# Image Browsing: Semantic Analysis of $NN^k$ Networks

Daniel Heesch and Stefan Rüger

Imperial College London, Department of Computing,
South Kensington Campus, London SW7 2AZ, UK
{daniel.heesch, s.rueger}@imperial.ac.uk

**Abstract.** Given a collection of images and a set of image features, we can build what we have previously termed $NN^k$ networks by representing images as vertices of the network and by establishing arcs between any two images if and only if one is most similar to the other for some weighted combination of features. An earlier analysis of its structural properties revealed that the networks exhibit small-world properties, that is a small distance between any two vertices and a high degree of local structure. This paper extends our analysis. In order to provide a theoretical explanation of its remarkable properties, we investigate explicitly how images belonging to the same semantic class are distributed across the network. Images of the same class correspond to subgraphs of the network. We propose and motivate three topological properties which we expect these subgraphs to possess and which can be thought of as measures of their compactness. Measurements of these properties on two collections indicate that these subgraphs tend indeed to be highly compact.

## 1 Introduction

The methodological framework in which the problem of content-based image retrieval has traditionally been investigated is that of query by example (e.g. [11] [2] [12] [6] [7]). The information-seeking users, on the one hand, provide a pictorial articulation of their information needs. The system, on the other hand, utilizes the supplied query specification to retrieve potentially relevant objects. The query by example paradigm has been somewhat broadened in recent years by utilizing relevance feedback as a way to increase retrieval performance (e.g. [20] [16] [17] [15]). An alternative to query by example is browsing. Remarkably, browsing has not quite met with the same enthusiasm although it has a number of advantages that make it superior to query by example in many contexts. Firstly, the query in content-based image retrieval typically takes the form of an image, which is not always easily at hand. For browsing, users do not require a physical instantiation of the query. Rather, the search is guided in the first place by the mental representation of a particular target image or a relevance class. Secondly, retrieval by example image not only presupposes that the users

W.-K. Leow et al. (Eds.): CIVR 2005, LNCS 3568, pp. 609–618, 2005.

have the means to formulate their information need pictorially, it also presupposes that users have an information need in the first place. This may often be doubted. The information need may rather develop in the course of and as a result of the interaction. In such situations methods based on example queries become otiose. Lastly, in order for retrieval technology to become of practical use, time complexity becomes as important an issue as retrieval effectiveness. Since a collection can be regarded as essentially static, the structure into which it is shaped for the purpose of browsing can be precomputed. Provided the structure can be loaded and displayed efficiently, online complexity can be made independent of collection size.

The key question of browsing is how to organize the images into a browsable structure. Our proposition ([9] [10]) is that of directed graphs with arcs between any two images if one is most similar to the other under *some* weighted combination of features. We term the resulting structure $\mathrm{NN}^k$ networks where NN stands for nearest neighbour and $k$ denotes the number of features used. The strength of $\mathrm{NN}^k$ networks stems from its fundamentally agnostic approach towards feature weighting. It does not organize images based on one particular feature, nor any particular combination of features, but according to all possible combinations at once. This is the key difference distinguishing our approach from all similar work. The rationale is that we often do not know which features are most useful for capturing semantic image similarity. The neighbours of an image in the graph can loosely be thought of as exemplifying the range of all those image meanings that lie within the representational scope of the features used. Some of these may not be relevant to a user, but if the features are any good, at least some will be. Because the set of neighbours in the graph is visually rather heterogenous, users can quickly navigate to different parts of the collection. We have previously shown [9] that the resulting networks exhibit so called small-world properties, a combination of low average distance between vertices even for large collections, and a high degree of local clustering, and have employed the structures very successfully in the search task of TRECVID [8]. In this paper, we continue our topological analysis by looking specifically at how semantically related images are distributed across the network. We hope that the distribution is compact in a sense that we will make more precise shortly.

The paper is structured as follows. Section 2 describes relevant work by other authors. Section 3 briefly introduces $\mathrm{NN}^k$ networks. Section 4 defines the notion of the compactness of a subgraph and uses this to present a structural analysis of $\mathrm{NN}^k$ networks in terms of the distribution of relevance classes. Section 5 concludes the paper.

## 2 Related Work

Research on browsing has remained remarkably scant. A major work dating back to the early years of information retrieval is that by Croft and Parenty [5]. They suggest to represent documents in a nearest neighbour network based on term similarity but does not conceive of the network as a structure for browsing. The idea was taken up by Cox [4] who recognized that nearest neighbour networks are ideal for interactive browsing of relational databases. Each field can be used to build a nearest neighbour network while individual records provide cross-linking between such networks. Remarkably, the idea was not taken up by the information retrieval community.

In CBIR, ideas about navigating through image collections began to surface with Rubner *et al.* [14]. Given a set of images, such as those returned for a query, an image feature and some distance metric, we can use the mutual distances between the images to derive an approximate two-dimensional embedding using multi-dimensional scaling. The visualization technique helps to inform users about the neighbourhood of the retrieved images and can also be used to display the entirety of small collections in a perceptually meaningful way.

The ostensive model by Campbell [3] supports browsing through a dynamically created tree structure. When an image is selected during browsing, the system tries to determine the optimal feature combination given the current query and the history of selected images. The results are displayed as nodes adjacent to the query image which can then be selected as the new query.

Another synthesis between query-based search and browsing is described in [19] and [18]. Similar to [14], the proposed system finds an embedding of the images in two dimensions that maximally preserves distances as computed under the current set of features. Relevance feedback is given by forming clusters of relevant images which the system utilizes to update the distance function resulting in a new configuration of images on the screen.

## 3 NN$^k$ Networks

NN$^k$ networks were introduced in [9] and we shall give only a very brief introduction. Given a collection of images and a set of image features, we define an image $p$ to be an NN$^k$ of image $q$ if and only if there exists at least one convex combination of feature-specific distances $d(p, q)$ for which $p$ has minimal distance to $q$. Formally, $p$ is an NN$^k$ of $q$ iff

$$\arg\min_i \left( \sum_j^k w_j \times d_j(i, q) \right) = p$$

for some $w = (w_1, w_2, \ldots, w_k)$ where $w_j \geq 0$ and $\sum w_j = 1$. No restriction is placed on how the distances between features are computed. We will often refer to the image $q$ as the focal image. The set of NN$^k$ can be thought of capturing the range of image meanings of the focal image. A collection can now be turned into a directed graph, or NN$^k$ network, by establishing an arc between each image and all its NN$^k$. For 8,000 images and 12 features, the average number of neighbours is around 50 which most screens can easily accommodate. The features used to build the networks are as described in [10].

## 4 Semantic Analysis

In [9] we have enquired into the topological structure of $\mathrm{NN}^k$ networks without reference to the meaning or relevance of images. We may call this the formal analysis. It has brought to light a few interesting properties, namely small average distance between any two images (number of arcs that need to be traversed to get from one to the other) and a high clustering coefficient (a measure of the extent to which an image's neighbours are themselves neighbours). These properties suggest that $\mathrm{NN}^k$ networks are ideal for efficient navigation. What has been missing is a *semantic* analysis of the browsing structure. By this we mean an analysis in terms of the distribution of particular subsets of images. The subsets we are interested in are the relevance classes that associate with a particular information need. Our hope is that two images that belong to the same relevance class are likely to be close to each other. We shall denote by $V_c$ the set of all those vertices in the network $G$ that belong to the same relevance class $c$. $G_c$ is the subgraph induced by $V_c$. Its vertex set is $V_c$ and its arc set includes all those arcs of $G$ that have their heads and tails in $V_c$. We call $G_c$ the relevance subgraph. We can view semantic analysis as the structural analysis of relevance subgraphs and hope that these subgraphs are in some sense *compact* such that semantically related images cluster. We take the next section to make the informal notion of compactness more precise.

### 4.1 Measuring Compactness of Subgraphs

In our endeavour to complement our formal analysis with a semantic analysis, the first problem we face is the question of how to measure compactness of subgraphs. Intuitively, a subgraph that is complete should be recognized as very compact, and as least compact a graph where no two vertices are connected. We will in the following motivate and describe three properties which we think capture this intuitive notion well.

**Average distance:** Perhaps the most intuitive property we would like relevance subgraphs to possess is that its constituent vertices have a small average distance (defined as the length of the shortest path between two vertices) in the original graph. This is loosely analogous to the observation that point sets in some metric space form clusters if the average distance between two points of that set is smaller than the expected distance between any two randomly chosen points. It is important to note that we are not interested in the distance within the subgraph but within the original graph since the user is not confined to the subgraphs when navigating through the network. The lowest possible average distance that can be attained is 1 and requires the subgraph to be a complete graph where each vertex is connected to every other. As we have shown in [9], the average distance between any two vertices in $\mathrm{NN}^k$ networks lies between three and four across a range of different collections. Given the lower bound of one we therefore do not expect any further reduction to be very substantial.

**Average vertex degree:** The average vertex degree of a vertex of $G_c$ tells us what proportion of images within the same relevance class are directly adjacent.

We hope that the average vertex degree of $G_c$ is larger than for a graph induced by a random set of vertices of $G$. To take extremes, the average vertex degree of a complete subgraph of order $|G_c|$ is $|G_c| - 1$, since each vertex is connected to every other. For a graph induced by a randomly chosen vertex set of $G$ the average vertex degree is to a good approximation equal to $|G_c| \times \bar{d}/|G|$ where $\bar{d}$ is the average vertex degree of $G$. This is typically much smaller than $|G_c| - 1$.

**Connectivity:** The last property which we shall consider is the order of the largest strongly connected component of the subgraphs. Strongly connected components can be thought of as a partitioning of $V$ into equivalence classes $V_i$, $1 \leq i \leq r$ such that vertices $v$ and $w$ are equivalent if and only if there is a path from $v$ to $w$. Let $E_i$, $1 \leq i \leq r$ be the set of arcs with head and tail in $V_i$. The graphs $G_i = (V_i, E_i)$ are called the strongly connected components of $G$.

This property may seem at first to be the least obvious. Upon reflection, however, it turns out to be perhaps the most desirable. If relevance subgraphs had the tendency to be strongly connected, or at least, that their largest strongly connected components comprised a large proportion of the vertex sets, users who found a relevant vertex, would be able to navigate to all other relevant images by following what might be called relevance paths, sequences of vertices each of which is relevant to their information needs. Connectivity is more important than average distance, for even if relevant vertices were only separated by a few clicks, the path connecting them would potentially lead through non-relevant territory. A number of very efficient algorithms have been devised to find strongly connected components. Perhaps the most elegant one has been reported in [1]. It involves two depth-first searches, the first carried out on $G$, the second on the graph that is obtained by reversing the direction of each arc in $G$.

Note that these three properties are only partly correlated. We can for example think of subgraphs with average distance of 2 with no path between any two vertices, or a subgraph with only one strongly connected component but with very low average vertex degree and large distance.

### 4.2 Image Collections and Relevance Classes

Our analysis is based on two collections. The first collection is a subset of the Corel 380,000 Gallery containing 6192 images classified into 63 categories [13]. A limitation of Corel results from the fact that images are assigned to only one category although alternative classifications are often conceivable. While this makes evaluation easier, it clearly fails to model an important aspect of images, which is their semantic ambiguity. To address this drawback, we decided to build a second collection of a more heterogenous kind with richer annotation than Corel. We have used the Getty collection accessible at http://www.getty.com as a source of such images. The system returns a thumbnail for each image along with a rich set of annotations pertaining to style, concepts and low-level characteristics. A broad search for "photography -nobody" retrieves images of the most diverse kind. Our Getty collection contains 8200 images with a raw vocabulary of a similar order. The vocabulary is reduced by retaining only those terms that are associated with at least 20 and at most 100 images. In addition, we discard terms that we deem either too hard (e.g. "freshness") or too easy (e.g. "blue") for retrieval purposes. After these pruning steps the vocabulary has shrunk to 100 terms which we now regard as class labels. Examples of classes are "Cityscapes", "Dunes", "Streetlights" and "Flock of birds" (images are available at http://faya.doc.ic.ac.uk:8800/images/getty).

### 4.3 Results

It is important to be clear about the reference with respect to which we judge the significance of the observed values. For the absolute values are by themselves relatively uninformative as they depend not only on the structure of the relevance subgraph but crucially on the structure of the whole network, which at the present stage is not our interest. Our hypothesis is that the compactness of relevance subgraphs as measured in terms of the three above properties differ from that of subgraphs induced by a randomly chosen set of vertices (to which we refer as random subgraphs). This hypothesis can easily be tested. For each relevance class we determine the average distance, average vertex degree and degree of connectivity not only of the corresponding relevance subgraph, but also on a random subgraph of the same order. We present the results for each of the three properties in turn. A summary of the results is found in Table 1.

**Average distance:** The average distance between any two vertices of the relevance subgraph turns out to be very similar to that for a random subgraph. The results are best displayed in the form of a boxplot (Figure 1). The upper and lower boundary of the box mark the 25th and 75th percentile of the data. The notches in the box are graphic confidence intervals about the medians. The average distance in the relevance subgraphs hovers around 2.8 for both collections, compared to 3 and 3.6 for random subgraphs. Although the magnitude of the difference is small, it is highly significant. The result tells us that the other members of one's class are on average only 2.8 vertices away. If we display during the browsing process the focal image and its nearest neighbours (as in [10]), this

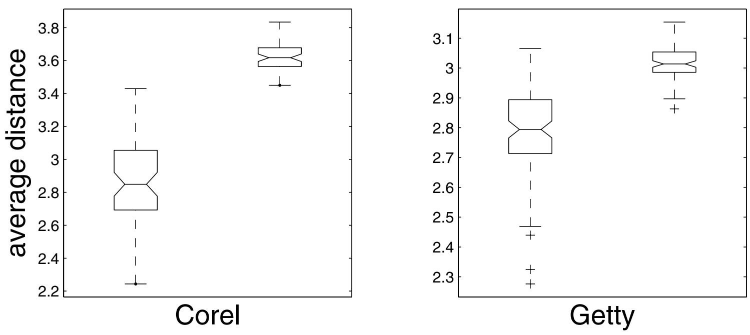


**Fig. 1.** Average distance between vertices in relevance subgraphs (left) and random subgraphs (right)

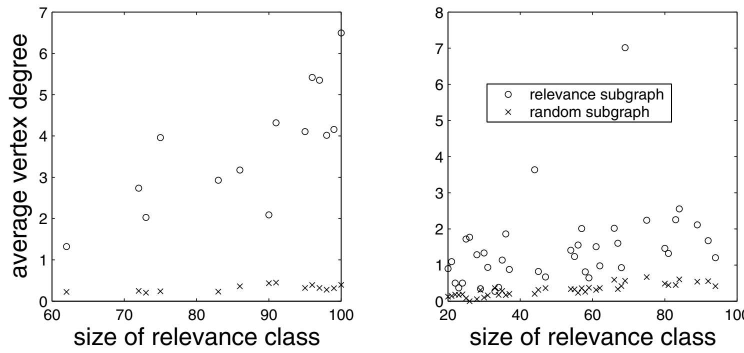


**Fig. 2.** Average vertex degree for Corel (left) and Getty (right)

means that after only two clicks the relevant image comes within sight. Note of course that the shortest path may well include non-relevant images and that it is not necessarily obvious which path one ought to follow to get to the target image.

**Average vertex degree:** We have just seen that vertices of a relevance subgraph tend to be closer than vertices in a random subgraph. It seems likely therefore that in a relevance subgraph a larger proportion of vertices are directly adjacent, and hence that the average vertex degree is higher. This is indeed the case. Figure 2 plots the average vertex degree against the size of the relevance class (i.e. the order of the subgraph). Because there are typically many classes having the same number of images (there are, for example, 42 classes of Corel of size 100), we only plot their average such that for each class size there is at most one point. Note that the vertex degree increases with the order of the subgraph, both for the relevance subgraphs and, less conspicuously, for the random subgraphs. Indeed, if we were to increase the class size further, the average degree would converge towards the average vertex degree of $G$. More importantly, note that the average degree of the relevance subgraphs is considerably larger than that for random subgraphs. For a random subgraph every vertex is connected on average to less than one other vertex: most vertices have no neighbours. In relevance subgraphs, each vertex is linked to at least one other vertex (1.4 and 5.6 for Getty and Corel, respectively), suggesting that the structures do form connected wholes.

**Connectivity:** The results so far give us some indication of whether the subgraphs tend to be strongly connected. The high average vertex degree of 5 for the Corel relevance subgraphs is particularly suggestive. Figure 3 plots the order of the largest strongly connected component against the size of the relevance class. As before, we average over all classes of the same size. Unlike random subgraphs, relevance subgraphs have remarkably large strongly connected components. The relative proportion of images of a class contained in its largest component lies around 85% for Corel and 26% for Getty. Strong connectedness is a rather strong condition for some images that do not belong to the largest component can nevertheless be

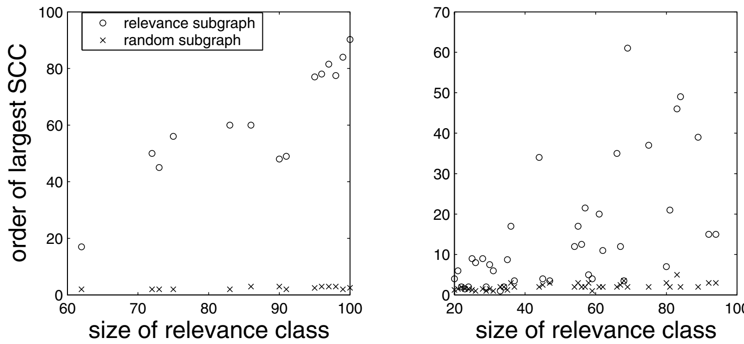


**Fig. 3.** Order of largest strongly connected component for Corel (left) and Getty (right)

reached from it. Thus, the actual number of images that are accessible from any one image of the component typically exceeds the order of the component. Figure 4.3 illustrates this. The subgraph corresponds to the "Flowers" class of Getty. Images that belong to the same strongly connected component are linked by black lines. Images without such lines form a component on their own. Gray lines are connections between components. The largest component is on the left with 14 of the 28 images. Of the remaining images, two can be reached from it via the gray lines.

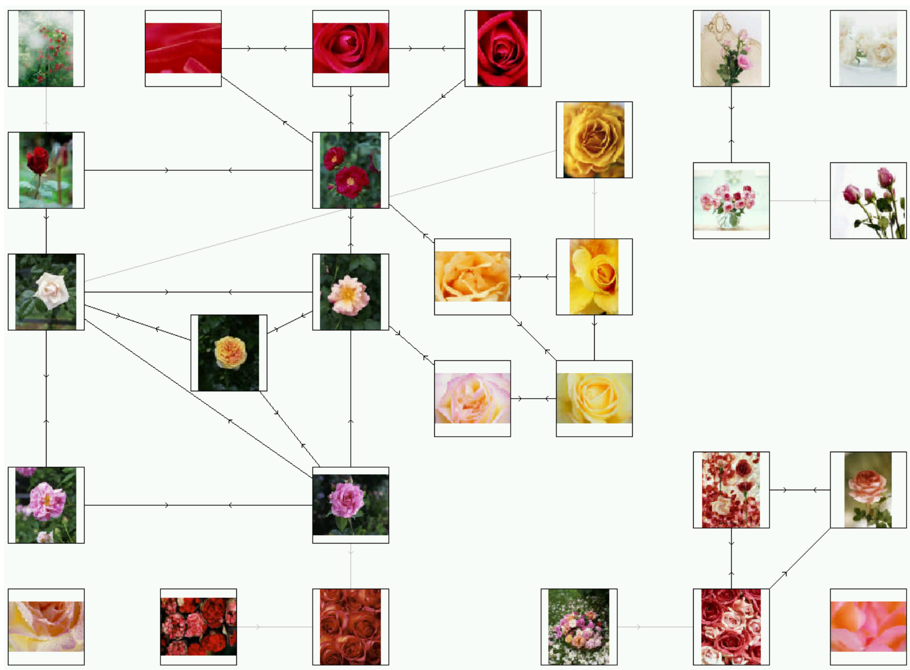

**Fig. 4.** The subgraph corresponding to the "Flowers" class of the Getty collection

**Table 1.** Synopsis of compactness measurements for the Getty and Corel collections

| | | Getty | Corel |
|---|---|---|---|
| **Average distance** | random | 3.0156 | 3.6251 |
| | relevance | 2.7863 | 2.8503 |
| | complete | 1 | 1 |
| | *p*-value | <.0001 | <.0001 |
| **Average vertex degree** | random | 0.2708 | 0.3896 |
| | relevance | 1.4394 | 5.6438 |
| | complete | 44 | 96 |
| | *p*-value | <.0001 | <.0001 |
| **Largest strongly connected comp.** | random | 0.0490 | 0.0253 |
| (proportion of images it contains) | relevance | 0.2583 | 0.8435 |
| | complete | 1 | 1 |
| | *p*-value | <.0001 | <.0001 |

## 5 Conclusions

We have previously proposed a novel browsing structure for searching image collections which we called $NN^k$ networks. We have had the chance to demonstrate their effectiveness on large image collections and attributed their strength to the presence of small-world properties [10]. This paper analyses the networks from a semantic viewpoint and reveals that the structural idiosyncrasies extend beyond small-world properties. The distribution of images belonging to the same semantic class is highly non-random and remarkably compact as judged by three compactness measures that we propose: average distance, average vertex degree and degree of connectivity. Even though the distance in the network between any two images of the same class is on average not much smaller than between any two randomly chosen images, the structure of the graph differs nonetheless substantially from random subgraphs. In particular, images of the same class tend to establish large strongly connected components that in the case of Corel contain an average of 85% of all class members. Hence, even though an image does not tend to be directly adjacent to a vast number of others of its class, most of these can be reached by following paths within the strongly connected component. This semantic analysis has provided a theoretical explanation for the usefulness of $NN^k$ networks as a structure for content-based image browsing.

## References


1. A V Aho, J E Hopcroft, and J D Ullman. *Data Structures and Algorithms*. AddisonWesley, 1983.
2. J R Bach, C Fuller, A Gupta, A Hampapur, B Horowitz, R Humphrey, and C Shu. Virage image search engine: An open framework for image management. In *SPIE Conf on Storage and Retrieval for Image and Video Databases IV*, volume 2670, pages 76–87, 1996.

3. I Campbell. *The ostensive model of developing information needs*. PhD thesis, University of Glasgow, 2000.
4. K Cox. Information retrieval by browsing. In *Proc 5th Int'l Conf on New Information Technology*, 1992.
5. B Croft and T Parenty. Comparison of a network structure and a database system used for document retrieval. *Information Systems*, 10:377–390, 1985.
6. M Flickner, H Sawhney, W Niblack, J Ashley, Q Huang, B Dom, M Gorkani, J Hafner, D Lee, D Petkovic, D Steele, and P Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, pages 23–32, 1995.
7. D Forsyth, J Malik, M Fleck, H Greenspan, and T Leung. Finding pictures of objects in large collections of images. In *Int'l Workshop on Object Recognition for Computer Vision*, 1996.
8. D Heesch, M Pickering, A Yavlinsky, and S Rüger. Video retrieval within a browsing framework using keyframes. In *Proc TRECVID 2003*, 2004.
9. D Heesch and S Rüger. $\mathrm{NN}^k$ networks for content-based image retrieval. In *Proc ECIR*, pages 253–266. LNCS 2997, Springer, 2004.
10. D Heesch and S Rüger. Three interfaces for content-based access to image collections. In *Proc CIVR*, pages 491–499. LNCS 3115, Springer, 2004.
11. W Niblack, R Barber, W Equitz, M D Flickner, E H Glasman, D Petkovic, P Yanker, C Faloutsos, G Taubin, and Y Heights. Querying images by content, using color, texture, and shape. In *SPIE Conf on Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–187, 1993.
12. A Pentland, R W Picard, and S Sclaroff. Photobook: content-based manipulation of image databases. *Int'l Journal on Computer Vision*, 18(3):233–254, 1996.
13. M Pickering and S Rüger. Evaluation of key-frame based retrieval techniques for video. *Computer Vision and Image Understanding*, 92(1):217–235, 2003.
14. Y Rubner, L J Guibas, and C Tomasi. The earth mover's distance, multidimensional scaling, and color-based image retrieval. In *DARPA Image Understanding Workshop*, 1997.
15. Y Rui and T S Huang. Optimizing learning in image retrieval. In *Proc IEEE Conf on Computer Vision and Pattern Recognition*, 2000.
16. Y Rui, T S Huang, and S Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proc IEEE Int'l Conf on Image Processing*, 1997.
17. Y Rui, T S Huang, M Ortega, and S Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. In *IEEE Trans on Circuits and Video Technology*, 1998.
18. S Santini, A Gupta, and R Jain. Emergent semantics through interaction in image databases. *IEEE Trans on Knowledge and Data Engineering*, 13(3):337–351, 2001.
19. S Santini and R Jain. Integrated browsing and querying for image databases. *IEEE MultiMedia*, 7(3):26–39, 2000.
20. A Smeulders, M Worring, S Santini, A Gupta, and R Jain. Content-based image retrieval at the end of the early years. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 22(12):1349–1379, 2000.