

图像混合的研究与实现

姓名 方阳 学号 2018201544 学院 电子信息学院

摘要

这篇文章将编写一个图像混合算法，并使用它来创建混合图像，该技术由奥利瓦、托拉尔巴和希恩斯于 2006 年发明，并在 SIGGRAPH 上发表了一篇论文。高频图像内容往往主导感知，但在一定距离内，只能感知低频内容。通过混合高频率和低频率的内容，我们可以创建一个混合图像，在不同的距离感知不同。

正文

一、引言

1.1 研究背景与意义

图像混合是基于人类处理视觉输入的方式，根据观看距离以两种不同方式之一感知的图像。艺术家们已经有效地使用了低空间频率的操作来引出一种感知，这种感知在依赖于周边视觉时发生了变化。受到这项工作的启发，Setlur 和 Gooch^[1]提出了一种技术，可以在不同的空间频率上创建具有冲突情绪状态的面部图像。这些图像会随着视线的变化产生微妙的表情变化。麻省理工学院的 Aude Oliva 和格拉斯哥大学的 Philippe G. Schyns 开发了一种技术^[2]，用于创建显示这种光学错觉的混合图像，这种方法最初是由希恩斯和奥利瓦在 1994 年提出的。在本文实现的算法中，演示了混合图像在创建具有两种不同可能解释的图像时的有效性。通过将两个图像叠加在两个来生成混合图像不同的空间尺度：通过滤波获得低空间尺度一个带低通滤波器的图像；获得了高空间尺度通过使用高通滤波器过滤第二个图像。最终的形象是通过添加这两个过滤图像组成的。混合图像将一幅图像的低空间频率与另一幅图像的高空间频率相结合，生成的图像具有随观看距离变化的解释。具体来说，人们可以在短距离看到高通滤波(锐化)图像，而在长距离看到低通滤波(平滑)图像。也许最常见的例子是爱因斯坦和玛丽莲梦露。从较近的距离看这张照片，你可以看到爱因斯坦的清晰图像，只有模糊失真的迹象，暗示着叠加图像的存在。从远处看，细节模糊了，清晰可见的玛丽莲梦露出现了。

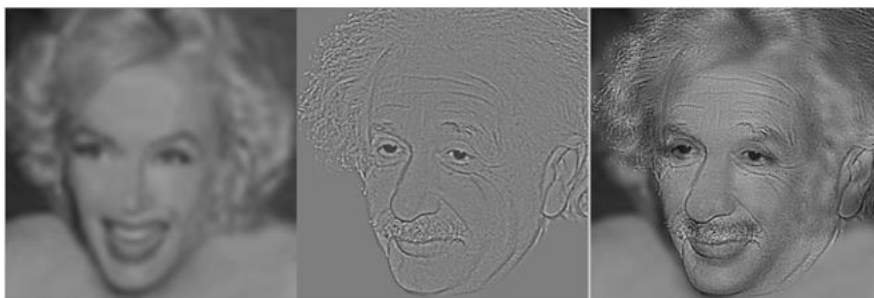


图 1-1 由玛丽莲梦露照片的低频分量和爱因斯坦照片的高频分量组成的混合图像

图像混合这一技术的应用挺多的，比如下图 1-2 所示的私人字体，我们可以使用混合图像来显示文本，这对于站在离屏幕一定距离的人来说是不可见的。保护用户隐私的商业产品通常依赖于头戴式显示器或偏振屏幕，其可见度随视角而降低。混合字体由两部分组成：高空间频率(将包含文本)和低空间频率(将包含掩蔽图像)。

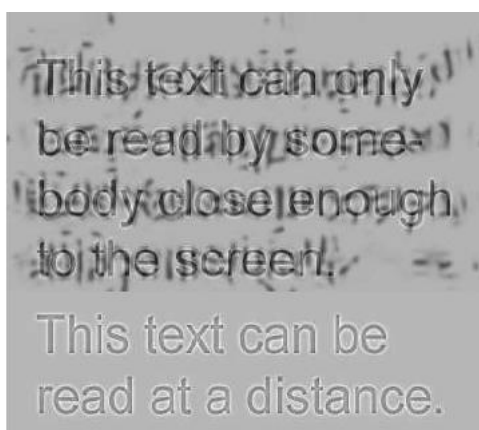


图 1-2 混合字体在几米内就看不见了，在相对较远的距离下，底部的文本仍然易于阅读

还有混合纹理，我们可以创建随着观看距离消失的纹理。图 1-3 显示了这个想法的一个例子。这张图展示了一个女人的脸，当她近距离看的时候变成了一只猫。注意，这种效果不能通过叠加妇女的脸和猫的脸使用透明。使用透明度(加法叠加)创建的人脸不会随着距离而改变。



图 1-3 当从几米远的地方看到这张照片时，与猫脸相对应的纹理就消失了

还有属于我们人脸面部的图像混合，当我们改变观看距离时，混合图像对于创建改变表情、身份或姿势的人脸图像特别有用。还可以混合同一场景的不同时间状态，比如通过结合两幅在两个不同时刻拍摄的照片来展示房子的两种状态。

1.2 研究现状

上述的混合方法目前用得差不多了，使用截断频率的做法来混合图像还是过于粗糙，生成的图像质量不佳，目前的图像混合技术已经向卷积神经网络（CNN）靠拢，相比手动去筛选特征，通过 CNN 网络自身学习合适的特征进行混合要更加好，得到的图像质量也更好。Gatys 等人^[3]的近期研究证实了 CNN 的力量：通过分离和重新组合图片内容与风格，CNN 可以创作出具有艺术魅力的作品。使用 CNN 将一张图片的语义内容与不同风格融合起来的过程被称为神经风格迁移（Neural Style Transfer）。从那以后，在学术研究和产业应用中，神经风格迁移已成为一个很受欢迎的主题，下图就是风格迁移的一个事例。



图 1-4 从左至右分别是内容图像，风格图像和生成图像

二、理论基础

视觉心理物理学研究表明，人类观察者能够在短短的一瞥之内理解一幅新图像的含义。在观看动作电影或音乐视频中的快速场景编辑时，可以体验到这种快速图像理解的非凡性能。对人类感知的研究表明，图像理解效率是基于对视觉输入的多尺度、全局到局部的分析，全球结构和组成部分之间空间关系的初步分析指导了对局部细节的分析。图像分析的全球优先假设暗示了一个图像的从粗到细的频率分析，其中由快速大细胞途径对比和携带的低空间频率成分主导早期视觉处理。

施因斯和奥利瓦 1994 年使用混合刺激测试了空间频带对自然图像解释的作用。当任务需要快速识别场景图像时，人类观察者在高空间频带(从 24 个周期/图像)之前解释低空间频带(在 8 个周期/图像的频率截止处):当仅显示 30 毫秒的混合图像时，观察者识别低空间尺度时，而在 150 毫秒的持续时间内，他们首先识别高空间尺度。有趣的是，参与者没有意识到视觉刺激有两种解释。同一个图像的性别，他们使用混合的低空间频率分量和高频率分量一样频繁。同样，参与者没有报告注意到这些图像中存在两种情绪或两种性别。这些结果表明频率的选择用于快速图像识别的频带是一种灵活的机制:图像分析仍然可以根据低到高的空

间尺度处理展开,但是人类观察者能够快速选择低频或高频频带,该频带传达了解决给定任务和解释图像的最多信息。重要的是,当选择空间频率时,观察者没有意识到其他空间尺度的信息。

在人类感知的研究中,混合图像允许表征不同频率通道在图像识别中的作用,并评估空间频率处理的时间过程。混合图像提供了一种新的模式,在这种模式下,可以通过播放观看距离或演示时间来调整图像解释。对于给定的观看距离或给定的时间频率,空间频率的特定频带支配视觉处理。混合图像的视觉分析仍然从整体感知展开到局部感知,但是在选定的频带内,对于给定的观看距离,观察者将首先感知混合图像的全局结构,并花费另外一百毫秒来将局部信息组织成一致的感知。

在成功的混合图像中,当一个感知占主导地位时,有意识地切换到另一种解释几乎是不可能的。只有当观看距离改变时,我们才能转换到另一种解释。在混合图像中,重要的是替代图像被感知为噪声(缺乏内部组织),或者与主导子带混合。感知分组规则调节混合图像的有效性。低空间频率(斑点)缺乏物体形状和区域边界的精确定义,这需要视觉系统将斑点组合在一起,以形成粗略尺度的有意义的解释。当观察者看到模棱两可的形式时,他们用最简单的方式解释元素。观察者更喜欢元素较少而不是较多的排列,对称而不是不对称的组成,并且通常尊重其他格式塔感知规则。

低空间频率模式的对称性和重复性不好:它们形成一种强烈的感觉,很难在感觉上消除。如果高空间频率的图像缺少相同的强分组提示,则即使从短距离观看,对应于低空间频率的图像解释也将总是可用的。通过引入偶然对齐,可以减少一个空间通道对另一个空间通道的影响。

一个有趣的观察是,当图像被恰当地构建时,观察者似乎将被掩蔽的图像感知为噪声。混合图像打破了真实世界自然图像的一个重要统计特性(图 2-1),即连续空间尺度下通带滤波器输出之间的相关性。图 2-1a 示出了自然图像的拉普拉斯金字塔的不同级别之间获得的互相关矩阵。在一个刻度上找到的边与上下刻度上找到的边相关联。当两幅图像叠加时,也可以获得同样的效果(附加透明度)。在这种情况下,没有简单的滤波器来分离两个图像(并且两个图像的感知被混合,而与我们观察图像的距离无关)。图 2-1c 示出了当图像被模糊,然后被附加白噪声破坏时获得的相关矩阵。相关矩阵揭示了哪些尺度受噪声支配,因为它们不具有我们从自然图像中预期的交叉尺度相关性。在混合图像的情况下,相关矩阵(图 2-1d)揭示了两组的存在。

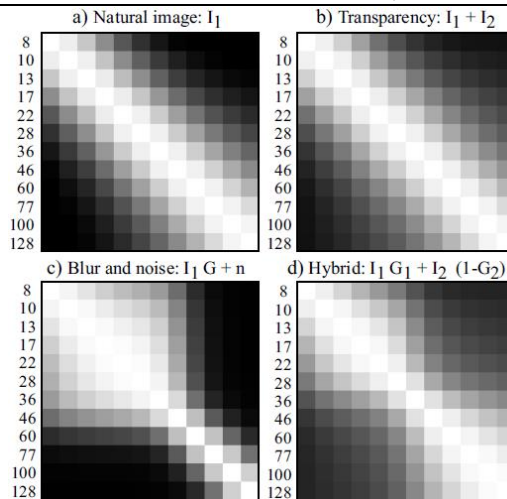


图 2-1 图像的拉普拉斯金字塔各层之间的相关性

三、本文方法

输入为两幅不同的图像，分别截取低频分量和高频分量，然后进行混合。以下是算法流程。

算法流程：

1. 首先读取两种需要混合的图像 `img1` 和 `img2`。
2. 然后设立截断频率的参数，通过 `fspecial` 函数得到相应的低通滤波器。
3. 然后是滤波函数，首先获得滤波器和图像的尺寸，根据滤波器的尺寸对图像进行相应的填充，填充尺寸为 $(\text{filter_size}-1)/2$ ，然后是卷积部分，由于实现二维卷积不简单，我决定用两次一维卷积来处理它。对每个通道，先与垂直的一维卷积核进行卷积，再与水平的一维卷积核进行卷积。在 `matlab` 中，我们可以使用有用的内建函数 `im2col` 和 `col2im` 对矩阵进行卷积。最后,输出滤波后的图像。
4. 得到一个图像的低频分量，然后求取第二个图像的低频分量，由于需要的是第二个图像的高频分量，所以这里用原图直接减去低频分量，这样就得到了第二个图像的高频分量。
5. 最后将第一个图像的低频分量与第二个图像的高频分量进行相加，得到混合图像。
6. 然后将两幅图像和混合图像显示出来，其中混合图像的显示是通过逐步对图像进行下采样，并将所有图像连接在一起。

四、实验与讨论分析

4.1 实验效果

Bird-Plane pair

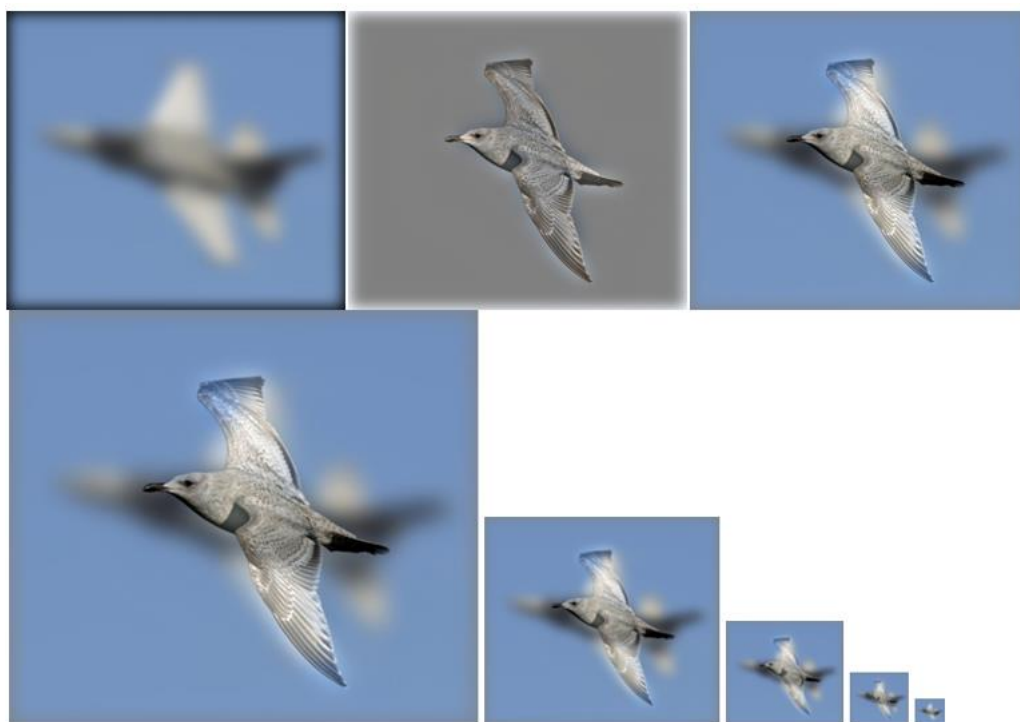


图 4-1 飞机与鸟

Motorcycle-bicycle pair



图 4-2 自行车与摩托车

Fish-Submarine pair

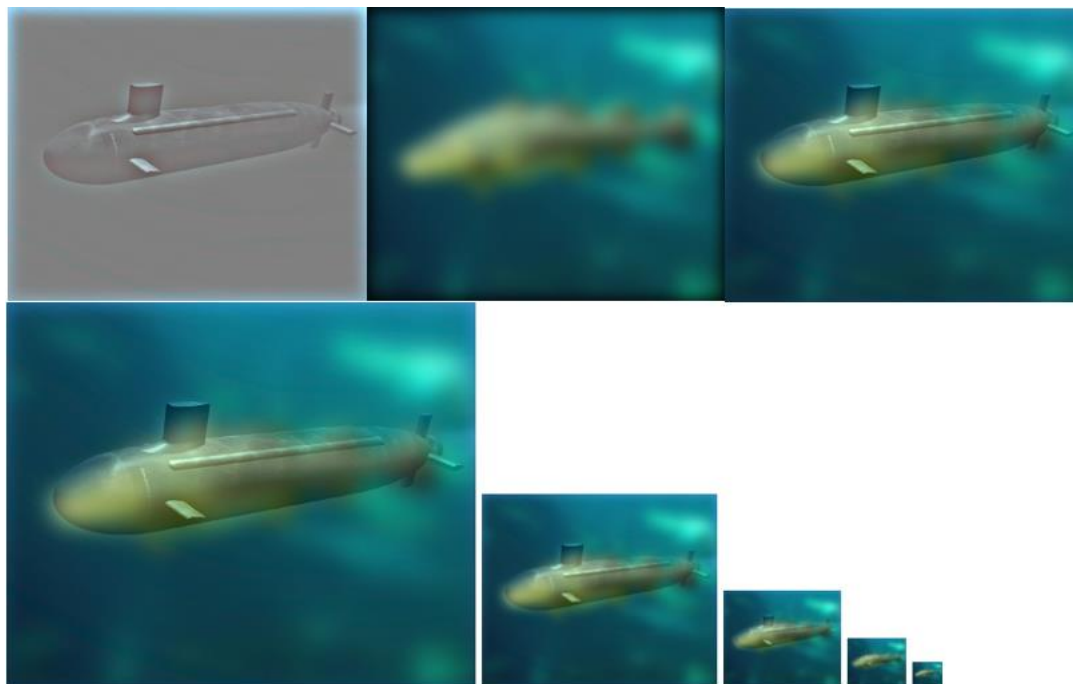


图 4-3 潜艇与鱼

4.2 实验分析

实验效果与预期一致，在近距离观看混合图像得时候，将是高频信息的图像，而在远距离观看时则是低频信息的图像。

另外，对于截断频率的选择需要主要，不能过大过小，如下图 4-4 所示，过大则低频信息将丢失掉了，过小则高频信息丢失掉了。

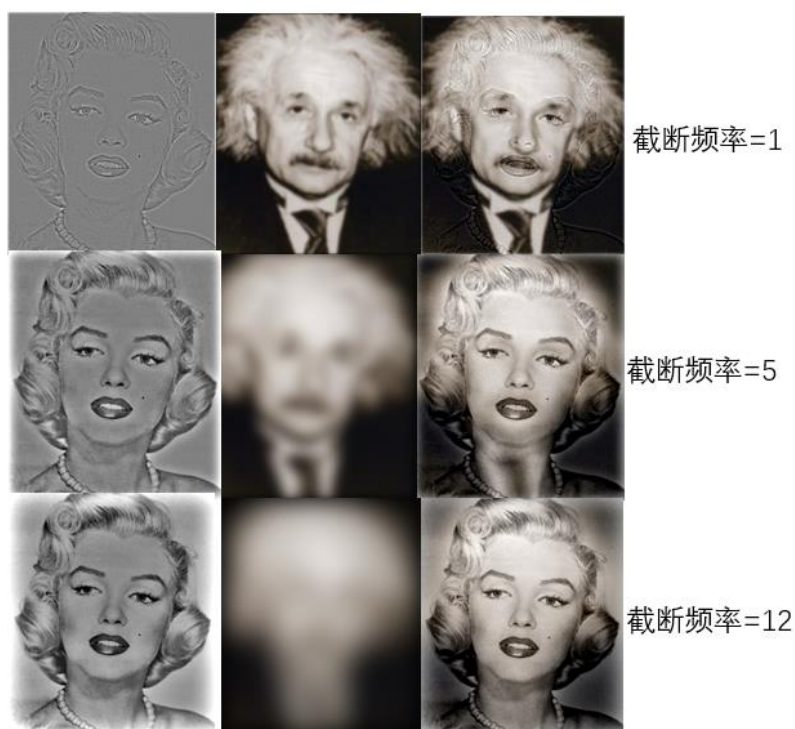


图 4-4 不同截断频率的合成图像

结束语

当看到这个效果图的时候，就感觉图像混合挺有趣，近看高频远看低频，人类的视觉感知拥有这种特性，着实很不可思议，视觉的奥秘依然很多，继续探究吧！

参考文献

- [1] Setlur V , Gooch B . Is that a smile? Gaze dependent facial expressions[C]// International Symposium on Non-photorealistic Animation & Rendering. DBLP, 2004.
- [2] Oliva A, Torralba A, Schyns P G. Hybrid images[J]. Acm Transactions on Graphics, 2006, 25(3):527-532.
- [3] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[J]. arXiv preprint arXiv:1508.06576, 2015.

附录

- 1. 字数统计 —— 3403
- 2. 建议与意见 —— 无