

# La visión artificial, un nuevo aliado para el análisis de imágenes artísticas

## Computer vision, a new ally for the analysis of artistic images

PILAR ROSADO-RODRÍGO  
Universidad de Barcelona  
prforma@gmail.com

EVA FIGUERAS-FERRER  
Universidad de Barcelona  
efigueras@ub.edu

Miquel Planas-Roselló  
Universidad de Barcelona  
miquelplanas@gmail.com

Ferran Reverter-Comes  
Universidad de Barcelona  
freverter@ub.edu

Recibido: 8 de abril de 2015

Aprobado: 8 de julio de 2015

### Resumen

En este estudio se evalúa el rendimiento de los métodos de *Bag-of-Visualterms* (BOV) para la clasificación automática de imágenes digitales de la base de datos del artista Miquel Planas. Estas imágenes intervienen en la ideación y diseño de su producción escultórica. Constituye un interesante desafío dada la dificultad de la categorización de escenas cuando éstas difieren más por los contenidos semánticos que por los objetos que contienen. Hemos empleado un método de reconocimiento basado en *Kernels* introducido por Lazebnik, Schmid y Ponce en 2006. Los resultados son prometedores, en promedio, la puntuación del rendimiento es aproximadamente del 70%. Los experimentos sugieren que la categorización automática de imágenes basada en métodos de visión artificial puede proporcionar principios objetivos en la catalogación de imágenes y que los resultados obtenidos pueden ser aplicados en diferentes campos de la creación artística. **Palabras clave:** visión artificial, descriptores SIFT, *Bag-of-Visualterms*, ideación artística.

Rosado-Rodrigo, P., Figueras-Ferrer, E., Planas-Roselló, M., Reverter-Comes, F. (2016): La visión artificial, un nuevo aliado para el análisis de imágenes artísticas. *Arte, Individuo y Sociedad*, 28(2) 325-338

### Abstract

This paper analyzes the automatic classification of scenes that are the basis of the ideation and the designing of the sculptural production of the sculptor Miquel Planas. The main purpose is to evaluate the performance of the Bag-of-Features methods, in the challenging task of categorizing scenes when scenes differ in semantics rather than the objects they contain. We have employed a kernel-based recognition method that works by computing rough geometric correspondence on a global scale using the pyramid matching scheme introduced by Lazebnik, Schmid and Ponce in 2006. Results are promising, on average the score is about 70%. Experiments suggest that the automatic categorization of images based on computer vision methods can provide objective principles in cataloging images. **Keywords:** artificial vision, Bag-of-visualterms, SIFT descriptors, images catalogation, automatic analysis of images.

**Sumario:** 1. Introducción, 2. Metodología, 2.1. Extracción de características locales (SIFT Descriptors), 2.2. Construcción del vocabulario visual, 2.3. Comparación teniendo en cuenta la estructura espacial, 2.4. Clasificación, 3. Resultados, 4. Conclusiones, 5. Propuesta para futuras aplicaciones del proyecto. Referencias.

Este artículo recoge resultados de la investigación y publicación realizados con el soporte de: *Ayudas para la realización de proyectos de investigación pre competitivos en Ciencias Sociales y Humanas (APPCSHUM)*, Vicerrectorado de Política Docente y Científica de la Universitat de Barcelona, 2011 – 2012.

## 1. Introducción

El presente artículo es el resultado del proyecto de investigación *Ideación y catalogación artística basada en métodos de visión artificial* (2012) de la Universidad de Barcelona, en el que participan investigadores del ámbito de la Estadística y de las Bellas Artes; con el objetivo de obtener un sistema de catalogación y ordenación de imágenes aplicable al campo de la creación artística, que, además de la gestión mecánica de la documentación, facilitara nuevos valores y relaciones, imperceptibles a primera vista al ojo humano y, paralelamente, la generación de nuevas propuestas, es decir, que permitiera aflorar nuevos elementos de conocimiento a partir de las imágenes realizadas desde mecanismos o sistemas analíticos objetivos, que tendieran a superar maneras de mirar básicamente personales y parciales. Al proponer una investigación de estas características, dentro del ámbito de las bellas artes, se planteó que los resultados obtenidos se podrían extrapolar a todo tipo de acciones entorno a la creación, en el que la comparación entre imágenes fuera su característica principal.

Para realizar esta investigación fue básico poder disponer de más de 3.500 fotografías digitales, realizadas por uno de los miembros del grupo, el escultor Miquel Planas, y utilizadas en sus procesos de creación artística, fundamentalmente en dibujo y escultura; de esta forma se disponía de la mirada subjetiva del propio creador como elemento crítico en la investigación.

Con la difusión generalizada de la tecnología digital, el conjunto de imágenes producido por el artista crecía exponencialmente y también la dificultad de ordenar y catalogar con criterios estables las imágenes que se habían generado. En nuestro proyecto pretendemos obtener un sistema de almacenamiento no únicamente acumulativo, ni lineal de forma cronológica, sino que devenga colección, que tenga interés como archivo, que supere el hecho acumulativo para así aportar conocimiento basado en la información.

Ana Maria Guasch, en su libro *Arte y Archivo 1920-2010*, diferencia entre el archivo “neutro” que mantiene relación directa con el orden en el que fue creado, primando la procedencia al propio significado, al archivo de “colección” que observa criterios distintos a su origen (Guasch, 2011, p. 16). Igualmente la autora expone la importancia que representa una determinada ordenación en un archivo fotográfico:

*La relación entre archivo y fotografía se da desde los inicios históricos de ésta, sean el caso de las instituciones que categorizan sus objetos a través de fotografía (...), sea el de los fotógrafos como Eugène Atget y August Sander, que construyen una taxonomía de objetos a través de su obra. En ambos supuestos el objetivo es el mismo: proporcionar un*

*corpus de imágenes que representen un objeto, un carácter, un género o una circunstancia específica, lo cual significa que las fotografías se dispongan en una determinada "condición de archivo" (Guasch, 2011, p 26-27).*

De aquí creemos que se puede avanzar hacia un aprovechamiento mayor de toda esta valoración taxonómica; a partir de la información que este análisis taxonómico nos ofrezca podremos acceder a otros grados de creación, como apunta Guasch recogiendo una cita del proceso de trabajo de Hannah Höch. La artista comenta: "En todas partes encuentro una inspiración y es la chispa. Entonces empiezo un trabajo serio y arduo: encontrar el elemento que se ajusta perfectamente. El azar no desempeña ningún papel. Se trata de buscar una manera disciplinada de ensamblar y de verificar" (Guasch, 2011, p. 35) Habitualmente es el artista quien ordena y cataloga sus obras a partir de sus propias apropiaciones y elaboraciones, fruto de sus procesos mentales (Guasch, 2005, p. 166) y por lo tanto se ve implicado el mismo en su ordenación, deviniendo arte y parte de la misma, por ello queremos ofrecer un sistema objetivo que le pueda aportar nuevos puntos de vista más allá de los propios personales sobradamente conocidos y de poca aportación a nivel informativo. Por esto la investigación ahonda en sistemas de comparación y ordenación neutrales, sin distorsión y desapasionados, que aporten nueva información al gestor, y que esta información pueda ser utilizada, manipulada, administrada y/o alterada en función de sus intereses artísticos, tal como proponía Hannah Höch.

## 2. Metodología

La metodología utilizada en nuestra investigación está basada en determinar las características locales que producen una representación de la imagen versátil y sólida capaz de mostrar el contenido global y local al mismo tiempo, y a la vez hacen robusta la descripción ante la oclusión parcial de objetos contenidos y la transformación de la propia imagen.

Para la construcción de un vocabulario visual en el que basar la descripción de las imágenes, seguimos un procedimiento análogo al que se utiliza en el análisis automático de textos. Se conoce como modelo "Bag-of-Words" (BOW) porque cada documento está representado como una distribución de frecuencias de las palabras presentes en el texto, sin tener en cuenta las relaciones sintácticas existentes entre ellas.

En el ámbito de las imágenes nos referiremos a representaciones "Bag-of-Visual Terms" (BOV). Este enfoque consiste en analizar las imágenes como un conjunto de regiones, describiendo solamente su apariencia e ignorando su estructura espacial. La representación BOV se construye a partir de la extracción y cuantización automática de descriptores locales y ha demostrado ser una de las mejores técnicas para resolver diferentes tareas en la visión por computador. La representación BOV fue implementada por primera vez en el desarrollo de una sistema experto de reconocimiento de objetos (Willamowski, Arregui, Csurka, Dance & Fan, 2004).

La construcción BOV requiere dos decisiones principales de diseño:

- a. La elección de los descriptores locales que aplicamos en nuestras imágenes.
- b. La elección del método que se utilice para obtener el vocabulario visual.

Ambas decisiones pueden influir en el rendimiento del sistema resultante, sin embargo la representación BOV es robusta, conserva su buen comportamiento en un amplio rango de opciones de los parámetros.

Esta representación de una imagen no contiene información acerca de las relaciones espaciales entre palabras visuales, del mismo modo que la representación BOW remueve la información relativa al orden de las palabras en los documentos.

No obstante, los métodos BOV, que representan una imagen como una colección desordenada de características locales, han demostrado impresionantes niveles de rendimiento en tareas de categorización de imágenes completas. Sin embargo, debido a que estos métodos no tienen en cuenta toda la información acerca de la disposición espacial de las características, se ha visto limitada su capacidad descriptiva. En particular, son incapaces de capturar formas o de separar un objeto de su fondo.

Para superar las limitaciones del enfoque BOV hemos implementado la metodología PHOW (Pyramid histogram of visual words). La pirámide de coincidencias trabaja mediante la colocación de una secuencia de cuadrículas cada vez más finas sobre la imagen obteniendo una suma ponderada de la cantidad de coincidencias que ocurren en cada nivel de resolución de la pirámide (Grauman, K. & Darrel, T., 2005).

Detallamos a continuación los pasos de la determinación:

### 2.1. Extracción de características locales (SIFT Descriptors)

El descriptor *Scale Invariant Feature Transform (SIFT)* fue desarrollado por Lowe (2004) como un algoritmo capaz de detectar puntos característicos (*keypoints*) estables en una imagen.

Estos puntos son invariantes frente a diferentes transformaciones como traslación, escala, rotación, iluminación y transformaciones afines.

Originalmente fue desarrollado para el reconocimiento de objetos en general y para realizar la alineación de imágenes.

El algoritmo *SIFT* se compone principalmente de cuatro etapas que se describen siguiendo la implementación de Lowe (2004):

A. Detección de extremos en el Espacio de Escala: la primera etapa del algoritmo realiza una búsqueda sobre las diferentes escalas y dimensiones de la imagen identificando los candidatos a keypoints. Esto se lleva a cabo mediante la función *DoG (Difference-of-Gaussian)* y construyendo una pirámide Gausiana (Fig. 1).

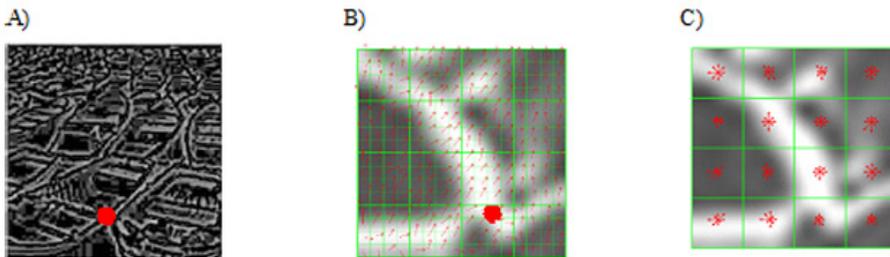


Figura 1. Pirámide Gausiana.

B. Localización de los *keypoints*: se seleccionan los *keypoints* a partir del conjunto de candidatos encontrados, aplicando una medida de estabilidad sobre todos ellos para descartar los que no sean adecuados. Un punto quedará seleccionado como *keypoint* sólo si es mayor que sus 26 vecinos o menor que todos ellos (Fig. 2).

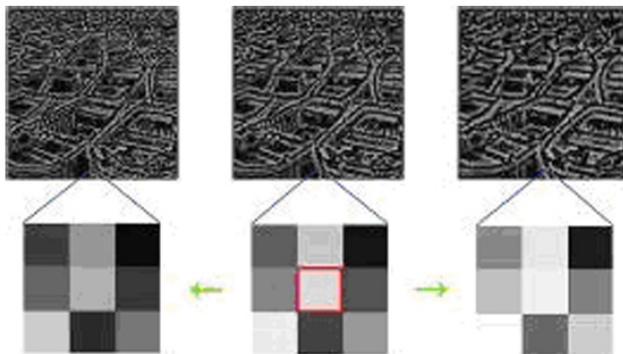


Figura 2. Vecinos anterior y posterior de escala.

C. Asignación de la orientación: se asignan una o más orientaciones a cada *keypoint* basándose en las direcciones locales presentes en la imagen gradiente. Definimos una región de 16x16 píxeles alrededor del punto donde vamos a determinar la orientación, y a cada uno de los píxeles se les calcula su gradiente (Fig. 3A y 3B). Éste viene determinado por su módulo e inclinación, ambos parámetros se calculan utilizando deferencias entre píxeles.

D. Descriptor del *keypoint*: la última etapa hace referencia a la representación de los *keypoints* como una medida de los gradientes locales de la imagen en las proximidades de dichos puntos clave y respecto de una determinada escala. Cada punto de interés corresponde a un vector de características compuesto por 128 elementos (Fig. 3C ).

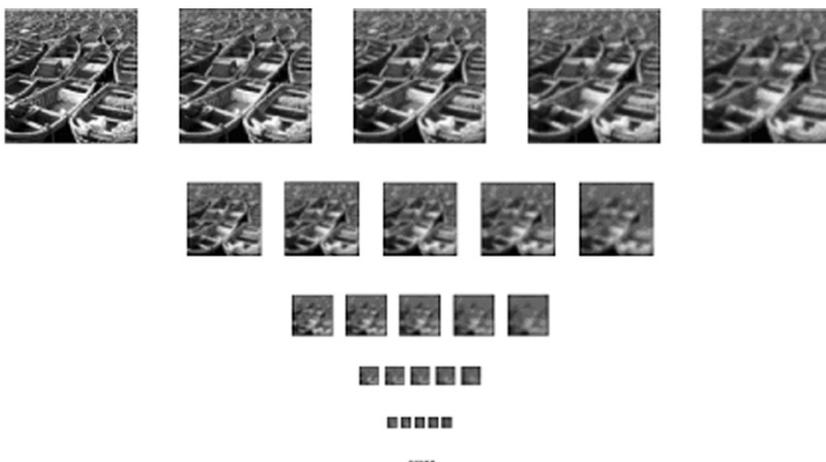


Figura 3A. Punto donde vamos a determinar la orientación. Figura 3B. Región de 16x16 píxeles alrededor del punto. Figura 3C. Descriptor de cada punto de interés formado por 128 elementos.

En algunos estudios de Lazebnik et al. (2006) y de Fei-Fei & Perona (2005), el cálculo de los descriptores locales *SIFT*, en vez de realizarse en los puntos de interés, se efectúa en los nodos de una malla regular sobreimpuesta en la imagen. Este enfoque es preferible con el fin de mejorar la capacidad de discriminación en implementaciones orientadas a la clasificación de escenas.

## 2.2. Construcción del vocabulario visual

El punto de partida para la construcción de un vocabulario visual es el conjunto de descriptores de la colección de imágenes y el punto al cual queremos llegar es un vocabulario de *visualterms*. Utilizaremos la expresión *visualterm* como equivalente a palabra visual, que sería la traducción literal al español.

Cada imagen ha quedado descrita mediante los descriptores *SIFT*. Considerando toda la colección de imágenes tenemos por tanto una gran colección de descriptores.

La construcción del vocabulario se realiza mediante agrupación (*clustering*). Más específicamente, aplicamos el algoritmo *k-means* a un conjunto representativo de descriptores locales extraídos de la colección de imágenes y tomaremos como palabras visuales los vectores de medias de cada clúster.

Usamos la distancia euclidiana en los procesos de agrupación y cuantización y elegimos el número de clústeres dependiendo del tamaño deseado de vocabulario.

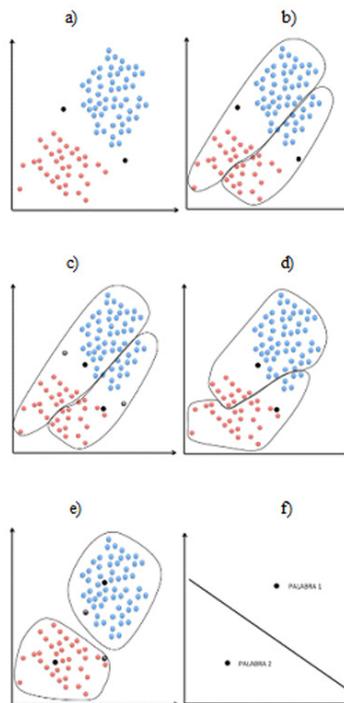


Figura 4A. Ejemplo del caso de descriptores bidimensionales y de dos palabras visuales.

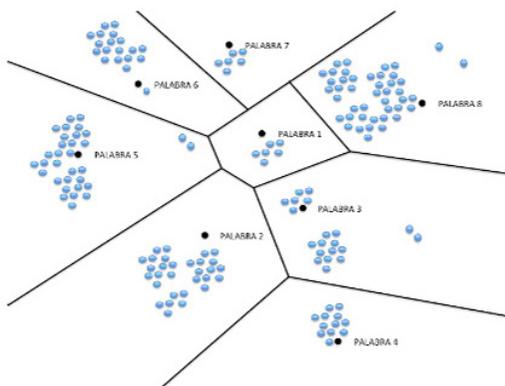


Figura 4B. Ejemplo para el caso de descriptores de más de dos palabras visuales.

A modo de ejemplo, supongamos el caso de descriptores bidimensionales y de dos palabras visuales (Fig. 4A). El algoritmo *k-means* establecerá una partición del espacio en dos regiones cada una asociada a una palabra.

a) Supongamos que los descriptores de la colección de imágenes configuran dos grupos separados (azul y rojo). El algoritmo empieza estableciendo dos centrosides al azar (negro).

b) Asignamos cada descriptor al centroide más cercano.

c) Recalculamos los nuevos centrosides de los grupos formados en la etapa anterior.

d) Repetimos la asignación de los descriptores al centroide más cercano.

e) El procedimiento prosigue recalculando los nuevos centrosides.

f) El proceso iterativo se detiene cuando no se produce cambio en los centrosides.

El segundo esquema (Fig. 4B) ilustra la partición del espacio de descriptores en el caso de un vocabulario de más palabras. Dado un descriptor calcularemos el centroide más cercano, y le corresponderá la palabra representada por dicho centroide.

Resumiendo, dada una colección de imágenes:

1) Inicialmente se calcularán los descriptores de las imágenes.

2) A continuación, se cuantizan los descriptores en  $M$  *clústeres*, los cuales definirán un vocabulario visual de  $M$  *visualterms*.

3) Una vez se dispone del vocabulario, los descriptores de cada imagen se asignan al *visualterm* más cercano.

4) Para obtener la representación BOV (*Bag-of-visualterms*) de una imagen dada, se calcula la frecuencia de cada *visualterm* en la imagen.

Esta representación de una imagen no contiene información acerca de las relaciones espaciales entre *visualterms*. No obstante, los métodos BOV, que representan una imagen como una colección desordenada de características locales, han demostrado impresionantes niveles de rendimiento en tareas de categorización de imágenes completas. Sin embargo, debido a que estos métodos no tienen en cuenta toda la información acerca de la disposición espacial de las características, se ha visto limitada su capacidad descriptiva. En particular, son incapaces de capturar formas o de separar un objeto de su fondo.

Para superar las limitaciones del enfoque BOV hemos implementado la metodología PHOW (Pyramid histogram of visual words). La pirámide de coincidencias trabaja mediante la colocación de una secuencia de cuadrículas cada vez más finas sobre la imagen obteniendo una suma ponderada de la cantidad de coincidencias que ocurren en cada nivel de resolución de la pirámide (Grauman et al. 2005).

### 2.3. Comparación teniendo en cuenta la estructura espacial

Para superar las limitaciones del enfoque *bag-of-visualterms* (BOV) hemos implementado la metodología PHOW (*Pyramid histogram of visual words*) introducida en Lazebnik et al. (2006) y Grauman et al. (2005). La pirámide de coincidencias trabaja mediante la colocación de una secuencia de cuadrículas cada vez más finas sobre el espacio de características obteniendo una suma ponderada de la cantidad de coincidencias que ocurren en cada nivel de resolución. (Fig. 5A y 5B, Histogramas de *Bag-of-visualterms*).

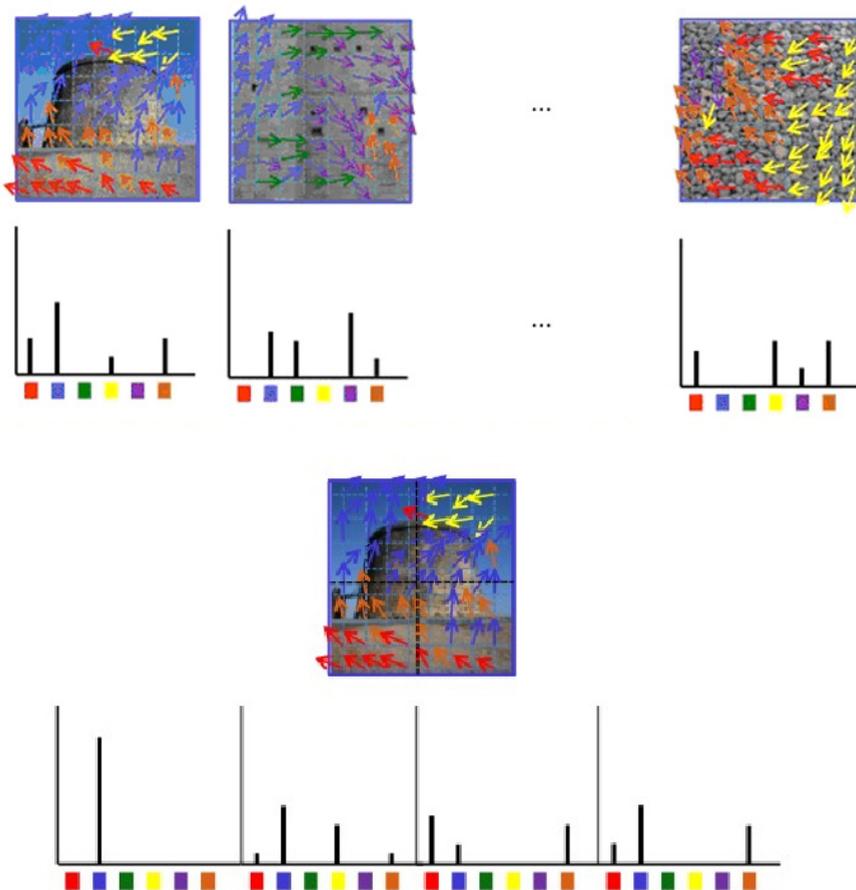


Figura 5A. Histogramas de *Bag-of-visualterms*. Nivel 0. Figura 5B. Concatenado de Histogramas de *Bag-of-visualterms*. Nivel 1 . (2D- image space).

### 2.4. Clasificación

Los histogramas espaciales se pueden utilizar como descriptores de imagen y alimentar con ellos un clasificador automático como por ejemplo los SVM (*Support Vector Machine*).

A continuación mostramos un esquema que resume todos los pasos del proceso (Fig.6A):

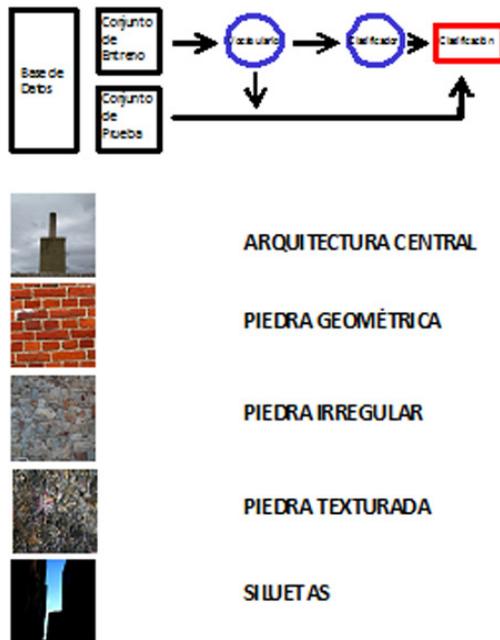


Figura 6A. Esquema que resume los pasos del proceso de clasificación.

### 3. Resultados

En el presente trabajo nos proponemos analizar automáticamente las imágenes de una base de datos de fotografías de Miquel Planas.

La resolución de las imágenes es de  $480 \times 480$  píxeles. La base de datos de prueba consta de 150 imágenes. Previamente se clasifican manualmente 75 de estas imágenes en 5 categorías (15 imágenes en cada categoría). Estas categorías corresponden a 5 tipologías diferentes identificadas en el conjunto de imágenes fotográficas por los miembros del grupo de investigación. Con este conjunto de datos hemos construido un vocabulario de 300 *visualterms*.

El objetivo de esta clasificación es entrenar al sistema y generar el vocabulario para después predecir la categoría a la que pertenece una nueva imagen problema. El esquema del proceso de clasificación sería el siguiente (Fig.6B):

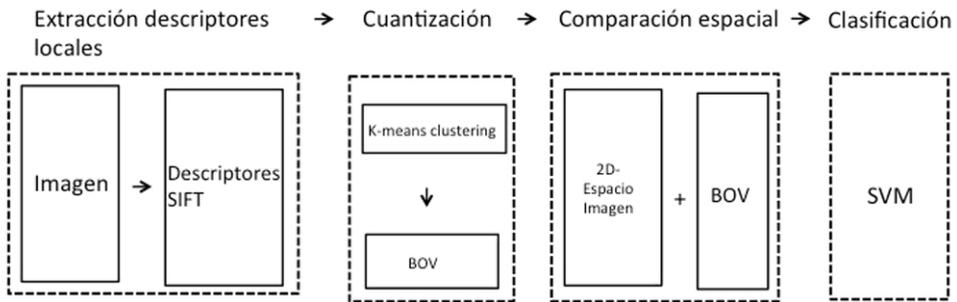


Figura 6B. Esquema del proceso de clasificación.

El proceso de clasificación de las fotos de prueba se repite 10 veces. A continuación acompañamos un esquema que muestra la proporción de clasificación de cada categoría. La fila es la categoría real y la columna es la predicción del sistema.

Arquitectura Central y siluetas son las categorías con una mayor proporción de clasificación correcta, 79% y 85% respectivamente.

Posteriormente, encontramos las categorías piedra texturada y piedra irregular con el 61% y el 57% de clasificación correcta. La mayoría de los errores de clasificación en esta categoría se deben a errores entre ambas categorías. La categoría de piedra geométrica tiene una proporción más baja de la clasificación correcta, el 41%. La mayoría de los errores se producen con la categoría de piedra irregular.

	AC	PG	PI	SI	PT	
AC	<b>79</b>	0	4	17	0	
PG	14	<b>41</b>	31	11	3	
PI	1	19	<b>57</b>	0	23	
SI	11	1	0	<b>85</b>	3	
PT	0	1	37	0	<b>61</b>	

El código eficiente para calcular los mapas de características está disponible como parte de la biblioteca de código abierto *VLFeat*. Si se desea ampliar los contenidos técnicos de este trabajo se recomienda consultar el artículo *Artistic ideation based on computer vision methods* de Reverter, Rosado, Figueras y Planas, publicado en *Journal of Theoretical and Applied Computer Science* 6 en 2012.

#### 4. Conclusiones

El problema de la clasificación de imágenes basadas en los objetos que contienen constituye un área de gran actividad en la investigación de la visión por computador. El conjunto de métodos disponibles en la actualidad que aborda el problema de la clasificación de imágenes en categorías es muy eficiente. En este trabajo se exploró el comportamiento de la técnica *bag-of-visualterms* cuando se enfrentan a una base

de datos de imágenes cuyas categorías están determinadas por aspectos semánticos implicados en el proceso de ideación artística. Hemos demostrado que estos métodos son adecuados para la clasificación de imágenes cuyas categorías se basan en aspectos semánticos. Los experimentos sugieren que la categorización automática de imágenes en base a métodos de visión por ordenador puede proporcionar principios objetivos en la catalogación de imágenes.

Los sistemas artificiales de visión aún están muy lejos del sistema visual humano. Para nosotros es muy fácil detectar objetos en una imagen o buscar relaciones y analogías. Disponemos de una habilidad increíble para interpretar las imágenes, pero esto se debe a que nuestro sistema visual utiliza mucha más información que la que le proporciona la imagen, como por ejemplo el conocimiento previo que tiene del entorno. A pesar de ello, actualmente hay un gran optimismo entre la comunidad científica en cuanto al futuro de la Visión Artificial. Ha sido posible desarrollar algoritmos para resolver problemas que hace 20 años parecían irresolubles, y es obvio que la tendencia seguirá al alza, dado el protagonismo actual de los dispositivos móviles con cámaras, tabletas y *smartphones*.

El proyecto que se presenta puede contribuir muy favorablemente al desarrollo del conocimiento científico del ámbito relacionado con el Arte. Una de las contribuciones más directas en la que se podría poner en práctica sería en la investigación artística iconográfica, necesitada de una metodología para la catalogación de obras de arte verdaderamente contrastada. No se intentaría substituir la intuición espontánea o el criterio experto, sino agudizar, reforzar o evidenciar nuevos elementos de utilidad para la mirada (o percepción visual (Rosado, P., Reverter, F. & Figueras, E., 2014).

Arnheim (1983, p.13) se quejaba de que “se ha dejado adormecer nuestra capacidad innata de entender con los ojos, y hay que volver a despertarla”. Quizá una forma de conseguirlo sea con la ayuda de los métodos de visión artificial que tenemos a nuestro alcance en la actualidad, y a ello pretendemos contribuir con este trabajo.

Muchas veces resulta difícil expresar con palabras lo que se percibe en una imagen. El artista visual hace uso de sus categorías formales para capturar desde lo particular aquello universalmente significativo, desde una forma necesariamente personal. Muchas veces podemos ver y sentir las cualidades de una obra sin poder explicarlas con palabras. El problema no sólo reside en el lenguaje, sino en que aún no hemos sido capaces de plasmar esas cualidades percibidas en las categorías adecuadas. Para poder nombrar algo debemos haberlo visto, oído, pensado o sentido con anterioridad (Arnheim, 1983). A esto se suma el hecho de que la mirada hacia el mundo es un juego de equilibrio entre las propiedades del objeto observado y la naturaleza del sujeto que observa.

## 5. Propuesta para futuras aplicaciones del proyecto

Ante la consolidación de las redes sociales, la aparición de nuevos medios de comunicación e interfaces en los sistemas operativos y dispositivos electrónicos, el ciudadano está cada vez más cerca de los sistemas interactivos en los que participa activamente, dejando a un lado el rol pasivo de simple receptor o espectador.

Ante esta evidencia, el proyecto que presentamos pretende ampliar de forma más dinámica la difusión y el conocimiento del arte, e intentar hacerlo de una manera más atractiva a través de la aproximación participativa del consumidor. En la mayoría de las instituciones artísticas ya sean galerías de arte, museos, pinacotecas, centros

artísticos, centros históricos, salas de arte, etc., habitualmente los visitantes se limitan a seguir un recorrido pautado contemplando las obras expuestas. De forma opcional, la visita puede ser acompañada de audio-guía o guía personal que complementa el recorrido visual e introduce al espectador en el universo contemplado, generalmente desde una vertiente historicista, y/o acompañado de unas referencias técnicas y estilísticas. Otro recurso clásico de transmisión del conocimiento de las obras es la palabra escrita en las cartelas y paneles informativos generales. Este sistema, a nuestro entender obsoleto, limita las fuentes documentales de carácter gráfico porque la información se transmite casi exclusivamente a través de la lectura o, como hemos comentado anteriormente, complementada oralmente con ayuda de las audio guías. Con estos formatos y canales predominantes, la información es sin ningún tipo de duda unidireccional y, a menudo, aburrida y poco significativa.

Partimos de la creencia de que si conseguimos hacer más interesante la contemplación de las obras expuestas en las instituciones citadas, probablemente aumente la afluencia de público y su grado de satisfacción. Para ello creemos importante reforzar el rol participativo del visitante. La interacción entre el visitante y la obra artística, generará *feedbacks* que llevarán a una mayor implicación e interés: El espectador, al intervenir en el proceso, adopta un cierto rol creativo, es un elemento activador.

Al ser un sistema creativo, generador de información, aporta nuevos conocimientos e interpretaciones y contribuye a establecer nuevas posibilidades o registros. Partimos de la idea de que al interactuar con las imágenes, salimos de la mirada arquetipo de la obra de arte, generando percepciones más abstractas e, incluso, inéditas de esa obra. El espectador será el que dirigirá su mirada y definirá los parámetros de la misma. Ya no será la mirada “oficial” o historicista o académica establecida. Entrará en juego una actitud lúdica, creativa, activa y expectativa del visitante.

El proyecto desarrollado puede constituir una pieza clave de un sistema CBIR (sistema de búsqueda para recuperar imágenes basándose en su contenido) que permitiría una aproximación nueva y diferente hacia instituciones artísticas. El espectador podría fijar su atención en una obra concreta y con un dispositivo móvil capturar la imagen, enviando ésta al sistema que sería capaz de acceder a una gran base de datos de imágenes para mostrarle aquellas de la colección que compartiesen con ella aspectos semánticos. El conocimiento y la interpretación no vienen sólo marcadas por los comisarios de las exposiciones o los responsables de dichos centros culturales, sino también por los mismos visitantes, que hacen sus lecturas e interpretaciones. El proyecto pretende facilitar, también, una difusión de las lecturas personales mediante las nuevas herramientas de comunicación a través de Internet, lo que puede replantear muchas de las normativas museísticas actuales, especialmente en lo que se refiere a la reproducción y difusión de las obras artísticas expuestas o en depósito.

El proyecto se convierte, además, en una propuesta divulgativa ante la dificultad que a veces supone la difusión de obras de difícil acceso o comprensión (obras más “cerradas”), debido a los conceptos o teorías formales, menos evidentes. Este sistema facilitaría la difusión de la información de los contenidos generales y especialmente de aquellos más específicos, implícitos y significativos, de cada una de las obras expuestas.

Ante la idea de exposiciones con documentación “hermética”, el sistema propone una fuente documental abierta, activa e interactiva con el usuario. La intención es que la obra de arte no se cierre en sí misma y se proponga ante el público como un elemento de experimentación por el cual el espectador puede extraer y comparar información para obtener diferentes lecturas. En este caso, el bagaje cultural e histórico del visitante no es determinante, puede acceder a nuevos o distintos grados de conocimiento en función de sus intereses, de la casuística, de la imaginación o de las ganas de experimentar... del espectador. El usuario no necesitaría conocer códigos específicos ni interpretaciones o traducciones. Se trataría de ofrecer una herramienta “universal” de lenguaje visual de fácil aplicación y sin la necesidad de conocimientos previos sofisticados.

El programa CBIR enfatiza aquellos aspectos más visuales, favoreciendo las comparaciones y las analogías de tipo formal (compositivo, cromático, organizativo, jerárquico...) y permitiendo la interactividad con las imágenes, alterando partes de la obra. En definitiva, el proyecto abre un nuevo campo en la mirada artística, aportando una visión más “abstracta” hacia la obra de arte haciendo evidente aquellas formas y relaciones que habitualmente se pueden escapar del ojo humano.

Concluyendo, este proyecto pretende hacer más atractiva y consciente la visita a las instituciones artísticas como pueden ser las exposiciones y los museos. Se dirige especialmente a las generaciones más jóvenes, las cuales están muy familiarizadas con el sistema operativo móvil y la red informática, y en consonancia con nuestra cultura visual. Pensamos que este tipo de proyectos puede constituir herramientas que potencien el estudio y la comparación entre obras artísticas, analizando sus elementos estéticos y formales. Puede aplicarse en los ámbitos de la docencia en general, considerándolo una herramienta práctica y lúdica para trabajar con niños, convirtiéndose en una aplicación informática de uso habitual y cotidiano, por su facilidad de aprendizaje y por su interactividad. También es un recurso recomendable en el ámbito de la docencia artística, como un medio para conocer e interactuar en el arte de todas las épocas y estilos, y, finalmente, pensamos que es de gran utilidad su incorporación en la investigación artística, especialmente cuando se trata de aplicar criterios de análisis formal de las obras de arte.

En un ámbito académico como el nuestro el sistema CBIR también puede ser de gran utilidad como recurso pedagógico para la mejora de la enseñanza artística. El docente y el discente pueden disponer de una herramienta de trabajo que les permita unos criterios de análisis formales, planteados de manera objetiva, de un conjunto de imágenes o de obras. Este recurso aumentaría la posibilidad de conectar y de establecer relaciones entre imágenes de distintos estilos, épocas, temáticas, etc., ya que se convierte en un mecanismo autónomo y no dependiente de los hechos habitualmente tratados de carácter más historicista.

El proyecto que se presenta puede contribuir muy favorablemente al desarrollo del conocimiento científico del ámbito relacionado con el arte. Una de las contribuciones más directas en la que se podría poner en práctica sería en las tesis doctorales dirigidas por los profesores pertenecientes a las áreas de conocimiento de Historia y Arte, favoreciendo sus investigaciones, ya que, exceptuando aquellas centradas en aspectos más técnicos, la mayoría se orientan hacia la investigación más iconográfica,

lo que hace realmente necesario una metodología para la catalogación de obras de arte verdaderamente contrastada.

En este sentido, por ejemplo, miembros del grupo de investigación estamos dirigiendo una tesis doctoral de análisis y catalogación de la obra gráfica de Josep Guinovart y creemos que la aplicación del sistema CBIR complementará la investigación y permitirá abrir nuevas posibilidades taxonómicas complementarias a las clasificaciones basadas en la cronología, las técnicas o en la temática.

Estamos llevando a cabo también investigaciones que apuntan hacia la utilización de estos aspectos formales detectados de manera automática en la producción artística. Disciplinas como la visión artificial y el arte generativo serán muy útiles para explorar las posibilidades que ofrecen las nuevas tecnologías al enriquecer el proceso creativo con la introducción de criterios computacionales en la generación de posibilidades expresivas.

## Referencias

- Arnheim, R. (1983). *Arte y percepción visual*. Madrid: Alianza forma.
- Fei-Fei, L. y Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. In *Proceedings of CVPR*, 2, pp. 524–531.
- Grauman, K. y Darrel, T. (2005). Pyramid match kernel: Discriminative classification with sets of image features. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Beijing.
- Guasch, A. M. (2005). Los lugares de la memoria: El arte de archivar y recordar, *Materia*, 5, pp. 157–183.
- Guasch, A. M. (2011). *Arte y Archivo 1920-2010. Genealogías, tipologías y discontinuidades*. Arte contemporáneo, 29. Madrid: Ediciones Akal.
- Lazebnik, S., Schmid, C. y Ponce, J. (2005). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of CVPR*, 1, pp. 26–33.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer*, 60(2), pp. 91-110.
- Reverter, F., Rosado, P., Figueras, E. y Planas, M.A. (2012). Artistic ideation based on computer vision methods, *Journal of Theoretical and Applied Computer Science*, 6(2), pp. 72-78.
- Rosado, P, Reverter, F. y Figueras, E. (2014). Intersecciones Semánticas entre Visión Artificial y Mirada Artística, *BR::AC Barcelona, Research, Art, Creation*, 2(1), pp. 1-54.
- Vedaldi, A. y Zisserman, A. (2010). Efficient Additive Kernels via Explicit Feature Maps. In *Proceedings of CVPR*. pp. 3539 – 3546.
- Vedaldi, A & Fulkerson, B. (2008). VLFeat - An open and portable library of computer vision algorithms. Retrieved from <http://www.vlfeat.org>
- Willamowski, J., Arregui, D., Csurka, G., Dance, C., & Fan, L. (2004). Categorizing nine visual classes using local appearance descriptors. In *Proceedings of LAVS Workshop, in ICPR'04*, Cambridge.