A large scale system for searching and browsing images from the World Wide Web

Alexei Yavlinsky[1], Daniel Heesch[2], and Stefan Rüger[1]

[1]Department of Computing, South Kensington Campus
Imperial College London, London SW7 2AZ, UK
[2]Department of Electrical and Electronic Engineering, South Kensington Campus
Imperial College London, London SW7 2AZ, UK
{alexei.yavlinsky, daniel.heesch, s.rueger}@imperial.ac.uk

Abstract. This paper outlines the technical details of a prototype system for searching and browsing over a million images from the World Wide Web using their visual contents. The system relies on two modalities for accessing images — automated image annotation and NN^k image network browsing. The user supplies the initial query in the form of one or more keywords and is then able to locate the desired images more precisely using a browsing interface.

1 Introduction

The purpose of this system is to demonstrate how simple image feature extraction can be used to provide alternative mechanisms for image retrieval from the World Wide Web. We apply two recently published indexing techniques — automated image annotation using global features [1] and NN^k image network browsing [2] — to 1.14 million images spidered from the Internet. Traditional image search engines like Google or Yahoo use collateral text data, such as image filenames or web page content, to index images on the web. Such metadata, however, can often be erroneous and incomplete. We attempt to address this challenge by automatically assigning likely keywords to an image based on its content and allowing users to query with arbitrary combinations of these keywords.

As the vocabulary used for automatically annotating images is inherently limited, we use NN^k image networks to enable unlimited exploration of the image collection based on inter-image visual similarity. NN^k networks have proven to be a powerful browsing methodology for large collections of diverse images [3]. The idea is to connect an image to all those images in the collection to which it is most similar under *some* instantiation of a parametrised distance metric (where parameters correspond to feature weights). This is unlike most image retrieval systems which fix the parameters of the metric in advance or seek to find a single parameter set through user interaction. By considering all possible parameter sets, the networks provide a rich and browsable representation of the multiple semantic relationships that may exist between images. In addition to showing the local graph neighbourhood of an image we extract a number of visually similar subgraphs in which that image is contained thus providing users with immediate access to a larger set of potentially interesting images.

Early experiments with our system are showing promising results, which is particularly encouraging given the 'noisy' nature of images found on the World Wide Web. In the next section we give short, formal descriptions of both indexing frameworks, and we conclude with a number of screenshots.

2 Large scale image indexing

2.1 Automated image annotation

We use a simple nonparametric annotation model proposed by [1] which is reported to perform on par with other, more elaborate, annotation methods.

14,081 images were selected from the Corel Photo Stock for estimating statistical models of image keywords, which were then used to automatically annotate 1,141,682 images downloaded from the internet. We compiled a diverse vocabulary of 253 keywords from the annotations available in the Corel dataset.

Global colour, texture, and frequency domain features are used to model image densities. The image is split into 9 equal, rectangular tiles; for each tile we compute the mean and the variance of each of the HSV channel responses, as well as Tamura coarseness, contrast and directionality texture properties obtained using a sliding window [4]. Additionally we apply a Gabor filter bank [5] with 24 filters (6 scales $\times$ 4 orientations) and compute the mean and the variance of each filter's response signal on the entire image. This results in a 129-dimensional feature vector for each image. Our choice of these simple features is motivated by results reported in [1] which demonstrate that simple colour and texture features are suitable for automated image annotation. Implementation details of Tamura and Gabor features used in this paper can be found in [6].

2.2 NN^k networks

NN^k Networks were introduced in [2] and analysed in [7] and [8]. They are based on a simple idea: instead of ranking images with respect to another image according to their similarity under some fixed metric, the metric is parametrised in terms of feature-specific weights and for *each* parameter setting the image is recorded that is closest under that particular instantiation of the metric. This set of top-ranked images are referred to as the NN^k. NN stands for nearest neighbours and k denotes the dimensionality of the feature space. The distance metric is a weighted sum of feature-specific distances.

$$D(p,q) = \sum_{f=1}^{k} w_f d_f(p,q).$$

Given a collection of images, we can use the NN^k idea to build image networks by establishing an arc between image q and p if p is the NN^k of q. The number of weights for which p is the NN^k of q defines the strength of the arc. The set of NN^k can be thought of as exemplifying the different semantic facets of the focal image that lie within the representational scope of the chosen feature set.

Structurally NN^k networks resemble the hyperlinked network of the World Wide Web (WWW), but they tend to exhibit a much better connectedness with only a negligible fraction of vertices not being reachable from the giant component. In collections of one million images, the average number of links between any two images lies between 4 and 5. Importantly, by being precomputed NN^k networks allow very fast interaction.

A soft clustering of the images in the networks is achieved by partitioning not the vertex set but the edge set. An image can then belong to as many clusters as it has edges to other images (for details see [9]).

3 Implementation

The search engine is implemented within the JavaServerPages framework and is served using Apache Tomcat[1]. Figure 1 shows the result of a query `‘tower structure sky’`. Below each image there are two links, one for using the image as a starting point for exploring the NN^k network and the other for viewing different clusters to which the image belongs. Figure 2 shows two steps in the image network after the second search result from the left has been selected as the focal point.

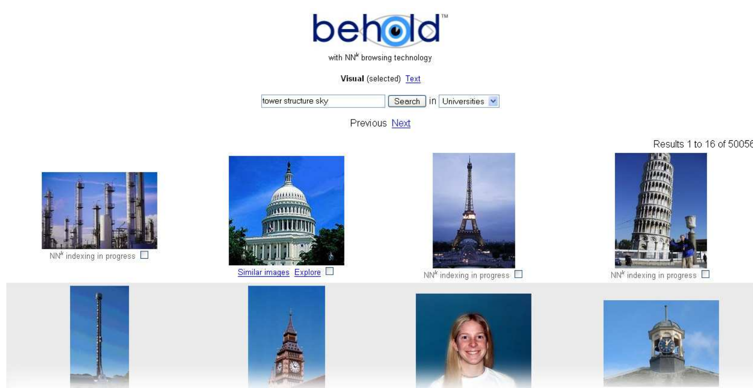

Fig. 1. Initial search using keywords (query: `‘tower structure sky’`)

4 Conclusions

Automated image annotation and image network browsing techniques appear to be promising for searching and exploring large volumes of images from the World Wide Web. It is particularly encouraging that representing images using very

[1] A live version of this demo can be found at `http://www.beholdsearch.com`

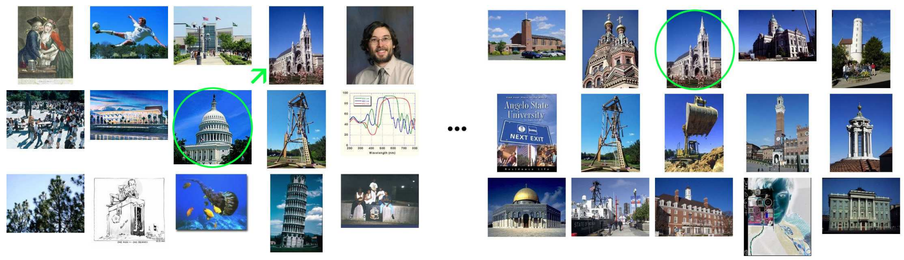

Fig. 2. Two steps of the NN^k browsing, circle indicates the focal image, other images are its neighbours

simple global features often yields meaningful search and visualisation results. Additionally, since most of the computation is done offline, the system is highly responsive to user queries — a desireable attribute for a content based image retrieval system.

References

1. A Yavlinsky, E Schofield, and S Rüger. Automated image annotation using global features and robust nonparametric density estimation. In *Proceedings of the International Conference on Image and Video Retrieval*, pages 507–517. LNCS 3568, Springer, 2005.
2. D Heesch and S Rüger. NN^k networks for content-based image retrieval. In *Proc European Conf Information Retrieval*, pages 253–266. LNCS 2997, Springer, 2004.
3. D Heesch, M Pickering, A Yavlinsky, and S Rüger. Video retrieval within a browsing framework using keyframes. In *Proc TREC Video*, 2004.
4. H Tamura. Texture features corresponding to visual perception. *IEEE Transactions. Systems, Man and Cybernetics*, 8(6):460–473, 1978.
5. B Manjunath and W-Y Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
6. P Howarth and S Rüger. Evaluation of texture features for content-based image retrieval. In *Proceedings of the International Conference on Image and Video Retrieval*, pages 326–334. LNCS 3115, Springer, 2004.
7. D Heesch and S Rüger. Three interfaces for content-based access to image collections. In *Proc Int'l Conf Image and Video Retrieval*, pages 491–499. LNCS 3115, Springer, 2004.
8. D Heesch and S Rüger. Image browsing: A semantic analysis of NN^k networks. In *Proc Int'l Conf Image and Video Retrieval*, pages 609–618. LNCS 3568, Springer, 2005.
9. D Heesch. *The NN^k technique for image searching and browsing*. PhD thesis, Imperial College London, 2005.