

How to spot a Dalmatian in a pack of dogs

A data-driven approach to searching unannotated images using natural language

Jonathon S. Hare¹, Paul H. Lewis¹, Peter G. B. Enser² & Christine J. Sandom²

(1) School of Electronics and Computer Science, University of Southampton

(2) School of Computer, Mathematical and Information Sciences, University of Brighton

MOTIVATION: THE SEMANTIC GAP

Much of the previous work on Content Based Image Retrieval uses a "Query By Example" paradigm and operates at the feature vector level. This is sometimes successful. However, it often fails due to the gap between feature vectors and semantics.

This semantic gap is the gap between information extractable automatically from the visual data and the interpretation a user may have for the same data. Typically this is the gap between low level features and the high-level semantics in which queries are posed.

In "The Bridging of the Semantic Gap in Visual Information Retrieval" project we are exploring how test-bed ontologies combined with content based techniques and annotation can help meet the needs of real users in limited domains.

Our research within the project has been two-fold; We have been working with a number of organisations who hold large image collections, and investigating and categorising how real queries are made and results retrieved. The second part of the project has been developing computational techniques that facilitate image retrieval, using ontological approaches when meta-data is available, and automatic annotation techniques for the cases where there is an absence of meta-data.

CASE STUDY: THE KENNEL CLUB PICTURE LIBRARY

One of our case studies has been in collaboration with the Kennel Club Picture Library. The picture library's image collection is relatively small, consisting of about 60,000 images. Of those 60,000, slightly over 7000 have been digitised, and about 3000 of those have subject metadata, mostly consisting of attached keywords. The remaining 4000 digitised images have no extra information.

The collection is growing rapidly. Each year, after the Crufts dog show they expect to receive an additional 4000-8000 (digital) images with no metadata other than date/time (and only then if the camera is set-up correctly). Manually annotating all of these images with subject metadata would be a very time consuming task.

Typical requests are for pictures of particular breeds of dog, however, some requests are more abstract:

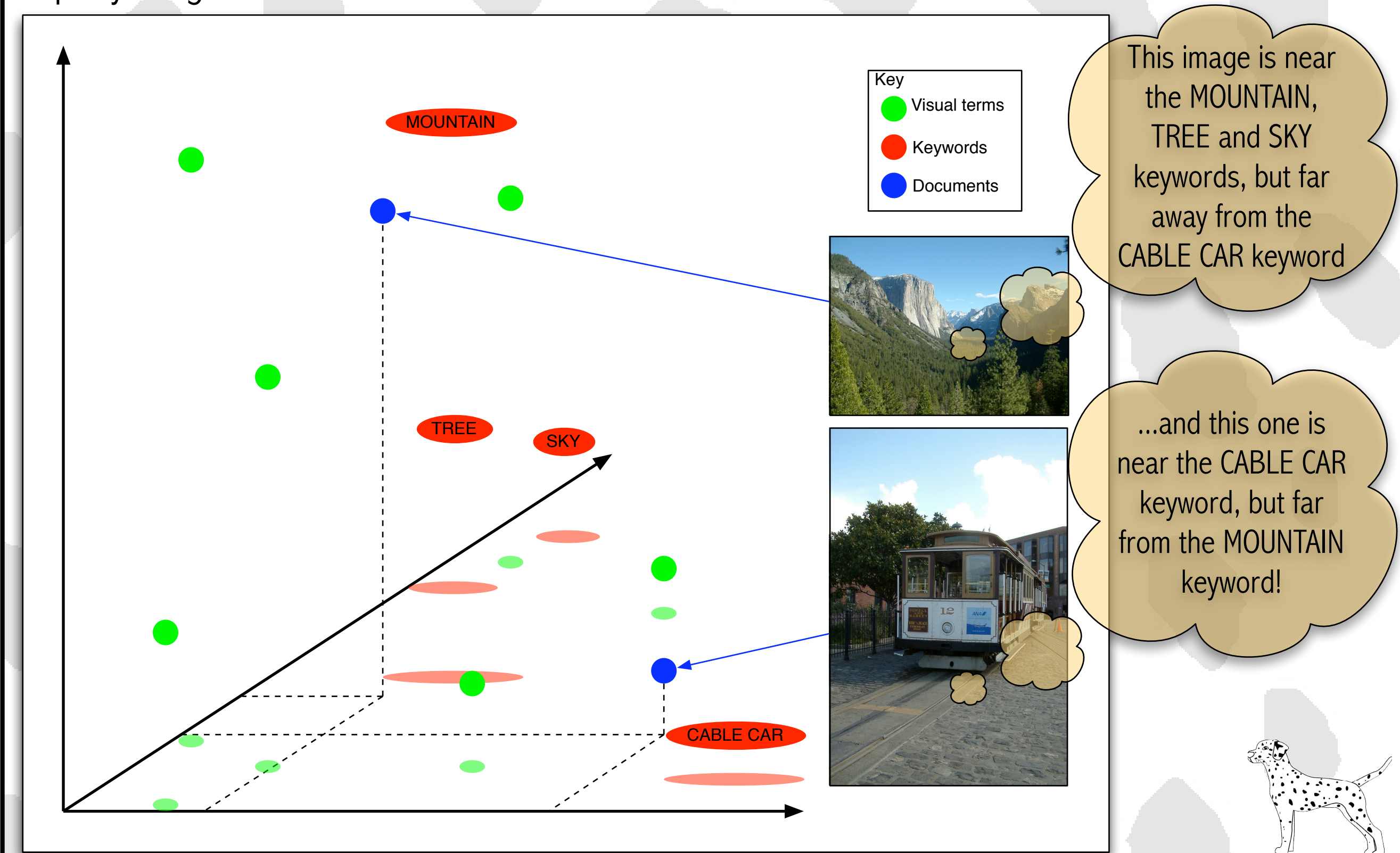
- "I need a photo of a cute dog, preferably a Heinz variety"
- "...some irresponsible parents still purchase dogs for presents for children. I was hoping you may have some images which we could use to illustrate the article piece"

TECHNOLOGY: THE SEMANTIC SPACE

Our Semantic Space approach to multimedia retrieval is based on a generalisation of an information retrieval technique called Cross-Language Latent Semantic Indexing (CL-LSI).

Conceptually, a Semantic Space is simply a large, multidimensional vector space in which images and terms (both keywords and visual-terms describing the images' content) representing the images, are placed. The placement of these images and terms is such that the images are placed 'near' to the terms that describe them.

Thus, in order to retrieve images given a keyword query, all that needs to be done is to locate the keywords coordinates in the space and find all the images that share similar coordinates. Of course, the space can also be used in reverse to determine which keywords most likely describe a query image.



WOOF!

Our "Poodle" and "K9" testbed image search engines allow the Kennel Club to search their annotated images using keywords with and without ontological mediation, and their un-annotated images using Semantic Space and content-based techniques.

