that this method achieves superior performance than the traditional Euclidean distance and histogram intersection methods.

Key words: Shot similarity; Ensemble similarity; Kernel methods; Probabilistic distance; Video retrieval

1 引言

由于视频数据巨大的数据量和丰富的内容，所以对视频数据库的检索一直都是一个具有挑战性的工作。通常，一段视频数据可以划分为几个场景(也叫做故事单元)，每个场景又包含一个到多个镜头。一个镜头是指一系列连续纪录的图像帧，用于表示一个时间段或相同地点连续的动作，由摄像机一次摄像的开始和结束所决定。一个视频场景结构是指一连串语义相关的镜头，它们一般发生在相同的时间和地点，出现相同的人物或事件。所以，视频数据可以按照由粗到细的顺序划分为4个层次结构：视频(video)、场景(scene)、镜头(shot)和帧(frame)。

目前,广泛接受的视频查询方式一般是基于关键字的检索和基于例子的检索。相对于基于关键字的检索方式，基于例子的检索在某些场合是一种必要方式，因为有时用户并不能清楚地用文字来表达他(她)的意愿，而是通过向查询系统提交一个图像或者视频例子，往往可以得到满意的查询结果。例子可以是一段视频、一幅图像、一个物体对象或者一些低级特征，比如颜色、纹理、形状等等[1–4]。对于基于例子的查询主要解决的一个问题是如何度量查询例子与视频数据库中相应视频信息的相似性。

本文主要考虑以一个镜头作为查询例子，查询结果是视频数据库中一组按内容相似度降序排列的镜头。很显然，一帧一帧地比较镜头之间的相似度是很困难的，需要很大的计算量。所以大多数方法都集中在基于关键帧的比较上。文献[1,2]用两镜头对应关键帧之间的平均距离来度量镜头的相似性。文献[3]按镜头内容变化把镜头分解为几个内容一致的子镜头，利用子镜头之间的的相似性来计算两个镜头的相似性。文献[4]将一种新的聚类方法——最近特征线法(NFL)用于基于镜头的检索。这些方法的缺点在于忽略了镜头的总体性，只是仅仅依靠几个关键帧，即几个点之间的距离来判断镜头是否相似。

组合相似性[5]是一种新的集合相似性度量方法，一个组合是指实体或者样本点的集合，组合相似性函数决定两个组合之间的相似性，它从总体上比较两个由若干样本点构成的组合之间的相似程度，克服了样本点之间点点比较的缺点。由于一个镜头是由一系列连续纪录的帧组成，所以可以把一个镜头看作由帧序列组成的组合，从而把组合相似性引入到视频检索中来，来度量两个镜头的之间的相似性，从总体上

2006-02-20 收到，2006-09-27 改回

比较两个镜头的相似程度，克服了上述方法的缺点。实验表明，在同等条件下，该方法检索时的查准率-查全率曲线明显优于传统的欧式距离和直方图交方法，取得了较好的检索效果。

2 算法描述

组合相似性的基本思想是把由若干个样本点组成的集合看作一个组合，然后假设组合中的样本点独立同分布于某个基本的概率分布，利用两个概率分布之间的距离作为组合相似性函数，因为概率距离计算的复杂性，只有当两个分布是正态分布时才能推导出解析表达式，而组合中样本点的分布一般不符合这个条件，为此，可以采用某种非线性映射将原空间映射到一个高维空间，在这个高维空间中，可以假设样本服从正态分布，从而得求得两个组合之间的概率距离，即可得到两个组合间的相似性。

镜头可以看作由帧序列构成的一个组合，帧是组合中的样本。颜色是描述视频内容的一个重要信息，为了方便比较，选用颜色特征来表示帧。$\boldsymbol{HSV}$ 颜色模型与人的视觉特征比较接近，它由色度 $\boldsymbol{H}$ 、饱和度 $\boldsymbol{S}$ 和亮度 $\boldsymbol{V}$ 等 3 个分量组成。由于人眼对视觉的分辨能力有一定的局限性，因此对整个颜色空间进行适当的量化是必要的，将 $\boldsymbol{H}$, $\boldsymbol{S}$, $\boldsymbol{V}$ 等 3 个分量按照人的颜色感知进行非等间隔的量化，把量化后的 3 个颜色分量合成一维特征矢量：

$$\boldsymbol{I} = 9\boldsymbol{H} + 3\boldsymbol{S} + \boldsymbol{V} \tag{1}$$

这样，$\boldsymbol{H}$ ，$\boldsymbol{S}$ ，$\boldsymbol{V}$ 等 3 个分量在一维矢量上分布开来，$\boldsymbol{I}$ 的取值范围是 [0,1,2,⋯,71]，将 3 个特征分量合成一个特征分量。为了消除帧大小的影响，在量化之后还要对直方图进行归一化处理，得到 72 位的一维直方图。每一帧在特征空间中对应于这样一个 72 位的特征向量。

2.1 镜头的组合表示

由于镜头可以看成帧序列组成的组合，所以两个镜头的相似性可以由相对应的两个帧序列组合的相似性来决定。

设镜头 $\boldsymbol{S}_1 = \{f_{11}, f_{12}, \cdots, f_{1n_1}\}$，$\boldsymbol{S}_2 = \{f_{21}, f_{22}, \cdots, f_{2n_2}\}$ 其中 f_{ij} 为 $\boldsymbol{S}_i$ 的第 j 个帧向量，$i = 1,2$，$j = 1,2,\cdots,n_i$。这里 n_1 和 n_2 不一定相同，一般的方法是抽取关键帧以后，利用各种基于图像之间相似度量方法[6]来计算 $d(f_{1i}, f_{2j})$，$i = 1,\cdots,n_1$，$j = 1,\cdots,n_2$，然后通过对 $d(f_{1i}, f_{2j})$ 取平均、最大、最小、中值化等操作得出镜头 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 的距离 $d(\boldsymbol{S}_1, \boldsymbol{S}_2)$。

本文将 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 看作两个帧组合，并假设 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 中的帧分别服从基本的概率分布 p_1 和 p_2，则有两组合相似性等同于两个概率分布之间的距离，用 vP 表示。因此，

$$d(\boldsymbol{S}_1, \boldsymbol{S}_2) = vP(p_1, p_2) \tag{2}$$

常用概率距离公式有Chernoff距离,Bhattacharyya距离，KL散度，Mahalanobis距离等。一般情况下计算这些概率距离不是一件易事，只有当两个概率 p_1 和 p_2 为正态分布时，才能计算出上述距离的解析表达式[7]。

为了解决这一问题，通过某种非线性映射 φ 把帧组合 $\boldsymbol{S}$ 映射到一个高维空间中。在这个高维空间中，可以假设帧向量服从正态分布[5, 8]。通过核技术，高维空间的点积形式可以用Mercer核[8, 9]来表示，不必知道非线性映射 φ 的具体形式，如下式所示：

$$k(f_i, f_j) = \varphi(f_i)^{\mathrm{T}} \varphi(f_j) \tag{3}$$

其中 $f_i, f_j \in \boldsymbol{S}$，$\varphi(f_i)$，$\varphi(f_j)$ 分别为 f_i，f_j 在高维空间的对应函数值。

可以看出，非线性映射，即帧在高维空间中的分布是由核函数来决定的，事实上任一个函数只要满足 Mercer 条件，就可以作为 mercer 核，同时可以分解为式(3)形式，目前核函数的选择依据尚没有定论，但是由于这里假设在映射后的空间中帧服从正态分布，所以核函数的选择很重要，通过实验，选择径向基函数作为核函数，如下式所示：

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2 / 2\sigma^2),\ \forall \boldsymbol{x}, \boldsymbol{y} \in R^d \tag{4}$$

这里 $\boldsymbol{x}, \boldsymbol{y}$ 指72维帧向量。这样，经过非线性映射 φ_i (i=1 或者 2) 映射后，在高维空间中帧组合分别为 $\{\varphi_1(f_{11}), \varphi_1(f_{12}), \cdots, \varphi_1(f_{1n_1})\}$ 和 $\{\varphi_2(f_{21}), \varphi_2(f_{22}), \cdots, \varphi_2(f_{2n_2})\}$，利用这些帧作为样本点，由最大似然估计估计出相应的均值和方差，再根据文献[5]给出的两正态分布之间的距离公式，即可求出两个帧组合之间的距离，从而得到两个镜头之间的相似性。但是由于镜头包含的帧的数目一般有上百个，使得计算十分复杂费时，因此，可以考虑在高维空间中对帧聚类，抽取关键帧，用关键帧代替帧计算概率距离。

2.2 结合关键帧的抽取

在高维空间中对帧聚类，抽取出关键帧，可以通过核聚类算法[10]实现。在这里，核函数就是上节中帧组合映射时所采用的核函数，这样，就可以实现在上述高维空间中的关键帧抽取，步骤如下：

(1) 判断镜头长短并确定聚类类别数 m ($2 \le m \le n$) 以及允许误差 $\mathrm{E}\max$，$g = 1$；这里 m 以 10:1 从镜头中来选取，n 为镜头长度；

(2) 确定初始聚类中心 $G_i(g), i = 1,\cdots,m$。在这里，将镜头序列平分为 m 段，每段的首帧为初始聚类中心；

(3)构造核函数映射，这里采用与上节相同的径向基核函数 $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2 / 2\sigma^2),\ \forall \boldsymbol{x}, \boldsymbol{y} \in R^d$，参数的选择也相同；

(4) 计算各样本点(帧)到聚类中心的距离；

(5) 修改核矩阵；

(6) 计算误差 $\boldsymbol{e}$；

(7) 如果 $e \le \mathrm{E}_{\max}$ 则转(8)，否则转(4)；

(8) 计算各类中心 $G_1, G_2, \cdots, G_m$ 并分别选取离类中心最近的帧作为关键帧。

在提取关键帧以后，在高维空间中用关键帧组合 $\{\varphi_1(F_{11}), \varphi_1(F_{12}), \cdots, \varphi_1(F_{1m_1})\}$ 和 $\{\varphi_2(F_{21}), \varphi_2(F_{22}), \cdots, \varphi_2(F_{2m_2})\}$，

其中 F_{ij} 为 $\boldsymbol{S}_i$ 的第 j 个关键帧向量，代替 $\{\varphi_1(f_{11}),\varphi_1(f_{12}),\cdots,\varphi_1(f_{1n_1})\}$ 和 $\{\varphi_2(f_{21}),\varphi_2(f_{22}),\cdots,\varphi_2(f_{2n_2})\}$ 来计算对应组合的概率距离。

2.3 镜头检索

由 $\{\varphi_1(F_{11}),\varphi_1(F_{12}),\cdots,\varphi_1(F_{1m_1})\}$ 和 $\{\varphi_2(F_{21}),\varphi_2(F_{22}),\cdots,\varphi_2(F_{2m_2})\}$，根据最大似然估计，分别估计出两个正态密度函数的均值和协方差矩阵。

计算距离的关键是核函数的使用，由式(3)，定义下式：

$$\begin{pmatrix}\boldsymbol{\Phi}_1^{\mathrm{T}}\\ \boldsymbol{\Phi}_2^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{\Phi}_1^{\mathrm{T}}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_1^{\mathrm{T}}\boldsymbol{\Phi}_2\\ \boldsymbol{\Phi}_2^{\mathrm{T}}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2^{\mathrm{T}}\boldsymbol{\Phi}_2\end{pmatrix}\equiv\begin{pmatrix}\boldsymbol{K}_{11} & \boldsymbol{K}_{12}\\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22}\end{pmatrix} \tag{5}$$

其中 $\boldsymbol{K}_{ij}=\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_j, i,j=1,2$，$\boldsymbol{\Phi}_t\equiv[\varphi(F_{t1},F_{t2},\cdots,F_{tm_t}]$，$t=1,2$ 且 $\boldsymbol{K}_{12}=\boldsymbol{K}_{21}^{\mathrm{T}}$，其中 $\boldsymbol{K}_{ij}=\left[k(F_{i,p},F_{j,q})\right]_{q=1,\cdots,m_j}^{p=1,\cdots,m_i}$，$i,j=1,2$，$k(\,,\,)$ 为核函数。

于是，可得到 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 之间概率距离，这里选择最常用的 Chernoff 距离和 KL 散度，对于 Chernoff 距离公式为：

$$vP_C=\alpha_1\alpha_2(w_{11}+w_{22}-w_{12}) \tag{6}$$

其中 α_1,α_2 为常数且 $\alpha_1+\alpha_2=1$，$w_{ij}=\{\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{K}_{ij}\boldsymbol{e}_j-\boldsymbol{e}_i^{\mathrm{T}}[\boldsymbol{K}_{i1}\boldsymbol{K}_{i2}]\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}[\boldsymbol{K}_{1j}^{\mathrm{T}}\ \boldsymbol{K}_{2j}^{\mathrm{T}}]^{\mathrm{T}}\boldsymbol{e}_j$，$\boldsymbol{e}_{m_i\times 1}\equiv\frac{1}{m_i}\mathbf{1}$；$\boldsymbol{L}_{m_i\times m_i}\equiv\frac{1}{\sqrt{m_i}}(\boldsymbol{I}_{m_i}-\boldsymbol{e}\mathbf{1}^{\mathrm{T}})$，$\mathbf{1}$ 指长度为 m_i 的列向量，$\boldsymbol{C}=\begin{pmatrix}\sqrt{\alpha_1}\boldsymbol{L}_1\boldsymbol{Q}_1 & 0\\ 0 & \sqrt{\alpha_2}\boldsymbol{L}_2\boldsymbol{Q}_2\end{pmatrix}$，$Q_i$ 是协方差阵的“缩影”，它保持了协方差阵的前 r_i 个特征对：$\{\lambda_{ij},v_{ij}\}_{j=1}^{r_i}$，$\boldsymbol{Q}_i\equiv\boldsymbol{V}_{r_i}(\boldsymbol{I}_{r_i}-\Lambda_{r_i}^{-1})^{1/2}$，$\boldsymbol{V}_{r_i}\equiv[v_{i1},\cdots,v_{ir_i}]$，$\boldsymbol{\Lambda}_{r_i}\equiv\mathrm{Diag}[\lambda_{i1},\cdots,\lambda_{ir_i}]$，$i=1,2$。

对于 KL 散度，公式为：

$$vP_K=\tau_{121}+\tau_{222}-\tau_{122}-\tau_{221}+tr[\boldsymbol{\Lambda}_{1\eta}]-\eta_{12} \tag{7}$$

其中 $\tau_{ijk}=(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{K}_{ik}\boldsymbol{e}_k-\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{K}_{ij}\boldsymbol{B}_j\boldsymbol{K}_{jk}\boldsymbol{e}_k)$，$\eta_{ij}=\mathrm{tr}[\boldsymbol{A}_i\boldsymbol{K}_{ij}\boldsymbol{B}_j\boldsymbol{K}_{ji}]$，$\boldsymbol{A}_i=\boldsymbol{L}_i\boldsymbol{V}_{r_i}\boldsymbol{V}_{r_i}^{\mathrm{T}}\boldsymbol{L}_i^{\mathrm{T}}$，$\boldsymbol{B}_i=\boldsymbol{L}_i\boldsymbol{V}_{r_i}\boldsymbol{\Lambda}_{r_i}^{-1}\boldsymbol{V}_{r_i}^{\mathrm{T}}\boldsymbol{L}_i^{\mathrm{T}}$，$i=1,2$。

则 vP_C 和 vP_K 的值给出了两镜头 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 之间的相似程度。另外根据参数选择的不同，vP_C 和 vP_K 会有所变化，而且 vP_C 和 vP_K 一般不同，实际使用中，一般根据情况分别选定阈值 λ_C,λ_K，当 $vP_C\le\lambda_C, vP_K\le\lambda_k$ 时就认为两者是相似的。

设用户查询镜头 $\boldsymbol{S}_q$，则通过计算该镜头与视频数据库中所有镜头之间的 Chernoff 距离或者 KL 散度来决定其与库中各个镜头之间的相似度，然后将所得到的查询镜头与所有镜头的距离按从小到大排序，距离最短的那个镜头就是在视频库中同查询点最相近的镜头。

3　实验结果

实验从国际影视检索测评(TREC Video Retrieval Evaluation, TRECVID)2003 提供的 CNN headline news 和 ABC World News Tonight 视频中随机选取若干视频段。建立一个包含 2060 个镜头的视频库。选择的视频段包含的视频内容非常丰富，有人物、事件、体育和影视等各方面的新闻内容。从视频库中随机挑选出 500 个镜头作为查询样本，并对每个查询镜头主观地选取一组视觉相似的镜头作为标准。

查准率(precision)和查全率(recall)是视频检索中常用的两个评价指标。查准率用检索到的与主观标准相符的镜头数与所有检索到的镜头数比值衡量，查全率用检索到的与主观标准相符的镜头数与主观选取所有镜头数比值。这里选用 500 个查询样本镜头的查准率和查全率作为检索评价指标。

经过实验比较，径向基核函数(RBF)效果比高斯核函数以及多项式核函数效果要好，所以这里选择它作为核函数。首先，利用核聚类算法进行关键帧抽取，根据镜头长短的不同，抽取出的关键帧数目变化很大，平均每个镜头的关键帧数约为 12 个。

很显然，概率距离依赖于特征对个数 r_1 和 r_2 (这里，取 $r_1=r_2=r$)以及 RBF 核宽度 σ，图 1 给出了任取的两镜头间距离随 σ 及 r 变化的情况，可以看出，当 $\sigma>2$ 时，各镜头间距离变得很小，可分性很差，所以在实际使用中，可选 $\sigma=0.5$ 左右。另外，r 的取值对距离影响不是很大，并不是说 r 越大距离就越大，实际使用时可以选择 r 为关键帧个数 m_1,m_2 中的较小者。

为了比较概率距离衡量的镜头检索的性能，这里选择常用的基于关键帧之间的欧几里德距离(Euclidean distance)方法[1]以及非距离度量方法的直方图的交(Histogram intersection)方法[11]作为比较。因为这里每个镜头的关键帧个数不一定相同，为了与概率距离方法比较，对两个镜头按欧式距离和直方图交方法进行相似性度量时，首先对两个镜头包含的所有关键帧两两按欧式距离和直方图交方法计算关键帧之间的度量值，最后取平均值作为两个镜头间的相似性度量。

图 2 所示为 500 个检索样本镜头的在 Chernoff 距离、KL 散度以及欧式距离度量下的查准率和查全率。从实验结果看，Chernoff 距离和 KL 散度的 recall-precision 曲线下面的面积明显大于 Euclidean 距离以及直方图交的 recall-precision 曲线下面的面积，而且当 recall 为 0.2 时，Chernoff 距离核 KL 散度方法的查准率达到 0.8 以上，而另外两种方法仅为 0.5 左右，当 recall 为 0.4 时，前两种基于概率距离的方法查准率为 0.2 左右，而此时两种传统方法的查准率趋于

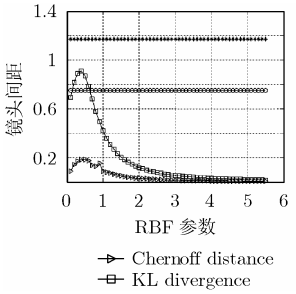

图 1　RBF 核宽度 σ 对各种距离的影响

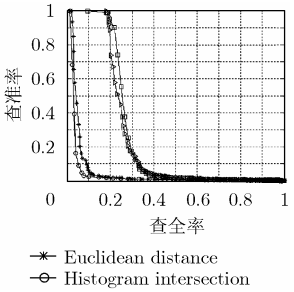

图 2　4 种方法得到的 precisionon-recall 图

0，这说明基于概率距离的 Chernoff 距离和 KL 散度的镜头检索效果明显好于欧氏距离方法以及直方图交方法。

为了更直观地说明这个问题，从 500 个查询镜头中随机选取 20 个镜头，这 20 个查询镜头的人工标记平均相似镜头数为 19.6 个。对每个查询镜头检索后，系统返回一系列按相似度降序排列的镜头，分别取前 10，50，100，200，300 个返回镜头，与人工标记的相似镜头进行比较，相同的为相关镜头，对于 4 种不同的方法，各种情况下 20 个查询镜头返回的平均相关镜头数如表 1 所示。结果表明，当返回镜头数为 10 时，前两种方法的 precision 分别为 45%和 40%，recall 为 23%和 20%,而后两种方法，precision 分别为 28%和 24%，recall 为 14%和 12%。此时前两种方法明显优于后两者。当返回镜头数大于 100 时，几种方法的返回的镜头中相关镜头数差别不大，但是此时的 precision 值很小。

表 1 不同返回镜头数目情况下各种方法的平均相关镜头数

返回的镜头数	Chernoff 距离	KL散度	欧氏距离	直方图交
10	4.5	4.0	2.8	2.4
50	6.1	6.6	4.8	4.3
100	7.2	7.8	6.4	5.5
200	8.4	9.1	7.8	7.4
300	9.8	10.1	9.0	8.5

综合看来，Chernoff 距离和 KL 散度方法的检索效果基本相同，明显优于欧氏距离和直方图的交方法，而后两者中欧氏距离方法略好约直方图的交方法。

4 结束语

本文研究了基于镜头的视频检索问题，提出了一种基于组合相似性来判断视频镜头相似与否的方法，这种方法从总体上比较两个镜头之间的相似度，克服了之前只是通过点点之间的比较相似性的缺点。并将之应用于视频检索中。与传统的欧式距离度量方法以及非距离度量的直方图的交方法相比，本文的方法在基于镜头的视频检索中取得较好表现。

未来的工作首先是把运动信息、空间信息与颜色信息融合起来，由于视频信息的复杂性，仅仅依靠颜色信息不能有效地表示镜头内容，造成检索性能不高；其次是增加镜头组合中样本数量，关键帧数目相对较少，对计算结果有很大的影响，虽然本文的实验结果表明基于关键帧的组合相似性检索效果比一般的方法要好，但是还是存在一定的缺陷，下一步考虑将对镜头进行亚采样，增加组合中样本个数。相信可以取得更好的检索效果。

参 考 文 献

[1] Jain A K, Vailaya A, and Wei X. Query by video clip. *Multimedia System*, 1999, 7(5): 369–384.

[2] Shan M K and Lee S Y. Content-based video retrieval based on similarity of frame sequence. In Proceedings of the IEEE Conference on Multimedia Computing and Systems, Austin, Texas, 5-7 Aug, 1998: 90–97.

[3] 林通, 张宏江, 封举富等. 镜头内容分析及其在视频检索中的应用. 软件学报, 2002, 13(8): 1577–1585.
Lin T, Zhang H J, and Feng J F, *et al.*. Shot content analysis for video retrieval application. *Journal of Software*, 2002, 13(8): 1577–1585.

[4] 赵黎, 祁卫, 李子青等. 基于关键帧提取的最近特征线 (NFL) 聚类算法的镜头检索方法. 计算机学报, 2000, 23 (12): 1292–1298.
Zhao Li, Qi Wei, and Li Zi-qin. Key-Frame Extraction Based Improved Nearest Feature Line(NFL) Classification Algorithm. *Journal of Computers*, 2000, 23(12): 1292–1298.

[5] Zhou S K and Chellappa R. From sample similarity to ensemble similarity: Probabilistic distance measure in reproducing kernel Hilbert space. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006, 28(6): 917–929.

[6] Missaoui R, Sarifuddin M, and Vaillancout J. Similarity measures for efficient content-based image retrieval. *IEE Proceedings Vision, Image and Signal Process*, 2005, 152(6): 875–887.

[7] Devijver P and Kitter J. Pattern Recognition: A statistical Approach. UK: Prentice Hall International, 1982: 120–146.

[8] Bach F and Jordan M I. Learning graphical models with Mercer kernels. In Advances in Neural Information Processing Systems15, Cambridge, MA, 2003, MIT Press.

[9] Kondor R and Jebara T. A kernel between sets of vectors. In Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC, USA. , August 21-24, 2003: 361–368.

[10] 张莉, 周伟达, 焦李成. 核聚类算法. 计算机学报, 2002, 25(6): 587–590.
Zhang Li, Zhou Wei-da, and Jiao Li-cheng. Kernel clustering algorithm. *Journal of Computers*, 2002, 25(6): 587–590.

[11] Swain M J and Ballard D H. Color indexing. International *Journal of Computer Vision*, 1991, 7(1): 11–32.

邓 丽： 女，1978 年生，博士生，研究方向为基于内容的视觉信息检索、模式识别、机器学习.

金立左： 男，1972 年生，副教授，研究方向为图像处理、模式识别、计算机视觉.

费树岷： 男，1961 年生，教授，博士生导师，研究方向为自适应控制、智能控制、自动目标识别.