

基于融合 MPEG-7 视觉描述符的图像分类方法

王松, 王卫红, 秦绪佳

(浙江工业大学软件学院, 杭州 310032)

摘要: 提出了利用融合不同的低层 MPEG-7 视觉描述符的方法来进行基于内容的图像分类的技术。目的在于通过融合几种描述符来改善机器学习分类器的性能, 包括 3 种方法来改善分类器的性能: 作用于支持矢量机 (SVM) 分类器的聚类融合, 作用于 K 近邻分类器的反向传播 (BP) 融合和作用于 FART 模糊神经网络的 BP 融合。将这些分类方法应用到海滩风景/城市风景的分类的实验中, 实验结果表明 BP 融合显示出更好的性能改善。

关键词: 视觉描述符; 分类器; 融合; 特征提取; 模糊神经网络

Image Classification Based on Fusing MPEG-7 Visual Descriptors

WANG Song, WANG Weihong, QIN Xujia

(College of Software, Zhejiang University of Technology, Hangzhou 310032)

【Abstract】 Several content-based image classification techniques based on fusing various low-level MPEG-7 visual descriptors are proposed in this paper. The goal is to fuse several descriptors in order to improve the performance of several machine-learning classifiers. Three approaches are described: A “merging” fusion combined with an SVM classifier, a back-propagation fusion combined with a K-Nearest Neighbor classifier and a Fuzzy-ART neurofuzzy network. In the latter case, fuzzy rules can be extracted in an effort to bridge the “semantic gap” between the low-level descriptors and the high-level semantics of an image. Experimental results on the beach/urban scenes classification problem show the best.

【Key words】 Visual descriptor; Classifier; Fusing; Feature extraction; Neurofuzzy network

基于内容的图像检索 (CBIR) 就是从大规模的多媒体数据库中检索出满足用户需求的图像或图像集。现在大多数的 CBIR 技术是直接从图像中对视觉信息 (颜色、纹理、形状等) 进行提取, 这些都是低层的特征提取^[1]。实际上, 图像检索系统的用户往往是根据图像的含义, 来直观地进行分类, 并判别图像之间的相似性, 因而这就包含了图像的高层语义特征。

但是, 图像简单视觉特征与用户检索丰富语义之间的“语义鸿沟”的减少现在仍然没有得到较好的解决。无论分类是否被监督, 无论使用什么样的分类模型 (例如, 最大期望值、向量量化、k-means^[2]、SVM^[3]或神经网络), 通常认为, 为了取得鲁棒的全局分类, 选择合适的描述符集合是至关重要的。视觉描述符通常必须捕获指定区域的独特属性和每种图像类的独立特征。例如, 局部颜色描述符和全局颜色直方图被用于图像的室内/室外分类中; 由于城市图像典型意义上包含水平和垂直边界, 因此边界直方图被用于城市风景/野外风景的分类中; 附加动态描述符可被用于体育视频快照的分类中^[4]。

在本文中, 几种 MPEG-7 描述符的融合是与 3 种不同的机器学习技术结合在一起的, 包括: 聚类描述符的融合与 SVM 分类器相结合; 一个 BP 前馈神经网络被训练, 用于基于低层描述符的两幅图像之间的距离计算, 把计算结果与 K 近邻分类器相结合; 最后, 为了提取模糊规则, 并且在图像的低层特征与高层语义特征之间解决“语义鸿沟”的问题, 使用 Falcon-ART 模糊神经网络。使用相同的训练集对这 3 种神经网络进行训练, 并且所有的 3 种分类策略都使用相同的测试集对其进行评估。

1 特征提取

为了能够提供 audio-visual (AV) 内容的标准描述, MPEG-7 标准^[4]制定了一组描述符, 每种描述符定义了某种基本视觉低层特征的语法和语义。本文中的图像分类是基于 MPEG-7 中的 3 个视觉描述符的。它们是利用 MPEG-7 标准开发工具——eXperimentation Model (XM) 进行提取的。为了对图像进行识别和分类, 我们使用了这些低层视觉描述符的子集。下面对这 3 个视觉描述符进行简单的介绍。

1.1 颜色布局描述符 (CLD)

CLD 是一个被定义在 YCrCb 颜色空间的压缩 MPEG-7 视觉描述符, 用于捕获一幅图像或任意区域的颜色空间分布。特征提取过程如下: 首先将图像分成 8×8 的小块, 然后对这些小块进行 DCT 变换; 得到的变换结果系数再进行 zig-zag 扫描, 这时图像小块可以由一个 12 维的向量表示 (其中 6 个系数是表示亮度的, 每个色度由 3 个系数表示); 最后, 将这些系数进行非线性量化。

1.2 可伸缩颜色描述符 (SCD)

SCD 是一个基于编码的 Harr 变换得出的颜色直方图, 用于测量整幅图像的颜色分布。颜色空间使用 HSV 颜色空间, 并且被统一量化为 256 块。

1.3 边界直方图描述符 (EHD)

EHD 捕获边界的空间分布, 甚至当纹理不均匀时, 仍然

基金项目: 国家自然科学基金资助项目 (60473024); 浙江省科技计划基金资助项目 (2004C30065)

作者简介: 王松 (1978 -), 男, 讲师、硕士, 主研方向: 图形图像, 多媒体技术等; 王卫红, 副教授、硕士; 秦绪佳, 副教授、博士

收稿日期: 2006-03-14 **E-mail:** ws@zjut.edu.cn

可以提供有效的边界描述,包括4个方向(0°,45°,90°,135°)的边界加上一个无方向边界被检测。输入的图像被分割成互相不覆盖的16块,应用基于块的提取策略进行5种边界的提取,并计算它们的相对群体,构成一个80维的向量。

2 基于 MPEG-7 视觉描述符的图像分类

为了同时处理以上的描述符来完成诸如相似性计算,距离向量形式化,分类器的训练等任务,融合描述符的不同元素就是理所当然的。

当基于某个 MPEG-7 视觉描述符比较两幅图像时,可以采用好几种距离函数来计算。然而,本文中比较时应当考虑上节所有的3个低层描述符,并且对其加上不同的权重。因为没有统一的方法进行距离的计算,这个问题的解决相当麻烦;除此之外,标准的距离函数的效率在每个给定数据库中也无法保证。本文将提供3种方法来解决这个问题,它们使用上节描述的 MPEG-7 描述符。

2.1 聚类融合/SVM 分类器

聚类融合是一种非常简单、直接的融合方法,就是将所有的描述符融合为一个唯一的向量。这种方法需要所有的特征都具有或多或少的同样的数值,以避免特征之间比例的失衡。MPEG-7 的描述符在提取时已经换算成相同量级上的整数,故在本文中不需要对不同特征值进行比例的重新调配。

2.2 使用 BP 融合的 K 近邻分类

此种方法是基于含有隐藏层的 BP 前馈神经网络的。该网络的输入是由需要计算距离的两幅图像的低层描述符来构成,输出是两幅图像距离的基于描述符的标准化估计。训练集来自于可用数据库中经过仔细选择的具有典型代表意义的图像构成。该网络在此假设下进行训练:当两幅图像属于同一类时,距离最小;当两幅图像属于不同类时,距离最大。在网络中的所有图像经过此训练之后,就可以建立一个矩阵。该矩阵的元素是由图像的行和列之间的估算距离组成。然后,该矩阵就可以被 K 近邻分类器利用,以便指定一幅图像的 K 近邻的多数属于哪一类,该幅图像就属于哪一类。

但是,此时存在一个问题,属于同一幅图像的描述符之间的距离并不是预期的0值,而是包括一些很小的非0值。当然,由于已经知道,同一幅图像描述符之间的距离应该是0值,因此可以用0来替换这些非0值。除此之外,即使一个网络被训练得很好,其输出仍然会有细微的差别,由上面的方法得出的距离矩阵也可能无法满足 K 近邻分类器所需要的距离基本属性。为了克服这个问题,我们将只采用距离矩阵的上三角部分的值或下三角部分的值。一种方法是距离矩阵中的新距离的计算按如下公式计算:

$$\hat{d}(i, j) = \hat{d}(j, i) = d(\max(i, j), \min(i, j))$$

另一种方法是距离矩阵中的新距离的计算按如下公式计算:

$$\hat{d}(i, j) = \hat{d}(j, i) = \frac{1}{2}(d(i, j) + d(j, i))$$

经过这两种修正之后,新距离矩阵就可以满足 K 近邻分类器所需的距离基本属性。

此外,还有一种方法可以有效地对不同视觉描述符进行融合:对独立的视觉描述符使用预先设定好的距离矩阵,并且对每个描述符设置不同的权重。通过成对图像之间的 CLD, SCD, EHD 的欧几里德距离来组成3个矩阵。此时,神经网络的输入就由这3个矩阵组成,输出是标准化的融合距离。作为结果的距离矩阵又可以被用于 K 近邻分类器的输

入。此种方法得到的距离矩阵是对称的,然而,由于矩阵对角线上的值不全为0,为了满足 K 近邻分类器的要求,对角线上的值需要重新设置为0。

2.3 使用 Falcon-ART 模糊神经网络的分类

以上的两种图像分类方法均不能提供实现分类时所需机制的语义分析,为了能够提取语义信息,我们采用了 Falcon-ART 模糊神经网络^[6],它是基于 Fuzzy-ART 聚集算法的^[7],其简单的描述如下:

网络的训练经过两个阶段:(1)结构学习阶段,Fuzzy-ART 算法被用来创建网络的结构;(2)参数学习阶段,网络的参数经过 BP 算法来改善。在输入空间中由 Fuzzy-ART 算法来建立超盒(hyperboxes),每个超盒对应一个合并集,并沿着属于对应类别的训练样本的方向增长。依照警戒标准新的超盒也可以被建立,并且在应用 Fuzzy-ART 算法之后,输入空间被合并,从输入的超盒与它们之间的连接中可以建立一个5层网络。参数和每层的节点之间的连接来源于 Fuzzy-ART 算法,作为结果的5层网络通过 BP 算法来训练。

包含所有以上3个 MPEG-7 视觉描述符的低层信息的融合描述符作为这个网络的输入,在经过训练之后,该网络的响应是输入分类的标准。然而,模糊神经网络的主要优势是能够提取语义模糊规则。因此,利用该方法,决定某幅图像属于哪一类的低层特征,显而易见,可以使用自然语言来表达。假设 $D = [d_1, d_2, \dots, d_n]$ 是一个低层描述符,根据这个描述符,可以推出以下的模糊规则:

IF d_1 是 low AND d_2 是 medium AND ... AND d_n 是 high

THEN 该幅图像属于 class 1

3 实验结果

本文采用的图像数据库是由海滩风景和城市风景两类高质量图像组成的,共有767幅。我们选择其中的60幅(40幅海滩图像,20幅城市图像)图像来对神经网络进行训练,剩余的707幅图像被用来测试不同分类方法的效率。3个视觉描述符的所有可能的组合都被考虑进本文提出的3种方法中去了,结果如表1所示。

表1 基于不同 MPEG-7 描述符 (EH,CL,SC) 的分类效率

分类	EH	CL	SC	EH+CL	EH+SC	CL+SC	EH+CL+SC
聚类/线性 SVM	79.5%	82.3%	83.6%	87.1%	88.7%	86.9%	89.0%
BP/KNN	81.9%	87.13%	85.86%	67.04%	90.1%	91.37%	86.28%
Falcon-ART	81.4%	84.7%	83.67%	82.4%	83.6%	86.3%	87.7%

(1)使用聚类融合的 SVM 分类器

直接将融合的向量作为带有第一级多项式核的 SVM 分类器的输入,与作为更高级多项式核的 SVM 分类器的输入,其结果相差不大。因此本文选择前者。40幅海滩图像和20幅城市图像被选择作为给定数据库的分类的样本,然后使用这些样本在 SVM 分类器中训练数据集。数据库中剩余的707幅图像被用来作测试。表1中的结果显示,当几种低层特征进行融合时,分类的效率会提高,3种一起融合时,达到最佳效率89%。

(2)使用 BP 融合的 K 近邻分类器

此种方法中,当两幅图像属于同一类时,它们之间的距离设为0;反之,设为1。由于网络的预期响应是独立于每对图像的排列顺序的,因此为了改善训练,每对图像都会建立正向和反向两个训练样本。可以建立2800个训练样本来训

练 BP 网络。剩余的 707 幅图像用于评价 BP 融合。把网络中所有可能的图像对表示出来，一个对称距离矩阵就建立了，并且可被用于 K 近邻分类器。

(3)Falcon-ART 神经网络

将上面 BP 融合的同样的 60 幅图像用于 Fuzzy-ART 分类器和 Falcon-ART 模糊神经网络的训练集，同时利用网络中随机输入的融合描述符来训练网络。在训练集中匹配成功率可以达到 95.8%，在测试集中可以达到 87.7%，并且 Fuzzy-ART 算法可以建立 8 个超盒，Falcon-ART 模糊神经网络被训练为 275 个初相。为了让 Fuzzy-ART 算法建立的规则更接近于人类感知，图像描述符的每一维都被分成 3 个相同部分，对应的值是：*low, medium, high*。由 Fuzzy-ART 算法建立的超盒引出的规则就使用这些值。

表 2 由 Falcon-ART 建立的模糊规则

图像部分	边界类型	规则 1	规则 2	规则 3	规则 4	规则 5
上部	0°	M	L	M-L	M-L	L
	45°	M	L	M-L	M	M
	90°	M	L	M	M	M
	135°	H	M	M	M	M
	无方向	M	M	M	M	M
中部	0°	M	L	M	M	M
	45°	M	M-L	H	M	H
	90°	M	M	M	M	H
	135°	H	M	M-L	H	H
	无方向	H	M	M	H	M
底部	0°	M	L	L	M	L
	45°	M	M	H	H	H
	90°	H	M	M	H	H
	135°	M	M-L	M	H	M
	无方向	M	M	M-L	H	M
类别		城市 风景	海滩 风景	城市 风景	海滩 风景	海滩 风景

使用第 2 节中表达的 EHD 描述符来作为分类时提取的特征。每幅图像被分成上，中，下 3 个子图像部分，每种边界

(上接第 174 页)

该可视化程序，可以实现新建决策集、打开决策集、关闭决策集、添加决策、删除决策、显示某条或全部决策、删除决策集等功能。

只要理解输入界面中各个因素中的数字代码的含义就可以编制覆盖各个决策的决策库。比赛时，将此决策库与客户程序的决策子系统连接，就能够进行比赛了。

4 结论

机器人足球比赛的成绩关键在于决策，但决策的编程却是青少年的一大难题。本文以菜单的形式实现了决策编程的可视化。

足球机器人比赛决策程序可视化编程技术已用于广州市科技计划项目“基于视觉的青少年机器人足球比赛系统”中，是其系统软件的一部分，为中小學生提供可视化的编程平台，降低了青少年机器人足球比赛决策编程的门槛，使青少年从繁杂的编程中解脱出来，从而致力于决策的开发，便于中小學生编制体现自己思想的比赛决策，激发他们爱科学、学科

都被赋上 *low, medium, high* 这 3 个定性的值。得到如下的模糊规则：

IF 图像上部分的 0° 边界值是 *low* AND 图像上部分的 45° 边界值是 *medium* AND ... AND 图像下部分的无方向边界值是 *high*

THEN 这个图像属于海滩类

利用 EHD 描述符，Falcon-ART 建立了 5 个模糊规则，详细描述如表 2 所示，其中 H 表示 *high*，M 表示 *medium*，L 表示 *low*。

4 结论

实验表明，使用 MPEG-7 描述符来进行图像分类的 3 个测试方法都可以成功应用，BP 融合显示出最好的分类结果。然而，由于使用 Falcon-ART 的融合可以提供潜在的分类机制的语义描述，因而也是很有用的。未来的工作，我们将会集中于使用更多的 MPEG-7 描述符和应用到更多的分类中。另外，这些分类策略也可以扩展到预定义对象模型的图像分割中。

参考文献

- 徐杰, 施鹏飞. 基于内容的图象检索技术[J]. 中国图象图形学报, 2003, 8(9): 977-983.
- Duda R O, Hart P E. Pattern Classification[M]. New York: John Wiley and Sons, 2001.
- Vapnik V. Statistical Learning Theory[M]. New York: John Wiley and Sons, 1998.
- Wang D H, Tian Q. News Sports Video Shot Classification with Sports Play Field and Motion Features[C]. Proc. of IEEE International Conference on Image Processing, Singapore, 2004: 2247-2250.
- Manjunath B, Ohm J R. Color and Texture Descriptors[J]. IEEE Trans. on Circuits and Systems for Video Technology, 2001, 11(6): 703-715.
- Carpenter G, Grossberg S, Rosen D. A Neural Network Realization of Fuzzy Art[R]. MA: Boston University, Technical Report: CAS/CNS-91-021, 1991.
- Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer-Verlag, 1995.

学、用科学的热情，有助于青少年机器人足球比赛的发展。

参考文献

- 王万民. 智能机器人[M]. 广州: 广东科技出版社, 2004.
- 徐爱平. 让智能机器人教育进入中小学[J]. 机器人技术与应用, 2004, (1): 14-16.
- 彭志刚, 杨艳丽. 微型机器人的设计与开发讲座——机器人足球与足球机器人[J]. 机器人技术与应用, 1999, (5): 29-31.
- 王学文, 赵妹颖, 孙萍等. 多智能足球机器人系统的关键技术[J]. 东北大学学报, 2001, 22(2): 192-195.
- Kwon Y D, Shin D M, Won J M, et al. Multi Agents Cooperation Strategy for Soccer Robots[C]. Proceedings of Conference on FIRA Robot World Cup 98, Paris, 1998.
- 唐华斌, 李响. Micro-robot World Cup Soccer Tournament FIRA[R]. 北京: 中国自动化学会机器人竞赛工作委员会, 2003: 200-206.
- 薛方正, 冯挺, 李平等. Simulative Robot World Cup Soccer Tournament FIRA[R]. 北京: 中国自动化学会机器人竞赛工作委员会, 2003: 147-154.