

基于内容的视频重组方法

程义民¹⁾ 沈 勇¹⁾ 王以孝¹⁾ 罗瑞江²⁾

¹⁾(中国科学技术大学电子科学与技术系, 合肥 230026)

²⁾(新加坡国立大学系统科学研究所)

摘 要 针对视频点播及快速测算时经常遇到的视频重现问题,提出了一种基于内容的视频重组方法,该方法是用整帧图象的彩色直方图、彩色矩等图象彩色信息来进行帧间图象的相似度量,在此基础上再进行分镜头切分,进而给出简化重组、比例重组和自适应重组等重组方法.该方法已在计算机上进行了模拟实验,获得了较好的结果.该方法可用于视频数据库及其他视频信息数据的快速预览,也可用于基于 Browser/Server 架构的 Web 型视频数据库的视频预览、视频摘要、视频检索以及低带宽条件下替代源视频对象等.

关键词 视频重组 彩色直方图 彩色矩 镜头分割

中图法分类号: TP941.1 TN948.6 **文献标识码**: A **文章编号**: 1006-8961(2001)04-0402-07

A Video Reorganization Method Based on Contents

CHENG Yi-min¹⁾, SHEN Yong¹⁾, WANG Yi-xiao¹⁾, LUO Rui-jiang²⁾

¹⁾(Dept. of Electronic Science and Technology, USTC, Hefei 230026)

²⁾(Institute of System Science, National University, Singapore)

Abstract This paper describes a method for video reorganization based on contents. The method uses the color information of images such as color histogram, color moment etc. to measure the similarity between frames, and to detect and cut the shots. Furthermore, three reorganization algorithms, "simplified reorganization", "proportional reorganization" and "self-adapting reorganization", are also described in detail. All these three algorithms get good results in the simulation. So it can be used for quick preview of the video-database or other video information data. It also can be used for video preview, video abstract, video retrieval and video substitution (in low net band) of Web-Video-Database based on Browser/Server technology.

Keywords Video reorganization, Color histogram, Color moment, Shots cut

0 引 言

随着信息技术的发展,Internet 网及信息数据库技术的广泛应用,视频信息在信息领域的作用越来越重要.同时在视频点播或浏览以视频形式存储的各种信息时,常需快速浏览相应的视频信息,因此视频重组问题的研究已引起了信息领域研究工作者的很大兴趣^[1,2].

传统的重组方法是,在视频序列图象中,按一定比例等间隔地抽取相应的图象帧,然后重组成新的序列图象,以达到快速浏览的目的.但这种方式,对于需要预览速度较快,且图象帧抽取间隔较大的浏

览,则有可能漏掉浏览者所关心的内容.

为了解决上述问题,可用基于内容的重组方法,但这这就要求从视频序列图象中,切分出视频内容的最小单元——分镜头,而且在视频重组中,还要保证每一分镜头有一定的图象帧,这样才能使快速浏览有比较好的效果.

Hongjiang Zhang 等人虽然将视频图象的彩色直方图成功地用于分镜头的切分^[3].但由于该方法分镜头的切分判决主要依据彩色直方图的均方差是否超过阈值,因此当整段视频序列图象使用同一阈值或阈值不当时,将影响分镜头切分,从而影响重组的效果.

为了避免上述问题, Sunxing Ding 等人则不对视频序列图象进行镜头分割操作, 而是先将整个序列图象按顺序分成若干小组, 再求出各组内首帧与末帧图象直方图的均方差, 并按照一定的比率将差别小的组删除, 而保留剩余的组, 然后将所有的剩余图象按顺序重新分组, 重复如上操作, 直到所剩的图象帧数满足预先指定的数量或比例为止, 从而得到重组的序列图象^[4]. 该方法对灰度图象的重组效果较好, 而用于彩色图象时, 由于数据量较大, 反复操作需要较多的处理时间.

1 系统概述

视频重组系统主要包括彩色直方图生成器、彩色矩生成器、镜头分割、视频重组、过滤器等部分(见图1). 该系统输入视频序列图象, 输出为经重组后的序列图象.

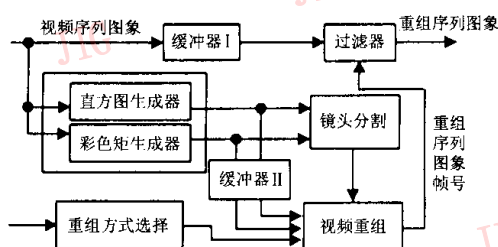


图1 视频重组系统框图

当视频图象序列中的每帧图象, 同时输入到缓冲器I及彩色直方图生成器、彩色矩生成器后, 该二生成器即生成该帧图象的彩色直方图和彩色矩数据流, 然后该数据流连同相应帧号即送入缓冲器II及镜头分割模块.

镜头分割模块的功能是检测数据流中每帧图象与分镜头头帧图象的彩色直方图及彩色矩, 当误差大于阈值时, 则该帧图象则为后一分镜头的头帧图象, 而该帧的前一帧图象为前一分镜头的末帧图象, 从而得到每一分镜头的头帧图象帧号、末帧帧号及镜头定位等信息, 再送入视频重组模块.

视频重组模块功能是, 根据每一分镜头的定位信息、重组方式及从缓冲区II输入的该分镜头内的帧号及相应每帧图象的彩色直方图、彩色矩信息, 进而确定该分镜头内的重组序列图象的帧号, 并将帧号送入过滤器. 而过滤器则是对从缓冲器I输入的视频序列图象, 进行过滤, 只让与视频重组模块输入帧号相同的图象帧通过, 从而生成重组序列图象.

2 彩色直方图及彩色矩计算

为了在重组序列图象中, 保留每一分镜头的‘内容’, 必须从序列图象中切分出每一个分镜头序列, 这样就必须获得每帧图象总体特征的描述. 该系统主要用彩色直方图、彩色矩等来描述序列图象中的每帧彩色图象.

2.1 彩色子空间

图象的直方图在一定意义上描述了该图象的总体特征^[5], 而对视频彩色图象, 由于彩色数据范围较大, 常包含 $2^{16}2^{24}$ 种彩色, 因此相对于每种彩色的直方图, 不仅需要大量的存储空间, 而且直方图的进一步处理也需较多时间.

Gargi 等人根据人眼对彩色感受特性, 在 HVC 彩色空间, 研究了彩色视频图象后发现, 对绝大多数彩色图象, 其象点的颜色往往集中在较少的几种主要彩色上^[6]. Yihong Gong 等人通过实验给出了适于人眼对图象识别的红、橙、黄、皮肤色、绿、青、蓝、紫、黑、灰、白等 11 种主要彩色(见表2)的 HVC 彩色空间分类^[7].

Gargi 等人的研究, 给人们一个重要启示, 即对每一类图象, 可选较少一个彩色子空间, 而每一子空间中心的彩色可作为该彩色子空间的代表, 在此基础上生成的‘彩色直方图’就可用较少数据给出整幅图象彩色特征的一种描述.

视频图象的彩色子空间, 是对彩色空间的一种划分, 可通过聚类的方式来实现.

2.1.1 初始聚类中心的选择

若彩色视频图象样本空间为 L 段彩色视频图象, 相应每段长度为 l_j 帧 ($j = 1, 2, \dots, L$), 每帧图象大小为 $m \times n$, 则样本空间可表示为一个集合 F

$$F = \{F_i\} \quad i = 1, 2, \dots, M; M = \sum_{j=1}^L l_j$$

其中, $F_i = \{f_{ir}(k, l), f_{ig}(k, l), f_{ib}(k, l)\}, k = 0, 1, \dots, m-1; l = 0, 1, \dots, n-1$ (红、绿、蓝, 简称为 R、G、B 彩色空间) 或 $F_i = \{f_{ih}(k, l), f_{iw}(k, l), f_{ic}(k, l)\}, k = 0, 1, \dots, m-1; l = 0, 1, \dots, n-1$ (HVC 彩色空间).

这里, 将该样本空间聚类到 N 个聚类中心, 是一个优化的划分过程, 其初始聚类中心, 可从样本空间任取一幅图象, 再从中任选 N 种彩色 $C_i = \{t_i, g_i, b_i\}, i = 1, 2, \dots, N$, 使其中任两种彩色之间的距离

大于阈值 T_1 ,即

$$\sqrt{(r_i - r_j)^2 + (g_i - g_j)^2 + (b_i - b_j)^2} > T_1, \quad \forall i \neq j, i, j \in \{1, 2, \dots, N\} \quad (1)$$

则这 N 种彩色集合 $C^{(0)}$ 为初始聚类中心,可表示为

$$C^{(0)} = \{C_i^{(0)}\}, i = 1, 2, \dots, N; C_i^{(0)} = \{r_i, g_i, b_i\}$$

2.1.2 彩色子空间的生成

上述样本空间的聚类,可由一迭代方法实现.若第 i 阶迭代计算得到的聚类中心为 $C^{(i)} = \{C_j^{(i)}\}, j = 1, 2, \dots, N; C_j^{(i)} = \{r_j^{(i)}, g_j^{(i)}, b_j^{(i)}\}$, 则第 $i+1$ 阶迭代的主要过程为,对每段彩色视频图象的每一帧图象 $F(k, l) = \{f_r(k, l), f_g(k, l), f_b(k, l)\}, (k = 0, 1, \dots, m-1, l = 0, 1, \dots, n-1)$ 的第 (k, l) 个象点,计算其到第 i 阶迭代生成的每一聚类中心的距离,并将其划分到与其最近聚类中心所对应的子集合

$$d_k = \min_j \left\{ [f_r(k, l) - r_j^{(i)}]^2 + [f_g(k, l) - g_j^{(i)}]^2 + [f_b(k, l) - b_j^{(i)}]^2 \right\}^{1/2} \quad \forall j = 1, 2, \dots, N \quad (2)$$

即,求离该帧图象的第 (k, l) 个象点 $\{f_r(k, l), f_g(k, l), f_b(k, l)\}$ 距离最近的聚类中心(第 k 个) $C_k^{(i)} = (r_k^{(i)}, g_k^{(i)}, b_k^{(i)}), k \in 1, 2, \dots, N$, 并将该象点注入 $C_k^{(i)}$ 所对应的子集合 $V_k^{(i)}$.

经过上述计算,样本空间的每段视频图象中的每帧图象的每个象点,都可被划分到一个子集合,这样,样本空间的所有象点都可被划分到与聚类中心 $C_j^{(i)} (j = 1, 2, \dots, N)$ 对应的 N 个子集合 $V_j^{(i)} (j = 1, 2, \dots, N)$ 中去,且第 j 个子集合可表示为

$$V_j^{(i)} = \{P_j(m)\}, m = 1, 2, \dots, M_j; \\ P_j(m) = (r_{jm}, g_{jm}, b_{jm})$$

其中, M_j 是第 j 个子集合中象点的个数, r_{jm}, g_{jm}, b_{jm} 分别为第 j 个子集合中第 m 个象点的 R, G, B 彩色分量,进而即可求出每一子集合的中心,并将其作为第 $i+1$ 阶迭代的聚类中心

$$r_j^{(i+1)} = \frac{1}{M_j} \sum_{m=1}^{M_j} r_{jm}, g_j^{(i+1)} = \frac{1}{M_j} \sum_{m=1}^{M_j} g_{jm}, b_j^{(i+1)} = \frac{1}{M_j} \sum_{m=1}^{M_j} b_{jm} \quad (3)$$

从而,第 $i+1$ 阶迭代产生的聚类中心为

$$C_j^{(i+1)} = \{r_j^{(i+1)}, g_j^{(i+1)}, b_j^{(i+1)}\}, j = 1, 2, \dots, N$$

将第 $i+1$ 阶产生的聚类中心与第 i 阶的聚类中心进行比较,即计算

$$d_j = \sqrt{[r_j^{(i+1)} - r_j^{(i)}]^2 + [g_j^{(i+1)} - g_j^{(i)}]^2 + [b_j^{(i+1)} - b_j^{(i)}]^2} \quad \forall j = 1, 2, \dots, N \quad (4)$$

当 $d_j < T_2, \forall j = 1, 2, \dots, N$ 时,收敛,否则,重复上述迭代计算.

收敛后,即可得到 N 个聚类中心 $C_j = (r_j, g_j, b_j), j = 1, 2, \dots, N$. 这样每一聚类中心,在 RGB 彩色空间即为一矢量,以这些矢量为基,即可组成一彩色子空间.若用于彩色象点分类,则每一矢量也可对应于一小子空间,其判决函数可表示为一球面方程:

$$(f_r - r_j)^2 + (f_g - g_j)^2 + (f_b - b_j)^2 < r^2 \quad (5)$$

其中, f_r, f_g, f_b 为任一象点的 RGB 彩色分量, r 为球面的半径.也可表示为 RGB 彩色空间上的一六面体方程:

$$\begin{cases} r_j - \Delta r < f_r < r_j + \Delta r \\ g_j - \Delta g < f_g < g_j + \Delta g \\ b_j - \Delta b < f_b < b_j + \Delta b \end{cases} \quad (6)$$

2.1.3 彩色直方图

若求得 N 个聚类中心 $C = \{C_j\} = (r_j, g_j, b_j), j = 1, 2, \dots, N$ 后,则可以这 N 个彩色为基础,对任一帧图象作象点分类,从而得到 N 种彩色直方图.

若将上述每一彩色 $C_j, j \in (1, 2, \dots, N)$ 归一化,即

$$[C_j] = \{R_j, G_j, B_j\} \\ = \left\{ \frac{r_j}{(r_j^2 + g_j^2 + b_j^2)^{1/2}}, \frac{g_j}{(r_j^2 + g_j^2 + b_j^2)^{1/2}}, \frac{b_j}{(r_j^2 + g_j^2 + b_j^2)^{1/2}} \right\}, \\ j = 1, 2, \dots, N \quad (7)$$

其中, $[C_j]$ 为第 j 种彩色的单位矢量.这样,任一帧图象中的任一彩色象点均可按其到每一彩色的投影分类,并可通过如下方法求得相应的彩色直方图.

对视频彩色序列图象中任一帧 $m \times n$ 大小的图象 $f(i, j) = \{f_r(i, j), f_g(i, j), f_b(i, j)\} (i = 0, 1, \dots, m-1; j = 0, 1, \dots, n-1)$, 其任一象点 $f(i, j)$ 可看作 RGB 彩色空间的一个矢量,首先计算

$$S_k = \max_k \{f(i, j) \cdot [C_j]\} \\ = \max_k [f_r(i, j), f_g(i, j), f_b(i, j)] \times [R_j, G_j, B_j]^T, \\ \forall j = 1, 2, \dots, N \quad (8)$$

即计算第 (i, j) 个象点到每一彩色的投影,然后再求出投影最大的彩色类 (k) , 即若其投影值超过阈值 T , 则将该点划归第 k 类, 否则划归其他类. 对一帧图象上的每一象点经过上述分类后, 即可得到该帧图象相对于这 N 种彩色的彩色直方图

$$h = \{h(0), h(1), \dots, h(N)\} \quad (9)$$

其中, $h(j), j = 1, 2, \dots, N$ 分别为该帧图象中第 j 类

彩色的象点数, $h(0)$ 为其他不可划归上述任一类彩色的象点数.

以上为针对 RGB 空间的算法, 对于 HVC 空间, 算法类似. 图 2 给出了一帧图象及相应的 HVC 彩色空间归一化直方图, 所用的 11 种彩色见表 1.

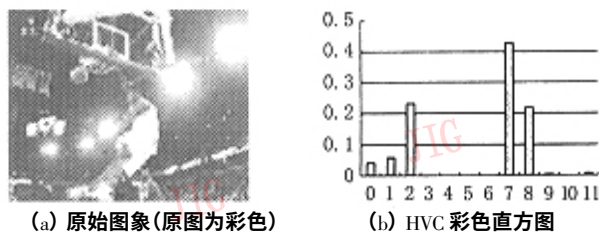


图 2 彩色直方图实例

表 1 11 种彩色的彩色子空间

称号	彩色名称	色度(Hue)	亮度(Value)	色饱和度(Chroma)
11	红	0 ~ 36	4 ~ 9	1.5 ~ 30
		36 ~ 64	4 ~ 9	15 ~ 30
10	橙	64 ~ 112	4 ~ 8	9 ~ 30
9	黄	80 ~ 112	9 ~ 10	1.5 ~ 30
8	皮肤色	36 ~ 64	4 ~ 9	1.5 ~ 15
		64 ~ 112	4 ~ 8	1.5 ~ 9
7	绿	112 ~ 196	4 ~ 10	1.5 ~ 30
6	青	196 ~ 256	6 ~ 8	1.5 ~ 30
5	蓝	256 ~ 312	4 ~ 8	1.5 ~ 30
4	紫	312 ~ 359	4 ~ 8	1.5 ~ 30
3	黑	-	< 3	-
2	灰	-	4 ~ 8	< 1.5
		-	3 ~ 4	-
1	白	-	> 9	< 1.5
0	其他	...	...	...

2.2 彩色矩

彩色矩在一定程度上反映了一帧图象的彩色及其分布特征. 对一帧 $m \times n$ 大小的红、绿、蓝彩色图象, $f(i, j) = \{f_r(i, j), f_g(j, j), f_b(i, j)\}$ ($i = 0, 1, \dots, m-1; j = 0, 1, \dots, n-1$), 其一阶彩色矩可表示为^[3]:

$$\begin{aligned}\mu_r &= \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} f_r(i, j) \\ \mu_g &= \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} f_g(i, j) \\ \mu_b &= \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} f_b(i, j)\end{aligned}\tag{10}$$

μ_r, μ_g, μ_b 分别为该帧图象中红(R)、绿(G)、蓝(B)彩色分量的平均值, 不失一般性, 仅考虑彩色的 r 分量时, 若其量化值为 $f_r = 0, 1, \dots, C$, r 分量为 f_r 的象

点数 n_r , 则 μ_r 可表示为

$$\mu_r = \frac{1}{m \times n} \sum_{f_r=0}^C n_r \cdot f_r$$

这样, μ_r, μ_g, μ_b 即反映整帧图象的平均色.

其二阶彩色矩可表示为^[3]

$$\begin{aligned}\sigma_r &= \left(\frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (f_r(i, j) - \mu_r)^2 \right)^{1/2} \\ \sigma_g &= \left(\frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (f_g(i, j) - \mu_g)^2 \right)^{1/2} \\ \sigma_b &= \left(\frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (f_b(i, j) - \mu_b)^2 \right)^{1/2}\end{aligned}\tag{11}$$

$\sigma_r, \sigma_g, \sigma_b$ 分别表示该帧图象中红(R)、绿(G)、蓝(B)彩色分量的均方差, 不失一般性, 仅考虑彩色的 r 分量时, 若其量化值 $f_r = 0, 1, \dots, C$, r 分量为 f_r 的象点数 n_r , 则 σ_r 可表示为

$$\sigma_r = \left(\frac{1}{m \times n} \sum_{f_r=0}^C n_r \cdot (f_r - \mu_r)^2 \right)^{1/2}$$

这里, $\sigma_r, \sigma_g, \sigma_b$ 即在一定程度上反映整帧图象彩色分量的分布特征及图象颜色的‘多彩’性.

为了弥补由于使用子空间彩色直方图对图象特征描述所产生的误差, 一阶、二阶彩色矩的计算可在整个红、绿、蓝彩色空间进行. 对于 HVC 空间, 其彩色矩计算式与式(10), (11)类似. 表 2 给出了图 2(a)原始图象 HVC 彩色空间计算的彩色矩数据.

表 2 样例图象的彩色矩

μ_H	μ_V	μ_C	σ_H	σ_V	σ_C
134.891	7.416	2.337	55.266	1.077	1.635

3 镜头分割

由于分镜头是视频序列图象中, 同一内容信息的最短序列图象, 常有相近的场景, 在空间及时间上有较强的相关性, 因此整帧图象的总体特征有较好的相似性.

3.1 相似性度量

由式(9)可知, 对任一帧 $m \times n$ 大小的彩色图象, 其象点彩色按 N 种彩色分类即可得到相应的直方图

$$h = \{h(0), h(1), \dots, h(N)\}\tag{12}$$

该式可看成为一个 $(N+1)$ 维矢量, 每一分量值即为该帧图象中相应类别彩色的象点个数, 而任两帧图象的直方图则可表示为两个 $(N+1)$ 维矢量, h_i

和 h_j , 那么

$$\begin{aligned} d_h &= (h_i - h_j)(h_i - h_j)^T \\ &= [(h_i(0) - h_j(0)), (h_i(1) - h_j(1)), \dots, (h_i(N) - h_j(N))] \cdot \\ &\quad [(h_i(0) - h_j(0)), (h_i(1) - h_j(1)), \dots, (h_i(N) - h_j(N))]^T \end{aligned} \quad (13)$$

即反映了两帧图象的相似程度, 且 d_h 越小, 两帧图象就越相似。

由式(10)和式(11)可知, 对任一 $m \times n$ 大小的彩色图象, 可由该两式求出 3 个一阶彩色矩 μ_r, μ_g, μ_b 及 3 个二阶彩色矩 $\sigma_r, \sigma_g, \sigma_b$, 并可将其表示为一个六维矢量

$$m = \{\mu_r, \mu_g, \mu_b, \sigma_r, \sigma_g, \sigma_b\} \quad (14)$$

如果任两帧彩色图象的彩色矩可表示为两个六维矢量 m_i 和 m_j , 那么

$$\begin{aligned} d_m &= (m_i - m_j)(m_i - m_j)^T \\ &= [(\mu_{ir} - \mu_{jr}), (\mu_{ig} - \mu_{jg}), \dots, (\sigma_{ib} - \sigma_{jb})] \cdot \\ &\quad [(\mu_{ir} - \mu_{jr}), (\mu_{ig} - \mu_{jg}), \dots, (\sigma_{ib} - \sigma_{jb})]^T \end{aligned} \quad (15)$$

也反映了两帧图象的相似度, 且 d_m 越小, 两帧图象越相似。

若考虑到这两种反映颜色特征的相似度, 则两帧图象的相似度函数可表示为

$$\begin{aligned} d_{ij} &= A(h_i - h_j)(h_i - h_j)^T + \\ &\quad (m_i - m_j)B(m_i - m_j)^T \end{aligned} \quad (16)$$

其中, A 为权值, B 为权值矩阵

$$B = \begin{bmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & b_6 \end{bmatrix}$$

3.2 镜头分割

对任一视频序列图象, 取第一帧 F_0 为起始帧, 将帧号 F_0 存入 F_{1s} , 先按式(12)和式(14)来计算 h_0 和 m_0 , 然后再计算下一帧 $F_i (i = 1, 2, 3, \dots)$ 的两个相应矢量 h_i 和 m_i , 按式(16)计算首帧与第 i 帧的相似度

$$d_{0i} = A(h_0 - h_i)(h_0 - h_i)^T + (m_0 - m_i)B(m_0 - m_i)^T \quad (17)$$

若 $d_{0i} \leq T$ 时, $i = i + 1$, 直至第 k 帧, $d_{0k} > T$ 时, 则将第 $(k - 1)$ 帧帧号存入 F_{1e} , 而将第 k 帧帧号存入 F_{2s} , 如此类推, 重复以上操作, 直至该视频序列图象结束, 其作后一帧图象即为最后一个分镜头的末帧。如果该段视频序列图象可分为 M 个分镜头,

则其分镜头序列的头帧末帧帧号可表示为

$$(F_{1s}, F_{1e}), (F_{2s}, F_{2e}), \dots, (F_{Ms}, F_{Me}) \quad (18)$$

4 视频重组

在得到一段视频序列图象的分镜头及每一分镜头首帧及末帧帧号序列后, 就可进行视频重组, 视频重组就是从分镜头中求出重组序列图象的帧号序列, 进而经滤波器得到重组序列图象。

该系统给出了简化重组、比例重组及自适应重组等 3 种重组模式。

4.1 简化重组

简化重组是从分镜头中按照时间轴等距地提取 N 帧图象 ($N \geq 1$), 以组成重组序列图象, 这种重组序列图象中即包含每一分镜头中的图象。

若 $N = 1$ 时, 则选取每一分镜头的首帧图象 (帧号 F_{is}), 以组成重组序列图象。

当 $1 < N < (F_{ie} - F_{is})$ 时, 重组序列的第 i 个分镜头第 j 帧帧号 F_{ij} 为

$$F_{ij} = \text{Int}[(j - 1)(F_{ie} - F_{is})/N + F_{is}] \quad (19)$$

$$1 \leq j \leq N$$

其中, $\text{Int}(\cdot)$ 为取整函数, F_{is}, F_{ie} 分别为第 i 个分镜头首、末帧帧号。

当 $N \geq (F_{ie} - F_{is})$ 时, 则 F_{is} 到 F_{ie} 所有帧都为重组序列。

简化重组计算简单、速度较快, 且不会漏掉一个分镜头, 因而较适用于比较快速的浏览, 但该重组方法对长短分镜头需取相同帧数, 因而不能很好地反映长分镜头的内容。

4.2 比例重组

比例重组是从每一分镜头中, 按照预先给定比例选取一定帧数来组成重组序列图象, 即较长的分镜头选取较多的帧数, 而较短的分镜头则选取较少的帧数。

若从第 i 个分镜头选取 $\alpha\%$ ($\alpha < 100$) 帧数重组序列图象, 则

$$N = \text{Int}[(F_{ie} - F_{is}) \cdot \alpha/100] \quad (20)$$

当 $N \leq 1$ 时, 则选取 F_{is} , 而当 $N > 1$ 时, 则重组序列图象帧号选取的主要过程如下:

首先将该分镜头内序列图象, 按每小段 m 帧 (如 $m = 3$ 时的分段如图 3 所示) 分段^[4], 最后若不足 m 帧也分为一段, 但每段至少 2 帧;

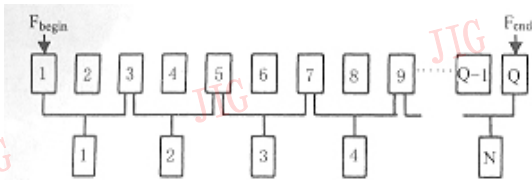


图 3 比例重组方式中的分组方法

再计算每一小段第一帧与最后一帧的相似度 d_{kl} (式 16),并将这些小段按相似度从大到小排列;然后删除相似度值较小的小段,并将留下的帧数与 N 比较,若保留下来的帧数大于 N ,则将其按帧号大小顺序重新排列,重新分段,如此重复,直至保留帧数小于或等于 N 为止.这样即可得到该分镜头重组序列图象的帧号序列.

应用该方法,能较好地反映长短分镜头的不同情况,从而便于控制重组序列的大小.

4.3 自适应重组

自适应重组可依据图象变化的剧烈程度来选择相应的帧数,以组成重组图象序列.图象变化越剧烈,选取的帧数就越多.具体算法如下:

定义 整型变量 F_K 为当前选取帧,用于存储当前选取帧帧号;整型变量 F_L 为当前判决帧,用于存储当前判决帧帧号;整型变量链表 H 用于存储重组序列图象的帧号序列.

(1)初值 对第 i 个分镜头,当前选取帧帧号 F_K 的初值为 $F_K = F_{is}$,当前判决帧帧号 F_L 的初值为 $F_L = F_{is} + 1$,链表 H 初始仅有一个节点,节点值为 F_{is} .

(2)按式(16)计算当前判决帧 F_L 与当前选取帧 F_K 的相似度 d_{KL} ;

(3)若 $d_{KL} \leq T_h$,则 $F_K = F_K, F_L = F_L + 1$;
若 $d_{KL} > T_h$,则 $F_K = F_L, F_L = F_K + 1$,向链表 H 中添加值为 F_K 的新节点;

(4)若 $F_K < F_{ie}$ 并且 $F_L \leq F_{ie}$,跳回到(2),否则算法结束.

算法结束后,从链表 H 即可得到重组序列图象的帧号序列.

该方法针对分镜头内场景变化能自适应性地控制重组序列的大小,故使变化细节得到了很好体现.

5 实验结果

为了验证上述方法,因而在一台 Pentium II PC

机上对文中所述方法进行了模拟.实验中,用的视频序列图象均来自中国科学技术大学电子科学与技术计算机视觉实验室广播电视节目样例数据库^[8].彩色直方图计算是使用 Yihong Gong 等人^[7]提出的方法,其 HVC 彩色空间中 11 种彩色子空间(见表 1),其彩色矩也是在 HVC 彩色空间中计算的.

实验 1 “简化重组”模拟,选用由 6 个分镜头组成的一段新闻报道序列图象,总长度为 213 帧,每个分镜头取一帧($N = 1$).图 4 给出了模拟获得的重组序列图象.



图 4 简化重组($N = 1$)的重组序列图象(原图象为彩色)

实验 2 “比例重组”模拟,选用分镜头较多,但场景变化慢的新闻报道节目以及分镜头多,且场景变化快的体育节目进行模拟.实验中, m 值取 3,删除因子 s 取 5%;15%.表 3 给出了实验 2 的统计结果.

表 3 “比例重组”实验结果统计表

内容	重组比例 α	重组压缩率 %	重组正确率 %
体育节目	5	约 5	95.5
	15	约 15	98.8
新闻报道节目	5	约 5	95.5
	15	约 15	99.0

表 3 中的统计数据定义如下:若设视频序列图象分割为 M 个分镜头,各分镜头长度为 $L_i, i = 1, 2, \dots, M$,其重组序列图象帧数为 $S_i, i = 1, 2, \dots, M$,则重组压缩率定义为 $\frac{1}{M} \sum S_i / T_i$,而重组正确率定义为 $\frac{1}{M} \sum S_i / \text{Int}(T_i \times s)$.

实验 3 “自适应重组”模拟,为了与比例重组相比较,本文还选用一组分镜头较多,但场景变化慢的新闻报道节目和一组分镜头多且场景变化快的体育节目组成的一段视频,同时作自适应重组和比例重组模拟,结果表明整个视频段的重组压缩率在两种方法下保持一致(15%).表 4 给出模拟实验中场景变化最大的 3 个镜头 $S_{F_1}, S_{F_2}, S_{F_3}$ 和变化最缓慢

的三个镜头 $S_{S_1}, S_{S_2}, S_{S_3}$ 的实验结果比较,其中 T 为镜头长度, N_1 为比例重组的帧数, N_2 为自适应重组的帧数. 统计结果显示出场景变化快的镜头其 N_1/N_2 都小于 1,而场景变化慢的镜头其 N_1/N_2 都大于 1,这表明,比例重组适于反映镜头长度,而自适应重组则适于反映镜头的场景变化速率.

表 4 “比例重组”和“自适应重组”比较统计表

	场景变化最快的 3 个镜头			场景变化最慢的 3 个镜头		
	镜头 S_{F_1}	镜头 S_{F_2}	镜头 S_{F_3}	镜头 S_{S_1}	镜头 S_{S_2}	镜头 S_{S_3}
T	71	62	60	103	86	77
N_1	11	9	9	16	13	12
N_2	18	15	15	9	7	6
N_1/N_2	0.61	0.6	0.6	1.78	1.86	2.0

实验结果表明,3 种重组都得到较好结果,但对于不同的视频对象应选择不同的重组原则,方能取得好的综合效果,例如,对于新闻报道节目这种分镜头内场景变化较小的视频对象,选择‘比例重组原则’比较合适,而对于体育节目这种分镜头内场景变化比较剧烈的视频对象,自适应重组则是比较好的选择.

6 结 论

本文描述了一种基于内容的视频重组方法,该方法是用整帧图象的彩色直方图、彩色矩等图象彩色信息,来进行帧间图象的相似度度量,并将其用于分镜头切分,进而给出简化重组、比例重组和自适应重组等重组方法.

上述方法,已在一台 Pentium II PC 机上进行了模拟,实验结果表明,对于将彩色视频序列图象中,各象点按其到彩色子空间(如表 2)投影进行分类的方法是可行的. 不仅可使彩色直方图简化,便于存储及进一步处理,也能满足分镜头切分对特征表示的需求. 若将彩色直方图、彩色矩用于各帧图象相似度的计算,则能够实现分镜头分割及分镜头的简化重组,比例重组,自适应重组等,且已获得较好的重组结果. 在基于内容方法模拟过程中,程序运行稳定可靠,各种重组方式都能获得预期的结果. 这表明该方

法用于视频序列图象的重组是可行的.

上述 3 种适用于不同要求和目的的重组方式,使整个视频重组系统有了较高的效率和灵活性,已可用于视频数据库及其他视频信息数据的快速预览,也可用于基于 Browser/Server 架构的 Web 型视频数据库的视频预览、视频摘要、视频检索以及低带宽条件下替代源视频对象等.

参 考 文 献

1 John R Smith. Digital video libraries and the internet. IEEE Communications Magazine, 1999, 37(1) :9297.

2 孙龙祥,程义民,王以孝等. 深度图象分析. 北京:电子工业出版社,1996.9.

3 Hongjiang Zhang, Jianhua Wu, Di Zhong *et al.* An integrate system for content-based video retrieval and browsing. Pattern Recognition, 1997, 30(4) :643658.

4 Xinding Sun, Kankanhalli Mohan S, Yongwei Zhu *et al.* Content-based representative frame extraction for digital video. Multimedia Computing and Systems, In IEEE international conference, Austin Teras Jun 28-Jul 1,1998 :190193.

5 程义民,丁红侠. 基于几何特征的深度图象识别. 中国图形图象学报(A 版),2000,5(7) :573579.

6 Gargi U, Oswald S, Kosiba D *et al.* Evaluation of video sequence indexing and hierarchical video indexing. In Proc. SPIE Storage Retrieval Image Video Databased III, 1995,2420 :144151.

7 Yihong Gong, Chua Hock Chuan, Guo Xiaoyi. Image indexing and retrieval based on color histograms. Multimedia Tools and Applications 2, 1996,2 :133156.

8 罗瑞江. 视频图象数据库系统设计方法研究[硕士论文]. 合肥:中国科学技术大学,1999.7.

程义民 1969 年毕业于中国科学技术大学无线电系计算机专业,现为中国科学技术大学电子科学与技术系教授. 目前感兴趣的领域为立体视觉、纹理分析、医学图象分析、深度图象分析、视频图象传输及多媒体技术等. 已发表论文 40 余篇.

沈 勇 1974 年生,中国科学技术大学电子科学与技术系电路与系统专业硕士研究生. 研究方向为视频处理、网络视频数据库等.

王以孝 1970 年毕业于中国科学技术大学无线电系,现为中国科学技术大学电子科学与技术系副教授.

罗瑞江 1972 年生,1999 年获中国科学技术大学硕士学位. 现在新加坡国立大学攻读博士学位.