Hacia una huella digital robusta para imágenes y video

J. Fernandez, N. Miranda, R. Guerrero, F. Piccoli
LIDIC- Universidad Nacional de San Luis
Ejército de los Andes 950
Tel: 02652 420823, San Luis, Argentina
{*jmfer, ncmiran, rag, mpiccoli*}*@unsl.edu.ar*

E. Chávez.
Universidad Michoacana de San Nicolás de Hidalgo
Morelia, México
elchavez@umich.com

Resumen

Consideremos como objetos digitales las modulaciones en el tiempo y el espacio de una señal digital. Nos interesa poder identificar dos variantes coherentes del mismo objeto. Para nuestros fines, en una dimensión el objeto es una señal de audio, en dos una fotografía y en tres dimensiones un video. Consideraremos que el objeto está formado por marcos o arreglos d-dimensionales ($d = 1, 2, 3$) de muestras de la señal. Las versiones de los objetos consisten en perturbaciones coherentes; es decir, sólo se permiten cambios locales en los marcos, no en su disposición. En otras palabras existe una biyección (dada por la función identidad) entre la disposición temporal y espacial de los marcos de ambos objetos a comparar. Aún con esta restricción, el problema resulta interesante para un gran número de aplicaciones. Entre las distorsiones coherentes podemos encontrar, contaminación por ruido, compresión con pérdida, ecualización, regrabación, etc. En este esquema, cada marco del objeto será representado por un vector o un escalar, que deberá ser invariante a la distorsión presentada.

En este trabajo, proponemos una solución al problema planteado, cada marco será representado por la entropía de la señal. Esta característica ha probado su efectividad en señales unidimensionales (de audio) y su generalización a fotografía y video se anticipa también eficiente. El problema es relevante dado el incremento del uso de los contenidos multimediales por el comercio electrónico y los servicios on-line, los problemas asociados con la protección de propiedad intelectual, administración de grandes bases de datos multimedia y la organización de su contenido.

Palabras Claves: Recuperación de Información, huellas digitales robustas, entropía, procesamiento de señales.

1. Introducción

El avance tecnológico ha permitido la digitalización de grandes cantidades de información multimedia, proveniente de fuentes de audio, discos, video, fotografías, etc. El almacenamiento de esta información multimedia no representa problemas tecnológicos (aun cuando se estima en varios zetabytes la información en línea). Por otro lado, la recuperación e identificación de la misma, si se convierte en un problema relevante. Identificamos dos problemas relacionados con la Recuperación de Información Multimedia (RIM). El primero consiste en identificar correctamente dos variantes de un mismo objeto y el segundo en hacerlo rápidamente. En esta primera parte de la investigación, nuestro objetivo será identificar correctamente los objetos bajo diferentes distorsiones.

El concepto de *búsqueda exacta* es central para los repositorios de información o bases de datos "tradicionales". Bajo este supuesto la información es almacenada y recuperada bajo la premisa de que no hay pérdida de información ni al almacenar ni al recuperar los datos. Si almaceno un documento,

espero poder recuperarlo íntegramente con cada caracter inalterado y en su lugar. Si esperáramos lo contrario (que hubiera modificaciones aleatorias en los datos almacenados) tendríamos que cambiar radicalmente nuestras estrategias de almacenamiento y recuperación de la información.

Por otro lado, las bases de datos multimedia almacenan información cuyas características son completamente diferentes a lo discutido para bases de datos tradicionales. No es que los objetos multimedia (canciones, películas, fotografías, etc.) cambien aleatoriamente al almacenarse; de hecho tienen las mismas propiedades de persistencia que los objetos de las bases de datos tradicionales. El comportamiento peculiar proviene al tratar de formalizar nuestra percepción de los objetos multimedia. Por ejemplo, si tomamos dos fotografías digitales de una misma escena bajo las mismas condiciones, al compararlas pixel a pixel observaremos que prácticamente ningún valor coincide. Sin embargo, al ojo humano las dos fotografías serían "iguales". Otro ejemplo representativo de este comportamiento lo encontramos en la digitalización de las señales de audio, los métodos de compresión con pérdida (e.g. con mp3, ogg, etc.) aplican la propiedad de que el cerebro no posee la habilidad para distinguir ciertos sonidos simultáneos. Si comparamos digitalmente una canción comprimida en mp3 y una sin comprimir, difícilmente encontraremos coincidencias digitales; pero si escuchamos las canciones, difícilmente podríamos distinguir la canción comprimida de la canción original. Lo mismo ocurre con el video y sus distintas representaciones (avi, mpeg, mov, etc.).

Cada registro en una base datos tradicional tiene un dato clave que permite recuperar todo el registro. En las bases de datos "full text", como es el caso de la recuperación de información, con cualquier segmento de un texto podemos recuperar el texto completo (un ejemplo son las búsquedas en Internet). Visto de otra manera, con una parte arbitraria del objeto almacenado podemos recuperar el objeto completo. En una base de datos multimedia deberíamos ser capaces de realizar lo mismo: con una parte del objeto almacenado se debería recuperar a todo el objeto. Esto es particularmente difícil si consideramos que los objetos multimedia no tienen una coincidencia digital, comparados bit por bit.

Un modelo estándar de búsqueda en bases de datos multimedia consiste en utilizar una medida de (dis)similaridad entre los objetos almacenados. Esta medida de distancia entre objetos debería modelar esencialmente el comportamiento de una persona al comparar dos objetos de esa naturaleza. Dos objetos *iguales perceptualmente* deberían estar a distancias pequeñas, mientras que dos objetos *perceptualmente distintos* a distancias grandes. Si la distancia está efectivamente diseñada, será posible localizar un objeto en la base de datos comparando el objeto muestra con todos los objetos de la base datos. Esto nos daría un mecanismo *correcto* de recuperación; sin embargo, este mecanismo correcto no es escalable respecto a la base de datos crece. Para hacer escalable la solución es necesario diseñar un índice que permita acceder a los mismas soluciones, sin comparar con todos los elementos de la base de datos. Ese sería un *índice multimedia*.

2. Características Estables de las Señales Digitales

Una característica de un patrón es un invariante, el resultado de la aplicación de una función cuya evaluación aplica objetos de clases iguales a clases iguales. Los clasificadores estadísticos realizan un *dark mapping*; en el sentido de que lo hacen como una caja negra; lo que sucede en el interior es resultado del llamado 'entrenamiento". Este es el caso de las redes neuronales, los modelos ocultos de Markov, las máquinas de soporte vectorial, etc.

Las técnicas estadísticas estándar no pueden ser aplicadas en este caso, dado que nuestra base de datos tendría tantas clases como elementos. Para solucionar el problema de identificación de señales, es necesario diseñar un *ligth mapping*, en donde no exista etapa de entrenamiento y la función que extrae la característica de los objetos pueda ser aplicada a nuevos objetos insertados en la base de datos.

Los objetos multimedia son complejos. Las características perceptibles por las personas en una fotografía son de mas alto nivel que simplemente niveles de brillo en un arreglo de pixels, una interpretación de alto nivel de una imagen es la combinación de la luminancia (la densidad angular

y superficial de luz) y crominancia (información del color: saturación y tinte). Un vídeo puede ser considerado una sucesión de imágenes y su análisis debe contemplar la conformación de arreglos tridimensionales (agregando el tiempo y la localización de los marcos que la componen)[1, 6, 8, 9].

El modelado de los objetos multimedia ha sido abordado de manera tradicional en [4, 5, 14, 16] en donde el objetivo es tener un modelo matemático cuyos parámetros permitan la comparación de los objetos. Dos objetos con parámetros cercanos serán semejantes. Nosotros estamos mas interesados en modelar a los objetos mediante huellas digitales, una aproximación de ello son [3, 13].

Una huella digital provee de un método de identificación de señales basado en su contenido perceptual. Dos objetos multimedia pueden ser identificados como "el mismo objeto" por una persona y pueden no coincidir en ninguno de sus bits (por ejemplo la compresión con pérdida).

Una huella digital extrae las características esenciales de una señal, intenta proveer un método confiable y rápido para la identificación de contenido. Obtener una huella digital significa extraer las características discriminantes de un objeto, identificándolo unívocamente. Dicho objeto puede ser una señal de audio, una imagen, un video, u otro elemento multimedia.

El fin de una huella digital es proveer un método confiable y rápido para la identificación de contenido. Obtener una huella digital significa extraer las características discriminantes de un objeto, identificándolo unívocamente. Dicho objeto puede ser una señal de audio, una imagen, un video, u otro elemento. Como una huella digital representa las características únicas de una señal [7], es habitualmente usada para medir el porcentaje de similitud entre señales. Idealmente la huella digital sería una invariante de la señal, aquellas características intrínsecas, no alteradas por su constante manipulación.

Existen en la bibliografía numerosos intentos de definir huellas digitales a través de marcas de agua, métodos estocásticos, métodos de procesamiento, entre otros. Aunque todos ellos son herramientas poderosas a fines específicos, también poseen inconvenientes asociados con la modificación del contenido del objeto y la seguridad. Una buena alternativa la constituye la entropía asociada a una señal, cualquiera sea: audio, imagen o video.

Nuestra objetivo es generalizar los resultados obtenidos en [7] para el tratamiento de señales de audio y utilizarlos en fotografía y video.

3. Huella Digital y Entropía

Un método de huella digital es generalmente diseñado para tratar con las distorciones naturales (compresión, codificación analógica, entre otros) y ataques maliciosos (adición de logo, distorción geométrica, cortes en la señal, entre otros). Una huella digital debería ser la misma antes y después de las alteraciones sufridas, siempre y cuando los ataques no cambien su contenido.

Una huella digital de una imagen puede ser una descripción global de la imagen o una descripción local de las características claves extraídas. En cambio, en un video, puede ser una descripcion global del video, un conjunto de huellas digitales para todos los frames del video o de sólo los frames claves del video.

La entropía ha sido usada en: señales de audio con ambientes de ruidos como una herramienta de segmentación [12], en la selección del tipo de frame deseado para el análisis de una señal de audio [15], en el umbralado de imágenes [2], en la representación del código de proteínas [10] .

En [11] se relaciona a la entropía con la incertidumbre o sorpresa que existe en cualquier experimento o señal aleatoria. Puede considerarse como la cantidad de "ruido" o "desorden" contenida o liberada por un sistema. Generalmente se considera a la entropía como la cantidad de información que lleva una señal. La medida de la entropía varía en el tiempo.

Por definición, sean v_1,v_2,v_3,...,v_n posibles valores de una muestra en una señal, donde cada v_i posee la probabilidad p_i de que ocurra. Toda la secuencia p_1,p_2,p_3,...,p_n es denominada función de distribución de probabilidad, la suma de todos los p_i da como resultado 1 ($\sum_{i=1}^{n} p_i = 1$).

La información I contenida en cada v_i depende únicamente de su probabilidad de ocurrencia denotada como $I(p_i)$. Un valor con menor probabilidad de ocurrencia posee mayor información que

aquel valor con mayor probabilidad de ocurrencia, esto significa que la máxima información es obtenida cuando no se tiene un conocimiento a priori, es decir mayor incertidumbre y por lo tanto la información es una función monótona decreciente de la probabilidad. La cantidad de información contenida en el valor v_i se define como:

$$I(vi) = ln(\frac{1}{pi}) = -ln(pi)$$

La entropía H es la información esperada en el contenido de una secuencia, esto es: el promedio de todo el contenido de la informacion influenciada por sus probabilidades de ocurrencia:

$$H = \sum_{i=1}^{n} p_i I(p) = -\sum_{i=1}^{n} p_i ln(p_i)$$

Como la entropía de una señal es la medida de cuan impredecible ella es, si la señal es constante, su entropía o impredecibilidad tendrá el valor 0 (cero), entropía mínima. Caso contrario, si la señal tiene una distribución uniforme su entropía es máxima teniendo el valor $ln(n)$.

$$H_{min} = -\sum_{i=1}^{n} p(k)ln(k) = -ln(1) = 0 \text{ para k constante}$$
$$H_{max} = -\sum_{i=1}^{n} \frac{1}{n} ln(\frac{1}{n}) = -ln(\frac{1}{n}) = ln(n)$$

De la experiencia de trabajos en audio, la modulación de la entropía es una herramienta robusta para la caracterización unívoca de las señales digitales unidimensionales y para resolver el problema de la identificación del audio y su recuperación. La idea aplicada en las señales unidimensionales puede ser el punto de partida para el estudio sobre su aplicación en señales bidimensionales como lo son las señales de vídeo y de imágenes.

4. Propuesta

El objetivo de la presente línea de investigación consiste en desarrollar un método robusto para la determinación de huellas digitales de fotografías y videos, que permita administrar bases de datos de contenidos multimediales en forma eficiente.

La huella digital obtenida debe ser una invariante de la representación de la imagen o stream de imágenes (video), resistentes a diferentes degradaciones, tales como: re-grabado, compresión con pérdida, conversiones análogo-digital/digital-análogo, cambio de escalas, desplazamiento de color/matiz.

Es también un punto a considerar el costo computacional involucrado en el proceso de obtención de la invariante, no sólo desde el punto de vista del tiempo implicado, sino también en los recursos requeridos para el procesamiento y transformación de los objetos.

Referencias

[1] V. Bhaskaran and K. Konstantinos. *Image and Video Compression Standards: Algorithms and Architecture*. Kluwer, Boston, Mass, 1997.

[2] C.-I. Chang, Y. Du, J. Wang, S.-M. Guo, and P.D. Thouin. Survey and comparative analysis of entropy and relative entropy thresholding techniques. In *Vision, Image and Signal Processing, IEE Proceedings*, volume 153, pages 837 – 850. IEEE, December 2006.

[3] R.R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Tran. Information Theory*, 38(2):713–718, 1992.

[4] Deng and B. S. Manjunath. Netra-v: Toward an object-based video representation. *IEEE Trans. on Circuits and Systems for Video Technology*, 8:616–627, 1998.

[5] J. Fan, W.G. Aref, A.K. Elmagarmid, M.S. Hacid, M.S. Marzouk, and X. Zhu. Multiview: multilevel video content representation and retrieval. *Journal of Electronic Imaging*, 10(4):895–908, 2001.

[6] R. Gonzalez and R. Woods. *Digital Image Processing, 2nd Edition.* Prentice Hall, 2002.

[7] J.A. Camarena Ibarrola. *Análisis digital de la señal de voz.* PhD thesis, Borrador - Universidad Michoacana de San Nicolás de Hidalgo, México, Agosto 2007.

[8] K.N. Ngan, T. Meier, and D. Chai. *Advanced Video Coding: Principles and Techniques.* McGraw-Hill, New York, 1999.

[9] C.A. Poynton. *A Technical Introduction to Digital Video.* John Wiley & Sons, New York, 1996.

[10] Harlan Robins, Michael Krasnitz, Hagar Barak, and Arnold J. Levine. A relative-entropy algorithm for genomic fingerprinting captures host-phage similarities. *Journal of Bacteriology*, 187(24):8370–8374, 2005.

[11] C. Shannon and W. Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1949.

[12] J.L. Shen, J.W. Hung, and L.S. Lee. Robust entropy-based endpoint detection for speech recognition in noisy environments. In *International Conference on Spoken Language Processing*, 1998.

[13] S. Thiemerta, H. Sahbib, and M. Steinebacha. Using entropy for image and video authentication watermarks. In Ping Wah Wong Edward J. Delp III, editor, *Security, Steganography, and Watermarking of Multimedia Contents VIII*, volume 6072. SPIE-IS T Electronic Imaging, 2006.

[14] Y. Wang, F. Makedon, J. Ford, L. Shen, and D. Goldin. Image and video digital libraries: Generating fuzzy semantic metadata describing spatial relations from images using the r-histogram. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 202–211, 2004.

[15] H. You, Q. Zhu, and A. Alwan. Entropy-based variable frame rate analysis of speech signal and its applications to asr. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2004.

[16] H.J. Zhang, J. Wu, D. Zhong, and S. Smoliar. An integrated system for content-based video retrieval and browsing. *Pattern Reconigtion*, 30:643–658, 1997.