基于 SVM 和 ICA 的视频帧字幕自动定位与提取

刘骏伟[1] 吴 飞[1] 庄越挺[1]

[1](浙江大学人工智能研究所,杭州 310027) (浙江大学医学院附属邵逸夫医院,杭州 310016)

摘 要 视频字幕蕴涵了丰富语义,可以用来对相应视频流进行高级语义标注,但由于先前视频字幕提取考虑的只是如何尽可能定义好字幕特征,而忽视了分类学习机自身的学习推广能力.针对这一局限性,提出了一种基于支持向量机和独立分量分析的视频帧字幕定位与提取算法.该算法是首先将原始图象帧分割成 $N \times N$ 大小子块,同时将每个子块标注为字幕块和非字幕块两类;然后从每个子块提取能够保持相互高阶独立的独立分量特征去训练支持向量机分类器;最后结合金字塔模型和去噪方法,用训练好的支持向量机来实现对视频字幕区域自动定位提取.由于支持向量机能够在样本不是很多的情况下,具有良好的分类推广能力以及能使独立成分特征之间彼此保持高阶独立性,与其他视频帧字幕定位提取算法比较的结果表明,该算法具有明显的优点.

关键词 模式识别(520·2040) 字幕定位 支持向量机 独立分量分析 金字塔模型

中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2003)11-1334-07

Automatic Caption Location and Extraction in Digital Video Frame Based on SVM and ICA

LIU Jun-wei[1], GUO Zhen-jiang[2], WU Fei[1], ZHUANG Yue-ting[1]

[1](*Institute of Artificial Intelligence, Zhejiang University, Hangzhou* 310027)

[2](*Sir Run Run Shaw Hospital affiliated to Zhejiang University, Hangzhou* 310016)

Abstract Video caption could be used to index video stream with high-level semantics since it implied lots of semantics inherently. The prior work of caption location and extraction considers how to define good caption features and neglects the self-generalization of classifier machine thereof. In order to overcome this limitation, an algorithm firstly localization and extraction video caption using support vector machine (SVM) and independent component analysis (ICA) is presented. In this algorithm, the raw video frame is segmented into N * N sub-blocks, and each block is identified either a caption block or a non-caption block; then mutually high-order independent ICA features are used to train a support vector machine classifier; finally the location and extraction of video caption can be finished automatically with pyramid model and de-noising techniques by each trained support vector machine classifier. Because support vector machine holds excellent generalization of classification with non-enough samples and independent component features are naturally high order independent each other, compared to other algorithms, the experiment data shows this method works well.

Keywords Caption location, Support vector machine, Independent component analysis, Pyramid model

0 引 言

由于视频流所包含的字幕表达了丰富语义,因此可以在原始视频流的分析理解过程中发挥有效作用,例如,在视频新闻报道中,字幕一般都概括叙述了所报道新闻发生的时间、地点、人物和主要事件等重要信息.由于视频字幕为相应的视频流提供了高度概括的语义,其可以在自动定位、提取和识别后对相应视频流进行分割标注,从而可方便基于语义视

基金项目:国家自然科学基金资助项目(60272031);教育部博士点科研基金项目(20010335049);
国家"十五"重大科技攻关项目(2001BA101A07-03);浙江省科技计划项目重点科研项目(2003C21010)
收稿日期:2002-06-17;改回日期:2003-06-13

频流的浏览和检索.

近年来，已经有许多学者提出了视频和图象字幕的动定位和提取算法：如 Lienhart 在文献[1]中，首先假设视频字幕自身基本上是同色的；然后利用视频字幕所固有的边界纹理特征，通过图象分割算法来得到候选的视频字幕区域；接着通过对视频相邻帧中的候选字幕区域之间运动信息的分析、字幕的横宽比，以及所占图象比例等统计信息来做出最后的判断. 在文献[2]中，Zhong 利用了字幕图象所特有的水平方向和垂直方向图象亮度的规律性变化在离散余弦变换(DCT)压缩域直接实现了 JPEG 图象(MPEG 的 I、P、B 帧图象)字幕的定位. Li 将视频字幕定位看作是一个二类模式分类问题，首先将图象分割为图象子块，然后对每一个子块通过训练好的神经网络分类器来判断它是否为字幕块，最后通过后期处理来完成字幕的定位工作[3]. 同时，文献[3]中也研究了针对视频流的视频字幕跟踪技术. 文献[4]则针对每一个图象子块将提取的关键像素值作为特征，使用支持向量机(SVM)作为字幕和非字幕的二类分类器，并结合金字塔模型和后期处理来进行视频字幕定位和提取.

视频流字幕可以分为如下两类：第 1 类是标注字幕，这种字幕是通过后期制作合成到视频流中去的，其包含了对当前视频流内容的语义描述；第 2 类可以称为场景字幕，这种字幕是录制中的环境和物体本身所携带的文字，例如路牌上的路名、服装上面的文字和产品上的商标等. 尽管有些场景字幕也蕴涵了语义信息，但是由于场景字幕出现具有偶然性，而且不同场景字幕之间的差异较大，相比较而言，其重要性就不如第 1 类字幕那么重要了，所以本文处理的视频字幕也主要针对第 1 类标注字幕(图 1).

图 1　待检测视频标注字幕示例

对于视频中的标注字幕而言，由于一般其在视频流中的前后帧之间是不会发生运动的，而只有出现和消失的可能，因此，对于视频流中这些标注字幕的定位和提取可以不必考虑其运动特征，即可在视频流的每一个单独图象帧中完成视频字幕的定位工作. 这样结合视频流的镜头切分和关键帧提取算法，在所提取的视频流关键帧中应用字幕定位算法，就完全可以达到视频内容标注的目的.

同文献[4]相似，视频字幕定位提取就是用提取出的图象子块特征来构造一个学习机，使之实现两类模式分类，即在视频中对字幕块与非字幕块进行分类. 该模式分类中有如下 3 点决定性因素：(1)从样本中所提取的特征向量之间要尽可能相互独立，如果提取的特征之间存在太多相关性信息，则会造成识别分类性能稳定性差，并将极大损坏分类识别结果[5]；(2)要考虑训练样本的多少以及规范性，由于样本过多会造成过学习问题，样本过少又会难以取得好的识别效果，同时由于在将图象分割为图象子块进行字幕识别的过程中，视频字幕中的每个文字本身不再保持完整性，而且对于类似中文的许多文字而言，文字自身结构也是千变万化的，因此很难保证训练样本的规范性和紧致性，而训练样本也只能包括较为有限的样本图象字幕块；(3)考虑基于这些有限样本的特征，如何能够找到一种分类机制，使之对测试样本和实际未知数据都能达到良好的分类目的，即具有好的学习推广能力. 文献[4]中使用 SVM 作为模式分类器，虽可以较好地解决第 3 个因素，但是缺乏对前两个方面的考虑. 在此基础上，本文提出了一种结合使用独立分量分析(Independent Component Analysis，ICA)和支持向量机(Support Vector Machine，SVM)的方法来实现在小样本情况下，提取尽量相互独立的视频字幕特征的方法，以便使 SVM 对提取视频字幕具有良好的识别分类推广能力.

基于独立分量分析和支持向量机的视频字幕定位提取算法如下：首先将训练样本中每幅视频图象按照大小切分成若干图象子块，然后把每个图象子块分别标注为字幕和非字幕两类，并按照 ICA 方法提取出互相独立的图象子块特征向量，接着用这些特征来训练识别视频字幕与非视频字幕的 SVM 分

类器. 在识别时,同样首先使用 ICA 方法对预处理后的测试视频子块进行无关独立特征提取,然后按照金字塔模型对提取的特征进行 SVM 分类识别,在经过去噪等后期处理后,即可得到定位结果. 这样把 ICA 和 SVM 结合起来就把上述模式识别过程中的 3 个因素都考虑到了,因为用 ICA 方法提取的相互独立特征向量能够让 SVM 在适宜样本数目下具有良好的分类学习推广能力,在后文中对此会有详细的叙述.

1 支持向量机原理

支持向量机起源于统计学习理论,它研究如何构造学习机来实现模式分类问题[6,7]. 支持向量机遵循结构风险最小化(Structural Risk Minimization, SRM 准则)原理来构造决策超平面,以便使每一类数据之间的分类间隔(Margin)最大. SVM 的思想就是在样本数目适宜的前提下,选取比较好的 VC 维,使实际风险变小. 这样,SVM 就能够在样本数目适宜的前提下,取得实际最好分类效果. 如今 SVM 已经在视频字幕提取中得到了应用,基于 SVM 视频字幕定位和提取的具体细节可以参考文献[4].

2 视频字幕独立分量特征提取

独立分量分析起源于盲源分离问题(Blind Source Separation, BSS),它与主分量分析(Principle Component Analysis,PCA)和奇异值分解(Singular Value Decomposition,SVD)均属于线性变换技术,但是后两者只能按能量大小对数据进行分解,消除数据之间的二阶相关性,而 ICA 能够消除输入数据的高阶相关性[8]. 在图象、视频和声音识别分类等应用中,不仅可以提取的特征很多,而且特征之间存在相关性,并且特征重要特性一般隐藏在高阶统计特性中,由于使用 ICA 方法不仅能够约减特征维数,而且能使特征保持高阶相互独立,因此它比只是消除 2 阶相关性的 PCA 和 SVD 方法更能提高识别正确率.

假设视频字幕图象块训练样本数为 L,每个训练样本像素点数目为 $N=n_1\times n_2$(表示每个样本为 n_1 行 n_2 列),则训练库 $\boldsymbol{X}$ 的 ICA 重构式为

$$\boldsymbol{y}=\boldsymbol{W}\boldsymbol{X} \tag{1}$$

其中,$\boldsymbol{W}$ 为分解方阵,可通过 ICA 算法得到. Bartlett 使用了如下两种不同的方法,并通过提取人脸 ICA 特征来进行人脸识别[9],从而取得了比 PCA 更高的识别正确率:第 1 种方法是首先通过 ICA 算法来得到相互独立的训练样本的人脸图象基,然后将提取的具有 ICA 特征的人脸图象投影到图象基上,其所得到的系数就作为这幅图象的 ICA 特征;第 2 种方法是首先对图象块的像素进行维数约减,然后对维数约减过的训练样本,并应用 ICA 算法来得到待提取的 ICA 特征. 与 Bartlett 方法类似,本文使用下面介绍的两种方法来提取图象视频字幕的 ICA 特征.

2.1 独立视频字幕基 ICA 特征

在这种方法中,每一个 $n_1\times n_2$ 的图象块均表示为一个行向量(记其维数为 N,$N=n_1\times n_2$),而数目为 L 的所有样本则组成形式为 $L\times N$ 的矩阵 $\boldsymbol{x}$(其中必须保证 $L\geq N$).

提取独立视频字幕基 ICA 特征的具体步骤如下:

(1) 计算 $\boldsymbol{x}^{\mathrm{T}}$ 的协方差矩阵 $\boldsymbol{C}$ 的特征向量和特征根,并将特征值按从大到小进行排序,然后选取前面 m 个特征值所对应的特征向量 $\boldsymbol{p}_i(i=1,\cdots,m,\boldsymbol{p}_i$ 是 $N\times 1$ 列向量)组成 N 行 m 列矩阵 $\boldsymbol{P}_m$,这一步也就是标准的 PCA 算法.

$$\boldsymbol{P}_m=[\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_m] \tag{2}$$

(2) 由于 $\boldsymbol{P}_m$ 含有 m 个与最大特征值对应的特征向量,也就含有原训练样本矩阵 $\boldsymbol{x}$ 最可能多的能量,因此用 $\boldsymbol{P}_m$ 的转置矩阵 $\boldsymbol{P}_m^{\mathrm{T}}$ 代替重构式(1)中的 $\boldsymbol{x}$,应用快速定点 ICA 算法[10],可以得到

$$\boldsymbol{y}=\boldsymbol{W}\boldsymbol{P}_m^{\mathrm{T}}\Rightarrow\boldsymbol{P}_m^{\mathrm{T}}=\boldsymbol{W}^{-1}\boldsymbol{y} \tag{3}$$

其中,$\boldsymbol{y}$ 的每一行代表一个独立视频字幕基,$m\times m$ 矩阵 $\boldsymbol{W}$ 可在训练中得到;

(3) 对于每个训练库样本可以用特征向量基坐标表示,即 $\boldsymbol{R}_m=\boldsymbol{x}\boldsymbol{P}_m$,$\boldsymbol{R}_m$ 是 $L\times m$ 矩阵,其第 1 行表示第 1 个样本对于 m 个特征向量基的坐标,最后一行表示第 L 个样本对于 m 个特征向量基的坐标,可使用最小平方误差法求 $\boldsymbol{x}$ 的逼近值 $\boldsymbol{x}_{\mathrm{mse}}$,并且将式(2)代入,得

$$\boldsymbol{x}_{\mathrm{mse}}=\boldsymbol{R}_m\boldsymbol{P}_m^{\mathrm{T}}=\boldsymbol{x}\boldsymbol{P}_m\boldsymbol{P}_m^{\mathrm{T}}=\boldsymbol{x}\boldsymbol{P}_m\boldsymbol{W}^{-1}\boldsymbol{y} \tag{4}$$

(4) 从式(4)可以看出,$\boldsymbol{x}\boldsymbol{P}_m\boldsymbol{W}^{-1}$中的第 i 行是第 i 个训练样本相对于 $\boldsymbol{y}$ 中独立视频字幕基的线性组合系数,由于这 m 个组合系数就是第 i 个训练样本的 ICA 特征,因此,对于任意测试样本 $\boldsymbol{I}_{1\times N}$,其独立视频字幕基 ICA 特征就是

$$c = IP_m W^{-1} \tag{5}$$

独立字幕基方法相当于认为字幕块是由相互独立的图象基的线性组合构成的，由于本质上来讲，图象字幕块是由具有水平或垂直分布的笔划或者笔划段构成的，而独立字幕基就相当于这些笔划段，是构成字幕的要素，因此独立字幕基方法在字幕块分类中是可行的.

2.2　独立系数 ICA 特征

用于独立成分分析的 ICA 特征方法和上一种方法不同，因为在这种方法中，每一样本图象块代表 x 中的一列. ICA 特征提取步骤如下：

(1) 对样本库中的图象块的维数进行约减，使总像素点 N 约减为待提取的 ICA 特征维数 m. x_m 表示进行了维数约减的图象训练样本，每一个样本图象块表示为 x_m 中的一列，x_m 为 $m \times L$ 的矩阵. x_m 可通过 P_m 来得到(P_m 定义与上面相同)，则

$$x_m = P_m^{\mathrm{T}} x \tag{6}$$

(2)应用 ICA 算法来提取 x_m 独立特征，其重构式为

$$y = W x_m = W P_m^{\mathrm{T}} x \tag{7}$$

(3) 对于任意测试图象块 I，I 为 $N \times 1$ 列向量，其独立系数 ICA 特征表示为

$$c = W P_m^{\mathrm{T}} I \tag{8}$$

其中，W 可由 ICA 算法计算得出. 如果令

$$U = W P_m^{\mathrm{T}} \tag{9}$$

则 U 中每一列表示的是与训练库所对应的每个图象的独立系数字幕基.

与独立字幕块算法不同，在独立系数算法中，字幕基之间不是独立的，即字幕基只相当于在样本中提取出的携带有较大相似特征的代表性图象块. 但是由于字幕块的不确定性和弱相似性，使其与人脸检测中人脸的高度相似性不同，因此这种方法应用在视频字幕检测时不会像在人脸检测应用中取得那么好的结果.

在实验中，取视频图象块大小为 11×11，即 $N = 121$，如果 m 取为 36，则每个样本提取的 ICA 特征维数大小为 36，并且从 ICA 特性知道，这些特征高阶统计相互独立.

3　视频字幕定位与提取

3.1　视频字幕 ICA 特征

对视频图象进行预处理，即首先将每幅图象分割为 $N \times N$ 子块(实验中，取 $N = 11$)，同时将每一图象子块标注为字幕块(+1)或者非字幕块(−1)两类；然后按照上面介绍的两种方法分别提取每个视频图象子块的 m 个 ICA 图象特征.

3.2　SVM 核函数的选择

SVM 分类学习机的结构包含输入层、隐含层和输出层 3 层，其中，输入层用于输入数据，在本实验中，输入层接受每个图象子块的 m 个 ICA 特征和对图象子块进行是否为字幕的标注；隐含层有如下两个功能：一是用非线性映射 Φ 把输入 ICA 特征从原始低维空间映射到高维特征空间，即使原始 ICA 特征成为高维特征，以达到高维可分目的；二是计算高维空间特征和支持向量的内积. 在实际处理中，这两步是通过核函数 K 一步来实现的，核函数满足

$$K(x, y) = (\Phi(x) \cdot \Phi(y)) \tag{10}$$

其中，K 是核函数，Φ 是高维非线性映射，$\cdot$ 是内积. 输出层将分类结果输出.

SVM 中研究最多的核函数主要有多项式、径向基函数(RBF)、多层 Sigmoidal 神经网络 3 类. 这里使用的是 RBF 核函数，其形式为

$$K(x, y) = \exp\left\{-\frac{|x - y|^2}{\sigma^2}\right\} \tag{11}$$

在实验中，由于是将字幕块定义为+1，非字幕块定义为−1，因此对每一个输入，如果输出为正，则该输入块被判定为字幕块，如果输出为负，则为非字幕块.

3.3　金字塔模型(Pyramid Model)

由于视频字幕大小经常变化，而且经常变化巨大，因此相同大小的子块可能只包含了某个字幕中的一个笔划，而在别的情况下确包含了多个小字幕. 为了解决这一问题，采取了金字塔模型[3]. 所谓 p 阶金字塔模型(p-step pyramid model)是指对原始图象分辨率依次进行 p 次缩小，例如 3 阶金字塔模型，总共对图象缩小 3 次，且在每一阶都将原图象长宽减少为原来的 $1/\sqrt{2}$，然后在每一阶都单独使用 SVM 进行字幕检测，最后通过合成，即通过将各阶的检测结果都还原到原始图象分辨率下合成来生成最终的检测结果.

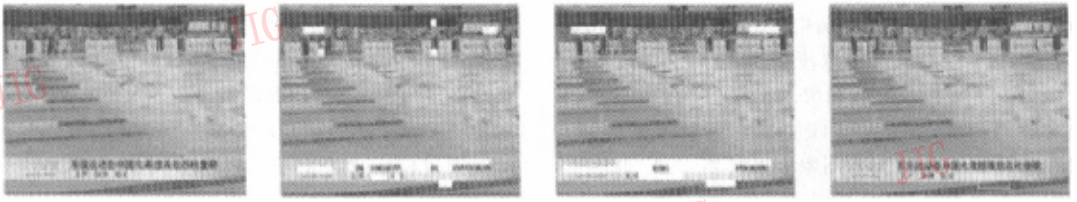
3.4　后期处理(Post-processing)

对每个图象子块都做出分类判断后，还要进行后期处理，其目的是为了抑制噪声和合并字幕区域[4]. 由于背景图象的复杂性，所以部分背景图象也会因体现出字幕块特性而被错判为字幕块，另外，字

形学知识也表明,字幕一般沿水平方向聚集在一起,因此应用这一性质,就可以通过"扩张-收缩"算法消除绝大多数孤立噪声块.其具体算法流程如下:

(1) 对每一图象子块做出判定之后,就可得到所有候选字幕块的集合;接着使用如下方法去构造每个候选字幕块(i,j)的扩充块$(\hat{i},\hat{j})$,即将(i,j)扩充为$(\hat{i},\hat{j})$,其中$(\hat{i},\hat{j})$包括(i,j)及与它相邻的两个子块$(i-1,j)$和$(i+1,j)$.如果某个候选字幕块(i,j)的扩充块$(\hat{i},\hat{j})$与任何一个其他候选字幕块的扩充块是连通的,则判断(i,j)为字幕块,否则为噪音块,并应从候选字幕块集合中去除(i,j);

(2) 去除噪声块后,就已经从水平方向上将所有真正字幕块连接起来,最后要做的工作就是确定出每一个联通字幕块集合的最大包围矩形,其中位于包围矩形中的区域就是最后确定的字幕区域(图 2).



(a) 原始图象 (b) 字幕检测结果 (c) 去除噪声和字幕合并 (d) 最终字幕区域

图 2 后期处理示意图

3.5 视频字幕识别

经过上述的步骤,即可得到视频图象中的字幕区域,若进一步应用 OCR 技术,就可以实现字幕识别.由于视频字幕自动定位可以避免对整幅图象进行 OCR 识别,从而提高了识别效率.

4 实验结果与分析

4.1 多种字幕子块特征提取方法识别结果比较

实验所用图象是从中央电视台选取的 1 000 帧不同的视频节目图象(主要为体育新闻,包括主持人画面,比赛画面,广告等),实验时,选取其中的 400 帧作为训练样本,其余为测试样本,在 Windows 2000 和 Matlab5.3 下进行仿真.

为了比较 SVM 对不同视频字幕特征的识别性能的差异,实验中提取了 5 种不同特征(见表 1).

表 1 视频字幕不同特征表示

特征类型	特征表示	维数	备注
原始灰度特征	Original Pixel	121	
关键灰度特征	Selected Pixel	43	按照文献[4]"米"字形结构提取
PCA 特征	PCA	36	按照文献[11]方法提取
ICA 独立字幕特征	ICA1	36	
ICA 独立系数特征	ICA2	36	

原始灰度特征指每个 11×11 像素大小子块的灰度值;关键灰度特征指对 11×11 像素大小子块按照"米"字形对角线结构提取的 41 个(即 $4\times N-3$)灰度值;ICA 分成两步进行识别:先对特征数据去均值和白化处理(由于其可以看成是普通的 PCA 算法,所以在某种程度上 PCA 是 ICA 的一部分);然后通过使用信息熵这个目标函数来训练矩阵 $\boldsymbol{W}$.

根据背景和字幕特性的不同,将 600 帧测试样本分为 3 类进行测试.为了考查各种方法的适用性,其中 200 帧选取的是与测试样本相差较大的字幕图象(如表 2 所示第 3 类):

表 2 3 类测试样本集

	第 1 类测试样本	第 2 类测试样本	第 3 类测试样本
字幕特性	字幕较少,相对集中	字幕一般,相对集中	字幕较多,分布广
背景特性	一般复杂,有一定干扰	很复杂,干扰较大	一般复杂,有一定干扰
与训练样本相似性	相似	相似	不相似

在对上述 5 种特征的识别效果进行对比的时候,为了体现出各种方法自身所能达到的识别准确率和可能产生的误判率,做出如下约定:(1)对所有特征,SVM 算法保证一致;(2)不使用金字塔模型,只在原始图象上进行字幕定位和识别;(3)在 SVM 输出结果后,即进行结果统计,不采取后期处理措施.上述 3 点约定就保证了能够在最公平的情况下评判各种特征的绝对效果.表 3 和表 4 分别给出了对于 3 类不同的测试样本,使用 5 种特征进行 SVM

字幕定位识别的准确率和误判率. 表 5 则是使用 5 种特征进行字幕识别定位总体上的准确率和误判率(对于第 1 类和第 2 类测试样本,由于自身所包含的字幕较少,所以误判比率就较高).

表 3　3 类测试样本在不同特征表示下的字幕识别正确率(%)对比

测试样本	采用特征				
	原始灰度特征	关键灰度特征	PCA 特征	ICA 独立字幕特征	ICA 独立系数特征
测试样本 1	85.27	90.2	88.24	86.27	80.4
测试样本 2	88	92	90	94	86
测试样本 3	71.56	83.97	78.66	85.54	69.71

表 4　3 类测试样本在不同特征表示下的字幕识别错误率(%)对比

测试样本	采用特征				
	原始灰度特征	关键灰度特征	PCA 特征	ICA 独立字幕特征	ICA 独立系数特征
测试样本 1	48.02	82.86	56.56	41.38	25.6
测试样本 2	21	44	33	15	12
测试样本 3	1.57	17.05	11.32	1.83	1.52

表 5　3 类测试样本在不同特征表示下的字幕识别总体性能对比

测试样本	采用特征				
	原始灰度特征	关键灰度特征	PCA 特征	ICA 独立字幕特征	ICA 独立系数特征
识别正确率(%)	81.78	88.69	84.96	88.8	79.04
识别错误率(%)	23.9	47.8	33.91	19.34	12.34

从上面的对比中可以看出:PCA 和 ICA 特征比图象灰度值特征有较优的识别性能;独立字幕基的 ICA 特征能够取得最好的识别准确率和较低的误判率;正如前文分析结果一样,独立系数 ICA 方法在识别准确率上较差.

总体上来说,在相同维数下,独立字幕基 ICA 特征比 PCA 特征取得了更好的识别效果,由于 SVM 通过核函数把低维特征向高维空间映射,而 ICA 特征又可在高阶统计上保持独立,从而保证了特征所携带信息的独立性.

在 SVM 的生成过程中,尽管没有用与第 3 类测试样本相似的字幕进行训练,但是从表 5 中可以发现,SVM 对第 3 类测试样本仍然取得了令人满意的识别结果,另外,SVM 具有很强的学习推广能力,即使训练样本较小,也可以取得较高的识别效果.

如果运用三层金字塔模型,对识别结果进行去噪和合并等后期处理,那么在使用独立字幕基 ICA 特征情况下,SVM/ICA 方法可以取得 96.3%的正确识别率和 18%的误判率,而文献[2]在大样本情况下的识别正确率为 99%,误判率为 36%,说明 SVM/ICA 在样本不是很多的情况下就可以取得较高识别率,而引起漏判的原因主要在于图象中字幕太小或字幕很少,且自身形成孤立块(如比赛分数),导致被误认为是噪声块而被清除,另外是由于一部分图象子块体现出了字幕块特性以及受到了场景字幕的干扰所致.

4.2　汉字字幕和非汉字字幕识别讨论

从语种上来分,中文字幕和英文字幕是最具代表性的两类视频字幕,它们具有截然不同的字形学特点. 英文字幕都是由 26 个英文字母的大小写形式以及若干其他辅助符号排列组合而成,相对而言,作为文字基本单元的字母,其字形变化不多,而且在组成单词时可保持各自的独立,拉丁语系的文字多和英语类似;而中文汉字的最小字形单元是笔划,与英文字母不同,其笔划在构成文字时不再相互独立,由于它是通过互相的重叠交错来组合成汉字,因此相对英文而言,汉字字形变化更加复杂,亚洲语系的文字大多具有类似中文的特点.

上述的差别使得必须考虑针对汉字和非汉字的文字字幕定位的 SVM/ICA 方法推广能力和适用性. 在 SVM/ICA 方法中,其实并没有考虑字母、字和词等文字形态学概念,字幕和非字幕的判断是分两步完成的,即首先基于分割出的图象子块做出是否是字幕块的判断,其所依据的只是图象子块是否呈现出字幕子块所特有的特征. 大家知道,无论是何种文字,均是由横,竖,斜线,圆弧等笔划组成,因此字幕子块具有相同的"微观"共性;其次在后期处理中,进行噪声过滤和字幕块合并步骤所依据的是视频字幕中文字的排列特征. 大家知道,在文字的排列上,各种语言的文字均体现出水平排列的特性,并且保持在一定区域的聚集性特征,从这个方面说,字幕子块在"宏观"上也具有相同的共性. 因此可以得出结论,SVM/ICA 方法对汉字字幕和非汉字字幕同样适用,实验结果也证实了这一结论(图 3).

5　结论和今后工作

由于先前视频字幕识别定位算法没有统一考虑字幕特征相关性、训练样本的有限性和分类学习机

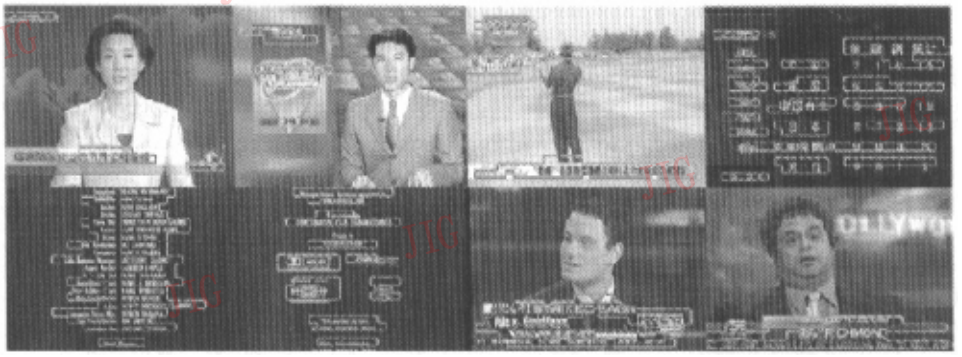


图 3 汉字字幕和非汉字字幕识别结果示例

自身学习推广能力 3 个因素,因此,本文提出使用 SVM/ICA 方法来进行视频字幕自动定位提取,并在样本不是很多的情况下,使用高阶相互独立的特征实现了较高精度的视频字幕定位提取.

今后的工作集中在如下两个方面:(1)虽然视频字幕含有极大语义,可以用来对视频语义进行标注,但是,并不是所有的视频图象都需要提取字幕.为了判别哪些视频图象需要提取字幕,哪些不需要提取,要充分利用多媒体视频流中的视觉、运动和听觉等信息,首先对视频流进行分类,然后只对新闻报道、体育、广告和电视对白等视频图象提取字幕,面对其他视频节目则不需要提取字幕;(2)研究在视频流中对视频字幕进行跟踪的技术.

参 考 文 献

1 Lienhart R, Stuber F. Automatic Text Recognition in Digital Videos[A]. In: Proceeding of Image and Video Processing[C], San Jose, CA, USA: SPIE Press, 1996: 2666~20.

2 Zhong Yu, Zhang Hongjiang, Jain Anil K. Automatic caption localization in compressed video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22 (4): 385~392.

3 Li H, Doermann D, Kia O. Automatic text detection and tracking in digital video [J]. IEEE Transactions Image processing, 2000, 9(1): 147~156.

4 庄越挺,刘骏伟,吴飞等. 基于支持向量机的视频字幕自动定位与提取[J]. 计算机辅助设计与图形学学报, 2002, 14(8): 750~753.

5 Proriol J. Selection of variables for neural network analysis[J]. Journal of Nuclear Instrument and Methods, 1996, 361: 581~585.

6 Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer Verlag, 1995.

7 Burges C. A tutorial on support vector machines for pattern recognition[J]. Knowledge Discovery and Data Mining, 1998, 2(2): 121~167.

8 Hyvarinen A. Survey on independent component analysis [J]. Neural Computing Surveys, 1999, 2: 94~128.

9 Bartlett M S, Lades H M, Sejnowski T J. Independent component representations for face recognition [A]. In: Proceedings SPIE Conference on Human Vision and Electronic Imaging III[C], San Jose, CA, USA: 1998: 528~539.

10 Aapo Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis [J]. IEEE Transactions on Neural Networks, 1999, 10(3): 626~634.

11 Pierre Baldi, Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima[J]. Neural Networks, 1989, 2(1): 53~58.

刘骏伟 1978 年生,2003 年获浙江大学计算机应用硕士学位.主要研究方向为多媒体处理与数字化图书馆.

吴 飞 1973 年生,2002 年获浙江大学计算机应用博士学位,讲师.主要研究方向为多媒体分析、计算机视觉和遥感处理.

庄越挺 1965 年生,1998 年获浙江大学计算机应用博士学位,教授,博士生导师.主要研究领域为多媒体数据库、人工智能、基于内容的图象/视频信息检索、及视频动画等.