

# 一种基于内容的视频情感类型识别算法<sup>①</sup>

闫乐林, 冯希叶

(齐鲁师范学院 计算机科学与技术系, 济南 250013)

**摘要:** 提出了一种新的视频语义分析算法, 着重对情感内容识别进行了研究。算法考虑了人的认知行为具有模糊性和不确定性的特点, 融合了未确知数学理论, 建立了视频低层特征和情感类型之间的关系模型。视频低层情感特征提取部分, 详细介绍了特征选取的依据和具体方法。视频情感类型判定部分, 描述了未确知情感测度矩阵的构建和情感类型判定方法。实验结果表明, 该算法是有效的、可行的。

**关键词:** 情感特征; 视频语义; 情感类型; 未确知测度

## A Content-Based Retrieval Algorithm for Video Emotion Types

YAN Le-Lin, FENG Xi-Ye

(Department of Computer Science and Technology, Qilu Normal University, Jinan 250013, China)

**Abstract:** A novel content-based retrieval algorithm is presented for video emotion types recognition in this paper. Considering the fuzziness and uncertainty of the human's cognitive behavior, we established the relation between low-level video features and high-level emotion types by the unascertained theory. The principle of video features selection and implementing method are introduced in detail in the part of video low-level emotional feature extraction. And the specific process of building the measure matrix and the degree of confidence method are discussed in the part of emotion type determination. Experimental results show that the proposed algorithm is feasible and effective.

**Keywords:** emotion feature; video semantic; emotion types; unascertained measure

随着网络和多媒体技术的发展, 数字视频逐渐成为信息系统中重要的媒体资源, 其数量以惊人的速度不断增长。如何对海量非结构化的视频数据进行有效的组织、表达和管理, 以方便人们的快速浏览和准确查询, 实现个性化信息服务, 成为目前迫切要解决的问题。视频情感内容是包含在视频中, 期望用户在观看过程中被诱发的情感强度和情感类型。目前, 国内外不少学者从事视频情感内容的研究, 主要通过模式分类器、规则推理等方法建立低层特征空间和基本情感类型空间之间的映射关系, 并依据此关系识别视频的情感类型<sup>[1,2]</sup>。然而, 情感属于心理学范畴, 具有典型的不确定性和未知性<sup>[3]</sup>。不同的观众对同一视频的情感内容的认同度有很大差异性。上述文献中注重对情感内容的识别, 但均未考虑情感的这种不确定性。本文基于未确知数学理论, 提出了一种基于未

确知测度的情感内容识别新算法, 用来识别视频中的情感类型。

### 1 视频情感低层特征提取

镜头是电影拍摄过程中使用的基本单位, 展现视频情感内容通常一个镜头难以实现, 因此, 本文选取视频场景片段作为研究情感内容的基本单位。统计发现, 视频(尤其是电影)情感常通过特定的形式被表达和渲染, 可以断定情感内容和视频的某些低层特征必定存在内在联系。比如, 导演通常使用灯光效果刻画场景氛围, 特效镜头表达特殊的情感体验以及使用冷色调展现演员的心情等<sup>[4]</sup>。鉴于此, 我们选用场景亮度、场景节奏和颜色能量作为分析视频情感内容的三个低层特征指标。为简单起见, 本文采用文献[5]的镜头分割算法来分割视频片段的镜头, 采用文献[6]的算

<sup>①</sup> 基金项目:山东省教育厅自然科学基金(J07WJ16)

收稿时间:2010-07-07;收到修改稿时间:2010-08-25

法进行视频场景分割提取。

### 1.1 场景亮度

对于一幅 RGB 模式或其它颜色模式图像,在图像处理中为降低复杂度常将其处理为对应的灰度图像。灰度图中所有像素不再有颜色信息而仅含一个亮度指标,称其为亮度分量(记为  $L$ )。亮度分量取值范围为 0 至 255,共 256 个值,对应像素点由黑到白的亮度变化。通过对图像主观实验发现,High-key 视频帧中的  $V$  值(HSV 颜色空间中的亮度)的平均值和标准差比较高,而 High-key 视频帧中值的平均值和标准差比较低<sup>[7]</sup>。因此,视频场景中第  $k$  帧的亮度线索具体计算为公式(1),用符号  $FL$  来表示。

$$FL(k) = \mu_k \times \sigma_k \quad (1)$$

其中  $\mu_k$  和  $\sigma_k$  分别表示第  $k$  帧中  $V$  值的平均值和标准差。本算法情感计算以视频场景为单位,因此在公式(1)基础上,场景亮度指标定义如下,记为  $ScL$ :

$$ScL(l) = \frac{\text{median}\{FL_{it}(KF)\}}{\max_{1 \leq l \leq Q} \{\text{median}\{FL_{it}(KF)\}\}} \quad (2)$$

其中  $FL_{it}(KF)$  表示第  $l$  个场景中第  $t$  个镜头的关键帧(镜头中间帧)的照明线索指标值, $Q$  表示视频片段内的场景总数,  $\text{median}()$  与  $\text{max}()$  分别为取中值和取最大值运算。

### 1.2 场景节奏

统计发现,视频场景中单镜头的持续时间和场景中镜头变化次数常作为电影等视频场景情感表达的一种重要手段。例如在电影中,当镜头切换较为频繁时,通常是故事情节比较激烈,相应引发观众的情感反应强度较明显。因此,本为引入一个参量,即场景节奏,作为一个视频情感的低层特征指标,反应视频场景的情感变化。设某视频场景有  $n$  个镜头,按视频的时间轴顺序将镜头长度记为  $x_1, x_2, \dots, x_n$ 。因此,视频场景场景节奏模型可由公式(3)定义,用符号记作  $ScR$ :

$$ScR = \frac{\min(x_1, x_2, \dots, x_n)}{\text{mean}(x_1, x_2, \dots, x_n)} \cdot e^{-\frac{1}{n}} \quad (3)$$

其中  $\text{mean}()$  是场景内所有镜头长度的平均值,函数  $\min()$  为取最小值计算,  $t$  是场景中镜头长度不小于镜头长度平均值的个数。

### 1.3 颜色能量

颜色是图像处理领域的一个重要要素,通过对颜

色变化的统计分析,能够得到图像的很多信息,如对象边缘检测、兴趣区域定位、相似图像匹配等。颜色对人的情感能产生重要的影响已是不争的事实,影视作品和绘画艺术中常使用特定色彩来表达导演或作者的情感意图。所谓颜色能量是指某一色彩对人类产生的相对美学冲击<sup>[1]</sup>。具体的颜色能量取决于:1)其色调、饱和度和亮度的属性;2)彩色区域的大小;3)前景和背景间色彩的对比度。同时,颜色属性(如色调、亮度、饱和度、颜色面积和对比度)和颜色能量之间也存在某些特定关系。在 HLS 颜色空间中,颜色能量提取模型如公式(4),用符号记为  $CP$ 。

$$CP(k) = \left( \sum_i \sum_j p(c_i) p(c_j) d(c_i, c_j) \right) \times \left( \sum_k^M E(h_k) s_k v_k \right) \quad (4)$$

其中  $c$  是图像中 HLS 颜色直方图区间数,  $i, j$  分别是区间数的索引号,  $p(\bullet)$  为直方图比例,  $d(c_i, c_j)$  为 HLS 颜色空间中范数,  $M$  是图像像素总和,  $s_k, v_k$  分别是饱和度和亮度,  $E(h_k)$  是满足与红色和绿色的角距离在  $[0.75, 1.25]$  内的像素色调的能量。

用公式(5)将场景的颜色能量归一化处理,使其值均分布在  $[0, 1]$  区间内,以便比较场景间的差异。第  $i$  个场景的颜色能量(用符号  $ScCP$  表示)可表示为:

$$ScCP(i) = \frac{\text{mean}(CP(F(t, n)))}{\max_n(\text{mean}(CP(F(t, n))))} \quad (5)$$

其中  $F(t, n)$  为第  $n$  个场景的第  $t$  个镜头的关键帧,  $\text{max}()$  为取最大值运算,  $\text{mean}()$  为取平均值运算。

综上所述,我们可以用一个三维特征向量来表示视频场景的情感内容,记为  $(ScL, ScR, ScCP)$ ,分量分别为视频场景的场景亮度、场景节奏和颜色能量。

## 2 视频场景情感类型判定模型

未确知测度情感判定模型由两部分组成:测度矩阵和置信度情感类型判定。未确知测度情感判定模型的输入是视频场景的三维低层情感特征向量,输出是视频的情感类型。

### 2.1 视频情感未确知测度矩阵

设  $x_1, x_2, \dots, x_n$  是  $n$  个视频场景对象,以  $X$  表示视频场景对象空间,则:

$$X = \{x_1, x_2, \dots, x_n\}$$

在本文中，要判定视频对象  $x_i$  情感类型需要测量 3 个指标：场景亮度、场景节奏、颜色能量，另外将视频的情感类型分为害怕、悲伤和高兴三种基本情感类型。令视频场景情感度量的指标空间  $I = \{I_1, I_2, I_3\}$ ，其中  $I_1$  表示场景亮度、 $I_2$  表示场景节奏、 $I_3$  表示颜色能量；情感判定空间  $U = \{c_1, c_2, c_3\}$ ， $c_1$  表示害怕、 $c_2$  表示悲伤、 $c_3$  表示高兴。若  $x_{ij} (i=1,2,\dots,n; j=1,2,3)$  表示第  $i$  个视频场景对象  $x_i$  关于第  $j$  个指标  $I_j$  的测量值，则  $x_i$  可表示为一个三维向量  $x_i = (x_{i1}, x_{i2}, x_{i3})$ 。设  $\mu_{ijk} = \mu(x_{ij} \in c_k)$  表示测量值  $x_{ij}$  属于第  $k$  个情感类型  $c_k$  的程度，若  $\mu$  满足非负性、归一性、可加性，称  $\mu$  为未确知测度，简称情感测度。

为了确定未确知情感测度函数，邀请 10 名受测者参与实验。对于一段给定的视频场景，首先根据前面的特征提取算法计算出视频情感特征向量 (ScL, ScR, ScCP)，再根据受测者观看影片时被诱发的情感类型报告，建立 (ScL, ScR, ScCP) 各分量到情感空间  $\{c_1, c_2, c_3\}$  各分量的映射，即确定情感测度  $\mu_{ijk}$ 。重复上面的识别训练过程，对于多个场景视频给予处理，通过数据的统计分析，得到场景亮度、场景节奏、颜色能量三个指标的单指标测度函数，见图 1。

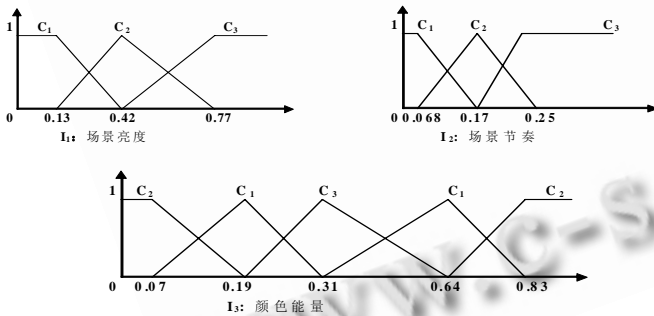


图 1 场景亮度、场景节奏、颜色能量的单指标测度函数

由上面三个单指标测度函数，可以构造矩阵  $(\mu_{ijk})_{3 \times 3}$ ，我们称矩阵  $(\mu_{ijk})_{3 \times 3}$  为单指标情感判定矩阵。

$$(\mu_{ijk})_{3 \times 3} = \begin{bmatrix} \mu_{i11} & \mu_{i12} & \mu_{i13} \\ \mu_{i21} & \mu_{i22} & \mu_{i23} \\ \mu_{i31} & \mu_{i32} & \mu_{i33} \end{bmatrix}$$

下面举例说明应用单指标测度函数计算情感判定矩阵的过程：

设有某一视频场景  $x_i$ ，根据上文的视频情感低层特征提取模型公式，计算出该场景对应特征向量为  $x_{ij} = (0.1865, 0.1204, 0.2903)$ ，即场景亮度  $I_1 = 0.1865$ 、场景节奏  $I_2 = 0.1204$ 、颜色能量  $I_3 = 0.2903$ 。因  $I_1 = 0.1865 (0.1865 \in [0.13, 0.42])$ ，由图 3-2 中场景亮度单指标测度函数图可知：该视频场景  $x_i$  的场景亮度在  $c_1$  (害怕) 和  $c_2$  (悲伤) 两种情感类型有度量值，而在  $c_3$  (高兴) 情感类型上无度量值。由场景亮度测度函数可计算

$$\mu_{i1k} \text{ (} \mu_{i1k} \text{ 表示视频场景 } x_i \text{ 的场景亮度指标 } I_1 \text{ 属于第 } k(k=1,2,3) \text{ 个情感类型 } c_k \text{ 的程度)}.$$

$$\mu_{i11} = \frac{0.1865 - 0.42}{0.13 - 0.42} = 0.8052, \mu_{i12} = \frac{0.1865 - 0.13}{0.42 - 0.13} = 0.1948,$$

$$\mu_{i13} = 1 - \mu_{i11} - \mu_{i12} = 0$$

类似的可计算场景节奏和颜色能量的测度函数值，从而得到此视频场景的单指标情感判定矩阵  $(\mu_{ijk})_{3 \times 3}$ ：

$$(\mu_{ijk})_{3 \times 3} = \begin{bmatrix} 0.8052 & 0.1948 & 0 \\ 0.4863 & 0.5137 & 0 \\ 0.1642 & 0 & 0.8358 \end{bmatrix}$$

## 2.2 置信度情感类型判定

对于任何多指标决策评价系统，各评价指标的相互重要程度即指标权重互不相同，不同的权重对应不同的评价结果，合理的确定指标权重对评价系统是十分重要的。本文采用信息熵<sup>[8]</sup>确定构建情感判定矩阵的三个情感低层特征指标的权重。

$$W_{ij} = 1 + \frac{1}{\log_2 3} \sum_{k=1}^3 \mu_{ijk} \log_2 \mu_{ijk} \quad (6)$$

反映第  $i$  个对象第  $j$  情感指标的重要程度。公式(7)计算各指标的相对权重值。

$$w_{ij} = \frac{W_{ij}}{\sum_{j=1}^3 W_{ij}}, \sum_{j=1}^3 w_{ij} = 1 \quad (7)$$

$w = (w_{i1}, w_{i2}, w_{i3})$  即为三个情感指标的权重向量。

因此，视频场景情感类型识别向量  $\mu_i = (\mu_{i1}, \mu_{i2}, \mu_{i3})$  可由公式(8)计算得到：

$$\mu_i = w \bullet (\mu_{ijk})_{3 \times 3} \quad (8)$$

本文情感空间为有序划分，最大隶属度识别原则不适用，故采用置信度识别准则：

设定置信度  $\lambda (0.65 \leq \lambda < 1)$ ，若

$$k_0 = \min \{k \mid \sum_{j=1}^k \mu_{ij} \geq \lambda, 1 \leq k \leq 3\} \quad (9)$$

则判定识别的视频场景的情感类型为  $C_{k_0}$ 。

### 3 实验结果和分析

本文采用主观评价的方法检验上述算法的性能，10 名测试者被要求参与实验过程。仿真环境为 Win XP(SP2)& Visual C++ 2005 平台，安装配置 Open CV 视频处理工具包。实验所用数据来源于 4 部电影，分别是《大腕》、《脸》、《男才女貌》、《暖春》，其总长度分别为：47:23、39:45、58:24、33:18。

首先，将电影按照场景分割为片段，并编号，影片总共被分割为 132 个场景。其次，请 10 个受测者分别观看场景片段，并将编号的场景分类为高兴、悲伤、害怕和其他四类<sup>[9]</sup>；然后，按照 70% 认同度将场景视频按情感统计归类；最终得到了情感特征明显的 35 个高兴场景、26 个悲伤场景、19 个害怕场景，另有 52 个场景情感特征不显著。

以韩国电影《脸》片段中第 1486 帧至第 2379 帧间场景为例(部分视频关键帧如图 2)。根据本文提出的算法及前面的各算法公式，可计算出该场景对应特征向量为(0.1869,0.1207,0.2132)，视频场景单指标情感测度矩阵为：

$$(\mu_{ijk})_{3 \times 3} = \begin{bmatrix} 0.8038 & 0.1962 & 0 \\ 0.4833 & 0.5167 & 0 \\ 0.8067 & 0 & 0.1933 \end{bmatrix}$$

三个底层情感特征指标的权重向量为(0.4723,0.3016,0.2261)，对应的情感类型识别向量为(0.7078,0.2485,0.0437)，取置信度  $\lambda = 0.65$  用未确知情感测度判定识别结果为害怕，这也与主观观察结果一致。



图 2 场景片段关键帧

选用情感分类后情感特征明显的视频片段作为样本进行仿真实验，用于算法的性能检验。实验结果见表 1。

表 1 实验结果

仿真 样本	识别结果		
	高兴	悲伤	害怕
高兴	26 (24)	2 (5)	3 (5)
悲伤	4 (6)	18 (18)	3 (3)
害怕	5 (4)	5 (3)	14 (12)

为验证算法性能，定义两个指标：精确度和识全率<sup>[10]</sup>。根据表 1，计算得到情感类型识别精度和识全率，见表 2。

表 2 实验情感类型检测精确度和识全率

情感类型	正确检测	丢失	虚假	精确度 (%)	识全率 (%)
高兴	26 (24)	9 (11)	9 (13)	74.3 (64.9)	74.3 (68.6)
悲伤	18 (18)	8 (8)	7 (8)	72 (69.2)	69.2 (69.2)
害怕	14 (12)	5 (7)	6 (8)	70 (60)	73.7 (63.2)

表 1、2 中的括号中的数据是采用文献 11 中的算法对相同电影场景数据进行的实验结果。表 2 数据表明，应用本文算法，实验样本中大部分情感内容能够被准确的识别出来，精度超过 70%，总体比算法文献 11 的性能要略好。

### 4 结论

本文基于未确知数学理论，使用场景亮度、场景节奏和颜色能量三个基本视频低层特征作为指标参数，构建了一种未确知测度的情感类型判定模型，提出了一种视频情感内容识别新算法。实验结果表明，我们提出的算法是有效的，研究的技术路线是可行的。本算法为研究视频情感类型的识别开辟了一个新视角，并相信，随着对于视频中声音、文字等特征情感属性的深入研究及系统融合，将来基于未确知测度的视频情感内容识别性能会有所提高。

#### 参考文献

- 1 Wang HL, Cheong LF. Affective Understanding in Film. IEEE Transactions on Circuits and Systems for Video Technology, 2003,16(6):689-703.
- 2 Hanjalic A, Xu LQ. Affective Video Content Representation and Modeling. IEEE transactions on multimedia, 2005,7(1):143-154.
- 3 刘开第,吴和琴,等.未确知数学.武汉:华中理工大学出版社, 1997.

(下转第 79 页)

目大于 100 时, 由于系统调度策略的影响等, CPU 利用率开始显著增长, 并开始有丢包, 此时聚合输出带宽仍然平滑增长; 最后, 当并发流数目升至 240 时, CPU 利用率达到 95%, 丢包率到达流媒体服务最大允许的丢包阈值 7%, 聚合输出带宽也达到峰值 280Mb; 随后再增多并发流数目, 服务器已经不能正常提供服务。

## 6 结论

随着支持 RTSP 协议的流媒体服务器和终端产品的相继推出, 如何全面评测这些产品的性能成为一个迫切需要解决的问题。本文提出了一种基于 RTSP 流媒体服务器测试工具的设计和实现方法, 测试工具能够根据具体需要设定测试用例, 并统计得出该测试过程中服务器各性能指标值, 其本身不能完全代替分析人员, 但在很大程度上为测试结果的确定提供了最直接客观的量化依据。需要说明的是, 该工具只针对服务器单个的基本性能参数进行的量化分析, 而不同的流媒体服务提供商因具体业务的不同而对服务器各个性能指标的要求程度也不同, 下一步的工作可以建立一个分析服务器整体性能的数学模型, 通过模型更加全面的量化各种流媒体服务器的综合性能。

(上接第 105 页)

- 4 林新棋. 基于模糊理论的电影情感内容识别. 北京: 北京邮电大学信息与通信工程学院, 2009.
- 5 Dai XW, Cai ZP, Zhong GY. Video Shot Detection Method Based on Wavelet Transform and Wavelet Entropy. *Journal of Southwest Jiaotong University*, 2008,43(3):314-318.
- 6 Wang XJ, Ding HT, Chen HX. A Shot Clustering Based Approach for Scene Segmentation. *Journal of Image and Graphics*, 2007,12(12):2127-2131.
- 7 Rasheed Z, Sheikh Y, Shah M. On the use of computable features for film classification. *IEEE Trans. Circuits Systems for Video Technology*, 2005,15(1):52-64.
- 8 Wu MS(translated). *Probability and Information*. Shanghai Science and Technology Press, 1964. 29-88.
- 9 Xu M, Chia LT, Jin J. Affective content analysis in comedy and horror videos by audio emotional event detection. *Proc. of IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands*, 2005,7:621-625.
- 10 Su JH, Huang YT, Yeh HH, et al. Effective content-based video retrieval using pattern-indexing and matching techniques. *Expert Systems with Applications*, 2010,37(11):5079-5085.
- 11 Lin XQ, Wen XM, Lu ZM, et al. Video Affective Content Recognition Based on Film Grammars and Fuzzy Evaluation. *Proc. of the 2008 International Conference on MultiMedia and Information Technology*. Three Gorges, China, 2008:264-267.

## 参考文献

- 1 Lee YJ, Min OG, Kim HY. Performance Evaluation Technique of the RTSP based Streaming Server. *Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)*, 2005,14-16:414-417.
- 2 Cherkasova L, Staley L. Building a Performance Model of Streaming Media Applications in Utility Data Center Environment. *Third IEEE International Symposium on Cluster Computing and the Grid (CCGrid'03)*, doi:10.1109/CCGRID.2003.1199352.
- 3 茅旭峰,李俊,吴刚.流媒体服务器服务能力基准实验与性能模型. *小型微型计算机系统*,2007,2(2):204-209.
- 4 李学森,李俊,奚宏生.流媒体服务器性能评价模型. *计算机工程*,2009,35(10):270-272.
- 5 张静,雷为民.一种支持媒体流的 SIP 性能测试工具的设计与实现. *小型微型计算机系统*,2006,27(10):1831-1836.
- 6 林浒,王阳,雷为民,于东海. SIP 协议测试方法和测试工具的研究. *小型微型计算机系统*,2008,29(3):406-410.
- 7 曾彬,张大方,张玫.基于主动测试的流媒体业务性能测量工具设计与实现. *计算机应用*,2008,28(4):832-836.
- 8 Schulzrinne H, Lanphier R, Rao A. Real time streaming protocol (RTSP). RFC 2326, Internet Engineering Task Force, Apr.1998.
- 9 Schulzrinne H, Casner S, Frederick R, Jacobson V. RTP: a transport protocol for real-time applications. RFC 3550, Internet Engineering Task Force, July 2003.