

浙 江 大 学

本 科 生 毕 业 论 文



题目_____基于图像检索的卡通人脸合成方法_____

姓 名_____姚 万 超_____

学 号_____3100104275_____

指导教师_____蔡 登_____

专 业_____计算机科学与技术_____

学 院_____计算机科学与技术_____

A Dissertation Submitted to Zhejiang
University for the Degree of Bachelor of
Engineering



TITLE: Cartoon Face Synthesis Based
On Image Retrieval

Author: Wanchao Yao

Supervisor: Deng Cai

Major: Computer Science and Technology

College: Zhejiang University

Submitted Date: May 31st, 2014

浙江大学本科生毕业论文(设计)诚信承诺书

1.本人郑重地承诺所呈交的毕业论文(设计),是在指导教师的指导下严格按照学校和学院有关规定完成的。

2.本人在毕业论文(设计)中引用他人的观点和参考资料均加以注释和说明。

3. 本人承诺在毕业论文(设计)选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文(设计)中对侵犯任何方面知识产权的行为,由本人承担相应的法律责任。

毕业论文(设计)作者签名:

_____年____月____日

摘要

卡通人脸是突出照片人脸面部特征的艺术形态，其创作依赖于人的主观判断，自动合成卡通人脸是具有挑战的任务。目前存在的方法，合成的卡通人脸和照片人脸相似程度不高。本文提出的卡通人脸合成方法主要基于图像检索和 LASSO 线性回归模型。我们收集了 1000 张卡通人脸图像，并剪切出其中 100 张的面部组件，为样本图块提取 PHOG 特征，并用随机 K-维树算法建立索引。输入人脸照片，先使用人脸对齐算法检测人脸，再对照片进行预处理，并截取出面部组件。通过两步实现人脸照片的卡通化，这两步均通过求解基于输入图块 K-近邻的 LASSO 线性回归模型来合成与输入图块相似的卡通图块。第一步将照片分割成不重叠图块，合成卡通图块再拼成初步的卡通图像。第二步通过 K-均值聚类提取出面部皮肤区域，并使之卡通化；再合成面部组件图块，使用泊松图像编辑将它们缝合到皮肤上。最后，通过后期处理消除瑕疵。实验评估表明，合成的卡通人脸和照片人脸相似程度较高，本文的卡通人脸合成方法具有良好的效果。

关键词 卡通人脸、图像检索、LASSO

Abstract

Cartoon face is an art form of photo face, which exaggerates the facial features. The creation of cartoon relies on subjective judgments of people. So automatic synthesis of cartoon face is a challenging task. The generated cartoon face of current methods is poorly similar to the input photo face. This paper presents a cartoon face synthesis method based on image retrieval and LASSO linear regression model. We collected 1000 cartoon face images, and cut out the facial components of 100 cartoon face images. We first compute the PHOG descriptor of sample patches, then use randomized K-d tree to build index. Input a face photo, we use face alignment algorithm to detect face and preprocess the photo. Afterwards, we cut out facial components. Cartoon synthesis is obtained through two steps. Both two steps are achieved by solving the LASSO linear regression model, which is based on K-nearest neighbors of the input patch. In the first step, the photo is be divided into non-overlapping patches, then synthesize all the patches and assemble them together to obtain a preliminary cartoon image. In the second step, we first use K-means algorithm to extract the facial skin area, and make it cartoonish; then synthesize the facial components and use Poisson image editing algorithm to stitch them onto the skin. Finally, we employ post-processing technique to reduce artifacts. Experimental evaluation shows the high similarity between generated cartoon face and photo face, which demonstrates the good performance of the proposed method on cartoon face synthesis in this paper.

Keywords Cartoon face, Image retrieval, LASSO

目录

摘要	I
Abstract.....	II
第 1 章 绪论	1
1.1 课题背景	1
1.2 本文研究目标和内容	2
1.2.1 研究目标	2
1.2.2 研究内容	2
1.3 本文结构安排	3
第 2 章 文献综述	4
2.1 半自动合成卡通人脸	4
2.2 全自动合成卡通人脸	5
2.3 本章小结	6
第 3 章 研究方案	7
3.1 图像预处理	7
3.2 特征提取	7
3.3 样本集的建立	8
3.4 初步合成卡通图像	8
3.5 合成卡通人脸	10
3.6 后期处理	12
3.7 本章小结	13
第 4 章 实验结果	14
4.1 CUHK 数据库的实验结果	14
4.2 其它人脸照片的实验结果	15
4.3 与其它合成方法的实验结果比较	16
4.4 本章小结	18

第 5 章 分析与讨论	19
5.1 近邻数目 K 的选择	19
5.2 与其它合成方法的比较	19
5.3 本章小结	20
第 6 章 本文总结与展望	21
参考文献	22
致谢	24

第1章 绪论

1.1 课题背景

随着计算机技术的迅速发展,人脸得到广泛的研究,是最近 30 年里模式识别、计算机视觉和图像处理中最热门的研究主题之一。人脸在很多领域得到应用,比如人脸检测、识别、跟踪等。近年来,人脸的自动绘制成为一个新兴领域,获得较大关注,比如素描人脸、肖像画、卡通人脸等。

“卡通”一词起源于中世纪,最初是指一件艺术品,如绘画、壁画、挂毯或彩色玻璃窗的草图,随着时间的推移,“卡通”具体的定义也发生了变化。在 19 世纪,它指杂志和报纸中幽默的插图;在 20 世纪初及以后,它指漫画和动画片。在本文的研究范围内,“卡通”指半写实的绘画。

卡通自动合成系统一方面有助于卡通电影和动画片制作者绘制卡通人物;另一方面,人们喜欢用自己的照片合成卡通图像,可以应用在网络聊天、电子相册等数字娱乐中。因此,卡通人脸自动合成技术对未来智能多媒体处理技术发展具有重大意义。

市场上有通过人脸照片合成卡通人脸的手机应用,主要分两类:一类是通过提取人脸轮廓、面部组件和发型的特征,自动匹配模板库中最相似的组件组合出卡通人脸,并提供交互界面,用户可以自行修改眼睛、鼻子等,但合成结果和人脸照片相似度不大;另一类是根据照片自动合成,用算法突出照片中人脸的边缘,抹平其它部分,但合成结果并不理想,有较大瑕疵。

艺术家往往是通过自己的主观判断,用简单、夸张的笔触突出人脸鲜明的面部特征(如宽额头、浓眉毛等)来绘制卡通人脸。不同的艺术家对同一个人绘制的卡通人脸也往往有较大不同。但是人们对人脸比较敏感,能够准确的识别出卡通人脸对应的真实人脸,这说明卡通人脸和真实人脸存在着隐含的、难以用语言表达的对应关系。照片和卡通人脸相比,在风格、线条、颜色等方面都有较大不同,所以由计算机根据人脸照片自动合成卡通人脸是一项具有挑战的任务。

1.2 本文研究目标和内容

1.2.1 研究目标

本研究的目标是实现由计算机根据正面人脸照片自动合成卡通人脸，并使得人眼能够辨别出照片人脸和卡通人脸的对应关系，研究主要基于图像检索和机器学习相关技术。

我们的样本集中有 1000 张从网络上收集来的卡通人脸图像，并剪切出了嘴巴、鼻子、左眼、右眼、左眉、右眉各 100 张图像。每次输入一张人脸照片，提取照片块和面部组件的 PHOG 特征，以此在样本集的卡通图像样本中检索 K-近邻，根据返回结果求解 LASSO 线性回归模型。合成分两步，第一步将根据照片块合成的卡通块拼接成初步的卡通化合成结果，第二步先合成卡通人脸皮肤，再分别合成面部组件图块，最后将面部组件图块使用泊松图像编辑缝合到卡通人脸皮肤上，进一步实现卡通化。在完成合成后，还要进行后期处理消除瑕疵，去除噪音和块效应。

1.2.2 研究内容

研究内容主要分为以下四个部分：

1) 卡通图像样本集的建立

研究如何根据不同风格、大小不一的卡通图像建立样本集用于检索和合成，重点在于卡通图像的处理、面部组件剪切、特征提取和建立索引。

2) 根据照片初步合成卡通图像

研究如何根据照片通过检索样本集成初步的卡通图像，使得合成结果与原始照片相近但初步实现卡通化，重点在于图像检索和合成模型的建立与求解。

3) 根据人脸特征合成卡通人脸

研究如何根据人脸的特征合成出卡通人脸，使得面部肤色统一，面部组件相似但进一步实现卡通化，能够识别出照片人脸和卡通人脸之间的对应关系，重点在于人脸对齐、皮肤分割、图像检索、合成模型的建立与求解以及面部组件缝合。

4) 合成结果的后期处理

研究如何去除合成的卡通图像中的瑕疵，比如噪音和块效应，重点在于去除瑕疵的同时能够保留有意义的信息。

1.3 本文结构安排

本文的结构安排如下：

第一章为绪论，包括本文的研究背景、研究目标和研究内容。

第二章为文献综述，介绍了目前文献中卡通人脸合成的相关技术，主要分为半自动合成卡通人脸和全自动合成卡通人脸。

第三章是研究方案部分，分步骤详细描述了我们提出的卡通人脸合成方案，主要基于图像检索和 LASSO 线性回归模型。

第四章展示了我们的卡通人脸算法的实验结果，通过和其它合成方法的结果比较证实了我们提出的算法的有效性。

第五章，分析、讨论了我们的合成方案与其它方案的异同。

最后，第六章总结全文，展望未来工作。

第2章 文献综述

2.1 半自动合成卡通人脸

在早期的研究中，人脸的研究远没有现在发达，人脸对齐、关键点检测等技术也没有如今成熟，所以并不能完全由计算机根据输入的人脸照片自动合成出卡通人脸图像，而是需要用户的交互，半自动的合成卡通人脸图像。

根据输入的人脸照片自动的合成出卡通人脸图像的工作可以追溯到 PicToon 系统(Chen H, Zheng N N, Liang 等人, 2002)[1]。PicToon 易于使用，只需要较少的用户交互，使用素描生成和笔触渲染从输入图像创建出个性化的卡通图像。PicToon 系统主要由三个部分组成：基于输入图像的卡通图像生成器，实现夸张效果的交互式卡通图像编辑器，以及语音驱动的卡通动画生成器。卡通图像生成可以分解为两个过程：素描生成和笔触渲染。首先，PicToon 采用基于例子的方法自动画出面部轮廓线。接着，一个连同面部模板拟合的非参数化采样方案被用来提取输入图像基于矢量的面部素描线。在笔触渲染这一步，用户可以添加不同的艺术笔触到卡通素描中。最后，卡通编辑器中有预先设计的模板，用户通过交互可以轻易的使卡通图像夸张或更具表现力。

Wen F, Chen S 和 Tang X 在 2008 年提出了一个基于家庭相册的卡通个性化系统——EasyToon[2]。EasyToon 使用最先进的计算机视觉和图形学技术，提供有效的 UI 设计，可以快速生成个性化的卡通图像。EasyToon 从家庭相册选择照片，将其中的人脸自然的混合到卡通图像中。个性化的卡通图像主要通过两个步骤简单和快速地获得。首先，交互的从相册选择最佳人脸候选图像。然后，将选择的人脸自动的混合到有趣的卡通形象来合成个性化的卡通形象。借助 EasyToon，人们不需要任何艺术背景，就可以简单的在几分钟内制作出高质量的个性化卡通图像。

除了通过照片合成卡通图像之外，一些工作专注于卡通动画，如视频卡通化。Wang J, Xu Y, Shum H Y 等人在 2004 年提出了 Video Tooning 系统，用于将输入视频转换成一个有一系列的风格的高度抽象的时空域连贯的卡通动画[3]。使用 Video Tooning 系统，用户只需要简单的画出关键帧中的物体轮廓。之后，一个均值平移引导的插值算法被用来创建三维语义区域，同时保持沿时间维度

的流畅轨迹。这些区域为创建嵌入时空域视频内的平滑的二维边缘表和笔触表提供了基础。这些区域、边缘表和笔触表是通过在特定的时间切片来渲染的。由语义平滑的区域和表提供的时间上的一致，导致了时间上一致的非真实感外观。

2.2 全自动合成卡通人脸

随着模式识别、计算机视觉和人脸研究的发展，计算机已经可以完全自动的根据输入的人脸照片合成出素描人脸、卡通人脸，不再需要人的辅助。合成素描人脸和卡通人脸有相同之处，也有不同之处，可以互相借鉴。

在目前的文献中，已经提出了许多方法来进行全自动的素描人脸的合成。一个基于示例的面部素描系统最初是由 Chen H, Xu Y Q, Shum H Y 等人在 2001 年提出的，它通过学习由一位艺术家绘制的特定风格的示例素描，从输入照片生成素描[4]。该方法采用非参数抽样系统学习图像及其素描之间的统计特征。Tang X 和 Wang X 于 2003 年提出了一种使用人脸照片合成除去头发区域的人脸素描的系统，主要采用了全局本征变换和马尔可夫网络两种方法[5]。这个系统把人脸照片的形状和纹理信息分开，并分别对它们应用了转换。

2010 年，Zhang W, Wang X 和 Tang X 提出了一种方法，从在不同的照明条件下或在不同的姿态下拍摄的人脸照片而不是素描样本集来合成一张人脸素描[6]。此方法使用一个多尺度马尔可夫网络产生局部素描图块，它主要包括三个步骤：第一步，引入针对面部组件的形状先验来减少由照明和姿态变化引起的瑕疵和变形；第二步，给定一个照片图块，使用对光照变化鲁棒性更强的新的素描图块的描述和度量来查找备选素描图块。最后，测量强度和梯度兼容性的平滑项被用来匹配马尔可夫网络上邻近的素描图块。

此外，人脸的分层-组合模型由 Xu Z, Chen H, Zhu S C 等人在 2008 年提出，它使用了三层“与-或图”来解释由尺度转换和状态转换引起的人脸的规律性和剧烈的结构多变性[7]。第一层将每个人脸当做一幅整体图像来处理，而第二层将局部面部组件当做一组单独的模板来共同细化。人脸进一步被分割成区域和细节的面部特征(例如眼角、皱纹等)在第三层中实现。这个模型有助于合成卡通人脸。

主动轮廓(或者蛇)是对于可变形物体提取有用的模型。Hsu R L 和 Jain A K.

在 2003 年提出用来迭代操纵多个蛇的方法，称为交互蛇，这个方法尽量减少轮廓上和封闭区域内的单独蛇之间的吸引能量泛函和多个彼此交互的蛇之间的排斥能量泛函[8]。使用交互蛇，一个从三维人脸模型的顶点子集导出的人脸语义图被提出用于构造卡通人脸。

2011 年，Li H, Liu G 和 Ngan K N.提出了一个从输入的正面人脸照片产生卡通人脸图像的引导合成方法[9]。该方法基于一个局部线性模型来执行卡通人脸图像合成，它采用来自照片-卡通对样本集的引导图像。为了实现鲁棒的合成，四个权重函数被定义来计算合成权重。第一个被称作照片-照片权重函数，其目的是计算一个输入照片和一个样本照片之间的相似性。第二个是照片-卡通权重函数，它用于通过计算一个卡通和一个输入照片之间的相似性来计算似然性。第三个权重函数是在合成的图像中计算的，它被设计用于施加邻近合成图块之间的平滑度的限制。最后的权重函数用来评估一个合成图块和一个照片图块基于空间距离的相似性。使用四个权重，在卡通合成中可以充分地利用人脸照片、合成的卡通、以及艺术家绘制的卡通之间的关系，这在现有的方法中通常是被忽略的。

2.3 本章小结

本章介绍了目前文献中卡通人脸合成的相关技术。在早期，计算机需要用户的交互来半自动的合成卡通人脸；随着模式识别、计算机视觉和人脸研究的发展，计算机可以全自动的根据输入的人脸照片合成出卡通人脸。

第3章 研究方案

3.1 图像预处理

为了进行卡通人脸合成,需要先对照片和卡通样本图像做预处理(照片和卡通图像中的人脸必须是正面的、自然表情的)。对于每一张人脸照片或者卡通人脸图像,需要先使用[10]的人脸对齐算法检测人脸,并标定脸部轮廓、嘴巴、鼻子、左眼、右眼、左眉、右眉的关键点。一个使用人脸对齐算法检测人脸的例子如图 3-1 所示,图 3-1 左侧为一张人脸照片,图 3-1 右侧为用[10]的人脸对齐算法检测人脸得到的结果,面部组件和人脸轮廓的关键点分别用不同颜色标识出。然后,通过调整左眼的位置到(31,63),右眼的位置到(63,63),来将图像截取并缩放至 128×96 大小,再将嘴巴、鼻子、左眼、右眼、左眉、右眉单独截取出来。此外,所有图像原始都在 RGB 色彩空间,为了进行卡通合成,需要将它们转换至 YCbCr 色彩空间,合成过程主要基于 Y 通道。



图 3-1 人脸对齐算法检测人脸的示例,左侧为人脸照片,右侧为人脸对齐的结果

3.2 特征提取

对于图像特征的提取,目前的文献存在着许多提取区域特征的方法,例如 SIFT[11], MSER[12], 以及 HOG[13]。SIFT 描述图像的局部特征,它在空间尺度中寻找关键点,并提取出其位置、尺度、旋转不变量。MSER 是指图像的最大稳定极值区域(使用不同的灰度阈值对图像进行二值化时得到的最稳定的区域),该区域对于图像灰度的仿射变化具有不变性。MSER 用来找到两幅视角不同的图像中图像元素的对应关系。HOG 使用了图像本身的梯度方向特征, HOG 在一个网格密集的大小统一的方格单元上计算,并且为了提高精确度使用了重叠的局部对比度归一化的方法。

在本研究中，我们首先将区域分解成不同的等级，并计算名为分层梯度方向直方图(PHOG)的特征[14]。因为该特征是基于一个区域分层表示来计算的，它可以捕获到边缘在不同尺度下的空间分布，所以 PHOG 常常用来测量形状/区域对应程度。图 3-2 展示了计算人脸照片图块 PHOG 特征的一个例子。在图 3-2 左侧所示的人脸照片中，红色方块表示一个候选的照片图块。通过 Canny 边缘检测得到的边缘图在图 3-2 中间给出，其中的边缘图块被标记为红色正方形，由边缘图块来计算 PHOG 特征。图 3-2 右侧展示出了在三个不同的尺度下，即 $L=1,2,3$ 的结果。在每个分层分辨率级别的每个网格单元计算一个 HOG 特征，一个图块最终的 PHOG 描述符是所有 HOG 描述符的连接。



图 3-2 计算图块 PHOG 特征的示例，左侧为人脸照片，中间为边缘图，右侧为计算 PHOG

3.3 样本集的建立

我们有 1000 张 128×96 大小的样本卡通图像，每一张图像被分割成 48 个互不重叠的 16×16 大小的图块，总计 48000 个 16×16 大小的图块。此外，还有截取出的嘴巴、鼻子、左眼、右眼、左眉、右眉图块各 100 个。

我们需要为卡通图像图块和面部组件图块分别提取 PHOG 特征，并分别为它们建立索引，来构成样本集。本研究采用随机 K-维树算法来建立索引和查找 K-近邻[15], [23]。原始的 K-维树是每个节点都为 K 维点的二叉树，所有非叶子节点在数据方差最大的维度上将数据分开，可以视作用一个超平面把空间分区成两部分。与此相比，随机 K-维树随机选取数据方差最大的前 D 维中的一维来将数据分开。

3.4 初步合成卡通图像

输入一张人脸照片，首先对其进行预处理，截取并缩放至 128×96 大小后，

我们就可以根据照片的 Y 通道来初步合成卡通图像。首先，与样本卡通图像一样，照片被分割成 48 个互不重叠的 16×16 大小的图块，为每一个图块提取 PHOG 特征，根据图块特征在样本集中检索 K-近邻。一方面由于检索精确 K-近邻时间复杂度较高，另一方面并不需要 100% 精确的 K-近邻，我们可以通过牺牲少量的精度来大幅度减少检索时间。对于每一个照片图块，得到它的近似 K-近邻后，我们通过求解一个基于 K-近邻的 LASSO 线性回归模型[16]，来合成该图块对应的卡通图块。

令 X 表示预处理后的人脸照片， Z 表示初步合成的卡通图像， D 表示 K-近邻图块组成的矩阵， α 表示候选样本图块系数组成的列向量，下标 p 表示照片图块的序号， λ 表示惩罚项权重，求解系数列向量如式 3-1 所示：

$$\hat{\alpha} = \arg \min \|X_p - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (\text{式 3-1})$$

合成图块 Z_p 如式 3-2 所示：

$$Z_p = D\hat{\alpha} \quad (\text{式 3-2})$$

初步合成的灰度卡通图像 Z 由所有的合成图块拼接而成，可以使用多种方法来着色。这里为了比较原始照片与合成图像，我们使用颜色保留方法为之着色。令 (i, j) 表示像素点位置，下标 $_{syn}$ 和 $_X$ 分别表示合成图像和输入照片，则合成图像每一个像素点在 YCbCr 色彩空间的值如下式所示：

$$\begin{aligned} Y_{syn}(i, j) &= Z(i, j), \quad i \in [1, 128], j \in [1, 96] \\ Cb_{syn}(i, j) &= Cb_X(i, j), \quad i \in [1, 128], j \in [1, 96] \\ Cr_{syn}(i, j) &= Cr_X(i, j), \quad i \in [1, 128], j \in [1, 96] \end{aligned} \quad (\text{式 3-3})$$

这一步的目的是将照片整体实现卡通化，包括背景以及人脸、头发的轮廓、纹理，同时保证合成的卡通图像与原始照片尽可能的接近，使得人眼能够识别出照片与卡通图像的对应关系。

一个人脸照片初步合成卡通图像的例子如图 3-3 所示，图 3-3 左侧为一张人脸照片，图 3-3 右侧为初步合成的卡通图像。可以看出，在初步合成卡通图像中，人物头发部分的细节纹理减少了，耳朵轮廓更加清晰了，合成图像与原始照片相似度较高。



图 3-3 初步合成卡通图像的示例，左侧为人脸照片，右侧为初步合成的卡通图像

3.5 合成卡通人脸

我们首先对人脸照片进行 K-均值聚类[17]，借助人脸轮廓关键点信息提取出人脸皮肤区域，再计算皮肤区域内亮度通道 Y 和两个色度通道 Cb、Cr 的均值，将它们的均值赋值给所有皮肤像素点，从而实现人脸肤色卡通化。此外，人脸卡通画除了面部肤色平均之外，还有脸型轮廓突出的特点，即用鲜明的笔触勾勒出脸部轮廓线，因而，我们也取了人脸下巴区域的 7 个关键点，用黑色线条勾勒出下巴轮廓线。令 Z_{face} 表示经由这一步合成的卡通人脸。图 3-4 展示了人脸面部肤色卡通化的一个例子。图 3-4 左侧为一张人脸照片，图 3-4 右侧为面部肤色卡通化的人脸。



图 3-4 面部肤色卡通化的示例，左侧为人脸照片，右侧为面部肤色卡通化的人脸 Z_{face}

在对输入的人脸照片进行预处理的阶段，通过人脸对齐确定了人脸的关键点，再提取出了人脸的轮廓和面部组件的图块，包括嘴巴、鼻子、左眼、右眼、左眉、右眉。令 S_T 表示面部组件的图块集合， T 取 X 表示输入照片的面部组件的图块集合， T 取 Z 表示合成卡通的面部组件的图块集合， S_T 如下式所示：

$$S_T = \{T_{mouth}, T_{nose}, T_{l-eye}, T_{r-eye}, T_{l-brow}, T_{r-brow}\}, \quad T = X, Z \quad (\text{式 3-4})$$

每一个面部组件，样本集中都有 100 张卡通图块。令下标 i 表示面部组件的序号， S_{Xi} 表示照片中的面部组件 i 的图块， S_{Zi} 表示合成的面部组件 i 图块。对

于每一个面部组件图块 S_{xi} ，与上一步骤中合成卡通图块类似，先计算 S_{xi} 的 PHOG 特征，然后在样本集中检索近似 K-近邻。得到 S_{xi} 的近似 K-近邻后，我们通过求解一个基于 K-近邻的 LASSO 线性回归模型，来合成 S_{xi} 对应的卡通图块。

令 D_i 表示面部组件 i 的 K-近邻图块组成的矩阵， α_i 表示面部组件 i 的候选样本图块系数组成的列向量， λ 表示惩罚项权重，求解系数列向量如下式所示：

$$\hat{\alpha}_i = \arg \min \|S_{xi} - D_i \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1, \quad i=1,2,3,4,5,6 \quad (\text{式 3-5})$$

合成的面部组件图块 S_{zi} 如下式所示：

$$S_{zi} = D_i \hat{\alpha}_i, \quad i=1,2,3,4,5,6 \quad (\text{式 3-6})$$

在按照上式合成完每一个面部组件 S_{zi} 后，得到合成卡通的面部组件的图块集合 S_z 。一个由人脸照片的面部组件图块集合 S_x 合成卡通人脸的面部组件图块集合 S_z 的例子如图 3-5 所示，图 3-5 左一为一张人脸照片，图 3-5 左二为提取出的人脸照片的面部组件图块集合 S_x ，图 3-5 右二为从样本集中检索到的候选图块集合，图 3-5 右一为合成卡通人脸的面部组件图块集合 S_z 。结果显示，合成卡通人脸的面部组件图块体现出了原始图块的鲜明特征，实现了卡通化，同时与原始图块有较高识别度。

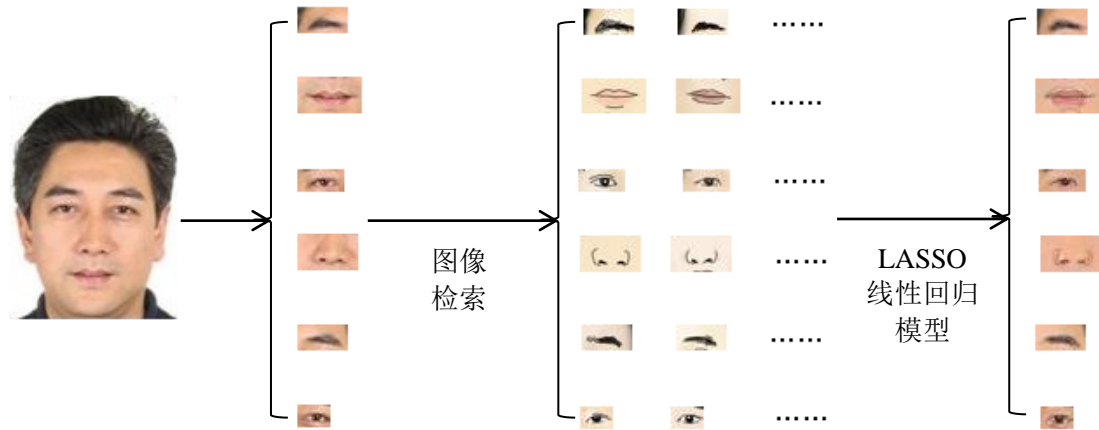


图 3-5 合成卡通人脸面部组件图块的示例，左一为人脸照片，左二为人脸照片的面部组件图块集合 S_x ，右二为检索到的候选图块集合，右一为合成的卡通面部组件图块集合 S_z

接下来就是将 S_z 中的每一个面部组件 S_{zi} 安装到 Z_{face} 上。需要注意的是， Z_{face} 的肤色是卡通化的，而由 LASSO 线性回归模型合成的面部组件 S_{zi} 中的肤色是接近照片中相应位置肤色的，所以直接将 S_{zi} 的像素值复制到 Z_{face} 上会导致

面部组件与周围肤色有色差的问题，这里我们使用泊松图像编辑无缝缝合算法[18]将每一个面部组件 S_{zi} 缝合到 Z_{face} 上，如图 3-6 所示。图 3-6 左侧为面部肤色卡通化的人脸 Z_{face} ，图 3-6 中间为合成的卡通面部组件图块集合 S_z ，图 3-6 右侧为将 S_z 通过泊松图像编辑无缝缝合到 Z_{face} 得到的结果。

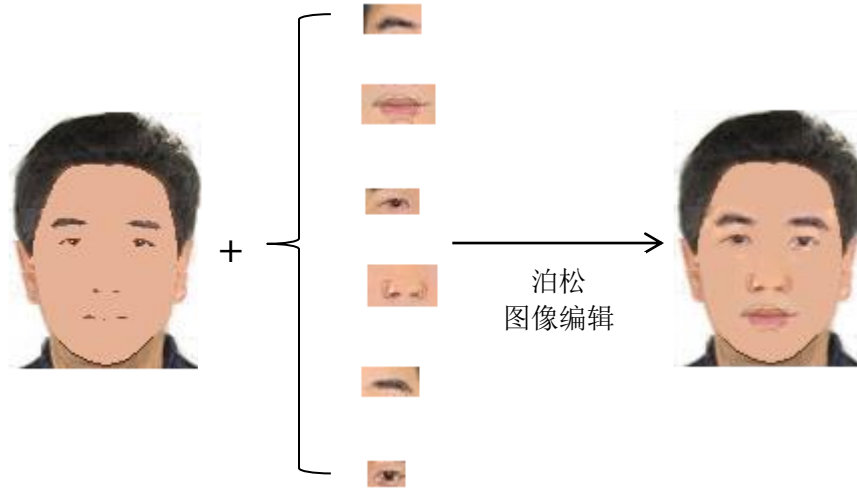


图 3-6 泊松图像编辑将卡通人脸面部组件图块缝合到卡通人脸皮肤的示例，左侧为面部肤色卡通化的人脸 Z_{face} ，中间为卡通面部组件图块集合 S_z ，右侧为将 S_z 缝合到 Z_{face} 的结果

3.6 后期处理

经过以上步骤后，我们已经得到了彩色的卡通人脸合成图像。但是，由于是计算机根据人脸照片全自动合成的卡通人脸图像，或多或少都会产生噪音和块效应等瑕疵，需要通过后期平滑处理来消除。我们比较了以下两种算法进行后期处理的结果：

1) 非局部均值算法

不像“局部平滑”滤波器，非局部均值(NL-means)算法[19]不通过平均周围像素点的值来更新当前像素点的值，而是使用被认为是最相像似的像素点的值加权平均后的值来更新。每个像素点的权重取决于它和目标像素点之间的距离。

2) 引导图像滤波算法

引导图像滤波器[20]通过考虑引导图像的内容计算滤波输出，引导图像可以是输入图像自身或另一个不同的图像。引导图像滤波器可以作为和双边滤波器类似的边缘保留平滑算子，但在边缘附近有更好的效果。

图 3-7 展示了上述两种后期处理算法应用于同一合成结果的示例，图 3-7 左侧为未经后期处理的卡通图像合成结果，图 3-7 中间为由非局部均值算法处理后的结果，图 3-7 右侧为由引导图像滤波算法处理后的结果。通过比较可以发现，两种后期处理算法都有效消除了眼睛、眉毛附近的块效应和鼻子、嘴巴处的多余噪音，不过引导图像滤波算法相比于非局部均值算法引入了更多的模糊效应。



图 3-7 两种后期处理算法结果的示例，左侧为未经后期处理的卡通图像合成结果，中间为由非局部均值算法处理后的结果，右侧为由引导图像滤波算法处理后的结果

3.7 本章小结

本章详细描述了我們提出的基于图像检索和 LASSO 线性回归模型的卡通人脸合成方案。我們首先需要建立样本集，为样本图块提取 PHOG 特征，用随机 K-维树算法来建立索引和查找 K-近邻。每输入一张人脸照片，先对照片进行预处理，使用人脸对齐算法检测人脸，将图像截取并缩放至 128×96 大小，再将嘴巴、鼻子、左眼、右眼、左眉、右眉单独截取出来，之后转换至 YCbCr 色彩空间，合成过程主要基于 Y 通道。通过两步实现人脸照片的卡通化，这两步都基于求解图块 K-近邻的 LASSO 线性回归模型。第一步将照片分割成 48 个互不重叠的 16×16 大小的图块，合成每一个卡通图块再拼接成初步的卡通图像。第二步先对人脸照片进行 K-均值聚类，提取出面部皮肤区域，使之卡通化；之后合成面部组件的图块，包括嘴巴、鼻子、左眼、右眼、左眉、右眉，再将合成图块使用泊松图像编辑无缝缝合到卡通化人脸皮肤上。最后，通过后期处理消除噪音和块效应等瑕疵，比较了非局部均值算法和引导图像滤波算法，两种后期处理算法都有效消除了块效应和噪音，不过引导图像滤波算法相比于非局部均值算法引入了更多的模糊效应。

第4章 实验结果

在本章中，我们以一些常用于人脸素描合成研究的人脸照片，以及我们从互联网收集的额外的人脸照片，来验证我们所提出的卡通人脸合成算法的性能。

4.1 CUHK 数据库的实验结果

我们首先用一组被[5]使用的 CUHK 数据库评估我们的方法。在图 4-1 中，展示了对一些测试人脸照片使用我们的方法合成的例子的结果。输入的照片如图 4-1 左一和右二所示，图 4-1 左二和右一展示了我们卡通人脸合成方法的结果。



图 4-1 CUHK 数据库的实验结果，左一和右二为人脸照片，左二和右一为卡通人脸合成结果

可以看出，我们的方法能够成功从测试人脸照片生成卡通。例如，对于图 4-1 左一下方的男生，虽然较强的照明可以从输入照片中观察到，卡通人脸合成的质量也比较好，它表现出与原始的照片相似的外观。此外，对于图 4-1 左一上方及右二下方的正面照明较昏暗的人脸照片，我们的方法也成功地生成了卡通图像。

4.2 其它人脸照片的实验结果

我们还用一些从互联网上的人脸照片测试了我们提出的方法，如图 4-2 所示。



图 4-2 其它照片的实验结果，左一和右二为人脸照片，左二和右一为卡通人脸合成结果

人脸照片如图 4-2 左一和右二所示，合成的卡通图像如图 4-2 左二和右一所

示。图 4-2 的人脸照片的光照、以及人物的肤色、发型之间的差异较大，但同样都取得了较好的合成结果。此外，与 CUHK 数据库的照片相比，可以从图 4-2 的人脸照片中观察到更多的小的姿态变化。值得注意的是，样本集中的嘴巴都是自然的、无微笑的，也没有露出牙齿，但人脸照片中的微笑姿态和露出的牙齿同样能够得到合成。结果证明了我们方法的有效性。

4.3 与其它合成方法的实验结果比较

首先，简单的介绍其它两个卡通人脸合成方法，如下：

1) 引导卡通人脸合成方法

引导卡通人脸合成方法基于一个局部线性模型来执行卡通人脸图像合成，它采用来自照片-卡通对样本集的引导图像[9]。这个模型定义了四个权重函数来计算合成权重。第一个被称作照片-照片权重函数，其目的是计算一个输入照片和一个样本照片之间的相似性。第二个是照片-卡通权重函数，它用于通过计算一个卡通和一个输入照片之间的相似性来计算似然性。第三权重函数是在合成的图像中计算的，它被设计用于施加邻近合成图块之间的平滑度的限制。最后的权重函数用来评估一个合成图块和一个照片图块基于空间距离的相似性。该方法将照片分割成多个相同大小的图块，相邻图块有覆盖，合成每一个图块时，为样本集的所有图库计算权重，合成图块为它们的加权平均。

2) 基于马尔可夫网络的卡通人脸合成方法

在这个方法中，马尔可夫网络被用来概率建模潜在的模式和观察到的数据之间的关系，以及相邻潜在的模式之间的关系。一个马尔可夫网络可以使用基于沿着网络的信息传递的学习和推理的阶段来解决。网络的参数可以从样本数据中学习，然后用来估计一个潜在的模式。马尔可夫网络已成功地应用于求解低层视觉问题，如超分辨率[21]和照片-素描合成[5]。给定一个图像对，这个模型采用贝叶斯置信传播找到局部最大后验概率。

接下来，我们比较我们的卡通人脸合成方法的实验结果与上述两个合成方法的实验结果，为了执行一个公平的比较，三个方法中图块大小都设置为 16×16 ，我们方法中近邻数目设置为 30，其它两个方法中重叠的图块均以 11 个像素的重叠距离按照光栅扫描顺序遍历。实验结果的比较如图 4-3 所示。

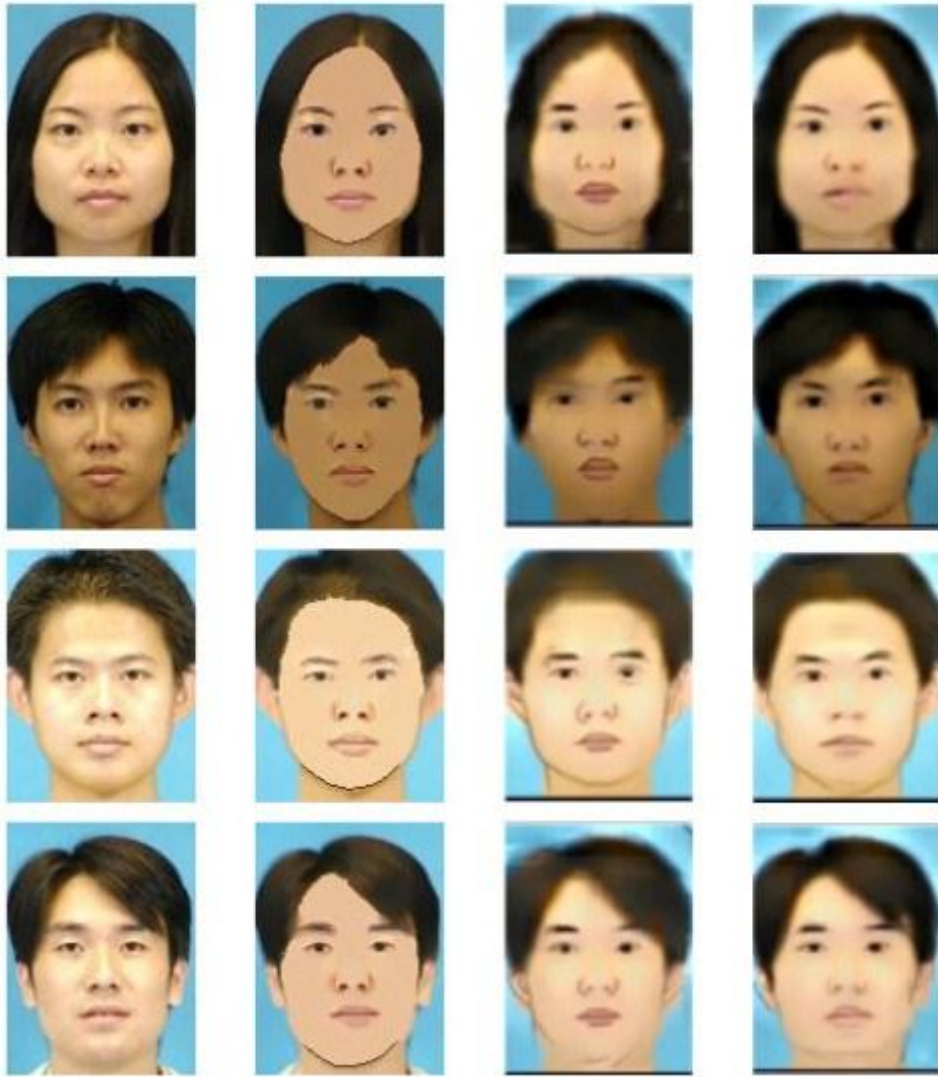


图 4-3 与其它合成方法的实验结果比较, 左一为人脸照片, 左二为我们的方法的合成结果, 右二为引导卡通人脸方法的合成结果, 右一为基于马尔可夫网络的方法的合成结果

图 4-3 左一为输入的人脸照片, 图 4-3 左二为我们的方法的合成结果, 图 4-3 右二为引导卡通人脸方法的合成结果, 图 4-3 右一为基于马尔可夫网络的方法的合成结果(图 4-3 右二和右一引用自[9])。通过合成结果的比较, 可以看出, 我们的方法具有较好的鲁棒性。面部组件的特征被我们的方法捕获, 在实现卡通化的同时与照片中的面部组件更为相近。此外, 我们的方法能够保留一些清晰的轮廓, 比如眼角, 而引导卡通人脸合成方法和基于马尔可夫网络的卡通人脸合成方法均引入较多的模糊。

4.4 本章小结

本章展示了我们的卡通人脸合成方法在 CUHK 数据库和网络上的人脸照片的合成结果，比较了我们的合成结果与引导卡通人脸合成方法和基于马尔可夫网络的卡通人脸合成方法的合成结果。比较显示，我们的方法具有较好的鲁棒性，更加突出面部组件的特征，且在实现卡通化的同时与照片中的面部组件更为相近。此外，我们的方法能够保留一些清晰的轮廓，而引导卡通人脸合成方法和基于马尔可夫网络的卡通人脸合成方法均引入较多的模糊。

第5章 分析与讨论

5.1 近邻数目 K 的选择

我们的卡通人脸合成是基于图像检索和 LASSO 线性回归模型的，检索每一个输入图块在样本集中的近似 K-近邻，之后求解基于 K-近邻的 LASSO 线性回归模型来合成对应的卡通图块。因而，样本集中的候选卡通图块数目(即近邻数目)影响合成图块与原始图块的相似程度。当一个小的近邻数目被选择，例如， $K=10$ ，合成的卡通图块往往会和原始图块有较大的视觉差异，导致失真。然而，近邻数目被设置为一个较大的值，例如， $K=90$ ，会有明显的过拟合，导致与原始图块太过相似从而失去了卡通特征，如图 5-1 右侧所示。从我们的实证研究，当近邻数目的值在 20 和 40 之间时，可以实现良好的性能。相应的结果如图 5-1 所示。

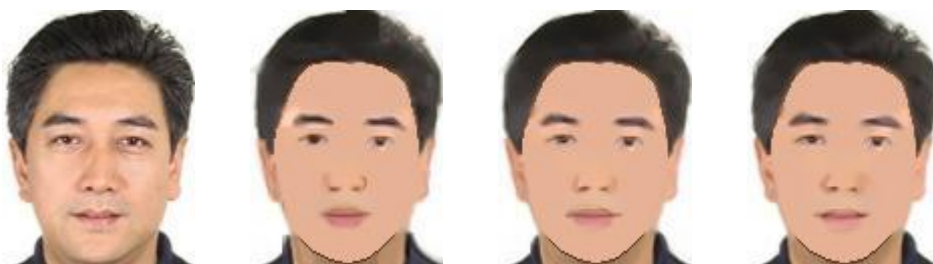


图 5-1 不同近邻数目 K 的合成结果比较,左一为人脸照片,左二到右一依次为 $K=10$, $K=30$, $K=90$ 的合成结果

图 5-1 左一为输入的人脸照片，图 5-1 左二到图 5-1 右一依次为近邻数目 $K=10$, $K=30$, $K=90$ 的合成结果。通过比较可以看出， $K=30$ 的结果效果最好，实现卡通化的同时能够识别出照片与卡通图像的对应关系； $K=10$ 的结果中的面部组件如嘴巴与照片中的有一定差别，且有较大的块效应瑕疵； $K=90$ 的结果中的面部组件如眉毛过拟合，卡通人脸整体上与照片人脸太过相似。

5.2 与其它合成方法的比较

在引导卡通人脸合成方法[9]中合成卡通图块时，同时基于四个一致性约束测量每个样本图块对照片合成的贡献。只有那些满足所有约束的图块，才被为

卡通合成施加较高的权重。给定一个的照片-卡通对样本集，该方法计算所有的样本集中的图块的加权平均值，而不是选择其中的一个或几个。

对于基于马尔可夫网络的方法[5]，它高度依赖于两种技术来合成卡通图块，即搜索/匹配算法和置信传播。第一个是在高维的搜索空间搜索和查找最佳匹配。后者是用来评价一个很好的匹配的空间的一致性。

我们的方法通过求解基于 K -近邻的 LASSO 线性回归模型来合成卡通图块，相比于岭回归模型，LASSO 能够将系数设置为 0 从而实现近邻图块的选择。在合成一个卡通图块时，引导卡通人脸合成方法为每一个样本图块都计算了一个权重，合成的卡通图块是所有样本图块的加权平均，合成的卡通图块往往与权重最高的样本图块相似，但权重最高的样本图块不一定与输入图块最相似，导致合成的卡通图块与输入图块相似度不高；我们的方法先检索输入图块的 K 近邻，再求解基于 K 近邻的 LASSO 线性回归模型，不仅合成速度上较快，且能够保证合成图块与输入图块的相似性。

在引导卡通人脸合成方法[9]和基于马尔可夫网络的方法[5]中，它们都采用了照片-卡通图像对作为样本，并且将输入的人脸照片分割为重叠的图块。我们的方法仅采用卡通图像作为样本，样本空间较容易构建和扩充；此外，我们将输入的人脸照片分割为互不重叠的图块和面部组件图块。虽然重叠的图块能够减少合成过程中出现的错误，但同样会引入模糊效应。

5.3 本章小结

本章首先分析了我们的合成方法中的近邻数目 K 对合成结果的影响，实证研究表面，当近邻数目的值在 20 和 40 之间时，可以实现良好的性能。接下来，分析、比较了我们的合成方法与引导卡通人脸合成方法[9]和基于马尔可夫网络的方法[5]的异同。

第6章 本文总结与展望

卡通人脸是突出照片人脸面部特征的艺术形态，其创作依赖于人的主观判断，自动合成卡通人脸是具有挑战的任务。目前存在的方法，合成的卡通人脸和照片人脸相似程度不高。

在这篇论文中，我们提出了一个新的方法从输入人脸照片合成卡通。合成是基于图像检索和 LASSO 线性回归模型的，它通过检索样本集中的卡通图像块来生成卡通。

我们首先需要建立样本集，样本集中有 1000 张从网络上收集来的卡通人脸图像，并剪切出了嘴巴、鼻子、左眼、右眼、左眉、右眉各 100 张图像。为样本图块提取 PHOG 特征后，用随机 K-维树算法来建立索引和查找 K-近邻。

每输入一张人脸照片，先对照片进行预处理，使用人脸对齐算法检测人脸，将图像截取并缩放至 128×96 大小，再将嘴巴、鼻子、左眼、右眼、左眉、右眉单独截取出来，之后转换至 YCbCr 色彩空间，合成过程主要基于 Y 通道。

通过两步实现人脸照片的卡通化，这两步都基于图块 K-近邻的 LASSO 线性回归模型。第一步将照片分割成 48 个互不重叠的 16×16 大小的图块，合成每一个卡通图块再拼接成初步的卡通图像。第二步先对人脸照片进行 K-均值聚类，提取出面部皮肤区域，使之卡通化；之后合成面部组件的图块，包括嘴巴、鼻子、左眼、右眼、左眉、右眉，再将合成图块使用泊松图像编辑无缝缝合到卡通化人脸皮肤上。最后，通过后期处理消除噪音和块效应等瑕疵，比较了非局部均值算法和引导图像滤波算法做后期处理的结果。

对一些人脸照片的实验评估表明，合成的卡通人脸和照片人脸相似程度较高，本文的卡通人脸合成方法具有良好的效果。

在未来，我们计划为卡通人脸合成将加入高层次语义特征，如 AAM[22]，我们可以以此构造外观模型来改善卡通合成性能。人们相信，一些人脸照片的高层次语义分析将对卡通生成很有用。

参考文献

- [1] Chen H, Zheng N N, Liang L, et al. PicToon: a personalized image-based cartoon system[C]//Proceedings of the tenth ACM international conference on Multimedia. ACM, 2002: 171-178.
- [2] Wen F, Chen S, Tang X. Easytoon: cartoon personalization using face photos[C]//Proceedings of the 16th ACM international conference on Multimedia. ACM, 2008: 1021-1022.
- [3] Wang J, Xu Y, Shum H Y, et al. Video tooning[C]//ACM Transactions on Graphics (ToG). ACM, 2004, 23(3): 574-583.
- [4] Chen H, Xu Y Q, Shum H Y, et al. Example-based facial sketch generation with non-parametric sampling[C]//Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. IEEE, 2001, 2: 433-438.
- [5] Tang X, Wang X. Face sketch synthesis and recognition[C]//Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003: 687-694.
- [6] Zhang W, Wang X, Tang X. Lighting and pose robust face sketch synthesis[M]//Computer Vision—ECCV 2010. Springer Berlin Heidelberg, 2010: 420-433.
- [7] Xu Z, Chen H, Zhu S C, et al. A hierarchical compositional model for face representation and sketching[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(6): 955-969.
- [8] Hsu R L, Jain A K. Generating discriminating cartoon faces using interacting snakes[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003, 25(11): 1388-1398.
- [9] Li H, Liu G, Ngan K N. Guided face cartoon synthesis[J]. Multimedia, IEEE Transactions on, 2011, 13(6): 1230-1239.
- [10] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2879-2886.
- [11] Lowe D G. Distinctive image features from scale-invariant keypoints[J].

International journal of computer vision, 2004, 60(2): 91-110.

[12] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and vision computing, 2004, 22(10): 761-767.

[13] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[14] Bosch A, Zisserman A, Munoz X. Representing shape with a spatial pyramid kernel[C]//Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, 2007: 401-408.

[15] Muja M, Lowe D G. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration[C]//VISAPP (1). 2009: 331-340.

[16] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1996: 267-288.

[17] Ravichandran K S, Ananthi B. Color Skin Segmentation Using K-Means Cluster[J]. International Journal of Computational & Applied Mathematics, 2009, 4(2).

[18] Pérez P, Gangnet M, Blake A. Poisson image editing[C]//ACM Transactions on Graphics (TOG). ACM, 2003, 22(3): 313-318.

[19] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 2: 60-65.

[20] He K, Sun J, Tang X. Guided image filtering[M]//Computer Vision—ECCV 2010. Springer Berlin Heidelberg, 2010: 1-14.

[21] Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision[J]. International journal of computer vision, 2000, 40(1): 25-47.

[22] Cootes T F, Edwards G J, Taylor C J. Active appearance models[J]. IEEE Transactions on pattern analysis and machine intelligence, 2001, 23(6): 681-685.

[23] Silpa-Anan C, Hartley R. Optimised KD-trees for fast image descriptor matching[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

致谢

弹指一挥间，四年大学生活已经接近了尾声。站在毕业的门槛上，回首往昔，欢笑与泪水已经成为最珍贵的回忆。值此毕业论文完成之际，我向所有帮助和关心我的人们表达最诚挚的感谢和最美好的祝愿。

感谢朱建科老师。从论文开题、资料查找、修改到最后定稿，您在每一个环节上都给予了我无私的帮助和指导。您严谨求是的治学态度对我影响深远，衷心感谢您对我学业道路上的帮助。

感谢何晓飞和蔡登老师。感谢你们付出了的大量辛劳，感谢你们对我的教导与帮助，更感谢你们对我的支持与照顾。你们的严格要求使我获益匪浅，也是我不断进步的动力。

感谢我的朋友和同学们。我们一起走过了难以忘怀的美好岁月，结下了真挚的友谊。感谢三位室友两年来的陪伴，因为有你们，我的大学充满了快乐与感动；感谢陈梦玲同学在背后默默的支持，因为有你，大学生活的最后一年很不一样。

最后要感谢我的父母，感谢他们对我学习的全力支持和生活上的照顾，他们是我坚实的后盾和最温暖的归宿。

本科生毕业论文(设计)任务书

一、题目：_____基于图像检索的卡通人脸合成方法_____

二、指导教师对毕业论文(设计)的进度安排及任务要求：

起讫日期 200 年 月 日至 200 年 月 日

指导教师(签名)_____职称_____

三、系或研究所审核意见：

负责人(签名)_____

年 月 日

毕 业 论 文(设计) 考 核

一、指导教师对毕业论文(设计)的评语:

指导教师(签名) _____
年 月 日

二、答辩小组对毕业论文(设计)的答辩评语及总评成绩:

成绩比例	文献综述 占(10%)	开题报告 占(20%)	外文翻译 占(10%)	毕业论文(设计)质 量及答辩 占(60%)	总 评 成绩
分 值					

答辩小组负责人(签名) _____
年 月 日