

EFFICIENT VISUAL SEARCH OF VIDEOS

→ Sivic & Zisserman, IEEE-PAMI 2009

Ejemplo : en google buscar 'videogoogle'

Explore movie Charade

zervs.robots.ox.ac.uk

→ ver frame 18300

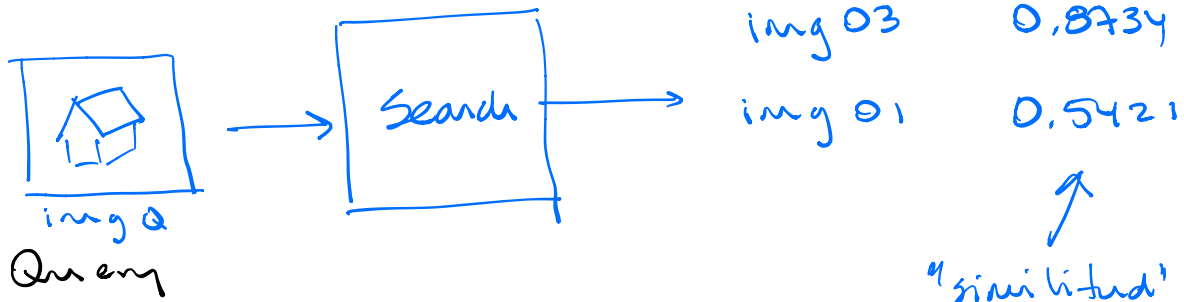
→ seleccionar Lufthansa

¿Cómo la búsqueda es tan rápida?

Método para encontrar fotos parecidas (basado en paper mencionado).



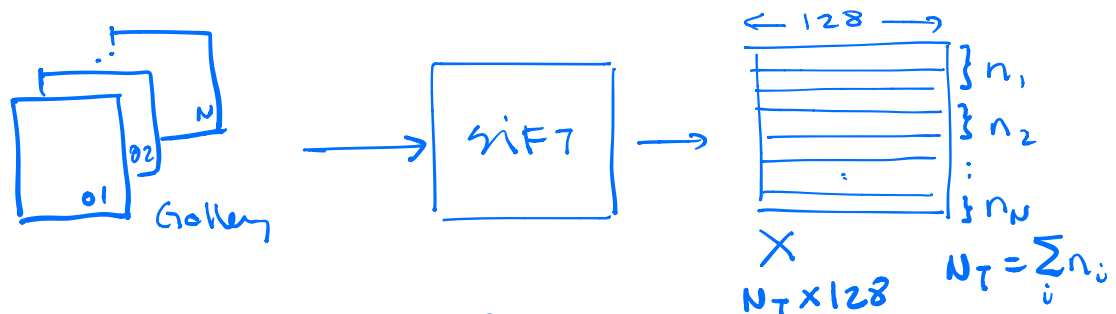
Gallery



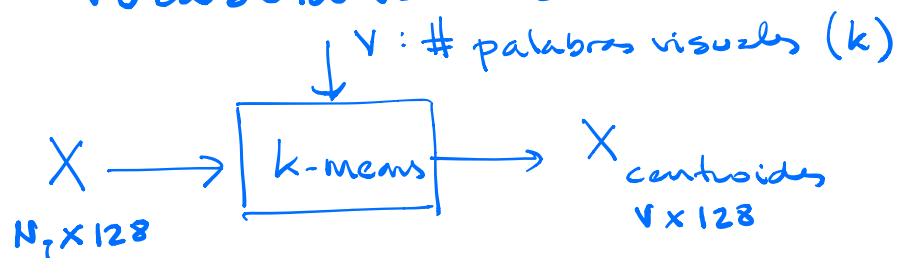
Paso 1: Extraer keypoints y correspondientes descriptores para cada imagen

Para cada imagen de la Galería:

- Detectar SIFT keypoints
- Extraer SIFT descriptores



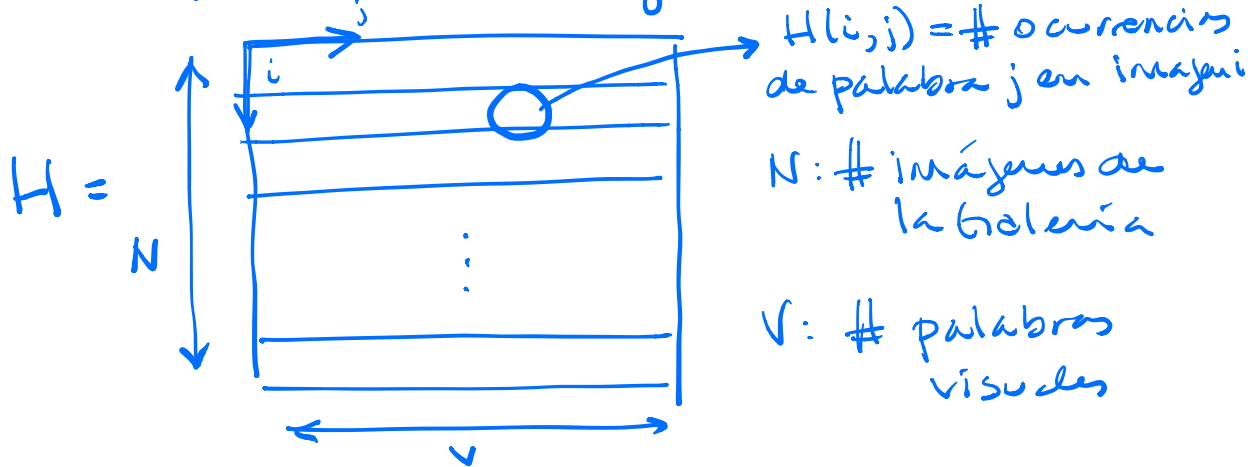
Paso 2: Construcción de Vocabulario Visual



Ej: $V = 1000$

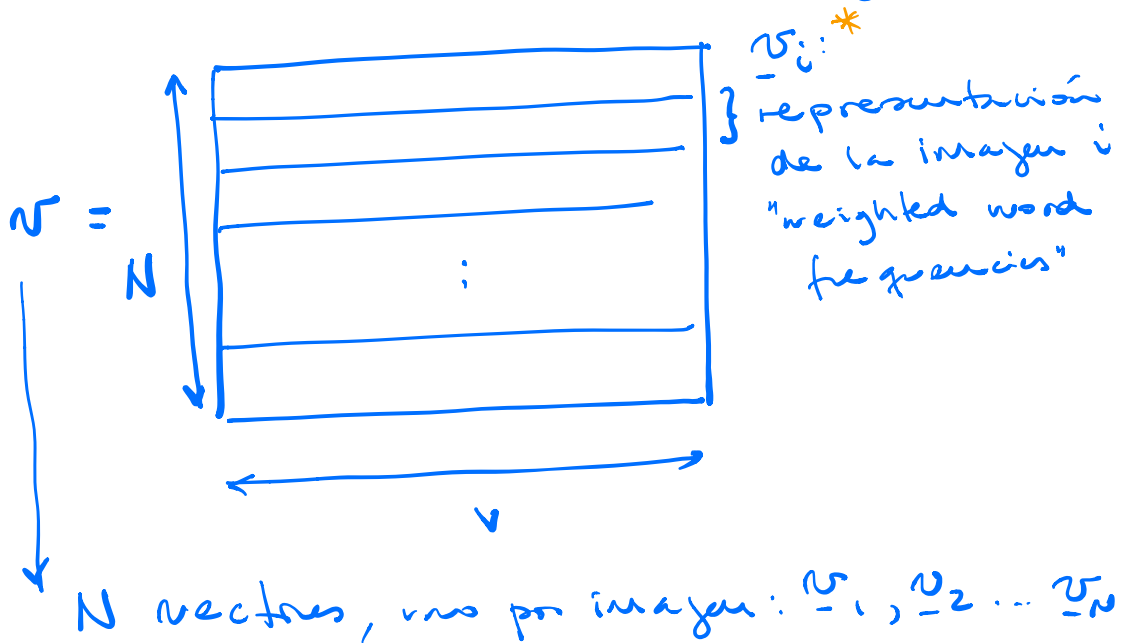
Paso 3:

- Construcción de Histograma de Frecuencias



Paso 4:

- Construcción de Visual Indexing



$$* \underline{v}_i = [t_1 \ t_2 \ \dots \ t_j \ \dots \ t_v]_i \quad \begin{matrix} i = 1 \dots N \\ j = 1 \dots V \end{matrix}$$

$$t_j = \left(\frac{n_{ij}}{n_i} \right) \left(\log \frac{N}{N_j} \right)$$

relevancia en imagen i relevancia en la galeria

Term Frequency -
Inverse Document
Frequency
 $t_f - i_d f$

donde:

n_{ij} : número de ocurrencias de la palabra j en la imagen i
 $= H(i, j)$

n_i : número de palabras encontradas en la imagen i
 $= \sum_j n_{ij}$

N : número de imágenes en la Galería

N_j : número de imágenes que tiene la palabra j
 $= \sum_i (n_{ij} > 0)$

$\frac{n_{ij}}{n_i}$: Word Frequency Ej: 0.5 significa que la mitad de las palabras de la imagen i son j .

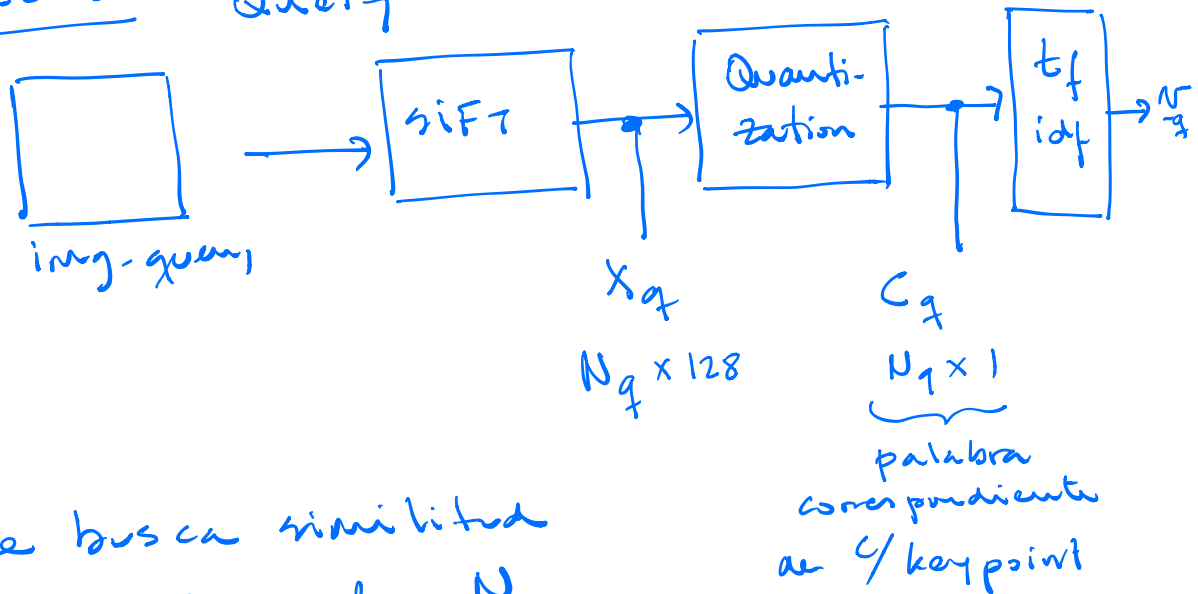
$\log \frac{N}{N_j}$: Index Document Frequency

Si $\frac{N}{N_j}$ es alto \Rightarrow palabra j es discriminativa
 (Si N_j/N es bajo)

E_j : $\frac{N}{N_j} = 1$: todos los imágenes contienen esta palabra, no sirve para diferenciar entre una imagen y otra: $\log(N/N_j) = 0$

LENTO OFF-LINE $\frac{N}{N_j} = 20$: 1 de cada 20 imágenes contiene esta palabra
 RÁPIDO ON-LINE $\log(N/N_j) \approx 3$

Paso 5: Query



Se busca similitud entre \underline{v}_q y los N vectores de la galería:

$$\text{score}_i = \frac{\langle \underline{v}_q, \underline{v}_i \rangle}{\|\underline{v}_q\| \|\underline{v}_i\|} \quad i = 1 \dots N$$

y se ordenan de mayor a menor.

→ ver Ejemplo PAT07-seq-query.m

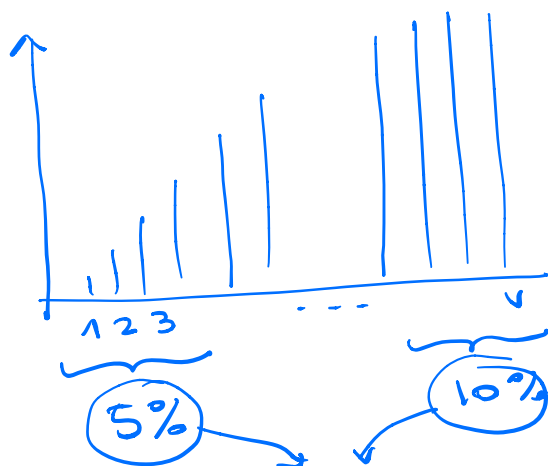
OPTIONAL: STOP LIST

Se eliminan del vocabulario las palabras no "discriminativas" y las "muy ruidosas" porque no aportan, o confunden.

El vocabulario tiene V palabras

n_{ij} : # ocurrencias de la palabra j en la imagen i

$$N_j = \sum_i (n_{ij} > 0) : \# \text{ imágenes de la galería que contienen palabra } j$$



← $\text{sort}(N_i)$
'ascending'

'stopped' : se eliminan