

文章编号:1001-9081(2010)06-1704-04

视频情感语义分析——类型-强度分解法

郭戈,平西建

(信息工程大学 信息工程学院, 郑州 450002)

(gouge800819@163.com)

摘要:提出一种视频情感语义分析方法,利用强度和类型两个独立分量构造的二维空间来建立人类情感模型,通过多模态的听觉、视觉特征,分别构造连续型的情感强度时序曲线和离散型的情感类型时序曲线,从而建立低层特征空间与情感空间之间的映射关系,实现对视频内容的情感注释和分析。类型-强度分解曲线可客观、真实地描述视频情感内容随时间的变化情况。实验结果验证了该方法的有效性。

关键词:情感语义;情感强度曲线;情感类型曲线;情感模型

中图分类号: TP391.4 **文献标志码:**A

Affective semantic analysis in video: type-intensity decomposition method

GUO Ge, PING Xi-jian

(Institute of Information Engineering, Information Engineering University, Zhengzhou Henan 450002, China)

Abstract: This paper looked into a new direction in video content, affective semantic analysis, which can be defined as the understanding of feelings or emotions expected to arise in viewers while watching video clip. According to the approach proposed in this paper, the affective semantic space can be decomposed by two independence elements, affective intensity and affective type. Thus the affective video content was mapped onto the 2-D emotion space by the model that bridges the gap between low-level features and high-level affective semantic meaning. A number of effective audiovisual cues were formulated to help establish the two element curves. The simulations verify the method's reliability and consistency between the emotions reflected by the video and the feelings observed on the viewers.

Key words: affective semantic; affective intensity curve; affective type curve; affective model

0 引言

随着网络和通信技术的迅猛发展,可应用和处理的视频数据呈几何级数增长,传统的人工标注方法无法适应快速增长的视频数据分类和检索需求。基于内容的视频语义分析技术应运而生,该技术通过建立视频低层物理特征空间和高层语义空间的映射关系,自动获取对视频语义内容的理解。对视频语义可从感知层和情感层两个层次空间进行分析,其中感知语义提取主要是对视频客观属性的研究,着眼于视频结构、事件、对象等客观语义信息的分析;情感语义则主要描述视频的主观属性,它关心的是视频通过声音、画面、事件等信息传达给观众的感受,即观众在观赏某段视频时会产生什么样的心理体验^[1]。在情感语义空间对视频进行分类和检索,更加符合人类的认知习惯。

目前情感分析的研究工作主要集中在语音和文本两个方向,近年来延伸至图像领域^[2-4]。然而针对视频情感内容分析的相关文献和成果较少,文献[5]提出基于情感的视频场景检索算法,根据情感的相似性实现对视频场景的检索;文献[6]利用隐马尔可夫模型将体育视频镜头划分为“Happy”和“Sad”镜头。总的来说,由于情感研究涉及到心理学、生理学、信息科学等多个学科内容,基于情感的视频分析方法仍不成熟,其评估、分类和检索技术还处于探索和起步阶段。

本文提出一种情感语义分析的新方法,首先建立强度连续、类型离散的情感空间模型,从而可将视频所表示的内容进行量化表示,通过提取视频内容中能够刻画情感倾向的特征,构造情感语义的强度曲线和类型曲线,最终实现对整个视频情感内容的自动标注。

1 情感空间模型

情感理论是实现自动情感识别的基础,人类情感是一种极其复杂的现象,对其给出准确的定义和描述并非易事,在心理学领域,虽然对情感的研究已有多年历史,但至今还没有一个统一的定义和理论模型。目前在情感信息处理领域影响最大的情感理论模型主要有两种^[7]:一种是维量模型,该模型认为可由Arousal、Valence 和 Control 三个维度对情感空间进行建模,在很多研究中表明在对以视频公共情感作为研究目标时,可合理地将三维情感空间简化为由“*Arousal*”和“*Valence*”分量构成的二维情感空间(VA空间模型)。另一种较为重要的是离散情感模型,该模型认为每种情感是一个离散的实体,可将情感划分为基本类和扩展类,扩展情感是由基本情感变化组合而成,这是当前情感识别研究中使用最多的模型,基于情感的多样性,人们常常将某几种基本情感作为研究目标。

VA情感空间将人类情感以连续化的数值表示,能够直

收稿日期:2010-01-06;修回日期:2010-03-03。

作者简介:郭戈(1980-),女,河南开封人,博士研究生,主要研究方向:视频处理、模式识别; 平西建(1953-),男,河南新乡人,教授,博士生导师,主要研究方向:图像处理、计算机视觉、模式识别。

观地反映视频中情感内容的变化情况,但利用该模型输出的情感的可区分性不强,也就是说由于 Valence 分量是连续的,使得多种情感之间的交叠处比较模糊从而难以区分;离散情感模型最终输出的是几种情感类型,这类模型虽然比较符合基于类型的情感检索要求,但却无法描述每种情感的强烈程度。结合两种模型对情感的表示与分析方法,本文提出强度连续-类型离散 (Intensity Continuous-Type Discrete, IC-TD) 模型来对视频情感内容建模,其中“强度”分量用于描述对视频内容在情感上反映的强烈程度,其值是连续分布的;“类型”分量则表示所分析的几种基本情感类型,其值为离散的。IC-TD 模型是合理的,这是因为对一段视频的情感分析结论最终归结于情感的“类型”,而强度则是对情感类型的加强或削弱。由于人们评判情感类型的语言是离散的,是无法刻画连续情感分类的,若将情感类型进行无限细的粒化对于检索来说是无意义的,因此在本文中认为离散的情感“类型”更为合理。

通过以上分析,视频的情感语义空间可以用两条随时间变化的曲线来描述,一条为连续型的情感强度时间曲线,一条为离散型的情感类型时间曲线。即:

$$\Psi(t) = \{IC(t), TD(t)\} \quad (1)$$

其中: $IC(t)$ 表示 t 时刻的情感强度值,表明情感强度随时间的变化; $TD(t) \in \Theta$ 表示 t 时刻的情感类型值,其中 $\Theta = \{\theta_i | i = 0, 1, \dots, N - 1\}$ 为预先定义的情感语义概念集合(N 为情感语义概念总数),则 $TD(t) = 2\pi i(t)/N$ 为 t 时刻的情感类型值($i(t) \in \{0, 1, \dots, N - 1\}$)。本文采用的情感语义概念集合 $\Theta = \{\text{Natural}, \text{Fear}, \text{Happy}, \text{Sad}, \text{Surprise}, \text{Anger}\}$ 。

图 1 表示某段电影视频情感强度和类型分量随时间的变化曲线,强度-时间曲线上,对应点的数值越大,表明该时刻情感强度越高;类型-时间曲线上,由于以镜头为基本单位进行处理,因此在一个镜头内的情感类型时间曲线为一直线。由强度和类型时序曲线可将视频在确定时刻的情感进行量化,同时也可获取情感的变化过程,即可采用强度时序曲线和类型时序曲线对视频的情感语义进行分析。

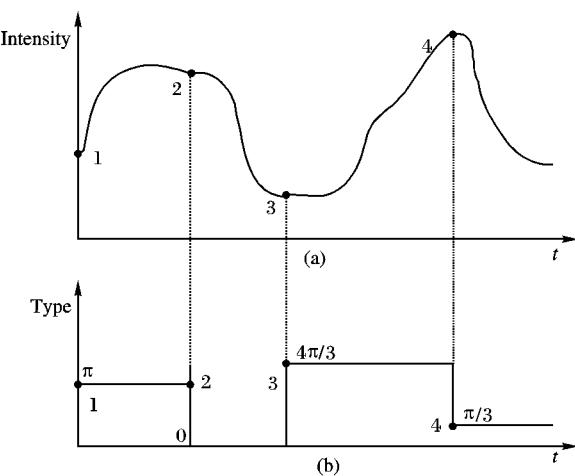


图 1 视频的强度和类型曲线

2 情感强度曲线生成

情感强度时序曲线描述了视频情感强度随时间的变化,视频中的听觉和视觉的多个特征可以用来反映这种变化过

程,假设共提取了 P 个特征参数来计算在 t 时刻时的情感强度,则视频情感强度时序曲线可定义为:

$$IC(t) = f_1(G_i(t)); i = 1, 2, \dots, P \quad (2)$$

这里用帧数来表示时间的进度,则上述曲线可转化为:

$$IC(k) = f_1(G_i(k)); i = 1, 2, \dots, P \quad (3)$$

其中: $IC(k)$ 为第 k 帧时的强度函数; $G_i(k)$ 代表每个强度特征参数随时间的变化函数,称为强度特征分量函数,由于各特征的取值范围有较大差异,需进行归一化处理; f_1 表示对这 P 个特征分量函数进行融合,称为强度融合函数。本文采用运动强度、镜头切换速率、音频能量作为特征参数来构建强度函数 $G_i(k)$ 。

2.1 运动强度

运动强度,即由一段连续视频影像的显著对象运动特性所获取的特征。运动特征与观众的情感具有密切的联系,在电影理论中也曾提及“运动特征可引起观众情感的急剧波动”,随着运动强度的增加,将直接引起情感强度的增加^[8]。这里采用标准的基于块运动估计方法来计算第 k 帧的运动矢量,以所有块(设为 B 块)运动矢量的平均幅度值进行归一化后作为第 k 帧的运动强度:

$$G_1(k) = \frac{1}{B |MV|_{\max}} \left(\sum_{i=1}^B |MV_i(k)| \right) \quad (4)$$

其中: $MV_i(k)$ 表示第 k 帧中第 i 块的运动矢量, $|\cdot|$ 表示对运动矢量取幅度值, $|MV|_{\max}$ 表示最大的运动矢量幅值,用来将 $G_1(k)$ 归一化至 $[0, 1]$ 。

2.2 镜头切换速率

镜头切换速率可直接反映一段视频影像的情感强度特征,例如在电影拍摄过程中,导演可利用镜头切换速度来控制情节的发展,当镜头切换较为频繁时,通常是激烈的故事情节,观众响应的情感强度也随之增加,缓慢的镜头切换速率对应于平缓的故事情节,情感强度也随之降低,镜头切换速率定义如下:

$$G_2(k) = e^{((1-(n(k)-p(k))/\delta))} \quad (5)$$

其中: $p(k)$ 和 $n(k)$ 分别表示第 k 帧左右两侧最近邻镜头边界的帧号, $n(k) - p(k)$ 为镜头的持续时间,该值直接决定 $G_2(k)$ 的大小,当 $n(k) - p(k)$ 变小, $G_2(k)$ 变大,即表明切换速率越高, δ 为常数,用于保证 $G_2(k)$ 分布在 $[0, 1]$,其值由 $n(k) - p(k)$ 的值决定。

2.3 音频能量

Murray 等人研究成果^[9]表明,音频流承载有大量的情感信息,音频的短时能量的大小可以用来衡量各类声音的强弱程度,音频能量参数定义为:

$$G_3(k) = \frac{1}{L |E|_{\max}} \sum_{i=1}^L |E_k(i)|^2 \quad (6)$$

其中: L 为一个音频片断中帧的个数, $E_k(i)$ 为 k 帧第 i 个音频帧的平均幅值, $|E|_{\max}$ 为帧平均幅值的最大值。

在取得上述 3 个特征参数后,定义融合函数 f 为各个特征参数的加权和,即:

$$IC(k) = \sum_{i=1}^3 \omega_i G_i(k) \quad (7)$$

为保证情感强度在相邻帧之间的连续性,减少镜头切换以及噪声对情感强度造成的影响,这里对上述强度曲线采用

Kaiser 窗函数 $k(l, \beta)$ 跟 $IC(k)$ 曲线卷积, 对情感强度曲线进行平滑, 最后将平滑后的强度曲线归一化到 $[0, 1]$ 。

3 情感类型曲线生成

人类的情感具有很大的主观性, 不同的视觉表现方式和视频的编辑手法会带给观测者不同的心理感受, 如镜头的切换速度越快表明视频内容越紧张等。由于情感类型具有短时平稳特性, 本文以镜头为基本对象, 镜头内所有的帧具有相同的情感类型, 如此得到一阶跃型的情感类型曲线, 记为:

$$TD(k) = f_2(C_i(k)); i = 1, 2, \dots, Q \quad (8)$$

其中: $TD(k)$ 为第 k 帧对应的情感类型; $C_i(k)$ 代表所提取的特征参数; Q 为提取的特征参数个数; f_2 表示通过某种关联模型将特征空间 C 映射至情感空间 TD , 即 $C \xrightarrow{f_2} TD$ 。本文中提取的用于表征情感类型的特征参数包括两个方面, 一是基于关键帧的特征, 主要有颜色特征和亮度特征; 二是时序特征, 主要有平均运动向量幅度值特征和镜头切换速率特征。

3.1 颜色特征

通常认为色彩的象征性可以引发人的联想, 如“热烈的”情感通常与红色的图像特征紧密相关, “冰冷的”则可以用蓝色和青色表示。这里利用主颜色来代表图像的色彩特征, 将 HSV 空间的 3 个分量分别进行量化, 根据量化结果, 将 3 个颜色分量合成为一个一维特征矢量:

$$q(h, s, v) = 9h + 3s + v \quad (9)$$

如此获得一个 72bin 的一维直方图, 将其进行归一化处理, 提取主颜色及其比例作为特征, 即 $C = \{(c_i, p_i); i = 1, \dots, 5; p_i \in [0, 1]\}$, 其中 c_i 表示前 5 个占主导地位的颜色, p_i 表示该颜色对应的百分比, 记为 $C_1(k) \sim C_{10}(k)$ 。

3.2 亮度特征

亮度是视频制作者向观众表达视频情感的一个重要手段, 如通过较明亮的图像, 给观看者带来愉悦、欢快的感觉, 而阴暗的图像则让观众感到悲伤的、压抑的、沮丧的, 提取图像的平均亮度特征, 记为 $C_{11}(k)$ 。

3.3 运动向量幅度值

运动是视频最独特的属性, 视频中的运动特征与精神状态之间有密切的相关性, 这一点在精神生理学中有较为明确的证明^[10]。平均运动向量的幅度值可以较好地表征视频帧间的运动情况, 当运动向量的幅度值较大时, 表现出的情感类型多为“紧张的”、“激烈的”, 而对于“平静的”、“舒缓的”情感类型则运动变化的强度较小, 该特征记为 $C_{12}(k)$, 计算方法见式(4)。

3.4 镜头长度

镜头的长度可以对观测者传达不同的情绪和情感, 这一特点在体育类视频以及电影类视频中表现尤为突出, 如“激烈的”和“紧张的”视频, 单个镜头的持续时间较短, 而且情感强度越强则切换速率越短, 反之对“平静的”、“舒缓的”、“高兴的”情感类型, 则多用较长的镜头, 提取镜头持续时间为长度特征, 记为 $C_{13}(k)$ 。

经过特征提取后, 所有得到的特征组成特征矢量 $C_i(k)$, 采用支持向量机分类机制进行情感类型的识别, 于是得到一个时间上的类别序列, 利用该序列可以对整段视频内容按照时间顺序进行情感内容标记, 从而得到情感类型时序曲线。

4 实验结果与分析

为验证强度-时序曲线模型对视频情感内容描述的有效性, 以电影视频为研究对象进行实验, 首先利用音频分离软件从视频中抽出音频信息, 并将其保存为 WAV 格式, 本文对所有的音频数据统一转化为相同的采样频率 22 050 Hz, 精度为 16 位。

对每个视频帧提取运动强度、镜头切换速率、音频能量 3 个特征参数, 经过融合函数 $f_1(\cdot)$ 后得到第 k 帧的情感强度曲线 $IC(k)$, 将每帧的情感强度归一化至 0 ~ 1, 这样无论对于任何内容的视频片段, 都可以对视频帧的情感强度差异进行定量的比较。观察原始情感强度曲线(图 2)可以看出, 曲线中包含了较多的毛刺, 这说明帧与帧之间在情感强度上具有较强的跳跃性, 而对人类情绪变化的研究指出, 人类的情感强度在很短的时间内波动不会太大, 考虑到视频中情感强度应具有一定的连续性, 我们采用 Kaiser 窗函数 $k(l, \beta)$ 对该曲线进行平滑, 图 3 为平滑后的情感强度曲线, 从图中可以看出, 经过平滑后情感曲线平滑了很多。

图 4 给出对上述数据中一个视频片段的情感强度曲线, 图中同时对情感强度发生较大变化的地方进行了注释, 在 A 点箭头所指示位置, 情感强度曲线具有较大的幅值, 具有较强的情感强度; 在 B 点箭头所指示位置, 情感强度曲线幅值小, 情感强度曲线出现较大波动的时间, 通常对应于视频中异常事件的发生。

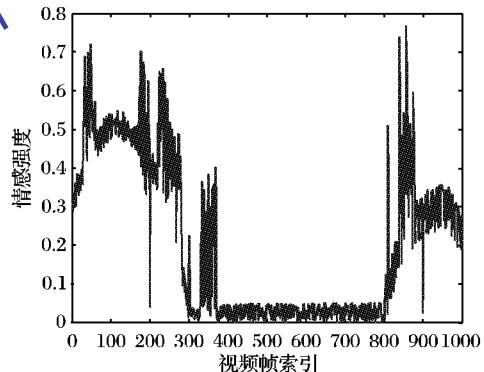


图 2 原始情感强度曲线

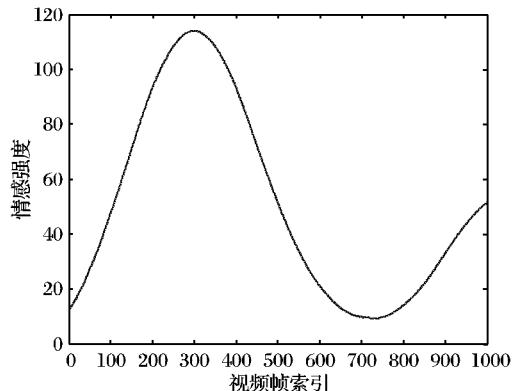


图 3 加窗后情感强度曲线

对视频片段首先进行镜头分割, 然后分别提取关键帧特征和时序特征, 图 5 为情感类型曲线的实验结果, 类型曲线可直观地显示出视频情感的变化特性。为了获取最终的视频类型评价, 将两条曲线对应的时间点进行合成, 将相同类型的情感强度进行相加处理, 归一化后的结果如图 6 所示。由图 6

可以看出,柱状图中具有最大幅值的分量为“Fear”情感类型,因此对该段视频最终的评价结果为“惊恐”,该评价结果良好反映了该段视频的真实情感,验证了本文方法的正确性。

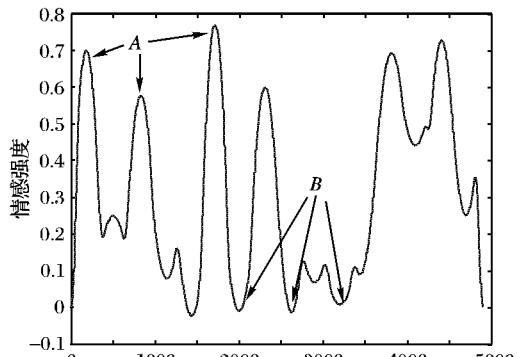


图4 情感强度曲线实例

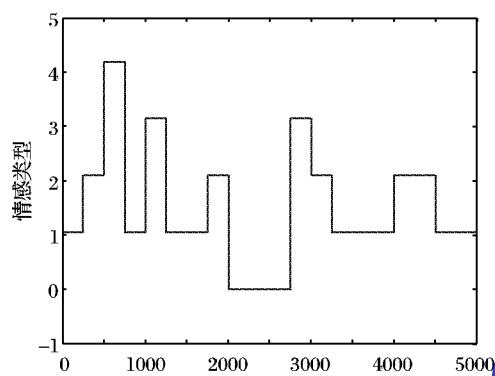


图5 情感类型曲线

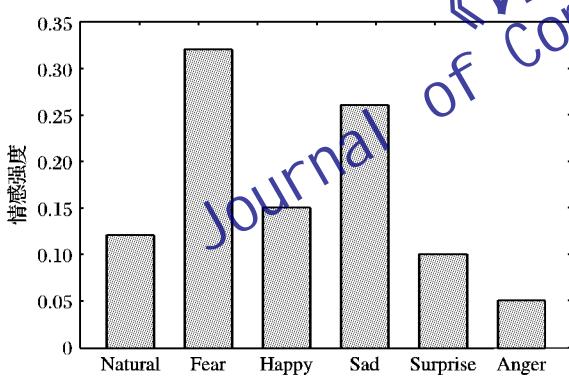


图6 情感类型评价

(上接第1703页)

4 结语

该图像采集系统进行了实际测试。采用串行 LVDS 传输时,可以以 150 fps 的速度连续采集 128×130 的图像。采用 HOTLink 传输时,可以以 500 fps 的速度采集 128×130 大小的图像,也可以 100 fps 的速度传输 320×250 的图像。图像采集系统工作稳定,可靠最大传输速率达到 320 Mbps。测试结果表明 Camera Link 是实现高速串行图像传输的便捷途径。同时利用 Camera Link 中的异步串行通信端口提供了上位机软件控制外部硬件的机制,大大提高了图像采集系统硬件的可配置性能。目前该图像采集系统已成功应用于某型号军工产品的测试设备中。

5 结语

本文提出一种视频情感语义的分析方法,即情感类型-强度分析方法。该方法将视频中的情感语义分解为连续型的情感强度和离散型的情感类型两个相互独立的分量,通过视频的多模态视觉、听觉的多个低层物理特征分别建立两个分量的时序曲线,最终通过情感类型和情感强度曲线的合成,实现对视频情感内容的自动标注。实验结果验证了本文所建立的情感模型能够有效地描述视频的情感内容,更好地符合了人类的认知过程。

参考文献:

- [1] 王上飞,王煦法.图像情感检索研究的进展与展望[J].电路与系统学报,2005,10(4):102-110.
- [2] CHENNOUKH S, GERRITS A, MIET G, et al. Speech enhancement via frequency bandwidth extension using spectral frequency [C]// Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Washington, DC: IEEE, 2001:665-668.
- [3] ESULI A, SEBASTIANI F. Determining the semantic orientation of terms through gloss classification [C]// Proceedings of ACM SIGIR Conference on Information and Knowledge Management. New York: ACM Press, 2005: 617-624.
- [4] COLOMBO C, DEL B A, PALA R. Semantics in visual information retrieval [J]. IEEE Multimedia, 1999, 6(3): 38-53.
- [5] YOO H-W, CHOI S-B. Video scene retrieval with interactive genetic algorithm [J]. Multimedia Tools and Applications, 2007, 34(3): 317-336.
- [6] WANG J, CHNG E, XU C, et al. Identify sports video shots with "happy" or "sad" emotions [C]// Proceedings of IEEE International Conference on Multimedia and Expo. Washington, DC: IEEE, 2006: 877-890.
- [7] 林奕琳,韦岗,杨康才.语音情感识别的研究进展[J].电路与系统学报,2007,12(1):90-98.
- [8] SIMONS R, DETENBER B H, ROEDEMA T M, et al. Emotion processing in three systems: The medium and the message [J]. Psychophysiology, 1999, 36(5):619-627.
- [9] MURRAY I R, ARNOTT J L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion [J]. Journal of the Acoustic Society of America, 1993, 93(2): 1097-1108.
- [10] SIMONS R, DETENBER B H, ROEDEMA T M, et al. Attention to television: Alpha power and its relationship to image motion and emotion [J]. Media Psychology, 2003, 5(3):283-301.

参考文献:

- [1] 李宁,汪骏发.基于 Camera Link 的高速数据采集系统[J].红外,2005(7):31-37.
- [2] Specifications of the Camera Link interface standard for digital cameras and frame grabbers [EB/OL].[2009-10-10]. <http://www.vision1.com/pdf/CameraLink5.pdf>.
- [3] 李志勇,袁魏华,杨镇华.基于 TMS320C6711 的 Camera Link 相机控制的实现[J].电子器件,2006(3):972-975.
- [4] National Instruments Corporation. NI PXI/PCI-1428 user manual [M]. Austin, Texas, USA: National Instruments Corporation, 2007.
- [5] Cypress Semiconductor Corporation. HOTLink transmitter/receiver (CY7B923/CY7B933) [M]. San Jose: Cypress Semiconductor Corporation, 1999.