INTEGRATED GLOBAL AND OBJECT-BASED IMAGE RETRIEVAL USING A MULTIPLE EXAMPLES QUERY SCHEMA

First Author Name, Second Author Name
Institute of Problem Solving, XYZ University, My Street, MyTown, MyCountry
f_author@ips.xyz.edu, s_author@ips.xyz.edu

Third Author Name
Department of Computing, Main University, MySecondTown, MyCountry
t_author@xy.mu.edu

Keywords: Content-based image retrieval, object-based image retrieval, multiple examples query, ROI extraction, interface.

Abstract: Conventional content-based image retrieval (CBIR) systems typically do not consider the limitations of the feature extraction-distance measurement paradigm when capturing a user's query. This issue is compounded by the complicated interfaces that are featured by many CBIR systems. The framework proposed in this work embodies new concepts that help mitigate such limitations. The front-end includes an intuitive user interface that allows for fast image organization though spatial placement and scaling. Additionally, a multiple-image query is combined with a region-of-interest extraction algorithm to automatically trigger global or object-based image analysis. The relative scale of the example images are considered to be indicative of image relevant and are also considered during the retrieval process. Experimental results demonstrate promising results.

1 INTRODUCTION

Among the different types of queries used in content-based image retrieval (CBIR) systems, the most widely adopted is *query by example* (QBE). In this approach, the user presents a *query image* (also known as an *example image*) to the system and expects similar or relevant images as the result. The framework of a general QBE system can be summarized in the following steps:

1. One or more features, such as color, texture, or spatial structure, are extracted from images in the image database and from the query image. These low-level characteristics are stored in a *feature vector* (FV).
2. A distance function compares the query image FV to all FVs in the database - the ultimate measure of image similarity.
3. The images in the database are sorted according to their calculated distances, from low (most similar) to high (least similar).
4. Finally, the first t most similar images are presented to the user. This is called the *retrieved set*. Usually t is returned by a cut function, but a constant can be also be used.

QBE is efficient because it is a compact, fast and generally natural way for specifying a query. While keyword or text based queries can be effective in very narrow domains, QBE is useful in broad databases where verbal specifications of the query are imprecise or impractical (Castelli and Bergman, 2002).

However, the QBE framework is not always accurate in capturing the user's true intentions for providing a particular query image. The main reasons for this are the limitations related to the feature extraction and distance measurement steps in the previous list. The user's intended query information is not always perceived in the extracted feature of the images and, hence, FV distances are not guaranteed to be correct. Similarity judgement based on a single extracted quantity (distance) is gross and ineffective reduction of the human user's desires.

Another point of weakness in QBE is the inherent difficulty in translating visual information into the semantic concepts that are understood by the human user. Indeed, this is one of the central challenges of the visual information retrieval field, commonly referred to as the *semantic gap* (Smeulders et al., 2000).

Several approaches have been purposed to overcome the semantic gap including the use of image descriptors that best approximate the way in which humans perceive visual information (Manjunath et al., 2001), (Renninger and Malik, 2004), or the inclusion of a user's willingness as a dynamic part of the system (Rui et al., 1998), (Santini and Jain, 2000).

Another approach to problem considers that, in many situations, the user is focused on an object, or *region-of-interest* (ROI) within the image. Such systems are named region- or object-based image retrieval and they perform a search based on local instead of global image features (Carson et al., 2002), (García-Pérez et al., 2006). Moreover, it is possible to use more than a single example image when performing a QBE. This technique is called *multiple example query* (Assfalg et al., 2000), (Tang, 2003).

In this work a new method for combining a multiple example query on both global- and region-based scenarios is presented. An attention-driven ROI extraction algorithm (Marques et al.,) and a new interface, the Perceptually-Relevant Image Search Machine (PRISM), are used. The novelty behind the PRISM interface is that it allows, in a simple way, to select images from a database and scale them according to user interest and perceived relevance.

The remainder of this paper presents and overview of the system (Section 2), experimental results (Section 3), and, finally, conclusions (Section 4).

2 ARCHITECTURE

A block diagram of our proposed QBE framework is given in figure 1.

2.1 Interface level: PRISM

PRISM is a general environment that allows the capture of a user's relative interest in particular images. PRISM also enables the incorporation of a variety of image retrieval including content-based image retrieval, content-free image retrieval, and semantic annotation (Mayron et al., 2006). The initial view of the PRISM desktop is shown in figure 2. The top part of the interface is the "filmstrip", the only source of new images. The user drags images from the filmstrip into the main content area. An image may be deleted from the filmstrip by dragging it to the trash can icon in the lower-right corner. When an image is removed from the filmstrip the vacant space is immediately replenished, ensuring that the filmstrip is always full. The lower section of the PRISM interface is the tabbed

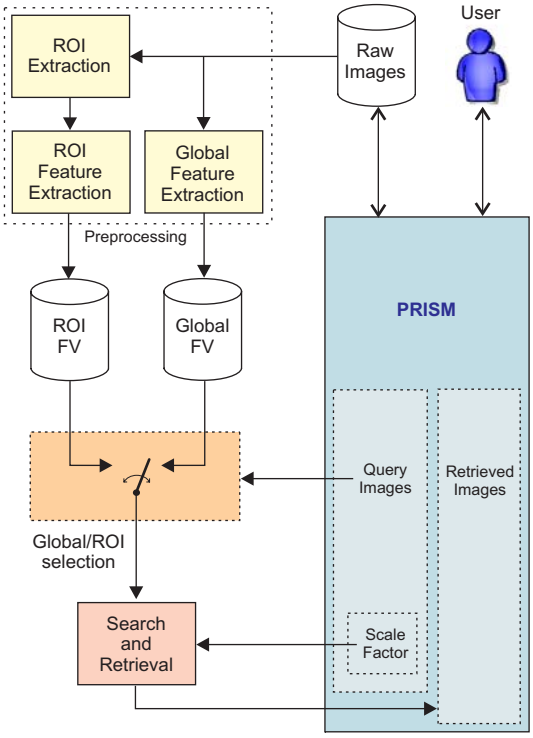

Figure 1: A general view of the system architecture

workspace. In PRISM tabs are used to organize individual groups of images, expanding the work area while avoiding overwhelming the user with too many visible images at one time.

Figure 2: The PRISM interface

The proposed CBIR system takes advantages of PRISM's ability to capture subjective user queries expressed by grouping and scaling images. Two or more example images are dragged from the filmstrip to the workspace. The selected images are than scaled by the user according to their relevance. That is, larger images indicate increasing relevance. From this point onwards a QBE is performed, taking into account user interest based on ROI (local) or global characteristics of the images as well as image scale. The system is

able to clearly capture user query concepts, deciding automatically between a global- or ROI-based search using image scale factors.

2.2 Preprocessing

In the offline preprocessing stage image are segmented by an attention-driven ROI extraction algorithm. This algorithm performs a set of morphological operation over the computed Itti-Koch (Itti et al., 1998) and Stentiford (Stentiford, 2001) models of visual attention. The result is a very good ROI (or object segmentation). See Figure 3 for an example. Since we are currently working with a database that contains *salient by design* objects, segmented salient regions should always correspond to the semantic objects in the images. However, occasionally, the ROI extraction output was refined by removing false positives.

Figure 3: Input vs. output example for the ROI extraction algorithm

Another interesting point of this algorithm is that it does not make use of any *a priori* object information, such as shape or color, running in a fully unsupervised way (a complete description can be found in (Marques et al.,)). After ROI extraction the global and ROI FVs are computed by *feature extraction* modules, figure 1. Both use the same descriptor: a 256-cell quantized HMMD (MPEG-7-compatible) color histogram (Manjunath et al., 2001). The computed FVs are stored in the *global* and *ROI* FV databases.

2.3 Global/ROI selection

If more than one query image is presented in the PRISM workspace a decision process takes place. The aim of this *global/ROI selection* decision is to select the global or ROI information for *search and retrieval* module input. This block compares the query images FVs and fires a global- or ROI-based search accordingly. Figure 4 depicts its operation.

In the case of the input example images in the top of Figure 4, the user's ROI-based search intention is clear, since the tennis ball's (ROIs) features are more similar between themselves than the global features. A simple approach based on the average coefficient of determination (squared correlation, r^2) is used for detecting the FVs degree of similarity. The r^2 ranges from 0 to 1 and represents the magnitude of the linear relationship between two vectors.

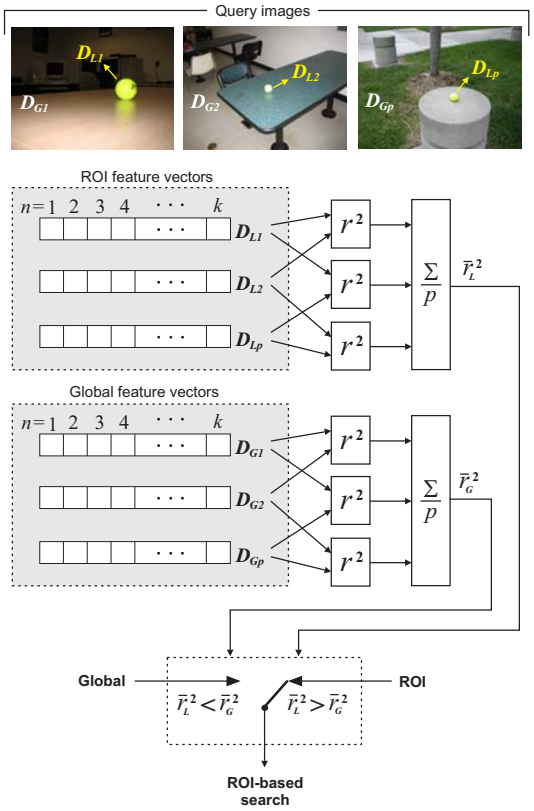

Figure 4: Functional diagram for Global/ROI selection and example for 3 input images (p=3). For these query images, an ROI-based search will be performed. G-Global, L-Local

In Figure 4, given $p(>1)$ query images, two independent groups of $k-positions$ ($k=256$) FVs are considered: one from the ROIs, $D_{Li}(n)$, and other from the global images, $D_{Gi}(n)$, where i is the query image, with $i \in \{1,\ldots,p\}$ and $n \in \{1,\ldots,k\}$. The coefficient of determination, $r_s^2(c)$, within each group, for all FVs pairs is given by equation (1).

$$r_s^2(c) = \frac{a}{ef} \tag{1}$$

where

$$a = \left[k\sum_{n=1}^{k} D_{sx}(n)D_{sy}(n) - \sum_{n=1}^{k} D_{sx}(n)\sum_{n=1}^{k} D_{sy}(n)\right]^2 \tag{2}$$

$$e = k\sum_{n=1}^{k}\left[D_{sx}(n)\right]^2 - \left[\sum_{n=1}^{k} D_{sx}(n)\right]^2 \tag{3}$$

$$f = k\sum_{n=1}^{k}\left[D_{sy}(n)\right]^2 - \left[\sum_{n=1}^{k} D_{sy}(n)\right]^2 \tag{4}$$

s denotes the group, with $s \in \{L,G\}$, c is the number of combinations of the p feature vectors, taken 2 at a time (x and y), $c \in \{1,\ldots,C_p^2\}$ and

$$C_p^2 = \frac{p!}{2(p-2)!}. \quad (5)$$

The average coefficients of determination, $\bar{r}_s^2$, of each group are then compared. Groups with high $\bar{r}_s^2$ value means that the FVs are more similar and hence more similar are the raw images.

2.4 Search and Retrieval

Once the search type is set the *search and retrieval* stage (Figure 1) can finally be performed. Figure 5 illustrates the operations for $p > 1$ general query images, Q_i.

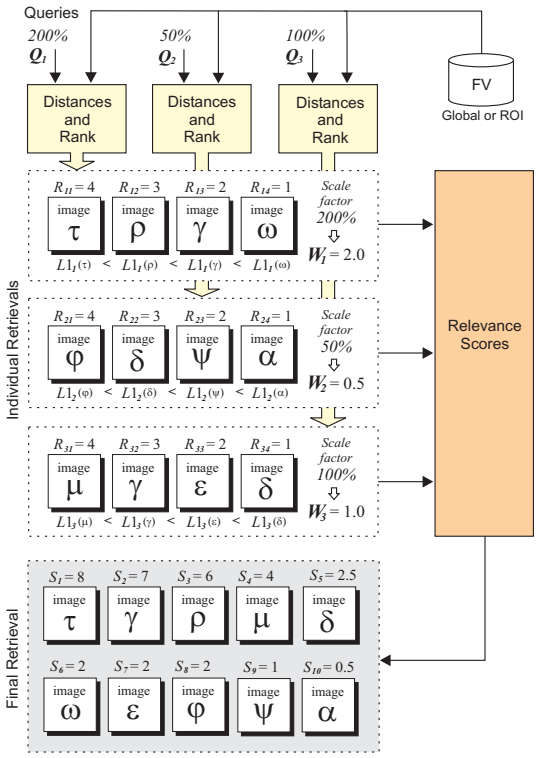

Figure 5: Functional diagram of the *search and retrieval* module of the system. Example for 3 queries ($p = 3$), 4 images retrieved per query ($t = 4$) and arbitrary scale factors of 200, 50 and 100%. Note that the image γ appears in individual retrievals 1 and 3, so their relevance scores are summed. A similar operation is done to image δ, that appears in retrievals 2 and 3. Images with the same S_j have relevances proportional to their Wi, as happens to images ω, ε and φ.

In the first step, individual retrievals of a fixed number of t images are made for each query. The distance between Q_i FV, $D_i(n)$, and all database images FVs, $D_b(n)$, is computed using the $L1$ measure:

$$L1_i(b) = \sum_{n=1}^{k} |D_i(n) - D_b(n)|, \quad (6)$$

where i is the query image and b the database image. The t most relevant images, R_{ih}, are ranked from the most (smaller distance) to the least similar, according to

$$R_{ih} = t - h + 1, \quad (7)$$

where h is the retrieved image, with $h \in \{1,\ldots,t\}$. Block *relevance score* in figure 5 groups these individual retrievals into a *final retrieval*. The system looks at user's subjective degree of relevance, represented by query images scales captured by PRISM. This is achieved using the scale factor (perceptual resize) of Q_i as a weight W_i, which is multiplied by each rank R_{ih}. The result of this weighting operation is a relevance score S_j, given by:

$$S_j = W_i R_{ih}, \quad (8)$$

where j is the image into the final retrieval, with $j \in \{1,\ldots,u\}$ and u is the number of different images among all individual retrievals. If the same image appears in different retrievals the S_j are summed, so as to increase it's relevance and assure a single occurrence of this image into the final retrieval.

In the case of images with the same S_j, their relevance is treated as follows: a) if they come from individual retrievals with different Wi, the one with the greater Wi is considered more relevant; b) if they come from individual retrievals with the same Wi, the most relevant is the one which was queried first (its correspondent query image was pushed first into the workspace).

Note that the use of a single example image does not make sense here since it is not possible to decide wether local or global features are to be inspected.

3 EXPERIMENTAL RESULTS

In this section, experimental results of the proposed system are shown. The examples in figure 6 cover different query scenarios, providing a good view of the system performance. The number of retrieved images per individual query is $t = 5$ for all experiments.

3.1 Database

The *raw images* database consist of 315 images with one salient object per image. In the database, there are five different semantic ROIs categories: mini basketball, blue plate, yellow sign, tennis ball and red ground objects. The use of a *salient by design* objects database is important for a meaningfully analysis of the system operation and results.

3.2 Discussion

In figure 6, Query a, two query images of outdoor red objects over different backgrounds were specified by the user. This clearly denotes users interest on local features of the images (red objects). The one with the red paper box, Q_1, was 200% resized, thus $W_1 = 2.0$. Q_2 was not resized, so it's scale factor is 100% and $W_2 = 1.0$. The first point to be observed about the retrieved set, is that the main concept delineated in the query by the user was correctly captured: "give me the images with red objects, no matters the background." Besides, users emphasis on Q_1, stating "red paper box are more relevant", has also been covered (since these objects appears first, with higher relevance scores S_j).

On the other hand, the example in figure 6, Query b, illustrates the case where global attributes of the query images are more relevant than the local. While the ROIs (orange mini basketball and tennis ball) exhibit significant differences in their features, the global features are more or less constant (concrete structures and grass). In the retrieved image set, images with similar global structures can be seen, regardless the different small salient objects present (a blue plate, mini basketball and tennis ball). We also highlight in this example, the strong emphasis on query Q_1, with $W_1 = 2.5$, and the *attenuation* on Q_2, with $W_2 = 0.5$. The gist of this search could be translated as: "I'm interested in outdoor concrete bases. Something such as this cylinder is ok, but a table like this would be better!" Again, the system was able to take into account the users query idea, by means of the *relevance scores* approach.

The example in figure 6, query c, shows a query with three images, where a tennis ball is the common feature. In spite of queries Q_2 and Q_3 share also global attributes (a blue table), the system was still able to correctly decide for a ROI-based search. As can be seen in retrieved set, all images contain a tennis ball, regardless the differences in their context (background). About the subjective scaling parameter of the query, the large scale factor on Q_1, against the reductions on Q_2 and Q_3, denotes that the tennis balls with a stamped black brand are of special interest. A close look in the first four images of the retrieved set shows that this query specification was attended. Finally, note that the system should return 15 images in the final retrieval, but 12 images were presented. In this example, this occurred because 3 of the 15 images in the *individual retrievals* (figure 5) appeared twice. So, that repeated images had their relevance scores summed, and appeared just once in the final retrieval.

4 CONCLUSION

The architecture presented in this paper incorporates different techniques for visual content access: object-based image retrieval, the ordinary global-based image retrieval and multiple examples query. Most CBIR systems use these approaches isolated or weakly integrated, while here they were truly combined. Moreover, by using the PRISM interface it is possible to get explicit information from the user about the relevance of each example image. This is done by means of images scale. Taking these characteristics together the system is able to more faithfully capture what users' have in mind when formulating a query, as was demonstrated by experimental results.

The obtained results should encourage CBIR developers to put effort not only towards the traditional feature extraction-distance measurement paradigm, but also into the improvement of the architectural aspects concerning the capture of user query concepts.

In the next stages of this work we intend to explore in more detail the information provided by the user to the interface, such as the arrangement of example images and the features of images dragged to the trash can.

ACKNOWLEDGEMENTS

This research was partially sponsored by UOL (www.uol.com.br), through its UOL Bolsa Pesquisa program, process number 200503312101a and by the Office of Naval Research (ONR) under the Center for Coastline Security Technology grant N00014-05-C-0031.

REFERENCES

Assfalg, J., Bimbo, A. D., and Pala, P. (2000). Using multiple examples for content-based image retrieval. In *IEEE International Conference on Multimedia and Expo (I)*, pages 335–338.

Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on PAMI*, 24(8):1026–1038.

Castelli, V. and Bergman, L. D. (2002). *Image Databases: Search and Retrieval of Digital Imagery*. John Wiley & Sons, Inc., New York, NY, USA.

García-Pérez, D., Mosquera, A., Berretti, S., and Bimbo, A. D. (2006). Object-based image retrieval using active nets. In *ICPR (4)*, pages 750–753. IEEE Computer Society.

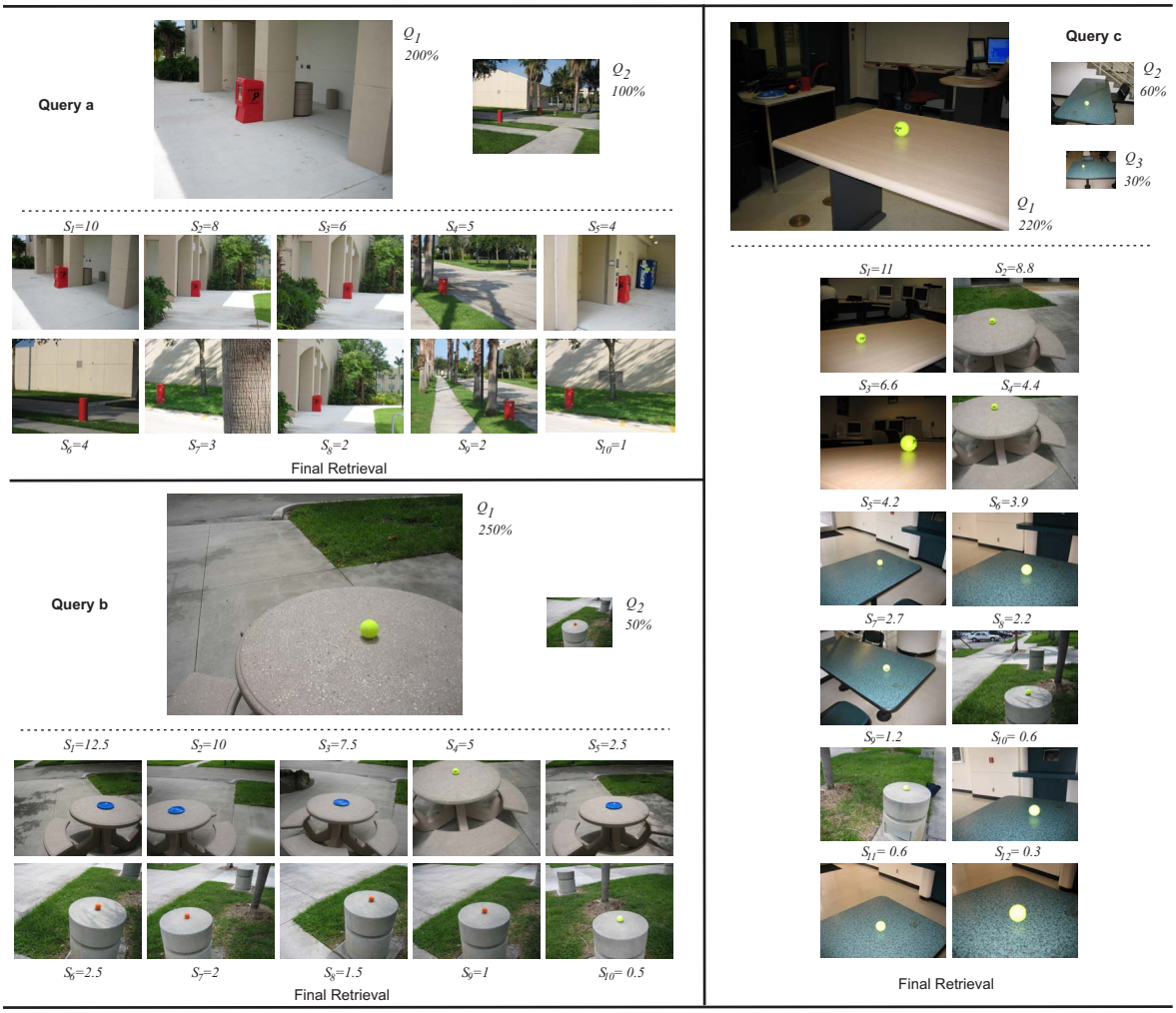

Figure 6: Three examples of queries. Queries a and c results in a ROI-based search, while query b in a global-based search.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on PAMI*, 20(11):1254–1259.

Manjunath, B. S., Ohm, J. R., Vinod, V. V., , and Yamada, A. (2001). Color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology, Special Issue on MPEG-7*, 11(6):703–715.

Marques, O., Mayron, L. M., Borba, G. B., and Gamba, H. R. An attention-driven model for grouping similar images with image retrieval applications. *EURASIP Journal on Applied Signal Processing (accepted)*.

Mayron, L. M., Borba, G. B., Nedovic, V., Marques, O., and Gamba, H. R. (2006). A forward-looking user interface for cbir and cfir systems. In *IEEE International Symposium on Multimedia (ISM2006)*, San Diego, CA, USA.

Renninger, L. W. and Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, 44(19):2301–2311.

Rui, Y., Huang, T., Ortega, M., and Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655.

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

Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. on PAMI*, 22(12):1349–1380.

Stentiford, F. (2001). n estimator for visual attention through competitive novelty with application to image compression. In *Picture Coding Symposium, pp. 25-27*, Seoul, Korea.

Tang, J. Acton, S. (2003). An image retrieval algorithm using multiple query images. In *ISSPA'03: Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, pages 193–196. IEEE.