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

Ricardo Omar Chávez García

Instituto Nacional de Astrofísica Óptica y Electrónica

14 de octubre de 2009

Resumen

¿Cuál es el problema?

El orden de la lista de imágenes, obtenida por un sistema de recuperación de imágenes (SRI), no siempre es el apropiado debido a los atributos utilizados para representarlas o a los mecanismos de recuperación.

¿Cómo se pretende resolver el problema?

Proponiendo un método que combine atributos visuales y textuales y un modelo que integre, además de la combinación anterior, el orden orginal del SRI y un enfoque de retroalimentación de relevancia.

Introducción (1)

Campo aleatorio de Markov:

- Un campo aleatorio es una colección de variables aleatorias indexadas por sitios.
- Un campo de Markov asigna probabilidad a cada configuración en el espacio de posibles configuraciones.
- El problema central es encontrar la configuración con mayor probabilidad.

Introducción (2)

Campo aleatorio de Markov

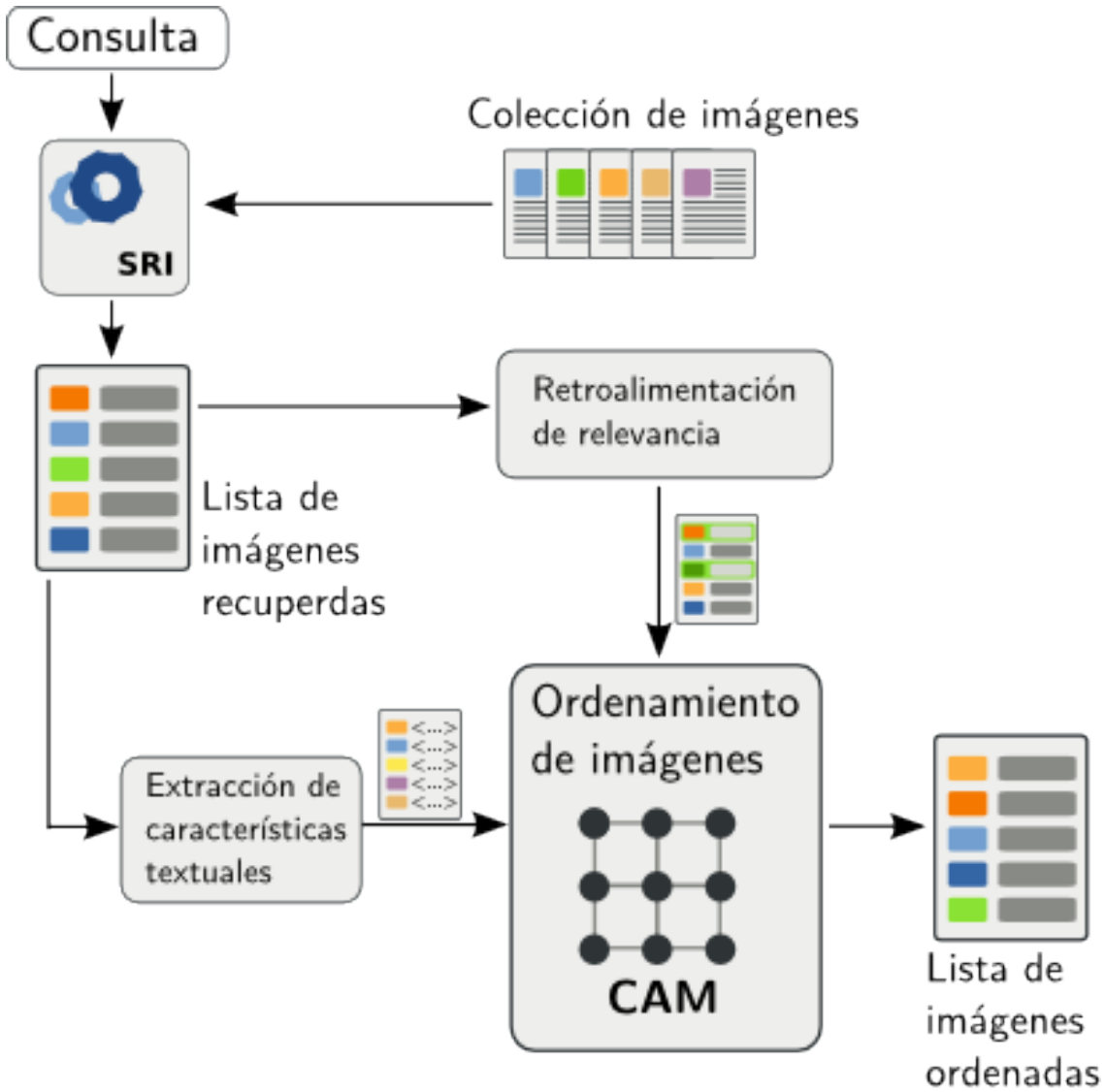
Sea un conjunto de variables aleatorias $F = \{F_1, ..., F_m\}$, asociadas a cada sitio del sistema de sitios S. Cada variable toma un valor f_i de un conjunto de posibles valores L. Entonces se dice que F es un campo aleatorio.

Un campo aleatorio de Markov es un campo aleatorio con las siguientes propiedades:

$$P(f) > 0$$

$$P(f_i|f_{s-i}) = P(f_i|vec(f_i))$$

Esquema general del método propuesto



CAM propuesto

Para construir el campo se definieron 6 parámetros.

- Variables. Cada imagen es representada por una variable aleatoria.
- Valores. Las variables tienen dos posibles posibles valores: relevante y no relevante.
- Esquema de vecindad. Cada variable tiene como vecinas al resto de las variables.
- Campo inicial. El campo inicial está definido por la retroalimentación de relevancia.
- Función de energía. Representa la información obtenida de los vecinos y la información *a priori*.
- Algoritmo de optimización. Se utilizó el algorimo ICM, el cual prefiere el valor con menor energía.

Campo inicial

- Si la imagen es seleccionada en la retroalimentación de relevancia tendrá el valor *relevante*.
- Si la imagen no es seleccionada en la retroalimentación de relevancia tendrá el valor *no relevante*.
- El valor de una variable seleccionada en la retroalimentación de relevancia no cambiará.

Atributos textuales

- Se utilizaron las palabras de los campos descriptivos de la imagen.
- Cada imagen fue representada por una bolsa de palabras binaria.
- La consulta fue tratada como una imagen.

Función de energía (1)

Está compuesta por dos potenciales: V_c es el potencial de interacción y V_a el potencial de asociación.

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

Función de energía (2)

El potencial V_c representa la información de los vecinos de la variable.

$$V_c(f) = \begin{cases} \bar{Y} + (1 - \bar{X}) & si\ f = no\ relevante \\ \bar{X} + (1 - \bar{Y}) & si\ f = relevante \end{cases}$$

$\bar{Y}$ es el promedio de las distancias entre la variable f y sus vecinas *no relevante*. $\bar{X}$ es el promedio de las distancias entre la variable f y sus vecinas *relevante*.
La similitud entre variables fue definida como: $1 - dice(f, g)$, donde $dice(f, g)$ representa el coeficiente Dice: $|f \cap q|/|f \cup q|$.

Función de energía (3)

El potencial V_a representa la información *a priori*, compuesta por el orden original del SRI y la similitud con la consulta.

$$V_a(f) = \begin{cases} (1 - sim(f, q)) * g(posinv(f)) & sif = no\ relevante \\ sim(f, q) * g(pos(f)) & sif = relevante \end{cases}$$

$sim(f, q)$ está definida por: $1 - |f \cap q|/|q|$.
La función $pos(f)$ devuelve la posición de la imagen f en la lista original, $posinv(f)$ devuelve el inverso de la posición.
La función seleccionada para el mapeo fue $g(x) = exp(x/20)/exp(5)$.

Función de energía (4)

Una vez descritos cada uno de los potenciales, la función de energía propuesta queda definida como:

$$U(f) = \begin{cases} \bar{Y} + (1 - \bar{X}) + \lambda(1 - sim(f, q)) * g(posinv(f)) \\ sif = no\ relevante \\ \\ \bar{X} + (1 - \bar{Y}) + \lambda sim(f, q) * g(pos(f)) \\ sif = relevante \end{cases}$$

Experimentos (1)

Se utilizó la colección IAPR TC-12 del foro Image CLEF2008 para la tarea de recuperación de fotografías.

```
<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>
```

Experimentos (2)

La tarea de recuperación de fotografías consta de 39 consultas.

```
<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>
```

Experimentos (3)

- El SRI base utilizado fue el SRI-TIATXTIMG.
- Se probó el método propuesto con 3 tipos de retroalimentación:
  - Manual (simulada): por los juicios de relevancia se conocen las imágenes relevantes para cada consulta.
  - Automática (ciega): los primeros N, y los primeros N después de un ordenamiento.
- Se realizaron experimentos para distintos valores de λ (50, 1.5, 1.0, 0.5, 0.3, 0, ∞) y tomando un número diferente de imágenes como relevantes (1, 3, 5, 8, 10).

Resultados (1)

Retroalimentación simulada.

Experimento	P5	P10	P20	MAP
TIA-TXTIMG	0.4769	0.4538	0.3910	0.2359
F1-L0.3	0.6103	0.5410	0.4833	0.2902
F1-LVA	0.5949	0.5000	0.4192	0.2538
F3-L0.3	0.7846	0.6128	0.4962	0.3070
F3-LVA	0.8000	0.5897	0.4474	0.2786
F5-L0.3	1.0000	0.7154	0.5474	0.3358
F5-LVA	1.0000	0.6923	0.4885	0.3015
F8-L0.3	1.0000	0.8821	0.6218	0.3706
F8-LVA	1.0000	0.8718	0.5641	0.3374
F10-L0.0	1.0000	0.9744	0.6551	0.3858
F10-LVA	1.0000	0.9744	0.6179	0.3580

Resultados (2)

Retroalimentación automática: primeros N.

Experimento	P5	P10	P20	MAP
TIA-TXTIMG	0.4769	0.4538	0.3910	0.2359
F1-L50	0.5179	0.4692	0.4038	0.2412
F1-L0.0	0.3949	0.3667	0.2974	0.2182
F3-L1.0	0.5077	0.4795	0.3974	0.2497
F3-L0.0	0.4872	0.4333	0.3628	0.2390
F5-L0.3	0.4769	0.4718	0.4128	0.2492
F5-LVA	0.4769	0.4615	0.4026	0.2383
F8-L0.5	0.4769	0.4692	0.4013	0.2483
F8-L0.0	0.4769	0.4590	0.3885	0.2378
F10-L0.0	0.4769	0.4538	0.4115	0.2443
F10-LVA	0.4769	0.4538	0.4038	0.2377

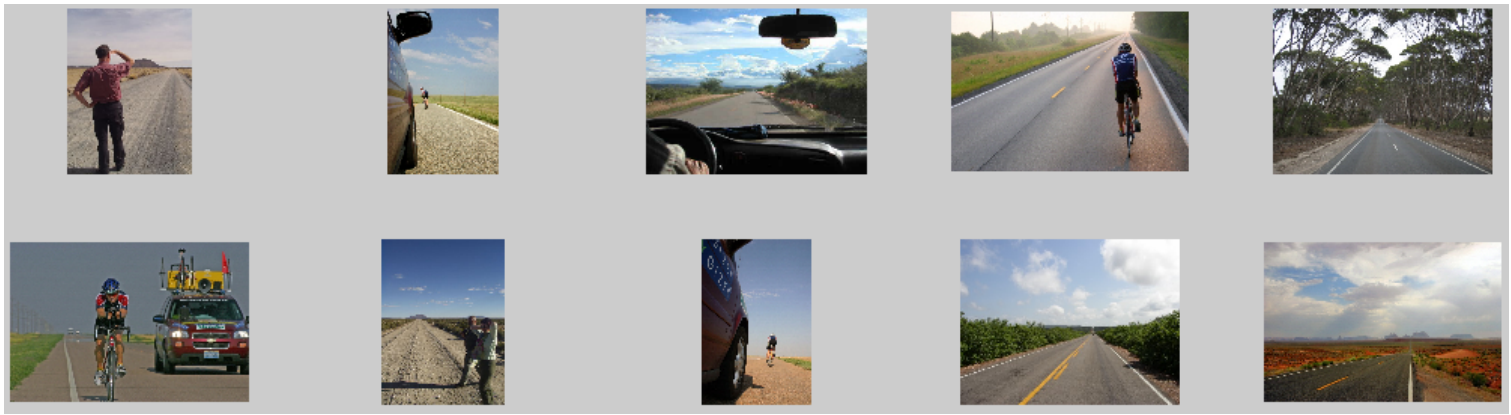
Resultados (3)

Retroalimentación automática: primeros N después de ordenarlos con la similitud *Dice*.

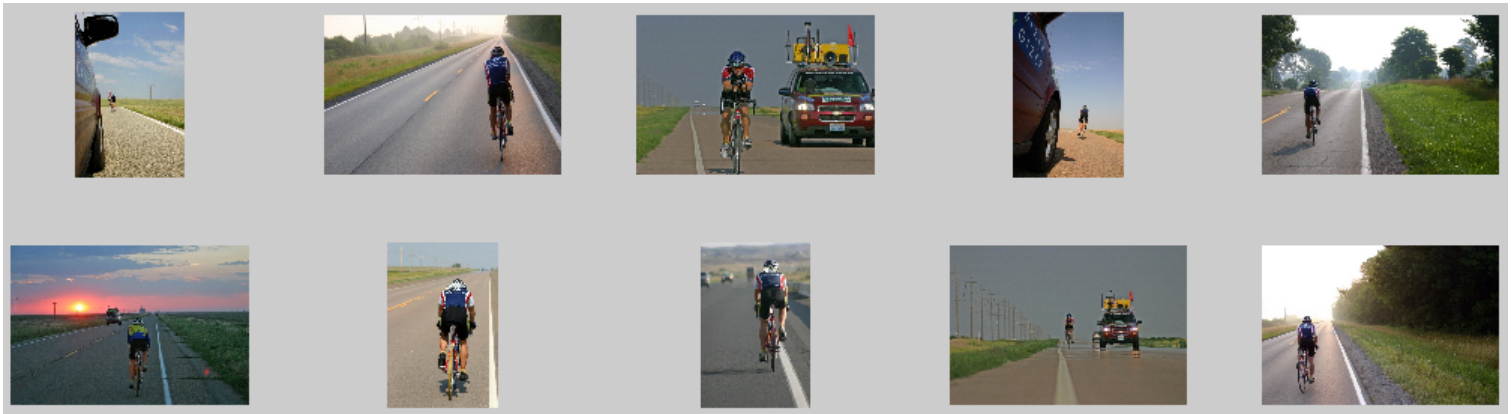
Experimento	P5	P10	P20	MAP
TIA-TXTIMG	0.4769	0.4538	0.3910	0.2359
F1-L1.5	0.4923	0.4744	0.4026	0.2521
F1-L0.0	0.4667	0.3923	0.3372	0.2325
F3-L1.5	0.5026	0.4795	0.4128	0.2531
F3-L0.0	0.4103	0.3769	0.3295	0.2229
F5-L1.0	0.4821	0.4590	0.4000	0.2558
F5-L0.0	0.4718	0.4128	0.3513	0.2381
F8-L1.0	0.4872	0.4769	0.4103	0.2590
F8-L0.0	0.4615	0.4359	0.3641	0.2385
F10-L0.5	0.5026	0.4744	0.3974	0.2588
F10-L0.0	0.4564	0.4282	0.3538	0.2291

Resultados (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*.



(a)



(b)

Resultados (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*

(a)

(b)

Conclusiones

- En el mejor de los casos el método propuesto mejoró en 63 % a la lista original, y en el peor un 7 % utilizando la retroalimentación de relevancia simulada.
- Los mejores resultados se obtienen al combinar ambos potenciales, en la mayoría de los casos dándole más peso a V_c.
- Para algunas consultas, la información textual de la imagen no es suficiente para separar las imágenes relevantes del resto.

Gracias por su atención

¿Alguna pregunta o comentario?