

Extracción de Información a partir de Datos No Estructurados No Textuales

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

1. Contexto

Esta propuesta de trabajo se lleva a cabo dentro de la línea de investigación “Informática Gráfica” del proyecto “Técnicas de Inteligencia Computacional aplicadas a Optimización, Minería de Datos y Coordinación de Agentes Inteligentes”. Dicho proyecto se desarrolla en el marco del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC), de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis.

2. Resumen

La toma de decisiones y la resolución de problemas complejos del mundo real requiere una visión generalizada del problema, para la cual se debe aportar la mayor cantidad de información relevante disponible. La información estructurada ha sido la principal materia prima utilizada en los sistemas computacionales hasta la fecha, pero restringirse sólo a ella conduce, muchas veces, a representar una visión parcial. Hoy comienza a reconocerse que gran parte de la información requerida para la toma de decisiones y la resolución de problemas de índole general proviene de información no estructurada, principalmente aquella almacenada en la forma de audio, imagen y video. En este contexto, el uso de información obtenida a

partir de datos no estructurados no textuales es de vital importancia. Para lograrlo es necesario replantear las etapas bien definidas de los procesos tradicionales, es decir se plantea la necesidad de revisar las etapas de adquisición, procesamiento y análisis de la información contenida en audio, imágenes y/o video. Esta propuesta de trabajo establece los lineamientos a seguir para incorporar información no estructurada en los procesos de resolución de problemas o toma de decisiones.

3. Introducción

Internet ha cambiado profundamente la forma en la que las personas se comunican, negocian y realizan el trabajo diario al tener acceso a infinitud de recursos e información. En los últimos años el uso de la información multimedial se fue convirtiendo en una combinación de facto al momento de la toma de decisiones y/o la resolución de problemas, mayormente de tipo empresarial. No obstante, el éxito de la Web ha originado uno de sus principales inconvenientes: la administración de la heterogeneidad de los datos en la que se encuentra disponible dicha información. Mas aún, la extracción de información para la toma de decisiones a partir de dichos datos, si bien es una tarea aparentemente simple para una persona, no lo es para una computadora, principalmente cuando el tipo de información es no estructurada (texto,

audio, imágenes y video).

Desde el punto de vista humano, el proceso de extracción de información es un proceso inherentemente global; a partir de información escasa, localizada y usualmente de mala calidad, nuestro cerebro puede reconocer los elementos básicos que la componen y sus interrelaciones, y elaborar una experiencia sensorial (cognitiva, visual, auditiva, etc.) coherente y, como resultado de ello, deducir la información necesaria.

Desde un punto de vista computacional, la mayoría de la información existente proviene de datos configurados en forma estructurada, lo cual permite que la misma sea fácilmente interpretada y directamente utilizada por una computadora una vez que la información es accedida desde estructuras clásicas de almacenamiento tales como matrices, registros de bases de datos, entre otros. Sin embargo, este tipo de organización de la información conduce a representar una visión parcial del problema y dejar fuera de consideración información que podría ser de gran importancia para la resolución efectiva del mismo.

En Ciencias de la Computación existen diversas áreas que intentan abordar el tratamiento de información no estructurada no textual, como es el caso de las áreas de procesamiento de señales, visión por computadora, minería de imágenes. Para el reconocimiento de un objeto/elemento cualquiera (visual o auditivo), es necesario previamente el entendimiento del escenario total en el cual éste se encuentra inmerso; es decir, implica la coordinación de muchas tareas diferentes entre las cuales se pueden mencionar: estimar las características del objeto/elemento (apariencia, forma, orientación, etc), reconocer el objeto/elemento (individualización involucrando la detección de objetos/elementos ocultos), categorizar el/los objetos/elementos reconocidos así como también el escenario total.

En los comienzos, la mayoría de las investigaciones han intentado reconocer un objeto/elemento en forma aislada mediante información local asociada al mismo, asumiendo que toda la información relevante de un

objeto/elemento se encuentra contenida dentro de una pequeña ventana en un espacio visual y/o auditivo [9, 10, 15, 23, 26, 27, 29]. Los errores típicos cometidos por dichos sistemas es que actúan como cámaras oscuras identificando objetos/elementos en forma individual sin considerar el escenario en el cual éste se encuentra inmerso. Es importante entonces, el poder aproximar el comportamiento humano observando fuera de la cámara oscura y considerando el escenario en su totalidad. En [9, 12, 14, 17, 21, 22] queda claramente bosquejado que la comprensión de un escenario es una tarea de reconocimiento global que va más allá de la simple identificación de los objetos/elementos interactuantes; debe además poder reconocer la interacción entre ellos.

Como ejemplos del uso de información no estructurada no textual se encuentran diversas aplicaciones en *reconocimiento de patrones musicales* [4, 16, 7, 8], *cartografía digital* [1, 2, 11, 13], *reconocimiento de rostros* [24, 28, 30], *conducción autónoma de robots* [9, 10, 18, 20, 25], *tours de realidad virtual* [3, 5, 6, 17, 26], *caminatas (walkthrough) en juegos de computadoras* [19, 23], entre otros.

4. Líneas de Investigación y Desarrollo

Trabajar con información no estructurada no textual implica trabajar con diferentes tipos de datos y en consecuencia con diferente complejidad intrínseca. Tal como se mencionó anteriormente, las fuentes de información no estructurada no textual pueden ser señales de audio, imágenes y/o videos; cada una de ellas en forma individual o en conjunto habilitarán diferentes líneas de investigación, algunas de ellas dependientes entre si. Entre las líneas propuestas a seguir se encuentran:

- *Identificación y representación de los aspectos perceptuales.* Básicamente asociados a las imágenes y video, esta línea pretende extraer la mayor cantidad de información perceptual posible. Usualmente las investigaciones se han concentrado

en métodos de segmentación y clustering tradicionales. No obstante, es necesario incorporar nuevos criterios perceptuales adecuados (sin pérdida de información) que permitan la representación de las imágenes no sólo considerando propiedades de colorimetría y luminancia, sino también en lo relativo a la identificación de los objetos que se encuentran en el escenario plasmado y su ubicación espacial relativa (profundidad 3D).

La correcta identificación y categorización de los objetos en forma semántica junto a la correcta reproducción de los lineamientos teóricos establecidos para la simulación del procesamiento realizado por el sistema visual humano, permitirá la reconstrucción del escenario completo capturado en una imagen o una colección de ellas.

- *Representación robusta de la información.* Identificar e individualizar una señal a pesar de las distorsiones naturales (compresión, codificación analógica, ruido, entre otros) y ataques maliciosos (adición de logo, distorsión geométrica, cortes en la señal, cambio en la colorimetría, entre otros) en forma eficiente conlleva a desarrollar un método robusto para la determinación de huellas digitales de señales (imagen, streamings de audio o video).

Para obtener una buena huella digital de una señal es necesario identificar como iguales dos variantes de un mismo elemento, lo cual implica el redefinir la manera en la cual se los ha venido identificando. Una alternativa es la inclusión del contenido perceptual, ya que dos elementos pueden ser identificados como “el mismo” por una persona y pueden no coincidir en su representación interna.

Idealmente la huella digital debe ser una invariante de la señal; aquellas características intrínsecas, no alteradas por su constante manipulación. Una huella digital de una imagen puede ser una des-

cripción global de la imagen o una descripción local de las características claves extraídas. En cambio, en un video, puede ser una descripción global del video, un conjunto de huellas digitales para todos los frames del video o de sólo aquellos claves del video. El mismo concepto se aplica a las señales de audio.

La determinación de una huella digital permitirá por ejemplo identificar objetos de audio a partir de segmentos, la detección de duplicaciones, plagios, el rotulado automático (MP3 modernos), la consulta por ejemplos y filtrados en redes p2p, entre otros. En el caso de los videos, a través de la huella digital se podrá identificar objetos o comportamientos específicos durante las transmisiones de canales de TV, o en bases de datos de video compartidas (Google, You Tube), seguridad a través de circuitos cerrados de video, derechos de autor, entre otros.

- *Recuperación de Información no estructurada:* El concepto de búsqueda exacta es central para los repositorios de información o bases de datos “tradicionales”. Trasladar el mismo concepto a repositorios de señales, no es simple. Por la naturaleza de los datos, el objetivo será resolver computacionalmente la similitud perceptual entre objetos similares. Los desarrollos actuales utilizan medidas de distancia, a través de las cuales se pretende reflejar la similitud perceptual. Dos objetos iguales perceptualmente deberían recibir distancias pequeñas; mientras que dos objetos perceptualmente distintos deberían recibir distancias grandes. Para ello es necesario contar con una representación robusta de cada dato del repositorio y definir una efectiva función de distancia, la cual permita establecer una valoración sobre la similitud o dis-similitud de dos objetos y relajar el concepto de búsqueda exacta por búsqueda por proximidad.

5. Resultados obtenidos / esperados

El principal aporte de esta propuesta será la incorporación de información no estructurada en los procesos de toma de decisiones y resolución de problemas que, normalmente, queda fuera de consideración en los enfoques clásicos.

Actualmente se está trabajando en la obtención de una huella digital robusta y eficiente para audio e imágenes así como también en la identificación y recuperación de información perceptual de profundidad. La eficiencia no sólo se relaciona con la capacidad de representar unívocamente una señal, sino con el tiempo que implica su cálculo. Para ello se están desarrollando algoritmos de procesamiento de imágenes, los cuales aplican técnicas de programación de alta performance.

6. Formación de Recursos Humanos

Como resultado de las investigaciones se cuenta con una tesis de maestría concluída y dos tesis doctorales y una de maestría en desarrollo; así como también varios trabajos de fin de carrera de la Licenciatura en Ciencias de la Computación.

Además las investigaciones se encuadran en el marco de un proyecto dentro del Programa de Promoción de la Universidad Argentina para el Fortalecimiento de Redes Interuniversitarias III en los que participa nuestra universidad junto con las universidades Michoacana (México) y de Zaragoza (España).

Referencias

[1] Armitage, R.P., Weaver, R. E., and Kent, M. "Remote sensing of semi-natural upland vegetation: the relationship between species composition and spectral response". In: *Vegetation Mapping: From Patch to Planet* (eds Millington, A. and Alexander, R.), John Wiley and sons (2000).

[2] Blaschke, T. and Hay, G., "Object-oriented image analysis and scale-space: Theory and methods for modeling and evaluating multi-scale landscape structure". *International Archives of Photogrammetry and Remote Sensing* 34: 22-29 (2001).

[3] A.L. Callahan and Dmitry and B. Goldgof and Ph. D and Ph. D and Thomas A. Sanocki and Melanie A. Sutton, "Function from visual analysis and physical interaction: a methodology for recognition of generic classes of objects", *Journal on Image and Vision Computing*, Vol 16, pp 745-763, 1998.

[4] A. Camarena-Ibarrola, E. Chavez, "On Musical Performances Identification, Entropy and String Matching", *MICAI 2006*.

[5] P. Carbonetto, N. de Freitas, and K. Barnard, "A statistical model for general contextual object recognition", in *Proc. ECCV*, 2004.

[6] (CUDA)-IEEE International Conference on Multimedia and Expo 2008 ? Pp 697:700 ? April 2008.

[7] Crockett, et al., "A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis", *AES Convention Paper 6416*, presented at the 118.sup.th Convention May 28-32, 2005, Barcelona, Spain. cited by other.

[8] Dixon, S.: Live tracking of musical performances using on-line time warping. *Proc of the 8th Int Conf on Digital Audio Effects (DAFx'05)* (2005)

[9] A. Ess, B. Leibe, and L. Van Gool. "Depth and appearance for mobile scene analysis". In *ICCV*, 2007.

[10] R. Fergus, P. Perona, and A. Zisserman. "Object class recognition by unsupervised scale-invariant learning". In *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[11] M.A. Fischler and R.C. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography". *Communications of the ACM*, 24(6):381-395, June 1981.

- [12] G. Kim, C. Faloutsos, and M. Hebert, "Unsupervised Modeling and Recognition of Object Categories with Combination of Visual Contents and Geometric Similarity Links", ACM International Conference on Multimedia Information Retrieval (ACM MIR), October, 2008.
- [13] Hay, G.J., Marceau, D. J., Dube, P., and Bouchard, A. "A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling". *Landscape Ecology* 16 (6): 471-490 (2001).
- [14] Hoiem, A.A. Efros, and M. Hebert, "Closing the Loop on Scene Interpretation", In CVPR 2008.
- [15] A. Hoiem, A. Efros, and M. Hebert. "Geometric context from a single image". In ICCV, 2005.
- [16] Ibarrola, A.C., Chavez, E.. "A robust, entropy-based audiofingerprint". IEEE, July 2006.
- [17] B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. "Dynamic 3d scene analysis from a moving vehicle". In CVPR, 2007.
- [18] Liu Yang, Rong Jin, Caroline Pantofaru, and Rahul Sukthankar, "Discriminative Cluster Refinement: Improving Object Category Recognition Given Limited Training Data", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 2007.
- [19] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 6, June 2002, pp. 780-788.
- [20] Pech-Pacheco, José L., Álvarez-Borrego, Josué, Cristóbal, Gabriel, Keil, Matthias S., "Automatic object identification irrespective of geometric changes", International Society for Optical Engineering (SPIE), Vol 42(2): 551-559 (2003).
- [21] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. "Objects in context". In ICCV, 2007.
- [22] A. Saxena, M. Sun, and A. Y. Ng. "Learning 3-d scene structure from a single still image". In ICCV 3dRR-07, 2007.
- [23] H. Schneiderman, "Learning a restricted bayesian network for object detection," in Proc. CVPR, 2004.
- [24] Torres M., "Is there any hope for face recognition?", Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2004, 21-23 April 2004, Lisboa, Portugal.
- [25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 6, June, 2007, pp. 929-944.
- [26] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in Proc. ICCV, 2003.
- [27] P. Viola and M. J. Jones, "Robust real-time face detection," IJCV, vol. 57, no. 2, 2004.
- [28] Xiaoxing Li; Mori, G.; Hao Zhang, Expression-Invariant Face Recognition with Expression Classification, The 3rd Canadian Conference on Computer and Robot Vision, Volume , Issue , 07-09 June 2006 Page(s): 77 ? 77, Digital Object Identifier:10.1FER03/CRV.2006.34. 2006.
- [29] W. Zhang, B. Yu, D. Samaras, and G. Zelinsky. "Object class recognition using multiple layer boosting with heterogeneous features". In IEEE Conference on Computer Vision and Pattern Recognition, 2005.
- [30] W.Y. Zhao, R. Chellappa, "Image-based Face Recognition: Issues and Methods, Image Recognition and Classification", Ed. B. Javidi, M. Dekker, 2002, pp. 375-402.