IMPLEMENTACIÓN DE ALGORITMO PARA MARCAS DE AGUA EN AUDIO

[1] FEDERICO FAZIO Y [2] MATEO GARCIA IACOVELLI

Universidad Nacional de Tres de Febrero, Procesamiento Digital de Señales, Buenos Aires, Argentina.
[1] faziofederico9@gmail.com
[2] mateogi99@gmail.com

RESUMEN *- En el siguiente trabajo, se desarrolla e implementa un algoritmo que permite la incorporación de una marca de agua a dos audios (un archivo de voz y un archivo musical), y su posterior recuperación. En este caso, se emplea una como marca de agua una imagen de dos colores únicamente. Para ello, se hace uso de principalmente dos herramientas: La Transformada de Fourier de Tiempo Reducido (STFT), y del algoritmo de Descomposición de Valores Simples (SVD). Luego, se evalúa en ambos audios la perceptibilidad de la imagen embebida por medio de la prueba Relación Señal Pico - Ruido (PSNR), y la Correlación Normalizada (NC). Ambas ofrecen un valor cuantitativo objetivo de perceptibilidad. Para la prueba de NC, se decide someter a los archivos de audio a dos ataques. Por un lado se les suma un ruido gaussiano, lo cual genera distorsión; y por otro lado, se le reemplazan componentes a los archivos originales por ceros.*

ABSTRACT *- In the next report, an algorithm is developed and implemented which allows the incorporation of a watermark into two audios (a voice file and a music file), and its subsequent recovery. In this case, a two-color image is used as a watermark.Two main tools are used: The Reduced Time Fourier Transform (STFT), and the Simple Value Decomposition (SVD). Then, the perceptibility of the embedded image is evaluated in both audios by means of the Peak Signal-Noise Ratio (PSNR), and the Normalized Correlation (NC). Both tests offer an objective quantitative value of perceptibility. For the NC test, it is decided to attack the audio files by adding Gaussian noise, which generates distortion to the signals; and to replace some original files components by zeros.*

1. INTRODUCCIÓN

En este trabajo se diseña un algoritmo que tiene como fin ocultar una imagen dentro de un archivo de audio. El objetivo del trabajo es obtener un archivo de audio que contenga incrustada a la imagen, pero que no presente cambios perceptibles en la escucha al momento de ser comparado con el archivo original. Para esto se aplica el concepto marca de agua de audio o audio watermarking. El audio watermarking es el proceso digital en el cual mediante la utilización de un algoritmo se embebe en un archivo de audio una secuencia de datos de forma que estos no sean detectables para el odio humano en el proceso de escucha.

Una vez obtenido el audio con la marca de audio incluido se aplica otro proceso con el fin de poder leer la marca de agua para así poder ver el mensaje o en este caso la imagen oculta en el mismo.

2. MARCO TEÓRICO

Se hace uso de principalmente dos herramientas: La Transformada de Fourier de Tiempo Reducido (STFT), y el método de Singular Value Decomposition (SVD).

2.1 TRANSFORMADA DE FOURIER DE TIEMPO REDUCIDO (STFT)

La STFT se encuentra relacionada con la Transformada de Fourier usada para determinar el contenido en frecuencia y de fase de una señal, así como sus cambios respecto al tiempo. En el caso del tiempo discreto, la información a ser transformada podría ser dividida en tramos (que usualmente se solapan unos con otros, para reducir irregularidades en la frontera). Cada tramo corresponde a una transformación de Fourier, y el resultado complejo se agrega a una matriz que almacena magnitud y fase para cada punto en tiempo y frecuencia. Esto se puede expresar en base a la siguiente ecuación:

$$STFT\{x[n]\} \equiv X(m,w) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}$$

Donde $x[n]$ es la señal y $w[n]$ es la ventana.

La STFT es invertible, lo cual significa que la señal original puede ser recuperada de la transformación por medio de la ISTFT.

2.2 DESCOMPOSICIÓN DE VALORES SINGULARES (SVD)

Por otro lado, se sabe que toda matriz A de orden M x N, puede ser factorizada de la siguiente forma:

$$A = U\,D\,V^T$$

Donde U es una matriz de dimensiones mxn, con columnas ortogonales; V tiene dimensiones nxn y D una matriz diagonal de dimensiones nxn, la cual contiene los valores singulares de A. A este resultado se lo conoce como Singular Value Decomposition (SVD). Los valores singulares de A son la raíces cuadradas de los autovalores de A^TA, y se denotan mediante $\sigma_1, ..., \sigma_n$. Es una convención acomodar los valores singulares de modo que $\sigma_1 \geq |\sigma_2 \geq ... \geq \sigma_n$. [1]

2.3 RELACIÓN SEÑAL PICO - RUIDO (PSNR)

Es una de las principales pruebas para evaluar la imperceptibilidad y robustez de una marca de audio sumergida en un determinado archivo anfitrión. Se la define como[2]:

$$PSNR = 10log_{10}(\frac{max(I(i,j))^2}{MSE})$$

Donde $max(I(i,\ j))$ corresponde a la componente de máximo valor de la matriz STFT del archivo de audio original, es decir, sin tener la marca de agua embebida. Por otro lado, el error cuadrático medio (MSE), entre el audio original I, y el audio con la imagen embebida I^w, se evalúa de la siguiente forma:

$$MSE = \frac{1}{M\ x\ N}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,\ j)\ -\ I^w(i,\ j))^2$$

Donde M y N son los números de filas y columnas de los archivos de audio.

Cuanto mayor sea el valor de PSNR, menor será la diferencia existente entre el audio original y el audio con marca de agua.

2.4 CORRELACIÓN NORMALIZADA (NC)

La correlación normalizada es un parámetro que mide la diferencia que hay entre la marca de agua empleada originalmente W, y aquella que se extrae del audio W^{new}. La misma se calcula de la siguiente forma[2]:

$$NC(W,\ W^{new}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j)-\mu_1)(W^{new}(i,j)-\mu_2)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j)-\mu_1)^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W^{new}(i,j)-\mu_2)^2}}$$

Donde μ_1 y μ_2 son los valores medios de W y W^{new} respectivamente.

Cuanto mayor sea la semejanza entre la marca de agua original y la extraída, el valor de NC tenderá a 1.

La marca de agua puede verse alterada si la señal anfitriona sufre determinados ataques, como puede ser el agregado de un ruido gaussiano, el reemplazo de ciertos valores que componen la matriz STFT del audio por ceros. Dichos ataques, provocarán una modificación en la marca de agua, por lo tanto, una vez extraída presentará diferencias con respecto a la marca original.

3. PROCEDIMIENTO

Se seleccionan dos señales de audio diferentes a las que se les aplicará el mismo proceso con el fin de poder comparar resultados. Por un lado se selecciona una señal correspondiente a una grabación de voz, la cual se enseña en la *Figura 1*,

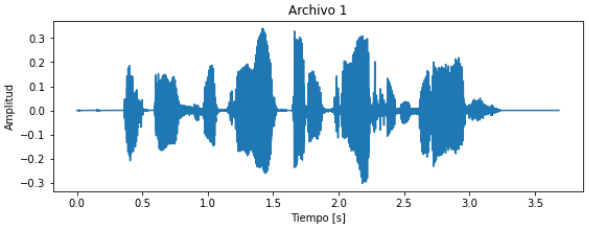

Figura 1 - Forma de archivo de voz.

y por otro lado, una señal correspondiente a una canción. Esta última se muestra en la *Figura 2.*

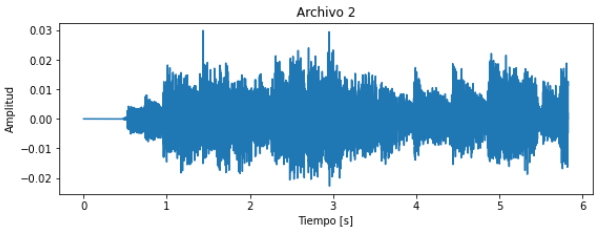

Figura 2 - Forma de onda archivo musical.

Luego se realiza el siguiente procedimiento.

En primer lugar, se calcula la STFT del archivo de audio seleccionado. Como resultado se obtiene una matriz de orden nxm, donde n corresponde a la cantidad de frecuencias presente en cada cálculo de la FFT para una determinada ventana y m la cantidad de frames, es decir, a la cantidad de ventanas total a las cuales se les calcula la FFT.

Por otro lado, se selecciona la imagen que se desea embeber en el archivo de audio. En este caso, se trata de una imagen de 200x200 pixels como se ve en la *Figura 3*.

Figura 3 - Imagen utilizada como marca de agua

El paso siguiente consiste en aplicar la descomposición de valores singulares, tanto a la matriz que contiene los valores de la STFT del audio, como también a la matriz que contiene la información binaria de la imagen. De este modo se obtiene, por un lado tres matrices U_{H}, S_{H} y $V^T{}_{\mathrm{H}}$ las cuales corresponden a la descomposición SVD de la STFT audio y por otro lado otras tres matrices U_{W}, S_{W} y $V^T{}_{\mathrm{W}}$ que corresponden a la descomposición de la matriz de la imagen.

3.1 EMBEBIDO

En este punto se genera una nueva matriz diagonal S_{new} de la siguiente forma

$$S_{New} = S_H + (S_W\ .\ \alpha)$$

Donde α corresponde al índice de robustez de la marca de agua. Luego se realiza el producto matricial entre U_{H} , S_{New} y $V^T{}_{\mathrm{H}}$ obteniéndose como resultado una nueva matriz H_{w} , por último se aplica la ISTFT a esta matriz para poder obtener un vector que contiene la información de la data del audio con la imagen embebida.

3.2 ATAQUES A LA MARCA DE AGUA

Al archivo resultante del proceso de embebido se le realizan dos tipos de ataques destructivos con el fin de ver cómo esto afecta a la marca de agua.

El primer ataque consiste en sumar al audio con la marca de agua un ruido aleatorio de distribución gaussiana de media cero y desviación estándar 0.01.

El segundo ataque se basa en la sustitución por cero de los primeros valores del archivo de audio que contiene a la marca de agua.

3.3 EXTRACCIÓN DE LA MARCA DE AGUA

Al audio que contiene la imagen embebida se le aplica la STFT, a la matriz resultante de la STFT se le realiza la descomposición en valores singulares de modo que se obtiene tres matrices U_{H}, S_{HW} y $V^{T}{}_{\mathrm{H}}$. Donde S_{HW} corresponde a la matriz que contiene los datos de la marca de agua previamente embebida. Se define una nueva matriz S_{w} de la forma

$$S_W = (S_{New} - S_{Hw})/\alpha$$

Idealmente S_{W} debería ser exactamente igual a la matriz diagonal que contiene los valores singulares producto de la descomposición de la matriz binaria de la imagen. A partir de esto se define reconstruye a la imagen embebida como el producto matricial entre U_{W}, S_{W} y V^{T} W.

4. ANÁLISIS DE RESULTADOS

En primer lugar, se analizó el comportamiento de la marca de agua para distintos valores de la constante de robustez α. Para esto se dejaron constantes los valores para el cálculo de la STFT. Tomando para ventanear una ventana tipo hanning, de longitud 25 ms con un solapamiento del 50%. Luego se procedió a variar los valores de α. Para $\alpha = 1$ se notaron grandes cambios desde el punto de vista perceptivo al reproducir el audio, este se escucha con una notable distorsión respecto del audio original. Para comprobar esto de manera numérica, se calculó los valores para el PSNR y se obtuvieron los resultados de -8dB para el audio 1 y -27dB para el audio 2, estos valores tan bajos reflejan la alta perecibilidad de la marca de agua a la hora de escuchar el audio. Por otro lado, para este α, la marca de agua presenta una gran robustez. Esto se evidencia en la *Figura 4*, donde se puede notar que, a pesar de los diferentes ataques realizados, la marca de agua permanece prácticamente idéntica en todos los casos. Además esto se comprueba calculando el NC entre la marca de agua obtenida del audio sin ataques y las obtenidas de los diferentes audios atacados, y en todos los casos se obtiene como resultado un valor para el NC igual a 1.

Figura 4- Marca de agua extraídas de los diferentes archivos de audio para un valor de $\alpha = 1$.

Para α =0.01 ocurre que, a nivel audible no se aprecian grandes diferencias entre el audio con la marca de agua y el original. Esto se ve reflejado en los valores del PSNR, obteniéndose un valor de 52 dB para el audio 1 y 33 dB para el audio 2. En cuanto a la robustez de la marca de agua se nota una clara desmejora respecto del valor de α más alto. En este caso, tal como se ve en la Figura 5.

Figura 5 -Marca de agua extraídas de los diferentes archivos de audio para un valor de $\alpha = 0.001$.

Se aprecia una clara diferencia entre la marca de agua obtenida directamente, y entre las obtenidas de los archivos a los que se le realizaron los ataques. De forma cualitativa esto se puede comprobar con el cálculo del NC como se muestra en la *Tabla 1*.

Correlacion Normalizada

	Attack 1	Attack 2
Audio 1	0.580935	0.942713
Audio 2	0.473851	0.997849

Tabla 1 - Valores del NC para $\alpha = 0.001$

En segundo lugar, se modificaron algunos parámetros involucrados en el cálculo de la STFT, dejando el valor de α igual a 0.001. Dejando constante el valor la longitud de la ventana en 25 ms, se varía los parámetros de solapamiento y tipo de ventana obteniéndose los resultados para el PSNR y el NC que se muestran en la Tabla 2.

	Solapamiento	Ventana	PSNR	NC	
				Attack 1	Attack2
Audio 1	50%	Rectangular	37	0.880929	0.888891
Audio 1	0%	Rectangular	37	0.879993	0.888891
Audio 1	50%	Hann	52	0.580935	0.942713
Audio 1	0%	Hann	52	0.579451	0.942713
Audio 2	50%	Rectangular	28	0.612997	0.921377
Audio 2	0%	Rectangular	28	0.611813	0.921377
Audio 2	50%	Hann	33	0.473851	0.997849
Audio 2	0%	Hann	33	0.471051	0.997849

Tabla 2- Valores de PSNR y NC para diferentes tipos de ventana y solapamiento.

Se puede ver que el tipo de ventana afecta el PSNR, lo que se traduce en un cambio de mayor o menor medida en cuanto a la perceptibilidad del audio. Analizando los valores de la tabla, se concluye que la utilización de una ventana tipo Hanning es beneficiosa para mantener una buena relación entre la señal y el ruido, es decir, no es tan destructiva para la señal audible. Por otro lado, si se analiza solamente la capacidad de mantener la robustez de la marca de agua frente a diferentes ataques la ventana rectangular consigue resultados más consistentes.

5. CONCLUSIÓN

Con el objetivo de incorporar una marca de agua a dos archivos de audio, se desarrolla un algoritmo que, por medio del uso de la Transformada de Fourier de Tiempo Reducido y la Descomposición de Valores Simples, permite embeber una imagen en dichos archivos, con la posibilidad de extraer de ellos nuevamente la marca de agua. También debe ser posible contar con un parámetro numérico que permita contemplar objetivamente que tan perceptible es la incorporación de la imagen en el archivo de audio, cuando este se reproduce. Para ello se utilizan los parámetros Relación señal pico - Ruido (PSNR), y la Correlación Normalizada (NC).

Se llega a la conclusión de que cuanto mayor robustez tenga la marca de agua, más perceptible resulta en la escucha su implementación, pero a la hora de su extracción, la imagen se conservará de mejor forma, lo cual resulta útil cuando se ataca al archivo de audio.

Por otro lado, los parámetros que definen al ventaneo para realizar la STFT (muestras, solapamiento y tipo de ventana), también influyen en la perceptibilidad de la marca de agua. El efecto de la marca de agua se vuelve más notorio bajo el uso de una ventana de tipo rectangular que con una de tipo Hanning. No se perciben cambios notorios a causa del solapamiento.

6. REFERENCIAS

1- Descomposición de Valores Singulares, 2010. Disponible online: http://www.mate.unlp.edu.ar/practicas/70_18_09112010 12951

2- Zainol, Z., Teh, J. S., Alawida, M., & Alabdulatif, A. (2021). Hybrid SVD-based image watermarking schemes: a review. *IEEE Access*, *9*, 32931-32968.

3- An SVD-Based Audio Watermarking Technique - Hamza Ozer, B¨ulent ¨ Sankur, Nasir Memon - MMSec ’05: Proceedings of the 7th workshop on Multimedia and securityAugust 2005 Pages 51–56.