



Multimodal Image Retrieval Based on Keywords and Low-level Image Features

Miran Pobar, Marina Ivašić-Kos

Department of Informatics, University of Rijeka

1st International KEYSTONE Conference IKC 2015
Coimbra Portugal, 8-9 September 2015

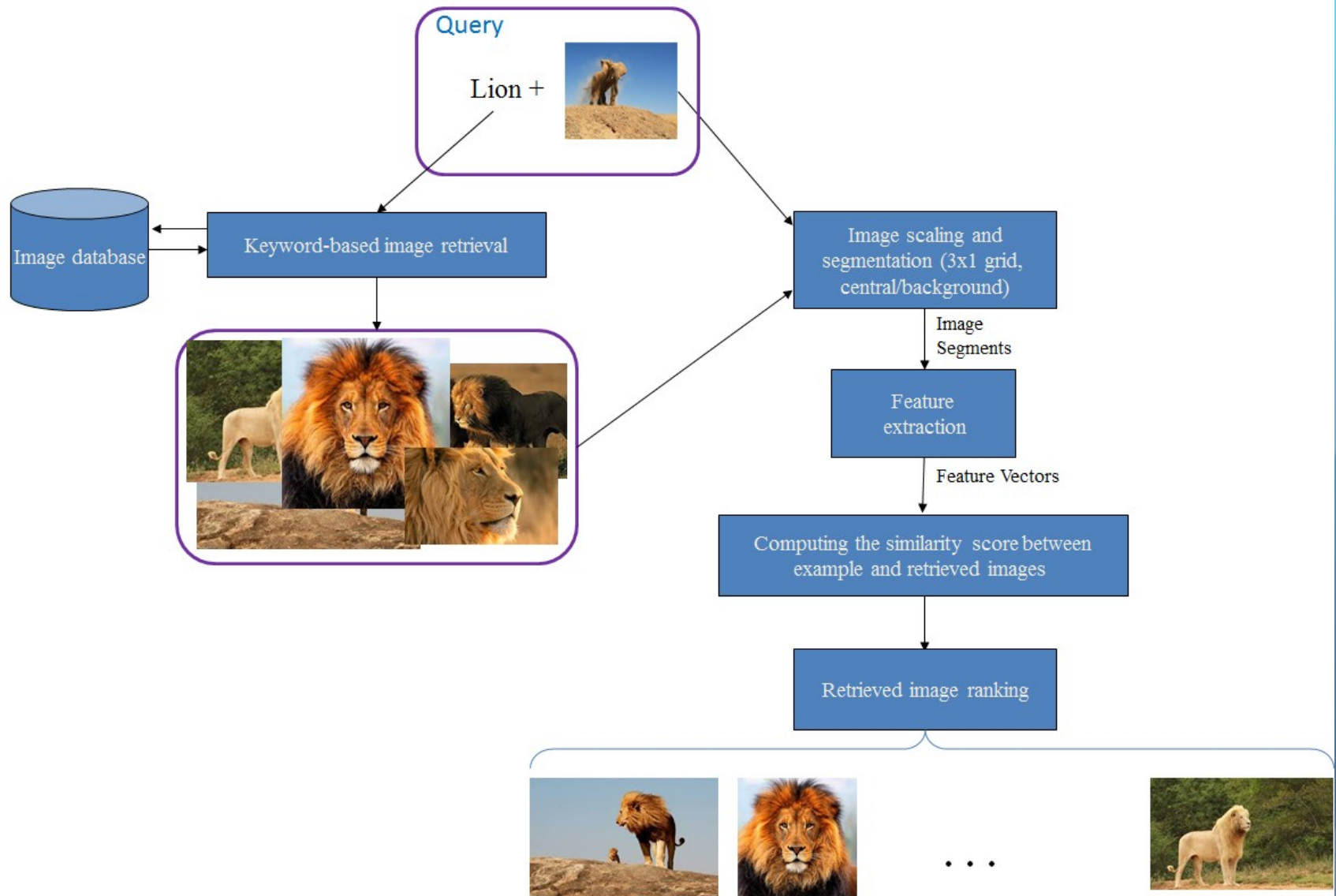
Outline

- Introduction
- Multimodal image retrieval framework
- Low level image features
- Content-based similarity
- Experiments
- Conclusion

Introduction

- **content-based image retrieval**
 - compare visual content - low-level image features
 - ranking based on visual similarity to a query image
 - appropriate for images with same semantics - medical images, criminalistics,...
- **text-based image retrieval**
 - relies on image annotation
 - matching text descriptions of images
 - easier in many everyday cases

Multimodal image retrieval framework



Low-level Image Features

- **pixel-based descriptors**

- color histograms, dominant colors
- robust in position, translation and rotation changes
- useful for rapid detection of objects in image databases
- to preserve the information about the color layout, computed on:
 - the whole image
 - on two centrally symmetric regions,
 - regions obtained by applying a 3x1 grid.

- **structure-based descriptors**

- GIST

Content-based similarity ranking

- Histogram comparison:
 - Bhattacharyya distance
 - histogram intersection
 - chi-squared histogram matching distance
- Dominant colors:
 - Jaccard distance
- GIST features:
 - Euclidean distance.

$$D_B(p, q) = -\ln(BC(p, q)),$$

$$BC(p, q) = \sum_{i=1}^n \sqrt{p_i q_i},$$

$$\text{histint}(p, q) = \sum_{i=1}^n \min(p_i, q_i),$$

$$X^2(p, q) = \frac{1}{2} \sum_{i=1}^n \frac{[p_i - q_i]^2}{p_i + q_i},$$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

$$D_J(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

Experiments

- keyword vocabulary of 27 natural and artificial objects ('airplane', 'bird', 'lion', 'train', ...)
- Data sets:
 - Images of natural scenes, annotated with a vocabulary keyword + other - multiple keywords
 - Corel image dataset
 - 500 images
 - Professional photographers
 - Flickr image dataset
 - 2700 images
 - Amateur photographers

Keyword-based retrieval

Keyword: tiger



Content based retrieval

Histogram intersection



target



$d = -2.8414$



$d = -2.8116$



$d = -2.8116$



$d = -2.7655$



$d = -2.75$



$d = -2.7425$



$d = -2.7303$



$d = -2.6945$

Multimodal retrieval

Keyword: lion

Euclidean distance, ...



target



$d = 1.1468$



$d = 1.147$



$d = 1.164$

Jacard distance



target



$d = 0.17647$



$d = 0.27778$



$d = 0.27778$

Multimodal retrieval

Keyword: tiger

Histogram intersection



target



$d = -2.8414$



$d = -2.5473$



$d = -2.1912$

Chi-squared distance



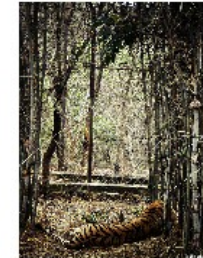
target



$d = 0.0053693$



$d = 0.013721$



$d = 0.019461$

Bhattacharyya distance



target



$d = 1.6499$



$d = -1.6233$



$d = -1.5881$

Multimodal retrieval

Keyword: wolf



target



$d = -1.5209$



$d = -1.4987$



$d = -1.4818$



target



$d = -1.6526$



$d = -1.6465$



$d = -1.6419$

Conclusion and Future Work

- **Proposed multimodal image retrieval framework**
 - integrating keyword-based image search with content-based ranking according to the visual similarity to a query image
- **Future work:**
 - more formal evaluation of features and measures for the task of image retrieval in general domain
 - Similar performance in top 10 results.
 - integrate automatic image annotation with multimodal image retrieval.