

# Mejora del ordenamiento en la recuperación de imágenes utilizando atributos textuales y un campo aleatorio de Markov

R. Omar Chávez, Luis E. Sucar, Manuel Montes  
*Instituto Nacional de Astrofísica Óptica y Electrónica*  
*Luis Enrique Erro # 1 72840, Santa María Tonantzintla, Puebla, México*  
*Email: {romarcg, esucar, mmontes}@inaoep.mx*

**Resumen**—Los sistemas de recuperación de imágenes (SRI) actuales no obtienen una lista de resultados ordenada apropiadamente debido a los atributos utilizados para describir las imágenes o los mecanismos de recuperación. En este trabajo se propone un método que, por medio de un campo aleatorio de Markov, combina los atributos textuales, el orden original de la lista de resultados y un enfoque de retroalimentación de relevancia para mejorar el orden de los resultados obtenidos por un SRI. Se realizaron experimentos con la colección IAPR TC-12 y un SRI del foro ImageCLEF 2008. Los resultados obtenidos muestran que se obtiene un orden más apropiado al utilizar el método propuesto, ya que mejora el orden de la lista original hasta en un 33 %, en el mejor de los casos, según la medida de evaluación *Mean Average Precision*. Además, los resultados obtenidos, sugieren trabajos futuros que utilicen una combinación de atributos textuales y visuales.

**Keywords**—Image re-ranking; Image retrieval; Markov random fields;

## I. INTRODUCCIÓN

En la tarea de recuperación de información, dada una consulta ingresada por un usuario, se deben recuperar todos los objetos de una colección que sean relevantes a la consulta de acuerdo con una medida de relevancia [1], [2]. Para la tarea de recuperación de imágenes los objetos a recuperar son imágenes.

Uno de los problemas a los que se enfrenta la tarea de recuperación de imágenes es el de describir completamente el contenido visual de una imagen. Tener una descripción completa permite recuperar todas las imágenes relevantes a una consulta. Se puede describir el contenido visual de una imagen utilizando una descripción textual, pero muchas veces esta descripción no incluye algunos detalles de la imagen que son relevantes para las consultas. Otras veces se utilizan características visuales como el color, la forma, la textura, entre otras, para describir una imagen, pero resulta computacionalmente costoso obtener las características visuales suficientes para describir completamente a la imagen.

Un sistema de recuperación de imágenes recibe una consulta (visual, textual o una combinación) dada por un usuario y obtiene una lista de imágenes ordenada, según un criterio de relevancia, de la más relevante a la menos relevante. En el caso ideal el SRI debe recuperar todas las imágenes relevantes, pero debido a los problemas de descripción no

todas las imágenes relevantes son recuperadas. Es por esto el SRI debe ofrecer una lista ordenada colocando en las primeras posiciones a las imágenes con más probabilidad de ser relevantes.

El orden apropiado en la lista de resultados de un SRI es importante porque facilita la tarea del usuario de buscar imágenes relevantes a una consulta.

Los sistemas de recuperación de imágenes actuales utilizan sólo un tipo de característica, o una combinación que no toma en cuenta la relación entre tipos de atributos, para almacenar, indexar o buscar imágenes. Estas consideraciones ocasionan que los resultados obtenidos no tengan un orden apropiado, ya que éstos dependen en gran medida del tipo de características tomadas en cuenta y de la intención de búsqueda del usuario al realizar la consulta [1].

Una forma de mejorar el orden de los resultados obtenidos por un SRI es utilizar un enfoque de retroalimentación de relevancia, ya sea manual (dada por el usuario) o automática. En la retroalimentación de relevancia, una vez obtenidos los resultados por el SRI, se hace una selección de cuales de las imágenes obtenidas son relevantes para posteriormente refinar la lista obtenida [3], [4], [5].

En este trabajo se propone un método que integra, mediante el uso de un campo aleatorio de Markov (CAM) [6], el orden original obtenido por un SRI, la similitud entre imágenes obtenida con las características textuales y la retroalimentación de relevancia, todo esto con el fin de generar un orden más apropiado para la lista de resultados de un SRI.

El CAM debe separar las imágenes relevantes de las que no lo son mediante el uso de una función de energía, la cual debe representar la similitud entre imágenes de la misma clase (relevantes y no-relevantes) haciendo uso de las características textuales. La función de energía también debe representar información *a priori* de la imagen, representada como la similitud con la consulta o el orden original del SRI. El utilizar la retroalimentación por relevancia permite incluir la intención de búsqueda en la función de energía.

Se realizaron experimentos tomando en cuenta las 39 consultas del foro Image CLEF 2008 hechas sobre el corpus IAPR TC-12 [7]. Se utilizó el SRI TIA-TXTIMG como SRI base para obtener la lista de resultados original [8].

Cada uno de los resultados obtenidos mejoran el orden de la lista original según las medidas de evaluación precisión y MAP (*Mean Average Precision*). En el mejor de los casos se obtiene una mejora de hasta el 33 %. Esto se debe principalmente a que las características textuales, utilizadas para medir la similitud, representan un factor importante al realizar el ordenamiento. Se obtuvieron resultados que demuestran que la descripción textual no es suficiente y sugieren el uso de atributos visuales para describir a las imágenes.

Este trabajo representa una primera etapa de un proyecto de tesis que propone mejorar el orden de las imágenes recuperadas por un SRI mediante la combinación de atributos textuales y visuales y el uso de un campo aleatorio de Markov.

Las secciones restantes del artículo están organizadas como sigue. En la sección 2 se hace una breve revisión del trabajo relacionado con el método propuesto. En la sección 3 se presenta y describe el método propuesto. En la sección 4 se describen los experimentos realizados así como la evaluación de los resultados obtenidos. Por último, en la sección 5 se presentan las conclusiones del trabajo realizado hasta el momento, y se sugieren trabajos futuros.

## II. TRABAJO RELACIONADO

La tarea de ordenar apropiadamente la lista de resultados obtenida por un SRI supone que, a pesar de la falta de un orden apropiado, existe un gran número de imágenes relevantes en la lista recuperada.

Existen diversos trabajos que tratan de resolver el problema de la falta de orden apropiado. Estos trabajos pueden ser clasificados según el tipo de características que toman en cuenta, el tipo de retroalimentación por relevancia o las técnicas utilizadas para generar el ordenamiento:

1. Según las características utilizadas:
  - a) Características visuales. Utilizan solamente las características visuales de las imágenes, por ejemplo características de forma, color y textura [4].
  - b) Características textuales. Utilizan características textuales relacionadas con las imágenes, por ejemplo descripciones textuales del contenido y títulos de las imágenes [1], [9].
  - c) Una combinación de tipos de características. Proponen combinar características visuales (por ejemplo el color y la textura), con características textuales de las imágenes (por ejemplo textos estructurados en XML) [10].
2. Según el tipo de retroalimentación de relevancia utilizado:
  - a) Relevancia dada por el usuario. Se le pide al usuario que seleccione algunas de las imágenes recuperadas que él considere relevantes para su intención de búsqueda [1].

- b) Relevancia obtenida automáticamente. Se hacen algunas suposiciones sobre la intención de búsqueda del usuario para seleccionar automáticamente imágenes relevantes [4], [5].

3. Según la técnica de ordenamiento utilizada:

- a) Construcción de clasificadores. Se seleccionan algunas imágenes como relevantes y otras como no relevantes para construir un clasificador y etiquetar las imágenes restantes [3].
- b) Funciones de similitud. Se proponen funciones de similitud, con base en ciertas suposiciones, para determinar las imágenes que están más relacionadas [4].
- c) Modelos probabilistas. Se proponen modelos de relevancia probabilista que permitan ordenar a las imágenes recuperadas [9].

El método propuesto toma en cuenta algunas de las ventajas de los trabajos relacionados, ya que considera utilizar atributos textuales para obtener la similitud entre imágenes. También toma en cuenta la retroalimentación de relevancia que permite obtener resultados más relacionados con la intención de búsqueda.

La aportación del método propuesto radica en utilizar un modelo gráfico probabilista (campo aleatorio de Markov) que permite incluir en una representación, la similitud entre imágenes dada por sus atributos, la retroalimentación de relevancia y el orden original obtenido por el SRI, todo esto con el fin de ordenar apropiadamente la lista de resultados original.

## III. MÉTODO PROPUESTO

El esquema general del método propuesto para ordenar las imágenes obtenidas por un SRI se muestra en la Figura 1. Dada una consulta, el SRI recupera de una colección de imágenes una lista de imágenes ordenadas por relevancia según los criterios de relevancia del SRI. De esta lista de resultados se seleccionan las imágenes relevantes mediante una técnica de retroalimentación de relevancia. Para cada una de las imágenes de la lista de resultados se extraen sus características textuales para generar una descripción de la imagen.

El conjunto de descripciones textuales, la consulta realizada por el usuario y la lista de imágenes seleccionadas como relevantes se introducen como parámetros al campo aleatorio de Markov para realizar el ordenamiento. El CAM, mediante el uso de una función de energía, separa las imágenes relevantes de las no-relevantes y con estas genera una lista en la que en primer lugar aparecen las imágenes relevantes y en segundo las imágenes no-relevantes. La lista obtenida por el método propuesto debe mejorar el ordenamiento de la lista original.

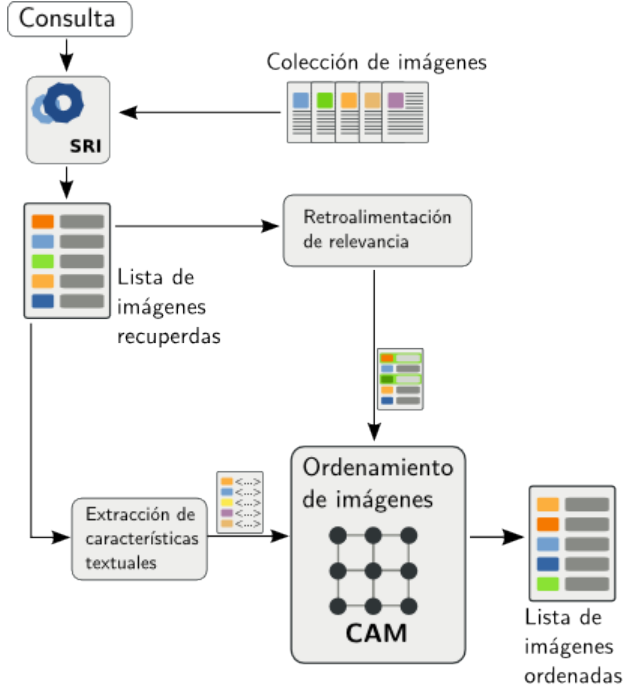


Figura 1: Esquema general del método propuesto. Como entrada, el método propuesto toma la lista original obtenida por un SRI. Después, mediante el uso de un CAM, separa las imágenes relevantes del resto para construir una nueva lista con un orden más apropiado que el de la lista original.

### III-A. Descripción de las imágenes

Una representación muy utilizada en la tarea de clasificación de textos es la representación vectorial. En esta representación cada atributo, con determinado peso, se representa como un elemento del vector que describe a la imagen. Para esta representación se decidió utilizar un pesado binario, el cual indica si una palabra está presente (valor 1) o ausente (valor 0).

La representación vectorial permite hacer uso de operaciones vectoriales como la distancia euclidiana, que nos dice qué tan cercanos (similares) son dos vectores (imágenes).

### III-B. Construcción del campo aleatorio de Markov

Los campos aleatorios de Markov modelan procesos en los que las variables aleatorias, además de tener una probabilidad asociada también se ven influenciadas por la probabilidad conjunta de sus variables vecinas. Dadas las probabilidades individuales de las variables y las probabilidades de sus vecinos, lo que se quiere encontrar es cada una de las posibles configuraciones de valores para las variables y, principalmente, la configuración más probable [6], [11].

Para construir el CAM propuesto se definieron 6 parámetros:

1. Un número de variables.
2. Los valores posibles para las variables.

3. Un esquema de vecindad que indique cuales serán los vecinos para cada variable.
4. Un algoritmo de iteración del campo.
5. Una función de energía que indique si una variable debe cambiar o no su valor, esta función debe tomar en cuenta los valores de los vecinos de la variable en cuestión e información del mundo exterior (observaciones).

Se representó a cada imagen como una variable aleatoria con 2 posibles valores: *relevante* y *no relevante*, lo que permite que las imágenes relevantes sean similares. Para comparar la relación entre todas las imágenes, se definió un esquema de vecindad en el que cada variable tiene como vecinas al resto de las variables.

Para iterar el CAM se utilizó el algoritmo ICM, el cual necesita de una configuración inicial, es en esta etapa donde se incluye la retroalimentación de relevancia, ya que la configuración inicial del CAM estará determinada por las imágenes seleccionadas como relevantes inicializadas con valor *relevante* y las no seleccionadas inicializadas con el valor contrario.

La función de energía propuesta toma en cuenta la similitud entre las imágenes, vista como la influencia entre variables vecinas, y la información *a priori* vista como el ordenamiento original del SRI y la consulta introducida al SRI. La función de energía se definió de la siguiente forma:

$$U(f) = V_c(f) + \lambda V_a(f)$$

Donde  $V_c$  es el potencial de interacción e indica la similitud entre la variable aleatoria  $f$  y sus vecinas.  $V_a$  es el potencial de asociación y representa la influencia de la información externa representada por el ordenamiento original del SRI y la consulta.  $\lambda$  es un factor que le da más relevancia al potencial de interacción ( $\lambda < 1$ ), al potencial de asociación ( $\lambda > 1$ ) o igual relevancia a ambos ( $\lambda = 1$ ).

El potencial  $V_c$  representa el nivel de apoyo, que dan los vecinos con el mismo valor, para que  $f$  cambie al valor que se está analizando. Si una variable tiene el valor *no relevante* y tiene una alta similitud con variables vecinas relevantes, estas apoyarán su cambio al valor de *relevante*. Si una variable tiene el valor *no relevante* y una alta similitud con variables vecinas *no relevantes*, estas evitarán su cambio al valor *relevante*, ya que apoyarán su valor *no relevante*. El potencial  $V_c$  fue definido como:

$$V_c(f) = \begin{cases} \frac{\sum_i^m \text{dist}(f, vnorel_i)}{m} & \text{si } f = \text{no-relevante} \\ \frac{\sum_i^m \text{dist}(f, vrel_i)}{m} & \text{si } f = \text{relevante} \end{cases}$$

Donde  $\text{dist}(f, vnorel_i)$  representa la distancia euclidiana entre la variable  $f$  y su vecina con valor *no-relevante*.  $\text{dist}(f, vrel_i)$  representa la distancia euclidiana entre la variable  $f$  y su vecina con valor *relevante*.

Debido a que puede no existir un número igual de vecinos relevantes y no relevantes, se obtiene el promedio de las

distancias de los vecinos con el mismo valor para calcular el potencial  $V_c$ .

El potencial  $V_a$  indica el nivel de confianza que se tiene para cambiar el valor de la variable  $f$  de acuerdo a qué tan parecida sea con la consulta y de acuerdo a la posición que tenga la imagen en la lista original obtenida por el SRI.

El potencial  $V_a$  se obtiene mediante dos factores. El primero indica qué tan similar,  $dist(f, q)$ , o diferente,  $(1 - dist(f, q))$ , es la variable  $f$  a la consulta  $q$  (representada como un vector de características textuales) introducida al SRI. El segundo factor representa el ordenamiento obtenido por el SRI, ya que la imagen será más relevante entre más cercana esté al inicio de la lista,  $\frac{\log(p)}{\log(n)}$ , o menos relevante en el caso contrario,  $\frac{\log(n-p)}{\log(n)}$ .  $n$  representa el número de imágenes a ordenar de la lista de resultados, y  $p$  la posición de una imagen en particular. El potencial  $V_a$  fue definido de la siguiente forma:

$$V_a(f) = \begin{cases} (1 - dist(f, q)) \frac{\log(n-p)}{\log(n)} & \text{si } f = \text{no-relevante} \\ dist(f, q) \frac{\log(p)}{\log(n)} & \text{si } f = \text{relevante} \end{cases}$$

Después de definir cada uno de los potenciales, la función de energía propuesta queda definida como:

$$U(f) = \begin{cases} \frac{\sum_i^m dist(f, v_{norel_i})}{m} + \lambda((1 - dist(f, q)) \frac{\log(n-p)}{\log(n)}) & \text{si } f = \text{no-relevante} \\ \frac{\sum_i^m dist(f, v_{rel_i})}{m} + \lambda(dist(f, q) \frac{\log(p)}{\log(n)}) & \text{si } f = \text{relevante} \end{cases}$$

La distancia euclidiana fue normalizada, utilizando la distancia más lejana entre variables, a valores entre 0 y 1.

#### IV. EXPERIMENTACIÓN Y EVALUACIÓN

Para realizar los experimentos se utilizaron los recursos del foro Image CLEF 2008, los cuales constan de la colección de imágenes IAPR TC-12, un conjunto de consultas para la tarea de recuperación de fotografías y las listas de resultados de uno de los participantes del foro (TIA-TXTIMG). Se eligió este corpus porque cuenta con juicios de relevancia para cada una de las consultas, esto permitió realizar una comparación con los resultados obtenidos por el método propuesto.

El SRI TIA-TXTIMG obtiene una lista de imágenes recuperadas mediante la combinación de las listas de resultados de varios métodos de recuperación, los cuales están basados en el contenido visual de las imágenes, en la descripción textual, en conceptos visuales, en anotaciones o en la recuperación multimodal que toma en cuenta una combinación de atributos visuales y textuales.

Se probó el método propuesto con las 39 consultas del foro Image CLEF 2008. Estas consultas constan de 3 imágenes de ejemplo y una descripción textual que incluye una narrativa sobre las imágenes relevantes a la consulta



```
<num> 2 </num>
<title> church with more than two
towers</title>
<cluster> city </cluster>
<narr> Relevant images will show a church,
cathedral or a mosque with three or more
towers.</narr>
<image> SampleImages/02/16432.jpg </image>
<image> SampleImages/02/37395.jpg </image>
<image> SampleImages/02/40498.jpg </image>
```

Figura 2: Ejemplo de una consulta para la tarea de recuperación de fotografías del foro Image CLEF 2008. Cada consulta incluye, entre otros campos, un campo *título* que resume el objetivo de la consulta, y un campo *narrativa* que especifica textualmente qué imágenes son relevantes para la consulta.

y algunas sugerencias sobre las imágenes que no son relevantes. Para el caso del método propuesto, no se tomaron en cuenta las imágenes de ejemplo debido a que, para esta primera propuesta, no se contempló utilizar atributos visuales. Tampoco se tomaron en cuenta las sugerencias sobre las imágenes no relevantes como parte de la consulta, debido a que no ofrecían información sobre las imágenes relevantes. La Figura 2 muestra un ejemplo de consulta para la tarea de recuperación de fotografías del foro Image CLEF 2008.

Cada una de las imágenes de la colección IAPR TC-12 tienen asignado un conjunto de campos descriptivos, de los cuales sólo se tomaron las palabras del título y de la descripción textual para representar a las imágenes. La Figura 3 muestra un ejemplo de una imagen de la colección IAPR TC-12 y sus campos descriptivos.

Como datos de entrada para el método propuesto, se tomaron las primeras 100 imágenes recuperadas por el SRI TIA-TXTIMG para cada una de las 10 consultas.

Se realizaron 5 experimentos de acuerdo con el valor de  $\lambda$ , para evaluar el comportamiento del método propuesto. El Cuadro I muestra el objetivo de cada uno de los



```

<TITLE>The Plaza de Armas</TITLE>
<DESCRIPTION>a yellow building with white
columns in the background; two palm trees
in front of the house; cars parked in
front of the house; a woman and a child
are walking over the square;</DESCRIPTION>
<NOTES>The Plaza de Armas is one of the most
visited places in Cochabamba. The locals are
very proud of the colourful buildings</NOTES>
<LOCATION>Cochabamba, Bolivia</LOCATION>

```

Figura 3: Ejemplo de una imagen de la colección IAPR TC-12 más su conjunto de campos descriptivos. Cada imagen de la colección tiene asignada algunos campos descriptivos que especifican el contenido visual de la imagen, por ejemplo el título, descripción y notas de la imagen.

experimentos realizados. Cada uno de los 5 experimentos se realizó tomando en cuenta 1, 3 y 5 imágenes como retroalimentación de relevancia.

Cuadro I: Objetivo de cada uno de los experimentos realizados para evaluar el método de propuesto según el valor de  $\lambda$ .

Valor de $\lambda$	Objetivo del experimento
$\lambda = 1,5$	Darle mayor peso a $V_a$ .
$\lambda = 1$	Darle igual peso a $V_c$ como a $V_a$ .
$\lambda = 0,5$	Darle mayor peso a $V_c$ .
$\lambda = 0$	Anular la aportación de $V_a$ .
$\lambda = \infty$	Anular la aportación de $V_c$ .

El enfoque de retroalimentación utilizado fue una retroalimentación simulada. Debido a que se conoce de antemano cuales son las imágenes relevantes de la lista recuperada (por los juicios de relevancia de la colección utilizada). Cabe mencionar que las imágenes seleccionadas como relevantes no necesariamente estaban en las primeras posiciones de la lista de imágenes recuperadas, éstas podrían estar en cualquier posición de la lista.

Cuando el CAM converge, las imágenes representadas como variables aleatorias, que terminaban con el valor de

relevante son colocadas en las primeras posiciones de la lista ordenada respetando el orden de aparición en la lista original. Las imágenes que terminaban con el valor de *no relevante* son colocadas, en la lista ordenada, después de haber colocado todas las imágenes con valor *relevante*.

Para evaluar los resultados obtenidos por el método propuesto se utilizaron las medidas de precisión y MAP. La precisión se define como el porcentaje de imágenes relevantes recuperadas del total de imágenes recuperadas. El MAP se define como la media del promedio de precisiones de varias consultas a diferentes niveles de recuerdo. El recuerdo se define como el porcentaje de imágenes relevantes recuperadas del total de imágenes relevantes. Entre mayor sea el valor del MAP más apropiada será la lista obtenida.

Una comparativa entre los resultados originales de la lista recuperada por el SRI TIA-TXTIMG y los resultados obtenidos con el método propuesto para diferentes valores de  $\lambda$  se muestra en el Cuadro II, donde los valores mostrados son un promedio de los valores obtenidos para las 39 consultas. Se puede ver que en cada una de las medidas consideradas, todas las variantes del método propuesto mejoraron los valores de la lista original, esto nos da evidencia empírica de que el método propuesto obtiene un orden más apropiado para las consultas realizadas.

Cuadro II: Tabla comparativa entre los resultados obtenidos por el SRI TIA-TXTIMG y por el método propuesto con todas sus variantes. El valor después de la letra F indica el número de imágenes tomadas como retroalimentación de relevancia. El valor después de la letra L representa el valor de  $\lambda$ .  $L_\infty$  indica que sólo se tomo en consideración al factor  $V_a$  para la función de energía. P5, P10 y P20 representan la precisión obtenida a las 5,10 y 20 imágenes recuperadas respectivamente; MAP es la medida *Mean Average Precision*.

	P5	P10	P20	MAP
TIA-TXTIMG	0.476	0.453	0.391	0.235
F1-L15	0.579	0.466	0.398	0.242
F1-L1	0.594	0.466	0.405	0.251
F1-L0.5	0.564	0.482	0.382	0.249
F1-L0	0.625	0.543	0.457	<b>0.264</b>
F1-L $\infty$	0.553	0.469	0.416	0.245
F3-L15	0.779	0.551	0.442	0.269
F3-L1	0.805	0.571	0.453	0.279
F3-L0.5	0.810	0.610	0.452	0.280
F3-L0	0.800	0.612	0.485	<b>0.285</b>
F3-L $\infty$	0.764	0.566	0.444	0.272
F5-L15	1.000	0.687	0.488	0.300
F5-L1	1.000	0.692	0.497	0.305
F5-L0.5	1.000	0.581	0.509	0.312
F5-L0	1.000	0.702	0.539	<b>0.314</b>
F5-L $\infty$	1.000	0.674	0.482	0.297

La mejora que obtiene el método propuesto se debe a las consideraciones tomadas en cuenta en la definición de los parámetros del CAM y a la función de energía. Considerar

esta información lleva a relacionar las imágenes relevantes entre sí. Las imágenes no relevantes son discriminadas por la poca similitud con la consulta, por la posición en la lista original, y por el mayor parecido con imágenes no relevantes.

Las imágenes seleccionadas como relevantes por la retroalimentación de relevancia, representan la información necesaria para que sus vecinas similares cambien su valor a *relevante*, sumado a esto se tiene el apoyo de la información *a priori* de la posición en la lista original.

Los resultados obtenidos al probar el método con  $\lambda = 0$  indican que, para la función de energía propuesta, se le debe dar más peso al potencial  $V_c$  o se debe reconsiderar el potencial  $V_a$  para que represente de una mejor manera la información *a priori*.

Los factores  $\frac{\log(p)}{\log(n)}$  y  $\frac{\log(n-p)}{\log(n)}$  dan una preferencia considerable a las imágenes en los primeros lugares o a los últimos según sea el caso. Esta preferencia puede provocar un sesgo al realizar el ordenamiento.

Un ejemplo de las primeras 20 imágenes de la lista original obtenida por el SRI para la consulta con título *straight road in the USA* y de lista ordenada obtenida por el método propuesto se muestra en la Figura 4, donde se puede ver que el método propuesto coloca en las primeras posiciones de la lista ordenada algunas imágenes que el SRI no había colocado. Para este ejemplo todas las imágenes colocadas por el CAM en las primeras 10 posiciones son relevantes, esto se debe, en gran parte, a la descripción textual del contenido de las imágenes y a que la función de energía pudo aprovechar esta información junto con el conocimiento *a priori*.

Los experimentos mostraron casos en los que la falta de información en la descripción textual provocó que la mejora del orden no fuera tan considerable como en el ejemplo de la Figura 4. La Figura 5 muestra las primeras 20 imágenes obtenidas por el SRI y las primeras 20 obtenidas por el CAM para la consulta con título *church with more than two towers*. Para este ejemplo pocas de las imágenes en las primeras posiciones en verdad son relevantes, debido a que en su descripción textual no se menciona información sobre el número de torres en el edificio, el cual es un factor importante que determina la relevancia de las imágenes para esta consulta. Estos resultados sugieren utilizar atributos visuales, en el método propuesto, para describir a las imágenes.

## V. CONCLUSIONES Y TRABAJO A FUTURO

En este trabajo se propuso un método para mejorar el ordenamiento de una lista de imágenes obtenida por un SRI. Se demostró experimentalmente que utilizando un CAM para integrar el orden original obtenido por el SRI, el enfoque de retroalimentación de relevancia y la similitud entre imágenes, se mejora el orden de la lista original hasta en un 33 % en el mejor de los casos. Ya que mediante esta

representación se pueden separar las imágenes relevantes de las que no lo son.

Se puede concluir, de los resultados obtenidos, que se debe reconsiderar el potencial  $V_a$  para que represente mejor la información externa ofrecida por el ordenamiento original, ya que los experimentos mostraron que se obtienen mejores resultados al anular este factor. También se puede concluir que el potencial  $V_c$  es de suma importancia para el método de ordenamiento.

Para algunas consultas la descripción textual no aporta información relevante sobre el contenido visual de la imagen, es por ello que se propone, como trabajo a futuro, probar el método de ordenamiento propuesto utilizando una combinación de atributos textuales y visuales de tal manera que se aprovechen las ventajas de ambos atributos.

Se pueden considerar también realizar experimentos utilizando otros tipos de retroalimentación de relevancia, por ejemplo retroalimentación obtenida automáticamente.

Otra propuesta de trabajo a futuro es la de probar el método propuesto en una colección de imágenes que no tenga una estructura bien definida para la información textual y visual de la imagen, por ejemplo se pueden utilizar imágenes recuperadas por un buscador Web de imágenes como Google Images.

Finalmente, para probar el método propuesto, se puede sustituir el SRI base por otros SRIs que consideren diferentes atributos o técnicas para la tarea de recuperación de imágenes.

## REFERENCIAS

- [1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, 2008.
- [2] A. W.M.Smeulders, MarcelWorring, SimoneSantini, AmarnathGupta, and RameshJain, "Content-based image retrieval at the end of the early years," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [3] T. Deselaers, R. Paredes, E. Vidal, and H.Ñey, "Learning weighted distances for relevance feedback in image retrieval," in *International Conference on Pattern Recognition 2008*, Tampa, Florida, USA, 08/12/2008 2008.
- [4] J. Cui, F. Wen, and X. Tang, "Real time google and live image search re-ranking," in *MM '08: Proceeding of the 16th ACM international conference on Multimedia*. New York, NY, USA: ACM, 2008, pp. 729–732.
- [5] A. Marakakis, N. Galatsanos, A. Likas, and A. Stafylopatis, "Application of relevance feedback in content based image retrieval using gaussian mixture models," in *ICTAI '08: Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 141–148.
- [6] S. Li, "Markov random field models in computer vision," *Computer Vision — ECCV '94*, pp. 361–370, 1994. [Online]. Available: <http://dx.doi.org/10.1007/BFb0028368>

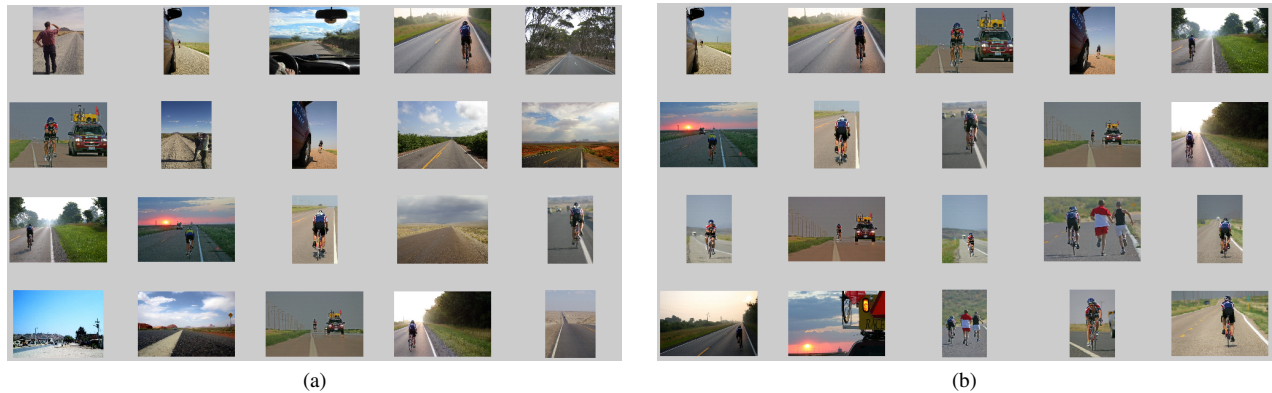


Figura 4: Primeras 20 imágenes de la lista obtenida por el SRI (a) y de la lista ordenada por el CAM (b) para la consulta *straight road in the USA*.

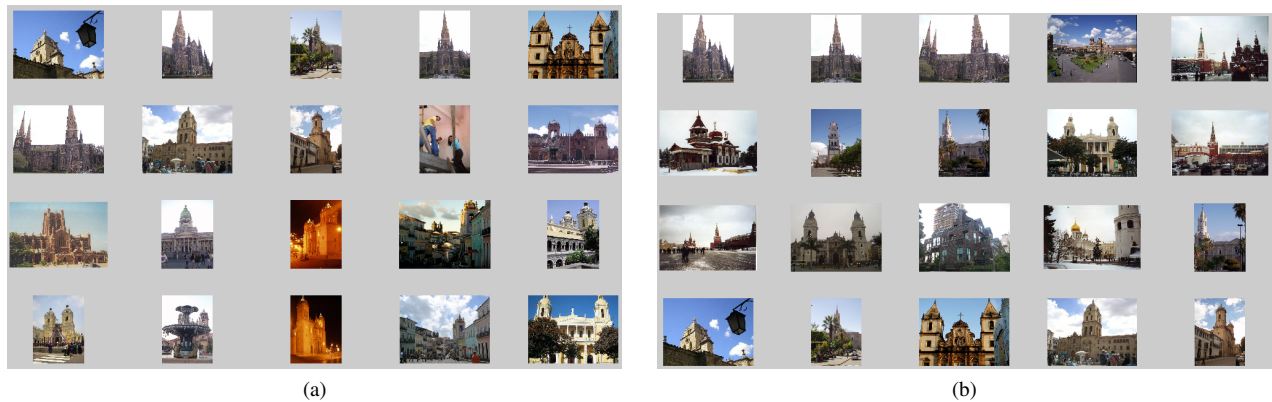


Figura 5: Primeras 20 imágenes de la lista obtenida por el SRI (a) y de la lista ordenada por el CAM (b) para la consulta *church with more than two towers*.

- [7] A. H. M. Grubinger, P. Clough and H. Müller, “Overview of the imageclef 2007 photographic retrieval task,” in *Springer Lecture Notes in Computer Science*. Budapest, Hungary: Springer, 03/2008 2008, pp. 433–444.
- [8] H. J. Escalante, J. A. Gonzalez, C. Hernandez, A. Lopez, M. Montes, E. Morales, E. Sucar, and L. Villasenor, “Tia-naoe’s participation at imageclef2008,” in *Working Notes of the CLEF 2008 Workshop*, Aarhus, Denmark, sep 2008.
- [9] W.-H. Lin, R. Jin, and A. Hauptmann, “Web image retrieval re-ranking with relevance model,” in *WI '03: Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*. Washington, DC, USA: IEEE Computer Society, 2003, p. 242.
- [10] J. T. J. T. Awg Iskandar, D. Pehcevski, “Combining image and structured text retrieval,” in *Advances in XML Information Retrieval and Evaluation*. Springer, 2006, pp. 525–539.
- [11] C. Hernández-Gracidas and L. E. Sucar, “Markov random fields and spatial information to improve automatic image annotation,” *Lecture Notes in Computer Science*, vol. 4872/2007, no. 4, pp. 879–892, 2007.