

混合图像



图1：混合图像是将一幅图像的低空间频率与另一幅图像的高空间频率组合在一起的图像，产生的图像的解释随观看距离而变化。在此图中，人们可能看起来很悲伤，近距离，但后退了几米并再次看了看表情。

摘要：我们提出了混合图像，一种产生具有两种解释的静态图像的技术，这些解释会随着观看距离的变化而变化。混合图像基于人类视觉系统对图像的多尺度处理，并且受到视觉感知中的蒙版研究的启发。这些图像可用于创建引人注目的显示，其中图像随着观看距离的变化而变化。我们表明，通过考虑感知分组机制，可以在每个距离上建立具有稳定感知的引人注目的混合图像。我们展示的示例中，混合图像用于创建仅在近距离观看时才可见的纹理，生成其表情随观看距离而变化的面部表情，并在单个图片中可视化随时间的变化。

关键词：混合图像，人类感知，尺度空间

1. 介绍

在这里，我们利用人类视觉的多尺度感知机制来创建视觉错觉（混合图像），通过改变观看距离或演示时间可以感知图片的两种不同解释。我们使用并扩展了Schyns和Oliva [1994; 1997; 1999]。图中显示了一个混合图像的示例，该图像是由两个图像组合而成的，其中面部显示出不同的情感。高空间频率对应于带有“悲伤”表情面部。低空间频率对应于具有“快乐”和“惊奇”情绪的相同面孔（即，情绪从左到右分别是：快乐，惊讶，快乐和快乐）。要从一种解释切换到另一种解释，可以距离图片几米进行观察。

艺术家已经有效地利用了低空间频率操纵来产生一种依赖周围视觉而改变的感觉（例如，[Livingstone 2000; Dali 1996]）。受这项工作的启发，Setlur和Gooch [2004]提出了一种技术，可以在不同的空间频率下创建具有冲突情绪状态的面部图像。图像随着注视变化产生微妙的表情变化。在本文中，我们展示了混合图像在创建具有两种截然不同的解释的图像中的有效性。

混合图像是通过在两个不同的空间比例上叠加两个图像而生成的：低空间比例是通过使用低通滤波器对一个图像进行滤波而获得的；高空间比例是通过使用高通滤波器对第二个图像进行滤波而获得的。通过将这两个过滤后的图像相加组成最终图像。请注意，混合图像是与图片马赛克不同的技术[Silvers 1997]。图片镶嵌有两种解释：一种是局部的（由组成该镶嵌的每张图片的内容给出），另一种是整体的（最好在预定义的距离处看到）。但是，混合图像包含两种相干的全局图像解释，其中一种是低空间频率，另一种是高空间频率。

我们用几个概念验证的例子来说明这种技术。我们展示了如何将该技术应用于创建随观看距离改变表情面部图片，在单个图片中显示场景的两种配置以及呈现远距离观看时消失的纹理。

2. 混合图像的设计

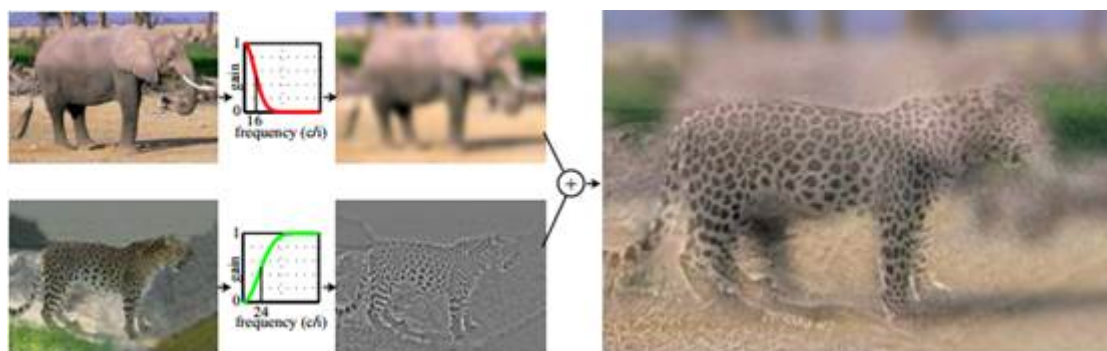


图2：通过在两个不同的空间比例上叠加两个图像来生成混合图像：低空间比例是通过使用低通滤波器对一个图像进行滤波而获得的，高空间比例是通过对具有高空间分辨率的第二幅图像进行滤波而获得的通滤波器。最终的混合图像是通过将这两个过滤后的图像相加而构成的。

通过组合两个图像 (I_1 和 I_2) 获得混合图像 (H)，一个图像用低通滤波器 (G_1) 滤波，第二个图像用高通滤波器 ($1 - G_2$) 滤波： $H = I_1 \cdot G_1 + I_2 \cdot (1 - G_2)$ ，操作在傅立叶域中定义。混合图像由两个参数定义：低分辨率图像的频率切割（在远距离处可以看到）和高分辨率图像的频率切割（可以近距离观察）。可以通过为每个频道引入不同的增益来添加其他参数。对于本文显示的混合动力，我们将两个空间通道的增益都设置为1。对于低通和高通滤波器，我们使用高斯滤波器 (G_1 和 G_2)。我们将每个滤波器的截止频率定义为滤波器幅度增益为1/2的频率。

图2说明了用于创建一个混合图像的过程。可以根据图像大小和滤波器的截止频率（以周期/图像1表示）来完全确定最佳地看到混合图像的每个分量的距离和混合感知交替的距离。在查看本文中的图像时，请通过距离图片几米的距离在各种解释之间切换。请注意，显示图像的尺寸越大，您就必须走的越远，才能看到替代图像的解释。

2.1 混合图像的感知



图3：边缘和斑点之间的感知分组。当近距离观看时，这三幅图像被视为老虎，而远处则被视为猎豹。这三个图像之间的差异是边缘和斑点之间的对齐程度。图像a) 包含两个没有对齐的重叠图像。在图像b) 中，眼睛对齐。并且在图像c) 中，头部姿势与眼睛和嘴巴的位置对齐。在适当对准的情况下，残余频带无法建立感知。当近距离观看时，很难看到猎豹的脸，该脸被老虎的脸完全遮盖了。从远处看，老虎的边缘与猎豹的脸融为一体。

在下一节中，我们将介绍混合图像背后的动机，因为它们与人类感知研究有关。我们将提供一个框架来理解涉及双重图像感知的机制。

视觉心理物理学研究表明，人类观察者能够在短时间内（100毫秒[Potter 1975]）理解新颖图像的含义。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惊人表现。对人类感知的研究表明，图像的理解效率基于对视觉输入的多尺度，全局到局部分析[Burt and Adelson 1983; Majaj等。 2002]：对全局结构和组件之间的空间关系的初步分析指导对局部细节的分析[Schyns and Oliva 1994;瓦特 (1987)]。图像分析的全局优先假设（“在树木前看森林”，[Navon 1977]）意味着对图像进行从粗到细的频率分析，其中低空间频率分量被快速的对比和携带。大细胞通路，主导早期视觉处理[Hughes等。 1996;林德伯格1993; Parker等。 1992年； Schyns和Oliva, 1994年； Sugase等。 1999]。

Schyns和Oliva [1994]使用混合刺激，测试了空间频带在解释自然图像中的作用。当任务需要快速识别场景图像时，人类观察者会在高空间频段（从24个周期/图像）之前解释低空间频段（以8个周期/图像的频率截止）：当显示混合图像仅30毫秒，观察者识别出较低的空间比例（例如，当他们收到图3的图像时会回答“猎豹”），而在150毫秒的持续时间内，他们首先识别出了较高的空间比例（例如，图3中的老虎）。有趣的是，参与者没有意识到视觉刺激有两种解释。其他实验表明，优先选择的用于解释图像的空间频带取决于观看者必须解决的任务。Schyns和Oliva [1999]使用类似于图5.b中的混合脸部，表明当要求参与者确定仅显示50毫秒（开心，生气或中立）的混合脸部图像时的情绪时，他们选择了低空间频率的脸部（图5.b中的角度），但是当他们必须确定同一张图片的性别时，他们使用混合频率的低空间频率分量和高频率一样。再次，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，选择用于快速图像识别的频率带是一种灵活的机制：图像分析可能仍会根据从低到高的空间比例处理而展开，但是人类观察者能够快速选择低频段或低频段。高，传达了最多的信息来解决给定的任务和解释图像。重要的是，当选择一个空间频率时，观察者并没有意识到其他空间尺度上的信息。



图4：高空间频率下的颜色用于增强自行车的近距离感。从远处看，一辆摩托车。摩托车的形状被解释为接近阴影。

在人类感知的研究中，混合图像可以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程。混合图像提供了一种新的范例，其中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，特定的空间频率带在视觉处理中占主导地位。混合图像的视觉分析仍然从全局感知到局部感知，但在选定的频段内，对于给定的观看距离，观察者将首先感知到混合图像的全局结构（图3中的图像代表头部）。，并花费额外的一百毫秒的时间将本地信息组织成一个连贯的感知（如果在很远的距离观看图像，则会组织斑点，而对于近距离观看则会组织边缘）。

2.2 感知分组和混合图像的规则

从理论上讲，可以将任意两个图像组合在一起以生成混合图像。在实践中，美观的混合图像需要遵循我们在本节中描述的一些规则。在成功的混合图像中，当一个感觉占优势时，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，我们才能切换到其他解释。在混合图像中，将替代图像感知为噪声（缺少内部组织）或将其与主要子带混合非常重要。

感知分组的规则调节混合图像的有效性。低空间频率（blob）缺乏对对象形状和区域边界的精确定义，这要求视觉系统将这些blob组合在一起以形成对粗尺度的有意义的解释。当观察者以歧义形式出现时，他们以最简单的方式解释元素。观察者更喜欢这样一种安排：元素要少而不是多，组成要对称而不是不对称，并且通常要遵守其他格式塔的感知规则。

在低间频率中，图案的对称性和重复性很差：它们形成了强烈的感觉，即很难从视觉上消除。如果高空间频率的图像缺少相同的强分组提示，则即使从近距离观看时，对应于低空间频率的图像解释也将始终可用。通过引入意外对准，可以减少一个空间通道对另一空间通道的影响。例如，在图2中，大象的顶部（低空间频率）与视界线（低空间频率和高空间频率）对齐。因此，当近距离观看图像时，大象

的顶部边缘可以用一些细边缘来解释。这降低了大象的显着性。图3示出了在低空间频率和高空间频率之间具有不同程度一致性的混合图像的几个示例。

颜色提供了非常强的分组提示，可用于创建更引人注目的幻觉。例如，在图4中，仅在高空间频率中使用颜色以增强自行车并加强近距离观看图像时摩托车作为阴影的解释。



图5：一个生气的男人还是一个体贴的女人？两种混合图像都是通过组合愤怒的男人（低空间频率）和严厉的女人（高空间频率）的面孔而产生的。您可以通过从几米处观看图片来切换感知。 a) 混合图像不良。由于滤镜重叠，图像从上至下看起来模棱两可。 b) 良好的混合图像。

正确选择滤波器截止频率的重要性如图5所示。在图5.a中，两个滤波器都有很强的重叠，因此，两个面之间没有清晰的过渡。对于图5.b上的混合图像，两个滤镜几乎没有重叠。结果是产生了更清晰的图像，产生了明确的解释（看起来像是近距离的女人，而像远方的男人）。当图像未完全对齐时，这一点尤其重要。

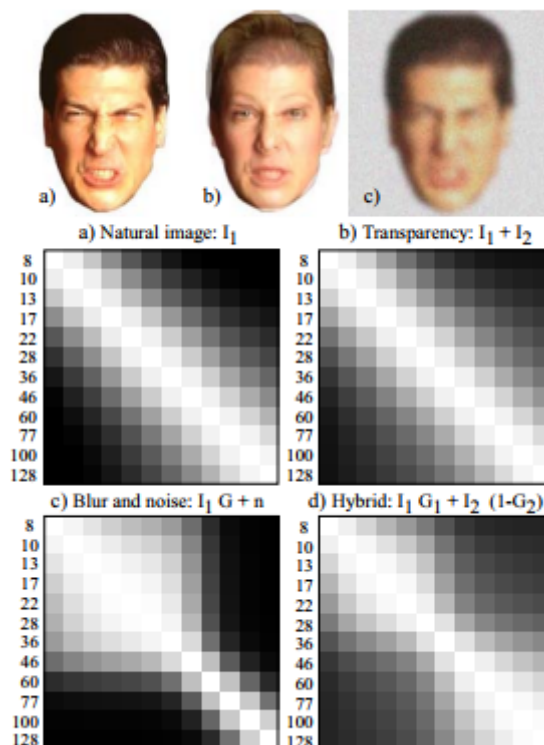


图6：经过多次处理后，图像在拉普拉斯金字塔各层之间的相关性。 a) 自然图像， b) 添加了两个图像， c) 具有加性白噪声的模糊图像，以及 d) 混合图像（ $f_1 = 16$ 个循环/图像， $f_2 = 48$ 个循环/图像）。

一个有趣的发现是，当正确构造图像时，观察者似乎感觉到被掩盖的图像是噪声。混合图像打破了现实世界中自然图像的一个重要统计属性（图6），即连续空间尺度上的通带滤波器输出之间的相关性。图6.a显示了自然图像在拉普拉斯金字塔的不同层之间获得的互相关矩阵。在一个比例尺上找到的边缘与在上下比例尺上找到的边缘相关。当两个图像叠加在一起时，会获得相同的效果（累加透明度）。在这种情况下，没有一个简单的过滤器来分离两个图像（并且两个图像的感知是混合的，与我们观察图像的距离无关）。图6.c显示了当图像模糊（截止频率为 $16c/i$ ）然后被加性白噪声破坏时获得的相关矩阵。相关矩阵可以揭示哪些标度受噪声控制，因为它们不具有我们希望从自然图像中获得的跨标度相关性。在混合图像的情况下，相关矩阵（图6.d）揭示了两组的存在。

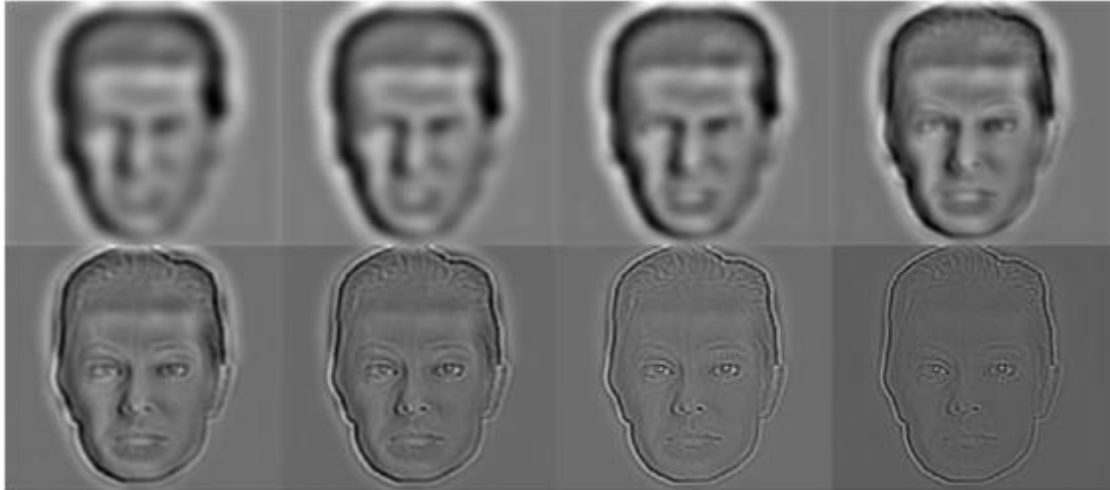


图7：拉普拉斯金字塔的输出，揭示了图5.b混合图像的组成部分。

图7显示了应用于图5.b混合图像的拉普拉斯金字塔的输出。低频频道和高频频道会看到不同的图像。注意，每个子带本身也是混合图像。如果您离开页面，您会看到子带——对应于小比例尺。在阅读距离处，第一行的四个图像被解释为一个愤怒的人；底部是一个严厉的女人。当您从图像退后一步时，您会看到愤怒的男人的脸开始出现在更多的子波段中。每个子带的比例越小，您必须走的越远才能看到图像的切换。

总之，可以利用两种主要机制来创建引人注目的混合图像。首先是使两个比例尺中的边之间的相关性最大化，以便它们融合。第二个事实是，与跨标尺的其他边缘不相关的其余边缘可以被视为噪声。在图5.b中就是这种情况，在这种情况下，跨比例尺的边缘融合非常引人注目，但是当近距离观看图像时，似乎存在一些低空间频率噪声。

2.3 尺度空间容量

到目前为止，已经通过混合两个图像获得了混合图像，但是是否可以合并两个以上图像，并且仍然具有随着我们改变观看距离而转变的连贯感知？在关于文本掩蔽的研究中，Majaj等人。[2002]创建了一个刺激，该刺激叠加了4个字母，每个字母包含不同空间范围的能量。当观察者远离刺激时，他们报告图像从一个字母转换为另一个字母。结果很有趣，但是在多个比例尺之间缺乏良好的分组提示会导致图像看起来失真。同样，在任何给定时间都可以看到多个字母。多个图像的叠加仍然是一个未解决的问题。

3. 应用

在本节中，我们讨论一些应用程序（有关其他示例，请参见对该文件进行补充的视频）。

专用字体：我们可以使用混合图像显示离屏幕一定距离的人看不见的文本。用于用户隐私的商业产品通常依赖于头戴式显示器或偏振屏，其可见度随视角而降低。混合字体包含两个组件：高空间频率（将包含文本）和低空间频率（将包含掩盖图像）。

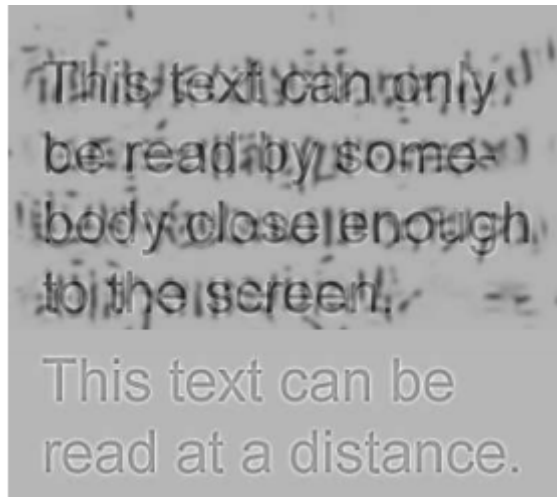


图8：混合字体在几米处变得不可见。底部文本在相对长的距离处仍易于阅读。

对于高通滤波器，我们使用调整宽度 (σ) 的高斯滤波器，使 $\sigma < np$ ，其中 np 是字母笔画的粗细（以像素为单位）。低频通道（掩蔽信号）包含类似文本的纹理[Portilla and Simoncelli 2000]。Solomon和Pelli [1994]已显示，每个字母3个周期的频带中的噪声会更有效地屏蔽字母。因此，我们将低通滤波器的截止频率调整为 $3 * n$ ，其中 n 为文本行中的字母数。目的是减少近距离观看时噪声对文本的干扰，同时从远处看时具有有效的掩盖噪声。在图8所示的示例中，只能从一米以下的距离读取文本。在大约两米的距离内，文本不可读。在产生这种效果时，低空间频率的屏蔽非常重要（图8）。底部的文本仅经过高通滤波，在低空间频率下没有遮罩，因此在相对较长的距离下仍易于阅读。



图9：右) 猫女：从几米远的地方观看图像时，与猫脸相对应的纹理消失了。左) 正在建造的房屋。在短距离查看图像时，可以看到房屋正在建造中，但是如果您离开图像，则将看到房屋的最终状态。

混合纹理：我们可以创建随观看距离消失的纹理。图9显示了这个想法的一个示例。该图显示了一个女人的脸，当靠近她的脸时变成猫的例子。请注意，通过使用透明度叠加女人的脸和猫的脸无法获得这种效果。使用透明度（加法叠加）可创建不会随距离变化的面。

改变面孔：混合图像在创建面孔图像时特别强大，这些面孔会随着我们改变观看距离而改变表情，身份或姿势。图1显示了一个令人信服的面部表情变化示例。多个比例尺的边缘融合在一起，产生的图像在所有距离上看起来都很自然。对于人脸图像，为了创建看起来没有变化的图片，人脸特征之间的正确对齐很重要。在未对准的情况下，最好是将低空间频率的脸部变形（仿射扭曲）。

时间变化：图9显示了一个示例，该示例使用混合图像通过组合在两个不同瞬间拍摄的两个图片来显示房屋的两种状态。

4. 总结

我们已经描述了混合图像技术，该技术允许创建具有两种解释的图像，这些解释会随着观看距离而变化。尽管该技术很简单，但这些图像仍对幼稚的观察者产生了非常引人注目的惊喜效果。它们还提供了一种有趣的新可视化工具，可将两个互补图像变形为一个。创建引人注目的混合图像是一个开放且具有挑战性的问题，因为它依赖于在不同空间尺度上相互作用的感知分组机制。