AusArt Journal for Research in Art. 4 (2016), 1, pp. 19-28
www.ehu.es/ojs/index.php/ausart
DOI: 10.1387/ausart.16670

ISSN 2340-8510
e-ISSN 2340-9134
UPV/EHU

DEL PÍXEL A LAS RESONANCIAS VISUALES: LA IMAGEN CON VOZ PROPIA

Pilar Rosado Rodrigo
Universitat de Barcelona, Dpto. Escultura
Eva Figueras Ferrer
Universitat de Barcelona, Dpto. Pintura
Ferrán Reverter Comes
Universitat de Barcelona, Dpto. Estadística

Resumen

Esta investigación aborda el problema de la detección de aspectos latentes en grandes colecciones de imágenes de obras de artista abstractas, atendiendo sólo a su contenido visual. Se ha programado un algoritmo de descripción de imágenes utilizado en visión artificial cuyo enfoque consiste en colocar una malla regular de puntos de interés en la imagen y seleccionar alrededor de cada uno de sus nodos una región de píxeles para la que se calcula un descriptor que tiene en cuenta los gradientes de grises encontrados. Los descriptores de toda la colección de imágenes se pueden agrupar en función de su similitud y cada grupo resultante pasará a determinar lo que llamamos *"*palabras visuales*"*. El método se denomina Bag-of-Words (bolsa de palabras). Teniendo en cuenta la frecuencia con que cada *"*palabra visual*"* ocurre en cada imagen, aplicamos el modelo estadístico pLSA (Probabilistic Latent Semantic Analysis), que clasificará de forma totalmente automática las imágenes según su categoría formal. Esta herramienta resulta de utilidad tanto en el análisis de obras de arte como en la producción artística.

Palabras-clave: VISIÓN ARTIFICIAL; MODELO BAG-OF-WORDS; CBIR (RECUPERACIÓN DE IMÁGENES POR CONTENIDO); PLSA (ANÁLISIS PROBABILÍSTICO DE ASPECTOS LATENTES); PALABRA VISUAL

FROM PIXEL TO VISUAL RESONANCES: IMAGES WITH VOICE

Abstract

The objective of our research is to develop a series of computer vision programs to search for analogies in large datasets—in this case, collections of images of abstract paintings—based solely on their visual content without textual annotation. We have programmed an algorithm based on a specific model of image description used in computer vision. This approach involves placing a regular grid over the image and selecting a pixel region around each node. Dense features computed over this regular grid with overlapping patches are used to represent the images. Analysing the distances between the whole set of image descriptors we are able to group them according to their similarity and each resulting group will determines what we call "visual words". This model is called Bag-of-Words representation Given the frequency with which each visual word occurs in each image, we apply the method pLSA (Probabilistic Latent Semantic Analysis), a statistical model that classifies fully automatically, without any textual annotation, images according to their formal patterns. In this way, the researchers hope to develop a tool both for producing and analysing works of art.

Keywords: ARTIFICIAL VISIÓN; BAG-OF-WORDS MODEL; CBIR (CONTENT-BASED IMAGE RETRIEVAL); PLSA (PROBABILISTIC LATENT SEMANTIC ANALYSIS); VISUAL WORD

Rosado Rodrigo, Pilar, Eva Figueras Ferrer & Ferrán Reverter Comes. 2016. "Del píxel a las resonancias visuales: La imagen con voz propia". *AusArt* 4(1): pp-pp. 19-28 DOI: 10.1387/ausart.16670

1. INTRODUCCIÓN

Flusser habla del tiempo circular de la magia en las imágenes y del tiempo lineal de la historia en los escritos:

> *"Las imágenes son superficies con significado. Normalmente señalan algo ubicado afuera en el espacio-tiempo, que han de hacer concebible en forma de abstracciones (reducciones de las cuatro dimensiones de espacio y tiempo a las dos de la superficie). Esta capacidad específica de abstraer superficies del espacio-tiempo y de reproyectarlas al espacio-tiempo la llamaremos imaginación. Ella es indispensable para la generación y el descifra-miento de imágenes; o, dicho de otro modo: para la capacidad de cifrar fenómenos en símbolos bidimensionales y de leer esos símbolos. El significado de las imágenes se encuentra en su superficie. Se aprehende con una sola mirada, si bien así permanece superficial. Si nos proponemos profundizar en el significado, es decir, reconstruir las dimensiones abstraídas, tendremos que pasear la mirada por la superficie, dejar que la explore. Esta exploración de la superficie de la imagen con la mirada la llamaremos escaneo. Al escanear, la mirada sigue un rumbo complejo marcado, por una parte, por la estructura de la imagen y, por otra, por las intenciones del contemplador. ...Este espacio-tiempo propio de la imagen no es otra cosa que el mundo de la magia, un mundo en el que todo se repite y todo participa de un contexto significativo . Este mundo se distingue estructuralmente de la linealidad histórica, en la que nada se repite y todo tiene causas y acarrea consecuencias"* (2009, 11-12).

Visualizar es la capacidad de formar imágenes mentales. El pensamiento en conceptos probablemente surgió del pensamiento en imágenes y mediante la abstracción permitió la simbolización y la escritura fonética. La evolución del lenguaje comenzó en las imágenes pero hoy en día existen numerosos indicios de que es necesario un retorno hacia la imagen, en el sentido de que es importante encontrar analogías con el lenguaje que puedan aplicarse a la información visual. Esto no es nada fácil dado que para conocer el significado de las palabras es necesario conocer las definiciones comunes que comparten, y este proceso trasladado a las imágenes corre el peligro de la sobredefinición. Pero si ha sido posible descomponer el lenguaje en elementos y estructuras ¿sería posible hacerlo también con las imágenes?

En el presente capítulo proponemos realizar esta aproximación a través de las matemáticas y aplicar un método extraído del análisis automático de textos que es capaz de representar, como veremos más adelante, el contenido de las imágenes en unidades discretas de información denominadas *palabras visuales.*

A raíz de la gran cantidad de imágenes que producimos y almacenamos en la actualidad, se están invirtiendo esfuerzos considerables dirigidos al diseño de sistemas automáticos que permitan la recuperación de imágenes basada en su contenido visual (*Content-Based Image Retrieval, CBIR*) y no en las anotaciones textuales que contengan. Forman parte de la comunidad CBIR personas de diferentes ámbitos; la visión por computador, el aprendizaje máquina, la recuperación de información, la minería de datos, la estadística, la psicología, etc. Los resultados obtenidos en la interpretación de imágenes por una máquina de escenas de la vida real son prometedores; sin necesidad de ninguna indicación externa pueden clasificar un gran conjunto de instantáneas en distintos grupos atendiendo a las características formales que posean; pueden agrupar por un lado las imágenes que contengan paisajes de costa, por otro las escenas urbanas, las que representan rostros, etc. Nuestro equipo de investigación trabaja en la aplicación de algunos de estos modelos de visión computacional sobre colecciones de imágenes abstractas que son utilizadas por artistas como base para la ideación de sus trabajos, obteniendo excelentes resultados de categorización (Reverter, Figueras, Planas & Rosado 2013; Rosado, Reverter, Figueras & Planas 2014; Rosado, Figueras & Reverter 2014).

El punto de partida de nuestro análisis es el convencimiento de que entre las obras de las colecciones de imágenes de artistas existen unos lazos de unión, de parentescos formales que hacen que constituyan una familia de significado común. El artista visual hace uso de sus categorías formales para capturar desde lo particular aquello universalmente significativo, de una forma necesariamente personal.

Kandinsky (1987), además de ser un artista apasionado, siempre se mostró deseoso de explicar las razones primeras y profundas de la creación artística y realizó aportaciones fundamentales que, además de contribuir a esclarecer el análisis de los elementos esenciales del quehacer pictórico, contribuyeron

a la búsqueda de un método genérico para las investigaciones de las ciencias artísticas.

> *"También existe la posibilidad de penetrar en la obra, participar en ella y vivir sus pulsaciones con sentido pleno. Y aunque no se tenga en cuenta su valor científico, que depende de un minucioso examen, el análisis de los elementos artísticos es un puente hacia la pulsación interior de la obra de arte"* (Kandinsky 1996, 15-6).

La oportunidad que nos brinda la imagen digital de describir las formas en términos matemáticos pone a nuestro alcance la posibilidad de descifrar la sintaxis visual, el problema del significado contenido en la imagen.

2. CONSTRUCCIÓN DEL VOCABULARIO VISUAL

La inteligencia artificial constituye una herramienta fundamental para la extracción automática de conocimiento. En el ámbito concreto de la visión artificial en el que se intenta que un ordenador "interprete" una imagen, los investigadores se enfrentan a dos grandes problemas: en primer lugar a las limitaciones que supone registrar la información que contienen las imágenes en un código abstracto, en segundo lugar a la dificultad de elaborar interpretaciones a partir de las imágenes. Para superar estos inconvenientes se han creado multitud de metodologías y se evalúan sus rendimientos.

En nuestra investigación hemos trabajado con colecciones de imágenes de artistas de contenido abstracto. Se ha programado un algoritmo basado en un modelo concreto de descripción de imágenes utilizado en visión artificial cuyo enfoque consiste en colocar una malla regular de puntos de interés en la imagen y seleccionar alrededor de cada uno de sus nodos una región de píxeles para la que se calcula un descriptor que tiene en cuenta los gradientes de grises encontrados (Lowe 2000, 20-31; 2004, 91-110).

Analizando las distancias entre el conjunto de descriptores de toda la colección de imágenes, se pueden agrupar en función de su similitud y estos grupos resultantes pasarán a determinar lo que llamamos "palabras visuales". El total de "palabras visuales" de una colección de imágenes genera un "vocabulario visual" concreto del conjunto. El método se denomina Bag-of-Words (bolsa de

palabras) porque representa una imagen como una colección desordenada de características visuales locales (Lazebnik, Schmid & Ponce 2006).

Teniendo en cuenta la frecuencia con que cada palabra visual ocurre en cada imagen, aplicamos el método pLSA (Probabilistic Latent Semantic Analysis), un modelo estadístico que clasificará de forma totalmente automática, sin anotación textual alguna, las imágenes según su categoría formal (Hofmann 2001, 177-96). Puede verse un esquema resumido de este método en la Figura 1.

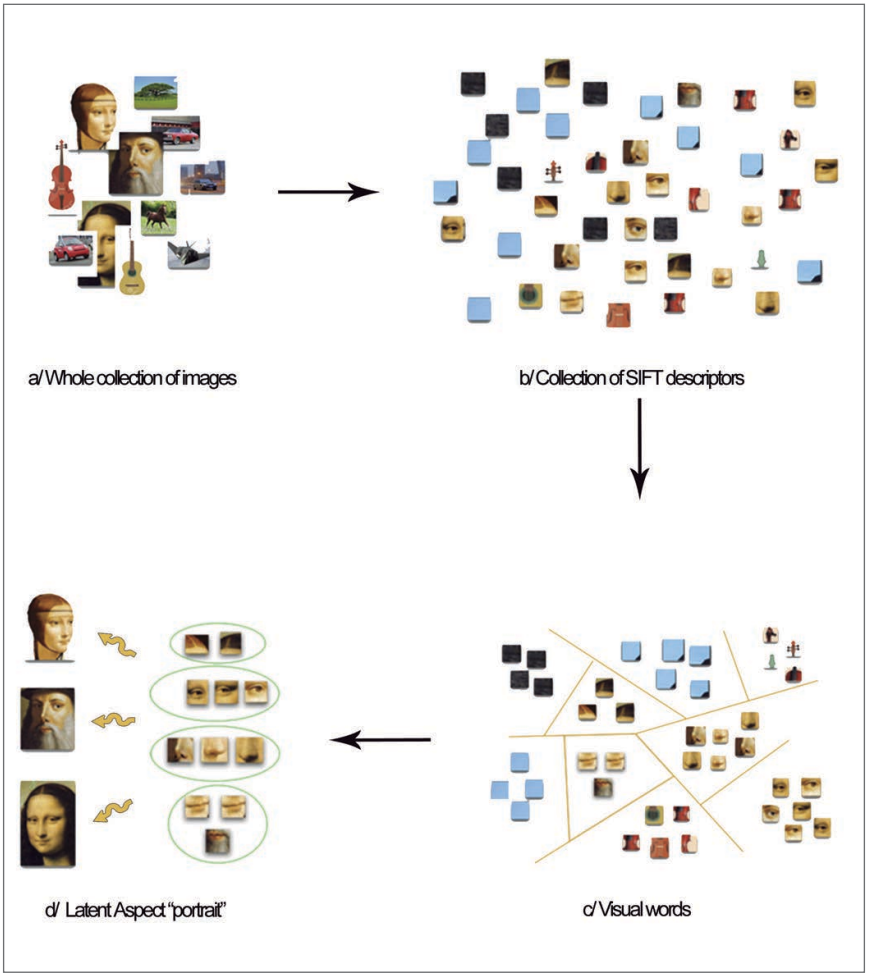

Fig.1: *Esquema de la metodología de visión artificial utilizada para la categorización de imágenes y la construcción del vocabulario visual. 1. Colección de imágenes. 2. Extracción de descriptores de pequeñas regiones de las imágenes. 3. Cuantización de las palabras visuales 4. Detección de aspectos latentes por co-ocurrencia de determinadas palabras visuales.*

3. RESULTADOS

Para comentar los resultados de agrupación automática obtenidos vale la pena comentar la experiencia de análisis de las imágenes presentadas por grupos; cuando aparecen ante nuestra vista por primera vez, pueden dejarnos expectantes. Es conveniente en primer lugar haber visualizado todos los grupos para, en adelante, teniendo ya conciencia del resultado total, percibir de inmediato que existen unos ritmos, unas características comunes a cada grupo. No obstante, si se fija la atención en concreto en alguna de las imágenes o si se intentan traducir a palabras textuales las cualidades generales percibidas en cada clase, la percepción de conjunto se desvanece. Se podría utilizar un paralelismo referido a la profundidad de campo en una cámara fotográfica; si el objetivo enfoca un punto, se desenfoca el resto, para tener una visión de conjunto es necesario mirarlas todas a la vez, y de esta forma sí, existe una intuición estética, una analogía de sentido que las atrae entre sí y las vincula en una misma categoría.

Cada conjunto resultante, en sí mismo, forma un corpus contenedor de un sentido que se pierde cuando se focaliza la atención en una única obra. Si un artista sólo tuviese ocasión de realizar un cuadro perderíamos la ocasión de percibir los potenciales valores contenidos en sus tanteos, exploraciones y decisiones. Es el conjunto de la obra de una artista consolidado el que nos abre la puerta a la visión de su núcleo íntimo de percepción. Y, a pesar de que toda clasificación por definición es subjetiva ya que se realiza según una determinada directriz, cuando podemos colocar unas obras próximas a otras en nuestro campo visual adquieren juntas un sentido que no poseen por separado.

Apelando a la teoría de la Gestalt, el principio básico de la organización perceptual es que el todo es más que la suma de las partes, es decir, que las propiedades de la totalidad no resultan de los elementos constituyentes, sino que emergen de las relaciones espacio-temporales del todo. (Köhler 1947; Koffka 1967).

Hemos aplicado el modelo descrito sobre colecciones de artista constituidas por imágenes digitales de escenas naturales de carácter abstracto (Rosado et al., 2014) obteniendo categorizaciones significativas. El reto al que nos enfrentamos, a diferencia de los estudios encontrados en la literatura que abordan la agrupación automática por contenido visual de imágenes de escenas reales y objetos cotidianos, es que aquellas tienen un contenido semántico universal-

mente asumido y en cambio, las bases de datos de arte abstracto que utilizamos en nuestro análisis son colecciones de imágenes de formas que el artista creador vincula porque considera que entre ellas existen analogías de sentido, y que por tanto suponen un reto de más difícil validación. Aclaramos que con el termino “abstracto” nos referimos al arte que no intenta imitar un modelo conocido, o sea, “no objetivo”.

En la Fig. 2 se muestran dos de las trece agrupaciones resultantes de aplicar la metodología descrita de representación de la imagen mediante Bag-of-Words, sobre el conjunto de obra de Antoni Tàpies (2001[1]). Observamos en el primer grupo de la izquierda un conjunto de imágenes de obras que transmiten idea de orden, composición y equilibrio. En bastantes de ellas aparecen figuras rectangulares o cuadradas bien definidas. Transmite también la sensación de simetría. Muy diferentes de las que se muestran en el grupo de la derecha que presentan trazos gruesos más enérgicos y gestuales.

Fig.7: *A la izquierda el aspecto: Equilibrio Compositivo y a la derecha el aspecto Trazo Grueso Denso. © Fundació Antoni Tàpies, Barcelona / Vegap. De la fotografia: © Gasull Fotografia.*

En la figura 3 se muestran 3 de las 300 palabras visuales utilizadas para realizar la categorización de la obra de Tàpies. Cada imagen corresponde a un grupo de pequeñas regiones pertenecientes a diferentes imágenes de la colección cuya similitud, en base a la distancia matemática que las separa, hace que el sistema las categorice como pertenecientes a la misma palabra visual.

Cada una de las palabras muestran conjuntos de pequeñas regiones que se corresponden con la zona de 16 x 16 píxeles alrededor del nodo que ha sido utilizado para calcular el descriptor de la zona. Estas pequeñas regiones de

las imágenes, al ser visualizadas, contribuyen a la comprensión de las características formales del aspecto al que corresponden. Al mirarlas agrupadas se perciben las constantes que han motivado el agrupamiento en la misma "palabra visual". La posibilidad de visualizar el vocabulario particular que utiliza una artista plástico al ejecutar sus obras y a la vez de medir la frecuencia del uso de unas palabras sobre otras, resulta muy significativo y de utilidad para la comprensión y el estudio de su producción. A su vez, la posibilidad de configurar un vocabulario visual complejo más amplio, compuesto por palabras de diversos artistas, es muy sugerente y sería también de gran utilidad como fondo para la creación digital de nuevas posibilidades estéticas.

Fig.8: *Palabras visuales constituidas por fragmentos de diferentes imágenes de la colección Tàpies.*

4. CONCLUSIONES

Las imágenes agrupadas presentan un tipo de conocimiento que resulta difícil de explicar de otro modo, especialmente complicado con palabras, pero que al presentarlo a la vista en su totalidad, proporciona una comprensión inmediata de sentido. De esta forma el lector tienen a su alcance una información que, tratando los mismos objetos de forma aislada, sería inaccesible. Este aspecto es especialmente valioso cuando se trata de imágenes de contenido abstracto, dado que en ellas el tema, el significado o el sentido no es producto de un acuerdo social, sino que se trata de resonancias visuales y sincronías que el artista creador relaciona y vincula desde su interior. Las nuevas tecnologías permiten la digitalización de contenidos visuales que de otro modo permanecerían guardados en cajones o en almacenes, pero dada la enorme cantidad de información, son cada vez más necesarias herramientas capaces de establecer relaciones genuinas entre imágenes, de propio contenido visual, que no requieran anotación textual alguna y no vengan condicionadas por los

conocimientos previos de quien realice la selección. La resolución de la histórica dialéctica entre textos e imágenes podría provenir de estos métodos que permiten a las imágenes hablar con su propia voz.

Los resultados obtenidos son considerados satisfactorios por expertos en arte y, lejos de pretender substituir el criterio de los entendidos, el sistema programado propone una herramienta de estudio para establecer analogías y buscar aspectos latentes en grandes colecciones de imágenes de arte abstracto, aunque también sería extensible el uso en obra figurativa. El sistema permite repetir los estudios sobre diferentes periodos del mismo artista, o sobre colecciones de distintos artistas o épocas, con los mismos criterios. De esta forma, los resultados obtenidos se pueden comparar sin riesgo de caer en interpretaciones subjetivas condicionadas por las preferencias o conocimientos previos.

Cabe destacar el interés de las herramientas presentadas desde el punto de vista del acceso simultáneo por parte de un artista a su colección de múltiples imágenes para poder analizar su trayectoria creativa, o desde el punto de vista de los teóricos del arte que podrían realizar estudios comparativos entre las obras de arte de diferentes artistas o épocas sin necesidad de mover de su emplazamiento ni una obra. Sin entrar a valorar la calidad estética de las agrupaciones que realiza la maquina, podemos concluir que las relaciones que establece, dada la cualidad matemática que le confiere la metodología utilizada para su realización, proporcionan nuevos puntos de vista libres de preconcepciones historicistas o vivenciales.

Referencias

Flusser, Vilém. 2009. Una filosofía de la fotografía. Traducción, Thomas Schilling. El Espíritu y la Letra 5. Madrid: Síntesis

Hofmann, Thomas. 2001. “Unsupervised learning by probabilistic latent semantic analysis”. *Machine Learning* 42

Kandinsky, Vasili Vasilievich (1912) 1987. *La gramática de la creación. El futuro de la pintura*. Ed. y notas de Philippe Sers. Barcelona: Paidós

— (1926) 1996. *Punto y línea sobre el plano: Contribución al análisis de los elementos pictóricos.* Traducción Roberto Echavarren. Barcelona: Paidós

Koffka, Kurt. (1935). 2014. *Principles of Gestalt Psychology.* Milano: Mimesis International

Köhler, Wolfgang. (1947) 1992. *Gestalt psychology: an introduction to new concepts in modern psychology.* New York: Liveright

Lazebnik, Svetlana, Cordelia Schmid & Jean Ponce. 2006. “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories”. *IEEE Computer Society Con-*

ference on Computer Vision and Pattern Recognition 2: 2169-78. Doi: doi.ieeecomputersociety.org/10.1109/CVPR.2006.68

Lowe, David G. 2000. "Towards a computational model for object recognition in IT cortex". En *Biologically motivated computer vision: First IEEE International Workshop, BMCV 2000 Seoul, Korea, May 15-17: Proceedings*, Seong-Whan Lee Heinrich H. Bülthoff & Tomaso Poggio, eds., 20-31. Berlin: Springer

Lowe, David G. 2004. "Distinctive image features from scale invariant keypoints". *International Journal of Computer Vision* 60(2): 91-110

Reverter Comes, Ferrán, Eva Figueras Ferrer, Miquel Planas-Rosselló & Pilar Rosado Rodrigo. 2013. *Ideación y catalogación artística basada en métodos de visión artificial*. Barcelona: Raima

Rosado-Rodrigo, Pilar, Ferrán Reverter Comes, Eva Figueras Ferrer & Miquel Planas Rosselló. 2014. "Semantic-based image analysis with the goal of assisting artistic creation". *Lecture Notes in Computer Science* 8671: 526-33. Doi: 10.1007/978-3-319-11331-9

Rosado-Rodrigo, Pilar, Eva Figueras Ferrer & Ferrán Reverter Comes. 2014. "Intersecciones entre visión artificial y mirada artística". *BRAC* 2(1): 1-54. Doi:10.4471/brac.2014.01

Notas

[1] Fundació Antoni Tàpies. Web oficial de la Fundación. Acceso 12 enero 2015. http://www.fundaciotapies.org. Agradecemos al Archivo de la Fundación Antoni Tàpies de Barcelona la posibilidad de acceder a la colección de imágenes digitales de obra del artista. Queda totalmente prohibida cualquier forma de reproducción, distribución, comunicación pública o transformación total o parcial de las imágenes sin el permiso escrito de los titulares de explotación.

(Artículo recibido 11.04.16; aceptado 19.05.16)