

山东大学

# 毕业论文 (设计)

论文题目： 基于纹理特征的遥感图像  
分类方法研究

学 院	控制科学与工程学院
专 业	自动化
年 级	2012 级
姓 名	XXX
学 号	201200xxxxxx
指导教师	XXX

2016 年 10 月 15 日



## 摘 要

这里是中文摘要。格式要求：(1) 居中打印“摘要”二字（三号黑体），二字之间空一格。(2) “摘要”二字下空一行打印摘要内容（小四号宋体），摘要内容每段开头缩进两个字。切忌将应在引言中出现的内容（如研究背景等）写入摘要，一般也不要对论文内容作诠释和评论（尤其是自我评价）。摘要中尽量少用特殊字符以及由特殊字符组成的数学表达式。

**关键词：**摘要内容下空一行，顶格位置打印“关键词”三字（小四号黑体），其后为关键词（小四号宋体）。每一关键词之间用逗号隔开，最后一个关键词后不打标点符号。关键字总数在 3-7 个为宜。



## ABSTRACT

This is the English Abstract. The standard format: (1) write "ABSTRACT" with the size of 16pt (three) and center it horizontally, skip two lines and type the abstract content with the font Times New Roman (12pt); (2) Leave four blank at the begin of each paragraph.

**Keywords:** Skip one line down and write "KEYWORDS", then put the keywords, separated by commas (no punctuation at the end). the number of keywords varies from 3 to 7.



# 目 录

摘 要 .....	I
ABSTRACT .....	II
第一章 关于 L <sup>A</sup> T <sub>E</sub> X .....	1
1.1 简介 .....	1
1.2 释义 .....	1
1.3 历史 .....	2
1.3.1 T <sub>E</sub> X 格式 .....	2
1.3.2 PlainT <sub>E</sub> X .....	2
1.3.3 L <sup>A</sup> T <sub>E</sub> X .....	2
1.3.4 LaTeX2e .....	3
1.3.5 L <sup>A</sup> T <sub>E</sub> X 各版本关系 .....	3
第二章 如何使用本模板 .....	4
2.1 关于本模板 .....	4
2.2 基本输入示例 .....	4
2.3 数学输入示例 .....	4
2.4 图片与表格插入 .....	5
2.4.1 插入表格 .....	5
2.4.2 插入图片 .....	5
第三章 其他内容 .....	7
3.1 如何获得 eps 格式 .....	7
3.2 如何引用参考文献 .....	8
3.3 使用算法环境 .....	8
结 论 .....	10
致 谢 .....	11
参考文献 .....	12
外文资料	
中文译文	

源代码



# 第一章 关于 L<sup>A</sup>T<sub>E</sub>X

## 1.1 简介

L<sup>A</sup>T<sub>E</sub>X<sup>1</sup>（音译“拉泰赫”）是一种基于 T<sub>E</sub>X 的排版系统，由美国计算机学家莱斯利·兰伯特（Leslie Lamport）在 20 世纪 80 年代初期开发，利用这种格式，即使使用者没有排版和程序设计的知识也可以充分发挥由 T<sub>E</sub>X 所提供的强大功能，能在几天，甚至几小时内生成很多具有书籍质量的印刷品。对于生成复杂表格和数学公式，这一点表现得尤为突出。因此它非常适用于生成高印刷质量的科技和数学类文档。这个系统同样适用于生成从简单的信件到完整书籍的所有其他种类的文档。

## 1.2 释义

【正式名称】：L<sup>A</sup>T<sub>E</sub>X

【纯文本名称】：LaTeX

【概述】：L<sup>A</sup>T<sub>E</sub>X 使用 T<sub>E</sub>X 作为它的格式化引擎，当前的版本是 LaTeX2<sub>ε</sub>。Leslie Lamport 开发的 L<sup>A</sup>T<sub>E</sub>X 是当今世界上最流行和使用最为广泛的 T<sub>E</sub>X 宏集。它构筑在 PlainT<sub>E</sub>X 的基础之上，并加进了很多的功能以使得使用者可以更为方便的利用 T<sub>E</sub>X 的强大功能。使用 LaTeX 基本上不需要使用者自己设计命令和宏等，因为 L<sup>A</sup>T<sub>E</sub>X 已经替你做好了。因此，即使使用者并不是很了解 T<sub>E</sub>X，也可以在短短的时间内生成高质量的文档。对于生成复杂的数学公式，L<sup>A</sup>T<sub>E</sub>X 表现的更为出色。L<sup>A</sup>T<sub>E</sub>X 自从八十年代初问世以来，也在不断的发展。最初的正式版本为 2.09，在经过几年的发展之后，许多新的功能，机制被引入到 L<sup>A</sup>T<sub>E</sub>X 中。在享受这些新功能带来的便利的同时，它所伴随的副作用也开始显现，这就是不兼容性。标准的 L<sup>A</sup>T<sub>E</sub>X 2.09 引入了“新字体选择框架” (NFSS) 的 L<sup>A</sup>T<sub>E</sub>X、SLiTeX，AMS-LaTeX 等等，相互之间并不兼容。这给使用者和维护者都带来很大的麻烦。为结束这种糟糕的状况，Frank、Mittel、bach 等人成立了 ATeX3 项目小组，目标是建立一个最优的，有效的，统一的，标准的命令集合。即得到 L<sup>A</sup>T<sub>E</sub>X 的一个新版本 3。这是一个长期目标，向这个目标迈出第一步就是在 1994 年发布的 LaTeX2<sub>ε</sub>。LaTeX2<sub>ε</sub> 采用了 NFSS 作为标准，加入了很多新的功能，同时还兼容旧 L<sup>A</sup>T<sub>E</sub>X 2.09。LaTeX2<sub>ε</sub> 每 6 个月更新一次，修正发现的错误并加入前，LaTeX2<sub>ε</sub> 将是标准的。

<sup>1</sup>以下介绍均来自百度百科

## 1.3 历史

### 1.3.1 $\text{T}_{\text{E}}\text{X}$ 格式

最基本的  $\text{T}_{\text{E}}\text{X}$  程序只是由一些很原始的命令组成，它们可以完成简单的排版操作和程序设计功能。然而， $\text{T}_{\text{E}}\text{X}$  也允许用这些原始命令定义一些更复杂的高级命令。这样就可以利用低级的块结构，形成一个用户界面相当友好的环境。

在处理器运行期间，该程序首先读取所谓的格式文件，其中包含各种以原始语言写成的高级命令，也包含分割单词的连字号安排模式。接着处理程序就处理源文件，其中包含要处理的真正文本，以及在格式文件中已定义了的格式命令。

创建新格式是一件需要由具有丰富知识的程序员来做的事情。把定义写到一个源文件中，这个文件接着被一个名叫  $\text{iniT}_{\text{E}}\text{X}$  的特殊版本的  $\text{T}_{\text{E}}\text{X}$  程序处理。它采用一种紧凑的方式存贮这些新格式，这样就可以被通常  $\text{T}_{\text{E}}\text{X}$  程序很快地读取。

### 1.3.2 Plain $\text{T}_{\text{E}}\text{X}$

Knuth 设计了一个名叫 Plain $\text{T}_{\text{E}}\text{X}$  的基本格式，以与低层次的原始  $\text{T}_{\text{E}}\text{X}$  呼应。这种格式是用  $\text{T}_{\text{E}}\text{X}$  处理文本时相当基本的部分，以致于我们有时都分不清到底哪条指令是真正的处理程序  $\text{T}_{\text{E}}\text{X}$  的原始命令，哪条是 Plain $\text{T}_{\text{E}}\text{X}$  格式的。大多数声称只使用  $\text{T}_{\text{E}}\text{X}$  的人，实际上指的是只用 Plain $\text{T}_{\text{E}}\text{X}$ 。

Plain $\text{T}_{\text{E}}\text{X}$  也是其它格式的基础，这进一步加深了很多人都认为  $\text{T}_{\text{E}}\text{X}$  和 Plain $\text{T}_{\text{E}}\text{X}$  是同一事物的印象。

Plain $\text{T}_{\text{E}}\text{X}$  的重点还只是在于如何排版的层次上，而不是从一位作者的观点出发。对它的深层功能的进一步发掘，需要相当丰富的编程技巧。因此它的应用就局限于高级排版和程序设计人员。

### 1.3.3 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$

Leslie Lamport 开发的  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  是当今世界上最流行和使用最为广泛的  $\text{T}_{\text{E}}\text{X}$  格式。它构筑在 Plain $\text{T}_{\text{E}}\text{X}$  的基础之上，并加进了很多的功能以使得使用者可以更为方便的利用  $\text{T}_{\text{E}}\text{X}$  的强大功能。使用  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  基本上不需要使用者自己设计命令和宏等，因为  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  已经替你做好了。因此，即使使用者并不是很了解  $\text{T}_{\text{E}}\text{X}$ ，也可以在短短的时间内生成高质量的文档。对于生成复杂的数学公式， $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  表现的更为出色。

$\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  自从二十世纪八十年代初问世以来，也在不断的发展。最初的正式版本为 2.09，在经过几年的发展之后，许多新的功能，机制被引入到  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  中。在享受这些新功能带来的便利的同时，它所伴随的副作用也开始显现，这就是不兼容性。标准的  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$  2.09，引入了“新字体选择框架”(NFSS)的  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ， $\text{SLiT}_{\text{E}}\text{X}$ ， $\text{AMSLaT}_{\text{E}}\text{X}$  等等，相互之间并不兼容。这给使用者和维护者都带来很大的麻烦。

#### 1.3.4 LaTeX2e

为结束这种糟糕的状况，Frank Mittelbach 等人成立了 L<sup>A</sup>T<sub>E</sub>X 3 项目小组，目标是建立一个最优的，有效的，统一的，标准的命令集合。即得到 L<sup>A</sup>T<sub>E</sub>X 的一个新版本 3。这是一个长期目标，向这个目标迈出第一步就是在 1994 年发布的 LaTeX2e。LaTeX2e 采用了 NFSS 作为标准，加入了很多新的功能，同时还兼容旧的 L<sup>A</sup>T<sub>E</sub>X 2.09。LaTeX2e 每 6 个月更新一次，修正发现的错误并加入一些新的功能。在 L<sup>A</sup>T<sub>E</sub>X 3 最终完成之前，LaTeX2e 将是标准的 L<sup>A</sup>T<sub>E</sub>X 版本。

#### 1.3.5 L<sup>A</sup>T<sub>E</sub>X 各版本关系

MiKTeX、fpTeX、teTeX、CTeX 是什么关系？T<sub>E</sub>X 在不同的硬件和操作系统上有不同的实现版本。这就像 C 语言，在不同的操作系统中有不同的编译系统，例如 Linux 下的 gcc，Windows 下的 Visual C++ 等。有时，一种操作系统里也会有好几种的 T<sub>E</sub>X 系统。目前常见的 Unix/Linux 下的 T<sub>E</sub>X 系统是 TeXlive，Windows 下则有 MiKTeX 和 fpTeX。CTeX 指的是 CTeX 中文套装的简称，是把 MiKTeX 和一些常用的相关工具，如 GSview，WinEdt 等包装在一起制作的一个简易安装程序，并对其中的中文支持部分进行了配置，使得安装后马上就可以使用中文。

## 第二章 如何使用本模板

### 2.1 关于本模板

本模板<sup>1</sup>是在天津大学的 L<sup>A</sup>T<sub>E</sub>X 毕业论文模板<sup>2</sup>的基础上修改完成的，其格式基本符合山东大学控制科学与工程学院学士学位论文的格式要求，但本人并不能保证完全符合，只是希望能通过自己的一点努力帮助大家学会这款出色的排版系统的基本使用。到目前为止该模板应该是已开源的唯一一个针对山大控院本科生定制的 L<sup>A</sup>T<sub>E</sub>X 模板，该模板与其他多数模板相比优点在于完全废弃了 CJK 宏包，在最新的 TexLive2015 (Windows7) 套装下使用自带的编辑器 TeXworks 能顺利编译通过。编译时需要执行四次编译，通过 xelatex+bibtex+xelatex+xelatex 就可以生成带有完整目录和参考文献信息的 PDF 文档。编译时也可以使用命令行切换到当前目录下，然后执行 Python 脚本 compile.py 或者 Windows 批处理文件 compile.bat 来一次性完成，该脚本可以实现自动对过程文件的清理以及自动重复编译。关于 LaTeX 的基本入门可以参考这篇文章[始终-一份其实很短的 LaTeX 入门文档](#)。

### 2.2 基本输入示例

文章的结构分为三个层次，分别是 chapter、section、subsection，当你想要开始一个新的章节时，首先要新建一个文件保存在 body 文件夹下，并在 sdumain.tex 文件中通过 \include{body/文件名} 的方式插入在合适的位置。一个新的章节应该以 \chapter{章节名} 开头，同样 section、subsection 也是如此。之后就可以进行正常的文本输入工作了，要注意换行 (\\ 或 \newline) 和重起一段 (两个空行) 的区别，在文本中输入的一个空行会被当做换行而不是换段处理。如果你需要修改封面内容，需要到 preface 文件夹下 cover.tex 文件里进行修改，如果需要修改论文题目可能就麻烦些，需要在 setup 文件夹下 format.tex 文件中大约第 232 行修改，并通过一些排版以适合自己的标题长度。

### 2.3 数学输入示例

行内公式使用 \(\text{公式代码}\) 或者 \$\text{公式代码}\$ 来输入。这是一个行内公式  $\lim_{x \rightarrow \infty} \frac{x}{\sin x}$ ，行内公式能够紧凑的插入到文字中间，对行间距不产生影响。行间公式使用 \[\text{公式代码}\] 或者 \$\$\text{公式代码}\$\$ 来输入。这是一个行间公式

$$\lim_{x \rightarrow \infty} \frac{x}{\sin x}$$

行间公式单独占一行。

<sup>1</sup>本模板 Git 仓库

<sup>2</sup>原始模板 Git 仓库

使用数学环境 `\begin{equation}` 公式代码 `\end{equation}` 可以输入带有编号的行间公式

$$\arg \max_{\theta_0, \theta_1, \theta_2, \sigma} \ln L = \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{dg_i - (\theta_0 + \theta_1 v(x_i) + \theta_2 s(x_i))}{2\sigma^2}} \right) \quad (2-1)$$

数学环境还可以实现对公式的引用，公式 (2-1) 就是一个例子。

要注意的是，所有在数学环境包括行内行间公式里的字符都将被视为变量名来根据数学逻辑进行排版，空格空行将被忽略。如果需要输入普通文本应该使用 `\text{rm}{..}` 来输入，比如 *dg* and **dg**，空格则用 `\quad`、`\qquad`、`\,` 来输入。数学公式里的黑体用 `\mathbf{..}` 输入，空心粗体用 `\mathbb{..}` 输入，比如 **R** and  $\mathbb{R}$ 。至于数学公式可以用 MathType 工具排好之后直接复制过来，但是要先修改预置菜单里的剪切与复制预置选项以支持 L<sup>A</sup>T<sub>E</sub>X 的语法。

## 2.4 图片与表格插入

### 2.4.1 插入表格

如2-1所示，要插入这样一个表格。首先要开始一个表格环境 `\begin{table}[htbp]`，参数含义为 h(这里),t(顶部),b(底部),p(浮动体)，设置标题 `\caption{BMI 指数分类详表}` 和字体。开始一个表格 `\begin{tabular}{c|cccc}`，参数 c 表示居中显示对应列，参数 | 表示对应列之间用竖线分隔，真正的表格部分以 `\toprule` 开始，输入列名后以 `\midrule` 分割，接下来是数据部分，各列之间用 `$` 分隔，各行之间用 `\\` 分隔，如果行与行直间需要用横线分隔使用 `\hline` 命令，最后一个表格以 `\bottomrule` 结束。

表 2-1 BMI 指数分类详表

BMI 分类	WHO 标准	亚洲标准	中国参考标准	相关疾病发病的危险性
体重过低	<18.5	<18.5	<18.5	低（但其它疾病危险性增加）
正常范围	18.5 ~ 24.9	18.5 ~ 22.9	18.5 ~ 23.9	平均水平
超重	25	23	24	增加
肥胖前期	25.0 ~ 29.9	23 ~ 24.9	24 ~ 26.9	增加
I 度肥胖	30.0 ~ 34.9	25 ~ 29.9	27 ~ 29.9	中度增加
II 度肥胖	35.0 ~ 39.9	30	30	严重增加
度肥胖	40.0	40.0	40.0	非常严重增加

### 2.4.2 插入图片

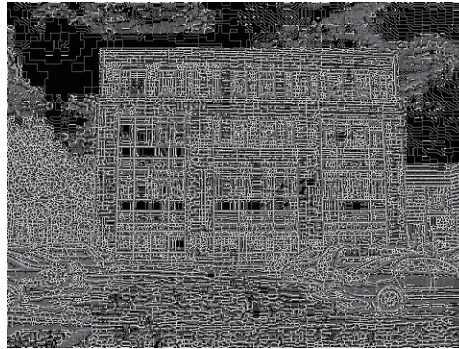
插入图片的方式和插入表格差不多，这里以在同一行插入一对图如图2-1来示范一下。代码如下：

`\begin{figure}[htbp]`

```
\centering
\subfigure[处理前]{
    \label{fig:subfig1:subsubfig1}
    \includegraphics[width=0.4\textwidth]{figures/beforedeal}}
\subfigure[处理后]{
    \label{fig:subfig1:subsubfig2}
    \includegraphics[width=0.4\textwidth]{figures/afterdeal}}
\caption{处理效果}\label{fig:subfig1}
\vspace{\baselineskip}
\end{figure}
```



a) 处理前



b) 处理后

图 2-1 处理效果

## 第三章 其他内容

### 3.1 如何获得 eps 格式

在文档中插入一张图片首先需要将图片转换为 eps 格式才能正确的显示，在这里推荐使用 Adobe Acrobat Pro 来进行转换。首先，假定你有了一张 \*.jpg 格式或者 \*.png 格式等的图片，正确安装 Adobe Acrobat Pro 后，在图片右键选择转换为转换为 Adobe PDF，转换完后使用文件另存为命令进行保存，保存类型选择内嵌式 PostScript(\*.eps)，点击下方的设置按钮按照图3-1进行设置即可，最后要删除掉 figures 文件夹下除 eps 以外的其他格式文件就可以在文档中显示图片了。

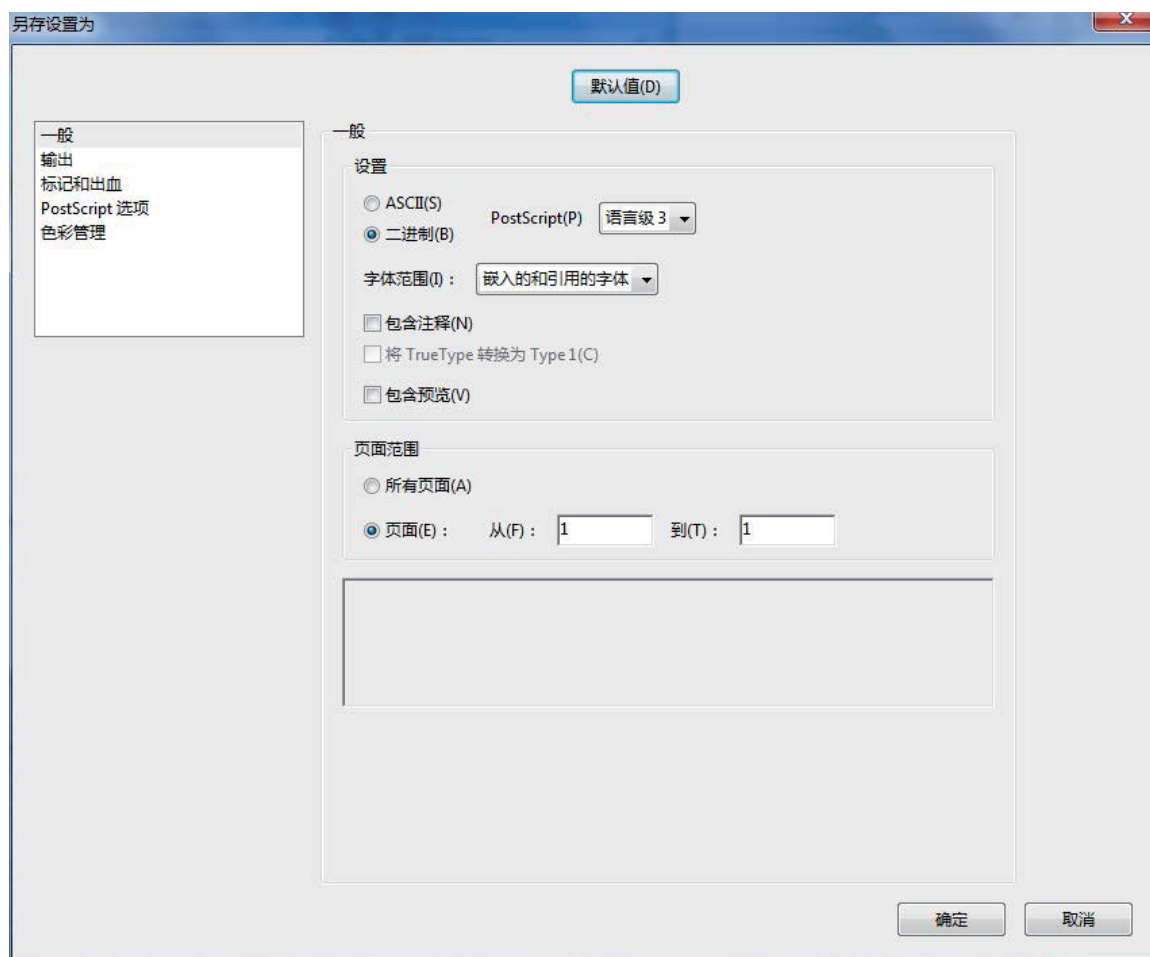


图 3-1 设置示意图

## 3.2 如何引用参考文献

由于一些原因，目前中国大陆无法使用 Google Scholar，但是还好我们还有 bing 学术，bing 学术同样有兼容 BibTeX 的参考文献信息。使用时也相当简单，如图3-2所示，只要先搜到你想要引用的论文，然后在图3-2-a 中点击下方的引用出现图 b，选择 BibTeX 即可出现引用信息如图3-3，将这几行文字复制到 references 文件夹下的 reference.bib 文件里就可以在文档中进行引用了，当然有时也需要少量修改才行，引用时使用命令 \cite{引用信息第一行的内容（必须是英文）}，如 \cite{zhu2015a}<sup>[1]</sup>。



图 3-2 获取参考文献信息

```
@article{zhu2015a,
title={A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior},
author={Qingsong Zhu,Jiaming Mai,Ling Shao},
journal={IEEE Transactions on Image Processing},
volume={24},
number={11},
pages={3522-3533},
year={2015}}
```

图 3-3 待修改的引用信息

## 3.3 使用算法环境

控制学院的同学做毕业设计不少遇到写代码的情况，在论文中阐述自己的算法实现也是常有的事，一些宏包也专门给我们提供了一个专门用来放置伪代码片段的一个环境。示例如下：



---

**Algorithm 1:** identifyRowContext

---

**Input:**  $r_i$ ,  $Backgrd(T_i)=T_1, T_2, \dots, T_n$  and similarity threshold  $\theta_r$

**Output:**  $con(r_i)$

$con(r_i) = \Phi$ ;

**for**  $j = 1; j \leq n; j \neq i$  **do**

    float  $maxSim = 0$ ;

$r^{maxSim} = null$ ;

**while** not end of  $T_j$  **do**

        compute  $Jaro(r_i, r_m)(r_m \in T_j)$ ;

**if**  $(Jaro(r_i, r_m) \geq \theta_r) \wedge ((Jaro(r_i, r_m) \geq r^{maxSim})$  **then**

            replace  $r^{maxSim}$  with  $r_m$ ;

**end**

**end**

$con(r_i) = con(r_i) \cup r^{maxSim}$ ;

**end**

return  $con(r_i)$ ;

---

## 结 论

山东大学(Shandong University)简称山大(SDU),是中华人民共和国教育部直属的综合性全国重点大学,是国家“211 工程”、“985 工程”重点建设院校。属于“111 计划”、“珠峰计划”、“2011 计划”、“卓越工程师教育培养计划”、“卓越法律人才教育培养计划”、第一批“地方立法研究服务基地”高等院校,也是重点建设的高水平大学之一。

山大创建于清光绪 27 年(公元 1901 年),初名山东大学堂,一百多年来,这所誉满海内外的百年名校历经山东大学堂、国立青岛大学、国立山东大学、山东大学等历史发展时期,迁徙分合、春华秋实,成为中国现代大学教育的重要发祥地。学校总占地面积 8000 余亩(含青岛校区约 3000 亩),形成了一校三地(济南、青岛、威海)八个校园的办学格局。截至 2013 年,学校有各类全日制学生达 6 万人,其中,全日制本科生 41437 人,研究生 16034 人,留学生 1737 人。

## 致 谢

值此论文完成之际，谨在此向多年来给予我关心和帮助的老师、同学、朋友和家人表示衷心的感谢！

同时也感谢众多的开源项目以及无数为开源而做出贡献的人们，谢谢！

.....

谨把本文献给我最敬爱的老师！

## 参考文献

- [1] Qingsong Z, Jiaming M, Ling S. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior [J]. IEEE Transactions on Image Processing, 2015, 24 (11): 3522–3533.
- [2] 樊冬进, 孙冰, 封举富, 等. 基于方差及方差梯度的指纹图像自适应分割算法 [J], 2008.
- [3] GUO W-j, YANG G-p, DONG J-l. A review of fingerprint image segmentation methods [J]. Journal of Shandong University (Natural Science), 2010, 7: 019.
- [4] 温苗利, 梁彦, 潘泉, *et al.* 基于 Gab0r 函数的小波域指纹图像增强算法 [J]. 计算机应用, 2006.
- [5] Dietterich T G. Machine-learning research [J]. AI magazine, 1997, 18 (4): 97.
- [6] Tucker A B. Computer science handbook [M]. CRC press, 2004.
- [7] Zhang M-L, Zhou Z-H. Exploiting unlabeled data to enhance ensemble diversity [J]. Data Mining and Knowledge Discovery, 2013, 26 (1): 98–129.
- [8] Zhang M, Zhou Z. A review on multi-label learning algorithms [J], 2013.
- [9] Cour T, Sapp B, Taskar B. Learning from partial labels [J]. The Journal of Machine Learning Research, 2011, 12: 1501–1536.
- [10] Hinton G, Osindero S, Welling M, *et al.* Unsupervised discovery of nonlinear structure using contrastive backpropagation [J]. Cognitive Science, 2006, 30 (4): 725–731.
- [11] Gonzalez R C, Woods R E. Digital image processing, 2nd [J]. SL: Prentice Hall, 2002.
- [12] Mehtre B M, Chatterjee B. Segmentation of fingerprint images—a composite method [J]. Pattern Recognition, 1989, 22 (4): 381–385.
- [13] Bazen A M, Gerez S H. Segmentation of fingerprint images [C]. In Proc. Workshop on Circuits Systems and Signal Processing (ProRISC 2001), 2001: 276–280.
- [14] 周明辉. 运动车辆车牌识别算法的研究与实现 [D]. [S. l.]: 昆明理工大学, 2013.
- [15] Duda R O, Hart P E, Stork D G. Pattern classification [M]. John Wiley & Sons, 2012.
- [16] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天, 今天和明天 [J]. 计算机研究与发展, 2013, 50 (9): 1799–1804.

## 外文资料

### The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features

Kristen Grauman and Trevor Darrell  
Massachusetts Institute of Technology  
Computer Science and Artificial Intelligence Laboratory  
Cambridge, MA, USA

#### Abstract

Discriminative learning is challenging when examples are sets of features, and the sets vary in cardinality and lack any sort of meaningful ordering. Kernel-based classification methods can learn complex decision boundaries, but a kernel over unordered set inputs must somehow solve for correspondences – generally a computationally expensive task that becomes impractical for large set sizes. We present a new fast kernel function which maps unordered feature sets to multi-resolution histograms and computes a weighted histogram intersection in this space. This “pyramid match” computation is linear in the number of features, and it implicitly finds correspondences based on the finest resolution histogram cell where a matched pair first appears. Since the kernel does not penalize the presence of extra features, it is robust to clutter. We show the kernel function is positive-definite, making it valid for use in learning algorithms whose optimal solutions are guaranteed only for Mercer kernels. We demonstrate our algorithm on object recognition tasks and show it to be accurate and dramatically faster than current approaches.

#### 1. Introduction

A variety of representations used in computer vision consist of unordered sets of features or parts, where each set varies in cardinality, and the correspondence between the features across each set is unknown. For instance, an image may be described by a set of detected local affine-invariant regions, a shape may be described by a set of local descriptors defined at each edge point, or a person’s face may be represented by a set of patches with different facial parts. In such cases, one set of feature vectors denotes a single instance of a particular class of interest (an object, scene, shape, face, etc.), and it is expected that the number of features will vary across examples due to viewpoint changes, occlusions, or inconsistent detections by the interest operator.

To perform learning tasks like categorization or recognition with such representations is challenging. While generative methods have had some success, kernel-based dis-

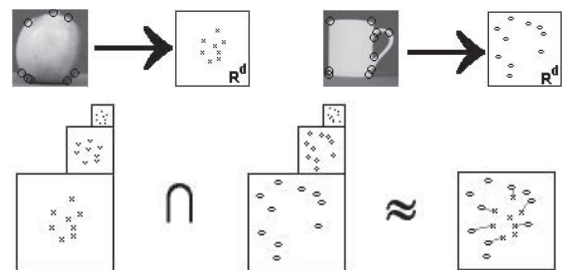


Figure 1: The pyramid match kernel intersects histogram pyramids formed over local features, approximating the optimal correspondences between the sets’ features.

criminative methods are known to represent complex decision boundaries very efficiently and generalize well to unseen data [24, 21]. For example, the Support Vector Machine (SVM) is a widely used approach to discriminative classification that finds the optimal separating hyperplane between two classes. Kernel functions, which measure similarity between inputs, introduce non-linearities to the decision functions; the kernel non-linearly maps two examples from the input space to the inner product in some feature space. However, conventional kernel-based algorithms are designed to operate on fixed-length vector inputs, where each vector entry corresponds to a particular global attribute for that instance; the commonly used general-purpose kernels defined on  $R^n$  inputs (e.g., Gaussian RBF, polynomial) are not applicable in the space of vector sets.

In this work we propose a *pyramid match kernel* – a new kernel function over unordered feature sets that allows them to be used effectively and efficiently in kernel-based learning methods. Each feature set is mapped to a multi-resolution histogram that preserves the individual features’ distinctness at the finest level. The histogram pyramids are then compared using a weighted histogram intersection computation, which we show defines an implicit correspondence based on the finest resolution histogram cell where a matched pair first appears (see Figure 1).

The similarity measured by the pyramid match approximates the similarity measured by the optimal correspon-

dences between feature sets of unequal cardinality (i.e., the *partial matching* that optimally maps points in the lower cardinality set to some subset of the points in the larger set, such that the summed similarities between matched points is maximal). Our kernel is extremely efficient and can be computed in time that is linear in the sets' cardinality. We show that our kernel function is positive-definite, meaning that it is appropriate to use with learning methods that guarantee convergence to a unique optimum only for positive-definite kernels (e.g., SVMs).

Because it does not penalize the presence of superfluous data points, the proposed kernel is robust to clutter. As we will show, this translates into the ability to handle unsegmented images with varying backgrounds or occlusions. The kernel also respects the co-occurrence relations inherent in the input sets: rather than matching features in a set individually, ignoring potential dependencies conveyed by features within one set, our similarity measure captures the features' joint statistics.

Other approaches to this problem have recently been proposed [25, 14, 3, 12, 27, 16, 20], but unfortunately each of these techniques suffers from some number of the following drawbacks: computational complexities that make large feature set sizes infeasible; limitations to parametric distributions which may not adequately describe the data; kernels that are not positive-definite (do not guarantee unique solutions for an SVM); limitations to sets of equal size; and failure to account for dependencies within feature sets.

Our method addresses all of these issues, resulting in a kernel appropriate for comparing unordered, variable-length feature sets within any existing kernel-based learning paradigm. We demonstrate our algorithm with object recognition tasks and show that its accuracy is comparable to current approaches, while requiring significantly less computation time.

## 2. Related Work

In this section, we review related work on discriminative classification with sets of features, using kernels and SVMs for recognition, and multi-resolution image representations.

Object recognition is a challenging problem that requires strong generalization ability from a classifier in order to cope with the broad variety in illumination, viewpoint, occlusions, clutter, intra-class appearances, and deformations that images of the same object or object class will exhibit. While researchers have shown promising results applying SVMs to object recognition, they have generally used global image features – ordered features of equal length measured from the image as a whole, such as color or grayscale histograms or vectors of raw pixel data [5, 18, 17]. Such global representations are known to be sensitive to real-world imaging conditions, such as occlusions, pose

Method	Complexity	C	P	M	U
Match [25]	$O(dm^2)$			x	x
Exponent [14]	$O(dm^2)$		x	x	x
Greedy [3]	$O(dm^2)$	x		x	x
Princ. ang. [27]	$O(dm^3)$	x	x		
Bhattach.'s [12]	$O(dm^3)$	x	x		x
KL-div. [16]	$O(dm^2)$	x			x
Pyramid	$O(dm \log D)$	x	x	x	x

Table 1: Comparing kernel approaches to matching unordered sets. Columns show each method's computational cost and whether its kernel captures co-occurrences (C), is positive-definite (P), does not assume a parametric model (M), and can handle sets of unequal cardinality (U).  $d$  is vector dimension,  $m$  is maximum set cardinality, and  $D$  is diameter of vector space. "Pyramid" refers to the proposed kernel.

changes, or image noise.

Recent work has shown that local features invariant to common image transformations (e.g., SIFT [13]) are a powerful representation for recognition, because the features can be reliably detected and matched across instances of the same object or scene under different viewpoints, poses, or lighting conditions. Most approaches, however, perform recognition with local feature representations using nearest-neighbor (e.g., [1, 8, 22, 2]) or voting-based classifiers followed by an alignment step (e.g., [13, 15]); both may be impractical for large training sets, since their classification times increase with the number of training examples. An SVM, on the other hand, identifies a sparse subset of the training examples (the support vectors) to delineate a decision boundary.

Kernel-based learning algorithms, which include SVMs, kernel PCA, or Gaussian Processes, have become well-established tools that are useful in a variety of contexts, including discriminative classification, regression, density estimation, and clustering [21]. More recently, attention has been focused on developing specialized kernels that can more fully leverage these tools for situations where the data cannot be naturally represented by a Euclidean vector space, such as graphs, strings, or trees.

Several researchers have designed similarity measures that operate on sets of unordered features. See Table 1 for a concise comparison of the approaches. The authors of [25] propose a kernel that averages over the similarities of the best matching feature found for each feature member within the other set. The use of the "max" operator in this kernel makes it non-Mercer (i.e., not positive-definite – see Section 3), and thus it lacks convergence guarantees when used in an SVM. A similar kernel is given in [14], which also considers all possible feature matchings but raises the similarity between each pair of features to a given power. Both [25] and [14] have a computational complexity that is quadratic in the number of features. Furthermore, both

## 中文译文

本文<sup>1</sup>提出一种基于局部特征的双空间金字塔匹配核 (bi-space pyramid match kernel, BSPM) 用于图像目标分类。利用局部特征在特征空间和图像空间建立统一的多分辨率框架, 以便较好地表达图像的语义内容。该方法同时在特征空间和图像空间建立金字塔型结构, 通过适当匹配可以得到正定核函数, 该函数具有线性计算复杂度, 可以运用于基于核的学习算法。将 BSPM 嵌入支持向量机对公共数据库中图像目标进行分类, 实验结果表明该方法对图像具有良好的分类能力, 优于词汇导向的金字塔匹配核和空间金字塔匹配核。

图像目标的分类、识别是计算机视觉和模式识别领域的一个重要研究问题。由于图像目标存在视角变化、亮度变化、尺度、目标变形、遮挡、复杂背景以及目标类内差别等影响, 使得图像目标的分类识别非常困难。针对这些问题, 已提出具有各种不变性的局部特征 [1-2], 如 SIFT(scale invariant feature transform)[3]。Fergus 等 [4] 提出基于局部特征的生成模型用于图像分类, Berg 等 [5] 和 Lazebnik 等 [6] 提出基于几何对应的图像分类方法, 这些方法没能较好地利用局部特征在特征空间的结构特性, 而且计算复杂度很高 [7]。最近提出基于特征包 (bags-of-features) 的方法对图像目标分类 [1,8-11], 取得了良好的效果, 但这些方法没有利用局部特征之间的空间关系。Grauman 提出的金字塔匹配核 (pyramid match kernel, PMK)[9] 具有比较优越的匹配、分类性能, 但不适用于高维特征。最近, Grauman 提出词汇导向的金字塔匹配核 (vocabulary-guided pyramid match kernel, VGPM)[10], 并取得良好的性能, 该方法首先利用一系列局部特征在特征空间建立金字塔型结构, 然后将图像的局部特征嵌入金字塔结构, 形成特征空间多分辨率直方图, 再计算直方图匹配, 得到核函数。为了利用局部特征在图像空间的位置关系, Lazebnik 等 [7] 借鉴 Grauman[9] 的金字塔匹配核的思想, 提出空间金字塔匹配核 (spatial pyramid matching kernel, SPM), 首先对局部特征量化, 在二维图像空间建立金字塔, 然后计算加权的子图像区域局部特征直方图交叉。

...

---

<sup>1</sup>基于双空间金字塔匹配核的图像目标分类. 陈海林, 吴秀清

## 源代码

file:main.c

```
1 | int main(int argc, char ** argv) {  
2 |     printf("Hello world!\n");  
3 |     return 0; /*'程序运行完毕'*/  
4 | }
```

...