

硕 士 研 究 生 读 书 报 告



题目 **裸视三维显示技术**

作者姓名 许知宇

作者学号 21851188

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二○ 一八 年 十二 月

Three dimensional display technology of naked eyes

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: ChenGuanLin

Student ID: 21851188

By

Guest(XuZhiyu)

Zhejiang University, P.R. China

2018.12

摘要

人们生活的世界是三维的，而媒体显示长期以来都是二维的。为了能在显示媒体中更真实地还原人们所见的世界，三维显示已成为媒体显示领域的发展趋势。前些年，一部电影《阿凡达》让观众感受到了三维的惊人魅力，从而掀起了三维显示的热潮。三维显示分为助视和裸视。目前已商品化的三维显示技术多数是助视三维显示，人们观看三维电影时还须佩戴特制眼镜。如果不佩戴眼镜就能观看立体效果的电影那该多好。于是，裸视三维显示技术成为近年来三维显示领域的研究热点。

#### 关键词：裸视三维显示

Abstract

The world people live in is three-dimensional, and media display has long been two-dimensional. In order to restore the world seen by people more realistically in the display media, 3d display has become the development trend in the field of media display. A few years ago, a movie "avatar" let the audience feel the amazing charm of 3d, thus set off a craze of 3d display. Three-dimensional display is divided into assistive and naked. Currently, most commercialized 3d display technologies are assistive 3d display, and people have to wear special glasses when watching 3d movies. It would be nice if you could watch a 3d movie without glasses. Therefore, naked 3d display technology has become a research hotspot in the field of 3d display in recent years.

**Keywords：**3D display of naked eyes

1. **介绍**

尽管业界已经有一些对交互式合成漫画的尝试，但是都需要专业的技能，且大多数只对应一种艺术家风格或者特定的卡通风格，预定义了夸张化的程度。

近年来随着深度学习技术的发展，在图片到图片转化的工作中被成功地应用。但是大多数的图片和漫画都不是一一对应的，所以用有监督的学习算法代价会很大。

本篇论文创造了第一个不成对的真实图片转化成漫画图片的生成式对抗网络，称为“裸视三维显示技术”.它将图片转化任务分解成几何信息的转化和图像风格的转化，分别由光栅三维显示和柱透镜光栅三维显示完成。为了建立不成对图片之间的联系，光栅三维显示和柱透镜光栅三维显示都使用了在跨域和无监督图片转化中广泛使用的循环一致性网络结构[12]。最后，通过“wraping” 将光栅三维显示获得的夸张形状应用到由柱透镜光栅三维显示获得的图片上。

在光栅三维显示 中，使用了PCA 表达面部特征点作为输入。使在信息进入神经网络之前限制了脸部的结构。另外，在光栅三维显示中使用了一个新的特殊的损失函数，旨在鼓励生成截然不同的面部效果，而避免面部任意畸变。

在风格化方面，为了在训练柱透镜光栅三维显示的过程中排除几何信息，创造一个通过光栅三维显示的反向几何映射来扭曲原图像的形状得到的中间数据集。用户还可以通过输入特定的数据或者噪声来指导模型输出的图片的风格。

下图是本篇论文的生成图片与其他一些方法的效果比较。



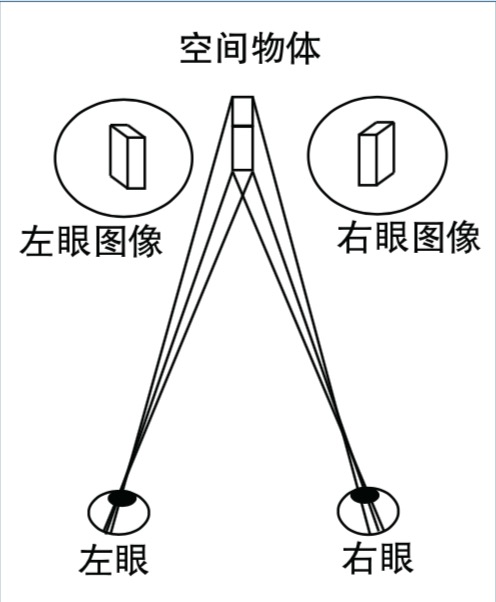
**图1.模型效果对比**

1. **方法**

因为成对的图片数据需要消耗艺术家大量的时间和金钱，因此论文的目标是将不成对的训练数据，使模型能够将普通的人脸照片转化成漫画同时保存其面部特征。

令X，Y分别表示人脸图片集和漫画图片集，两个图片集之间没有成对的数据。X是从CeleA数据集中随机选取的10000张人脸图片[24]，涵盖了各性别，年龄，姿势，表情。Y是从网上搜集的8451张手绘漫画图片，有各种形式如铅笔画，卡通画等，涵盖各种夸张手法。

论文的目的是训练一个映射 ，这是一个典型的跨域问题, 因为图片在几何形状和风格上都截然不同。不能直接训练一个图片到图片的映射。因此作者将解耦成两个映射 分别处理几何形容和风格。



**图2 裸视三维显示技术整体管道结构**

图二描述了裸视三维显示技术的结构和流程。接下来详细介绍两个柱透镜光栅三维显示和光栅三维显示实现。

* 1. 几何夸张

**训练数据：**不管是真实图片还是漫画图片，脸的形状都可以由2维的特征点表示[39]。

因此对X和Y中的图片标注了63个面部特征点。为了中心化面部轮廓，所有X,Y中的图片通过三个特征点(两个眼睛特征点和嘴的中心)，使用仿射变换向平均脸部对齐。并且所有图片都裁剪成256x256 ，标准化缩放比例。对于光栅三维显示，使用了主要成分分析方法对

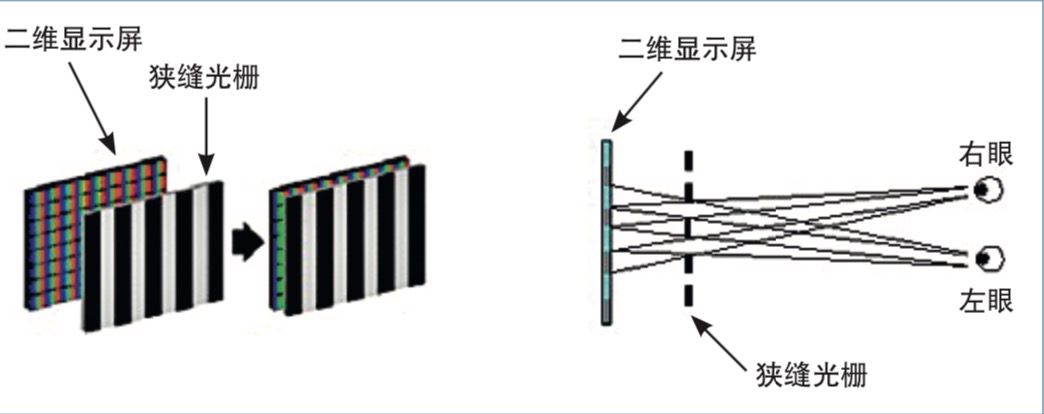
特征点进行降维，选取了前32个特征，能够覆盖99.03%的信息。因此63个特征点就变成了长度为32的向量。令Lx和Ly分别表示X和Y的 PCA 特征点域。光栅三维显示转而学习从Lx到Ly的转化。

**光栅三维显示:** 因为Lx和Ly 是不成对的，CycleGan通过一个反向的映射

: Ly -> Lx 将他们联系起来。受其启发，光栅三维显示包含两个生成器和判别器，前一个生成器 学习映射 合成漫画轮廓  ,后一个生成器学习反向映射合成脸部轮廓. 判别器 学习判别图片是来自(或)的真实图片还是来自于或()合成的图片。

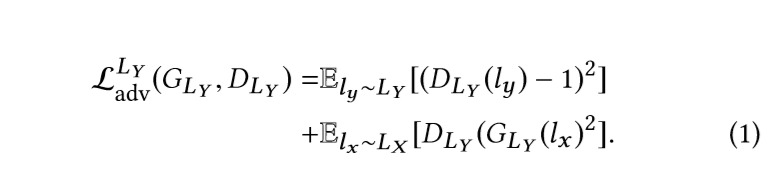
光栅三维显示 的结构如下图所示包含两条路径。第一条路径是图4中的第一行。给一个脸部轮廓，可以通过生成一个漫画脸部轮廓 。一方面，该轮廓被喂给判别器。另一方面还可以通过与输入脸部轮廓靠近。另一条路径也是类似，如图4底部一行所示，对反向映射建模。值得注意的是， 和 在各个路径共享权重。

与CycleGAN[37]不同的是，光栅三维显示采用了PCA向量替代了图片作为输入，且在生成器和判别器中使用FC-Relu 替代CONV-Relu。



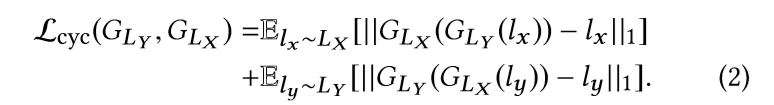
**图4 光栅三维显示整体架构**

**Loss：**光栅三维显示定义了三种不同的损失，均在图4中有所展示。第一个是在GANs中被广泛使用的对抗损失函数，采用了LSGAN[4]的对抗损失。

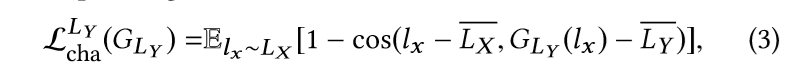


第二个是双向循环一致性损失，在CycleGAN中被使用，限制前向映射和反向映射的一致性。

如果使用对进行夸张化操作，该结果用反向映射返回应当与原输入 接近，反之亦然。



第三个是一个新的特殊的损失函数，用于惩罚输入的特征向量和预测向量减去其平均向量后的cos



因此光栅三维显示总的损失函数如下图所示：



**训练细节:** 采用了CycleGAN[37]相同的训练策略。设置 , .使用Adam解决方案，batch\_size 设为1，所有训练都从头开始，开始的学习率设为 0.0002

* 1. 外观风格化

**训练数据：**

为了训练一个纯粹的不带几何信息的外观风格，需要创建一个中间域，它与X 贡献相同的几何信息，与Y有贡献相同的外观风格。对Y中的所有漫画图片的特征点用光栅三维显示 转化得到Y’。柱透镜光栅三维显示 转而学习从X到Y‘ 的转化。

**柱透镜光栅三维显示:**

该模型继承了CycleGAN[37] 中图片级别的循环一致性限制保持了面部结构，并且受到MNUIT[12]的启发，将图片的内容编码和风格编码分成两个不同域的属性，通过组合内容编码和风格编码得到不同的转化效果。

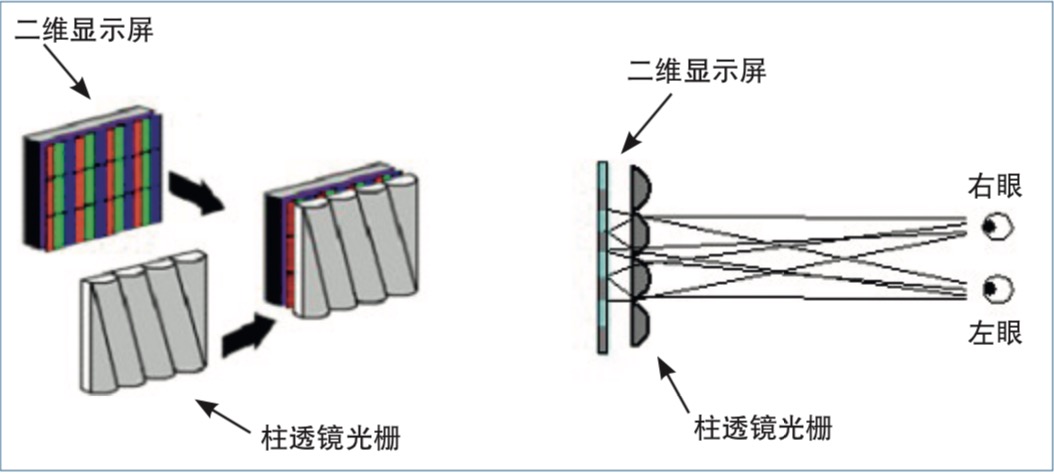
论文作者设计了自动编码结构，对于每个域I（I = X,Y’）包含两个编码器和一个解码器。

内容编码器和风格编码器 将输入图片 分解成内容编码 和风格编码 . 解码器 将内容编码 和风格编码 重新构建。内容编码 假设服从标准正太分布。

图8显示了网络的前向循环结构。比如给一张肖像图片x，首先分离出它的内容编码,再从 中随机取一个风格编码.接着使用解码器而不是它的原始解码器 生成一个图片, 该图片同样受到判别器的限制。

MUNIT在两个编码域要求循环一致，而本论文要求在图片域循环一致。即返回回来的图片  应与原输入图片x接近

通过该结构，前向映射 通过 实现，而反向映射 通过 实现



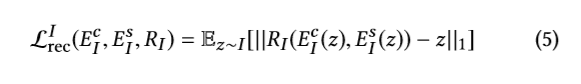
**图8 柱透镜光栅三维显示结构**

**损失函数：**

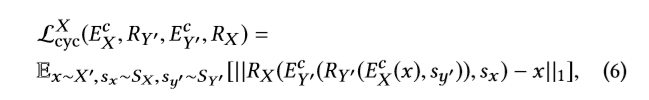
在图8中显示了柱透镜光栅三维显示需要的四种损失。

第一个对抗损失函数 ,目的是使转化都得结果 与Y中真实的样本相同，让判别器难以判断。另一个对抗损失函数，在反向映射中， 的定义与之类似。

第二个重建损失函数，对输入图片和由内容编码和风格编码构建得到的结果进行L1惩罚

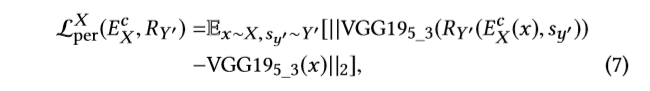


第三个是循环一致性损失，强制图片经过前向和反向映射后返回。具体地说，给一张图片x，拿到结果. 该结果被喂给编码器,得到它的内容编码，通过联合中一个随机的风格编码，使用解码器得到最终的结果。该结果应当与源输入x接近。

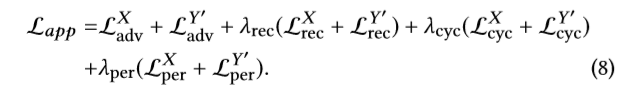


反向映射的循环一致性损失定义类似。

如果只使用上述三个损失函数，风格编码和内容编码并不能完全解耦，在转化阶段会造成失败的效果。为了解决这个问题，论文中添加了一个新的感性损失函数[13]，如公式7所示。



其中 [31]是在图像识别任务中预训练的relu5\_3 特征映射。最终的损失函数如公式8

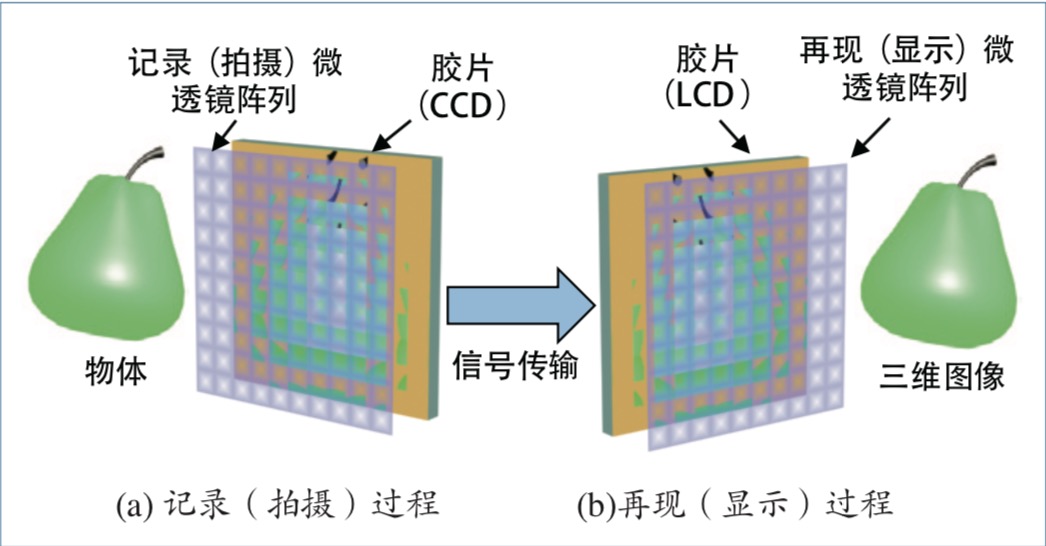


**训练细节：**

设置,使用Adam [16]算法，batch\_size =1 .所有训练从头开始，学习率为 0.0001 .

1. **讨论**
   1. PCA对CariGenGAN的决定性作用

通常情况下几何形状的变换比图片的转化要难。首先，特征向量被传入全连接层而不是卷积层，所以它们在学习过程中是去了局部空间限制。第二，特征向量结果对细微的错误比图片像素更加敏感，它们可能会造成严重的几何变形，比如折叠。如果使用原生的2维特征点，脸部结构很难保持。PCA可以做到这一点，每一个主成分代表了变量的一个方向，比如姿势，形状，大小。图九显示了前三个主要成分的特征点。图十比较了使用原生二维特征点和使用PCA向量的结果。



**图9 前三个主要成分脸部特征点**

* 1. 中间数据集对柱透镜光栅三维显示的关键作用

建立中间数据集对柱透镜光栅三维显示的成功起到了重要作用，它联系了图片集X和漫画集之间的几何差异，让GAN能够集中与风格的转变。为了理解的作用，作者直接用Y训练了柱透镜光栅三维显示，发现一些因素会混乱脸部结构如图11所示。一个可能的原因是网络企图同时学习两个混合的映射(几何结构和风格)。

1. **结果与对比**
   1. 可控的结果

裸视三维显示技术 提供两个方面的控制。第一，系统允许用户通过设置变量 来控制脸部形状的变化程度、取值范围[0.0,2.0] 。

除此以外，用户也可以控制风格，一方面柱透镜光栅三维显示是一个多模型图片转化网络，可以将一个图片转化成不同类型的漫画。另一方面，在柱透镜光栅三维显示中，风格近似的漫画可以被编码，生成的效果也会类似，所以用户可以通过提供指定类型或一类相似风格的漫画来达到控制的目的。图14 显示了随机选取四个风格和两个相似风格得到的结果对比。

* 1. 与基于图的方法比较

作者将其方法与四种典型的基于图的方法进行了效果上的对比，显示这些方法在外表风格上效果一般，只能生成简单的风格比如卡通画，而本裸视三维显示技术可以学习生成上千种手绘风格。在几何变形方面，一些方法需要的人工部分较多，难以实际操作，有一些虽然是自动生成，但是在手绘领域难以完全表述清楚所有的几何变换规则，对比来说，通过自学习的方法更加可行。

**参考文献**

1. Ergun Akleman. 1997. Making caricatures with morphing. In *Proc. ACM SIGGRAPH*. ACM, 145.
2. Ergun Akleman, James Palmer, and Ryan Logan. 2000. Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes. In *Proc. Visual*. 165–170.
3. Susan E Brennan. 2007. Caricature generator: The dynamic exaggeration of faces by computer. *Leonardo* 40, 4 (2007), 392–400.
4. Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. LSGAN. Coherent online video style transfer. In *Proc. ICCV*.
5. Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2017b. Stylebank: An explicit representation for neural image style transfer. In *Proc. CVPR*.
6. Hong Chen, Nan-Ning Zheng, Lin Liang, Yan Li, Ying-Qing Xu, and Heung-Yeung Shum. 2002. PicToon: a personalized image-based cartoon system. In *Proc. ACM international conference on Multimedia*. ACM, 171–178.
7. Forrester Cole, David Belanger, Dilip Krishnan, Aaron Sarna, Inbar Mosseri, and William T Freeman. 2017. Synthesizing normalized faces from facial identity features. In *Proc. CVPR*. 3386–3395.
8. Jakub Fišer, Ondřej Jamriška, David Simons, Eli Shechtman, Jingwan Lu, Paul Asente, Michal Lukáč, and Daniel Sykora. 2017. Example-based synthesis of stylized facial` animations. *ACM Trans. Graph. (Proc. of SIGGRAPH)* 36, 4 (2017), 155.