

局部回归模型驱动的人脸素描自动生成

吉娜焯^{1,2)}, 柴秀娟¹⁾, 山世光¹⁾, 陈熙霖¹⁾

¹⁾(中国科学院计算技术研究所中国科学院智能信息处理重点实验室 北京 100190)

²⁾(中国科学院大学 北京 100049)

(naye.ji@vip1.ict.ac.cn)

摘 要: 为了兼顾生成人脸素描与对应照片间的相像性和艺术性,提出一种利用局部回归模型学习照片-素描间的映射关系自动生成人脸素描的方法.首先在图像局部区域上采用回归技术获得照片-素描对应局部区域之间的映射关系;然后将映射关系应用到输入的人脸照片上,考虑结构性特征并模拟铅笔素描笔触来合成对应的素描图像.在 CUFS 数据库上的实验结果表明,采用该方法合成的素描保持了原始人脸照片的身份信息,并且体现出较好的铅笔素描纹理.

关键词: 人脸素描;局部回归;近邻;lasso;线积分卷积
中图法分类号: TP391

Local Regression Model for Automatic Face Sketch Generation

Ji Naye^{1,2)}, Chai Xiujuan¹⁾, Shan Shiguang¹⁾, and Chen Xilin¹⁾

¹⁾(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(University of Chinese Academy of Sciences, Beijing 100049)

Abstract: This paper proposes an automatic face sketch generation approach which learns the mapping relationship from photo-sketch pair examples through the local regression model. The proposed method takes into account both face fidelity and aesthetics. First, the relationship between a face photo and its corresponding sketch is learned on image patch level by exploring some regression techniques. Then, by applying the learned relationship to the input face photo, the final face sketch is generated with the consideration of structural information and the simulation of pencil stroke texture. Experiments conducted on CUFS database have shown that our results are more appealing than previous methods for preserving of both face identity and aesthetics.

Key words: face sketch; local regression; nearest neighbor; lasso; line integral convolution

人脸素描画像是肖像画的一种重要艺术形式.对于没有专业绘画技巧的人来说,绘制一幅生动的素描肖像画会很难;对于艺术家而言,也需要半个小时到几个小时的时间才能绘制一幅栩栩如生的肖像素描.近年来,鉴于肖像素描在数字娱乐和法律实施等方面的应用,越来越多的人开始关注人脸素描的

自动生成.借助于自动肖像素描合成技术,非专业的用户可以很简便地完成一幅素描肖像;同时,艺术家可以节约很多时间来绘制具有创意的个性化肖像画.另外,自动肖像素描合成有利于跨模态的人脸识别.在法律实施方面,通常先由画师根据目击证人口述绘制嫌犯素描,再与人脸库中由照片转换成的素描

匹配,而自动肖像素描生成系统有助于警方据此来搜寻嫌犯的照片。

目前,已有很多从图像生成素描的研究工作。一类工作是针对通用的各种图像,不特定于人脸图像。早期经典的从单幅图像生成素描的方法表现为钢笔墨水的风格。Salisbury 等^[1-2]实现了由用户从库中选择素描笔触再由计算机依据笔触生成图像对应的钢笔风格素描的系统,这 2 种钢笔素描生成方法是交互式的且不针对人脸,因此要得到一幅较好的素描肖像,需要大量的交互操作并要求用户有一定的艺术背景;随后,Mao 等^[3]提出一种用线积分卷积(line integral convolution, LIC)滤波的方法从图像生成铅笔素描;在 LIC 算法的基础上,谢党恩等^[4]提出一种基于卷积算子的铅笔滤镜生成算法,可实时转换视频图像为铅笔素描风格,莫晓斐等^[5]使用双边滤波增强形体特征从而生成了较好的铅笔素描画;Lu 等^[6]根据图像结构信息提取轮廓边缘,再根据纹理信息模拟铅笔笔触,从而生成自然美观的铅笔画。以上几种利用图像处理生成铅笔画的方法对场景、物体图像的素描化效果看起来不错,但是在人脸图像上的效果还不够令人满意。

另一大类工作是从图像生成素描,其侧重于复杂的人脸图像,依据人脸照片-素描对的数据集学习特定的素描风格,从而自动合成人脸素描^[7]。Tang 等^[8]应用特征变换全局地从一幅照片合成素描,这种全局线性模型无法很好地合成头发等变化多样的细节。为了克服此种限制,Liu 等^[9]提出了基于小块重构的方法,其中每个小块独立合成;但由于这种方法忽视了块与块之间的空间关系,最终的合成结果不能较好地反映人脸结构且有很强的块模糊效应。随后,一种利用多尺度的马尔可夫随机场(Markov random field, MRF)模型的方法^[10]在良好条件下合成的素描效果较优,与文献[8-9]相比,该方法消除了一些模糊、失真的缺陷。在 MRF 模型的基础上,Zhang 等^[11]利用亮度和梯度之间的相容性来匹配相邻小块,从而实现了在不同姿态和光照条件下人脸素描的合成;Zhou 等^[12]将 MRF 模型扩展为可以合成新素描块的马尔可夫权重场模型(Markov weight field, MWF),较好地合成了有姿态等情况下的人脸素描。近年来,人们提出了一系列在图像小块局部学习照片-素描词典对来合成人脸素描的方法,一类假定学习得到的词典对之间的关系相同^[13-14],另一类假设词典对之间的关系不同^[15]。最近,Wang 等^[16]采用一种贝叶斯推断的概率模型来估算待合成的素

描图,可以得到比已有方法更好的人脸素描合成结果。然而,与艺术家绘制的素描肖像画相比,以上这些统计学习方法合成的人脸素描有如下不足之处:一是由于没有很好地捕捉个人特征,导致合成的素描和原始的人脸不够相像;二是合成的素描缺乏艺术性,部分方法合成的结果有明显的模糊和块效应^[8-9,13,15],绝大多数方法的结果由于过度平滑丢失了铅笔素描的排线风格笔触^[10-12,14]。

本文同样应用了基于样例的学习策略,提出一种简单而直接的方式从照片合成素描。本文从训练集中学习出照片及其对应素描之间的映射关系,将该映射赋予一幅新的人脸照片得到其对应的人脸素描。由于映射关系整体来讲是非线性的,因此本文选择在局部图像区域上处理该问题,而非直接学习全局的人脸结构。假设 2 个照片图像小块很相似时其对应的素描图像小块也很相似,则一旦学习得到映射关系,就可以由人脸照片推断出人脸素描对应的小块。在素描合成时,首先利用训练集中的照片图像小块表示待生成的照片图像小块,然后将同样的表示关系应用于训练集中的素描图像小块上,以估计待合成的素描图像小块。本文借助一些回归技术,如最小二乘(least squares, LS)^[17]、岭回归(ridge regression, RR)^[18]、lasso^[17],可以很好地表示人脸照片和人脸素描之间的映射关系。为了保持邻接小块之间的关联性,本文将人脸图像采样为有重叠的区域,划分的区域大小决定了要学习的人脸结构的尺度,这些图像小块的不同尺度大小以及块与块之间重叠的多少会造成不同的合成效果。我们在香港中文大学人脸素描数据库(Chinese university of Hong Kong face sketch database, CUFS)^[10]上验证了本文方法的有效性。从合成的素描结果上可以看出,相比之前的方法,本文方法更具有艺术性并且更能把握原始照片中人脸的特性。

本文是文献[19]的扩展,应用回归模型之后,引入 LIC 方法生成了有素描风格的纹理,与局部回归模型合成的初步结果叠加增强了最终合成人脸素描的艺术感;增加了在 AR^[20]、XM2VTS^[21]照片-素描数据库上从照片到素描的合成实验;增加了和马尔可夫权重场方法^[12]、半耦合词典学习方法^[14]等代表性方法的对比实验。

1 本文方法流程

自动肖像素描合成的任务可以描述如下:对于

一幅给定的人脸照片 P , 可以从 P 自动生成对应的人脸素描 S , 得到的 S 需要满足如下要求: 1) 具有铅笔素描的艺术特性, 具体表现为和训练集中的素描样例 $\{S'_j\}_{j=1}^N$ 的风格相近, 其中 N 是照片-素描训练的总数; 2) 和原始照片 P 相像, 即素描 S 和给定照片 P 可以识别为同一个人. 假设 P 和 S 之间的映射关系为 F , 用 P'_j 表示训练集中第 j 幅照片图像, S'_j 表示 P'_j 对应的素描图像, 满足如下约束

$$\begin{cases} S=FP, \\ P \sim^{\chi_P} \{P'_j\}_{j=1}^N, \text{ s. t. } \chi_P \approx \chi_S; \\ S \sim^{\chi_S} \{S'_j\}_{j=1}^N, \end{cases}$$

其中, χ_P, χ_S 表示从训练集到测试集之间的映射关系, 分别是人脸照片之间的映射关系和人脸素描之间的映射关系. 由于映射 F 在图像整体上的非线性, 且在局部图像小块上表达的人脸特征比在全局

图像上更为准确, 因此本文在局部图像小块上应用回归模型, 将每幅图像均匀划分为 n 个小块区域, 每个小块大小为 $m \times m$ 像素. 本文所实现的人脸素描自动生成的步骤如下:

Step1. 预处理. 所有的照片和素描图像首先转换成灰度图像, 然后对人脸图像进行归一化处理, 经旋转、平移以及放缩处理使得人脸的两眼中心位于固定位置.

Step2. 局部回归映射. 利用局部回归模型, 从训练集的小块 $\{\{p'_i\}_{i=1}^n\}_j\}_{j=1}^N$ 中寻找输入照片小块对应的表示 (标记为 $\hat{p}_i$), 要求 $\hat{p}_i$ 可以最好地表示 p_i 的特性, 优化求解得到的 $\hat{p}_i$ 对应的 $\hat{s}_i$ 可以用来表示对应的素描小块 s_i .

Step3. 素描感增强. 根据估计出的素描图像小块 $\{\hat{s}_i\}_{i=1}^n$ 组合拼贴成整幅合成的图像, 并增强素描纹理以得到最终的人脸素描图像 S .

根据以上的步骤合成素描的流程如图 1 所示, 下面详细叙述 Step2 和 Step3 的实现.

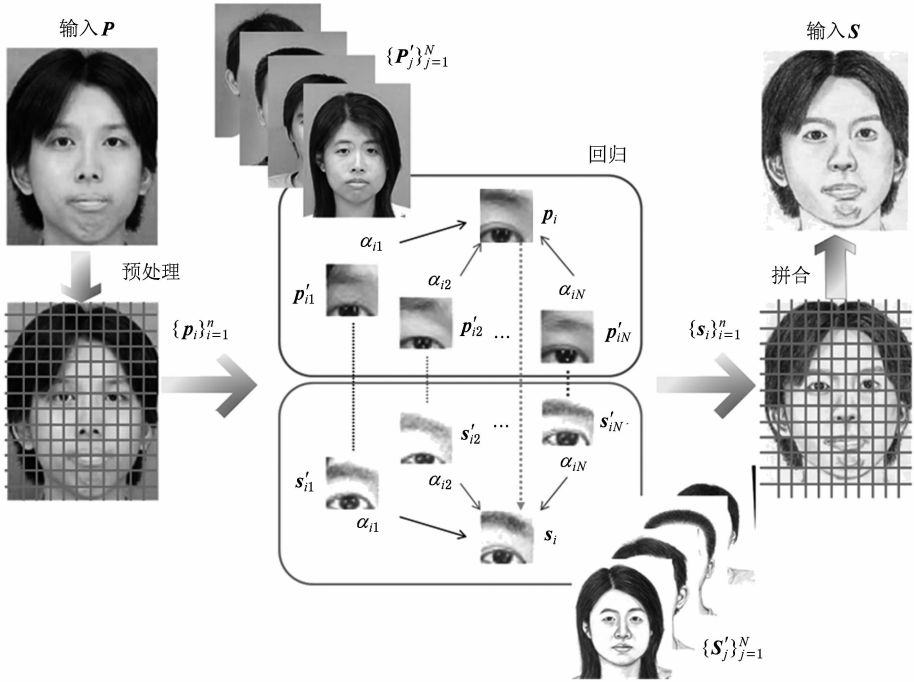


图 1 本文方法流程图

2 算法步骤描述

2.1 局部回归模型

在预处理步骤中, 原始的人脸图像被分为有重叠的小块区域, 图 2 所示为 2 种可能的重叠情况. 人脸照片图像小块之间的重叠约束着相邻小块之间的融合关系, 同时, 合成的素描图像的相邻小块像素要求在重叠区域有着近似的亮度值. 经过预处理, 人脸素描合成问题转换为: 对于已划分的输入人脸照片

图像的各个小块 p_i 对应的素描图像小块 s_i , 需要从训练集中的照片小块中寻找人脸照片小块 p_i 的最优表示 $\hat{p}_i$. 假设每个图像小块都可以由一系列相关联的图像小块线性表示, 若输入照片的第 i 个小块 p_i 和训练样本中找到的某个小块 p'_{ij} 有着相似的外观, 则 p'_{ij} 对应的素描小块 s'_{ij} 被认为是待合成的素描小块 s_i 最优的一个候选. 因此, 本文构建局部回归模型, 在训练集中寻找每个输入人脸照片小块的最优表示, 从而最优地构建出待合成的各个人脸素描小块.

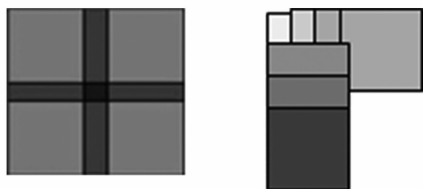


图 2 相邻小块之间 2 种不同的重叠模式

1) 近邻回归

构建 k 近邻(nearest neighbors, NN)回归模型来从训练集中选取最相近的照片图像小块以重构各个素描小块. 对于一幅输入照片的第 i 个小块 $\mathbf{p}_i$, 从训练集中选择其最相近的 k 个小块 $\mathbf{p}'_{ij}$ 来重构 $\mathbf{p}_i$, 为训练集中每个照片小块赋予权重, 则第 j 个样本的权重 $\hat{\omega}_{ij}^{\text{NN}}$ 可以形式化为

$$\hat{\omega}_{ij}^{\text{NN}} = \begin{cases} \text{SIM}(\mathbf{p}_i, \mathbf{p}'_{ij}), & \mathbf{p}'_{ij} \in k\text{NN of } \mathbf{p}_i \\ 0, & \text{otherwise} \end{cases};$$

其中,

$$\mathbf{p}'_{ij} = \frac{1}{N_j} \sum_{l=1}^{N_j} \mathbf{p}'_{ij,l},$$

且 $\text{SIM}(\cdot)$ 是小块 $\mathbf{p}_i$ 和 $\mathbf{p}_j$ 的相似度(similarity, SIM)度量. 本文使用了 2 种相似度度量方法, 将在第 2.2 节中讨论.

2) LS 回归

NN 回归方法仅仅使用一个或多个近似的小块去替代输入照片的小块, 会失去一些显著特征, 原因是输入照片图像小块的某些特征在有限的训练样本集中不一定存在一致的图像小块. 为了获得更多的局部特征, 本文使用线性回归模型

$$\hat{\mathbf{p}}_i = \sum_{j=1}^N \mathbf{p}'_{ij} \hat{\alpha}_j = \mathbf{p}'_i^T \hat{\alpha} \quad (1)$$

来寻找输入照片小块 $\mathbf{p}_i$ 更好的表示. 其中, $\hat{\alpha} = \{\alpha_j, j=1, \dots, N\}$ 是系数向量.

本文采用 LS 方法来较优地拟合训练集中的局部小块, 该方法是迄今为止最流行的线性方法. 对系数 α 求取最小化的乘方残差(residual sum of squares, RSS)

$$\text{RSS}(\alpha) = \left\| \mathbf{p}_i - \sum_{j=1}^N \mathbf{p}'_{ij} \alpha_j \right\|^2 =$$

$$\sum_{i=1}^m (\mathbf{p}_{it} - \mathbf{p}'_{it}^T \alpha)^2 = (\mathbf{p}_i - \mathbf{P}\alpha)^T (\mathbf{p}_i - \mathbf{P}\alpha).$$

令 $\mathbf{P}'$ 表示一个 $m \times n$ 的矩阵(m 为图像小块大小, n 为一幅图像小块的数目), 其中每列为训练集中一幅照片图像对应的小块向量, 即 $\mathbf{p}_i$ 是一个 n 维的向

量. 当 $\mathbf{P}'^T \mathbf{P}'$ 非奇异时, RSS 的唯一解为

$$\hat{\alpha}^{\text{LS}} = (\mathbf{P}'^T \mathbf{P}')^{-1} \mathbf{P}'^T \mathbf{p}_i \quad (2)$$

将式(2)代入式(1), 可以得到输入照片的第 i 个小块 $\mathbf{p}_i$ 对应的最匹配的估计 $\hat{\mathbf{p}}_i$.

3) RR

由于 LS 方法是一个离散过程, 不可避免地会出现一些噪声和模糊效应, 本文引入收缩方法来减少整个模型的预测误差. 通过对系数增加惩罚项来收缩回归系数, RR 或 lasso 回归这 2 种方法可以得到更优的回归系数, 从而减少合成结果的假缺陷. 其中, RR 通过增加正则项而达到偏差-方差均衡

$$\hat{\alpha}^{\text{RR}} = \arg \min_{\alpha} \left\{ \left\| \mathbf{p}_i - \sum_{j=1}^N \mathbf{p}'_{ij} \alpha_j \right\|^2 + \lambda \sum_{j=1}^N \alpha_j^2 \right\} \quad (3)$$

将式(3)写成矩阵形式为

$$\text{RSS}(\lambda_{\text{RR}}) = (\mathbf{p}_i - \mathbf{P}'\alpha)^T (\mathbf{p}_i - \mathbf{P}'\alpha) + \lambda_{\text{RR}} \alpha^T \alpha,$$

推导求解 RR 情况下最优的回归系数

$$\hat{\alpha}^{\text{RR}} = (\mathbf{P}'^T \mathbf{P}' + \lambda_{\text{RR}} \mathbf{I})^{-1} \mathbf{P}'^T \mathbf{p}_i,$$

其中 $\mathbf{I}$ 是一个 $m \times m$ 的单位阵.

4) Lasso 回归

将 RR 中的 l^2 惩罚项 $\sum_1^m \alpha_i^2$ 替换为 l^1 惩罚项

$\sum_1^m |\alpha_i|$, 即加入稀疏约束并令部分系数为 0, 则式(3)变为 lasso 回归的形式

$$\hat{\alpha}^{\text{lasso}} = \arg \min_{\alpha} \left\{ \frac{1}{2} \left\| \mathbf{p}_i - \sum_{j=1}^N \mathbf{p}'_{ij} \alpha_j \right\|^2 + \lambda \sum_{j=1}^N |\alpha_j| \right\}.$$

该 lasso 回归的求解是一个二次规划问题, 本文使用最小角回归算法^[22]来有效地求解该 lasso 系数.

将如上的回归模型作用到输入人脸照片小块以及训练集人脸照片小块上, 可以得到输入人脸照片小块与训练集中人脸照片小块之间的映射 χ_P . 由于训练集中包含了很好对应的照片-素描对, 即人脸照片上的映射关系可以复用到人脸素描上($\chi_P \approx \chi_S$, χ_S 表示人脸素描图像从训练集到测试集之间的映射关系), 并且输出的素描图像必须和输入照片近似, 因而本文将映射 χ_P 转移到训练集中的人脸素描上, 即可以预测出待合成的所有人脸素描图像小块. 其中, 第 i 个人脸素描小块

$$\hat{\mathbf{s}}_i = \hat{\alpha} \mathbf{s}'_i.$$

2.2 小块匹配度量

在 NN 回归的小块匹配步骤中, 可以有多种度量方式来衡量 2 个图像小块之间的相似程度. 本文采用了一种有效的度量方法——余弦相似度(cosine

similarity, CSIM)^[23],来度量 2 个图像小块之间的相似程度. CSIM 度量法中,2 个照片小块之间的相似度由 2 幅图像的亮度向量之间的夹角余弦求得. 具体地,图像小块 p_i 和 p'_i 之间的相似度

$$\text{SIM}^{\text{CSIM}}(p_i, p'_i) = \cos \theta = \frac{p_i \cdot p'_i}{\|p_i\|^2 \cdot \|p'_i\|^2}.$$

CSIM 度量法或经典的欧氏距离度量仅仅关注每个小块中像素的亮度值,与人的复杂的视觉感知有一定差距. 为了获得更高质量的合成结果,NN 回归模型在小块匹配步骤中同时考虑了图像小块的结构化信息. 进一步,本文使用结构相似性(structural similarity index, SSIM)度量法^[24]来衡量 2 个图像小块之间的相似度. 在 SSIM 度量法中,图像的保真度由亮度、对比度和结构 3 个不同的因素共同决定. 本文将 SSIM 度量应用在输入照片图像小块 p_i 和待匹配的照片图像小块 p'_i 之间,通常它们的相似度为

$$\text{SIM}^{\text{SSIM}}(p_i, p'_i) = \frac{(2\mu_{p_i}\mu_{p'_i} + c_1)(2\sigma_{p_i p'_i} + c_2)}{(\mu_p^2 + \mu_{p'_i}^2 + c_1)(\sigma_{p_i}^2 + \sigma_{p'_i}^2 + c_2)};$$

其中, $\mu_{p_i}, \mu_{p'_i}$ 是 p_i, p'_i 的均值, $\sigma_{p_i}^2, \sigma_{p'_i}^2$ 是 p_i, p'_i 的方差,且 $\sigma_{p_i p'_i}$ 是 p_i, p'_i 的协方差, c_1, c_2 是结构化相似度所包含的参数. 本文将在第 3.1 节中展示使用 SSIM 度量法和 CSIM 度量法对合成素描结果的对比.

2.3 素描感增强

在应用回归模型之后,通过进一步后处理可以

得到最终接近画家手绘风格的人脸素描图. 当待合成人脸素描的各个小块获得之后,需要将这些小块拼合成整个素描图像. 为了减少模糊效应,在合成时本文对图像小块与小块之间重叠部分的像素值计算平均的亮度,从而保证邻接小块之间有着平滑的亮度变化. 经过如上处理,可以得到有较少块效应的拼合图像.

然而,以上小块拼合后的图像会因为多次平滑而丢失素描排线笔触,且面部轮廓、五官等关键特征被弱化. 因此,本文引入 LIC 算法^[25]来模拟铅笔排线风格的纹理,并进一步提取边缘特征以增强面部轮廓以及五官特征. 如图 3 所示,对原始人脸照片的灰度图应用阈值化处理求取对应的白噪声图,并构建向量场来反映铅笔排线方向,对白噪声图和向量场应用 LIC 算法得到有铅笔排线样式的素描纹理图. 不同于之前利用 LIC 算法素描化直接叠加纹理和边缘的方法^[3],本文是在局部回归模型得到的合成图像上应用 LIC 算法模拟的素描纹理增强. 其中,向量场的构建是对图像过分割得到的局部区域分别求取各子区域的方向,采用一种吸附边缘很好的超像素分割法^[26]划分区域,同时利用 SIFT 算子^[27]统计子区域的方向. 最终,通过叠加模拟的素描纹理图 I_T , Canny 边缘提取图 I_E 到局部回归模型得到的合成图像 I_L 上,调整明暗,得到素描感增强的人脸素描图 $I_S, I_S = \varphi_1 I_T + \varphi_2 I_E + \varphi_3 I_L$.

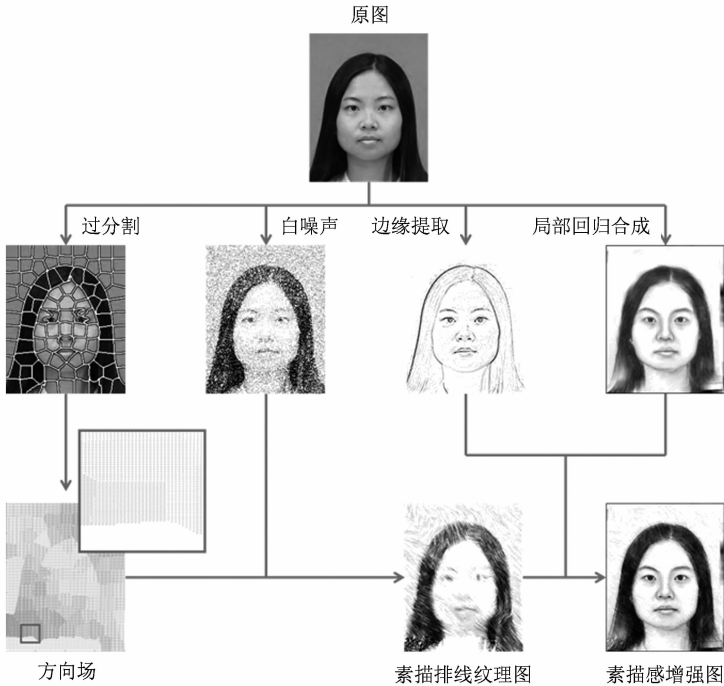


图 3 LIC 算法素描感增强示意图

3 实验及结果分析

本文在 CUFS 数据库^[10]上进行实验,以验证本文算法的有效性. CUFS 数据库是人脸素描合成研究工作中通用的一个数据库,库中的每个人有 1 幅正面姿态、正面光照、中性表情且没有遮挡情况下的人脸照片,同时有 1 幅对应的由画师根据此照片绘制的人脸素描. 本文从 CUFS 数据库中的 CUHK 库中选取 88 个人的照片-素描对作训练,余下的 100 个人的照片-素描对用作测试;另外,对 CUFS 数据库中的 AR 数据库^[20]中的 195 对以及 XM2VTS 数据库^[21]中的 295 对照片-素描图像采用留一测试. 在预处理时,以两眼中心(75,125)以及(125,125)来对齐人脸区域,并将人脸图像统一归一到 200×250 像素大小;然后将这些图像分割成有着不同大小且不同重叠方式的图像小块,以建立后续的局部回归模型. 本文所有实验是在 2.93 GHz Core 2 Duo CPU,

4 GB 内存的 PC 机上用 Matlab 实现的,整个素描合成过程仅需要 1~10 min(精确的合成时间取决于具体选取的参数).

3.1 不同参数情况下的合成结果

本节给出了本文提出的局部回归模型在不同参数、未经过素描感增强情况下的合成结果. 图 4 所示为使用 4 种不同回归方法合成的人脸素描结果,图 4c~4f 的各行是采用不同的小块大小和小块重叠大小的结果. 其中,图 4c 合成的人脸素描的参数 $k=1$,小块匹配度量使用 SSIM 度量法,图 4e,4f 的正则参数 $\lambda=0.1$. 第 1 行的小块大小是 20×20 重叠大小是 15 像素,第 2~4 行的小块大小分别为 30×30 , 40×40 和 50×50 且都采用 20 像素的重叠大小. 可以看出,图 4c,4f 的素描合成结果从人的视觉感知上看起来像一幅素描画像,其中图 4c 中分割的小块小些(例如 20×20)、图 4f 中分割的小块大些(例如 40×40)的时候块效应较少,结果看起来自然比较贴近原始人脸照片.

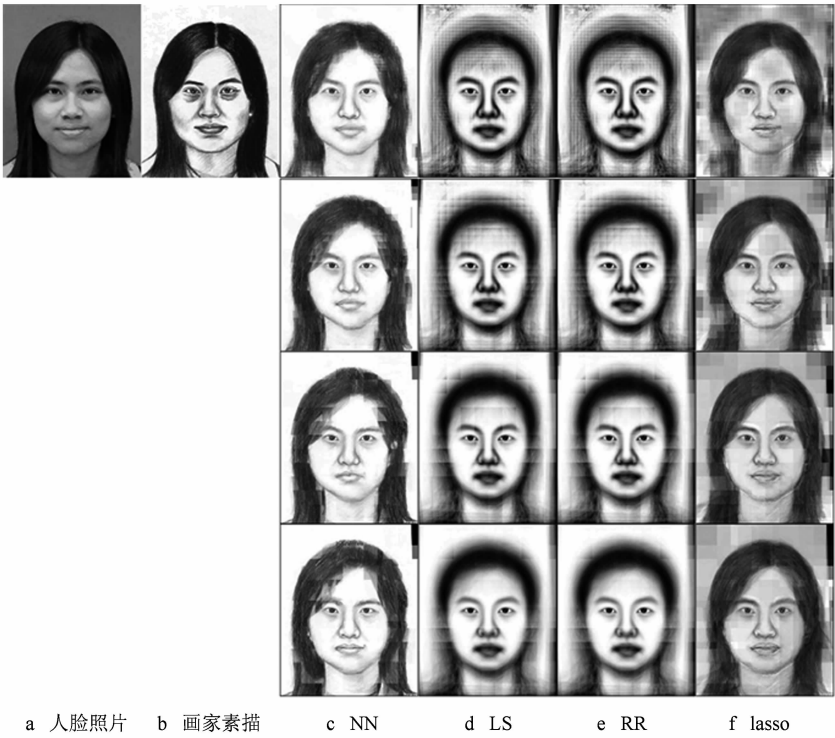


图 4 4 种回归方法合成的素描比较

图 5 所示为更多的 NN 方法下的结果,其中图像划分的小块大小比图 4 的情况下小. 图 5c~5f 中的各行的小块之间的重叠大小分别为 3,4,14 和 18. 由于图像小块重叠的大小必须 $\leq$ 图像划分的小块本身的大小,导致一些不存在的结果在图中留白. 当小块的重叠大小为定值时,如果小块大小较小(例

如 5×5),更多的细节信息会保留,但此时合成的结果噪声过多,且合成时间会因小块数目的过多而拉长;而当小块的大小为定值时,马赛克状的块效应会随着小块重叠大小的增加而减少. 总的来说,当小块大小为 15×15 、重叠大小为 14 时,以及小块大小为 20×20 、重叠大小为 18 时的结果令人满意,且每幅

素描像的合成时间 ≤ 10 min.

素描结果. 由于使用整幅图像全局构造素描图像时会丢失一些细节特征, 图 6 c 所示为 lasso 回归情况

图 6 所示为更多的使用 lasso 回归合成的肖像

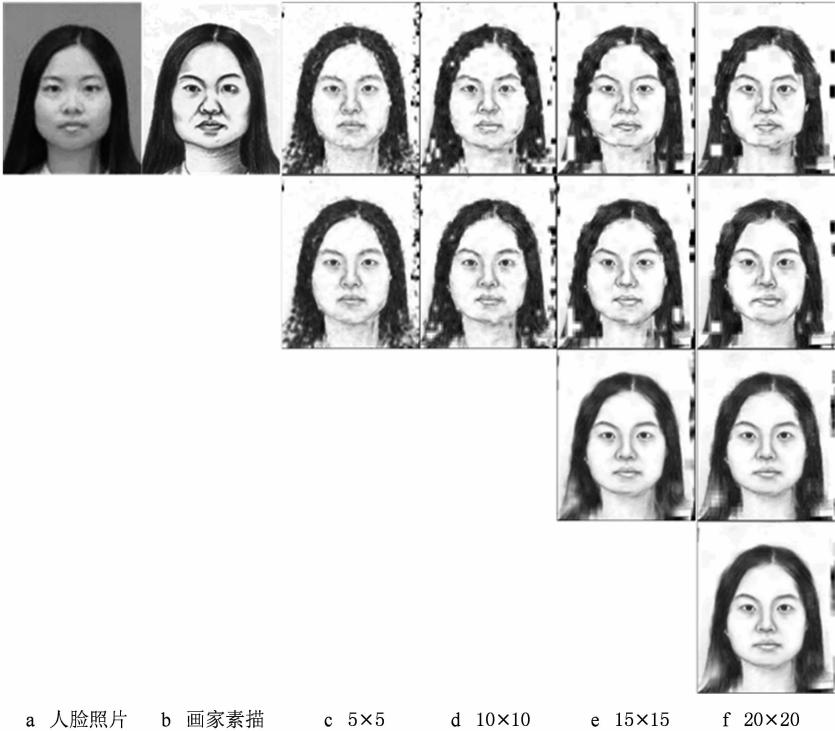


图 5 不同参数时 NN 回归合成素描的比较

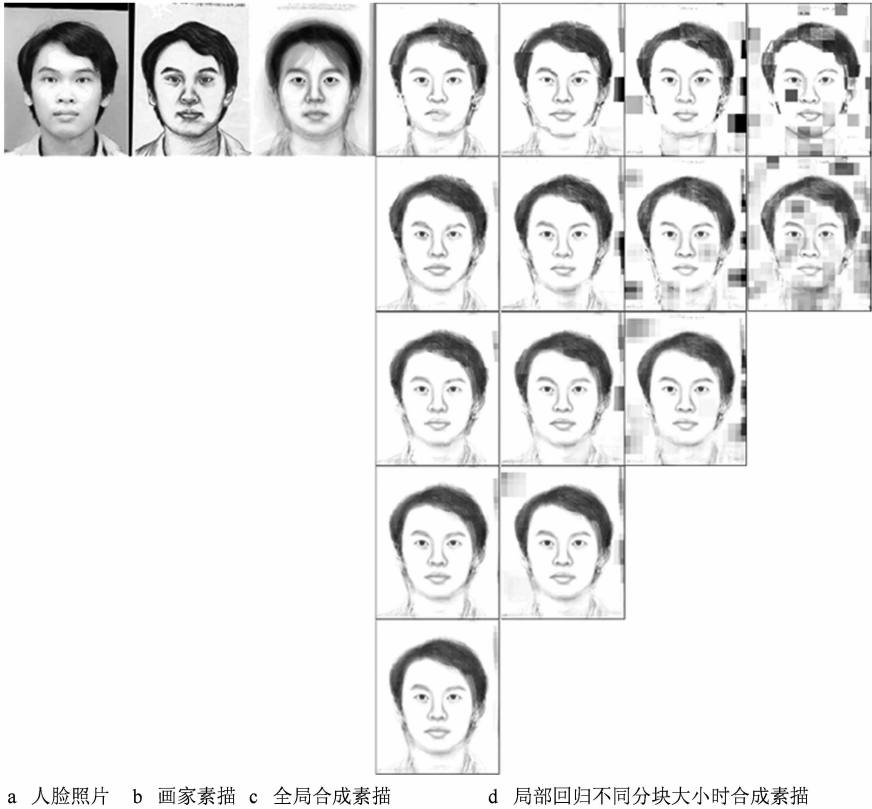


图 6 不同参数时 lasso 回归合成素描的比较

下的全局合成结果,合成的素描和原始人脸照片非常不相像且有很严重的重影效应.因而,本文在局部图像小块上应用 lasso 回归方法,图 6 d 中各列的分块大小分别为 50×50 , 40×40 , 30×30 和 20×20 ,而各行的小块重叠大小为 $0 \sim 40$ 像素(以 10 像素逐渐递增).可以看出,当图像小块大小为 50×50 以及 40×40 ,重叠大小为 30 时结果很好.与 NN 方法在分割的小块较小时结果较优不同,lasso 回归方法在分割的小块较小时(例如 20×20)会在局部小块上产生不自然的人工缺陷.比较 lasso 回归和 NN 回

归 2 种方法合成的素描效果可以看出,lasso 回归合成的素描比 NN 方法损失一些细节特征,这是由于 lasso 方法多个小块融合造成的过度平滑效应.尽管如此,lasso 回归耗时少的优势对于肖像素描的合成仍然适用,这是因为 NN 方法在小块匹配过程中会耗费大量的时间.

图 7 所示为使用 lasso 回归时,正则化参数 λ 不同时各种素描的合成结果,可以看出,由于 lasso 回归比较强的稀疏性,正则化参数的取值对合成结果的影响很小.

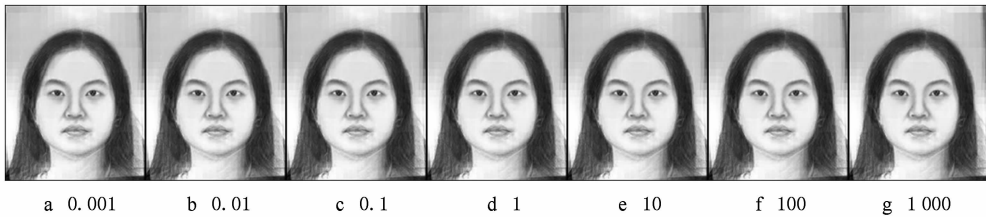


图 7 不同 λ 时 lasso 回归合成素描的比较

本文对比了 NN 方法在小块匹配过程中使用第 2.2 节的 2 种度量方法合成的素描结果,如图 8 所示.图 8 中素描合成结果均是在小块大小为 20×20 、小块重叠大小为 4 的情况下实现的,其中使用 SSIM 度量^[24]时图像滤波参数设定为 $K = [0.01$

$0.03]$, $window = \mathbf{I}_{4 \times 4}$, $L = 100$. 可以看出,用 CSIM 衡量小块相似性会造成比 SSIM 度量更多的错误匹配结果,尤其是在高亮区域以及纯黑色的头发区域.由此可以验证,在度量 2 个图像小块的相似度时,结构化信息的引入会比只使用图像亮度的度量更有效.



图 8 2 种度量方法对素描合成结果影响的对比

作为文献[19]的扩展,本文引入的 LIC 等局部特征增强方法可以增强最终合成人脸素描的艺术感.图 9 所示为本文方法与文献[19]的对比结果,其中设定的经验参数如下:线积分卷积的流线长度

$length = 5$,各合成图的叠加系数 $\varphi_1 = \varphi_2 = 0.2$, $\varphi_3 = 0.5$.可以看出,局部特征增强之后的人脸面部五官的特征更为鲜明,头发有着更清晰准确的光照和纹理特征,例如,女性头发分发处的分界线更为清晰,

男性的卷发纹理更为清晰准确。



a 人脸照片 b 局部回归结果 c 素描感增强结果

图 9 与文献[19]素描合成结果的对比

3.2 和其他方法的素描合成结果的对比

本节对比了本文提出的局部回归方法和之前有代表性的人脸素描合成方法合成的素描结果在主观上的不同。为了和其他方法展示的结果保持一致以利于对比,本文将图像裁剪为 155×200 的大小,即从顶部、底部、左侧、右侧分别裁剪 25,25,25 和 20 像素。

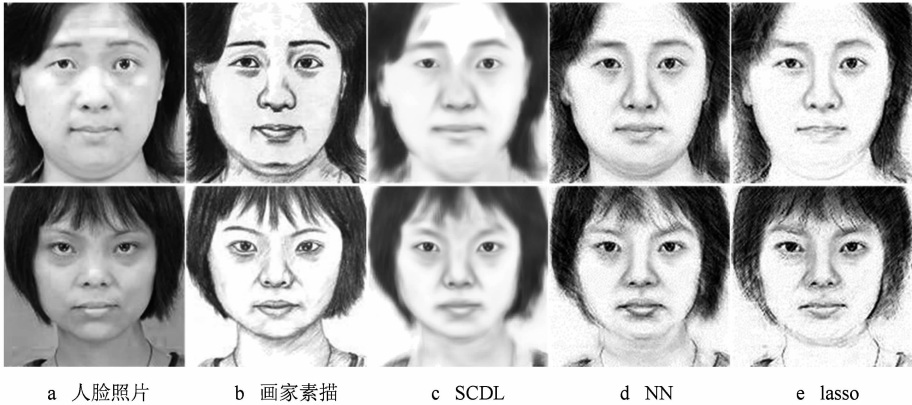
图 10 所示为与全局特征变换(global eigen-transformation, GE)^[8]和局部线性嵌入法(locally linear embedding, LLE)^[9]2 种方法的比对。可以看出,与 GE 和 LLE 方法相比,本文方法可以较好地合成头发区域,且本文中合成的人脸区域更清晰、没有噪声干扰。



d LLE e NN f lasso

图 10 和 GE,LLE 方法合成素描结果对比

图 11 所示为与最具代表性的稀疏表示的双词典学习方法,即半耦合词典学习(semi-coupled dictionary learning, SCDL)方法^[15]的对比。为了与文中给定的



a 人脸照片 b 画家素描 c SCDL d NN e lasso

图 11 和 SCDL 方法合成素描结果对比

结果一致,统一裁剪图像至 150×150 的大小.可以看出,SCDL 方法得到的面部区域比较浅淡不清晰,非常类似于原始照片的灰度化效果;而本文合成的素描结果五官清晰且明暗分明,更具有素描画像的艺术化特性.

图 12 所示为与马尔科夫模型相关方法合成的人脸素描在 AR 库上对比结果的展示. 尽管 MRF 方法^[10]是最经典的人脸素描合成方法,但其合成结果偏向平均人脸而与原始照片不够相像,且丢失了不少素描笔触纹理. MWF 方法^[12]合成的五官比 MRF 方法刻画更为精确,但同样存在缺乏素描艺术感的缺陷. 特别地,从眉毛的刻画以及阴影区域可以看出, MRF 方法的结果由于多层的重叠显得特别光滑,而 MWF 方法的合成结果在放大的情况下可以看到在边缘很多地方有明显的块效应. 相比较而言,本文合成的结果和画师的手绘结果更为接近,表现为保留了更多的铅笔素描的排线风格笔触,且在五官等特征的刻画上更为清晰.



图 12 和 MRF, MWF 方法合成素描结果对比

图 13 中展示了更多的从人脸照片合成素描的结果,可以看出,本文提出的局部回归方法在这 3 个经典的照片-素描库上都能得到不错的结果.

CUHK 库



AR 库



XM2VTS 库



a 人脸照片 b 画家素描 c 本文合成素描

图 13 更多的素描合成结果

4 总 结

本文提出一种基于样例的学习方法——基于局部回归模型的素描肖像画生成方法. 该方法以一种简单而直接的方式从人脸照片合成素描,采用局部回归模型为主的数据驱动方式并融合笔触增强的物理模拟方式;从训练集中学习出给定人脸照片的映射关系,将该映射关系迁移到待合成的人脸素描. 本文方法可以在几分钟内自动合成一幅给定的人脸照片

对应的人脸素描. 实验结果显示, 本文方法比已有的素描合成方法更能体现素描像的艺术性并保留与原照片的相似性.

本文方法目前仅仅考虑了正面姿态、较优光照情况, 光照、姿态以及表情变化情况下的人脸照片对应的素描合成还有待后续研究; 另外, 一些增强合成素描像的艺术感和自然性的工作有待进一步尝试. 今后, 我们将考虑引入铅笔笔触图形化的渲染来增强素描纹理并减轻合成块效应.

参考文献 (References):

- [1] Salisbury M P, Anderson S E, Barzel R, *et al.* Interactive pen-and-ink illustration [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 1994: 101-108
- [2] Salisbury M P, Wong M T, Hughes J F, *et al.* Orientable textures for image-based pen-and-ink illustration [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 1997: 401-406
- [3] Mao X Y, Nagasaka Y, Imamiya A. Automatic generation of pencil drawing using LIC [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 2002: 149
- [4] Xie Dangen, Zhao Yang, Xu Dan. A method for generation of pencil filter and its implementation on GPU [J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(1): 26-31 (in Chinese)
(谢党恩, 赵 杨, 徐 丹. 一种铅笔滤镜生成算法及其在 GPU 上的实现[J]. 计算机辅助设计与图形学学报, 2008, 20(1): 26-31)
- [5] Mo Xiaofei, Ding Youdong. Pencil sketch generation with shape features [J]. Journal of Image and Graphics, 2013, 18(2): 219-224 (in Chinese)
(莫晓斐, 丁友东. 利用形体特征的铅笔素描画生成[J]. 中国图象图形学报, 2013, 18(2): 219-224)
- [6] Lu C W, Xu L, Jia J Y. Combining sketch and tone for pencil drawing production [C] //Proceedings of the Symposium on Non-photorealistic Animation and Rendering. Aire-la-Ville: Eurographics Association Press, 2012: 65-73
- [7] Wang N N, Tao D C, Gao X B, *et al.* A comprehensive survey to face hallucination [J]. International Journal of Computer Vision, 2013, 106(1): 9-30
- [8] Tang X O, Wang X G. Face sketch synthesis and recognition [C] //Proceedings of the 9th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2003: 687-694
- [9] Liu Q S, Tang X O, Jin H L, *et al.* A nonlinear approach for face sketch synthesis and recognition [C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005, 1: 1005-1010
- [10] Wang X G, Tang X O. Face photo-sketch synthesis and recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(11): 1955-1967
- [11] Zhang W, Wang X G, Tang X O. Lighting and pose robust face sketch synthesis [M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2010, 6316: 420-433
- [12] Zhou H, Kuang Z H, Wong K Y K. Markov weight fields for face sketch synthesis [C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 1091-1097
- [13] Chang L, Zhou M Q, Han Y J, *et al.* Face sketch synthesis via sparse representation [C] //Proceedings of the 20th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2010: 2146-2149
- [14] Wang N N, Gao X B, Tao D C, *et al.* Face sketch-photo synthesis under multi-dictionary sparse representation framework [C] //Proceedings of the 6th International Conference on Image and Graphics. Los Alamitos: IEEE Computer Society Press, 2011: 82-87
- [15] Wang S L, Zhang L, Liang Y, *et al.* Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis [C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2216-2223
- [16] Wang N N, Tao D C, Gao X B, *et al.* Transductive face sketch-photo synthesis [J]. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24(9): 1364-1376
- [17] Hastie T, Friedman J, Tibshirani R. The elements of statistical learning [M]. Heidelberg: Springer, 2001
- [18] Tibshirani R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society Series B: Methodological, 1996, 58(1): 267-288
- [19] Ji N Y, Chai X J, Shan S G, *et al.* Local regression model for automatic face sketch generation [C] //Proceedings of the 6th International Conference on Image and Graphics. Los Alamitos: IEEE Computer Society Press, 2011: 412-417
- [20] Martinez A M. The AR face database [R]. Columbus: The Ohio State University, CVC Technical Report 24, 1998
- [21] Messer K, Matas J, Kittler J, *et al.* XM2VTSDB: the extended M2VTS database [C] //Proceedings of the 2nd International Conference on Audio and Video-based Biometric Person Authentication. Heidelberg: Springer, 1999: 72-77
- [22] Efron B, Hastie T, Johnstone I, *et al.* Least angle regression [J]. The Annals of Statistics, 2004, 32(2): 407-499
- [23] Tan P N, Steinbach M, Kumar V. Introduction to data mining [M]. Boston: Addison-Wesley Longman, 2005

[24] Wang Z, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612

[25] Cabral B, Leedom L C. Imaging vector fields using line integral convolution [C] //Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH. New York: ACM Press, 1993: 263-270

[26] Achanta R, Shaji A, Smith K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282

[27] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110