ABSTRACT

Human-computer interaction is increasingly recognised to be an indispensable component of image retrieval systems. A typical form of interaction is that of relevance feedback whereby users supply relevance information on the retrieved images. This information can subsequently be used to optimise retrieval parameters. The first part of the chapter provides a comprehensive review of existing relevance feedback techniques and also discusses a number of limitations that can be addressed more successfully in a browsing framework. Browsing models therefore form the focus of the second part of this chapter where we will evaluate the merit of hierarchies and networks for interactive image search. This exposition aims to provide enough detail to enable the practitioner to implement most of the techniques and to find ample pointers to the relevant literature otherwise.

Interaction models and relevance feedback in image retrieval

Keywords
information retrieval, content-based image retrieval, query by example, human-computer interaction, relevance feedback, browsing, image networks, image polysemy

Daniel Heesch

Department of Electrical and Electronic Engineering

South Kensington Campus

Imperial College London

Tel: +44 (0) 207 594 6165

Fax: +44 (0) 207 594 6302

Email: daniel.heesch@imperial.ac.uk

Stefan Rueger

Department of Computing

South Kensington Campus

Imperial College London

Tel: +44 (0)20 7594 8298

Fax: +44 (0)20 7581 8024 SW7 2AZ

Email: srueger@doc.ic.ac.uk

INTRODUCTION

Similarity in appearance is often revealing about other, and potentially much deeper, functional and causal commonalities between objects, events and situations. Things that are similar in some respect are likely to behave in a similar way and owe their existence to similar causes. It is because of this regularity that similarity is fundamental to many cognitive tasks such as concept learning, object recognition and inductive inference.

Similarity-based reasoning requires efficient modes of retrieval. It is perhaps only in experimental settings that subjects have direct sensory access to the patterns that they are asked to compare. In most situations an observed pattern is evaluated by comparing it with patterns stored in memory. The efficiency with which we can classify and recognise objects suggests that the retrieval process is itself based on similarity. According to Steven Wolfram (2004) the use of memory "underlies almost every major aspect of human thinking. Capabilities such as generalization, analogy and intuition immediately seem very closely related to the ability to retrieve data from memory on the basis of similarity." He extends the ambit of similarity-based retrieval to the domain of logical reasoning, which ultimately involves little more than "retrieving patterns of logical argument that we have learned from experience" (p. 627).

The notion of similarity is clearly not without problems. Objects may be similar on account of factors that are merely accidentals and that, in fact, shed no light on the relationship that one could potentially unveil. The problem of measuring similarity largely reduces, therefore, to one of defining the set of features that matter. The problem of estimating the relative significance of different features is one of information retrieval in general. It is however greatly compounded in the case of image retrieval in two significant ways: First, documents readily suggest a representation in terms of its constituent words. Images do not suggest such a natural decomposition into semantic atoms with the effect that image representations are to some extent arbitrary. Secondly, images typically admit to a multitude of different meanings. Each semantic facet has its own set of supporting visual features and a user may be interested in any one of them.

These challenges have traditionally been studied in the context of QBE. In this setting the primary role of users is to formulate a query, the actual search is taken care of by the computer. This

division of roles has its justification in the observation that the search is the computationally most intensive part in the process, but is questionable on the grounds that the task of recognising relevance is still best solved by the human user. The introduction of relevance feedback into QBE systems turns the problem of parameter learning into a supervised learning problem. Feedback on retrieved images can help to find relevant features or better query representations. Although the incorporation of relevance feedback techniques can result in substantial performance gains, it does not overcome the more fundamental limitations of the QBE framework in which they have been formulated. Often users may not have an information need in the first place and wish to explore an image collection. Moreover, the presence of an information need does not mean that a query image is readily at hand to describe it. Also brute force nearest neighbour search is linear in the collection size and the sub-linear performance achieved through hierarchical indexing schemes does not extend to high dimensional features spaces with more than 10 dimensions.

Browsing provides an interesting alternative to QBE but has, by comparison, received surprisingly scant attention. Browsing models for image search tend to cast the collection into some structure that can be navigated interactively. Arguably one of the greatest difficulties of the browsing approach is to identify structures that are conducive to effective search in the sense that they support fast navigation, provide a meaningful neighbourhood for choosing a browsing path and allow users to position themselves in an area of interest.

The first part of this chapter will examine relevance feedback models in the QBE setting. After a brief interlude in which we discuss limitations of the QBE framework, we shift the focus to browsing models that promise to address at least some of these. Each section concludes with a summary table that juxtaposes many of the works that have been discussed.

QUERY BY EXAMPLE SEARCH

Query by example systems return a ranked list of images based on similarity to a query image. Relevance feedback in this setting involves users labelling retrieved images depending on their perceived degree of relevance. Relevance feedback techniques vary along several dimensions which makes any linear exposition somewhat arbitrary. We structure the survey according to how relevance feedback is used to update system parameters: *Query adaptation* utilises relevance information to compute a new query for the next round of retrieval. *Metric optimisation* involves an update of the distance function that is used to compute the visual similarities between the query and database images. *Classification* involves finding a decision function that optimally separates relevant from non-relevant images.

Query adaptation

Query adaptation describes the process whereby the representation of an initial query is modified automatically based on relevance feedback. Query adaptation was among the first relevance feedback techniques developed for text retrieval (Salton & McGill, 1982) and has since been adapted to image retrieval (Rui, Huang & Mehrotra, 1997; Ishikawa, Subramanya & Faloutsos, 1998; Porkaew, Chakrabarti & Mehrotra, 1999; Zhang & Su, 2001; Aggarwal, Ashwin and Ghosal, 2002; Urban, Jose & Rijsbergen, 2003; Kim & Chung, 2003). The two most important types of query adaptation are query point moving and query expansion. We will be dealing with each in turn.

Query point moving

Query point moving is a simple concept and illustrated in Figure 1. Relevant (+) and non-relevant (o) objects are displayed in a two-dimensional feature space with the query initially being in the bottom right quadrant (left plot). The images marked by the user correspond to the bold circles. The goal of query point moving is to move the query point towards the relevant images and away from the non-relevant images. Clearly, if relevant images form clusters in feature space, this technique should improve retrieval performance in the next step (right plot). Techniques differ in how exactly this movement in feature space is achieved.

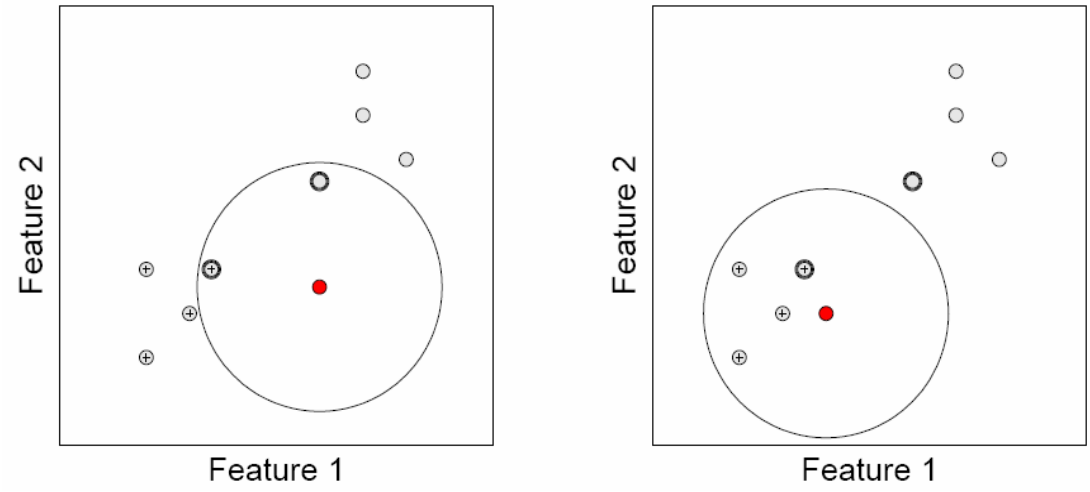

Figure 1 Moving the query point towards positive examples.

In Urban et al.(2003), for example, images are represented in terms of a set of keywords and a colour histogram. Given a set of images for which the user has indicated some degree of relevance, the visual part of the query is computed as the weighted average over the relevant images. Meanwhile, Rui et al. (1997) the representation of the query is altered using both relevant and non-relevant images. The method employs Rocchio's formula (1971) originally developed for text retrieval. In particular, given sets R and N of feature vectors of relevant and non-relevant images, respectively, the learned query vector is computed as

$$q(t+1) = \alpha q(t) + \beta\left(\frac{1}{|R|}\sum_{x\in R} x\right) - \gamma\left(\frac{1}{|N|}\sum_{x\in N} x\right),$$

where α, β and γ are parameters to be chosen. For $\alpha = \gamma = 0$, the new query representation is the centroid of the relevant images. The goal of the method is to move the query point closer towards relevant images and further away from non-relevant images.

Yet another approach is taken by Ishikawa et al. (1998) and Rui and Huang (2000) who find the best query point as that point that minimises the summed distances to all relevant images. The optimal query representation turns out to be the weighted average of the representations of all relevant images.

Query expansion

Query point moving suffers from a notable limitation: If relevant images form visually distinct subsets (corresponding to multiple clusters in feature space), the technique may easily fail to move the query into a better position as relevant images suggest multiple and mutually conflicting directions. The problem arises from the requirement to cover multiple clusters with only one query. A simple modification of the above approach that alleviate this problem involves replacing the original query point by multiple query points each located near different subsets of the relevant images. This modification turns query point moving into what may more aptly be described as query expansion (Porkaew et al, 1999; Kim & Chung, 2003; Urban & Jose, 2004a. Again, differences between techniques are down to details, in particular to the question of how to choose the precise locations of the query points.

In Porkaew et al (1999), for example, relevant images are clustered and the cluster centroids chosen as new query points. The overall distance of an image to the multi-point query is computed as the weighted average over the distances to each query point, i.e.

$$D(x,Q) = \sum_{q \in Q} w_q d(x,q) .$$

The weights are taken to be proportional to the number of relevant images in each cluster. Thus, query points that seem to represent the user need better have a greater weight in the overall distance computation. One should note, however, that this scheme retains the feature it seeks to overcome by linearly averaging over individual distances. In fact, it can be shown that the overall result is equivalent to query point moving where the new query point is given by $\sum_{q \in Q} w_q q$. If d is taken to be the Euclidean distance, for example, then the iso-distance lines for a multi-point query remain circles now centred at the new query point.

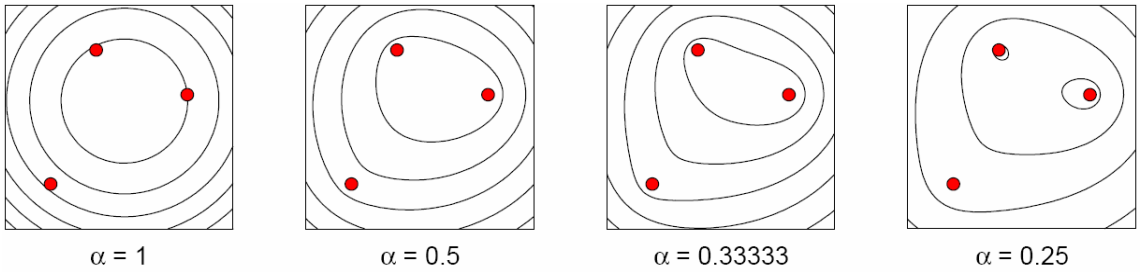

Figure 2 Multi-point queries: From query point moving towards disjunctive queries.

This is shown on the left plot in Figure 2. The method reduces therefore to the solution suggested by Ishikawa et al (1998) and Rui and Huang (2000). To properly account for the cluster structure it seems more reasonable to treat the multi-point query as a disjunctive query, an approach taken in more recent work by Wu, Faloutsos, Sycara and Payne (2000), Kim and Chung (2003) and Urban and Jose (2004a). In the second work, for example, the others suggest a remedy in the form of

$$D(x,Q) = \sum_{q \in Q} d(x,q)^{\alpha}$$

where q is the representation of a cluster and α is a parameter that confers the desired non-convexity. The iso-distance lines of the model with α ranging from 1 to 1/4 are shown in Figure 2. For α = 1, the model reduces to that of Porkaew et al. (1999).

The above methods have in common that they are concerned with combining distance scores. An alternative approach to multi-point queries is taken by Urban and Jose (2004) who adapt a rank-based aggregation method, median-rank aggregation (Fagin, Kumar & Sivakumar, 2003), to multi-point image queries and establish superior performance over the simple score-based method of Porkaew et al (1999). The rank of an image p is the number of images whose distance scores are smaller or equal to that of p. This particular method involves computing the image ranks with respect to each of the different query points. The median of those ranks becomes the final rank assigned to that image. If an image has a large distance to only a small number of query points, these will have no effect on the final rank of that image. This provides support for more disjunctive queries, but the method is equally robust against unusually small distances.

Distance metric optimisation

Introduction

The second large group of relevance feedback techniques is concerned with modifying the distance metric that is used to compute the visual similarities between the query and the database images. As noted earlier, one of the problems pertaining to the notion of similarity relates to the question of how to weigh different features. A feature that is good at capturing the fractal characteristics of natural scenes may not be good for distinguishing between yellow and pink roses. How can we infer

from relevance feedback which feature is important and which one isn't?

A naive method of computing the distance between two representations is to concatenate all individual feature vectors into one and measure the distance between two vectors x and y. In hierarchical models, distances are computed between individual features and the resulting distances are aggregated. In both models, we have to compute the distance between two vectors. In image retrieval, commonly used distance metrics are instances of the general Minkowski metric,

$$D(x,y) = \left[\sum_i | x_i - y_i |^{\alpha} \right]^{\frac{1}{\alpha}}, \alpha > 0.$$

This reduces to the Euclidean metric for $\alpha = 2$ and to the L_1 metric for $\alpha = 1$. The advantage of the Minkowski metric is that it can readily be parameterised by adding a weight to each component-wise difference. For $\alpha = 2$, we obtain a weighted Euclidean distance

$$D(x,y) = \left[\sum_i w_i (x_i - y_i)^2 \right]^{\frac{1}{2}},. \qquad (1.1)$$

where one typically constrains the weights to sum to one and to be non-negative. Relevance feedback can now help to adjust these weights so that relevant images tend to be ranked higher in subsequent rounds. The idea is illustrated in Figure 3. Under the weighted Euclidean metric with equal weights the iso-distance lines in a two-dimensional vector space are circles centred at the query vector. In this example, the one image marked relevant is much closer to the query with respect to the second feature. On the assumption that a relevant image has more relevant images in its proximity, we wish to discount distances along the dimension along which the relevant image differs most from the query. Here this is achieved by decreasing the weight for feature 1 with the effect that the iso-distance lines become ellipsoids with their long axes parallel to that of the least important feature.

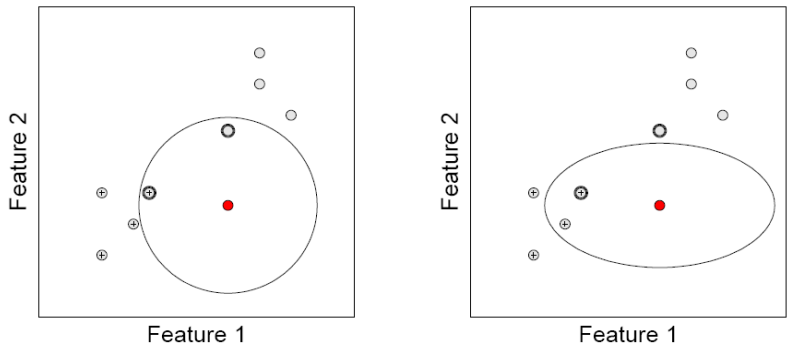

Figure 3: Changing parameters of the metric.

In hierarchical models these distances need to be combined. By far the most popular aggregation method is the weighted linear sum that we encountered earlier in the context of multi-point queries,

$$D(x,q) = \sum_{i=1}^{k} w_i d_i(x,q), \tag{1.2}$$

where we now sum over k features rather than over query points. The great majority of relevance feedback methods are concerned with adjusting weights of each individual feature component either in a flat (Ishikawa et al., 1998, Peng, Bhanu and Qing, 1999) or in a hierarchical feature model (Sclaroff, Taycher & La Cascia, 1997; Rui, Huang, Ortega & Mehrotra, 1998; Schettini, Ciocca & Gagliardi, 1999; Rui & Huang, 2000; Heesch and Rueger, 2003; Urban & Jose, 2004b). We shall refer to the two types of weights as component weights and feature weights, respectively.

Early models

In the hierarchical model proposed by Rui et al. (1998) the weight of a component is taken to be inversely proportional to the standard deviation of that component among relevant images. This heuristic is based on the intuition that a feature component that shows great variation among the relevant images does not help to discriminate between relevant and non-relevant images. Although any function that monotonically decreases with the variance would appear to be a good candidate, it turns out that dividing by the standard deviation agrees with the optimal solution that was later derived in the optimisation framework of Ishikawa et al. (1998). The feature weights are adjusted by taking into account both negative and positive examples using a simple heuristics. Although experiments suggest a substantial improvement in retrieval performance on a database containing more than 70,000 images the figures should be treated with great caution as the number of images

on which relevance feedback is given lies in the somewhat unrealistic range of 850 to 1100.

Optimising generalised Euclidean distances

An elegant generalisation of the relevance feedback method of Rui et al. (1998) was developed by Ishikawa et al. (1998). It is motivated by the observation that relevant images may not necessarily align with one of the feature dimensions and so the weighted Euclidean distance used in Rui et al. (1998) cannot fully account for their distribution (its iso-distance lines form ellipsoids whose diameters are parallel to the coordinate axes). This limitation can be addressed by considering a generalisation of the Euclidean distance, introduced by Chandra Mahalanobis under the name D-statistic in the study of biometrical data and now simply known as the Mahalanobis distance. It is more conveniently written in matrix notation as

$$D(x, y) = (x - q)^T M (x - q),$$

where M is any square matrix. If M is a diagonal matrix, then the expression reduces to Equation (1.1) with the weights corresponding to the diagonal elements. If M is a full matrix, then the expression contains products between the differences of any two components. In two dimensions with

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

this writes as

$$D(x, q) = a(x_1 - q_1)^2 + (b + c)(x_1 - q_1)(x_2 - q_2) + d(x_2 - q_2)^2$$

and similarly for higher dimensions. The iso-distance lines of the general Mahalanobis metric are ellipsoids that need not align with the coordinate axes. The components of M are found by minimising the sum of the distances between the images marked relevant and the query. The interesting twist of the model is that the query itself is re-estimated at each step. The optimisation thus determines not only M but also the query with respect to which the distances are minimised. The method integrates the two techniques of query point moving and metric update in one optimisation framework. The objective function is

$$\min_{M,q} \sum_{i=1}^{N} v_i (x^i - q) \;\; M (x^i - q), \tag{1.3}$$

where N is the number of relevant images, and v_i a relevance score given by the user. Note that x^i

here denotes the vector of the *i*th image, not the *i*th component of some vector x. Under the additional constraints that det(M) = 1 and that M is symmetric the solutions for q and M are

$$q = \frac{\sum_{i=1}^{N} v_i x_i}{\sum_{i=1}^{N} v_i} \qquad \text{and} \qquad M = [\det(C)]^{\frac{1}{n}} C^{-1},$$

where C is the covariance matrix of the positive examples. In order for the inverse of C to exist, relevance feedback needs to be given on at least as many images as there are feature components. If this is not the case, a pseudo-inverse can be used instead.

Based on our preceding discussion, some of the limitations of the approach taken by Ishikawa et al. (1998) should be evident: First, the approach tackles the problem of query point moving but does not support multi-point queries; secondly, it exploits only positive feedback which might be rather scarce at the beginning of the search; thirdly, it assumes a flat image representation model with all features for one image concatenated into one single vector. This inflates the number of parameters to be learned with the effect of rendering parameter estimation less robust.

To address the last shortcoming, Rui and Huang (2000) extend the optimisation framework of Ishikawa et al. (1998) by adding feature weights. For each feature, distances are computed using the generalised Euclidean metric and the overall similarity is obtained according to Equation (1.2). Like in Ishikawa et al. (1998) the aim is to minimise the summed distances between relevant images and the query. The objective function takes the form of Equation (1.3) except for an additional inner sum,

$$\min_{M,q,w} \sum_{i=1}^{N} v_i \sum_{j=1}^{k} w_j (x_{ij} - q_j)^T M (x_{ij} - q_j),$$

where, as before, v are relevance scores, w are feature weights and x_{ij} is the *j*th feature vector of the *i*th relevant image. The optimal solutions for q and M are the same as in Ishikawa et al. (1998) while the feature weights are given by

$$w_j \propto \frac{1}{\sqrt{\sum_{i=1}^{N} v_i d(x_{ij}, q_j)}}$$

where the squared denominator is the sum of the weighted distances between the query and all

relevant images under feature j.

Optimisation with negative feedback

The above methods only make use of positively labelled examples despite the fact that negative feedback has repeatedly been shown to prevent the retrieval results from converging too quickly towards local optima (Mueller, Mueller, Squire, Marchand-Maillet & Pun, 2000; Vasconcelos & Lippman, 2000; Heesch and Rueger, 2002; Mueller, Marchand-Maillet & Pun, 2002). An innovative method that takes explicit account of negative examples is by Aggarwal et al. (2002). They adopt the general framework of Ishikawa et al. (1998) by minising Equation (1.3) but the extra constraint is added that there are no non-relevant images within some small neighborhood of q. This is achieved by automatically modifying the relevance scores v. In particular, given some solution q and M of Equation (1.3) with an initially uniform set of relevance scores, the relevance score of the relevant image that is farthest from the current query point q is set to zero and the scores of any other positive image set to the sum of its quadratic distances from the negative examples. Minimising the objective function again with the thus altered scores yields a new solution q and M, which is more likely to contain only relevant images. This scheme is iterated until the neighborhood contains only relevant images.

Another example of metric optimisation involving negative feedback is by Lim, Wu, Singh and Narasimhalu (2001). Here, users are asked to re-rank retrieved images and the system subsequently minimises the sum of the differences between the user-given ranks and the computed ranks. Because of the integral nature of ranks, the error function is not analytic and numerical optimisation is required to find the feature weights.

A method that admits to an analytic solution is proposed in Heesch and Rueger (2003). Relevance feedback is given by positioning retrieved images closer to or further away from the query that is originally situated at the centre (Figure 4 left and middle). The user provides a real-valued vector of new distances, and the objective function is the sum of the squared errors between the distances computed by the system and the new distances supplied by the user. The distance function that minimises the objective function is used for the next retrieval step (Figure 4 right).

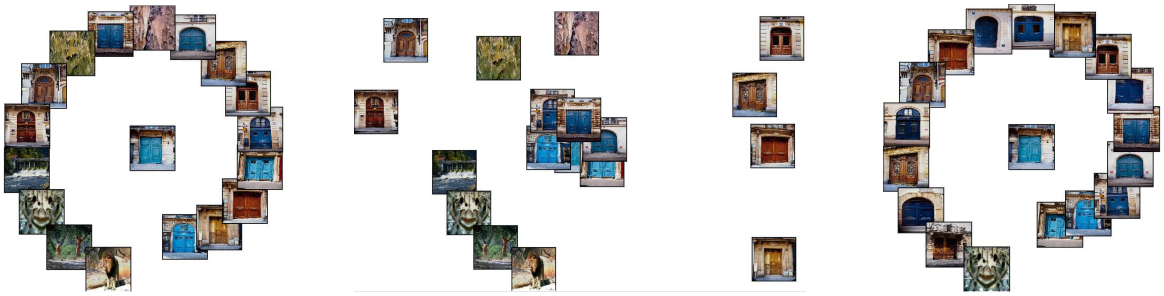

Figure 4: In search for blue doors. Left: initial display with default distance metric; Middle: display after user feedback; Right: display when retrieving with newly learned distance metric.

Multi-dimensional scaling

The methods discussed thus far retrieve a ranked list of images, often organised on a two-dimensional grid or as in Heesch and Rueger (2003) in the form of a spiral around the query. Crucially, however, the mutual distances within the set of retrieved images are not taken into account for the display, i.e. returned images that are visually similar may not necessarily be displayed close to each other. Rubner, Guibas and Tomasi (1997) apply multi-dimensional scaling (Kruskal, 1964) to the search results to achieve a more structured view. Given a set of objects and their mutual distances, we can place each object in a high-dimensional metric space such that the distances are exactly preserved. For practical purposes, the preferred dimensionality of the space is two for which distances can only be approximated. The result is an approximate two-dimensional embedding that preserves as far as possible the distances between objects. The technique can be applied both to the set of retrieved images but can also be used as a means to display the entirety of small collections in a perceptually meaningful way. Navigation through the collection can be achieved by letting the user select one of the retrieved images as the new query.

Another attempt of a synthesis between automated search and browsing is described in Santini and Jain (2000) and Santini, Gupta and Jain (2001). Similar to Rubner et al. (1997), the proposed system seeks a distance-preserving projection of the images onto two dimensions. As well as selecting an image from the display as the new query, users move images to new positions. In Santini et al. (2000), the system finds feature and component weights that minimise the mismatch between the relations imposed by the user and the computed distances. The system thus uses information about

the desired relative distances between images.

Similarity on manifolds

Up to now, we have only considered global metrics. Distances for all images are computed using the same, possibly parameterised distance metric. On the assumption that relevant images fall on some manifold in the Euclidean feature space, a better approach would be to find the best local metric. He, Ma and Zhang (2004) propose to approximate the metric structure of the manifold at the location of the query. The approximation makes use of positive examples, which are assumed to be close to the query under the geodesic distance. The algorithm proceeds by computing the k-nearest neighbours of each of the positive examples. The union of these sets constitutes the set of candidates from which we shall eventually retrieve. The geodesic distance is approximated by the topological distance on a graph whose vertices correspond to elements of the 'candidate' set along with the query and the positive examples. Edges are constructed between any two images if their unweighted Euclidean distance does not exceed some threshold. The geodesic distance is then approximated by the topological distance on the graph, that is, the length of the shortest path between two images. Retrieval on the manifold returns the set of images with the smallest topological distance to the query.

Similarity search as classification

A third class of techniques treats the problem of similarity search as one of classification. The techniques are similar to the class of metric optimisation discussed in the preceding section and some can be interpreted as estimating parameters of some similarity function.

Probabilistic approaches

Methods that approach the classification problem from a Bayesian perspective explicitly model probability densities. The aim of these methods is to assign class probabilities to an image based on the class-specific feature densities estimated from relevance feedback. Let p be an image, x its feature representation and R and N be the sets of relevant and non-relevant images. By Bayes' rule we have

$$P(p \in R \mid x) = \frac{P(x \mid p \in R)P(p \in R)}{P(x)}.$$

In Nastar, Mitschke and Meilhac (1998) the feature density of relevant images P(x|p ϵ R) is assumed to be Gaussian, and features are assumed to be independent so that P(p ϵ R|x) is a product of Gaussians,

$$P(p \in R \mid x) \propto \prod_{i=1}^{k} P(x_i \mid p \in R).$$

If we were only to consider relevant examples, the mean and standard deviation can readily be found using the principle of maximum likelihood. Nastar et al. (1998) suggest an iterative technique that takes into account negative examples. It does this by determining the proportion of negative examples falling into a 3σ confidence interval around the current mean and the proportion of positive examples falling outside of it. The error is simply the sum of the two terms. To better account for multi-modality a mixture of Gaussians can be used, an extension that has the slight disadvantage of requiring numerical optimisation for parameter estimation (Vasconcelos & Lippman, 2000; Yoon & Jayant, 2001).

Meilhac and Nastar (1999) drop the assumption of Gaussianity of feature densities and use a Parzen window for non-parametric density estimation. Feature densities are estimated for both relevant and non-relevant images and the decision rule is

$$I(x_i) = -\log[P(x_i \mid p \in R)] + \log[P(x_i \mid p \in N)]$$

for each feature. Assuming independence of features we obtain

$$I(x) = \sum_{i=1}^{k} (-\log[P(x_i \mid p \in R)] + \log[P(x_i \mid p \in N)])$$

The additiveness of this density estimation method makes it incremental, i.e., at every round a fixed number of terms is added to the decision function making the algorithm cost-effective.

The Bayesian framework developed by Cox, Miller, Omohundro and Yianilos (1998) and Cox, Miller, Minka, Papathomas and Yianilos (2000) for target search is based on an explicit model of what users would do given the target image they want. The system then uses Bayes' rule to predict the target given their action.

Discriminant classifiers

An alternative approach to classification that does not require an explicit modelling of feature densities involves finding a discriminant function that maps features to class labels using some labelled training data.

An increasingly popular classifier is the support vector machine or SVM (Vapnik, 1995). SVMs typically map the data to a higher-dimensional feature space using a possibly non-linear transform associated to a reproducing kernel. Linear discrimination between classes is then attempted in this feature space. SVMs have a number of advantages over other classifiers that make them particularly suitable for relevance feedback methods (Hong, Tian & Huang, 2000; Chen, Zhou & Huang, 2001; Tong & Chang, 2001; Jing, Li, Zhang, Zhang & Zhang, 2003; He, Li, Zhang, Tong & Zhang, 2004; Crucianu, Ferecatu & Boujema, 2004). Most notably, SVM avoid too restrictive distributional assumptions regarding the data and are flexible as prior knowledge about the problem can be taken into account by guiding the choice of the kernel.

In the context of image retrieval the training data consists of the relevant and non-relevant images marked by the user. Learning classifiers reliably on such small samples is a particular challenge. One potential remedy is that of active learning (Cohn, 1994). The central idea of active learning is that some training examples are more useful for training the classifier than others. It is guided by the more specific intuition that points close to the hyperplane, that is, in regions of greater uncertainty regarding class membership, are most informative and should be presented to the user for labelling instead of a random subset of unlabelled points. Applications of active learning to image retrieval are found in Tong and Chang (2001) and He et al. (2004). In the former work a support vector machine is trained over successive rounds of relevance feedback. In each round the system displays the images closest to the current hyperplane. Once the classifier has converged, the system returns the top k relevant images farthest from the final hyperplane. Although the method involves the user in several rounds of potentially ungratifying feedback, the performance of the trained classifier improves over that of alternative techniques such as query point moving and query expansion.

A summary of much of the above can be found in the table below. For each system, we note the kind of information communicated through feedback, the part of the system that is modified in response.

Author	Type of Feedback	Range	Objective
Rui-97	+/-	Binary	Query point moving
Rui-98	+	Real	Query point moving
Rui-98	+	Discrete	Metric optimisation
Rui-00	+	Real	Query point moving
Porkaew-99	+	Binary	Query expansion
Ishikawa-98	+	Real	Query point moving
Nastar-98	+/-	Binary	Distribution of relevant
Meilhac-99	+/-	Binary	Distribution of relevant
Lim-01	+/-	Discrete	Metric optimisation
Ishikawa-98	+	Real	Metric optimisation
Heesch-03	+/-	Real	Metric optimisation
Urban-03	+	Binary	Query point moving
Urban-04	+	Binary	Query expansion
Kim-03	+	Binary	Query expansion
Tong-01	+/-	Discrete	Discrimant classifier
He-04	+	Binary	Metric optimisation
Aggarwal-02	+/-	Real	Metric optimisation

Table 1: Overview of relevance feedback systems developed in a QBE setting

INTERLUDE

Let us now take a step back and assess the merit of the general methodology described above. The reported performance gains through relevance feedback are often considerable even though any performance claims must be judged carefully against the experimental particulars, especially the database size, the performance measures, and the type of queries. Below we suggest two major problems with the relevance feedback methodology.

Parameter initialisation

The utilisation of relevance feedback for query expansion and multi-modal density estimation has attracted much attention and appears justified on the ground that the feature distributions of most relevance classes tend to be multi-modal and form natural groups in feature space. But unless the

query itself consists of multiple images representing these different groups, we should not reasonably expect images from different groups to be retrieved in response of the query. If anything, the retrieved images will contain images from the cluster to which the query image is closest under the current metric.

But not only do relevance classes often form visually distinct clusters, images often belong to a number of relevance classes. This is an expression of the semantic ambiguity which pertains in particular to images and which relevance feedback seeks to resolve. With queries consisting of single images, the question to resolve is which natural group the query image belongs to, not so much which the different natural groups belonging to the relevance class of the query. But while some systems cater for multi-modality, none explicitly deal with polysemy. By initialising parameter values, systems effectively impose a particular semantic interpretation of the query.

The problem of parameter initialisation has so far received insufficient attention. One notable exception is the work by Aggarwal et al. (2002) which we had mentioned earlier in a different context. The system segments the query image, modifies each segment in various ways and displays a set of modified queries to the users who mark segments that continue to be relevant. The feature weights are then computed similar to Rui et al. (1997) by considering the variance among the relevant segments.

Another method of parameter initialisation that is very similar in spirit to that of Aggarwal (2002) is developed in Heesch (2005). The method seeks to expose the different semantic facets of the query image by finding all images that are most similar to it under Equation (1.2) for some weight set w. As we vary w, different images will become the nearest neighbour of the query. For each such nearest neighbour we record its associated w, which we may regard as a representation of one of the semantic facets users may be interested in. Users select a subset of these nearest neighbours and thereby implicitly select a set of weights. These weights are then used to carry out a standard similarity search. The method outperforms relevance feedback methods that retrieve with an initially uniform weight set w but is not inexpensive computationally. NN^k Networks which we shall discuss in the next section provide another attempt to tackle the initialisation problem.

Exploratory search

With very few exceptions, the methods described above rely on the assumption that users know what they are looking for. The methods are designed to home in on a set of relevant items within a few iterations and do not support efficient exploration of the image collection. We shall see in the second half of this chapter that more flexible interaction models may address this issue more successfully.

SEARCH THROUGH BROWSING

Browsing offer an alternative to the conventional method of query by example but have received surprisingly little attention. Some of the advantages of browsing are as follows:

- Image browsing requires but a mental representation of the query. Although automated image annotation (Lavrenko, Manmatha & Jeon, 2003; Feng, Manmatha & Lavrenko, 2004; Zhang, Zhang, Li, Ma & Zhang, 2005; Yavlinsky, Schofield and Rueger, 2005) offers the possibility to reduce visual search methodologically to traditional text retrieval, there may often be something about an image which cannot be expressed in words leaving visually guided browsing a viable alternative.
- Retrieval by example image presupposes that users already have an information need. If this is not the case, enabling users to navigate quickly between different regions of the image space becomes of much greater importance.
- For large collections, time complexity becomes an issue. Even when hierarchical indexing structures are used, performance of nearest neighbour searches has been shown to degrade rapidly in high-dimensional feature spaces. For particular relevance feedback techniques, approximate methods may be developed that exploit correlations between successive nearest neighbour searches (Wu & Manjunath, 2001), but there does not exist a universal cure. Meanwhile, browsing structures can be precomputed allowing interaction to be very fast.
- The ability of the human visual system to recognise patterns reliably and quickly is a marvel yet to be fully comprehended. Endowing systems with similar capabilities has proven an exceedingly difficult task. Given our limitations in understanding and emulating human cognition, the most promising way to leverage the potential of computers is to combine their strengths with those of users and achieve a synergy through interaction. During browsing users

are continuously asked to make decisions based on the relevance of items to their current information need. A substantial amount of time is spent, therefore, by engaging users in what they are best at, while exploiting computational resources to render interaction fast.

Hierarchies

Hierarchies have a ubiquitous presence in our daily life: examples include the organisation of files on a computer, the arrangement of books in a physical library, the presentation of information on the web, employment structures, postal addresses and many more.

To be at all useful for browsing, hierarchical structures need to be sufficiently intuitive and allow users to predict in which part of the tree the desired images may reside. When objects are described in terms of only a few semantically rich features, building such hierarchies is relatively easy. The low-level, multi-featural representation of images renders the task substantially more difficult.

Agglomerative Clustering

The most common methods for building hierarchies is by way of clustering either by iteratively merging clusters (agglomerative clustering) or by recursively partitioning clusters (divisive clustering), see Duda (2001) for an overview.

Early applications of agglomerative clustering to image browsing are described in Yeung and Liu (1995), Yeung and Yeo (1997), Zhang and Zhong (1995) and Krishnamachari and Abdel-Mottaleb (1999}. The first two papers are concerned with video browsing and clustering involves automated detection of topics and for each topic the constituent stories. Stories are represented as video posters, a set of images from the sequences that associate with repeated or long shots and act as pictorial summaries. In Zhang and Zhong (1995) and Yang (2004} the self-organising map algorithm (Kohonen, 1995) is applied to map images on a two-dimensional grid. The resulting grid is subsequently clustered hierarchically. One of the major drawbacks of the self-organising map algorithm (and neural network architectures in general) is its computational complexity. Training instances often need to be presented multiple times and convergence has to be slow in order to achieve good performance, in particular so for dense features

Chen, Bouman and Dalton (1998; 2000) propose the concept of a similarity pyramid to represent image collections. Each level is organised such that similar images are in close proximity on a two-dimensional grid. Images are first organised into a binary tree through agglomerative clustering based on pairwise similarities. The binary tree is subsequently transformed into a quadtree which provides users a choice of four instead of two different child nodes. The arrangement of cluster representatives is chosen such that some measure of overall visual coherence is maximised. Since the structure is precomputed, the computational cost incurred at browsing time is slight.

Divisive clustering

Agglomerative clustering is quadratic in the number of images. Although this can be alleviated by sparsifying the distance matrix, this method becomes inaccurate for dense feature representations and is more amenable to key-word based document representations.

A computationally more attractive alternative is divisive clustering whereby clusters are recursively split into smaller clusters. One popular clustering algorithm for this purpose is k-means. In Pecenovic, Do, Vetterli and Pu (2000) it is applied to 6,000 images with cluster centroids being displayed according to their position on a global Sammon map. However, compared to agglomerative clustering, the divisive approach has been found to generate less intuitive groupings (Yeung & Yeo, 1997; Chen et al., 2000) and the former has remained the method of choice in spite of its computational complexity.

Networks

Nearest neighbour networks

A significant work on interlinked information structures dates back to the mid-1980s (Croft & Parenty, 1985). It proposes to structure a collection of documents as a network of documents and terms with accordingly three types of weighted edges. The authors suggest to keep only links between a document and the document most similar to it, and similarly for terms. Term-term and document-document links thus connect nearest neighbours and each document gives rise to what a star cluster comprising the document itself and all adjacent nodes. Although the structure is

intended for automated search, the authors are aware that "as well as the probabilistic and cluster-based searches, the network organisation could allow the user to follow any links in the network while searching for relevant documents. A special retrieval strategy, called browsing, could be based on this ability." (p. 380). However, the number of document-document edges does not exceed by much the number of documents, and star clusters are disconnected rendering browsing along document-document nodes alone impractical.

Importantly, the work has inspired subsequent work by Cox (1992; 1995). Cox motivates associative structures for browsing by observing that "people remember objects by associating them with many other objects and events. A browsing system on a static database structure requires a rich vocabulary of interaction and associations." His idea is to establish a nearest neighbour network for each of a set of the different object descriptors. Being aware that different features may be important to different users, Cox realises the importance of interconnecting nearest neighbour networks to allow multi-modal browsing.
Unfortunately, Cox's work has not become as widely known as perhaps it should have. What may partly account for this is that content based image retrieval was then in its very early beginning and the first research programme that grew out of the initial phase of exploration happened to be that of query by example pushing browsing somewhat to the periphery.

NN^k Networks

The problem with many of the above structures is that the metric underlying their construction is fixed. The advantage of fast navigation therefore comes at a prize: users are no longer in a position to alter the criterion under which similarity is judged. The structures thus deride the principal tenet that motivates relevance feedback techniques. Zhou and Huang (2001) arrive at a similar conclusion when they observe that ‘the rationale of relevance feedback contradicts that of pre-clustering.’

A browsing structure that has been designed with this in mind are NN^k Networks (Heesch, 2005; Heesch, Pickering, Yavlinsky & Rueger (2004); Heesch & Rueger, 2004; 2005). The structure is a directed graph where an arc is established from p to q if q is the nearest neighbour of p under at least one combination of features (represented in terms of index i in Equation (1.2)). Instead of imposing a particular instance of the similarity metric, NN^k Networks expose the different semantic

facets of an image by gathering all top-ranked images under different metrics. During browsing users select those neighbours in the graph that match their target best. NN^k Networks exhibit small-world properties (Watts & Strogatz, 2000) that make them particularly well suited for interactive search. Relevant images tend to form connected subgraphs so that a user who has found one relevant image is likely to find many more by following "relevance trails" through the network. The screenshots below illustrate the diversity among the set of neighbours for three different positions in a network of 32,000 Corel images. The size of the image is a measure of the number of different metrics under which that image is more similar to the currently selected image than any other.

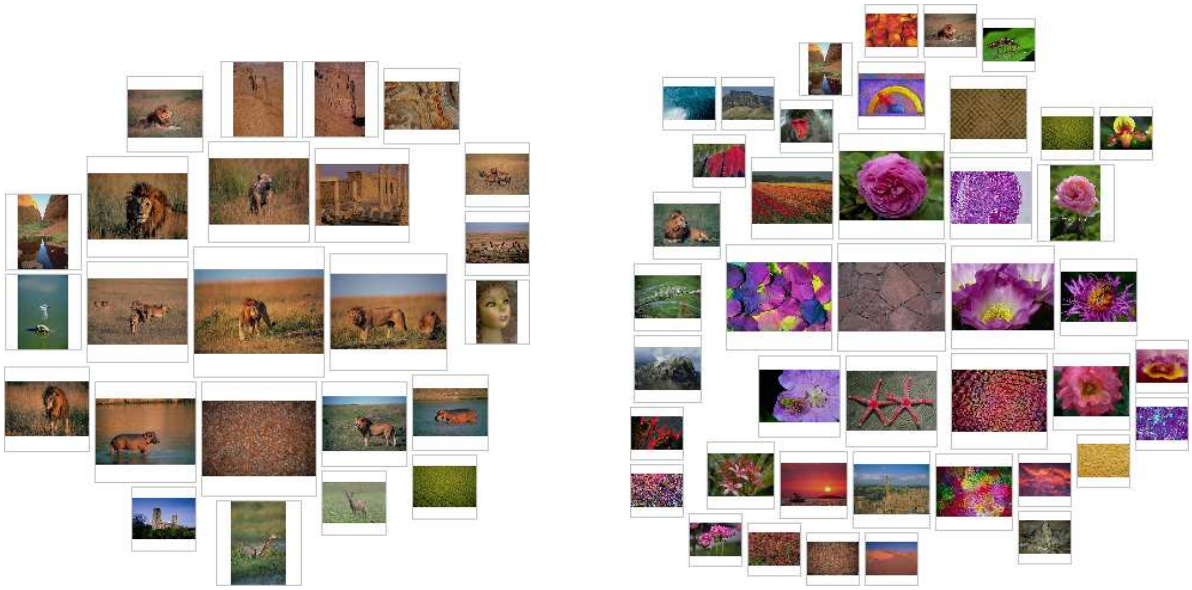

Figure 5: The set of NN^k in a network of 32,000 Corel images for three different positions.

Pathfinder networks

For browsing at least parts of the network need to be visualised. The large number of links in a network may prevent users from recognising structural patterns that could aid navigation. A practical strategy is to reduce the number of links. The pathfinder algorithm is one example of a link-reduction algorithm (Dearhold & Schvaneveldt, 1990). It is not concerned with constructing the original network but converting a network of any kind to a sparser network. The pathfinder algorithm removes an edge between vertices if there exists another path of shorter length. An application of pathfinder networks to the problem of organising image collections is found in Chen, Gagaudakis and Rosin (2000) but the scope for interaction is limited. Indeed, it seems that the principal application domain of pathfinder networks has so far been visual data mining, not

interactive browsing. The reason is quite likely to be found in the computational complexity that is prohibitive for collection sizes of practical significance. Moreover, visualisation and navigation places somewhat different structural demands on the networks. While visualisation requires the extraction of only the most salient structure, retaining some degree of redundancy renders the networks more robust for navigation purposes.

Dynamic trees: ostensive browsing

The ostensive model of Campbell (2000) is iterated query by example in disguise but the query only emerges through the interaction of the user with the collection. The impression for the user is that of navigating along a dynamically unfolding tree structure. While originally developed for textual retrieval of annotated images, the ostensive model is equally applicable to visual features (Urban et al., 2003). It consists of two components: the core component is the relevance feedback model, the other is the display model.

Relevance feedback takes the form of selecting an image from those displayed. A new query is formed as the weighted sum of the features of this and previously selected images. In Urban et al. (2003) images are described by colour histogram. Given a sequence of selected images, the colour representation of the new query is given as the weighted sum of individual histograms with weights taking the form of $w_i = 2^{-i}$ (i = 0 indexing the most recent image).

The display model is that of an unfolding tree structure: images closest under the current query are displayed in a fan-like pattern to one side of the currently selected image. Users can select an image from the retrieved set, which is placed in the centre, and a new set of images are retrieved in response. Since previous images are kept on the display the visual impression of the user is that of establishing a browsing path through the collection. In Urban et al. (2003) the browsing path is displayed in a fisheye view (see Figure 6).

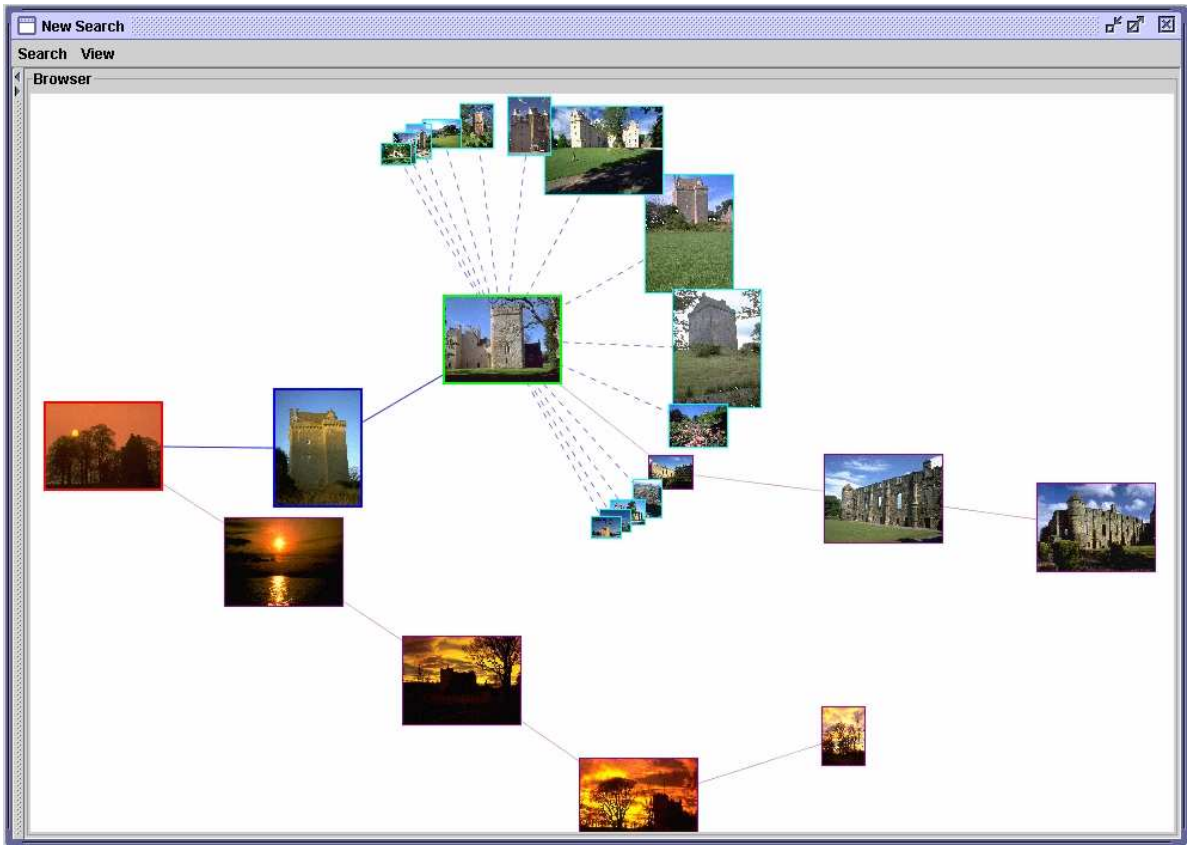

Figure 6: The interface of the ostensive browser by Urban et al. (2003)

The ostensive model attempts to track changing information needs by continuously updating the query. Which set of images are retrieved depends on which path the user has travelled to arrive at the current point. Because the number of such different paths grows quickly with the size of the image collection, it is impractical to compute a global structure beforehand. Nonetheless, for the user the impression is one of navigating in a relatively unconstrained manner through the image space. Unlike many other relevance feedback systems, users do not have to rank or label images, or change their relative location. The interaction is thus light and effective. Again, a summary of the models is given below (Table 2).

Author	Structure	RF	Flexible Metric	Offline	Online	# images
Cox-95	Networks	No	Yes	$O(n^2)$	O(1)	< 100
Heesch-04	Networks	No	Yes	$O(n^2)$	O(1)	32,000
Chen-00	Networks	No	No	$O(n^4)$	O(1)	279
Urban-03	Dynamic Trees	Yes	Yes	O(1)	O(n)	800
Zhang-95	Hierarchies	No	No	$O(n^2)$	O(1)	unavailable
Chen-00	Hierarchies	No	No	$O(n^2)$	O(1)	10,000

Table 2: Overview of browsing network models

CONCLUSIONS

It has become clear over the past decade that content-based image retrieval can benefit tremendously from letting the user take on a greater role in the retrieval process. In this chapter we have examined the different forms of user involvement in two contexts: query by example and interactive browsing. In the former setting, users initiate a search by submitting a query image and wait for images to be retrieved for them. A standard method of involving users in the subsequent stages of the process is to ask for relevance feedback on the retrieved images. The relevance information can be used to automatically modify the representation of the original query (query update), to adjust the function that is used to compute similarities between images and the query, or to learn a classifier between non-relevant and relevant images.

The query by example setting has a number of limitations. Most importantly, it assumes that users already have an information need and a query image at their disposal. Systems of this category do not generally support free exploration of a collection. The second part of this chapter has examined a number of browsing models where the user becomes the chief protagonist. In addition to requiring only a mental representation of the query, browsing structures have the advantage that they may be precomputed so that user interaction is fast. Browsing structures often take the form of hierarchies or networks and browsing takes place by moving between vertices of the graph. Hierarchies can readily be constructed through hierarchical clustering and support search from the more general to the more specific thus affording an impression of progressive refinement. However, it may equally

create a sense of lost opportunities if navigation is restricted to the vertical dimension.
Networks have the advantage over hierarchies that navigation may be less constrained. At the same time, it is more difficult to provide a global overview of the content so that it becomes increasingly important to organise objects in the network such that the local neighbourhood of the currently selected object contains sufficient information for users to decide where to go next.
There is a more general problem with precomputed structures that affects most of the models discussed. By being precomputed, users are not generally in a position to remould the structure according to their own preferences. This seems necessary, however, as the structures are almost always constructed by fixing the distance metric and applying that same metric across the entire collection. The advantage of fast navigation comes at the price that users can no longer impose their own perception of similarity.

There remain a number of exciting and important problems, a solution to which should lead to a new generation of smarter, more versatile systems for visual search. For example, while searching interactively for images users continuously provide implicit relevance feedback. In addition to exploiting this information for the current search session, one should clearly wish to endow systems with some form of long-term memory. Also, large collections will take an appreciable amount of time to be cast into a browsable structure. This seems acceptable provided the effort needs to be expended only once but many collections are dynamic with new images regularly being added and others removed. An update should not involve a complete recomputation of the structure but the extent to which the above models lend themselves to an efficient update is seldom investigated.
Finally, most of the systems we have discussed either involve a precomputed structure or initiate a new query at every step. Systems of the first kind are often too rigid, systems of the second too slow for large collections. What may hold promise are hybrid structures that are partially precomputed but flexible enough to remain responsive to relevance feedback.

REFERENCES

Aggarwal G, Ashwin T and Ghosal S (2002). An image retrieval system with automatic query modification. IEEE Trans Multimedia, 4(2):201-213

Bang H and Chen T (2002). Feature space warping: An approach to relevance feedback. In Proc Int'l Conf Image Processing

Campbell I (2000). The ostensive model of developing information-needs. PhD thesis, University of Glasgow

Chen C, Gagaudakis G and Rosin P (2000). Similarity-based image browsing. In Proc IFIP World Computer Congress

Chen J-Y, Bouman C and Dalton J (2000). Hierarchical browsing and search of large image databases. IEEE Trans Image Processing, 9(3)

Chen Y, Zhou X and Huang T (2001). One-class SVM for learning in image retrieval. In Proc Int'l Conf Image Processing

Cohn D, Atlas L and Ladner R (1994). Improving generalization with active learning. Machine Learning, 15(2):201-221

Cox I, Miller M, Minka T, Papathomas T and Yianilos P (2000). The Bayesian Image Retrieval System, PicHunter: Theory, implementation, and psychophysical experiments. IEEE Trans Image Processing, 9(1):20-38

Cox I, Miller M, Omohundro S and Yianilos P (1998). An optimized interaction strategy for Bayesian relevance feedback. In Proc IEEE Conf Computer Vision and Pattern Recognition, pages 553-558

Cox K (1992). Information retrieval by browsing. In Proc Int'l Conf New Information Technology

Cox K (1995). Searching through browsing. PhD thesis, University of Canberra

Croft B and Parenty T (1985). Comparison of a network structure and a database system used for document retrieval. Information Systems, 10:377-390

Crucianu M, Ferecatu M and Boujemaa N (2004). Relevance feedback for image retrieval: a short survey. In: State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction including Datamodels and Languages

Dearholt D and Schvaneveldt R (1990). Properties of Pathfinder networks, In Schvaneveldt R (Ed.), Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex

Duda R, Hart P and Stork D (2001). Pattern Recognition. Wiley, New York.

Fagin R, Kumar R and Sivakumar D (2003). Efficient similarity search and classification via rank aggregation. In Proc ACM Int'l Conf Management of Data, pages 301-312

Feng S, Manmatha R and Lavrenko V (2004). Multiple Bernoulli relevance models for image and video annotation. In Proc Int'l Conf Computer Vision and Pattern Recognition

Fowler R, Wilson B and Fowler W (1992). Information navigator: An information system using associative networks for display and retrieval. Department of Computer Science, Technical Report NAG9-551, 92-1

He J, Li M, Zhang H-J, Tong H and Zhang C (2004). Mean version space: a new active learning method for content-based image retrieval. In Proc Int'l Workshop on Multimedia Information Retrieval in conjunction with ACM Multimedia, pages 15-22

He X, Ma W-Y and Zhang H-J (2004). Learning an image manifold for retrieval. In Proc ACM Multimedia, pages 17-23

Heesch D (2005). The NN^k idea for image searching and browsing. PhD Thesis. Imperial College London

Heesch D, Pickering M, Yavlinsky A and Rüger S (2003). Video retrieval within a browsing framework using keyframes. In Proc TREC Video Retrieval Evaluation (TRECVID)

Heesch D and Rüger S (2002). Combining features for content-based sketch retrieval - a comparative evaluation of retrieval performance. In Proc European Conf Information Retrieval, pages 42-51. LNCS 2291, Springer

Heesch D and Rüger S (2003). Performance boosting with three mouse clicks - relevance feedback for CBIR. In Proc European Conference on Information Retrieval, pages 363-376. LNCS 2633, Springer

Heesch D and Rüger S (2004). NN^k Networks for content-based image retrieval. In Proc European Conf Information Retrieval, pages 253-266. LNCS 2997, Springer

Heesch D and Rüger S (2005). Image browsing: A semantic analysis of NN^k Networks. In Proc Int'l Conf on Image and Video retrieval, pages 609-618. LNCS 3568, Springer

Hong P, Tian Q and Huang T (2000). Incorporate support vector machines to content-based image retrieval with relevant feedback. In Proc IEEE Int'l Conf Image Processing

Ishikawa Y, Subramanya R and Faloutsos C (1998). Mindreader: querying databases through multiple examples. In Proc Very Large Data Bases Conf, pages 433-438

Jing F, Li M, Zhang L, Zhang H-J and Zhang B (2003). Learning in region-based image retrieval. In Proc IEEE Int'l Symp Circuits and Systems

Kim D-H and Chung C-W (2003). Qcluster: relevance feedback using adaptive clustering for content-based image retrieval. In Proc ACM SIGMOD Int'l Conf Management of Data, pages 599–610

Kohonen T (2001). Self-organizing maps. Springer Series in Information Sciences (Volume 30).

Kruskal J (1964). Multi-dimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. Psychometrika, 29:1-27

Lavrenko V, Manmatha R and Jeon J (2003). A model for learning the semantics of pictures. In Int'l Conf Neural Information Processing Systems

Lim J, Wu J, Singh S and Narasimhalu D (2001). Learning similarity matching in multimedia content-based retrieval. IEEE Trans Knowledge and Data Engineering, 13(5):846-850

Meilhac C and Nastar C (1999). Relevance feedback and category search in image databases. In Proc Int'l Conf on Multimedia Communications Systems, pages 512-517

Mueller H, Marchand-Maillet S and Pun T (2002). The truth about Corel - evaluation in image retrieval. In Proc Int'l Conf on Image and Video Retrieval, pages 38-49. LNCS 2383, Springer

Mueller H, Mueller W, Squire D, Marchand-Maillet S and Pun T (2000). Strategies for positive and negative relevance feedback in image retrieval. In Proc Int'l Conf Pattern Recognition

Nastar C, Mitschke M and Meilhac C (1998). Efficient query refinement for image retrieval. In IEEE Conf Computer Vision and Pattern Recognition

Pecenovic Z, Do M, Vetterli M and Pu P (2000). Integrated Browsing and Searching of Large Image Collections. In Proc Int'l Conf Advances in Visual Information Systems, pages 279-289. LNCS 1929, Springer.

Peng J, Bhanu B and Qing S (1999). Probabilistic feature relevance learning for content-based image retrieval. Computer Vision and Image Understanding, 75(12):150-164

Porkaew K, Chakrabarti K and Mehrotra S (1999). Query refinement for multimedia similarity retrieval in Mars. In Proc 7th ACM Int'l Conf Multimedia, pages 235-238

Rocchio J (1971). The SMART Retrieval System. Experiments in Automatic Document Processing. Prentice Hall

Rubner Y, Guibas L and Tomasi C (1997). The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In DARPA Image Understanding Workshop

Rui Y and Huang T (2000). Optimizing learning in image retrieval. In Proc IEEE Conf on Computer Vision and Pattern Recognition

Rui Y and Huang T and Mehrotra S (1997). Content-based image retrieval with relevance feedback

in Mars. In Proc IEEE Int'l Conf on Image Processing

Rui Y, Huang T, Ortega M and Mehrotra S (1998). Relevance feedback: A power tool for interactive content-based image retrieval. In IEEE Trans Circuits and Video Technology

Salton G and McGill M (1982). Introduction to Modern Information Retrieval. McGraw-Hill Book Company

Santini S, Gupta A and Jain R (2001). Emergent semantics through interaction in image databases. In IEEE Trans Knowledge and Data Engineering, 13(3):337-351

Santini S and Jain R (2000). Integrated browsing and querying for image databases. IEEE MultiMedia, 7(3):26-39

Schettini R, Ciocca G and Gagliardi I (1999). Content-based color image retrieval with relevance feedback. In Proc Int'l Conf Image Processing, Japan

Sclaroff S, Taycher L and La Cascia M (1997). Imageroever: A content-based image browser for the world wide web. In IEEE Int'l Workshop Content-based Access of Image and Video Libraries

Tong S and Chang E (2001). Support vector machine active learning for image retrieval. In Proc ACM Int'l Conf Multimedia, pages 107-118, New York, NY, USA, ACM Press.

Urban J and Jose J (2004). Ego: A personalised multimedia management. In Proc ICMP

Urban J and Jose J (2004). Evidence combination for multi-point query learning in content-based image retrieval

Urban J, Jose J and Rijsbergen K (2003). An adaptive approach towards content-based image retrieval. In Proc Int’l Workshop on Content-Based Multimedia Indexing pages 119-126

Vapnik V (1995). The Nature of Statistical Learning Theory. Springer

Vasconcelos N and Lippman A (2000). Bayesian relevance feedback for content-based image retrieval. In IEEE workshop on Content-Based Access of Image and Video Libraries, South Carolina, page 63

Watts D and Strogatz S (1998). Collective dynamics of small-world networks. Nature 393:440-442

Wolfram S (2004). A new kind of science. Wolfram Ltd.

Wu L, Faloutsos C, Sycara K and Payne T (2000). Falcon: Feedback adaptive loop for content-based retrieval. In Proc Conf Very Large Data Bases, pages 297-306

Wu P and Manjunath B (2001). Adaptive nearest neighbour search for relevance feedback in large image databases. In Proc ACM Multimedia, pages 89-97

Yang C (2004). Content-based image retrieval: a comparison between query by example and image browsing map approaches. Journal of Information Science. 30:3, pages 254-267.

Yavlinsky A, Schofield E and Rüger S (2005)**:** Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation. Int'l Conf on Image and Video Retrieval, pages 507-517, Springer LNCS 3568

Yeung M and Liu B (1995). Efficient matching and clustering of video shots. In Proc IEEE Int'l Conf Image Processing, pages 338-341

Yeung M and Yeo B (1997). Video visualization for compact presentation and fast browsing of pictorial content. IEEE Trans Circuits and Systems for Video Technology, 7:771-785

Yoon J and Jayant M (2001). Relevance feedback for semantics based image retrieval. In Proc Int'l Conf Image Processing, pages 42-45

Zhang H and Zhong D (1995). A scheme for visual feature based image indexing. In Proc SPIE/IS&T Conf Storage and Retrieval for Image and Video Databases III, volume 2420, pages 36-46

Zhang H-J and Su Z (2001). Improving CBIR by semantic propagation and cross modality query expansion. In Proc Multimedia Content-based Indexing and Retrieval, pages 79-82

Zhang R, Zhang Z, Li M, Ma W-Y and Zhang H-J (2005). A probabilistic semantic model for image annotation and multi-modal image retrieval. In Proc Int'l Conf Computer Vision

Zhou X and Huang T (2001). Relevance feedback in image retrieval: a comprehensive review. ACM Multimedia Systems