Scale-Space Matching and Image Retrieval*

S. Ravela and **R. Manmatha** and **E. M. Riseman**
Computer Vision Research Laboratory
and
Center for Intelligent Information Retrieval
University of Massachusetts, Amherst, MA 01003
{ravela,manmatha}@cs.umass.edu

Abstract

The retrieval of images from a large database of images is an important and emerging area of research. Here, a technique to retrieve images based on appearance that works effectively across large changes of scale is proposed. The database is initially filtered with derivatives of a Gaussian at several scales. A user defined template is then created from an image of an object similar to those being sought. The template is also filtered using Gaussian derivatives. The template is then matched with the filter outputs of the database images and the matches ranked according to the match score. Experiments demonstrate the technique on a number of images in a database. No prior segmentation of the images is required and the technique works with viewpoint changes up to 20 degrees and illumination changes.

1 Introduction

The advent of multi-media and large image collections in several different domains brings forth a necessity for image retrieval systems. These systems will respond to visual queries by retrieving images in a fast and effective manner. The application potential is enormous; ranging from database management in museums and medicine, architectural and interior design, image archiving, to constructing multi-media documents or presentations[Gudivada 95].

Simple image retrieval solutions have been proposed, one of which is to annotate images with text and then use a traditional text-based retrieval engine. While this solution is fast, it cannot however be effective over large collections of complex images. The variability and richness of interpretation is quite enormous as is the human effort required for annotation.

To be effective an image retrieval system should exploit image attributes such as color distribution, motion, shape [Flickner 95], structure, texture or perhaps user drawn sketches or even abstract token sets (such as points, lines etc.). Representations of these attributes can be matched to gauge similarity and perhaps also be used to index the images. Image retrieval can be viewed as an ordering of match scores that are obtained by searching through the database. Therefore, the key challenges in building a retrieval system are the choice of attributes, their representations, query specification methods, match metrics and indexing strategies.

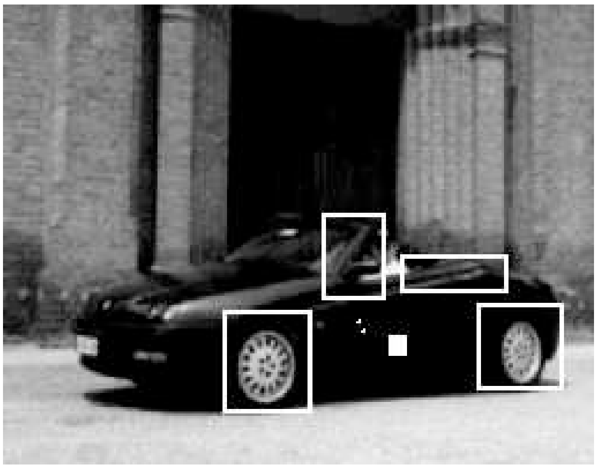


Figure 1: Construction of a query begins with a user marking regions of interest in an image, shown by the rectangles.

In this paper a method for retrieving images based on appearance is presented. Without resorting to token feature extraction or segmentation, images are retrieved in the order of their *similarity in appearance* to a *query*.

Query construction begins with the user selecting regions in an image. An example is shown in Figures 1 and 2. Here, the user wishes to retrieve images similar in view and shape (appearance) to the car shown in Figure 1. In order to do so, the user

* This work was supported in part by the Center for Intelligent Information Retrieval, NSF Grants IRI-9208920, CDA-8922572 and ARPA N66001-94-D-6054

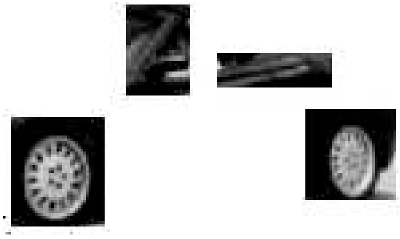


Figure 2: The regions of interest and their spatial relationships define a query.

outlines salient regions (in his or her opinion) on the image(shown as rectangles in Figure 1). These regions along with their spatial relationship are conjunctively called as the query(Figure 2)[1].

Similarity of appearance is quite simply the similarity of shape under small view variations. While the proposed definition constrains view variations, there is however, no constraint imposed on scale variations. That is, the image of an object in the database can be very different in size from the image of the object in the query. This could happen due to variations in resolution of the image or due to the object or scene being imaged from different distances. The variation in scale is particularly important in image databases since no control can be exerted over the image acquisition process. Any appearance based retrieval system must therefore address this fundamental issue. In order to measure the similarity of appearance between a query and an image, two issues must be addressed. First, appropriate representations of images must be chosen and second, a mechanism for matching these representations must be developed.

Filtered versions of images are used as representations of appearance. In particular a *vector representation*(VR) of an image is obtained by associating each pixel with a vector of responses to Gaussian derivative filters of several different orders. A single VR is the basic representation that can be used to retrieve images but, under a fixed scale. To retrieve images under varying scale a representation over the scale parameter is required and scale-space representations [Lindeberg 94] are a natural choice. Lists of VRs generated using banks of Gaussian derivative filters at several different scales form a scale-space representation of the object. This scale-space representation is used to retrieve objects under large (but not arbitrary) scale variations. In particular, this paper demonstrates retrieval for scale changes up to a factor of 4 (1/4 to 4 times the query size). The choice of Gaussians and their derivatives to derive representations of appearance (VRs) is motivated by a number of considerations. It has been argued by Koenderink and others that the structure of an image may be represented using Gaussian derivatives [Koenderink 87]. Hancock et al [Hancock 92] have shown that the principal components of a set of images containing natural structures may be modeled as the outputs of a Gaussian and its derivatives at several scales. That is, there is a natural decomposition of an image into Gaussian derivatives at several scales. Gaussians and their derivatives have, therefore, been successfully used for matching images of the same object under different viewpoints [Bergen 92, Werkhoven 90a, Werkhoven 90b, Kass 88, Manmatha 94, Rao 95]. This paper is an extension to matching "similar" objects using Gaussian derivatives.

Images are matched by correlating their vector-representations. VR matching is robust to lighting variations and tolerates small variations in view. In addition, well-designed queries have yielded significant variation in retrieved shapes(see Section 6). It is quite likely that structures similar to that of a query are present in the database at a different scale. As described, the VR matching cannot account for gross changes in scale. VRs generated from filters at several scales are used to search over scale-space for possible scale variations of the query. The range of scale variation as well as the step size is a user-controlled parameter. Scale-space matching is described in detail in Section 4. The entire process of retrieval can be viewed as the following three-step process. The first is an off-line computation step that generates vector-representations of database images for matching. The second is construction of queries and their VRs. The third is an ordering of images ranked by the correlation of their VRs with that of the query.

While one is tempted to argue that retrieval and recognition problems have a lot in common, one should also note the sharp contrasts between the two paradigms. First, putting a user in the "loop" , shifts the burden of the determination of feature saliency to the user. For example, only regions of the car in Figure 1 (namely, the wheels, side-view mirror and mid-section) considered salient by the user are highlighted. Second, user interaction can be used in a retrieval system of sufficient speed to evaluate the ordering of retrieved images and reformulate queries if necessary. Thus, in the approach presented in this paper, alternate regions could be marked if the retrieval is satisfactory. Third, a hundred percent accuracy of retrieval is desirable but not at all critical (for comparison the best text-based re-

[1] The retrieved images for this case are shown in Figure 4.

trieval engines have retrieval rates less than 50%). The user ultimately views and evaluates the results, allowing for tolerance to the few incorrect retrieval instances.

The remainder of this paper is organized as follows. In Section 2 other related approaches are examined. In Section 3 VR matching is described. In Section 4 VR matching is extended to account for scale variations. Then, in Section 5 query construction is discussed. In Section 6 a retrieval is demonstrated on a database with over 300 images containing automobiles, locomotives (steam and diesel), apes and houses. These images obtained mainly over the internet have uncontrolled lighting and viewing geometry.

2 Related Work

This paper is related to a number of threads in the literature. The first concerns matching with Gaussian derivative filters at multiple scales.

The idea of using Gaussian derivatives for matching and recovering local structure was suggested among others by Koenderink [Koenderink 87]. Among filter representations, Gaussian derivatives have a number of advantages - they are steerable [Freeman 91] and separable. The use of multiple derivative filters requires that correlation be performed between vectors. This is discussed by Granlund et al [Granlund 95].

Some of the earliest uses of scale in matching go back to the Gaussian and Laplacian pyramids constructed by Burt and Adelson [Burt 83] and Crowley [Crowley 87]. These pyramids have been used to do coarse to fine matching under translation, affine or more general transforms (see [Bergen 92]). The pyramids speed up the computation as well as performing matching at the appropriate scales. However, as Lindeberg [Lindeberg 94] points out in his extensive discussion of scale space and its properties they do not form a true scale space.

Kass [Kass 88] used the Gaussian and its derivatives at multiple scales for stereo matching. The notion of matching across Gaussians of different scales was used by Manmatha [Manmatha 94] for matching image patches under similarity and affine transforms. He also used the idea of comparing the outputs of Gaussians at different standard deviations to compute large scale changes. Rao and Ballard [Rao 95] used Gaussian derivatives at multiple scales to match a moving object when the viewpoint change was small.

The second thread to which our work is related is the area of image indexing and retrieval. To the best of our knowledge, retrieval on the basis of appearance or shape is almost entirely based on prior segmentation of the object. Examples include the QBIC project at IBM [Flickner 95], the photo book project [Pentland 94] and shape retrieval [Mehrotra 95]. These methods all require knowledge of the contour or binary shape of the object. For specific objects like faces, principal component analysis has been used successfully for representation [Kirby 90] and retrieval [Turk 91]. Using texture measures, Picard et al [Picard 94] are able to classify images into a few distinct categories (e.g. city scene, country scene).

3 Matching Vector Representations

The key processing involves obtaining and matching vector-representations of a sample gray level image patch S and a candidate image C. The steps involved in doing this will now be described:

Consider a Gaussian described by it's coordinate $\mathbf{r}$ and scale σ

$$G(\mathbf{r}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mathbf{r}^2}{2\sigma^2}} \qquad (1)$$

A vector-representation $\vec{V}$ of an image I is obtained by associating each pixel with a vector of responses to partial derivatives of the Gaussian at that location. Derivatives up to the second order are considered. More formally, $\vec{V}$ takes the form $\langle I_x, I_y, I_{xx}, I_{xy}, I_{yy} \rangle$ where I_x, I_y denote the the filter response of I to the first partial derivative of a Gaussian in direction x and y respectively. I_{xx},I_{xy} and I_{yy} are the appropriate second derivative responses.

In this paper, only the first and second derivatives of Gaussians are used. Let us consider 1-D Gaussian derivatives. The odd derivatives are all correlated with each other. This also holds true for the even derivatives which are correlated with each other. However, for the same σ, the first derivative of a Gaussian is uncorrelated with the second derivative of a Gaussian [Kass 88]. Thus picking only the first and second derivatives of Gaussians insures that maximal information is extracted from the image. Gaussian (as opposed to Gaussian derivatives) filters are not used because they are sensitive to the actual intensity value.

The correlation coefficient η between images $\vec{C}$ and $\vec{S}$ at location (m, n) in $\vec{C}$ is given by:

$$\eta(m, n) = \sum_{i,j} \hat{C_M}(i, j) \cdot \hat{S_M}(m - i, n - j) \qquad (2)$$

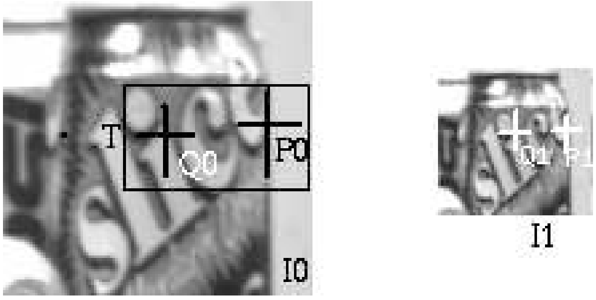


Figure 3: I1 is half the size of I0. To match points p_0 with p_1, Image I_0 should be filtered at point p_0 by a Gaussian of a scale twice that of the Gaussian used to filter image I_1 (at p_1). To match a template from I_0 containing p_0 and q_0, an additional warping step is required. See text in Section 4.

where

$$\hat{S_M}(i,j) = \frac{\vec{S}(i,j) - S_M}{\left\|\vec{C}(i,j) - C_M\right\|}$$

and S_M is the mean of $\vec{S}(i,j)$ computed over S. $\hat{C_M}$ is computed similarly from $\vec{C}(i,j)$. The mean C_M is in this case computed at (m,n) over a neighborhood in C (the neighborhood is the same size as S).

Vector correlation performs well under small view variations. It is observed [Ravela 96] that typically for the experiments carried out with this method, in-plane rotations of up to 20^o, out-of plane rotation of up to 30^0 and scale changes of less than 1.2 can be tolerated. Similar results in terms of out-of-plane rotations were reported by [Rao 95].

4 Matching Across Scales

The database contains many objects imaged at several different scales. For example, the database used in our experiments has several diesel locomotives. The actual image size of these locomotives depends on the distance from which they are imaged and shows considerable variability in the database. The vector correlation technique described in Section 3 cannot handle large scale changes, and the matching technique, therefore, needs to be extended to handle large scale changes.

In Figure 3 image I_1 is half the size of image I_0 (otherwise the two images are identical). Thus,

$$I_0(\mathbf{r}) = I_1(s\mathbf{r}) \tag{3}$$

where $\mathbf{r}$ is any point in image I_0 and $s\mathbf{r}$ the corresponding point in I_1 and the scale change s = 0.5. In particular consider two corresponding points p_0 and p_1 and assume the image is Gaussian filtered at p_0 Then by substituting for I_0 using equation 3 we have:

$$\int I_0(\mathbf{r})G(\mathbf{r}-\mathbf{p_0},\sigma)d\mathbf{r} = \int I_1(s\mathbf{r})G(s\mathbf{r}-\mathbf{p_1},s\sigma)d\,(s\mathbf{r}) * s^{-1} \tag{4}$$

But it can be shown that $G(\mathbf{r},\sigma) = G(s\mathbf{r}, s\sigma)$ [Manmatha 94]. Thus,

$$\int I_0(\mathbf{r})G(\mathbf{r}-\mathbf{p_0},\sigma)d\mathbf{r} = \int I_1(s\mathbf{r})G(\mathbf{r}-\mathbf{p_1},\sigma)d\,(s\mathbf{r}) \tag{5}$$

In other words, the output of I_0 filtered with a Gaussian of scale σ at p_0 is equal to the output of I_1 filtered with a Gaussian of scale $s\sigma$ i.e. the Gaussian has to be stretched in the same manner as the image if the filter outputs are to be equal. This is not a surprising result if the output of a Gaussian filter is viewed as a Gaussian weighted average of the intensity. A more detailed derivation of this result is provided in [Manmatha 94].

The derivation above does not use an explicit value of the scale change s. Thus, equation 5 is valid for any scale change s. The form of equation 5 resembles a convolution and in fact it may be rewritten as a convolution

$$I_0(\mathbf{r}) \star G(.,\sigma) = I_1(s\mathbf{r}) \star G(.,s\sigma) \tag{6}$$

Similar derivations may also be carried out for higher derivatives of Gaussians (see [Manmatha 94]). Here the results for the first and second derivatives of Gaussians are listed. Define the normalized first derivative of Gaussian by

$$\mathbf{G}'(\mathbf{r},s\sigma) = s\sigma \quad dG(\mathbf{r},s\sigma)/d\mathbf{r} \tag{7}$$

The first derivative of the Gaussian has been energy normalized by the term $s\sigma$ so that its energy is the same as that of the Gaussian filter [Werkhoven 90a].

The normalized second derivative of Gaussian may be similarly defined by

$$\mathbf{G}''(\mathbf{r},s\sigma) = (s\sigma)^2 \quad d^2G(\mathbf{r},s\sigma)/d(\mathbf{r}\mathbf{r}^\mathbf{T}) \tag{8}$$

where the term $(s\sigma)^2$ again ensures that the energy of the second derivative Gaussian filter is the same as the energy of the first derivative Gaussian filter and the Gaussian filter.

Note that the first derivative of a Gaussian is a vector and the second derivative of a Gaussian a 2 by 2 matrix.

Then the Gaussian derivatives are related by (see [Manmatha 93])

$$I_s \star \mathbf{G}^{'}(.,\sigma) = I_0 \star \mathbf{G}^{'}(.,s\sigma) \tag{9}$$

and,

$$I_s \star \mathbf{G}''(.,\sigma) = I_0 \star \mathbf{G}''(.,s\sigma) \qquad (10)$$

The above equations are sufficient to match the filter outputs (in what follows assume only Gaussian filtering for simplicity) at corresponding points (for example at $\mathbf{p_0}$ and $\mathbf{p_1}$). A further complication is introduced if more than one point is to be matched while preserving the relative distances (structure) between the points. Consider for example the pair of corresponding points $\mathbf{p_0}, \mathbf{q_0}$ and $\mathbf{p_1}, \mathbf{q_1}$. The filter outputs at points $\mathbf{p_0}, \mathbf{q_0}$ may be visualized as a template and the task is to match this template with the filter outputs at points $\mathbf{p_1}, \mathbf{q_1}$. That is, the template is correlated with the filtered version of the image I_1 and a best match sought. However, since the distances between the points $\mathbf{p_1}, \mathbf{q_1}$ are different from those between $\mathbf{p_0}, \mathbf{q_0}$ the template cannot be matched correctly unless either the template is rescaled by a factor of 1/2 or the image I_1 is rescaled by a factor of 2. The matching is, therefore, done by warping either the template or the image I_1 appropriately.

Thus, to find a match for a template from I_0, in I_1, the Gaussians must be filtered at the appropriate scale and then the image I_1 or the template should be warped appropriately. Now consider the problem of localizing a template T, extracted from I_0, in I_1(see Figure 3). For the purpose of subsequent analysis, assume two corresponding points ($\mathbf{p_0}$,$\mathbf{q_0}$) of interest in T and I_1 ($\mathbf{p_1}, \mathbf{q_1}$) respectively. To localize the template the following three steps are performed.

1. *Use appropriate Relative Scale:* Filter the template and I_1 with Gaussians whose scale ratio is 2. That is, filter T with a Gaussian of scale 2σ and I_1 with σ.

2. *Account for size change:* Sub-sample T by half. At this point the spatial and intensity relationship between the warped version (filtered and sub-sampled) of template points p_0 and q_0 should be exactly same as the relationships between filtered versions of p_1 and q_1.

3. *Translational Search:* Perform a translational search over I_1 to localize the template.

This three step procedure can be easily extended to match VRs of T and I_1 using Equations 9 and 10. In step(1) generate VRs of T and I_1 using the mentioned filter scale ratios. In step(2) warp the VR of T instead of just the intensity. In step(3) use vector-correlation(Equation 2 at every step of the translational search.

Without loss of generality any arbitrary template T can be localized in any I_1 that contains T scaled by a factor s.

4.1 Matching Queries over Unknown Scale

The aforementioned steps for matching use the assumption that the relative scale between a template and an image is known. However, the relative scale between structures in the database that are similar to a query cannot be determined *a priori*. That is, the query could occur in a database image at some unknown scale. A natural approach would be to search over a range of possible relative scales, the extent and step size being user controlled parameters.

One way of accomplishing this is as follows. First, VRs are generated for each image in the database over a range of scales, say $\frac{1}{4}\sigma$,$\frac{1}{2\sqrt{2}}\sigma$,...,$4\sigma$. Then, a VR for the query is generated using Gaussian derivatives of scale σ. The query VR is matched with each of the image VRs, thus traversing a relative scale change of $\frac{1}{4}$...4, in steps of $\sqrt{2}$. For each scale pairing the three step procedure for matching VRs is applied. In the warping step of this procedure either the query or the image is warped depending on the relative scale. If the relative scale between the query and a candidate image is less than 1 the candidate VR is warped and if it is greater than 1 the query VR is warped. After the query is matched with each of the image VRs, the location in the image which has the best correlation score is returned.

In practice, VR lists are generated both for the query and database images to save computational cost, memory, and to avoid running in to filter discretization problems. For the experiments carried out in this paper the scales of the filters used for both the query and database images are in the range $[0.8 \cdots 3.2]$ in steps of $\sqrt{2}$.

It is instructive to note that VR lists over scale are scale-space representations in the sense described by Lindeberg [Lindeberg 94] and by [Granlund 95]. By smoothing an image with Gaussians at several different scales Lindeberg generates a scale-space representation. While VR lists are scale-space representations, however, they differ from Lindeberg's approach in two fundamental ways. First VRs are generated from derivatives of Gaussians and second, an assumption is made that smoothing is accompanied by changes in size (i.e. the images are scaled versions rather than just smoothed versions of each other). This is the reason warping is required during VR matching across scales.

On the other hand, the VR list approach should not be confused with pyramidal representations [Burt 83]. While pyramidal representations are also generated by filtering and sub-sampling images, there is an important distinction. Pyramids are generated as a translational search reduction mechanism for use in coarse-to-fine matching. Pyramid matching assumes that the scale of the template and the image within which it is being localized is the same. Therefore, matching the coarsest level of the image and template first followed successively by the finer representations yields reductions in translational search. However, the relative scale between the query and the image is never known, forcing a true search across the scale parameter. As Lindeberg points out recursive application of filters and sub-sampling as is done in pyramidal schemes is not in general a scale-space representation [Lindeberg 94]. VR lists, which are not generated recursively, are proper scale-space representations and the matching occurs across scale-space.

5 Constructing Query Images

The query construction process begins with the user marking salient regions on an object. VRs generated at several scales within these regions are matched with the database in accordance with the description in Section 4. Unselected regions are not used in matching. One way to think about this is to consider a composite template, such as one shown in Figure 2. The unselected regions have been masked out. The composite template preserves inter-region spatial relationships and hence, the structure of the object is preserved. Warping the composite will warp all the components appropriately, preserving relative spatial relationships. That is, both the regions as well as distances between regions are scaled appropriately. Further, there are no constraints imposed on the selection of regions and the regions need not overlap.

Careful design of a query is important. It is interesting to note that marking the entire object does not work very well. Marking extremely small regions has also not worked with this database. There are too many coincidental structures that can lead to poor retrieval.

Many of these problems are, however, simplified by having the user interact extensively with the system. Letting the user design queries eliminates the need for detecting the saliency of features on an object. Instead, saliency is specified by the user. In addition, based on the feedback provided by the results of a query, the user can quickly adapt and modify the query to improve performance.

6 Experiments

The choice of images used in the experiments was based on a number of considerations. It is expected that when very dissimilar images are used the system should have little difficulty in ranking the images. For example, if a car query is used with a database containing cars and apes, then it is expected that cars would be ranked ahead of apes. This is borne out by the limited number of experiments done. Much poorer discrimination is expected if the images are much more 'similar'. For example, man-made vehicles like cars, diesel and steam locomotives should be harder to discriminate. It was therefore decided to primarily use images of cars, diesel and steam locomotives as part of the database.

The database used in this paper has digitized images of cars, steam locomotives, diesel locomotives, apes and a small number of other miscellaneous objects such as houses. Over 300 images were obtained from the internet to construct this database. About 215 of these are of cars, diesel locomotives and steam locomotives. There are about 80 apes and about 12 houses in the database. These photographs, were taken with several different cameras of unknown parameters, and, under varying but uncontrolled lighting and viewing geometry. The objects of interest are embedded in natural scenes such as car shows, railroad stations, country-sides and so on.

Prior to describing the experiments, it is important to clarify what a correct retrieval means. A retrieval system is expected to answer questions such as 'find all cars similar in view and shape to this car' or 'find all steams similar in appearance to this steam engine'. To that end one needs to evaluate if a query can be designed such that it captures the appearance of a generic steam engine or perhaps that of a generic car. Also, one needs to evaluate the performance of VR matching under a specified query. In the examples presented here the following method of evaluation is applied. First, the objective of the query is stated and then retrieval instances are gauged against the stated objective. In general, objectives of the form 'extract images similar in appearance to the query' will be posed to the retrieval algorithm.

Questions of this form are interesting to answer in the context of the types of images present in the database. Diesel locomotives, steam engines and cars are all man made objects and can be expected to be similar. From several experiments performed

	No. Retrieved Images				
Query	1-10	11-20	21-30	31-40	41-50
Car	8	6	1	0	1
Steam	7	2	1	0	2
Diesel	7	5	5	6	4

Table 1: Correct retrieval instances for the Car, Steam and Diesel queries in intervals of ten. The number of "similar" images in the database as determined by a human are 16 for the Car query, 12 for the Steam query and 30 for the Diesel query.

with this database it is observed that queries can be constructed, such that vector-matching does a good job of ordering the dissimilarities in appearance of these objects. For example, a car query that intuitively captures distinguishing features on a car ranks cars of similar appearance higher than other objects. Additionally, good discrimination is easily obtained between fairly dissimilar objects such as apes and engines for example. Several different queries were constructed to retrieve objects of a particular type. It is observed that under reasonable queries at least 60% of m objects underlying the query are retrieved in the top m ranks. Best results indicate retrieval results of up to 85%.

Several experiments were carried out with the database [Ravela 96]. The results of the experiments carried out with a car query, a diesel query and a steam query are presented in table 6. The number of retrieved images in intervals of ten is charted in Table 6. The table shows, for example, that there are 16 car images "similar" in view to the car in the query and 14 of these are ranked in the top 20. For the steam query there are 12 "similar" images (as determined by a person), 9 of which are ranked in the top 20. Finally, for the diesel query there are 30 "similar" images, 12 of which are found in the top 20 retrievals.

Due to space limitations only the results of the *Car retrieval* are displayed (Figure 4) and analyzed in detail (for the others see [Ravela 96]).

The car image used for retrieval is shown in the top left picture of Figure 4. The objective is to 'obtain all similar cars to this picture'. Towards this end a query was marked by the user, highlighting the wheels, side view-mirror and mid section. The results to be read in text book fashion in Figure 4 are the ranks of the retrieved images. The white spots indicate the location of the centroid of the composite template at best match. In the database, there are exactly 16 cars within a close variation in view to the original picture. Fourteen of these cars were retrieved in the top 16, resulting in a 87.5% retrieval. All 16 car pictures were picked up in the top 50. The results also show variability in the shape of the retrieved instances. The mismatches observed in pictures labeled '15.tif' and '19.tif' occur in VR matching when the relative scale between the query VR and the images is $\frac{1}{4}$.

Wrong instances of retrieval are of two types. The first is where the VR matching performs well but the objective of the query is not satisfied. In this case the query will have to be redesigned. The second reason for incorrect retrieval is mismatches due to the search over scale space. Most of the VR mismatches result from matching at the extreme relative scales.

Overall the queries designed were also able to distinguish steam engines and diesel engines from cars precisely because the regions selected are most similarly found in similar classes of objects. As was pointed out in Section 5 query selection must faithfully represent the intended retrieval, the burden of which is on the user. The retrieval system presented here performs well under it's stated purpose: that is to extract objects of similar shape and view to that of a query.

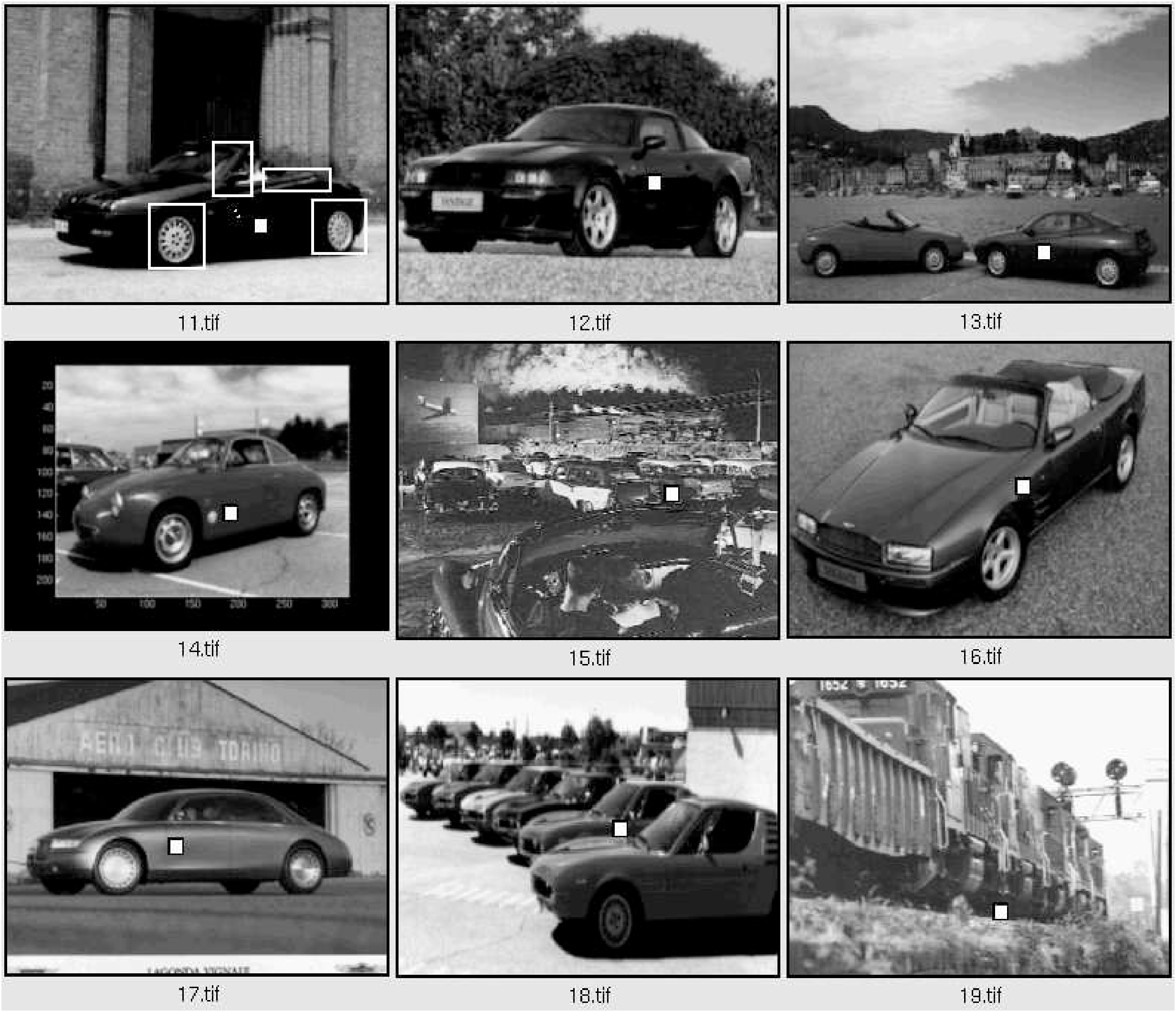
7 Conclusions and Limitations

A method to retrieve images based on shape properties of images was presented. The vector-correlation algorithm is robust to lighting changes and small deformations. Vector-Correlation was extended to incorporate gross scale changes. Thus, the resulting representation of images is a proper scale-space representation and matching is performed over this space.

Using this technique objects of similar appearance were retrieved. There are several factors that affect retrieval results, including query selection, and the range of scale-space search. The results indicate that this method has sufficient accuracy for image retrieval applications.

One of the limitations of our current approach is the inability to handle large deformations. The filter theorems described in this paper hold under affine deformations and a current step is to incorporate it in to the vector-correlation routine.

While these results execute in a reasonable time they are still far from the high speed performance desired of image retrieval systems. Work is on-going towards building indices of images based on local shape properties and using the indices to reduce the amount of translational search.



11.tif 12.tif 13.tif

14.tif 15.tif 16.tif

17.tif 18.tif 19.tif

Figure 4: Retrieval results for Car.

Acknowledgments

The authors thank Prof. Bruce Croft and the Center for Intelligent Information Retrieval (CIIR) for continued support of this work. We also thank Jonathan Lim and Robert Heller for systems support. The pictures of trains were obtained from John Hurst's home page at http://www.cs.monash.edu.au/image_lib/trains/. The pictures of cars were obtained from the Team.Net home page at ftp://ftp.team.net/ktud/pictures/.

References

[Bergen 92] J. R. Bergen, P. Ananadan, K. J. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation", *Proc. Second European Conference on Computer Vision*, pp. 237-252, 1992.

[Burt 83] P. J. Burt, and E. H. Adelson, "The Laplacian pyramid as a compact image code", *IEEE Transactions on Communications*, 9:(4), pp. 532-540, 1983.

[Crowley 87] J. L. Crowley, and A. C. Anderson, "Multiple Resolution Representation and Probabilistic Matching of 2-D Gray-scale Shape", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):113-121, 1987.

[Damasio 94] Antonio R. Damasio, "Descartes' Error",*G. P. Putnam's Sons*, New York, 1994

[Freeman 91] William T. Freeman, and E. H. Adelson, "The Design and Use of Steerable Filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891-906, September 1991.

[Flickner 95] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovix, David Steele, and Peter Yanker,"Query By Image and Video Content: The QBIC System", *IEEE Computer Magazine*, pp.23-30, September 1995.

[Granlund 95] Gösta H. Granlund, and Hans Knutsson, "Signal Processing in Computer Vision",*Kluwer Academic Publishers*, ISBN 0-7923-9530-1, Dordrecht, The Netherlands, 1995.

[Gudivada 95] Venkat N. Gudivada, and Vijay V. Raghavan, "Content-Based Image Retrieval Systems", *IEEE Computer Magazine*, pp.18-21, September 1995.

[Hancock 92] P. J. B. Hancock, R. J. Bradley and L. S. Smith, "The Principal Components of Natural Images", *Network*, 3:61-70, 1992.

[Kass 88] M. Kass, "Linear Image Features in Stereopsis", *International Journal of Computer Vision*, Vol. 1, pp. 357-368, 1988.

[Kirby 90] M. Kirby, and L. Sirovich, "Application of the Karuhnen-Loeve Procedure for the Characterization of Human Faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103-108, January 1990

[Koenderink 87] J. J. Koenderink, and A. J. van Doorn, "Representation of Local Geometry in the Visual System", *Biological Cybernetics*, vol. 55, pp. 367-375, 1987.

[Kosslyn 92] Stephen M. Kosslyn, and Oliver Konig, "Wet Mind: The new cognitive neuroscience", *The Free Press*, 1992.

[Lindeberg 94] Tony Lindeberg, "Scale-Space Theory in Computer Vision", *Kluwer Academic Publishers*, ISBN 0-7923-9418-6 , Dordrecht, The Netherlands, 1994.

[Manmatha 94] R. Manmatha, "Measuring Affine Transformations Using Gaussian Filters", *Proc. European Conference on Computer Vision*, vol II, pp. 159-164, 1994.

[Manmatha 93] R. Manmatha and J. Oliensis, "Measuring Affine Transform - I, Scale and Rotation", *Proc. DARPA IUW*, pp. 449-458, Washington D.C., 1993.

[Mehrotra 95] Rajiv Mehrotra and James E. Gary, "Similar-Shape Retrieval In Shape Data Management",IEEE Computer Magazine, pp. 57-62, ,September 1995.

[Pentland 94] A. Pentland, R. W. Picard, and S. Sclaroff,"Photobook: Tools for Content-Based Manipulation of Databases", Proc. Storage and Retrieval for Image and Video Databases II, Vol.2, 185, SPIE, pp. 34-47, Bellingham, Wash. 1994.

[Picard 94] Monika M. Gorkani, and Rosalind W. Picard, "Texture Orientation for Sorting Photos "at a Glance"", *TR-292, M.I.T., Media Labortory, Perceptual Computing Section*, 1994.

[Rao 95] R. Rao, and D. Ballard, "Object Indexing Using an Iconic Sparse Distributed Memory", *Proc. International Conference on Computer Vision*, pp. 24-31, 1995.

[Ravela 96] S. Ravela, R. Manmatha and E. M. Riseman "Retrieval from Image Databases Using Scale-Space Matching", *Technical Report UM-CS-95-104*, Dept. of Computer Science, Amherst, MA 01003. To appear in *Proc. European Conference on Computer Vision*, 1996.

[Shepard 82] R. N. Shepard, and L. A. Cooper, "Mental Images and their transformations", MIT Press, Cambridge, MA, 1982.

[Turk 91] M. Turk, and A. Pentland, "Eigen Faces for Recognition", *Journal of Cognitive Neuroscience*, vol 3., pp.-71-86, 1991.

[De Valois 88] R. L. De Valois, and K. K. De Valois, "Spatial Vision", *Oxford University Press*, New York, 1988.

[Werkhoven 90a] P. Werkhoven and J. J. Koenderink, "Extraction of Motion Parallax Structure in the Visual System 1", *Biological Cybernetics*, 1990.

[Werkhoven 90b] P. Werkhoven and J. J. Koenderink, "Extraction of Motion Parallax Structure in the Visual System 2", *Biological Cybernetics*, 1990.