

基于相位奇点包的图像分类方法研究

王星星¹ 乔宇^{1,2}

¹ (中国科学院深圳先进技术研究院 深圳 518055)

² (香港中文大学 香港 999077)

摘要 相位奇点是复信号中的零点。在此前的研究中,我们发现相位奇点可提供丰富的图像信息,并对变换噪声等具有鲁棒性。本论文将相位奇点与视觉词袋模型相结合,构建图像表示的相位奇点包表示方法,并在此基础上开发图像的分类方法。与传统基于SIFT检测子的图像表示方法相比,基于相位奇点的表示方法不仅可以找到更多的兴趣点,而且可以根据相位奇点的正负属性进行单词的预分类。实验结果表明作者提出的方案是可行的,并在PASCAL2005图像分类的结果上表现的比基于SIFT的方法更好。

关键词 相位奇点; 图像分类; 视觉词袋模型

Image Classification Using Bag of Phase Singularities

WANG Xing-xing¹ QIAO Yu^{1,2}

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(The Chinese University of Hong Kong, Hong Kong 999077, China)

Abstract Phase singularities (PSs) refer to the zero points in complex signals. In previous studies, we found that PSs can provide rich information for image representation and are robust to image transformation, noise etc. This paper makes use of PS and bag of visual words (BoVW) model to construct bag of PS representation for images. Then we use SVM to classify bag of PS representations. Compared with previous works using SIFT points, our new representations not only use more interested points, but also allow us to pre-classify the words according to the sign property of PSs. The experimental results show that the proposed methods achieve better performance on PASCAL2005 image classification tasks than SIFT detectors.

Keywords phase singularity; image classification; bag of visual words

1 引言

图像分类是计算机视觉和图像处理中的一个基本问题。近年来伴随着互联网和信息技术的发展,以及数码相机DV等设备的普及,人们日常拍摄、存储和使用数字图像的机会和数目日益增多。与此同时对智能高效的自动图像分类技术的需求也越发强烈,并引起研究人员持续和强烈的关注。

图像分类中的一个基本问题是如何建立有效的图像表示方法。一方面,我们希望图像表示能提供丰富

的内容和类别信息,以表达图像的高层语义;另一方面,我们也希望图像表示对类内变化,如光照、拍摄角度、同类图像的不同个体等有较强的鲁棒性。传统的图像分类方法,往往针对某类特定的对象如文字、人脸等,这些方法利用对象的特点设计特征和分类方法,对象不同所使用的特征和分类方法也不同。近年来,一般图像的分类方法,越发受到关注。与传统的特定对象识别相比,进行非特定和一般图像的分类和识别需要建立统一的图像表示方法。广大的研究人员在这一领域开展了大量的工作,其中受自然语言处理中词袋模型启发的视觉词袋(Bag of Visual Words)模型^[3]

被广泛的应用，并取得了巨大的成功。

将视觉词袋用于图像分类主要包括以下几个步骤，首先从图像中抽取局部特征作为基本表示元素，然后对局部特征聚类形成视觉单词的字典，我们可以利用这个字典对局部特征进行量化。对于某种特定的图像，我们可以通过统计图像中各视觉单词出现的个数来构成直方图作为图像的视觉词袋表示。可以看出，在上述过程中，图像的局部特征是构建视觉词袋表示的基本单元，起着重要的作用。此前的研究常常使用SIFT检测算子从图像中检测兴趣区域，并从中抽取SIFT特征作为构建词袋模型的基本特征。这一方法在早期的图像识别中取得了一定的成功^[1]。

在参考文献[2]中，我们提出基于相位奇点的图像表示和分析方法。相位奇点指图滤波后得到的复信号中的零点，在这些点无法计算相位，因而被称为相位奇点。在传统工作中，相位奇点一般被认为是应该被去除的不利因素，但是我们通过研究发现，相位奇点的位置和分布中含有丰富的图像信息。我们还分析并验证了相位奇点对图像平移缩放，以及噪声的稳定性。本文利用相位奇点作为构建图像词袋表示的基本元素，开发了相位奇点包模型。由于相位奇点可以分为正负两类，于此对应，相位奇点包模型也可以分为正负两种。本文引入的相位奇点相对于SIFT检测子在同一图像上可以检测到更多的兴趣点，同时相对于密集采样有着的生成图像特征计算代价少和兴趣点信息冗余度相对较小等优点。

本文的第2节介绍了相位奇点的性质和检测方法以及对应使用的描述子，第3节介绍图像分类使用到的BoVW模型，第4节介绍实验设计，第5节介绍结果与将来的工作。

2 相位奇点及其基本属性

2.1 相位奇点的定义

数学上相位奇点(Phase Singularities, 简称PS)是指复数信号值中实部和虚部均为零的点，在这些点无法计算相位。因为数字图像本身的像素值是实数。要寻找相位奇点，首先要对图像进行处理使之成为复数信号。我们将二维图像 $I(x,y)$ 与二维复数滤波器 $Q(x,y,\sigma)=Q_r(x,y,\sigma)+iQ_i(x,y,\sigma)$ 进行卷积运算从而产生复数图像 $\hat{I}(x,y,\sigma)$ ，其实部虚部相同位置上的值等于0的点即为相位奇点。

二维图像与二维复数滤波器进行卷积运算的公式

如下：

$$\begin{aligned}\hat{I}(x,y) &= I(x,y) * Q(x,y) \\ &= I(x,y) * Q_r(x,y) + iI(x,y) * Q_i(x,y)\end{aligned}$$

其中‘*’表示卷积， σ 表示尺度因子，选择不同的 σ 可以构建不同尺度的二维复数滤波器，用不同尺度的滤波器对图像滤波可得到不同尺度的滤波图像。

本文所使用的二维复数滤波器由选定的母滤波器 $T(x,y,\sigma)$ 分别对 x,y 方向求导得到：

$$\begin{aligned}Q_r(x,y,\sigma) &= \frac{\partial T(x,y,\sigma)}{\partial x} \\ Q_i(x,y,\sigma) &= \frac{\partial T(x,y,\sigma)}{\partial y}\end{aligned}$$

母滤波器可以选择LoG, Gabor, Hermite等。

由此复数图像可以分解为：

$$\begin{aligned}\hat{I}(x,y,\sigma) &= I(x,y) * T(x,y) \\ &= I(x,y) * \frac{\partial T(x,y,\sigma)}{\partial x} + iI(x,y) * \frac{\partial T(x,y,\sigma)}{\partial y} \\ &= E_x(x,y,\sigma) + iE_y(x,y,\sigma)\end{aligned}$$

这里 $E(x,y,\sigma)=I(x,y)*Q$,

$$\begin{aligned}E_x(x,y,\sigma) &= \frac{\partial E(x,y,\sigma)}{\partial x} \\ E_y(x,y,\sigma) &= \frac{\partial E(x,y,\sigma)}{\partial y}\end{aligned}$$

PS点是 $\hat{I}(x,y,\sigma)=0$ 的点，即 $E_x(x,y,\sigma)=0$ 和 $E_y(x,y,\sigma)=0$ 的位置。我们采用

$$PS(x,y,\sigma) = \left\{ (x,y,\sigma) \mid \hat{I}(x,y,\sigma) = 0 \right\}$$

表示固定尺度下图像的相位奇点集合，相位奇点也可以看成是零交叉曲线集合 $E_x(x,y,\sigma)=0$ 和 $E_y(x,y,\sigma)=0$ 的交点。以上得出的相位奇点具有位置 (x,y) 和尺度 σ 信息。

2.2 相位奇点的正负性

在向量微积分中，雅克比(Jacobian)矩阵是指一阶偏导数以一定方式排列的矩阵，Hessian矩阵是指以自变量为向量的实值函数的二阶偏导数组成的方块矩阵。那么可以得到 $\hat{I}(x,y)$ 的实部和虚部的Jacobian矩阵等于 $E(x,y)$ 的Hessian矩阵。

$$\begin{aligned}J(\hat{I}(x,y,\sigma)) &= \begin{bmatrix} \text{Re}(\hat{I}(x,y,\sigma))_x & \text{Re}(\hat{I}(x,y,\sigma))_y \\ \text{Im}(\hat{I}(x,y,\sigma))_x & \text{Im}(\hat{I}(x,y,\sigma))_y \end{bmatrix} \\ &= \begin{bmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{bmatrix} = H(E(x,y))\end{aligned}$$

由此可以根据Hessian的行列式的值将相位奇点分为正负2种。从数学定义可知，正相位奇点对应于

$E(x,y)$ 的极值点, 负相位奇点对应于 $E(x,y)$ 的鞍点。

如图1所示, 其中的红色代表检测到的正相位奇点, 蓝色代表检测到负相位奇点, 圆半径的大小表示相位奇点的尺度 σ 。



图1 左图表示当尺度固定的时候, 正负相位奇点相间分布
右图显示的是全尺度情况下相位奇点的分布

2.3 相位奇点的定位

相位奇点附近的复数信号的实部和虚部可以由两个平面近似^[5, 6]:

$$\begin{aligned} E_x &\approx a_r x + b_r y + c_r, \\ E_y &\approx a_i x + b_i y + c_i \end{aligned}$$

根据上述两个平面近似和PS点的定义: $\hat{I}(x,y,\sigma) = 0$ 的点, 即 $E_x(x,y,\sigma) = 0$ 和 $E_y(x,y,\sigma) = 0$ 可以初步求解相位奇点的空间位置 (x,y,σ) 。

相位奇点的精确定位需要进行寻找滤波图像的极值点或鞍点, 在这我们采用滤波图像 $E(x,y)$ 的Taylor二阶展开来进一步精确定位。滤波图像函数的Taylor展开式。

$$\begin{aligned} E(X) &= E + \frac{\partial E^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 E}{\partial X^2} X \\ \hat{X} &= -\left(\frac{\partial^2 E}{\partial X^2}\right)^{-1} \frac{\partial E}{\partial X} \\ E(\hat{X}) &= E + \frac{1}{2} \frac{\partial E^T}{\partial X} X \end{aligned}$$

$X=(x,y,\sigma)^T$ 代表相位奇点的初始位置, $E(X)$ 代表的是滤波图像对应的 X 位置上的值, $\hat{X}=(x,y,\sigma)^T$ 表示修正后点的位置, 得到修正后位置滤波图像的值 $E(\hat{X})$, 当 $|E(\hat{X})|$ 小于给定的阈值时便认为该点不稳定, 去除该点。否则新的的位置 \hat{X} 既精确定位的结果。

文献[2]研究了相位奇点具有的性质, 相位奇点对于图像平移和旋转具有不变性。尽管小的噪声干扰或者图像失真会影响相位奇点的位置, 由于相位奇点本身特殊的拓扑稳定性, 小的图像扰动如噪声和变形通常不会使图像产生或消除新的相位奇点, 论文[2]中通过实验证明了相位奇点对实数噪声和图像变形是稳定的。

相位奇点数量的多少是由图像的内容和滤波器的尺度决定, 尺度越大点数越少, 尺度变换时同一PS

点的位置不会改变。相位奇点的尺度空间分析中可以看到随着尺度参数不断变大, 一些相位奇点会消失。将尺度空间中的相位奇点连成曲线, 称为相位奇点曲线, 如图2^[2]所示:

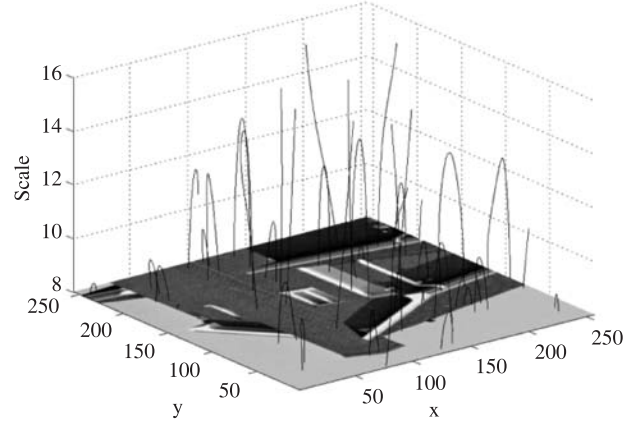


图2 相位奇点在尺度空间的分布

从图中可以看出, 大多数相位奇点曲线与底面垂直, 表明尺度变换时相位奇点的位置是稳定的。同时可以看出同一条相位奇点曲线在不同尺度下的相位奇点的位置是相关的, 可以通过对归一化导数进行最大化或最小化来选择关键的相位奇点。在两幅图像的特征匹配过程中, 通过找相位奇点曲线上的极点可以找到两幅图像中的关键相位奇点和关键尺度。

3 基于相位奇点包的图像分类方法

我们利用视觉词袋^[3]模型(Bag of Visual Word, 简称为BoVM)建立图像表示的相位奇点包模型, 并在此基础上开发图像分类方法。这一过程分为训练和测试两个过程, 具体描述如下:

- (1) 从给定的训练图像集合中抽取相位奇点, 并根据正负属性将相位奇点分为正负两个集合。
- (2) 计算相位奇点周边图像块的局部描述子, 这里我们采用SIFT特征作为局部描述子。
- (3) 对得到的正负相位奇点集的描述子分别进行 k -means聚类, 把聚类中心作为BoVW模型中的视觉单词。构建正负相位奇点包的视觉字典。
- (4) 对训练集中的每张图片, 检测相位奇点并计算描述子。将相位奇点的描述子依据正负视觉字典, 进行量化。统计图片中, 视觉单词出现的直方图。然后对直方图做L1归一化处理。归一化后的直方图作为图像的相位奇点包表示。由于我们有正负两个字典, 所以共有正负两个直方图表示。
- (5) 用训练集图片的相位奇点包表示和类别信

息,对每一类训练一个one-one的SVM分类器,既这一类的图像作为正样本,其他类的图像做负样本,训练SVM分类器。

(6)输入待分类的图像,根据步骤4所述的方法获得其正负相位奇点包表示,然后用每一类图像训练得到SVM分类器对这张图像进行分类,每一类图像训练得到的SVM会输出一个分类分数,我们取输出分

类分数最高的那一类SVM的类别作为待分类图像的类别。由于我们有正负两个相位奇点字典,因而获得相位奇点包图像表示也有两个。如图3所示,在实验中,我们可以将二者连成一个特征向量用于SVM训练,也可以分别对其训练SVM并将其进行融合。我们将在实验中比较这两个方法。

整个流程如图3所示。

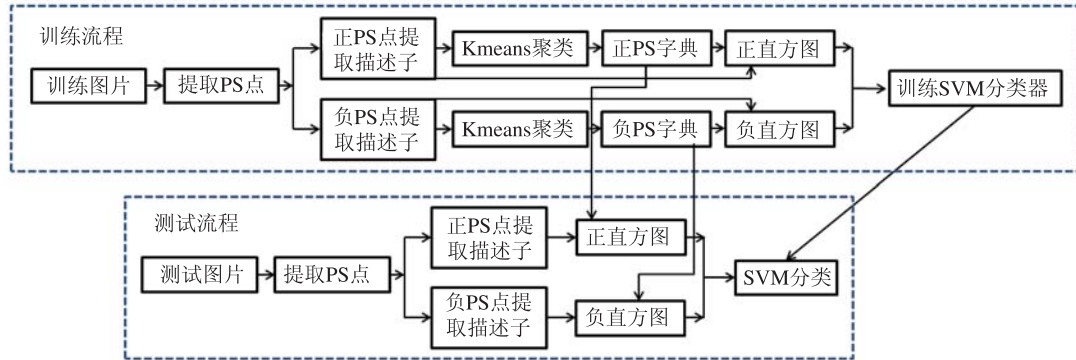


图3 本文使用的BOVW模型的训练与测试流程

4 实验

我们使用的是PASCAL2005¹的图像分类数据,PASCAL2005数据库共有4类图像,分成1个训练集和1个测试集,训练集有684张图像,测试集有689张图像,其中测试集与训练集中同类别的样本数量大约为1:1。如无特殊说明,实验二维复滤波器的母滤波器采用的是Laplacian of Gaussian (LoG)。

$$LoG(x, y, \sigma) = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

PASCAL 2005数据库采用的是AP(Average Precision)来衡量分类器的性能,AP的定义如下:

$$AP = \int_0^1 p(r) dr,$$

这里: $p(r)$ 即Precision被表示成Recall的函数,recall即查全率, Precision即查准率。即对每一个recall计算其对应的Precision,求其平均就能得到AP。

4.1 正负统计直方图的使用方法

对于每张图像的我们有正负2个统计直方图,我们采用了2种方法来组合这2个统计直方图进行识别,一类称为直方图组合(Histogram combine),另一类分数组合(Score combine)。

直方图组合是将正负相位奇点包表示(既正负直方图)拼接起来作为新的图像特征,然后利用新特征训练SVM;分数组合是分别对正负2个统计直方图训

练出不同的SVM分类器,将2个SVM训练分类器的分数加起来作为该图像的总分。流程如图4所示:

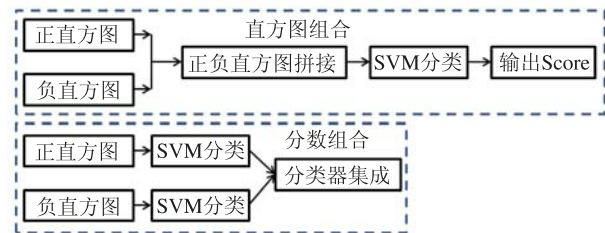


图4 直方图组合和分数组合SVM分类流程

我们选取了4000作为视觉单词的数量,既正负各2000个视觉单词。实验结果如表1,从中可以看出这两类方法的结果相近。分数组合的方法,略好于直方图组合。

表1 PASCAL 2005, 分别采用Hist Combine和Score Combine组合正负统计直方图的实验结果

类别	AP	
	分数组合	直方图组合
Motorbike	69.45	69.1
Bicycle	83.48	83.6
People	87.78	87.66
Cars	63.28	63.1
Average	76	75.88

4.2 SIFT检测子与相位奇点检测子的比较

在第二个实验中。我们比较相位奇点检测子与SIFT检测子。公平起见,两种方法我们选取4000作

¹ <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2005/index.html>

为视觉单词的数量。其中因为相位奇点分为正负两大类, 我们对每一类各使用2000个视觉单词。因为实验的目的是为了比较兴趣点检测子, 所以我们对两种兴趣点都采用SIFT的描述子。

实验结果如图5:

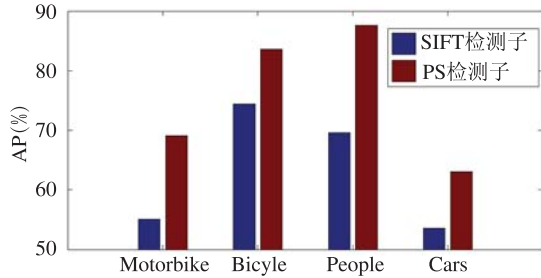


图5 PASCAL 2005, 分别采用PS和SIFT检测子的实验结果

由图5中的结果可知, PS检测子相对于SIFT检测子在每一类上都拥有更好的分类性能。PS检测子相对于SIFT检测子能检测到更多稳定的兴趣点, 由PS检测子得到的统计直方图表达能力比SIFT检测子的统计直方图表达能力更强。

4.3 视觉词典大小变化对分类的影响

在第三个实验中, 我们测试视觉词典大小对分类结果的影响。我们分别选取了4000, 6000, 8000进行实验。实验结果如图6:

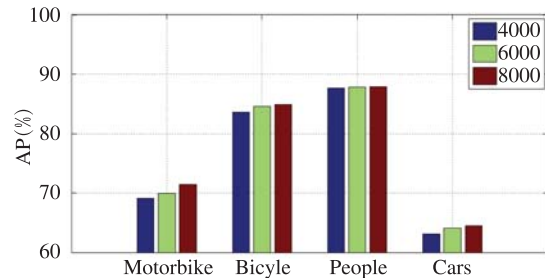


图6 PASCAL 2005, 采用不同大小Vision Word数量的PS检测子的实验结果

由图6的结果可知AP随着视觉单词个数的增加而增加, 但是对于特定类别如People增加有限。

4.4 采用不同母滤波器对分类的影响

在第四个实验中, 我们分别选取了LOG滤波器和Gaussian滤波器作为PS检测子的母滤波器, 并比较二者的结果。视觉词典的大小设为4000。

实验结果如表2。

由表2可知, LoG(Laplacian of Gaussian)滤波器的效果好于Gaussian滤波器。LoG滤波器对因的滤波图像的值是图像中边缘强度的表现, 而Gaussian滤波器只是起到一个低通滤波器的作用。图像的边缘一般是不容易变化的部分, 有较好的鲁棒性, 且边缘周围富

表2 PASCAL 2005, 分别采用LOG和Gaussian作为母滤波器的实验结果

类别	AP	
	LOG	Gaussian
Motorbike	69.1	42.17
Bicycle	83.6	67.52
People	87.66	58.96
Cars	63.1	41.33
Average	75.88	52.5

含图像纹理与细节变化。这些因素部分解释了LOG滤波器为何表现由于Gaussian滤波器。

5 结论和未来的工作

本文提出了一种基于相位奇点包的图像表示和分类方法。我们利用视觉词典模型(BoVW)分别对正负相位奇点进行建模, 构建正负相位奇点字典已经相应图像表示。并利用SVM对相位奇点包模型进行分类。通过对PASCAL VOC 2005数据集的实验表明, PS检测子相对与SIFT检测子有着更好的性能。在将来我们会在更大的数据集上测试我们的方法, 并加入更多的新的编码方式。

参考文献

- [1] David G L. Object recognition from local scale-invariant features [C] // International Conference on Computer Vision. Corfu, Greece, 1999: 1150-1157.
- [2] Qiao Y, Wang W, Minematsu N, et al. A theory of phase singularities for image representation and its applications to object tracking and image matching [J]. IEEE Transactions on Image Processing, 2009, 18(10): 2153-2166.
- [3] Li F F, Perona P. A bayesian heirarchical model for learning natural scene categories [C] // Computer Vision and Pattern Recognition, 2005, 2: 524-531.
- [4] Smeulders A W M, Worring M, Santini, S, et al. Content-based image retrieval at the end of the early years [J]. Pattern Analysis and Machine Intelligence, 2000, 22(10): 1349-1380.
- [5] Wang W, Yokozeki T, Ishijima R, et al. Optical vortex metrology based on the core structures of PS in LG trans-form of a speckle pattern [J]. Optics Express, 2006, 14(22): 10195-10206.
- [6] Wang W, Yokozeki T, Ishijima R, et al. Optical vortex metrology for nano metricspeckle displacement measurement [J]. Optics Express, 2006, 14(1): 120-127.
- [7] Chai D, Bouzerdoum A. A Bayesian approach to skin color classification in YCbCr color space [C] // TENCON, 2000, 2: 421-424.