

权利要求书

1. 一种人脸声纹特征融合验证的方法，其特征在于，包括以下步骤：

将输入的声音文件解析成声音的时域信号；

通过短时傅里叶变换和加窗分帧将所述时域信号转变为频域信号；

通过对数频谱变换将所述频域信号中的频率转换成人耳能感知的线性关系；

通过倒频分析，采用 DCT 变换将转换后的频域信号中的直流信号分量和正弦信号分量分离；

提取声音频谱特征向量，将所述向量转换成图像；

将所述图像与二维人脸图像融合，具体为：将倒频谱图顺时针旋转 90 度，如果拼接后的图像的横轴长度和旋转 90 度后的倒频谱图的横轴长度不一致，则缩放二维人脸图像，使二者横轴长度一致，再将二者进行拼接。

2. 根据权利要求 1 所述的一种人脸声纹特征融合验证的方法，其特征在于，所述通过短时傅里叶变换和加窗分帧将所述时域信号转变为频域信号，具体为：

选择一个时频局部化的窗函数 $h(t)$ ，通过短时傅里叶变换计算出各个不同时刻的功率谱，所述短时傅里叶变换的公式为：

$$F(\omega, t) = \int_{-\infty}^{+\infty} f(\tau) h(\tau - t) e^{-j\omega\tau} d\tau$$

其中， $f(\tau)$ 代表输入音频的时域信号， τ 表示积分变量， t 表示所述不同时刻。

3. 根据权利要求 2 所述的一种人脸声纹特征融合验证的方法，其特征在于，所述窗函数为海明窗。

4. 根据权利要求 1 所述的一种人脸声纹特征融合验证的方法，其特征在于，所述通过对数频谱变换将频率转换成人耳能感知的线性关系，具体为：

通过下式将频率标度转化为对数频率标度，使人耳对频率的感知度为线性感知关系：

$$mel(f) = 2595 * \log_{10}(1 + f / 700)$$

其中， $mel(f)$ 表示对数频率， f 表示短时傅里叶变换后得到的频率。

5. 根据权利要求 1 所述的一种人脸声纹特征融合验证的方法，其特征在于，所述通过倒频分析，采用 DCT 变换将转换后的频域信号中的直流信号分量和正弦信号分量分离，具体为：

$$mfcc(u) = c(u) \sum_{i=0}^{N-1} mel(i) \cos\left[\left(\frac{i+0.5}{N}\right)\pi u\right]$$

其中，

$$c(u) = \begin{cases} \sqrt{\frac{1}{N}}, u = 0 \\ \sqrt{\frac{2}{N}}, u \neq 0 \end{cases}$$

其中，mfcc(u)表示倒频谱，mel(i)表示对数频率，N 表示频率点的数量，u 表示倒频谱的频率点。

6. 根据权利要求 1 所述的一种人脸声纹特征融合验证的方法，其特征在于，所述提取声音频谱特征向量，将所述向量转换成图像，具体为：

将输出向量的范围：

$$mfcc \in [\min, \max]$$

线性变换到图像的范围：

$$pixel \in [0, 255]$$

$$pixel = \frac{mfcc - \min}{\max - \min} \times 255$$

这样就得到了声音的倒频谱图，所述声音的倒频谱图的横轴为时间，纵轴为频率；其中，mfcc 表示倒频谱，min 表示 mfcc 的最小值，max 表示 mfcc 的最大值，pixel 表示转换为图像之后的像素。