(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102867513 A
(43)申请公布日 2013.01.09

(21)申请号 201210278724.3

(22)申请日 2012.08.07

(71)申请人 西南交通大学
地址 610031 四川省成都市二环路北一段 111 号西南交通大学科技处

(72)发明人 王宏霞 刘正辉

(74)专利代理机构 成都信博专利代理有限责任公司 51200
代理人 张澎

(51)Int.Cl.
G10L 17/00(2013.01)

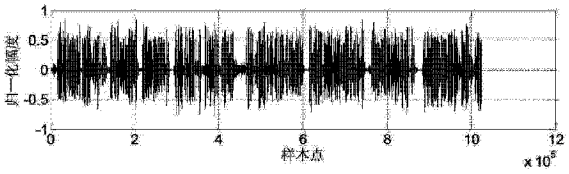
权利要求书 1 页 说明书 5 页 附图 2 页

(54)发明名称

一种基于伪 Zernike 矩的语音内容认证方法

(57)摘要

本发明公开了一种基于伪 Zernike 矩的语音内容认证方法，水印嵌入时将原始语音信号 A 分为 P 帧，每帧分为 N 段。然后，由每帧前 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值的平均值生成水印 W，并通过量化每帧的后 N/2 段 DCT 低频系数的伪 Zernike 矩来嵌入水印，得到含水印语音 A′。本发明充分利用了语音信号 DCT 低频系数的伪 Zernike 矩幅值与语音内容紧密相关的特性，以及对常规语音信号处理的鲁棒性，既保证了对恶意篡改攻击的敏感性，又保证了良好的容忍一定常规语音信号处理能力。



CN 102867513 A

1. 一种基于伪 Zernike 矩的语音内容认证方法，用以区分常规信号处理操作和恶意攻击，同时对恶意攻击能够有效地篡改定位，具体步骤包括：

(1) 水印嵌入：首先从语音信号的第 K 个样本点开始将原始语音信号 A 分为 P 帧，并将每帧分为 N 段；然后计算每帧前 N/2 段离散余弦变 DCT 低频系数的 n 阶伪 Zernike 矩幅值之和，并求出伪 Zernike 矩幅值的均值，由均值生成水印 W；将得到的水印通过量化 DCT 低频系数的伪 Zernike 矩嵌入在每帧的后 N/2 段中，得到含水印语音 A′；

(2) 语音内容认证过程：与水印嵌入过程类似，首先从待检测语音信号 A^* 的第 k_1 个样本点开始将语音分为 P 帧，每帧分为 N 段。计算每帧前 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值之和，并求其均值，由均值生成水印 W′；计算每帧后 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值，由 Zernike 矩的幅值提取出水印 W^*；比较 W^* 和 W′，判断对应位不同的地方为语音信号被篡改过的位置，从而实现了语音内容真实性和完整性认证。

一种基于伪 Zernike 矩的语音内容认证方法

技术领域

[0001] 本发明涉及一种语音识别，尤其是语音内容真实性和完整性认证问题的解决方法。

背景技术

[0002] 近年来，数字化语音通信的迅猛发展和各种语音产品的广泛普及，以及各种功能强大的音频处理软件的出现，使得数字语音的传输与应用日益变得频繁与广泛。与此同时，篡改传输和存储的语音内容数据变得相对容易。例如，一段重要的法庭证词录音，在存储、传输过程中如果要害部分内容被恶意篡改，其后果可想而知！。因此，如何鉴别一段重要或敏感的语音内容是否被篡改过，哪里被篡改了，语音记录来源是否真实、可信，这些涉及数字语音真实性的认证问题，引起了国内外学者极大的研究兴趣。音频水印技术作为一种保护音频的技术手段，从上世纪 90 年代出现就受到了人们的重视，并成为信息安全研究领域的热点。

[0003] 与音频信号相比，语音信号具有采样率低，对常规信号处理更加敏感等特点。因此，现有的很多音频内容认证算法无法用于语音内容认证，或者用于语音内容认证的效果不是很理想。现实生活中，针对音频更多的是解决版权保护问题，而针对语音则更多的是解决内容真实性和完整性认证问题。基于数字水印的语音内容认证技术，如果嵌入的水印与语音自身内容无关，一方面会增加信息的传输量，另一方面也存在一定的安全隐患，所以基于语音自身特征或内容来生成水印的语音认证算法就更具有研究意义和实用价值。

[0004] 伪 Zernike 矩 (Zernike 矩) 的幅值具有旋转不变性的特征，该特征已广泛地应用于图像表示、图像检索和图像水印等领域，而在音频上的应用还很少。文献“Robust audio watermarking based on low-order Zernike moments”(Xiang Shi-jun, Huang Ji-wu, Yang Rui, 5th International Workshop on Digital Watermarking, pp226-240, Oct.2006) 首先对音频进行一维到二维的变换，然后对相应的二维信号进行 Zernike 变换。通过实验证明了 Zernike 矩的幅值对常规信号处理具有很强的鲁棒性；同时分析了 Zernike 矩的幅值和音频样本值的线性关系，由此提出了基于低阶 Zernike 矩的鲁棒音频水印算法。文献“A pseudo-Zernike moments based audio watermarking scheme robust against desynchronization attacks”(Wang Xiang-yang, Ma Tian-xiao, Niu Pan-pan, Computers and Electrical Engineering, vol.37, no.4, pp.425-443, July 2011) 首先在时域基于统计均值嵌入同步码，然后量化伪 Zernike 矩的幅值嵌入水印，提出了基于伪 Zernike 矩的抗同步攻击的音频水印算法。对于上述的基于伪 Zernike 矩 (Zernike 矩) 的水印算法，一方面，需要计算所有样本点的伪 Zernike 矩，计算量较大，耗费的时间较长。水印的嵌入是通过同比例地缩放各音频段的样本值来完成的。分析表明，直接缩放音频样本值对原始音频的改变量较大，对原始音频信号的质量造成较大的破坏；另一方面，水印的嵌入位置和方法是公开的，各音频帧的特征 (伪 Zernike 矩) 的计算也是已知的。于是，攻击者可以找到各音频帧的位置并计算每帧的特征，重新量化伪 Zernike 矩来去除嵌入的水

印，使算法失去保护版权的作用。或者，攻击者可以使用其它的音频段来替换含水印的音频，然后量化替换后的音频内容，使其满足水印正确提取的条件，对其内容实施攻击。因此，研究基于内容的抗攻击能力强的语音内容认证算法具有重要的现实意义。

发明内容

[0005] 鉴于现有技术的不足，本发明的目的在于提供一种基于伪 Zernike 矩的语音内容认证算法，该算法能够有效地区分对语音的常规信号处理操作和恶意攻击，并能有效定位语音内容恶意篡改的位置，从而实现语音内容的真实性和完整性认证。

[0006] 为实现这样的目的，本发明以 DCT 低频系数的伪 Zernike 矩幅值对常规信号处理的鲁棒性为依据，设计了一种新的水印生成和嵌入方法。

[0007] 一种基于伪 Zernike 矩的语音内容认证方法，能够有效区分常规信号处理操作和恶意攻击，同时对恶意攻击能够有效地篡改定位。从而实现语音内容的真实性和完整性认证，包括如下具体步骤：

[0008] (1) 水印嵌入：首先从语音信号的第 K 个样本点开始将原始语音信号 A 分为 P 帧 (K 作为水印系统的密钥)，并将每帧分为 N 段。然后计算每帧前 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值之和，并求出伪 Zernike 矩幅值的均值，由均值生成水印 W。将得到的水印通过量化 DCT 低频系数的伪 Zernike 矩嵌入在每帧的后 N/2 段中，得到的含水印的语音信号记为 A′

[0009] (2) 语音内容认证过程：与水印嵌入过程类似，首先从待检测语音信号的第 k_1 个样本点开始将 A^* 分为 P 帧，每帧分为 N 段。计算每帧前 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值之和，并求其均值，由均值生成水印 W′ 。计算每帧后 N/2 段 DCT 低频系数的 n 阶伪 Zernike 矩幅值，由 Zernike 矩的幅值提取出水印 W^*。比较 W^* 和 W′ ，判断那些对应位不同的地方为语音信号被篡改过的位置，从而实现了语音内容真实性和完整性认证。

[0010] 与现有的用于内容认证的语音水印算法相比，本发明利用语音的内容来生成水印，接收端在收到语音信号的同时也收到了嵌入在语音信号中的水印。从而减少了传输带宽，节约了资源；同时也增强了水印传送的安全性。水印的嵌 入只需要对 DCT 低频系数进行伪 Zernike 变换，提高了算法的效率和水印容忍常规信号处理的能力。于是本发明更易于实际应用。

[0011] 附图说明

[0012] 图 1 为本发明实施例的含水印语音信号图。

[0013] 图 2 为对图 1 部分语音内容静音攻击后的语音信号图。

[0014] 图 3 为对图 1 部分内容替换攻击后对应的语音信号图。

[0015] 图 4 为图 2 的篡改定位结果。

[0016] 图 5 为图 3 的篡改定位结果。

[0017] 图 6 为不可听性测试结果列表。

[0018] 图 7 为对常规信号处理的鲁棒性测试结果列表。

具体实施方式

[0019] 以下结合附录和实施例对本发明的技术方案作进一步描述。

[0020] 1、水印的生成和嵌入：

[0021] (1) 语音数据的分帧以及每帧语音段的划分。将原始语音信号 A = {a(1), 1 ≤ 1 ≤ LA+K} 分为 P 帧 (K 作为水印系统的密钥)，每帧长为 I = LA/P，第 i 帧记为 A(i)(i = 1,2,...,P)。每帧等分为 N 段，每段的长为 I/N，第 i 帧第 j 段记为 A(i,j)，1 ≤ i ≤ P，1 ≤ j ≤ N。

[0022] (2)DCT 变换。对 A(i,j) 做 DCT 变换，D(i,j) 表示第 i 帧第 j 段的 DCT 系数，取 i 帧前 N/2 段的 DCT 系数记为 $D_1(i,j)$。

[0023] (3) 计算 n 阶 m 重伪 Zernike 矩。将 $D_1(i,j)$ 的前 $m_1 \times m_1$ 个低频系数变换为二维信号。按照如下方法计算其 n 阶 m 重伪 Zernike 矩：

[0024] 记 $\{V_{nm}\}$ 为伪 Zernike 多项式，它是一系列复值多项式构成的集合，$\{V_{nm}\}$ 构成单位圆内的完备正交基，其定义如下式

[0025] $V_{nm}(x,y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)\exp(im\theta)$

[0026] 其中 n 为非负整数，m 为满足 | m | ≤ n 的整数。记坐标原点到点 (x, y) 的向量为 1，ρ = | 1 |，θ 为 x 轴正向到向量 1 逆时针方向的夹角。$R_{nm}(\rho)$ 为径向多项式，即

[0027] $$R_{nm}(\rho)=\sum_{s=0}^{n-|m|}\frac{(-1)^s(2n+1-s)!}{s!(n+|m|+1-s)!(n-|m|-s)!}\rho^{n-s}$$

[0028] 坐标平面内的二维信号 f(x, y)($x^2+y^2 \leq 1$) 可以表示为 $V_{nm}(x, y)$ 的线性组合，如下式

[0029] $$f(x,y)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}A_{nm}V_{nm}^{*}(x,y)$$

[0030] 其中 $V_{nm}^{*}(x,y)$ 和 $V_{nm}(x,y)$ 互为共轭复数，A_{nm} 为 n 阶 m 重伪 Zernike 矩，定义如下：

[0031] $$A_{nm}=\frac{n+1}{\pi}\sum_{x}\sum_{y}f(x,y)V_{nm}^{*}(x,y),x^2+y^2\leq 1$$

[0032] (4) 语音水印的生成。取各帧的前 N/2 段来生成水印。记 $C_1(i,j)=\sum_{m=-n}^{n}|A_{nm}|$，1 ≤ i ≤ P，1 ≤ j ≤ N/2 为 n 阶伪 Zernike 矩的幅值之和，计算 $C_1(i,j)$ 的均值 $\overline{C}_1(i)=\sum_{j=1}^{N/2}C_1(i,j)\Big/N/2$。记 $\lfloor \overline{C}_1(i) \rfloor$ 的最高位为 $M_1(i)$，$M_1(i)$ 对应的二进制设为 $W_1(i)$ = $\{w_1(i,t), 1 \leq t \leq N/2\}$，$W_1(i)$ 即为 i 帧生成的水印。

[0033] (5) 水印的嵌入。取 i 帧后 N/2 段的 DCT 系数记为 $D_2(i, j)$，N2+1 ≤ j ≤ N，将 $D_2(i, j)$ 的前 $m_2 \times m_2$ 个低频系数变换为二维信号，并计算其 n 阶伪 Zernike 矩幅值之和，记为 $C_2(i, j)$。记 $\lfloor C_2(i,j) \rfloor$ 的最高位为 $M_2(i, j)$，水印按照下面的方法嵌入：

[0034] 当 $w_1(i, t) = 1$ 时

[0035] $$M_2'(i,j)=\begin{cases}M_2(i,j), & M_2(i,j)\bmod 2=1\\ M_2(i,j)+1, & M_2(i,j)\bmod 2=0\end{cases}$$

[0036] 当 $w_1(i, j) = 0$ 时

[0037] $$M_2'(i,j)=\begin{cases}M_2(i,j), & M_2(i,j)\bmod 2=0\\ M_2(i,j)+1, & M_2(i,j)\bmod 2=1\end{cases}$$

[0038] 上式中，当 $M_2(i,j)=9$ 时，$M_2'(i,j)=M_2(i,j)-1$；$j=t+N2$，$1\leqslant t\leqslant N2$。用 $M_2'(i,j)$ 替换 $C_2(i,j)$ 整数部分最高位，并将次高位量化为 5，对应的值记为 $C_2'(i,j)$。

[0039] 将 $D_2(i,j)$ 的前 $m_2\times m_2$ 个低频系数扩大 $\alpha_2(i,j)$ 倍，对应的值记为 $D_2'(i,j)$，$\alpha_2(i,j)$ 可由下式得到：

[0040] $$\alpha_2(i,j)=\frac{C_2'(i,j)}{C_2(i,j)},\quad N/2+1\leq j\leq N$$

[0041] 对 $D_2'(i,j)$ 做逆 DCT，得到的信号即为 i 帧的后半部分内容，i 帧前半部分和后半部分结合在一起即为 i 帧含水印语音信号。

[0042] (6) 对 P 个语音帧依次进行这样的嵌入，直至嵌完所有语音帧，便得到含水印语音 A′。

[0043] 2、语音内容认证：

[0044] (1) 类似水印生成及嵌入过程的步骤 (1) ～ (4)，对待检测的语音信号 A^* 从 K 个样本点开始等分为 P 帧，每帧等分为 N 段，第 i 帧记为 $A^*(i)(i=1,2,\ldots,P)$，第 i 帧第 j 段记为 $A^*(i,j)$，$1\leqslant j\leqslant N$；对 $A^*(i,j)$ 做 DCT，对应的 DCT 系数记为 $D^*(i,j)$。取 i 帧前 N/2 段的 DCT 系数记为 $D_1^*(i,j)$，将 $D_1^*(i,j)$ 的前 $m_1\times m_1$ 个低频系数变换为二维信号，并计算其 n 阶伪 Zernike 矩幅值之和，记为 $C_1^*(i,j)$，$1\leqslant j\leqslant N/2$。计算 $C_1^*(i,j)$，$1\leqslant j\leqslant N/2$ 的均值 $\overline{C}_1^*(i)$，$\overline{C}_1^*(i)=\sum_{j=1}^{N/2}C_1^*(i,j)\Big/N/2$。记 $\lfloor\overline{C}_1^*(i)\rfloor$ 的最高位为 $M_1^*(i)$，$M_1^*(i)$ 二值化为 $W_1^*(i)=\{w_1^*(i,t),1\leq t\leq N/2\}$，$W_1^*(i)$ 即为 i 帧生成重构的水印。

[0045] (2) 取 i 帧后 N/2 段的 DCT 系数记为 $D_2^*(i,j)$，将 $D_2^*(i,j)$ 的前 $m_2\times m_2$ 个低频系数变换为二维信号，并计算其 n 阶伪 Zernike 矩幅值之和，记为 $C_2^*(i,j)$，N2+1 ≤ j ≤ N。记 $\lfloor C_2^*(i,j)\rfloor$ 的最高位为 $M_2^*(i,j)$，进行如下计算获得提取的水印

[0046] $$\hat{W}_1^*(i)=\{\hat{w}_1^*(i,t),1\leq t\leq N/2\}$$

[0047] $$\hat{w}_1^*(i,t)=\begin{cases}1 & M_2^*(i,t+N/2)\bmod 2=1\\ 0, & M_2^*(i,t+N/2)\bmod 2=0\end{cases}$$

[0048] (3) 定义认证序列 TA(i) 为

[0049] $$TA(i)=\sum_{t=1}^{N/2}\hat{w}_1^*(i,t)\oplus w_1^*(i,t),T\in\{0,1\}$$

[0050] 如果 TA(i)=0，则表明第 i 帧语音内容是真实的，否则，TA(i)=1 表明第 i 帧语音内容被篡改。

[0051] 本发明方法的效果可以通过以下的性能分析验证：

[0052] 1、不可听性

[0053] 选取采样率为 22.05kHz，样本长度为 1024078，16 位量化的单声道语音信号来做不可听性测试。图 6 给出了 3 种语音类型的 SNR 值，由测试结果可以看出本文算法具有很好的不可听性。

[0054] 2、对常规信号处理的鲁棒性

[0055] 用误码率 BER(bit error rate) 来测试本文算法对常规信号处理的鲁棒性，BER 的

定义如下式

[0056] $$BER=\frac{E}{T}\times 100\%$$

[0057] 其中，E 为提取水印错误比特数，T 为语音信号所嵌水印总比特数。BER 值越小说明算法对常规信号处理的鲁棒性越强。

[0058] 图 7 列出了成年男声在经过一些常规信号处理后的 BER 值（其它类型语音信号的测试结果与此相似），可以看出本发明方法对 MP3 压缩、低通滤波、重采样等常规语音信号处理具有较强的鲁棒性。

[0059] 3、恶意篡改定位

[0060] 对如图 1 所示的含水印语音信号分别进行了静音和替换攻击。攻击后的语 音信号分别如图 2 和图 3 所示，对应的篡改定位结果分别如图 4 和图 5 所示。图 4、图 5 中，TA(i)=1 的帧表示被恶意攻击的部分，TA(i)=0 的帧表示没有恶意攻击的部分。从篡改定位的结果来看，本发明方法对恶意攻击能够有效地篡改定位。

[0061] 上述针对较佳实施例的描述过于具体，本领域的普通技术人员将会意识到，这里所述的实施例是为了帮助阅读者理解本发明的原理，应被理解为发明的保护范围并不局限于这样的特别陈述和实施例。

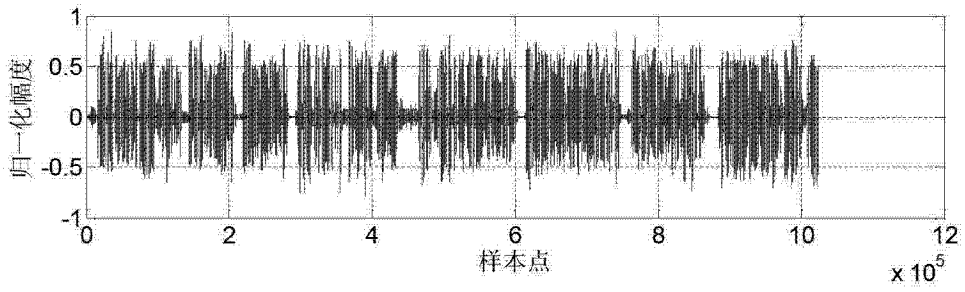


图 1

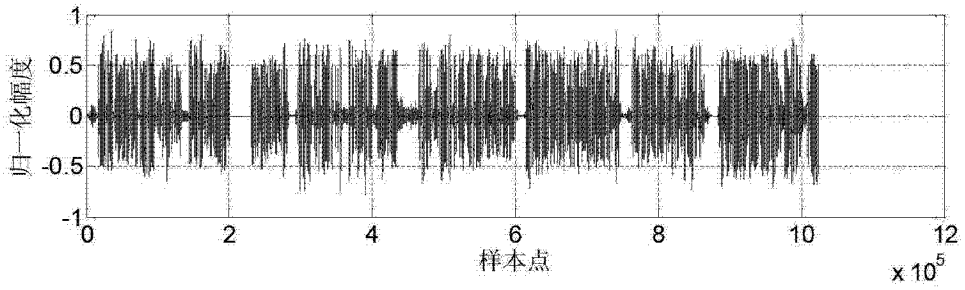


图 2

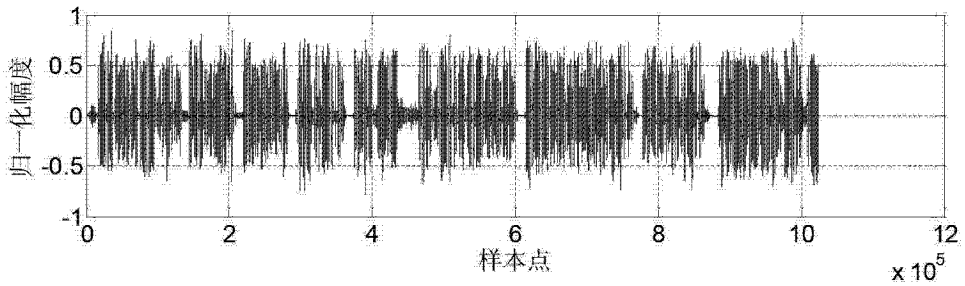


图 3

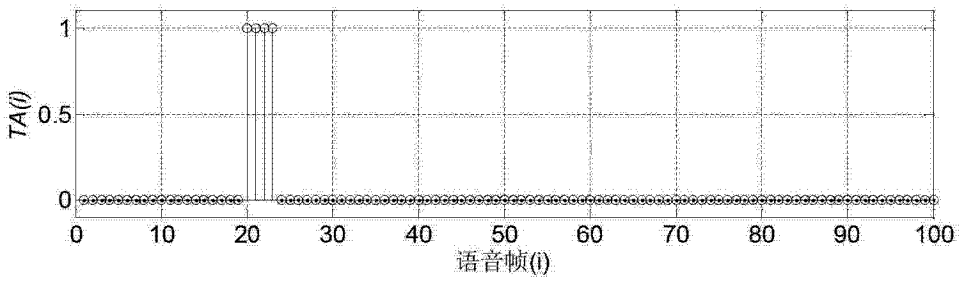


图 4

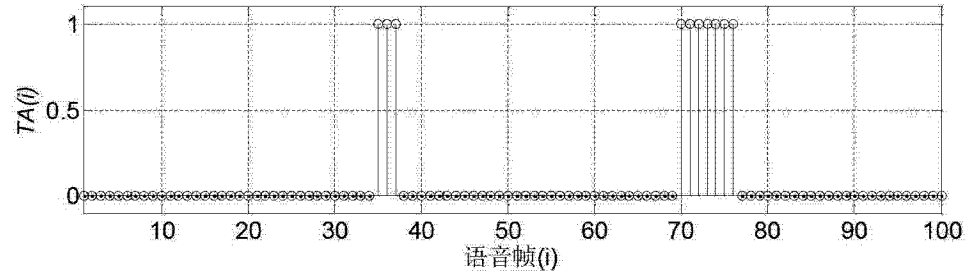


图 5

语音类型	成年男声	成年女声	童声
SNR(dB)	43.0606	42.0573	41.4364

图 6

常规信号处理	BER
MP3 压缩(64kbps)	0.0328
MP3 压缩(96kbps)	0
MP3 压缩(128kbps)	0
重采样(44.1→8→44.1kHz)	0
重采样(44.1→11.025→44.1kHz)	0
低通滤波(11.025kHz)	0

图 7