DOI: 10.11992/tis.201804030
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.TP.20180928.2216.008.html

视听觉跨模态表面材质检索

刘卓锟[1]，刘华平[2]，黄文美[1]，王博文[1]，孙富春[2]
（1. 河北工业大学 省部共建电工装备可靠性与智能化国家重点实验室，天津 300130；2. 清华大学 智能技术与系统国家重点实验室，北京 100084）

摘　要：针对文本图像特征有时无法满足对物体材质进行真实准确分析的情况，本文在视听领域使用跨模态检索方法进行表面材质检索。首先提取声音的梅尔频率倒谱系数 (MFCC) 特征，使用卷积神经网络 (CNN) 提取图像特征，然后利用典型相关分析将两种特征映射到子空间并用欧氏距离进行检索，并在慕尼黑工业大学触觉纹理数据集上进行实验验证，实现了使用声音检索图像的跨模态检索过程。实验结果表明，所提出的方法在材质检索方面有较好应用效果。

关键词：跨模态检索；特征提取；典型相关分析；子空间映射；材质分析；卷积神经网络；梅尔频率倒谱系数；欧式距离
中图分类号：TP 391　　**文献标志码：**A　　**文章编号：**1673−4785(2019)03−0423−07

中文引用格式：刘卓锟，刘华平，黄文美，等. 视听觉跨模态表面材质检索 [J]. 智能系统学报，2019，14(3)：423–429.
英文引用格式：LIU Zhuokun，LIU Huaping，HUANG Wenmei，et al. Audiovisual cross-modal retrieval for surface material[J]. CAAI transactions on intelligent systems，2019，14(3)：423–429.

Audiovisual cross-modal retrieval for surface material

LIU Zhuokun[1], LIU Huaping[2], HUANG Wenmei[1], WANG Bowen[1], SUN Fuchun[2]
(1. State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300130, China; 2. State Key Lab of Intelligent Technology and Systems, Tsinghua University, Beijing 100084, China)

Abstract: Text and image features sometimes do not allow for true and accurate analysis of the material. To solve this problem, a cross-modal method for surface material retrieval in an audiovisual field is proposed. First, the sound feature is extracted using mel frequency cepstral coefficients (MFCCs), and the image feature is extracted using convolutional neural network (CNN). Then, these two features are mapped to the subspace using canonical correlation analysis and are further retrieved via Euclidean distance. Experimental validation performed using the tactile texture dataset of the Technical University of Munich showed that the proposed method has a good application effect on material retrieval.

Keywords: cross-modal retrieval; feature extraction; canonical correlation analysis; subspace mapping; material analysis; convolutional neural network; Mel-frequency cepstral coefficients; Euclidean distance

面对多媒体信息数据量的激增和模态复杂多样化的挑战，跨模态检索因其可以处理不同模态的数据成为国内外学者研究的重要课题。跨模态检索应用得比较成熟的领域主要为计算机视觉、模式识别、文本图像检索等[1-4]，其研究的重点依然放在图像和文本两种模态之间。但是图像反映的颜色、纹理等信息和文本对物体的描述有时不能带给我们足够的信息量，比如在网购过程中，消费者仅通过浏览购买商品的文字和图片信息，有时不能在大脑完整地构建商品的特征信息，因而会购买到与需求不符的商品；在深海和太空探索领域，由于视频和图像受环境因素影响较大，仅凭摄像机反馈回来的视频和图像不足以让人们确定未知物体的材质信息；在日常生活中，当我们购买家具或西瓜时，仅通过视觉信息并不能准

收稿日期：2018−04−18.　**网络出版日期：**2018−09−30.
基金项目：国家自然科学基金重点项目 (U1613212)；河北省自然科学基金项目 (E2017202035).
通信作者：刘华平. E-mail：hpliu@tsinghua.edu.cn.

确判断家具所用木材质量的好坏或西瓜是否熟透，常常通过敲击其表面产生的声音来辅助判定。

引入声音模态在某些方面可以解决文本和图像信息量不足的问题。目前关于声音的检索技术大多涉及的是与语音和音乐相关的检索技术，其中声音特征采用梅尔频率倒谱系数 (Mel-frequency cepstral coefficients, MFCC)。梅尔频率倒谱系数模仿人耳的感知特性[5]，该方法具有很好的识别性和可靠性，是应用最广泛的声音特征之一。另一方面，图像特征采取卷积神经网络 (convolutional neural network, CNN) 提取。卷积神经网络的出现使得图像识别领域发展迅速，国外已有研究将卷积神经网络应用于跨模态检索的图像特征提取[6]。

1　跨模态检索

不同于相同模态之间的检索，在跨模态检索中，检索结果和查询的模态是不同的。如何在不同模态之间建立相关性成为跨模态检索的关键。目前，应用在跨模态检索中的方法有典型相关分析法[7-9]、偏最小二乘法[10]、耦合字典学习法[11] 等。对比其他方法，典型相关分析 (canonical correlation analysis, CCA) 因其简单高效的特点在跨模态检索领域应用十分广泛，文献 [7] 提出多标签典型相关分析，可以处理多标签信息量大的数据集的情况。文献 [8] 提出多视图典型相关分析方法，利用不同视图的互补和相关信息可以处理多视图数据。文献 [9] 提出核典型相关分析，解决了非线性情况下不同模态间相关性的问题。

然而，传统的典型相关分析在应用时要求两组变量间符合一一配对关系。当两组变量间出现多个对应关系或配对形式为组配对时，上述方法将不再适用。针对上述情况，本文引入聚类典型相关分析方法。首先使用梅尔频率倒谱系数声音特征和卷积神经网络提取的图像特征，然后利用聚类典型相关分析将两种特征映射到子空间并用欧氏距离进行检索，最后在慕尼黑工业大学触觉纹理数据集上进行验证，实验结果表明所述方法适用于材质检索，具体流程如图 1 所示。

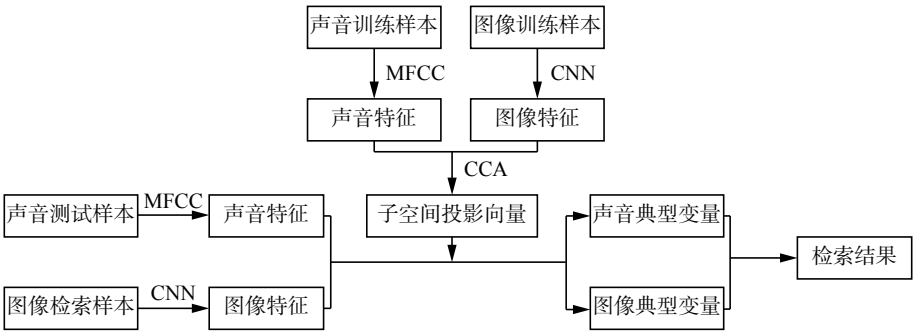


图 1　检索流程

Fig. 1　The retrieval process

2　声音和图像特征提取

本文的声音特征使用梅尔频率倒谱系数特征，图像特征使用卷积神经网络提取得到。

2.1　梅尔频率倒谱系数

梅尔频率倒谱系数是语音处理中最常用的特征之一。文献 [12] 对敲击物体产生的声音提取梅尔频率倒谱系数特征，并应用于声音的分类。本文求得梅尔频率倒谱系数的一阶和二阶差分特征系数，结合标准梅尔频率倒谱系数[13]，最终得到 39 维梅尔频率倒谱系数特征。图 2(a) 、(b) 所示为训练集中敲击竹木和红色羊毛毡的声音时域信号，图 2(c)、(d) 所示为经过上述过程得到的声音特征。

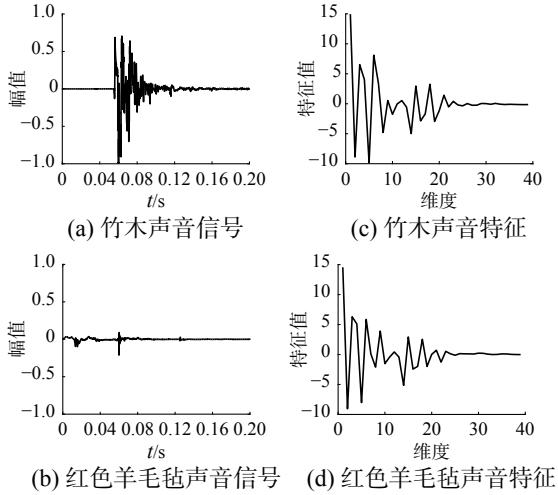
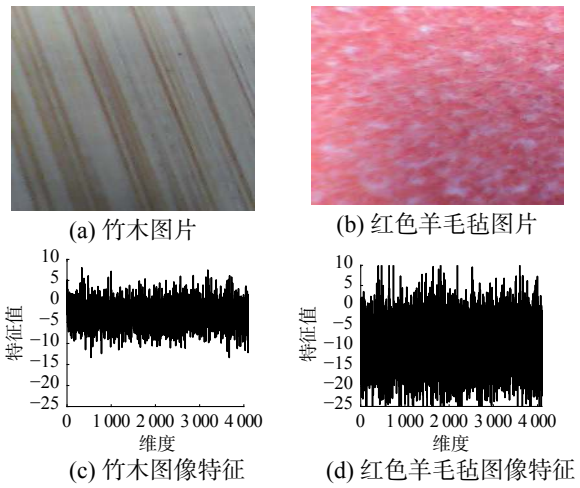


图 2　竹木和红色羊毛毡声音信号和声音特征

Fig. 2　Sound signals and features of bamboo and red fleece

2.2　卷积神经网络

近年来，卷积神经网络已经被广泛地应用于图像的识别检测领域。本文选用的网络为预先训练好的 AlexNet 网络[6]，包含 5 个卷积层和 3 个完全连接层。将图片分辨率调整为 256×256 输入到文献 [6] 所述模型之中，最终得到 4 096 维图像特征。图 3(a)、(b) 所示为训练集中敲击竹木和红色羊毛毡的图片，图 3(c)、(d) 所示为经过上述过程得到的图像特征。



(a) 竹木图片　(b) 红色羊毛毡图片

(c) 竹木图像特征　(d) 红色羊毛毡图像特征

图 3　竹木和红色羊毛毡图片和图像特征

Fig. 3　Image features of bamboo and red fleece

3　典型相关分析

典型相关分析作为一种灵活有效、可扩展能力强的数据分析方法，在跨模态检索领域占据着重要地位。典型相关分析不仅可以最大化两组变量在投影空间的相关性，还能对复杂特征进行降维处理。本文使用这种方法对声音特征和图像特征进行相关性分析处理。

3.1　典型相关分析基本原理

使用典型相关分析对声音特征矩阵 $\boldsymbol{X}=[x_1\ x_2\cdots x_n]$ 和图像特征矩阵 $\boldsymbol{Y}=[y_1\ y_2\cdots y_n]$ 进行处理。将 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 表示为各自特征的线性组合，$\boldsymbol{U}=\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{X}$ 和 $\boldsymbol{V}=\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{Y}$，通过研究 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的关系来代替 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的关系，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的相关系数 ρ 表达式为

$$\rho=\frac{\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{\Sigma}_{XY}\boldsymbol{\omega}_y^{\mathrm{T}}}{\sqrt{(\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\omega}_x)}\sqrt{(\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\omega}_y)}} \tag{1}$$

式中：$\boldsymbol{\omega}_x$ 和 $\boldsymbol{\omega}_y$ 为两组变量对应的投影向量；$\boldsymbol{\Sigma}_{XX}$ 和 $\boldsymbol{\Sigma}_{YY}$ 分别表示特征集 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差矩阵；$\boldsymbol{\Sigma}_{XY}$ 表示 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的互协方差矩阵：

$$\boldsymbol{\Sigma}_{XY}=\frac{1}{n}\sum_{p=1}^{n}\boldsymbol{x}_p\boldsymbol{y}_p^{\mathrm{T}} \tag{2}$$

$$\boldsymbol{\Sigma}_{XX}=\frac{1}{n}\sum_{p=1}^{n}\boldsymbol{x}_p\boldsymbol{x}_p^{\mathrm{T}} \tag{3}$$

$$\boldsymbol{\Sigma}_{YY}=\frac{1}{n}\sum_{p=1}^{n}\boldsymbol{y}_p\boldsymbol{y}_p^{\mathrm{T}} \tag{4}$$

通过构造拉格朗日等式，在约束条件 $\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\omega}_x=1$ 和 $\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\omega}_y=1$ 下，找到合适的投影向量 $\boldsymbol{\omega}_x$ 和 $\boldsymbol{\omega}_y$，使 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的相关性达到最大化：

$$L=\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{\Sigma}_{XY}\boldsymbol{\omega}_y-\frac{\lambda}{2}(\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\omega}_x-1)-\frac{\theta}{2}(\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\omega}_y-1) \tag{5}$$

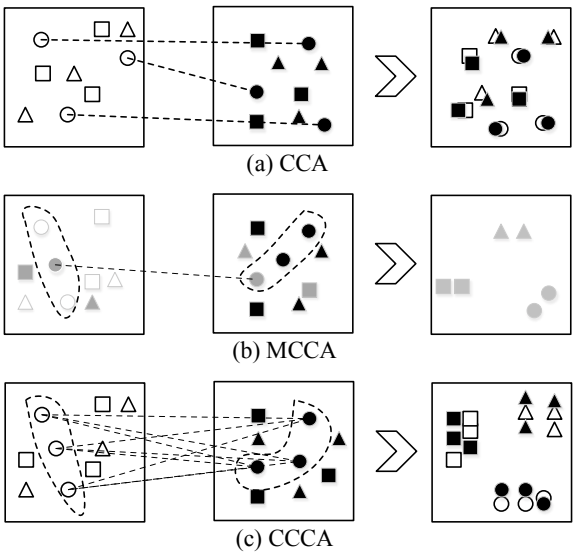
式中：L 为构造的拉格朗日函数；λ 和 θ 为引入的系数变量。

将求解转化为常规的特征值问题，$\boldsymbol{\omega}_x$ 和 $\boldsymbol{\omega}_y$ 可以通过其对应最大特征值的特征向量找到：

$$\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}=\lambda^2\boldsymbol{\omega} \tag{6}$$

3.2　改进的典型相关分析

当样本变量不再是一一对应关系时，雅虎和微软研究院的 Rasiwasia 等[14] 改进典型相关分析，提出均值典型相关分析 (mean canonical correlation analysis, MCCA) 和聚类典型相关分析 (cluster canonical correlation analysis, CCCA)，相应的子空间对应关系如图 4 所示，不同的形状代表不同的种类，相同形状代表同一种类中的不同物体。



(a) CCA

(b) MCCA

(c) CCCA

图 4　3 种方法的子空间对应关系

Fig. 4　The subspace correspondences of the three methods

对于本文使用的声音数据集 $\boldsymbol{X}=[X_1\ X_2\cdots X_C]$ 和图像数据集 $\boldsymbol{Y}=[Y_1\ Y_2\cdots Y_C]$，其中 C 表示数据集的总类别数，$\boldsymbol{X}_c$ 和 $\boldsymbol{Y}_c$ 是属于类别 c 对应的数据 $\boldsymbol{X}$、$\boldsymbol{Y}$ 的子集。

$$\boldsymbol{X}_c=[x_1^c\ x_2^c\cdots x_{|X_c|}^c] \tag{7}$$

$$\boldsymbol{Y}_c=[y_1^c\ y_2^c\cdots y_{|Y_c|}^c] \tag{8}$$

式中：$|X_c|$ 和 $|Y_c|$ 分别为相应第 c 类数据个数。

3.2.1　均值典型相关分析

均值典型相关分析较为简单，首先求得每个子集的平均值，然后求得投影向量来建立子集均值之间的相关关系，最后寻找相关系数最大时的投影向量，即

$$\rho = \frac{\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{M}_{XY} \boldsymbol{\omega}_y^{\mathrm{T}}}{\sqrt{(\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{M}_{XX} \boldsymbol{\omega}_x)}\sqrt{(\boldsymbol{\omega}_y^{\mathrm{T}} \boldsymbol{M}_{YY} \boldsymbol{\omega}_y)}} \tag{9}$$

其中，$\boldsymbol{M}_{XY}$、$\boldsymbol{M}_{XX}$ 和 $\boldsymbol{M}_{YY}$ 定义分别为

$$\boldsymbol{M}_{XY}=\frac{1}{C}\sum_{c=1}^{C}\boldsymbol{\mu}_x^c(\boldsymbol{\mu}_y^c)^{\mathrm{T}} \tag{10}$$

$$\boldsymbol{M}_{XX}=\frac{1}{C}\sum_{c=1}^{C}\boldsymbol{\mu}_x^c(\boldsymbol{\mu}_x^c)^{\mathrm{T}} \tag{11}$$

$$\boldsymbol{M}_{YY}=\frac{1}{C}\sum_{c=1}^{C}\boldsymbol{\mu}_y^c(\boldsymbol{\mu}_y^c)^{\mathrm{T}} \tag{12}$$

式中：$\boldsymbol{\mu}_x^c$、$\boldsymbol{\mu}_y^c$ 为相应的第 c 类子集的平均值，$\boldsymbol{\mu}_x^c=\frac{1}{|\boldsymbol{X}_c|}\sum_{i=1}^{|X_c|}\boldsymbol{x}_i^c$，$\boldsymbol{\mu}_y^c=\frac{1}{|\boldsymbol{Y}_c|}\sum_{j=1}^{|Y_c|}\boldsymbol{y}_j^c$。

3.2.2　聚类典型相关分析

聚类典型相关分析不再建立子集间均值的关系，而是建立子集中每一个数据点和对应子集所有数据点的关系，此时相关系数表达式为

$$\rho = \frac{\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{R}_{XY} \boldsymbol{\omega}_y^{\mathrm{T}}}{\sqrt{(\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{R}_{XX} \boldsymbol{\omega}_x)}\sqrt{(\boldsymbol{\omega}_y^{\mathrm{T}} \boldsymbol{R}_{YY} \boldsymbol{\omega}_y)}} \tag{13}$$

其中，$\boldsymbol{R}_{XY}$、$\boldsymbol{R}_{XX}$ 和 $\boldsymbol{R}_{YY}$ 定义如下：

$$\boldsymbol{R}_{XY}=\frac{1}{T}\sum_{c=1}^{C}\sum_{i=1}^{|X_c|}\sum_{j=1}^{|Y_c|}\boldsymbol{x}_i^c(\boldsymbol{y}_j^c)^{\mathrm{T}} \tag{14}$$

$$\boldsymbol{R}_{XX}=\frac{1}{T}\sum_{c=1}^{C}\sum_{i=1}^{|X_c|}|\boldsymbol{Y}_c|\boldsymbol{x}_i^c(\boldsymbol{x}_i^c)^{\mathrm{T}} \tag{15}$$

$$\boldsymbol{R}_{YY}=\frac{1}{T}\sum_{c=1}^{C}\sum_{j=1}^{|Y_c|}|\boldsymbol{X}_c|\boldsymbol{y}_j^c(\boldsymbol{y}_j^c)^{\mathrm{T}} \tag{16}$$

式中：T 为建立对应关系的总对数，$T=\sum_{c=1}^{C}|\boldsymbol{X}_c||\boldsymbol{Y}_c|$。

4　实验结果及分析

本实验所用的数据集为慕尼黑工业大学建立的触觉纹理数据集[15]。数据集中包含 108 种不同的物体，按照材质和表面特征分为固体网状物、石头、玻璃陶瓷、木材、橡胶、纤维、泡沫、塑料纸片、纺织面料等九大类，具体每类物体的图像如图 5 所示，图 5 中数字表示该类材质第一个物体的起始位置。训练集包括声音集和图片集，声音集中每个声音样本由一个人敲击待测物体表面 1 次所得，其长度为 0.2 s。将 108 种待测物体每种重复敲击 10 次，共得到 1 080 个声音样本。图片集每张图片分辨率为 320×480，在不打开闪光灯情况下，同样由一个人重复拍摄待测物体 10 次所得，共得到 1 080 张图片样本。测试集数据数量和样本大小与训练集相同，不同之处在于采集数据的过程有所差别，测试集中声音和图片样本不是由同一个人重复 10 次完成，而是由 10 个不同的人每人采集 1 次所得。整个数据集的特点是采集数据的过程均为人工完成，没有施加约束条件，例如敲击物体表面时，没有限制施加力的大小。

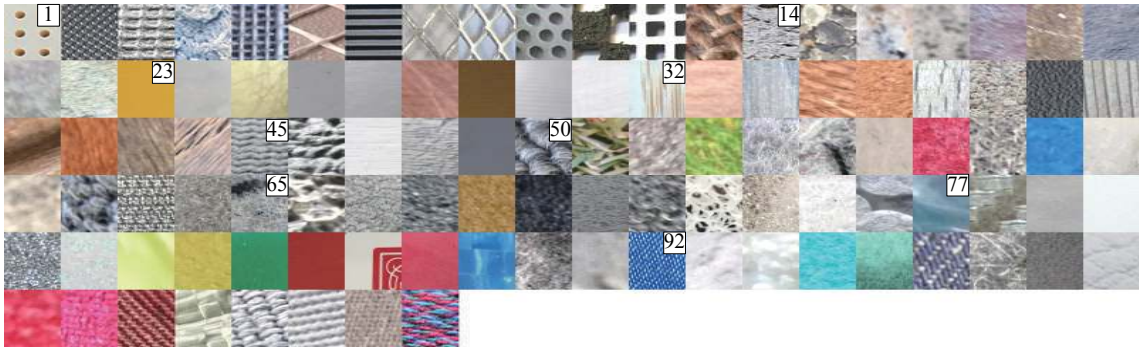


图 5　数据集中包含的所有材料

Fig. 5　Materials included in the data set

根据第 2 章得到的 39 维声音特征和 4 096 维图像特征，应用于第 3 节所述典型相关分析方法，找到训练集中声音特征和图像特征典型相关分析子空间，然后将测试集中的声音特征和图像特征映射到典型相关分析的子空间，即可使用子空间的声音特征去检索图像特征，通过计算欧氏距离度量样本特征的相似性。

实验最终在测试集上执行从声音到图像的跨模态信息检索。常用的信息检索的评价指标有查准率 P、查全率 R 和平均准确率 (mean average precision, MAP) 等。PR 曲线比较直观地显示出检索效果的好坏，MAP 则考虑到检索结果的排名情况。PR 曲线与坐标轴围成的面积越大，MAP 值越高，则检索效果越好。本文使用 MAP 和 PR 曲线对 RCCA (同种物体声音图像随机匹配)、MCCA 和 CCCA 3 种方法的实验结果进行评价。图 6 所

示为 3 种不同方法的 MAP 值的大小随子空间维度的变化, 从图 6 可以得到, 子空间维度为 5 时, 3 种方法效果最好, 且 CCCA 的 MAP 值明显优于其他 2 种方法。

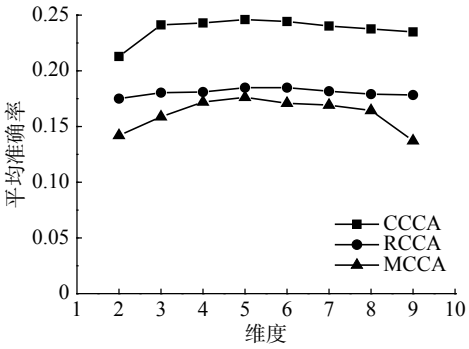


图 6　不同方法的 MAP 值随子空间维度的变化

Fig. 6　Variation of the MAP of different methods with subspace dimensions

图 7 所示为子空间维度为 5 时, 3 种方法的 PR 曲线, 从中可以看出, CCCA 的 PR 曲线与坐标轴围成的面积最大, 检索效果最好。由于所使用的数据集中的数据不符合传统意义上的一一配对关系, RCCA 和 MCCA 的检索效果不如 CCCA。

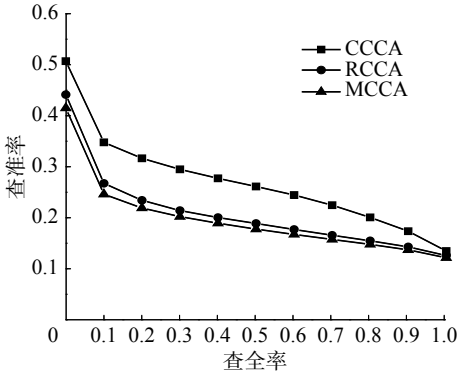


图 7　PR 曲线

Fig. 7　PR curve

图 8 所示为数据集中纤维、泡沫和塑料 3 种材料图像和声音数据的低维映射, 其中蓝色代表纤维, 黄色代表泡沫, 红色代表塑料。从图 8 中可以看出, CCCA 对这 3 类材料的区分度要强于 RCCA 的效果。

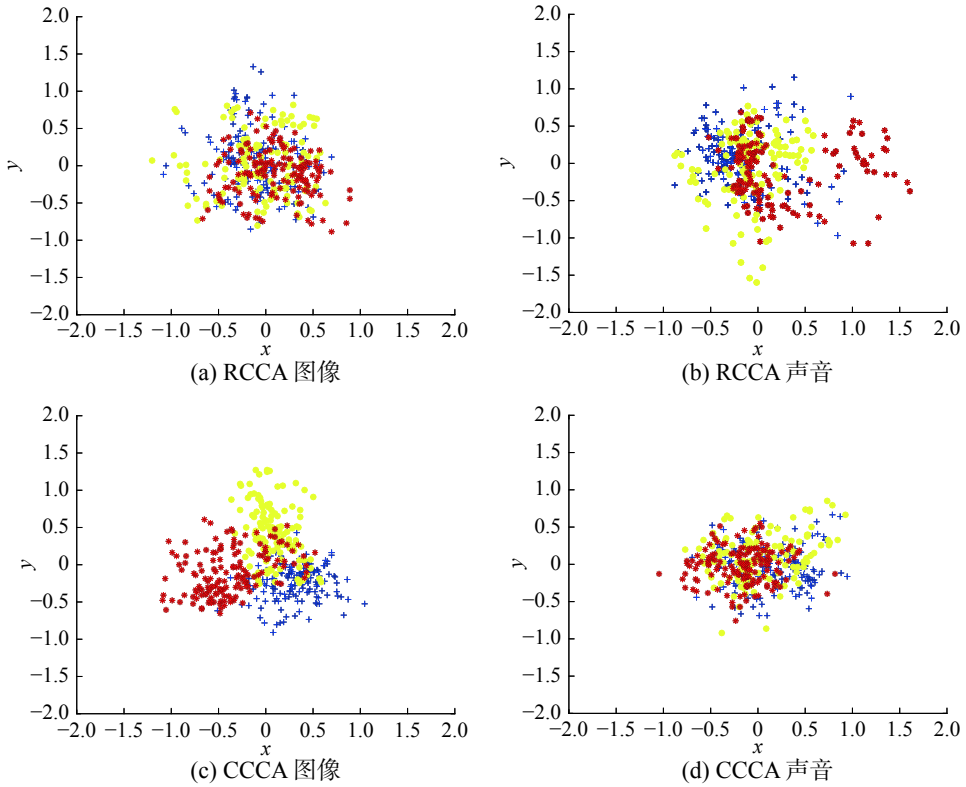


图 8　3 种材料的低维映射图

Fig. 8　Low-dimensional mapping of three materials

表 1 为 3 种方法下不同材质类别的 MAP 大小, 图 9 为对应的柱形图。整体结果显示, 本文引入的 CCCA 在硬质材质 (固体网状物、石头、玻璃陶瓷等) 的检索效果比软质材料 (橡胶、纤维、泡沫等) 好, 这主要由于本文所使用的声音数据是由敲击物体表面所得, 而实验过程中待测物体放置在实验台上, 采集数据时容易受到实验台影响。特别是, CCCA 在石头这类材料测试中的表

现尤为出色，MAP 值达到 0.32，比 RCCA 和 MCCA 高 50%。

表 1　不同材质类别的 MAP

Table 1　MAP of different categories of material

材质类别	CCCA	RCCA	MCCA
固体网状物	0.30	0.21	0.19
石头	0.26	0.13	0.13
玻璃陶瓷	0.26	0.24	0.13
木材	0.32	0.23	0.19
橡胶	0.09	0.09	0.09
纤维	0.24	0.18	0.20
泡沫	0.15	0.13	0.13
塑料纸片	0.19	0.16	0.17
纺织面料	0.26	0.19	0.19

图 10(a) 所示为测试集一个竹木图片，图 10(b) 为敲击这种竹木的声音样本，使用 CCCA 进行检索，检索得到图 10(c) 所示的 10 张图片，从左到右依次为落叶松木、纺织网、石瓦片、铝板、樱桃树木、压缩木板、落叶松木、山毛榉木、压缩木材、银橡木。从实验结果可以看出，与测试集竹木样本最相似的 10 个结果有 7 个和测试样本属于同一类别，检索正确率达到 70%，可见 CCCA 在木材类材质识别效果较好。

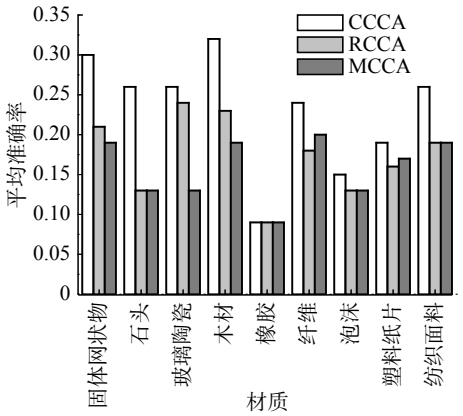
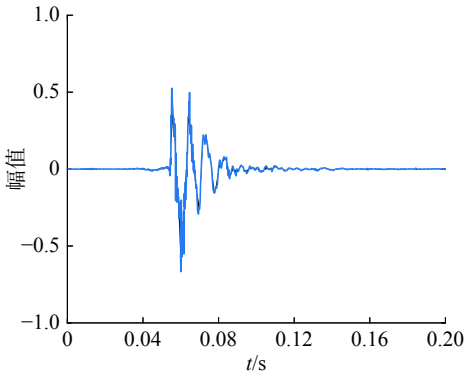


图 9　不同材质类别的 MAP

Fig. 9　MAP of different categories of material

(a) 竹木图片



(b) 声音样本

(c) 检索结果

图 10　使用竹木声音样本的检索结果

Fig. 10　Retrieval result of bamboo sound sample

5　结束语

本文跨越不同模态之间的限制，结合声音图像特征与典型相关分析方法，将跨模态检索方法应用于材质检索领域，在慕尼黑工业大学触觉纹理数据集上取得较好效果。虽然通过实验验证该方法目前的效果存在一定的局限性，但随着不同

模态信息的不断加入和特征提取的方法不断改进，未来该方法的应用前景必定更加广阔。

参考文献：

[1] MANDAL D, BISWAS S. Query specific re-ranking for improved cross-modal retrieval[J]. Pattern Recognition Letters, 2017, 98: 110–116.

[2] WANG Kaiye, HE Ran, WANG Liang, et al. Joint feature selection and subspace learning for cross-modal retrieval [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(10): 2010–2023.

[3] DENG Cheng, TANG Xu, YAN Junchi, et al. Discriminative dictionary learning with common label alignment for cross-modal retrieval[J]. IEEE transactions on multimedia, 2016, 18(2): 208–218.

[4] ZHANG Liang, MA Bingpeng, LI Guorong, et al. Metric based on multi-order spaces for cross-modal retrieval[C]// Proceedings of 2017 IEEE International Conference on Multimedia and Expo. Hong Kong, China, 2017: 1374-1379.

[5] 张毅, 谢延义, 罗元, 等. 一种语音特征提取中 Mel 倒谱系数的后处理算法 [J]. 智能系统学报, 2016, 11(2): 208–215.
ZHANG Yi, XIE Yanyi, LUO Yuan, et al. Postprocessing method of MFCC in speech feature extraction[J]. CAAI transactions on intelligent systems, 2016, 11(2): 208–215.

[6] WEI Yunchao, ZHAO Yao, LU Canyi, et al. Cross-modal retrieval with CNN visual features: a new baseline[J]. IEEE transactions on cybernetics, 2017, 47(2): 449–460.

[7] RANJAN V, RASIWASIA N, JAWAHAR C V. Multi-label cross-modal retrieval[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 4094-4102.

[8] SHARMA A, KUMAR A, DAUME H, et al. Generalized multiview analysis: a discriminative latent space[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2160-2167.

[9] HARDOON D R, SZEDMAK S, SHAWE-TAYLOR J. Canonical correlation analysis: an overview with application to learning methods[J]. Neural Computation, 2004, 16(12): 2639–2664.

[10] CHEN Yongming, WANG Liang, WANG Wei, et al. Continuum regression for cross-modal multimedia retrieval[C]//Proceedings of the 19th IEEE International Conference on Image Processing. Orlando, USA, 2013: 1949-1952.

[11] MANDAL D, BISWAS S. Generalized coupled dictionary learning approach with applications to cross-modal matching[J]. IEEE transactions on image processing, 2016, 25(8): 3826–3837.

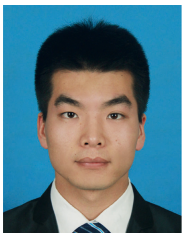
[12] STRESE M, SCHUWERK C, IEPURE A, et al. Multimodal feature-based surface material classification[J]. IEEE transactions on haptics, 2017, 10(2): 226–239.

[13] CAO Jiuwen, ZHAO Tuo, WANG Jianzhong, et al. Excavation equipment classification based on improved MFCC features and ELM[J]. Neurocomputing, 2017, 261: 231–241.

[14] RASIWASIA N, MAHAJAN D, MAHADEVAN V, et al. Cluster canonical correlation analysis[C]//Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland, 2014: 823-831.

[15] STRESE M, BOECK Y, STEINBACH E. Content-based surface material retrieval[C]//Proceedings of 2017 IEEE World Haptics Conference. Munich, Germany, 2017: 352-357.

作者简介：



刘卓锟，男，1994 年生，硕士研究生，主要研究方向为新型磁性材料与器件、触觉感知与模式识别。

刘华平，男，1976 年生，副教授，博士生导师，IEEE Senior Member、中国人工智能学会理事，中国人工智能学会认知系统与信息处理专业委员会秘书长，主要研究方向为机器人感知、学习与控制、多模态信息融合。主持国家自然科学基金 5 项。发表学术论文 200 余篇，被 SCI 检索 100 余篇。

黄文美,女,1969 年生,教授,主要研究方向为磁性材料与器件、电机及其控制技术。完成国家自然科学基金项目 4 项、河北省自然科学基金项目 2 项。发表学术论文 40 余篇,被 SCI、EI、ISTP 检索 20 余篇。