

基于感知模型的水印探究

阳小丽 (湖南工程职业技术学院 长沙 410151)

摘要:文章分析了感知模型一般形式上要解决的问题,并结合人感觉系统的特点,提出了一种可测量嵌入水印感知度的模型,从感知的角度提出了一种检测水印的方法,同时这种模型也增强了水印被载体作品隐藏的效果。

关键词:数字水印 感知模型 灵敏度 掩蔽 合并

1 引言

数字水印作为解决数字产品版权问题的有效手段,近年来受到了人们的广泛关注,并逐渐成为了当今学术界的研究热点。一个有效的数字水印技术必须具有三个典型的特点:透明性、鲁棒性和安全性。透明性就是指水印应该是不可感知的,那么怎样去衡量水印的可感知性?怎样能够在—幅作品中嵌入水印而不被察觉和感知?文章将通过对感知模型的描述来回答这两个问题。

2 感知模型的一般形式

人类的感知机制不能象均方误差评价方法一样对作品的所有变化一视同仁,人类听觉系统(HAS)会根据输入的频率和响度的不同而做出不同的反应,人类的视觉系统的反应也会因其输入的空间频率、亮度以及颜色的不同而不同,因此水印所有成分的感知程度不是完全一致的。

感知上的变化是可以测量的,且可以针对这些变化构建模型。感知模型一般都试图解决三种基本现象:灵敏度、掩蔽及合并。

灵敏度指的是眼睛或耳朵对直接激励的反应。通过测量灵敏度的实验得知,尽管眼睛或耳朵会对一个信号的很多不同方面敏感,但最基本的激励或者说被测量的特性经常是频率和响度(或亮度),颜色和方向在图像和视频数据中也非常重要。

2.1 频率灵敏度

人类听觉系统和视觉系统对于输入信号所做出的反应都是与频率有关的。对于听觉,频率变化被感知为不同的音调,实验表明人耳对 3KHZ 左右的频率最

为敏感,对于过低(20HZ)或者过高(20KHZ)的频率,敏感性就会降低。

对于视觉,有三种不同形式的频率响应:空间频率、光谱频率和时间频率。

空间频率响应通常是指对于作为空间频率函数的亮度对比的敏感度,对比敏感度函数表明,我们最为敏感的是中等频率上的亮度差异,在较高和较低频率时,敏感度会下降,这与人类听觉系统的情况是类似的。

光谱频率是以颜色的形式被感知的。最底层的颜色视觉由三个独立的颜色系统构成。实验表明,低频响应(蓝色信道)要比其他两种信道低很多,正因如此,很多彩色水印系统都将 RGB 图像的大部分水印信号分配在蓝色信道。

时间频率通常以运动或者跳跃的形式被感知。人眼对不同时间频率反应的实验测量结果表明,当频率高于 30KHZ 时,敏感度就会迅速减小,这就是电视和电影的传输速度都不会超过 60 fps(帧每秒)的原因。

2.2 响度/亮度敏感度

在很多关于人类听觉的研究中已经测定过作为强度函数的最小可察觉的强度变化。尽管实验的方法手段有所不同,但都得出同一个一般性结果:平均强度越大,所能识别的变化就越小。人眼的情况刚好相反,对于越亮的信号眼睛的敏感度越低。

2.3 掩蔽

环境背景会对感知过程造成影响。我们能听出具有某个特定声音强度的独立音调,但是如果在这个频率附近存在另外一个更响的音调,前者可能就会完全听不见了。同样地,一个孤立情况下很容易被辨识的纹理,当被附加到另一个具有更丰富纹理的图像上时,其识别难度也可能会增大,这就是说,一个信号的存在

能够隐藏或者掩蔽另一个信号,掩蔽就是在另一个“掩蔽”信号同时存在的情况下,观察者对某一激励所做响应的量度。

2.4 合并

敏感度模型和掩蔽模型可以用于估计某一特征变化的可感度。然而如果发生变化的不止是单一频率而是多个频率同时变化,就必须知道应如何对各个频率的敏感度和掩蔽信息进行综合。在感知距离模型中,对多个不同失真可感知性的综合往往归结为对作品整体变化的评价,这就是合并。合并通常可以用下面的等式来表示:

$$D(C_o, C_w) = (\sum |d[i]|^p)^{\frac{1}{p}} \quad (1.1)$$

式中 $d[i]$ 表示观察者觉察到 C_o 和 C_w 之间某个参数变化的可能性,这些参数可以是时间抽样、空间像素或者傅里叶频率参数。式(1.1)通常称为明科斯基或者 L_p 范数。在音频情况下,可以取 $P=1$ 得到一个比较合适的线性和,而对于图像而言 P 更多地被取值为 4。

3 一个音频感知模型

这里所讲的音频模型在参考文献[1,93]被用于加入水印,它提供了给定的音频信号听力阈值的度量,该模型对每一频率下被感知的最小声音能量进行估计,根据这些估计值可以测出一个给定频率的特定变化能否被听出来。

为了得到听力阈值,首先需要将信号划分成相互重叠的帧或者窗口,这通常可以汉明窗滤波得到。在傅里叶变换域中对各个音频帧 $C[t]$ 独立地进行分析,并使用敏感度和频率掩蔽函数对识别某一频率为 f 的音调所需的最小声音能量进行估计,得到估计值 $t[f]$ 。

3.1 敏感度

声压水平(SPL)表示的是相对于 $20\mu Pa$ 的参照强度而言的声音强度,单位为分贝(db):

$$SPL = 20 \log_{10} \frac{p}{20} \quad (2.1)$$

式(2.1)中 p 表示激励的声压,单位为 Pa(帕斯卡)。绝对听力阈值被定义为无噪声环境下一个纯音调被人类检测到所需要的最小强度,用 db SPL 表示。绝对听力阈值可以用如下函数表示:

$$T_q(f) = 3.64 \left(\frac{f}{1000}\right)^{-0.8} - 6.5 \exp^{-0.6(f/1000 - 3.3)^2} +$$

$$10^{-3} \left(\frac{f}{1000}\right)^4 \quad (2.2)$$

阈值 $T_q(f)$ 近似等于安静环境下可识别的最小变化值。

3.2 掩蔽

人耳可以建模为一系列重叠的带通滤波器,这些滤波器的标志特性为其临界带宽,可以具体描述为:考察一个特定响度的窄带噪声源,当其带宽增大时,我们感到响度一直保持不变,直到超过临界带宽时,我们感到响度增加了。临界带宽随频率变化而变化。通常将人耳看做一系列离散的临界带宽滤波器组成,临界带宽用 z 来表示,这里 $1 \leq z \leq Z_1$,其中 Z_1 为临界带宽的总数。

每个临界频带的阈值是该频带中视在能量的函数,同时还与该帧听起来是像噪声还是像音调有关,而各频带的视在能量又取决于相邻频带的真实能量。因此,得到各个频率的掩蔽阈值的具体步骤为:

(1) 确定各个临界频带的能量。

(2) 确定由各个临界频带之间的能量传播引起的每个临界频带的视在能量。

(3) 确定当前音频帧是类噪声的还是类音调的。

(4) 确定掩蔽阈值。

每个临界频带中的能量 $B[z]$ 为:

$$B[z] = \sum_{f=L[z]}^{H[z]} C[f] | \quad (2.3)$$

式(2.3)中 $L[z]$ 和 $H[z]$ 分别表示频带宽度 z 中频带的上限和下限,如表 1.1 所示, $|C[f]|$ 表示在谱 C 中的频率 f 对应的幅值。

临界频带间的能量传播可以通过基膜传播函数 $SF(z)$ 计算得到:

$$SF(z) = 15.81 + 7.5(z + 0.474) - 17.5 \sqrt{1 + (z + 0.474)^2} \quad (2.4)$$

能量传播之后得到的各频带的视在能量 $Ba[z]$ 为:

$$Ba[z] = \sum_{z'=z-z_0}^{z+z_1} B[z-z'] SF(z') \quad (2.5)$$

式(2.5)中 z_0 和 z_1 表示能量传播显著的领域范围。

掩蔽不仅取决于临界频带中视在能量的多少,还与能量是以类噪声还是类音调为主有关。为了确定音频帧类噪声或类音调的程度,下面计算频谱平坦性量

度 SFM_{db} :

$$SFM_{db=10\log_{10}} \left[\frac{\left[\sum_{z=1}^Z B_o[z] \right]^+}{\frac{1}{Z} \sum_{z=1}^Z B_o[z]} \right] = \frac{\mu_g}{\mu_o} \quad (2.6)$$

式(2.6) μ_g 和 μ_o 分别是各个临界频带功率谱密度的几何平均值和算术平均值。由频谱平坦性量度, 可以导出音调系数 α :

$$\alpha = \min\left(\frac{SFM_{db}}{-60}, 1\right) \quad (2.7)$$

当 α 趋近于 1 时, 音频窗就是类音调的, 而当 α 趋近于 0 时, 就是类噪声的。音调系数又可以用来计算偏移量 $O[z]$, 公式为:

$$O[z] = \alpha(14.5 + z) + (1 - \alpha)5.5 \quad (2.8)$$

将每个临界频带的视在能量 $B_o[z]$ 转换成以分贝表示, 再减去偏移量 $O[z]$ 就得到该频带的掩蔽值, 因此, 用分贝表示的阈值就为:

$$t_{db}[z] = 10\log_{10} B_o[z] - O[z] \quad (2.9)$$

用帕斯卡表示则为:

$$t_{spl}[z] = 10^{t_{db}[z]/10} \quad (2.10)$$

要想知道每个频率 f 在一个临界频带中带宽能做多少改动, 就必须根据对应频带中的离散频率数 $T_{spl}[z]$ 进行归一化, 从而得到归一化检测阈值 $t_n[f]$:

$$t_n[f] = \frac{t_{spl}[z_f]}{Nz_f} \quad (2.11)$$

式中 $Z[f]$ 中包含频率 f 的临界频带, Nz_f 为该频带中的点数。

将所得到的归一化阈值 $t_n[f]$ 和听力绝对阈值 $t_q[f]$ 相比较, 最大者即为最终掩蔽阈值 $t[f]$:

$$t[f] = \max(t_n[f], t_q[f]) \quad (2.12)$$

合并

现在我们已经求得了各个频率的掩蔽阈值, 即, 可被识别的最小声压水平。如果一个频率 f 中的真实能量 $|C[f]|$ 低于这个阈值 $t[f]$, 就不能被听见, 因此, 信号 $|C[f]|$ 可以在不超出掩蔽阈值的允许范围内变化。如果真实能量超出对应的掩蔽阈值, 该模型不能确定信号在不被识别的情况下变化的范围。因为一个频率的变化对于其他频率的变化是独立的, 这就等效于用 L_∞ 范数进行合并。

4 结论

作品中嵌入的水印应该是不能被感知的, 在真实能量不超出对应的掩蔽阈值的情况下, 可以通过设计的感知模型来进行测量, 这为我们检测作品中是否有水印提供了一种方法。这个感知模型的一个典型的用处是对水印能够进行局部缩放, 即对某些区域削弱而对另外一些区域放大, 从而使得水印能够更好的被载体作品隐藏, 增强了作品中嵌入水印的不可感知性。

参考文献

- 1 M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney. "Robust Audio Watermarking Using Perceptual Masking," *Signal Processing*, 66(3):337-355, 1998.
- 2 信息安全丛书《数字水印》[美] Ingemar J. Cox、Mathew L. Miller、Jeffrey A. Bloom 著 王颖等译, 2003. 7.
- 3 赵海燕, 《脆弱性水印与版权保护研究综述》信息技术, 2005 年第 6 期.