

天津大学

毕业设计（论文）说明书



学院名称 计算机科学与技术

专 业 计算机科学与技术

年 级 2007

姓 名 张翔

指导教师 刘世光

2011 年 6 月 16 日

天津大学

毕业设计（论文）任务书

题目：灰度图像与视频自动着色技术研究

学生姓名 张翔

学院名称 计算机科学与技术

专 业 计算机科学与技术

学 号 3007216026

指导教师 刘世光

职 称 副教授

一、原始依据（包括设计或论文的工作基础、研究条件、应用环境、工作目的等。）

目前在各种影视、图片档案、医疗等领域中存在着大量的灰度或者黑白图像和视频资料，将它们彩色化能够增强细节特征，帮助人们更好地识别和利用。然而，传统的手工着色方法耗费大量的人力物力，而且未必能得到良好的效果。如果在提供一幅物象和内容类似的彩色图像的条件下，让计算机自动地对目标灰度图像或者视频进行合理可信的着色，将大大提高这项工作的效率。

互联网上有大量的图片和视频资源，能够为我们的研究提供素材；同时，在目前的学术检索系统中已有不少对灰度图像着色方法的论文文献，可以为我们研究方案的设计和实现提供参考；此外，多种成熟的基于计算机视觉的图像分析和视频处理手段亦能够为我们系统的实现提供重要的借鉴。

该课题的研究能够提高电视节目编辑和影视制作特效的效率，改进生物医学图片（如 X 光显影图片、MRI 核磁共振扫描图片）的辨识度，还可以对物理、化学中的实验结果进行可视化增强，在摄像摄影和可视化产业的应用前景也十分广阔。

此项工作能够使学生了解计算机图像处理和计算机视觉技术的应用，促使学生对掌握相关领域的基本知识，增强实践锻炼和创新能力。

二、参考文献

- [1]Charpiat G, Hofmann M., Schölkopf B, Image Colorization via Multimodal Predictions[A], Proceedings of the 10th European Conference on Computer Vision[C], 2008, 126-139.
- [2]Li J, Hao P, Transferring Colours to Grayscale Images by Locally Linear Embedding[A], Proceedings of the British Machine Vision Conference[C], Leeds, UK, 2008.835-844.
- [3]Morimoto Y, Taguchi Y, Naemura T, Automatic Colorization of Grayscale Images using Multiple Images on the Web[A], Proceedings of SIGGRAPH Posters[C], 2009 Article 32.
- [4]Bae S, Paris S and Durand F, Two-scale Tone Management for Photographic Look[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2010.
- [5]Paris S and Durand F, A Fast Approximation of the Bilateral Filter using a Signal Processing Approach[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2006.
- [6]贾云涛, 胡事民. 基于图切分的交互式图像染色算法[A]. 计算机学报[J], 2006, 29(3): 508-513.
- [7]胡伟, 秦开怀. 高分辨率灰度图像的快速多分辨率着色[A]. 计算机学报[J], 2009, 32(5): 1062-1068.

三、设计（研究）内容和要求（包括设计或研究内容、主要指标与技术参数，并根据课题性质对学生提出具体要求。）

该课题设计内容主要包括以下几点：

1. 阅读国内外相关领域的学术论文，掌握方法和原理，有所见地；并据此进行算法设计和实现；
2. 基于图像颜色和纹理统计信息的自动分区技术和自动匹配技术研究并实现图像的自动着色；
3. 结合视频处理技术实现灰度视频的自动着色技术；
4. 与目前已知算法在同样实验材料条件下的结果比较与可信度分析；在一定的计算机硬件条件下测试的内存和时间性能，并根据所测结果进一步改进算法性能；
5. 综合上述算法，采用 Matlab 或者 C/C++进行编程实现，并设计合理的用户操作界面。

指导教师（签字）

年 月 日

审题小组组长（签字）

年 月 日

天津大学本科毕业设计（论文）开题报告

课题名称	灰度图像与视频的自动着色技术研究		
学院名称	计算机科学与技术	专业名称	计算机科学与技术
学生姓名	张翔	指导教师	刘世光
<p>（内容包括：课题的来源及意义，国内外发展状况，本课题的研究目标、研究内容、研究方法、研究手段和进度安排，实验方案的可行性分析和已具备的实验条件以及主要参考文献等。）</p> <p>一、来源及意义：</p> <p>近年来，在计算机图形学和图像处理领域中，灰度图像和视频的着色技术得到了广泛的研究。但是，现有的方法大多需要手工交互，费时费力并且精度不高，这严重影响了着色技术在影视制作、游戏动漫等方面的实际应用。本项目拟基于微分几何和图像颜色编辑技术，研究灰度图像和视频的自动着色技术。现有的基于颜色迁移、色彩统计和图像分区的方法，为我们自动化的灰度图像着色方案提供了参考。互联网络上大量的图像资源可以为我们的算法实现提供丰富的素材。</p> <p>二、国内外发展状况：</p> <p>灰度图像和视频的着色技术研究是一个非常具有挑战性的课题。当前已有较为广泛的研究，但现有的方法大多数是半自动方法，它们需要用户进行交互一定的交互操作；仅有的一些全自动着色的方法，也存在需要多幅颜色来源图像、精度不高和计算速度慢等问题，而且视频的全自动着色还少有研究。</p> <p>在全自动图像着色方法的研究方面，Charpiat 等人^[1]提出一种基于多模型预测的着色方法。但这一方法仍旧需要用户提供颜色标记来保证着色结果的有效性。Li 等人^[2]提出一种基于机器学习的方法。该方法将训练数据中的输入灰度样本和输出彩色图像样本当作特征向量，分布于两个流型中并假设它们具有相似的结构。这一自动着色的方法可以用于批量的图像处理，但它需要有足够大的训练样本数据才能保证着色的结果。Morimoto 等人^[3]实现了基于多幅图像来源的自动灰度图像着色方法，这些图像来源是从互联网上获取的。该方法受制于从网络上搜索得到的成千上万的图像样本，因此当颜色来源图像的结构类似但内容不同时会造成不符合实际的着色结果。</p> <p>三、研究目标：</p> <p>基于图像颜色和纹理统计方法与计算机视觉学习算法的有关理论，研究灰度图像和视频的自动着色算法，建立着色结果评估与分析模型，并实现相关的程序系统。</p>			

四、研究内容：

根据课题的研究目标，本课题的研究内容主要分为图像的自动着色和视频的自动着色两部分，并对着两部分构造相应的数学模型和优化方法，具体而言：

1. 基于图像颜色和纹理统计信息的自动分区技术和自动匹配技术研究；
2. 基于视频处理技术的视频自动着色算法研究；
3. 建立着色结果的比较分析与评估模型；
4. 算法的加速优化及实时着色算法研究；
5. 程序实现及用户界面技术研究。

五、研究方法：

图像自动着色和视频自动着色两部均需要不同的研究方法，同时，这两部分的结合也需要有相应的数学工具进行实验，因此本课题的研究方法主要有：

1. 应用图像局部熵、局部对比度、纹理度和亮度直方图等图像颜色和纹理统计信息，研究图像分区的自动匹配技术；
2. 应用视频处理的多种光流运动分析和关键帧检测技术，结合降维学习和聚类匹配算法，对视频进行逐帧自动匹配和着色，并充分利用视频的内容连续进行算法优化；
3. 结合算法评估和用户评估，对着色结果给出评估模型，并基于此模型进行与其它算法的比较分析；
4. 利用计算机硬件在快速傅里叶变换、高维向量化处理方面的优势研究算法的加速技术；
5. 结合 Matlab 数学计算工具、OpenCV 程序库等编程环境，开发可用的应用程序并设计相应的用户界面。

六、研究手段：

本课题需要利用相应的数学软件和编程语言进行建模效果的实验和改进，同时本课题还需要有相应的数学与算法理论进行分析，因此其研究手段主要有：

1. 基于 Matlab 数学软件平台进行算法建模和试验，高效地实践上述各种算法方案；
2. 结合数学模型、统计学、优化学、算法设计等数学和计算机理论，严格推导和阐释所研究的模型和算法；
3. 参考图形软硬件的优化处理方法，借助通用 CPU 和图形处理硬件提供的向量计算、矩阵处理、浮点运算等工具库，使用编程语言实现算法的加速优化。

七、进度安排：

根据指导老师的要求，本课题的进度安排主要是：

1. 2月28日-3月27日：阅读相关参考文献和技术资料，补充图像处理、视

频处理和计算机视觉相关理论知识,着手建立图像着色算法的数学模型并进行初步实现;

2. 3月28日-4月24日:完成视频算法数学模型的相关构思和实现,并对模型进行测试与完善,逐步改进算法的效果和效率;
3. 4月25日-5月22日:实现图像和视频自动着色的独立程序,建立效果评估模型,同其它算法进行比较,开始撰写最终论文;
4. 5月23日-6月19日:进一步完善算法和程序,在实际环境中对算法进行测试和评估,完成最终论文及其相关材料的整理,准备答辩。

八、实验方案的可行性分析和已具备的实验条件:

在计算机视觉领域本课题已经得到了初步研究,同时现存的各种理论工具也为其奠定了现实基础,具体而言:

1. 对颜色空间基本理论、时域和频域数字信号处理、统计学、优化学、数值分析、机器学习和计算机视觉等相关领域的常用方法和数学原理已有基础,并对相关的算法优化技术形成了初步认识;
2. 通过图像局部熵、局部对比度、主元分析法和纹理度等多种纹理分析手段实践了创新性的纹理分析方法,为下一步取得纹理和颜色一致性的着色方法提供了保证;
3. 基本建立无交互自动着色系统的数学模型,以该模型为基础可以基本保证算法能够同时保证上色的速度和纹理与颜色的一致性;
4. 目前在视频处理领域有多种光流运动分析和关键帧检测技术,可以应用于对视频自动着上色的研究中。

九、主要参考文献:

- [1]Charpiat G, Hofmann M., Schölkopf B, Image Colorization via Multimodal Predictions[A], Proceedings of the 10th European Conference on Computer Vision[C], 2008, 126-139.
- [2]Li J, Hao P, Transferring Colours to Grayscale Images by Locally Linear Embedding[A], Proceedings of the British Machine Vision Conference[C], Leeds, UK, 2008.835-844.
- [3]Morimoto Y, Taguchi Y, Naemura T, Automatic Colorization of Grayscale Images using Multiple Images on the Web[A], Proceedings of SIGGRAPH Posters[C], 2009 Article 32.
- [4]Bae S, Paris S and Durand F, Two-scale Tone Management for Photographic Look[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2010.
- [5]Paris S and Durand F, A Fast Approximation of the Bilateral Filter using a Signal Processing Approach[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2006.
- [6]贾云涛, 胡事民. 基于图切分的交互式图像染色算法[A]. 计算机学报[J], 2006, 29(3): 508-513.
- [7]胡伟, 秦开怀. 高分辨率灰度图像的快速多分辨率着色[A]. 计算机学报[J], 2009, 32(5): 1062-1068.
- [8]Mitchell T.M. Machine Learning[M]. McGraw-Hill, 1997. 236-238.

- [9]Zhang J. Face Recognition: Eigenface, Elastic Matching and Neural Nets[A], Proceedings of IEEE[J], Vol. 85, No. 9, September 1997
- [10]Chatzigiorgaki M., Skodras A.N. Real-time Keyframe Extraction towards Video Content Indentification[A], School of Science and Technology, Hellenic Open University, 26222 Patras, Greece.

选题是否合适： 是☐ 否☐

课题能否实现： 能☐ 不能☐

指导教师（签字）

年月日

选题是否合适： 是☐ 否☐

课题能否实现： 能☐ 不能☐

审题小组组长（签字）

年月日

摘 要

着色的目的是向灰度的图像或视频来源自动地添加颜色。这一问题本身具有不确定性,因为在没有任何先验知识的情况下向灰度图像或视频进行着色是歧义性的。大多数前人的方法使用或多或少的手工的操作来辅助着色过程,而且视频的着色技术还少有研究。本论文提出一种全自动的灰度图像着色方法,并基于这一方法来探讨灰度视频自动着色的实现。对于灰度图像的着色,首先假设用户能够提供一幅同目标灰度图像内容类似的彩色图片,然后计算局部化的图像纹理映射。通过对灰度图像和彩色图像的纹理与亮度分区和匹配,提取出局部和全局颜色对应关系,据此实现对灰度图像的加权着色。在视频的着色方面,假设用户能够提供与目标灰度视频内容类似的图像库,利用图像配准和视频剪辑算法,再结合关键帧技术实现速度较快的灰度视频着色。

关键词: 图像着色; 图像配准; 视频剪辑; 视频着色; 纹理度

ABSTRACT

Colorization aims to adding colors to a grayscale image or video. This task is ill-posed in the sense that assigning the colors to a grayscale image without any prior knowledge is ambiguous. Most of the previous methods require some amount of user interventions. This paper proposed a novel automatic grayscale image and video colorization method, based on the assumption that user can provide an image or a collection of images with similar contents as that in the target image or video. In image colorization, we compute the spatial map using textureiness method. And then, we perform automatic segmentation on both the spatial map and the intensity map, so that we can match these segmentations between the color source and the target image. With this information we can compute the local and global luminance-color correspondence, which can later be used directly for colorizing the target. In video colorization, we perform a simple image registration algorithm on the image collection, so that we can match a proper color source to each key-frame detected through our video clipping algorithm. After the colorization of key-frames, we can utilize the intermediate data to colorize rest of the frames.

Key words: Image Colorization; Image Registration; Video Clipping;
Video Colorization; Textureiness

目 录

第一章 概述.....	1
1.1 研究的目的和背景.....	1
1.2 相关工作	1
1.3 论文章节安排	2
第二章 灰度图像的自动着色.....	3
2.1 灰度图像自动着色模型设计.....	3
2.2 纹理映射和双边分解.....	3
2.3 统计自适应伽玛矫正.....	5
2.4 局部加权回归和映射分区.....	7
2.5 灰度图像自动着色的实现.....	9
第三章 灰度视频的自动着色.....	12
3.1 灰度视频自动着色模型设计.....	12
3.2 基于主元分析的图像配准与搜索	12
3.3 奇异值分解计算主元分析.....	14
3.4 灰度视频的自动剪辑.....	16
3.5 灰度视频自动着色的实现.....	18
第四章 性能分析与改进.....	19
4.1 快速双边滤波算法及其替代算法	19
4.2 对亮度-色彩对应关系的量化.....	21
第五章 结论与展望.....	23

5.1 总结	23
5.2 工作展望	24
参考文献.....	25
外文资料	
中文译文	
致谢	

第一章 概述

1.1 研究的目的是背景

目前在各种影视、图片档案、医疗等领域中存在着大量的灰度或者黑白图像和视频资料，将它们彩色化能够增强细节特征，帮助人们更好地识别和利用。然而，传统的手工着色方法耗费大量的人力物力，而且未必能得到良好的效果。如果在提供一幅物象和内容类似的彩色图像的条件下，让计算机自动地对目标灰度图像或者视频进行合理可信的着色，将大大提高这项工作的效率。

着色是一种计算机辅助的过程，目的是要对灰度图像或视频赋予色彩。它的应用主要体现在对黑白影像效果的提升上，这些影像包括黑白照片、旧式电影和科学成像结果等。由于不同的颜色可能会产生相同的亮度值，因此着色并不是一种有确定性解的问题。大多数当前的方法需要一定量的用户交互来产生满足需要的结果，但这不可避免地使得着色成为一个非常费时费力的过程。同时，这些方法也难于对着色结果进行有效的评估。本课题的目的是提出一种新的图像着色方法，同时基于这一方法探讨视频着色方法的实现，解决之前方法的问题并对新的方法做出分析与评估。

该课题的研究能够提高电视节目编辑和影视制作特效的效率，改进生物医学图片（如 X 光显影图片、MRI 核磁共振扫描图片）的辨识度，还可以对物理、化学中的实验结果进行可视化增强，在摄像摄影和可视化产业的应用前景也十分广阔。

1.2 相关工作

目前的着色方法主要是颜色迁移（Color Transfer）和图像类比（Image Analogies）这两种技术的延伸。这些方法均需要用户提供一幅或多幅与目标图像内容类似的彩色图像作为参考，我们将这些彩色图像称作颜色来源。尽管彩色图像和目标图像的内容类似性越强，着色的结果可能越好，但这些方法需要的内容匹配仍需要耗费大量的劳动和时间，使得灰度视频的自动着色技术无从谈起。

在全自动图像着色方法的研究方面，Charpiat 等人^[1]提出一种基于多模型预测的着色方法。但这一方法仍旧需要用户提供颜色标记来保证着色结果的有效性。Li 等人^[2]提出一种基于机器学习的方法。该方法将训练数据中的输入灰度样本和输出彩色图像样本当作特征向量，分布于两个流型中并假设它们具有相似的结构。这一自动着色的方法可以用于批量的图像处理，但它需要有足够大的训练样本数据才能保证着色的结果。Morimoto 等人^[3]实现了基于多幅图像来源的自动灰度图像着色方法，这些图像来源是从互联网上获取的。该方法受制于从网络上搜索得到的成千上万的图像样本，因此当颜色来源图像的结构类似但内容不同时会造成不符合实际的着色结果。

本论文的自动灰度图像着色方法在借鉴前人方法的基础上，解决了上色过程中亮度和纹理的一致性问题的。其中纹理的分析方法来源于 Bae 等人^[4]提出的“纹理度”，在本文的方法中称之为“纹理映射”。为了实现全自动的亮度与纹理映射分区，本文的方法还利用了局部加权回归（Locally Weighted Linear Regression, LWLR）^[8]的方法对直方图进行了处理，它是机器学习和数值计算学科的经典方法。在视频的处理方面，本论文首先利用主元分析方法（Principal Components Analysis, PCA）^[9]对图像集进行基于内容的配准，然后使用 Chatzigiorgaki 等人总结的简化关键帧检测与视频剪辑方法^[10]对视频进行剪辑，从而提取出对应的关键帧和剪辑片断，利用灰度图像自动着色方法对关键帧着色的数据来实现对整个视频的着色。

1.3 论文章节安排

本论文以下章节的安排如下：

第二章和第三章分别讨论灰度图像的自动着色和灰度视频的自动着色。第四章对算法中关键的部分进行性能分析并提出多种提升算法性能的方案，第五章则是对全文的总结。

第二章 灰度图像的自动着色

2.1 灰度图像自动着色模型设计

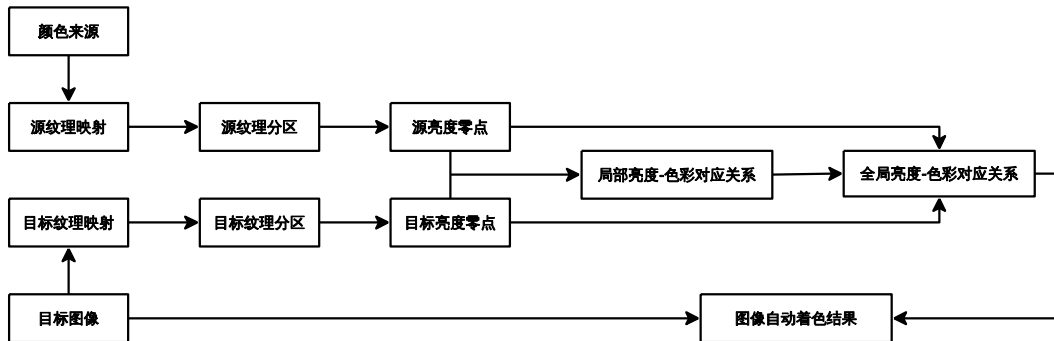


图 2-1：灰度图像自动着色模型

图 2-1 是灰度图像自动着色模型的主要框架。这里包含有七个部分，六个过程。其中，图像（包括目标图像和颜色来源）、纹理映射、纹理分区和亮度零点是给定的颜色来源和目标图像分别计算的。此模型的目的是利用颜色来源所提供的信息对目标图像进行着色，其基本假设是颜色来源和目标图像具有类似的内容。图像的纹理映射是对图像本身的函数，它表示的是每一像素附近的亮度空间分布特点。本课题定义亮度的空间分布特点是人类视觉对图像局部的抽象认知。实际上，本课题采用了 Soonmin 等人提出的纹理度^[4]方法作为纹理映射的表达。

纹理分区是基于纹理映射的分区结果。这从纹理映射到纹理分区这一步需要一种合适的分类或聚类方法。本论文采用的是简单的局部加权回归^[8]来对量化后纹理映射的直方图进行拟合，获取其极低点作为分区拆分点。在每一个纹理分区内部，为了能够对仅包含亮度信息的目标图像进行着色，还需要将色彩和亮度对应起来。本论文同样利用了局部加权回归来解决这一问题，即获取每个纹理分区内亮度直方图拟合零点，并将其作为颜色匹配的索引。这些零点在配适之后，便可以从颜色来源中获取每个亮度值对应的平均色彩。我们称之为每个纹理分区之内的局部亮度-色彩对应（Local Luminance-color Correspondence）。

全局亮度-色彩对应关系（Global Luminance-color Correspondence）有每个纹理分区内的局部亮度-色彩对应关系加权获得。它的实际上对每个纹理映射值和亮度值给出了一种颜色，从而能够利用这两个信息对一幅目标图像进行着色，并保证其色彩和纹理的一致性。

2.2 纹理映射和双边分解

纹理映射表示每一像素附近亮度空间分布的特征。纹理度表示的是图像的局部高频内容，即每一像素的局部纹理强度，因此纹理度是一种较好的纹理映射。为了获取纹理度，首先需要得到图像 I 的高频内容 H 并进行平滑化，然后使用统

计自适应伽马矫正来优化纹理度，使其能够作为纹理映射来方便其后的处理。

2.2.1 双边滤波及其快速算法

双边滤波器能够在保留重要边界的情况下平滑化输入图像。每一个像素都是其临近像素的加权平均，但这一权值随着空间距离和亮度差的增大而减小。假设 $g_{\sigma}(x) = \exp(-\frac{x^2}{\sigma^2})$ ，即高斯函数，则应用于图像 I 上的双边滤波函数在像素 p 处的值定义为：

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) I_q$$

其中

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

其中 σ_s 控制空间邻域，而 σ_r 则影响着亮度距离， k 则用于归一化权重。双边滤波常用于创建双范围的分解，其中滤波器的输出作为了大范围层（即基本层），而两幅图像的不同则被称作细节层。本论文采用了 Paris 等人提出的双边滤波的快速方法^[5]，其实质是将图像内容投影到三维的空间内，然后对三维空间做一次快速的高斯滤波。

2.2.2 双边分解

本论文采用了类似于 Durand 和 Dorsey 的双边分解方法^[4]。由于对比度是一种乘积效应，因此双边分解在对数空间中进行处理。从图像 I 中得到基本层 B 和细节层 D 的过程如下所示：

$$B = \text{bf}(I), D = I - B$$

σ_s 和 σ_r 的选的选择对于算法十分关键。 σ_s 描述了空间的尺度，在实验中采用 $\sigma_s = \min(\text{width}, \text{height}) / 16$ 能够获得一致的效果。 σ_r 能够区分重要边界和细节。本论为的双边滤波方法还依赖于梯度范数来估计输入的边界值。如果用 p_{90} 来表示百分之 90，那么 $\sigma_r = p_{90}(\|\nabla I\|)$ 能够获得较为理想的结果。这些设置对于空间和亮度范围都是非常合理的。

Durand 和 Dorsey 指出当边界不够精细时，一些附带的人工效果会出现^[4]。他们介绍了一种“修正”方法，即利用平滑后的基本层来检测不确定性的像素，但他们亦指出了这种方法并不总是令人满意。本论文要求一种可信的无光晕效果的细节层。

本论文通过直接限制分解的梯度来避免逆转的发生。具体而言，我们迫使细节导数 $\frac{\partial D}{\partial x}$ 和 $\frac{\partial D}{\partial y}$ 与输入导数有着同样的方向，并且不会有大于它们范数的值出现。为了能够达到这个目的，我们构建了一个梯度场 $v = (x_v, y_v)$ ：

$$x_v = \begin{cases} 0, & \text{if } \text{sign}\left(\frac{\partial D}{\partial x}\right) \neq \text{sign}\left(\frac{\partial I}{\partial x}\right) \\ \frac{\partial I}{\partial x}, & \text{if } \left|\frac{\partial D}{\partial x}\right| > \left|\frac{\partial I}{\partial x}\right| \\ \frac{\partial D}{\partial x}, & \text{otherwise} \end{cases}$$

y 部分的 y_v 可以类似地定义。矫正后的细节层可以通过解对应的泊松方程来构建。我们依照之前的方法来获取基本层： $B = I - D$ 。这个方法能够保证结果细节层是高保真的，因为它直接解决了梯度逆转问题。

2.2.3 纹理度

Soonmin 等人介绍一种能够表示局部高频内容程度的方法^[4]，从而能够区分细节区域和平滑区域。他们采用同样的边界变量 σ_s 计算图像的高频版本 H 。可以知道，由于低频内容被去掉，该高频图像的平均值是 0。然后，本论文的方法会使用其绝对值来进行滤波。频谱映射或活动映射即这一大小值的低频滤波。这一的映射提供了一个高频纹理和平滑区域良好的对比。图 2-2 是纹理度求解的结果实例。

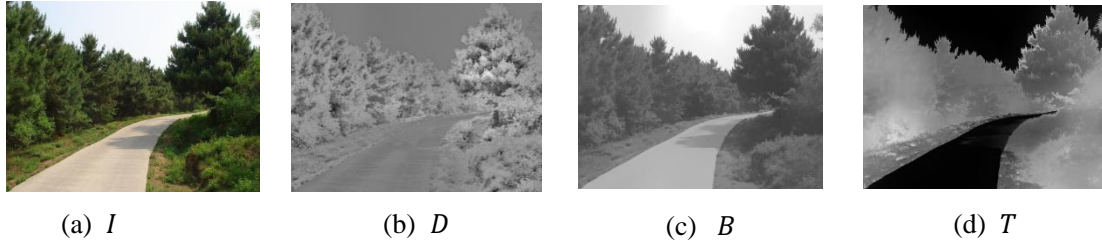


图 2-2：双边分解和纹理度

这一的空间变换操作会在边界区域导致人工效果，即同线性图像分解同样的光晕效果。这一问题同边界保留分解所要解决的问题是一致的，因此可以通过交叉双边滤波^[6]定义纹理度。即：

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) |H|$$

其中

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

这一交叉滤波器与双边分解有同样的 σ_r 值，但 σ_s 在实际情况中是原先的 8 倍，藉此便能够保证平滑的纹理度变换。

2.3 统计自适应伽玛矫正

在得到纹理度 T 之后，其值分布在较为不确定的范围之内。为了统一化处理，有必要将其变换到一个确定的范围中。为此，本论文提出了统计自适应伽玛矫正 (Statistical Adaptive Gamma Correction, SAGC) 方法来将纹理度适当地变

换到区间 $[0,1]$ 内。统计自适应伽玛矫正之前，首先会以 T 的最大值除之。

统计自适应伽玛矫正的本质，是要根据给定的图像均值目标 μ 计算一个合适的伽玛矫正值 γ 。假设给定一幅值分布在 $[0,1]$ 内的图像，伽玛矫正问题定义为

$$\mu = \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y)]^\gamma$$

通过数值分析的方法我们可以解决上述问题的计算方法。该问题等价于解决下面的等式：

$$l(\gamma) = \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y)]^\gamma - \mu = 0$$

因此，按照牛顿法（Newton's Method）可以得到该问题的解。 $l(\gamma)$ 的导数为

$$l'(\gamma) = \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y)]^\gamma \ln[f(x, y)]$$

故可以建立下面的迭代了解 γ ：

$$\gamma_{k+1} = \gamma_k - h \frac{l(\gamma_k)}{l'(\gamma_k)}$$

其中 h 是算法的步长。在实际过程中，如果结果离目标解距离较远则会使用较大的步长 h ，反之亦然。步长 h 的调整亦是一个迭代过程，在每个迭代中如果 γ_{k+1} 并未过度计算，则 h 会加倍，否则我们会每次将 h 减半直到 γ_{k+1} 变得可以接受。在这里，我们将 γ_{k+1} 的过度计算定义为 $l(\gamma_{k+1})l(\gamma_k) < 0$ 。这实际上是一种变步长的牛顿法，它在实际运行过程中有精确的结果和较快的速度。

这一迭代的初始值可以按照符合伽玛矫正数学意义的经验公式来选定，即

$$\gamma_0 = \frac{\frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y)}{\mu}$$

统计自适应伽玛矫正除了能够用于纹理映射的预处理，还可以用于许多其他的应用中，包括颜色迁移、颜色协调化和图像清晰化等，这些内容都将是本论文之后可能探讨的领域。在此，我们将简略地证明算法的收敛性。

观察 $l'(\gamma)$ 形式可知， $l'(\gamma) < 0$ ，因此 $l(\gamma)$ 是一个单调递减函数。同时，我们可以知道 $l(0) = 1 - \mu > 0$ 且 $l(\infty) = -\mu < 0$ ，根据连续函数的性质，可知最终的解 ξ 满足 $\xi > 0$ 。

$l(\gamma)$ 的二阶导数为：

$$l''(\gamma) = \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y)]^\gamma \ln^2[f(x, y)]$$

因此可以知道 $l''(\gamma) > 0$ 。根据泰勒级数定理（Taylor's Theorem），在 γ_k 展开 $l(\xi)$ ，有：

$$0 = l(\xi) = l(\gamma_k) + (\xi - \gamma_k)l'(\gamma_k) + (\xi - \gamma_k)^2 \frac{l''(\eta)}{2}$$

因此

$$|\xi - \gamma_k| = \left| -\frac{l(\gamma_k)}{l'(\gamma_k)} - (\xi - \gamma_k)^2 \frac{l''(\eta)}{2l'(\gamma_k)} \right|$$

再由牛顿迭代法公式（假设 $h = 1$ ），可得：

$$|\xi - \gamma_{k+1}| = \left| -(\xi - \gamma_k)^2 \frac{l''(\eta)}{2l'(\gamma_k)} \right|$$

根据前面的条件 $l'(\gamma) < 0$ 和 $l''(\gamma) > 0$ ，结合上述两个公式可以得到：

$$\frac{|\xi - \gamma_{k+1}|}{|\xi - \gamma_k|} = -\frac{(\xi - \gamma_k)^2 \frac{l''(\eta)}{2l'(\gamma_k)}}{-\frac{l(\gamma_k)}{l'(\gamma_k)} - (\xi - \gamma_k)^2 \frac{l''(\eta)}{2l'(\gamma_k)}}$$

可以看出上述公式满足

$$\frac{|\xi - \gamma_{k+1}|}{|\xi - \gamma_k|} < 1$$

因此当 $k \rightarrow \infty$ 时，通过牛顿迭代法，可以使得 $\gamma_k \rightarrow \xi$ 。证毕。

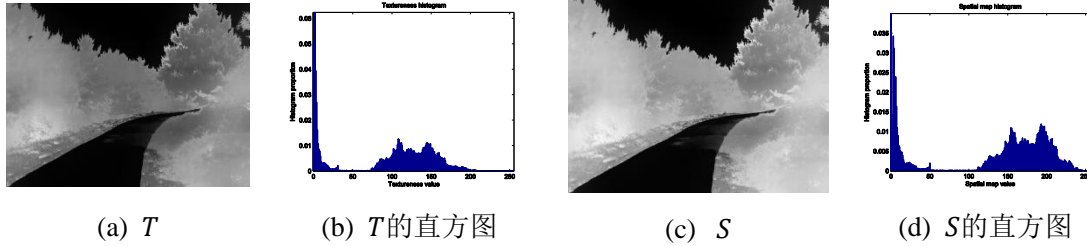


图 2-3：应用统计自适应伽玛矫正于纹理映射上的结果

统计自适应伽玛矫正的特点之一，是它对过度曝光的图像处理效果较好。从数学上，这种情况表示

$$\mu \leq \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y)$$

这一性质可能是由指数函数本身的性质导致的。因此，对于曝光不足的图像，可以用 1 减去其值再进行矫正，矫正完毕后再将数值还原。这一过程叫做互补统计自适应伽玛矫正（Complementary SAGC, cSAGC），如下面的公式所示：

$$S = 1 - \mathbf{SAGC}(1 - T)$$

纹理映射（即在 $\mu = 0.5$ 下的矫正结果）对于本论文的后续处理具有非常好的特性，如图 2-3 所示。

2.4 局部加权回归和映射分区

本论文的算法中大量使用了局部加权回归（Locally Weighted Linear

Regression, LWLR) 的方法。我们定义直方图的回归结果是其每个参数值局部的斜率和 y 截距。可以知道, 这样的回归结果能够通过 在每个参数值上运行局部加权回归来得到, 其中每一次回归都采用对应的参数值作为中心, 其余参数值作为回归影响因子。

$$\theta_{\text{LWLR}} = \operatorname{argmin}_{\theta} \sum_{j=1}^m w(j) (\theta^T x(j) - y(j))^2$$

此处 $w(j)$ 是非负的权重值。在机器学习领域, 一个较为标准的权重是 Bell 曲线, 即:

$$w(j) = \exp\left(-\frac{\|x(j) - x\|^2}{2\tau^2}\right)$$

由于局部加权回归是一种局部算法, 我们必须对每一个 x 均运行算法进行计算。如果我们能够按照下面的方法来构造矩阵 X 和向量 y :

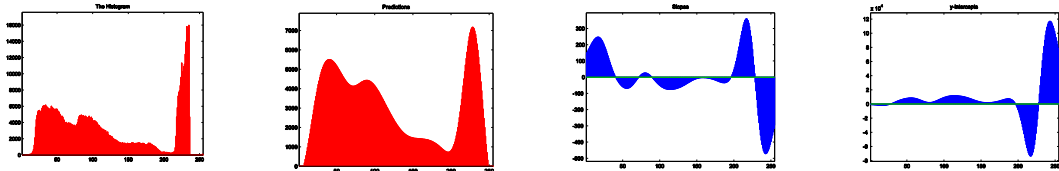
$$X = \begin{bmatrix} (\sqrt{w(1)}x(1))^T \\ (\sqrt{w(2)}x(2))^T \\ \vdots \\ (\sqrt{w(m)}x(m))^T \end{bmatrix}, y = \begin{bmatrix} \sqrt{w(1)}y(1) \\ \sqrt{w(2)}y(2) \\ \vdots \\ \sqrt{w(m)}y(m) \end{bmatrix}$$

那么我们可以用下面的公式来得到回归的解:

$$\theta = (X^T X)^{-1} X^T y$$

从直观上来讲, 图像直方图较低的点可以作为图像内容的拆分点。这一点对于纹理映射和亮度信息来说是一致的。直方图是一种离散函数, 它的分布常常是非常不均衡的。为了能够找到适当地区分图像内容的直方图特征点, 有必要首先对直方图进行平滑化, 这便是我们介绍局部加权回归的目的。

从上面的推导中可以知道, 对于每一个 j , 直方图拟合的结果包括两部分, 分别是局部的 y 截距 $\theta_1(j)$ 和局部的斜率 $\theta_2(j)$ 。局部斜率的作用同一阶导数的意义是类似的, 但它是一种离散拟合的结果。我们可以使用 $h'(j) = \theta_2(j)$ 来表示这一类比。如果我们想找到零点 (直方图的极小或者极大值点), 只要找到 $h'(j)$ 中最接近 0 的点即可。



(a) 直方图 (b) 局部加权回归 (c) 局部斜率 $\theta_2(j)$ (d) 局部 y 截距 $\theta_1(j)$

图 2-4: 直方图局部加权回归的例子

如果颜色来源和目标图像直方图的零点不匹配, 有必要对其进行配适操作。

其基本思想是从零点较多的直方图中删去一部分零点。为了尽量保证两幅直方图的匹配性，本论文制定了下面的规则来对零点进行配适：

1. 如果两个零点集合元素个数差是奇数，则需要从较大的集合中删去前端或者后端的一个点。实际删除的点是与极大-极小更迭性不匹配的点。
2. 如果较大的零点集合包含一种特定的递增或递减趋势下的零点对，则删去这样的点对，直到找不到这样的点对或者两个集合已经适配。
3. 如果找不到如 2 所言的点对，则从较大的零点集合中删去直方图值最近的两个相邻零点，直到两个零点集合的零点个数相等。

我们称上述的算法为零点配适。这一技术将会同样地运用于之后计算亮度-色彩对应关系的过程中，这是因为这一步也需要用局部加权回归来处理直方图。在这一算法过程中，较难的部分是第二步，因为在这一步骤我们需要给定零点集合的零点对是否处于递增或递减趋势中。 z_i 和 z_{i+1} 是满足这一关系的充要条件是：

$$\min\{h(z_i), h(z_{i+1})\} > \min\{h(z_{i-1}), h(z_{i+2})\}$$

且

$$\max\{h(z_i), h(z_{i+1})\} < \max\{h(z_{i-1}), h(z_{i+2})\}$$

能够拆分直方图的点是在颜色来源和目标图像中均未极小值的点。假设 j 是一个拆分点，那么它必须满足 $h'(j)$ 局部接近于 0 且 $h'(j) < h''(j+1)$ 。这是由于零点导数的递增性表示的是局部极小值。

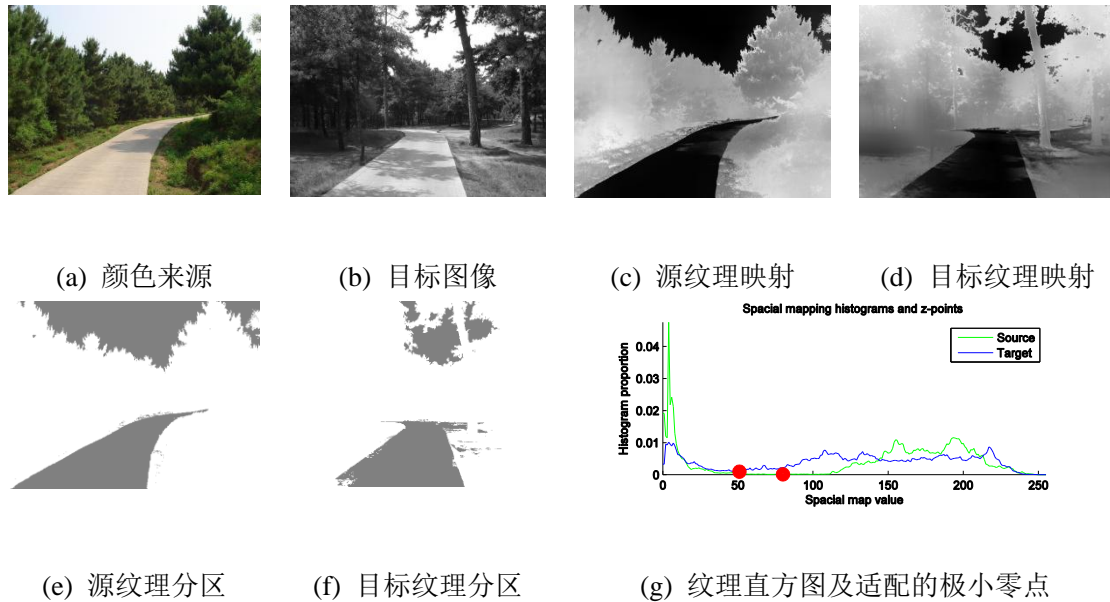


图 2-5：直方图对纹理的分区结果

2.5 灰度图像自动着色的实现

为了保证着色结果中色彩和纹理的一致性，图像的亮度和纹理映射均会作为着色的索引来使用。为了达到这一目的，首先需要利用直方图的局部加权回归与零点适配算法来对纹理映射和亮度图像进行分区；其次，根据每个分区的信

息可以建立其局部亮度-色彩对应关系；再次，根据局部的亮度-色彩对应关系可以加权得到全局亮度-色彩对应关系；最后，根据全局亮度-色彩对应关系，结合目标图像的亮度信息和纹理映射，便可以对其进行整体的着色。

本论文的方法所采用的颜色空间是 JFIF 定义的 YCbCr 空间。与其它空间相比，从 RGB 向这一空间的变换是线性的，而且可以直接地量化到 8 位数据之内。在亮度-颜色对应关系的计算过程中，很有可能某些亮度值不配对于任何一种颜色信息，因此为了能够保留亮度，我们必须假设 Cb 和 Cr 通道的默认值是 128。

为了计算局部亮度-色彩对应关系，首先必须得到色彩来源每一个纹理分区内每个亮度的平均色彩 c_s 。然后需要对每个纹理分区内的亮度直方图运行局部加权回归与零点适配。这些适配的零点便可以作为指引点来生成每一个纹理分区内的亮度-色彩对应关系。

假设给定对应适配零点对之间的某亮度区域， η 是最小的亮度值， δ 是两个适配点对之间的亮度差，那么可以按照下面的方法来算得目标图像的亮度-色彩对应关系 c_T ，其中 c_s 是颜色来源的平均色彩：

$$c_s = \frac{\sum_{i=0}^{255} w \left(i, (j - \eta_T) \frac{\delta_s}{\delta_T} + \eta_s, \frac{\delta_s}{\delta_T} \right) c_s(i)}{\sum_{i=0}^{255} w \left(i, (j - \eta_T) \frac{\delta_s}{\delta_T} + \eta_s, \frac{\delta_s}{\delta_T} \right)}$$

从上面的公式可以知道，局部亮度-颜色对应关系是按照一种加权的方式进行计算的。这能够避免纹理分区内颜色突变带来的失真和人工效果。这种计算仅仅是对亮度的色彩分布进行了平滑化，并没有改变纹理的强弱信息，因此不能等价于对图像的模糊化操作。

在有了局部亮度-色彩对应关系之后，即可以在不同的纹理分区之间进行一次色彩对应关系的高斯平滑化。这次平滑化是在着色边界点上进行的，其中着色边界点指的是纹理直方图拟合分区过程中的拆分点。在这一步骤完成后，便生成了可以直接完成目标图像着色的全局亮度-色彩对应关系（Global Luminance-Color Correspondence, GLCC）。

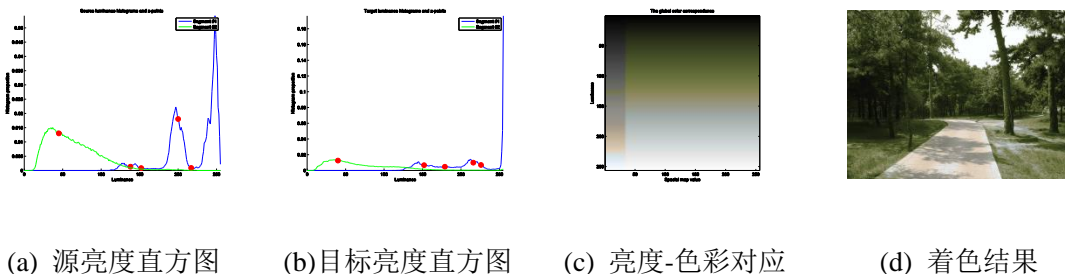


图 2-6：灰度图像自动着色的例子

全局亮度-色彩对应关系本身亦可看成是一幅二维彩色图像，其宽度和高度分别表示亮度范围和纹理映射范围。这一映射可以直接用来对原图像进行着色，具体而言，通过一个简易的循环遍历纹理映射和图像亮度，根据这两个值对应的

色彩对目标图像进行着色即可。图 2-6 是之前例子的着色结果。

第三章 灰度视频的自动着色

3.1 灰度视频自动着色模型设计

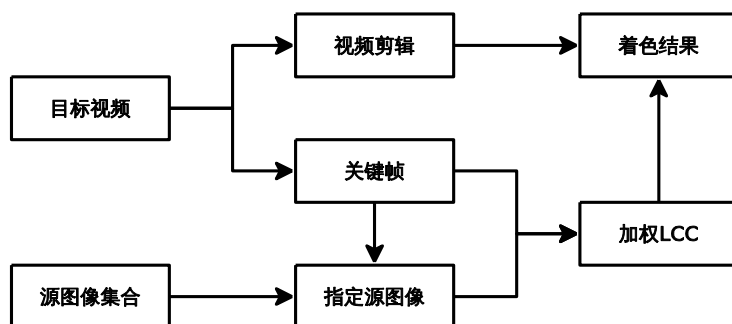


图 3-1：灰度视频自动着色模型

灰度视频的自动着色模型以灰度图像自动着色为基础。这体现在灰度视频自动着色的模型上，如图 3-1 所示。灰度视频的自动着色模型以关键帧(Key-frame)为主要技术，利用灰度图像自动着色在关键帧上的运行数据来使对应视频剪辑(Video Clip)着色。

灰度视频的自动着色模型在灰度图像自动着色模型的基础上，增加了两种关键性的技术，即图像配准(Image Registration)和视频的关键帧检测(Key-frame Detection)。图像集合的配准是指对图像集进行分析，并将图像数据转换到新的空间中，以便于对图像进行基于内容的检索或识别。视频的关键帧检测是指对视频的图像序列进行分析，将视频按照时间进行拆分，提取出关键帧及其对应的视频剪辑。

在图像集的配准与搜索中，本论文选用基于主元分析(Principal Components Analysis, PCA)的方法^[9]和支持向量机(Support Vector Machines)，其中主元分析可以直接由像素点作为数据进行内容分析和数据投影，图像配准速度较快，图像检索或识别的代价较小；而支持向量机以多种图像特征(Features)为数据进行聚类，图像配准速度较慢，但目的性较强，图像检索或识别的代价一般。本论文所采用的视频及图像集类别未知，特征提取不具有普适性，故采用主元分析法。

如果视频中的帧图像逐一进行着色，耗时巨大，因此我们引入了关键帧方法，以提取视频中的关键帧并提取对应的视频剪辑。根据对关键帧图像着色的主要参数，可以对其对应的视频剪辑进行统一着色。这一方法可以极大地提高处理视频着色的速度。

3.2 基于主元分析的图像配准与搜索

主元分析的目的是将数据 $x \in \mathbb{R}^n$ “大致地”投影到某 k 维子空间中，其中 $k \ll n$ 。也就是说，主元分析尝试识别数据投影的子空间信息。主元分析所需要

的整个计算核心是一个特征向量计算过程，相对而言比其它的基于统计或概率模型的方法更易于理解。此外，这个特征向量计算过程可以通过一个奇异值分解（Singular Value Decomposition, SVD）算法来得到。目前各种线性代数计算工具库对奇异值分解算法的实现已经非常高效。

在运行主元分析之前，通常会将数据进行预处理，即进行归一化操作：

1. 令 $\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$;
2. 用 $x^{(i)} - \mu$ 替代 $x^{(i)}$;
3. 令 $\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_j^{(i)})^2$;
4. 用 $x_j^{(i)} / \sigma_j$ 替代 $x_j^{(i)}$ 。

步骤 1-2 将数据的平均值变成 0，但对于已知均值为 0 的数据可以忽略。步骤 3-4 按照每个坐标的方差进行缩放，使其具有单位方差，保证了不同的特征均按照一致的“程度”来进行处理。如果事先已经知道各个维度是一致的，则步骤 3-4 亦可忽略。在本论文的图像处理过程中不需要进行步骤 3-4 的计算。

完成归一化之后，下一步需要决定“方差主要维度” u 。这一问题可以理解为寻找合适的单位向量 u ，当数据投影到 u 上时，数据的在 u 方向的方差得到了最大化。直观而言，数据在刚开始便具有一定的方差，我们希望能够找到一个方向 u ，使得我们能够最佳地仅通过 u 便能近似地表示数据。

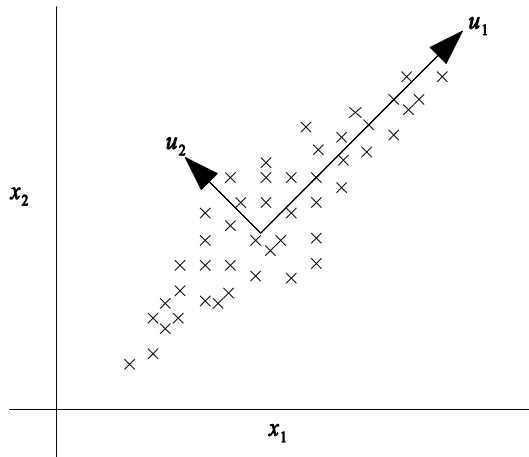


图 3-2：主元分析表示

为了将此问题形式化，首先可以知道给定一个单位向量 u 和一个数据点 x ，那么将 x 投影到 u 方向上的长度是 $x^T u$ 。也就是说，如果 $x^{(i)}$ 是一个数据集中的数据点，那么它在 u 方向上的投影是距离原点 $x^T u$ 长度的向量。因此，最大化投影方差可以通过选择合适的单位向量 u 来完成：

$$\frac{1}{m} \sum_{i=1}^m (u^T x^{(i)})^2 = \frac{1}{m} \sum_{i=1}^m u^T x^{(i)} (x^{(i)})^T u = u^T \left(\frac{1}{m} \sum_{i=1}^m x^{(i)} (x^{(i)})^T \right) u$$

容易知道，在 $\|u\|_2 = 1$ 的条件下最大化上面的值实际上会得到矩阵 $C = \frac{1}{m} \sum_{i=1}^m x^{(i)} (x^{(i)})^T$ 的特征向量。其中 C 恰好是数据的经验协方差矩阵。

总而言之，如果希望找到一个 1 维的子空间来逼近给定的数据集，则可以令 u 成为 C 的主要特征向量。推广到更一般的情况，如果希望将数据投影到 k 维空间中 ($k < n$)，那么只要选择 C 的前 k 个特征向量 u_1, \dots, u_k 作为投影方向即可。这些向量实际上形成了数据的一个新的正交基。

如果希望在新的基中表示数据 $x^{(i)}$ ，只需要计算相应的向量即可：

$$y^{(i)} = \begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \vdots \\ u_k^T x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

尽管 $x \in \mathbb{R}^n$ ，向量 y 现在成为一个低维度 (k 维) 的 x 近似。因此，主元分析又称作维度约减算法。向量 u_1, \dots, u_k 称作数据的主 k 元。

本论文的目的是利用主元分析完成图像集的配准与搜索^[9]。学界对主元分析在图像集的配准与搜索方面已经有较多研究，其中大多数集中于人脸识别方面。本课题的实践表明主元分析对于图像集的配准与搜索方法已经足以满足本论文的设计需求。具体地说，图像集的配准是按照下面的步骤完成的：

1. 将图像转换成堆栈化向量 (Stacked Vectors)，在这个过程中可以对图像进行大尺度的压缩；
2. 对每个向量中的值，删去其对应维度所有向量的平均值 (均值变为 0)；
3. 将得到的向量组合成数据矩阵，运行奇异值分解 (Singular Value Decomposition)，获得这些数据协方差矩阵的特征值 (Latents) 和特征向量 (Coefficients)，并获得在变换空间中的结果 (Score)。

如果需要进行图像的搜索及匹配，只要将图像按照上述参数转换到新的空间中，获取最近的数据 (Nearest Score) 即可。这一过程是非常快速的。

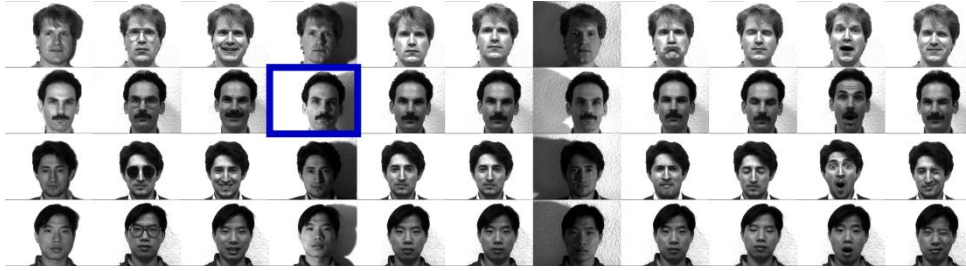
3.3 奇异值分解计算主元分析

奇异值分解 (Singular Value Decomposition, SVD) 是计算主元分析的重要手段，这是因为目前利用计算机进行奇异值分解的算法已经十分成熟，具有较好的性能和较快的速度，且运行奇异值分解不需要计算主元分析中的经验协方差矩阵本身。

假设 M 是一个 $m \times n$ 矩阵，那么对其存在一个如下方式的分解：

$$M = U \Sigma V^T$$

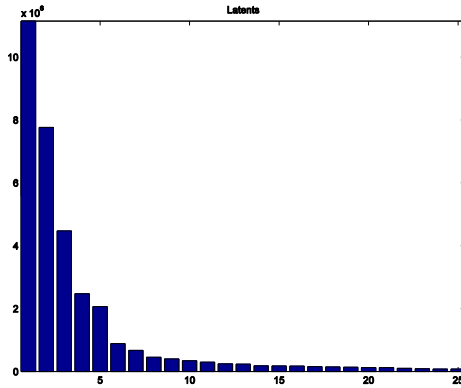
其中 U 是一个 $m \times m$ 单位矩阵， Σ 是一个 $m \times n$ 对角矩阵，其对角元素均为非负数， V^T 表示的是 $n \times n$ 单位矩阵矩阵 V 的转置。这样的一个分解叫做矩阵 M 的奇异值分解。



(a) 待配准图像集



(b) 匹配图像排序



(c) 主元分析 Latent



(d) 主元分析 Score(25 个主元)

图 3-3: 对耶鲁大学人脸识别图像集的测试

基于上面的定义，我们可以将 $M^T M$ 写成如下的推导：

$$M^T M = (U \Sigma V^T)^T U \Sigma V^T = V \Sigma^T U^T U \Sigma V^T$$

容易知道，由于 U 是一个 $m \times m$ 单位矩阵，因此 $U^T U$ 是对角幺元矩阵，即 $U^T U = I$ 。由于我们已经定义了 $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ ，其中 $k = \min\{m, n\}$ ，因此我们有 $\Sigma^2 = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$ 。

继续推导 $M^T M$ 可知：

$$M^T M = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T I \Sigma V^T = V \Sigma^2 V^T$$

由特征向量分解的定义可知，一个 $n \times n$ 矩阵 A 可以表示为 $A = S \Lambda S^T$ ，其中 S 是一个 $n \times n$ 单位矩阵，包含 A 的单位特征向量； Λ 是包含 A 对应特征值的对角矩阵。

由此容易知道，在 $M^T M$ 推到过程中得到的 $V \Sigma^2 V^T$ ，实际上 V 是 $M^T M$ 的特征向量矩阵， Σ^2 是 $M^T M$ 的特征值矩阵。同样地，可以知道 U 是 MM^T 的特征向量矩阵。

回到主元分析的问题上来，实际上我们需要计算的是经验协方差矩阵 $C = \frac{1}{m} \sum_{i=1}^m x^{(i)} (x^{(i)})^T$ 的特征向量。如果定义矩阵 X 为：

$$X = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix}$$

那么可以证明 $C = \frac{1}{m} X^T X$ 。

通过线性代数的知识可以知道，对 $X^T X$ 乘以 $\frac{1}{m}$ 不改变其特征向量。因此计算 C 的特征向量归约成对 X 求解奇异值分解，其中 $X = U \Sigma V^T$ ，而 V 中即含有所求的向量。

图 3-3 给出了利用耶鲁大学人脸识别数据库进行人脸识别的图像配准结果。其中蓝色图像是目标匹配图像。由于我们利用了奇异值分解进行快速主元分析计算，实际性能下 44 幅 512*384 灰度图像的配准时间在 17 秒之内，利用 25 个维度进行配准的时间为 500 毫秒。

3.4 灰度视频的自动剪辑

对灰度视频进行着色最直观的方法就是对视频进行逐帧着色。但对视频的每一帧都进行图像搜索与着色耗时巨大，难于有效地完成视频自动着色的任务，因此本论文采用视频关键帧的技术来实现较好的视频着色性能。



(a) 关键帧 1/509



(b) 关键帧 295/509



(c) 关键帧 360/509



(d) 关键帧 439/509

图 3-4：关键帧检测和视频剪辑的一个例子

关键帧是指能够代表视频某时间段内容的关键性图像，关键帧检测技术在视频压缩、视频传输和其他视频处理方法中都有大量地应用。本论文采用的关键帧

检测技术是基于帧数据差方法^[11]的简化版，检测速度较快，精度一般，但对于我们的目的来说已经足够。

该方法的基本思想是，帧与帧之间的像素值差可以作为帧之间内容差别的指示，同时较大的像素值差意味着新的关键帧和新的视频剪辑。因此，本论文的关键帧检测结果包含两种数据，一是关键帧本身，二是对应于关键帧的视频剪辑。图 3-4 说明了本方法的检测结果。

在实际测试中，500 帧长度的 720*576 视频可以在 1 分钟之内完成关键帧检测与视频剪辑，说明本论问的方法有可以接受的性能。图 3-5 是本方法的流程图。

有了关键帧检测的算法，结合图像集的搜索与配准方法，即可通过利用关键帧图像着色数据来对对应视频剪辑进行统一着色。

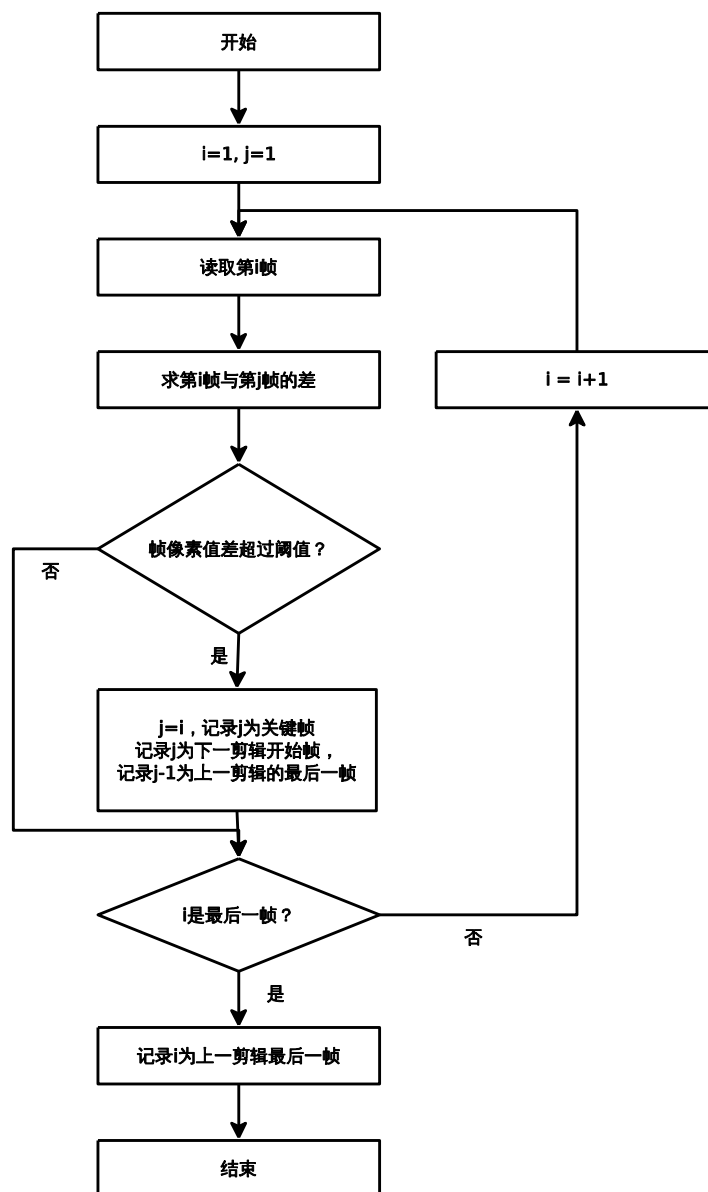
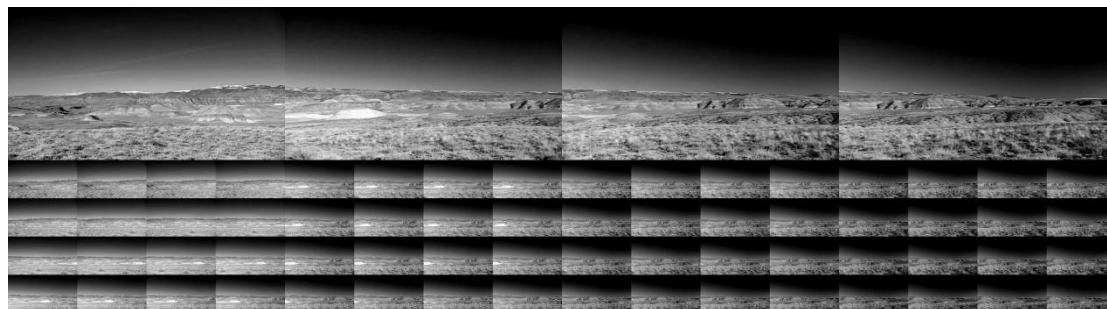


图 3-5：关键帧检测和视频剪辑算法流程图

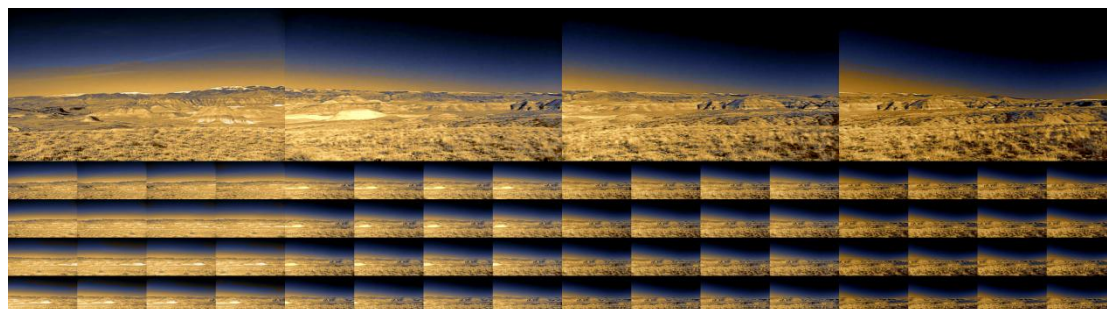
3.5 灰度视频自动着色的实现

灰度视频自动着色是以关键帧为着色对象，利用对关键帧的着色数据来对关键帧所对应的视频剪辑进行着色，从而实现速度较快的视频整体着色过程。

由于本论文视频自动着色的假设是用户能够提供一套与目标灰度视频内容类似的图像集，因此我们首先会将这一图像集进行基于主元分析的配准，并利用配准后的数据进行基于内容的检索，对每一幅关键帧匹配一幅颜色来源，使得关键帧的着色过程自动化。



(a) 检测出的关键帧及其对应剪辑中的帧



(b) 对关键帧及其对应剪辑的着色

图 3-6: 的自动着色结果

在关键帧着色过程结束之后，要对视频自动剪辑得到的剪辑片断进行着色，在这个着色过程中需要利用关键帧的图像着色结果。之前已经提到，图像着色的中间数据包括有零点配适结果和关键帧的全局亮度-色彩对应关系等。在对视频剪辑每一帧的处理过程中，仍然需要计算每一帧图像的纹理度，然后利用帧着色过程中产生的纹理映射直方图和亮度直方图的零点配适结果对视频剪辑帧进行处理，获取其零点匹配程度。若零点匹配类似性在预先定义的范围内，则可以用关键帧着色过程中产生的全局亮度-色彩对应关系对目标图像进行着色。着色完成之后将所有帧写入视频即可得到目标着色结果。本课题测试的视频大小为 720*576，平均处理时间约 8s/帧。

第四章 性能分析与改进

4.1 快速双边滤波算法及其替代算法

本论文提出的图像着色算法中利用了纹理度方法来计算纹理映射，其中关键性步骤是双边滤波算法。双边滤波是一种关键性边界保留的算法，它的这一特性使之广泛应用于高动态范围图像处理、图像与视频摘要等新兴的图形图像课题之中。但由于纹理度算法本身敏感于亮度的差，因此双边滤波算法受限于其性能的缺陷。本论文采用了类似于 Paris 等人^[5]提出的快速双边滤波的方法，其速度是原先直接实现的 50-60 倍。这一速度提升并非由于我们提升了算法的复杂性，而是因为我们利用了现代 CPU 的向量化（Vectorization）计算和处理能力。有关向量化的详细技术本论文不做探讨，但读者可以参阅 Intel SSEx 系列指令集的说明。

之前已经提到，双边滤波算法可以表示为：

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) I_q$$

其中

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

$g_{\sigma} = \exp(-\frac{x^2}{\sigma^2})$ 表示高斯核。我们通过一个高维变换实现快速双边滤波，其基本思想是将二维图像投影到包含亮度维度的三维空间之中，从而使得亮度高斯核能够用一致地用统一的空间高斯核表示。

假设 J 是对图像 I 的高维变换，满足：

$$J(p, l) = \begin{cases} I(p), & \text{if } I(p) = l \\ 0, & \text{otherwise} \end{cases}$$

则有：

$$\begin{aligned}
 \text{bf}(I)_p &= \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I(p) - I(q)|) I(q) \\
 &= \frac{1}{k} \sum_{q \in I} \exp\left(-\frac{\|p - q\|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I(p) - I(q)|^2}{\sigma_r^2}\right) I(q) \\
 &= \frac{1}{k} \sum_{q \in I} \exp\left(-\frac{\|p - q\|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I(p) - I(q)|^2}{\sigma_r^2}\right) I(q) \\
 &= \frac{1}{k} \sum_{q \in I} \exp\left(-\frac{\|p - q\|^2}{\sigma_s^2} - \frac{|I(p) - I(q)|^2}{\sigma_r^2}\right) I(q) \\
 &= \frac{1}{k} \sum_{q \in I, l \in L} \exp\left(-\left\|\left(\frac{p - q}{\sigma_s}, \frac{l_p - l}{\sigma_r}\right)\right\|^2\right) J(q, l)
 \end{aligned}$$

因此，若假设存在三维的高斯核 $g_J(x, l) = \exp\left(-\left\|\left(\frac{x}{\sigma_s}, \frac{l}{\sigma_r}\right)\right\|^2\right)$ ，且有如下的线性滤波：

$$\text{bf}(J)_{p,l} = \frac{1}{k} \sum_{q \in I, o \in L} g_J(\|p - q\|, \|o - l\|) J(q, o)$$

其中

$$k = \sum_{q \in I, o \in L} g_J(\|p - q\|, \|o - l\|)$$

那么可以知道

$$\text{bf}(I)_p = \text{bf}(J)_{p, I(p)}$$

通过上面的推导可知，如果想对图像 I 进行双边滤波，可以通过对提升了维度的数据 J 进行单一的高斯滤波，从而利用现代计算机的向量化计算能力来实现高速的双边滤波。实际应用过程中，由于双边滤波本身是一种低通滤波，可以首先对数据进行一定程度的量化，而不会影响最终的滤波结果。这更大程度上地提升了算法的效率。

按照刚才的结论，假设图像的大小是 $M \times N$ ，亮度维度的大小是 L ，那么原双边滤波的算法复杂度是 $O((MN)^2 \log(MN))$ ，而新的高速算法复杂度是 $O(MNL \log(MNL))$ 。从复杂度上难以分析计算速度的提升程度，但一次性的滤波可以利用快速傅立叶变换对三维数据整体来计算，充分利用了 CPU 的向量化能力和内存交换能力，因此获得了几十倍的速度提升，使得视频着色成为可能。表 4-1 是双边滤波算法的性能比较。

表 4-1 双边滤波算法性能

图像大小	非线性算法耗时	替代算法耗时
128*96	3.9711s	0.2375s
256*192	5.7994s	0.3558s
512*384	17.1844s	0.8787s
1024*768	120.5443s	2.5774s
2048*1536	533.4667s	10.0776s

4.2 对亮度-色彩对应关系的量化



(a) 颜色来源 (b) 目标图像 (c) 量化之前的效果 (d) 量化之后的效果

图 4-1：对亮度-色彩对应关系的量化

在得到全局亮度-色彩对应关系之后，可以通过遍历亮度图像和纹理映射的值来对图像进行着色。在遍历过程中，实际上是根据亮度图像和纹理映射的值来对图像的像素进行取色，从而实现着色过程。在之前的计算过程中，由于我们将图像和纹理映射均量化为 8 位，全局亮度-色彩对应关系是一幅 256*256 的图像，遍历一次需要的迭代数是 65536。这一过程对一幅 1024*768 的图像大约会消耗 8 分钟的时间，性能不高。而通过试验，我们发现如果对全局亮度-色彩对应关系进行进一步量化，则可以在基本上不改变图像着色效果的情况下加快处理速度，使根据全局亮度-色彩对应关系进行着色的时间缩短到 5 秒之内。

结合之前快速双边滤波算法对性能的提升，最终关键帧图像着色的时间大约是 20 秒(1024*768)，非关键帧着色的时间大约 4 秒，性能在可接受的程度之内。本论文的测试图像大小为 1024*768，其着色时间为 24 秒左右。图 4-2 给出了更多图像着色结果。

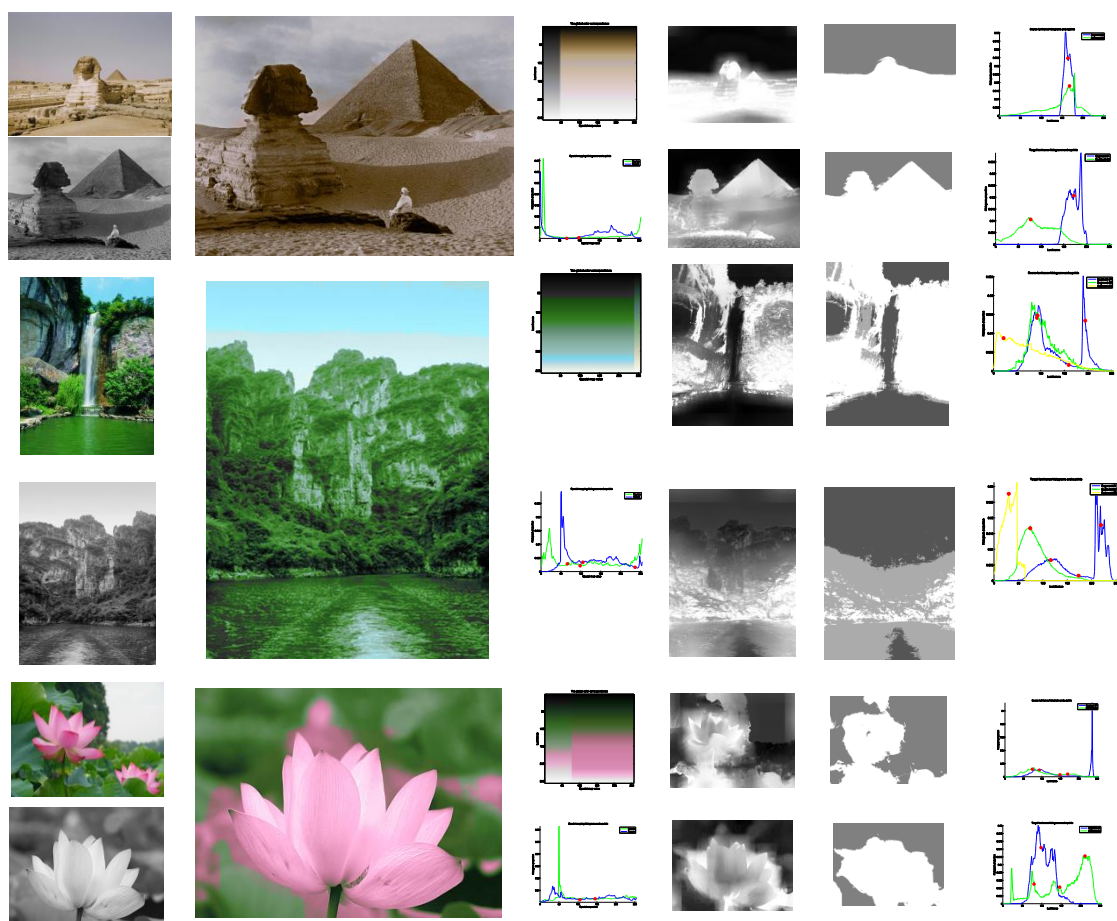


图 4-2：更多图像着色结果实例

第五章 结论与展望

5.1 总结

本论文探讨了灰度图像和视频自动着色技术的实现。本论文的研究主要包括灰度图像的自动着色研究和灰度视频的自动着色研究。

灰度图像的自动着色研究从亮度和纹理两方面出发，体现了着色过程保持亮度及纹理一致性的思想。其主要创新技术包括纹理映射和纹理度方法，基于直方图局部权重回归的图像分区，局部和全局亮度-色彩对应关系，以及快速双边滤波与亮度-色彩对应关系量化的加速方法。

纹理映射和纹理度方法是我们保证纹理与颜色一致性的基础手段，它们作为原图像的函数，用于表达图像中纹理的局部分布信息，我们实际采用的纹理度方法是对纹理程度的表示，其核心技术包括双边滤波和泊松重建等。

直方图用于表示图像中亮度的统计分布情况，对于纹理映射来说，它表示纹理映射中纹理映射值的统计分布情况。对直方图进行加权局部回归可以得到局部化的斜率与 y 截距，其中斜率可以用于分析直方图中零点的分布状况，从而能够寻找出直方图的最低点来对图像进行分区。在零点查找的过程中还会对零点进行匹配，以保证颜色来源和目标图像内容的一致性。根据纹理映射直方图和亮度直方图的双重零点适配实际上同时考虑了图像的亮度和纹理度信息，亦可用作亮度和纹理协调化的分区方法。

根据每个纹理分区内的亮度零点可以得到局部亮度-色彩对应关系，然后再由一次加权计算即可得到全局亮度-色彩对应关系。这一对应关系可以直接地用于进行着色。但这样实现的图像自动着色技术性能较差，为了提升其性能，我们对性能瓶颈双边滤波和全局亮度-色彩对应关系进行了优化。其中双边滤波使用了高维的一次高斯滤波计算来加速，而全局亮度-色彩对应关系则是进行了一个简单的量化过程。

本论文灰度图像的自动着色为灰度视频的自动着色打下的坚实的基础。灰度视频的自动着色以灰度图像的自动着色为基础，并包含两个关键性的技术：图像集的配准与搜索，以及视频关键帧检测与视频剪辑技术。

图像集的配准与搜索是以图像内容为目标的过程，它的目的是对原图像集进行空间变换，从而能够将新的图像在图像集内匹配到符合条件的目标图像。这个过程我们采用的是主元分析方法，其本质是对原数据的空间变换。主元分析可以利用奇异值分解算法进行快速计算，它在我们的整个算法过程中是性能较高的部分。

视频的关键帧检测和视频剪辑是为了加快视频着色的处理过程而采用的。我们首先会对关键帧进行着色，然后利用着色的中间数据对该关键帧对应的视频进

行统一着色。在关键帧检测和视频剪辑算法过程中，帧之间的像素差值可以作为帧差别的指示，而较大的像素值也意味着新的关键帧和新的视频剪辑。这种方法是一种简化的基于关键帧差的方法，实现简易，性能也较高。

5.2 工作展望

在本论文的实验过程中，仍然有效果不佳的实例，这是本研究接下来可能要进行的方向。本论文的研究过程中有许多创新性的内容也可以作为新的研究方向。这包括了利用纹理映射进行其它的需要考虑纹理的图像处理过程进行操作，利用直方图的局部加权回归实现简易的图像分区，以及统计自适应伽玛矫正的多种用途。这些内容都将作为我们接下来学习和研究的指导来参考。

参考文献

- [1]Charpiat G, Hofmann M., Schölkopf B, Image Colorization via Multimodal Predictions[A], Proceedings of the 10th European Conference on Computer Vision[C], 2008, 126-139.
- [2]Li J, Hao P, Transferring Colours to Grayscale Images by Locally Linear Embedding[A], Proceedings of the British Machine Vision Conference[C], Leeds, UK, 2008.835-844.
- [3]Morimoto Y, Taguchi Y, Naemura T, Automatic Colorization of Grayscale Images using Multiple Images on the Web[A], Proceedings of SIGGRAPH Posters[C], 2009 Article 32.
- [4]Bae S, Paris S and Durand F, Two-scale Tone Management for Photographic Look[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2010.
- [5]Paris S and Durand F, A Fast Approximation of the Bilateral Filter using a Signal Processing Approach[A], Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2006.
- [6]贾云涛, 胡事民. 基于图切分的交互式图像染色算法[A]. 计算机学报[J], 2006, 29(3): 508-513.
- [7]胡伟, 秦开怀. 高分辨率灰度图像的快速多分辨率着色[A]. 计算机学报[J], 2009, 32(5): 1062-1068.
- [8]Mitchell T.M. Machine Learning[M]. McGraw-Hill, 1997. 236-238.
- [9]Zhang J. Face Recognition: Eigenface, Elastic Matching and Neural Nets[A], Proceedings of IEEE[J], Vol. 85, No. 9, September 1997
- [10]Chatzigiorgaki M., Skodras A.N. Real-time Keyframe Extraction towards Video Content Identification[A], School of Science and Technology, Hellenic Open University, 26222 Patras, Greece.

外文资料

Two-scale Tone Management for Photographic Look

Soonmin Bae Sylvain Paris Fredo Durand

Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

Abstract: We introduce a new approach to tone management for photographs. Whereas traditional tone-mapping operators target a neutral and faithful rendition of the input image, we explore pictorial looks by controlling visual qualities such as the tonal balance and the amount of detail. Our method is based on a two-scale non-linear decomposition of an image. We modify the different layers based on their histograms and introduce a technique that controls the spatial variation of detail. We introduce a Poisson correction that prevents potential gradient reversal and preserves detail. In addition to directly controlling the parameters, the user can transfer the look of a model photograph to the picture being edited.

1 Introduction

Much research has been dedicated to tone mapping for the display of high-dynamic-range images. These tools focus on contrast reduction, seeking a neutral reproduction, and are ideal when fidelity is needed. However, tone manipulation is also useful when the input has normal dynamic range, and many users seek to obtain a certain “look” for their pictures to convey a mood or an aesthetic. This is particularly significant for black-and-white photography where strikingly distinctive styles can be achieved. We present a new tone management approach that offers direct control over the “look” of an image for both high- and normal-dynamic-range inputs.

The “look” of images has been addressed in Non-Photorealistic Rendering and recent analogy approaches enable the imitation of texture or stylized images in a purely data-driven fashion, e.g. [Hertzmann et al. 2001]. However, to the best of our knowledge, no approach enables the imitation of a photographic “look” such as the ones achieved by master black-and-white photographers.

We argue that a large part of such a look deals with the management of tones, which advanced photographers perform through elaborate lighting, tedious work in the darkroom, or using photo editing software. Unfortunately, such painstaking work and advanced skill is out of reach of casual users. In addition, the issues of workflow and efficiency are becoming prevalent among professional users. The workflow describes the full process from image capture to printing and can include multiple software stages and manual retouching, all requiring much effort. Reducing the user work is critical to professionals, and many manuals and tools are dedicated to optimizing and automating all steps. For example, a wedding photographer takes hundreds of pictures and needs to give them a consistent look and to swiftly deliver

them to clients. Individual retouching is hardly an option, and creative control over the look of each image is often sacrificed. Recently-introduced software such as Apple’s Aperture and Adobe’s Lightroom focuses on workflow optimization but offers little interactive editing capabilities.

To address these difficulties, we propose a tone-management technique dedicated to both casual and professional photographers. We focus on the tonal aspects of photos decoupled from their content. Issues such as framing and topic selection are out of the scope of our work. We nevertheless demonstrate the wide range of looks that our approach can produce. We provide simple controls and enable both global and local tone management. In addition to direct manipulation, users can transfer the look of a model picture, thereby “showing” the desired look. This also allows professionals to apply the rendition of previous prints to new photographs.

This paper makes the following contributions.

Large-scale Tonal Balance Management: We control the large-scale spatial tonal variation over an image.

Spatial Detail Variation: We manipulate the amount of high-frequency detail or texture and its spatial variation. In particular, we introduce a computation of textureiness that measures local high-frequency content while respecting strong edges.

Gradient Constraint: We employ a gradient reconstruction step to prevent gradient reversal and preserve detail.

Our exposition focuses on transfer between images because it demonstrates the relevance and robustness of the features we manipulate. However, direct control through the curve interface is equally powerful, though perhaps more suited to advanced users.

1.1 Related work

Tone Mapping Tone-mapping seeks the faithful reproduction of high-dynamic-range images on low-dynamic-range displays, while preserving visually important features [Reinhard et al. 2005]. Our work builds on local tone mapping where the mapping varies according to the neighborhood of a pixel [Pattanaik et al. 1998; Tumblin and Turk 1999; Reinhard et al. 2002; Durand and Dorsey 2002; Fattal et al. 2002; Li et al. 2005]. The precise characteristics of film have also been reproduced [Geigel and Musgrave 1997; Reinhard et al. 2002]. However, most techniques seek an objective rendering of the input, while we want to facilitate the exploration and transfer of particular pictorial looks.

Conversion to Grayscale Gooch et al. [2005] convert color images to grayscale while preserving salient features. They also seek fidelity to the original picture, whereas we explore stylistic variations. Their approach is nonetheless complementary to ours because it extracts compelling contrast from color images.

Gradient Image Processing A number of recent techniques have characterized images by their gradient and used Poisson reconstruction to perform tone mapping

[Fattal et al. 2002] and montages [Pérez et al. 2003; Agarwala et al. 2004]. We also exploit the Poisson approach to ensure the quality of our result, because it naturally allows us to combat gradient reversal, a traditional plague of aggressive multi-scale manipulation.

Style Transfer and Stylization Style transfer has been explored for the textural aspects of non-photorealistic media, e.g. [Hertzmann et al. 2001; Drori et al. 2003], and DeCarlo et al. stylize photographs based on saliency [2002]. In contrast, we seek to retain photorealism and control large-scale effects such as tonal balance and the variation of local detail. In addition, our parametric approach leads to continuous changes supported by interactive feed-back and enables interpolations and extrapolations of image look.

Visual Equalizer Our work is inspired by the ubiquitous visual equalizer of sound devices. Similarly, the modification of frequency bands can alter the “mood” or “style” of motion data [Bruderlin and Williams 1995]. The equivalent for images is challenging because of the halos that frequency decomposition can generate around edges. Our work can be seen as a two-band equalizer for images that uses non-linear signal processing to avoid halos and provides fine tonal and spatial control over each band.

1.2 Achieving a Photographic Look

The traditional darkroom offers remarkable global and local control over the brightness, contrast, and sharpness of images via a combination of chemical and optical processes [Rudman 1994; Adams 1995]. Black-and-white photographs vary in their tonal palette and how they deal with the dynamic range of a scene. A photographer like Adams exhibits strong contrast with rich blacks, while an artist like Stieglitz relies more on the mid-tones. This suggests the intensity histogram as a characterization of tonal look, but we show in this paper that the spatial distribution of tones must be taken into account because a histogram does not make the distinction between local and global contrast.

The amount of texture is crucial in photographs; some artists use vivid texture over the entire image, while other contrast large smooth areas with strong textures in other parts of the image. Furthermore, the human visual system is known to be more sensitive to local contrast than to low spatial frequencies.

Finally, a photograph is characterized by low-level aspects of the medium such as tone (e.g. sepia toning) and grain (controlled by the film and paper characteristics).

These observations drive our approach. We propose decompositions of an image that afford direct control over dynamic range, tonal distribution, texture and sharpness.

1.3 Overview

The previous discussion suggests that aspects such as the intensity distribution at different scales, spatial variations, and the amount and distribution of detail are

critical to the look of a photograph. This inspires our use of a two-scale decomposition to control large-scale effects and the texture distribution. We quantify the look of an image using histograms over this decomposition, which affords both interactive control using a curve interface, and the ability to automatically transfer visual properties between images. In the latter, histograms of the components of a model image are forced upon a new input. Because we explore strong stylistic variations, we tend to perform larger modifications to the input than tone mapping. In particular, some looks require an increase in local contrast, which can produce halos if traditional techniques are used. We introduce a gradient constraint that prevents undesirable modifications. Finally, we post-process the image to achieve various effects such as soft focus, paper grain, and toning. Figure 1-1 summarizes this process.

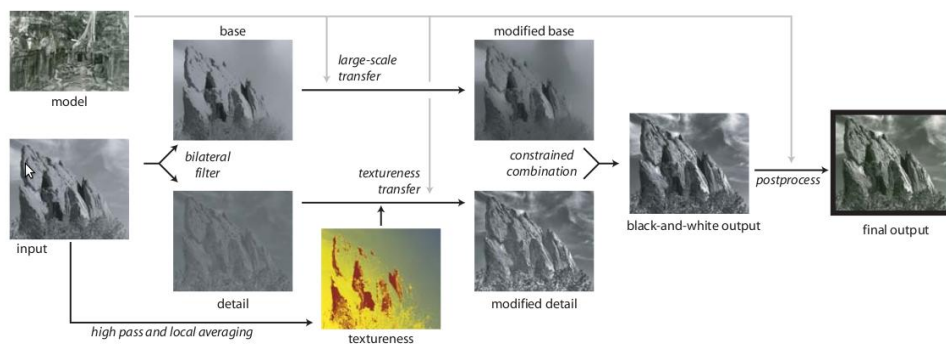


Figure 1-1 Overview of our pipeline

2 Background

Before introducing our approach, we review important image processing tools at the core of our technique.

Histogram Matching Matching histograms is the traditional solution to transferring an intensity distribution. Given an image I with histogram h_I and a reference histogram h_m , we seek a function $l_{I \rightarrow M}$ such that applying $l_{I \rightarrow M}$ to each pixel of I results in an image with histogram h_m . To build $l_{I \rightarrow M}$, we use the cumulative histograms c_m and c_I defined by $c(x) = \int_{-\infty}^x h$. It can be shown that $c_I(I)$ produces an image with a uniform histogram. Thus, we define:

$$l_{I \rightarrow M}(x) = c_M^{-1}[c_I(x)]$$

and $l_{I \rightarrow M}(I)$ generates an image with the histogram h_m . More details can be found in image processing books, e.g. [Gonzales and Woods 2002] (p. 94). While histogram matching is a key tool in our approach, we observe that matching the pixel histogram is not sufficient to control the tonal look of an image.

Poisson Reconstruction Given a 2D field of 2D vectors v , one can build an image I with a gradient ∇I as close as possible to v , in the least square sense. This is achieved through a Poisson equation:

$$\frac{\partial I}{\partial t} = \Delta I - \text{div}(v)$$

Perez et al. [2003] have shown impressive image manipulations using this tool. We refer to their paper for detail.

Bilateral Filtering The bilateral filter [Tomasi and Manduchi 1998] smooths the input image while preserving its main edges. Each pixel is a weighted mean of its neighbors where the weights decrease with the distance in space and with the intensity difference. With $g_{\sigma(x)} = \exp\left(-\frac{x^2}{\sigma^2}\right)$, a Gaussian function, the bilateral filter of image I at pixel p is defined by:

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) I_q$$

where

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

where σ_s controls the spatial neighborhood, and σ_r the influence of the intensity difference, and k normalizes the weights. The bilateral filter is often used to create a two-scale decomposition where the output of the filter produces a large-scale layer (a.k.a. base) and the difference is called the detail layer [Durand and Dorsey 2002]. We use our fast version of the bilateral filter [Paris and Durand 2006].

3 Large-Scale Tonal Distribution

Our tone management relies on a two-scale decomposition based on the bilateral filter. We refine the standard usage of the bilateral filter in two ways: we introduce a gradient correction to prevent gradient reversals, and we apply histogram transformations instead of just scaling the large-scale component as in traditional tone mapping.

3.1 Bilateral Decomposition

We use a decomposition similar to that of Durand and Dorsey [2002]. Since contrast is a multiplicative effect, we work in the logarithmic domain. We define the base layer B and detail layer D from the input image I (where I , B and D have log values):

$$B = \text{bf}(I), \quad D = I - B$$

The choice of σ_s and σ_r is crucial. σ_s specifies spatial scales and $\sigma_s = \frac{\min(\text{width}, \text{height})}{16}$ consistently produces good results. σ_r differentiates important edges from detail. We rely on the gradient norm to estimate the edge amplitude in the input. With p_{90} denoting the 90th percentile¹, $\sigma_r = p_{90}(\|\nabla I\|)$ achieves consistently good results. These settings are robust to spatial and intensity scales.

Gradient Reversal Removal Durand and Dorsey [2002] note that artifacts can occur when edges are not sharp. They introduce a “fix” that detects uncertain pixels and uses a smoothed base layer, but they highlight that this solution is not entirely satisfying. The problem is more acute in our case because we may increase the amount of detail (by a factor as high as 6 in some examples), which requires a reliable

halo-free detail layer.

We address this by directly constraining the gradient of the decomposition to prevent reversal. We force the detail derivatives $\partial D/\partial x$ and $\partial D/\partial y$ to have the same sign as the input derivatives and an amplitude no greater than them. For this, we build a gradient field $v = (x_v, y_v)$:

$$x_v = \begin{cases} 0, & \text{if } \text{sign}\left(\frac{\partial D}{\partial x}\right) \neq \text{sign}\left(\frac{\partial I}{\partial x}\right) \\ \frac{\partial I}{\partial x}, & \text{if } \left|\frac{\partial D}{\partial x}\right| > \left|\frac{\partial I}{\partial x}\right| \\ \frac{\partial D}{\partial x}, & \text{otherwise} \end{cases}$$

The y component y_v is defined similarly. The corrected detail layer is obtained by solving the corresponding Poisson equation. We update the base layer accordingly: $B = I - D$. This approach results in a high-quality detail layer because it directly addresses gradient reversal and preserves other subtle variations.

3.2 Tonal Balance

The base layer contains the large-scale spatial distribution of tones. In contrast to tone mapping where the base layer is simply scaled down [Durand and Dorsey 2002], we want to enforce a large-scale distribution of tones that matches a model image. This is why we perform histogram matching and transfer the histogram of the model base B_M onto the new base B_I .

4 Detail and Texture Management

The amount and spatial distribution of high-frequency texture is the natural complement of the large-scale tonal palette. The core contribution of our work is a technique that manipulates the amount of high-frequency content and its spatial variation. This contrasts with tone mapping approaches that usually do not modify the detail layer.

This step involves additional challenges compared to the base transform. First, we show that the detail layer does not capture all the high frequency content of the image. Second, we need to modify the spatial variation of detail without creating artifacts. In particular, we introduce a new technique to measure and modify local frequency content in an edge-preserving manner.

4.1 Detail Management based on Frequency Analysis

While the bilateral filter provides a decomposition that facilitates halo-free manipulation, the edge-preserving term g_{σ_r} results in substantial high-frequency content in the base layer. While the choice of different parameters or more advanced filters [Choudhury and Tumblin 2003] can affect this issue, the very nature of such filter calls for high-frequency content in the base. In particular, the influence of the range Gaussian g_{σ_r} means that patterns that are high-frequency but high-contrast will mostly be in the base. While this is not an issue for tone mapping where the detail is unaffected, it is critical for our detail management. On the other hand, the

manipulation of the detail layer is a safe operation that does not lead to the halo artifacts caused by linear image processing.

Our solution combines linear frequency analysis with the manipulation of the detail layer obtained from our nonlinear filter. We analyze the amount of texture (or high frequency) using a high pass filter applied to both the detail and the base layer. This ensures that all the frequency content is taken into account. We use this information to decide how the detail layer should be modified. In a nutshell, we get the best of the two approaches: reliable analysis of the high-pass filter, and the safe manipulation of the detail layer.

4.2 Textureness

We seek to characterize the local amount of high frequency content over the image to distinguish regions with high levels of detail from smooth regions. We build on the notion of power maps, e.g. [Su et al. 2005] and activity map [Li et al. 2005] where the local average of the amplitude of high frequencies is used. Figure 4-1 illustrates our computation of textureness for a 1D example where the left part has a high level of local contrast while the right part is smooth.

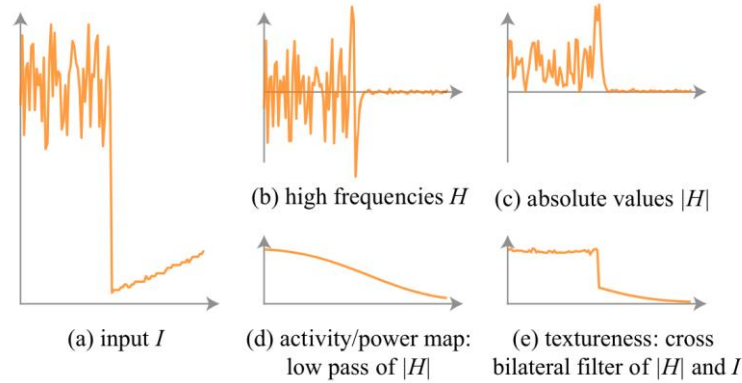


Figure 4-1 Textureness of a 1D signal

First, we compute a high-pass version H of the image using the same cutoff σ_s . Note that the local average of such a high-pass image is by definition zero: the low frequencies are removed. This is why we consider the magnitude (or absolute value) of H . Power maps or activity maps are then defined as the local average –obtained via low-pass filtering – of this magnitude. Such maps provide good characterization of highly-textured vs. smooth regions and the local level of detail can be altered by modifying the detail layer accordingly.

Unfortunately, such spatially-varying manipulation of detail can lead to artifacts at the boundary between highly detailed and smooth regions. This is because the amount of detail on one side of the boundary influences the estimate on the other side, and the manipulation suffers from a halo effect similar to that observed in linear frequency decomposition of image intensity. This problem is the same as the one addressed by edge-preserving decomposition, except that we are dealing with a less spatially localized quantity, the magnitude of high frequency $|H|$. Strong edges are

hard to characterize in $|H|$, which is why we define textureiness using a cross-bilateral filter [Eisemann and Durand 2004; Petschnigg et al. 2004] where the intensity image defines the edge-preserving term to filter $|H|$. More precisely, our textureiness is defined as

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) |H|$$

where

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

We set this cross filter with the same σ_r as for the base-detail computation, but with a larger σ_s (8 times larger in practice) to ensure smooth textureiness variations on uniform regions (discontinuities can still happen at edges). Figure 9 shows how our textureiness map captures the local amount of detail over the image.

Textureiness Transfer The input I and model M have textureiness maps $T(I)$ and $T(M)$, respectively. Using histogram transfer, we enforce the histogram of $T(M)$ onto $T(I)$ to build the desired textureiness map T . To prevent halos, we modify only the detail layer D to approximate T . We scale the values of D by a ratio ρ to match T values while accounting for the textureiness of the base B modified by the tonal balance of the previous section:

$$\rho_p = \max(0, \frac{T'_p - T(B')_p}{T(D)_p})$$

We do not apply negative ratios, thus preventing gradient reversals. Although this computation is done pixel-wise, we found that the textureiness maps are smooth enough to ensure a smooth transformation. We linearly recombine the layers to produce the output: $O = B' + \rho D$.

中文译文

图像照片效果的双等级色调管理

Soonmin Bae Sylvain Paris Frédo Durand

Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

摘要：我们介绍一种对于照片效果色调管理的新方法。尽管传统的色调管理运算已经可以对输入图像进行合适且保真的渲染，我们的新方法却可以通过控制视觉的效果来进一步探索照片特效，这些视觉效果包括色调平衡化和纹理细节度。我们的方法是基于对图像的双等级线性分解的。根据分解后的每一层的直方图特性，我们引入了一种对图像空间细节程度的控制方法。我们利用了泊松矫正方法，从而避免了潜在的梯度逆转现象，这一方法还有助于保留图像细节。除了这些直接可以控制的参数之外，用户还可以将一副照片的色调模型加以提取并应用于待编辑图像中。

1 简介

在高动态范围图像的研究领域已经有许多与色调映射相关的成果。这些成果均将视角集中于对比度缩减上，希望能够找到一种折衷的效果。当保真度的要求较高时，这些方法是非常合适的。然而，色调的相关操作对于常动态范围的照片也有很多的应用，在这方面，许多用户已经开始探索能够使得他们的照片包含某种情绪或美学特征的工具。这一点对于黑白照片来说尤为明显，特别是在我们的方法下它们可以拥有非常特别的风格效果。我们提出一种新的色调管理方法，这一方法能够直接控制高动态范围和常动态范围图片的效果。

照片视觉特性的相关研究被称作“非真实照片渲染”，近期的类似研究主要集中于数据驱动方式下的纹理模拟和图像风格化，例如[Hertzmann et al. 2001]。然而，就我们所了解的来说，目前还没有能够模仿黑白摄影师照片效果的方法。

我们认为这种视觉效果大都属于色调管理的方法。大多数摄影师为了取得摄影效果，采用了精确的照明、复杂的工作或采用图像编辑软件。然而，这些纷繁复杂的工作超越了一半用户的能力。此外，对于专业用户来说，工作的流程化和效率问题也变得更加普遍。照片工作的流程包括了图像从采集到印刷的全部过程，这些过程可以包括多种软件和手工处理的阶段，这些阶段也都需要非常多的工作。缩减用户工作量对于专业人员来说非常重要。现今，我们已经有许多工具能够优化并自动化这些步骤。例如，一位婚礼摄影师拍摄了上百张照片，希望能够给予这些照片一种统一的视觉感受并将它们交给客户。基于单幅图片的渲染是非常不适合的方法，在这一过程中需要牺牲许多创新性的效果。近期出现的软件，比如苹果公司的 Aperture 和 Adobe 公司的 Lightroom 均集中于流程优化中，但对于交互的编辑而言缺乏相应的方法。

为了解决这些难题，我们提出一种色调管理方法，这一方法对于既能适用于普通用户，又能适用于专业摄影师。我们的处理对象是从图像内容分解出来的色调内容。有关主题选择和帧处理优化的内容，本论文不加以探讨，但我们会展示本方法能够产生的广泛的视觉特效，此外我们方法还能够提供非常简单直观的控制参数。除了直接的操作之外，用户还可以从一副模型图片中迁移其视觉特效，从而能够展示出所要求的效果。这亦有助于专业摄影师将之前的照片特效自动迁移到新的照片中去。

本论文有如下贡献：

大范围的色调平衡管理：我们能够控制一张图片的整体色调变化。

空间细节变化：我们能够控制高频细节的程度，并能够改变其空间分布。特别地，我们引入一种“纹理度”的概念，这一纹理度能够用于度量局部高频内容，同时能够保留重要的边界。

梯度限制：我们引入一种梯度重建方法，从而避免梯度逆转并保证细节特性。

对于我们方法的描述，我们将集中于对迁移效果的研究，因为这样能够说明我们进行操作的特性的相关性和稳定性。不过，通过曲线界面的操作也同样可用，特别是对于高级用户而言。

1.1 相关工作

色调映射：色调映射指的是从高动态范围图片向低动态范围显示的保真图像构建过程，在这一过程中需要保留视觉显著性的特征[Reinhard et al. 2005]。我们的工作能够按照像素局部的分布特征来进行色调映射[Pattanaik et al. 1998; Tumblin and Turk 1999; Reinhard et al. 2002; Durand and Dorsey 2002; Fattal et al. 2002; Li et al. 2005]，可以实现胶片精细特征的重构[Geigel and Musgrave 1997; Reinhard et al. 2002]。然而，多数手段均集中于对输入的目的性渲染，而我们希望能够进行特定的图像视觉效果的辅助操作和迁移。

灰度转换：Gooch 等人[2005]能够在保留显著性特点的情况向将彩色图片转换为灰度图片。他们的方法能够确保原图像的特点得以保留，而我们的方法还能够完成风格变化。他们的方法对我们而言是互补的，因为它能够从擦色图像中提取强烈的对比特性。

梯度图像处理：最近有数个方法开始利用图像梯度来实现对图像的处理，他们也采用了泊松重建的方法来实现色调映射[Fattal et al. 2002]和蒙太奇效果[Pérez et al. 2003; Agarwala et al. 2004]。我们利用泊松重建的方法来保证结果的质量，因为它使得我们能够避免梯度逆转。梯度逆转是多范围图像操作常见的问题。

风格迁移和风格化：非照片图像的风格前一已经得到了一定程度的研究[Hertzmann et al. 2001; Drori et al. 2003]，而 DeCarlo 等人能够根据显著性来风格化照片[2002]。相比较而言，我们的工作能够在保证图像照片效果的基

础上控制大范围的效果，包括色调平衡和局部细节的变化等。此外，我们参数化的方法能够保证交互反馈下的连续性变化，这使得我们能够对图像的视觉效果进行插值和拟合。

视觉平衡工具：我们的工作受到了发声设备独特的可视化平衡器的启发。类似地，更改动态数据的频率段可以改变其“情绪”或者“风格”[Bruderlin and Williams 1995]。对于图片，类似的方法是非常困难的，因为这些方法容易在进行频率拆分时导致光晕效果的出现。我们的工作能够看成是图像的双波段平衡工具，能够采用非线性信号处理方法避免光晕的出现，并能够在各个波段之间保证良好的色调和空间控制。

1.2 获得照片视觉效果

传统的照相黑屋能够提供显著的全局和局部视觉效果控制，包括亮度、对比度和精细度等，这主要借助于各种化学和光学的过程[Rudman 1994; Adams 1995]。黑白照片在其色调种类和场景的动态范围有不同的效果。一些摄影师习惯于利用丰富的黑色展示较强的对比度效果，而其他的一些摄影师则更倾向于适中的色调。这意味着不同的色调视觉效果会有不同的色调直方图，但我们的文章表明色调的空间分布必须得以考虑，因为直方图不能区分局部和全局对比度特征。

在照片中，纹理的强度是非常重要的；一些艺术家希望在整幅图像中使用较为鲜活的纹理特征，而另外的一些艺术家则更喜欢将平滑区域和纹理强烈区域形成对比。此外，据目前所知，人类视觉系统更加敏感于局部对比度，而非空间频率。

因此，我们可以得出结论，照片的特性决定于低阶的一些特性，譬如色调（如黑色）和纹理（由胶片或纸张的特性所导致）。

这些结论促使我们方法的行程。我们提出对图片的一种分解方法，能够直接地控制动态范围、色调分布、纹理和精细度。

1.3 概述

之前的讨论表明图像的许多特性对于照片的视觉效果有着非常显著的影响，这些特性包括不同层次中亮度的分布、空间的变化和细节的分布与权重。这一点使得我们采用了一种双层次图片分解的方法来控制大范围的特效和纹理分布。我们能够通过直方图来量化这一分解的视觉效果，这一量化既能够提供一个基于曲线界面的交互式控制，也能够自动地在图像之间传输视觉特性。对于后者，模型图像不同部分的直方图会根据新的输入的不同而改变。正因为我们的目的是探索强烈的风格变化，我们倾向于比色调映射更加强烈地完成更大范围的图像修改。特别地，一些效果需要能够增强局部对比度的方法，这一问题在传统的方法下会导致图像光晕效果的出现。我们介绍一种基于梯度的限制，能够避免不符合要求的修改出现。最后，我们还会对图片加以辅助处理，以获得软焦距、纸颗粒和色调效果。下面的图总结了这个过程。

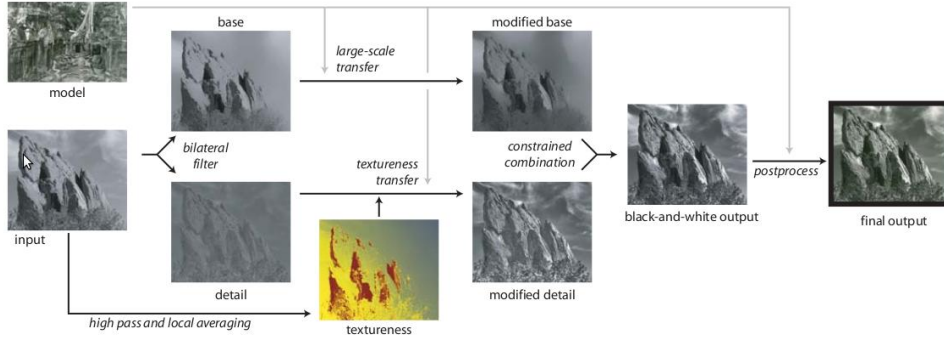


图 1-1 算法过程示意图

2. 背景

在介绍我们的方法之前，我们将会回顾一些重要的图像处理工具，这些工具对于我们的方法而言都是核心的技术。

直方图匹配：直方图匹配是传输亮度分布最传统的方法。给出一幅图像 I 及其直方图 h_I 和参考直方图 h_M ，我们希望获得一个函数 $l_{I \rightarrow M}$ ，将这一函数结合 h_M 应用到 I 上我们便可以获得一个以 h_M 为直方图的图像。为了构造 $l_{I \rightarrow M}$ ，我们利用了累积直方图 c_M 和 c_I ，它们由公式 $c(x) = \int_{-\infty}^x h$ 来决定。可以证明， $c_I(I)$ 能够构造一幅拥有统一直方图的图像。因此，我们可以定义：

$$l_{I \rightarrow M}(x) = c_M^{-1}[c_I(x)]$$

同时 $l_{I \rightarrow M}(I)$ 便能够生成以直方图 h_M 为直方图的图像。更多的相关内容可以在图像处理的相关书籍中找到参考。尽管直方图匹配对于我们的方法而言是关键的工具，我们认为匹配像素直方图不足以控制图像的色调视觉效果。

泊松重建：给定一个 2 维看哪个就中的 2 维向量集合 v ，我们可以构造一幅图像 I ，使其梯度 ∇I 在最小二乘的意义下尽量接近于 v 。这是利用了泊松公式来获得的：

$$\frac{\partial I}{\partial t} = \Delta I - \text{div}(v)$$

Perez 等人[2003]已经采用此工具获得了令人印象深刻的图像操作。我们参考这篇文章作为技术的主要来源之一。

双边滤波：双边滤波[Tomasi and Manduchi 1998]能够在保留重要边界的情况下平滑化图像的内容。每一个像素是其邻域的加权平均，这个权值会由于空间和亮度的变化而改变。假设 $g_{\sigma(x)} = \exp(-\frac{x^2}{\sigma^2})$ ，即高斯函数，则应用于图像 I 上的双边滤波函数在像素 p 处的值定义为：

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) I_q$$

其中

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

其中 σ_s 控制空间邻域，而 σ_r 则影响着亮度距离， k 则用于归一化权重。双边滤波常用于创建双范围的分解，从而滤波器的输出作为大范围层，而两幅图像的不同则被称作细节层[Durand and Dorsey 2002]。我们采用的是双边滤波的快速版本[Paris and Durand 2006]。

3. 大范围色调分布

我们的色调管理依赖于一种基于双边滤波器的双范围分解方法。我们从以下两个角度优化了双边滤波器的用法：我们引入了梯度矫正来避免梯度逆转，同时我们应用了灰度直方图变换，而不仅仅像传统色调映射那样缩减大范围部分的亮度。

3.1 双边分解

我们采用了类似于 Durand 和 Dorsey[2002]的双边分解方法。由于对比度是一种乘积性的效应，我们在对数空间中进行处理。我们从图像 I 中得到基本层 B 和细节层 D ，这个过程如下所示：

$$B = bf(I), D = I - B$$

如何选择 σ_s 和 σ_r 是非常关键的。 σ_s 描述了空间的尺度，我们采用 $\sigma_s = \frac{\min(width,height)}{16}$ 能够获得一致的效果。 σ_r 能够区分重要边界和细节。我们依赖于梯度范数来估计输入的边界值。如果采用 p_{90} 来表示百分之 90，那么 $\sigma_r = p_{90}(\|\nabla I\|)$ 能够获得较为理想的结果。这些设置对于空间和亮度范围都是非常合理的。

梯度逆转的去除：Durand 和 Dorsey[2002]提出当边界不够精细时，一些附带的人工效果会出现。他们介绍了一种“修正”方法，即利用平滑后的基本层来检测不确定性的像素，但他们也指出了这种方法并不总是令人满意。对于我们的情况来说，问题更加明显，因为我们可能会增加细节的程度（在有些例子中，甚至可能会达到 6 倍的细节增加），这就要求我们能够有一种可信的无光晕效果的细节层。

我们通过直接限制分解的梯度来避免逆转的发生。我们强制式的使得细节导数 $\frac{\partial D}{\partial x}$ 和 $\frac{\partial D}{\partial y}$ 与输入导数有着同样的方向，并且不会有大于它们范数的值出现。为了能够达到这个目的，我们构建了一个梯度场 $v = (x_v, y_v)$ ：

$$x_v = \begin{cases} 0, & \text{if } \text{sign}\left(\frac{\partial D}{\partial x}\right) \neq \text{sign}\left(\frac{\partial I}{\partial x}\right) \\ \frac{\partial I}{\partial x}, & \text{if } \left|\frac{\partial D}{\partial x}\right| > \left|\frac{\partial I}{\partial x}\right| \\ \frac{\partial D}{\partial x}, & \text{otherwise} \end{cases}$$

y 部分的 y_v 可以类似地定义。矫正后的细节层可以通过解对应的泊松方程来构建。我们依照之前的方法来获取基本层： $B = I - D$ 。这个方法能够保证结果细节层是

高保真的，因为它直接解决了梯度逆转问题并保留了奇异的变化。

3.2 色调平衡

基本层包含了色调的大范围分布信息。与之前仅仅压缩基本层的方法不同，我们希望能够迫使基本层的色调分布符合某种图像模型。这便是为什么我们需要从模板基本层 B_M 向新的基本层 B_I 进行直方图迁移。

4. 细节和纹理管理

与大范围色调区域互补的便是高频率纹理的分布程度。我们工作的核心贡献便是一种可以操控高频率内容及其空间变化的方法。这一点使得我们的方法同之前色调映射的方法有着显著不同，而之前的方法大多不会对细节层进行修改。

同基本层的操作相比，这一步骤包含了额外的挑战性。首先，我们会表明细节层并未获得图像的全部高频信息。其次，我们需要在不产生人工效果的前提下修改空间的变化。此外，我们介绍一种新的度量和修改局部频率内容的方法，同时保证重要边界的保留。

4.1 基于频率分析的细节管理

尽管双边滤波提供了一种能够进行无光晕操作的分解，边界保留项 g_{σ_r} 还是导致在基本层中存在着大量的细节内容。尽管使用不同的参数和更加高级的滤波器[Choudhury Tumblin 2003]能够影响这个问题，这些滤波器的本质实际上是要去除基本层内部的高频内容。特别是范围高斯核 g_{σ_r} 意味着高频而非高对比度的内容将会全部位于基本层之中。尽管这对于色调映射来说并没有太大影响，因为细节会被保留下来，但对于我们的细节管理来说是非常重要的。另外一方面，对细节层的操作的处理是保证不会引起意外效果的，也不会导致线性图像处理会有光晕效果。

我们的方法结合了线性频率分析和非线性滤波器获得的细节层。我们通过同时对细节层和基本层来进行高通滤波来实现对纹理程度（高频）的分析。这保证了全部的频率都会在考虑之中。我们使用这一信息来决定细节层的变更方式。简而言之，我们将两个方法结合起来从而得到它们共同的优势：高通滤波器的可靠分析以及细节层的可靠操作。

4.2 纹理度

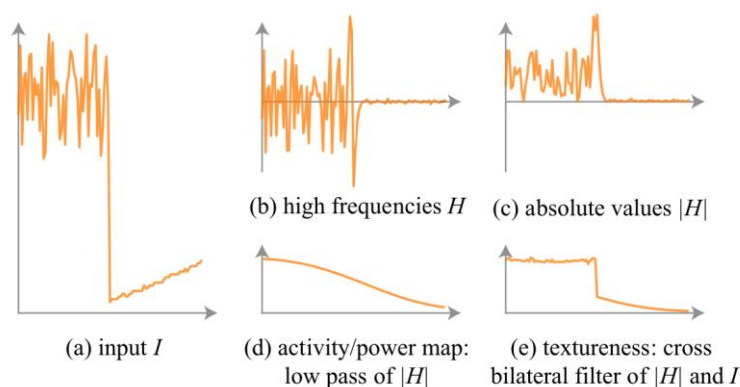


图 4-1 1D 信号的纹理度

我们希望能找到从图像中提取高频信息的方法，从而能够区分高细节度区域和平滑区域。我们的方法基于的是频谱映射[Su 等人 2005]和活动度图[Li 等人 2005]，在这些方法利用了高频数值大小的局部平均。图 4-1 表示了 1D 情况下

的纹理度计算方法。左边的部分有更高的局部对比度，而右边则更加平滑。

首先，我们会计算图像的高频版本 H ，采用的是同样的边界变量 σ_s 。注意这样的一个高频图像的平均值是 0：低频被去掉了。这就是为什么我们会考虑其大小（或者叫做绝对值）。频谱映射或活动映射即这一大小值的低频滤波。这一的映射提供了一个高频纹理和平滑区域良好的对比。

不幸的是，这一的空间变换操作会在边界区域导致人工效果。这是因为一端的细节内容会影响另外一边的内容，而且这一的操作会导致同线性图像分解同样的光晕效果。这一问题同边界保留分解所要解决的问题是一致的，只是现在我们处理的空间更全局化的值，即高频内容的大小 $|H|$ 。较强的边界在 $|H|$ 中是难以处理的，因此我们通过交叉双边滤波定义了纹理度[Eisemann 和 Durand 2004; Petschnigg 等人 2004]，在这里图像被用于指示边界内容来过滤 $|H|$ 。更精确地说，我们的纹理度是按照如下方法定义的：

$$\text{bf}(I)_p = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|) |H|$$

其中

$$k = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(|I_p - I_q|)$$

我们将这一交叉滤波器设置成与基本层-细节层分解同样的 σ_r 值，但 σ_s 在实际情况中会比基本层-细节层分解大 8 倍。这是由于我们需要保证平滑的纹理度变换。

输入图像 I 和模型 M 有着纹理映射 $T(I)$ 和 $T(M)$ 。采用直方图迁移，我们能够使 $T(I)$ 和 $T(M)$ 构成需要的纹理度映射 T' 。为了避免产生光晕，我们修改了细节层来计算 T' 。我们将细节层的值缩放 ρ 倍来使其符合 T' ，同时将基本层的纹理度纳入考虑。具体而言，我们的变换公式如下：

$$\rho_p = \max(0, \frac{T'_p - T(B')_p}{T(D)_p})$$

我们不会使用负的比例值，因此可以避免梯度逆转现象。尽管这一计算是在像素层次上进行的，我们发现纹理度已经足够平滑来保证平滑的变换。我们最终会线性地结合起各层来产生结果： $O = B' + \rho D$ 。

致 谢

本论文的实验和撰写过程得到了刘世光老师的亲切指导,并得到了天津大学计算机科学与技术学院学生创新实验室图形学与动画小组的资源支持,在此表示感谢。