Image Retrieval through Qualitative Representations over Semantic Features

Zia Ul-Qayyum, A.G. Cohn
zia@comp.leeds.ac.uk, A.G.Cohn@leeds.ac.uk

Abstract

We propose a qualitative knowledge-driven semantic modelling approach for image retrieval based on qualitative relations over local semantic concepts of images. The relative similarity of two images is proportional to their qualitative similarity. The similarity measure is calculated for each query by exploiting the notion of conceptual neighbourhood – a measure of closeness between qualitative relations. The approach is motivated by the need to perform semantic querying using qualitative relations and bridge the semantic gap between a human user and that of CBIR systems. Three qualitative representations (and several variants) and a corpus of 700 natural scene images have been used to evaluate the effectiveness of image retrieval using this approach.

1. Introduction

Advances in digital technologies along with the growth of the Web have resulted in universal access to very large archives of digital data. This has lead to an increasing requirement for systems with more flexible and robust techniques to handle dynamic and complex visual content at a higher semantic level. Content based image classification and retrieval systems have thus gained more importance and have become an active research area [1]. In all such systems, image interpretation and understanding plays a vital role. Most of the research in this area is primarily based on use of low level image features like colour, texture, shape etc [9, 18]. Although low level image processing algorithms and methodologies are quite mature, such systems are hard to be used effectively by a novice due to the semantic gap between user perception and understanding, and system requirements. Bridging this gap between low level synthetic features and high level semantic meanings is, therefore, generally regarded as an open problem [1]. Humans tend to describe scenes using natural language semantic keywords/concepts like sky, water etc and specify queries like “an image with water next to fields and sky above....” or “… has a small lake with high peaks of mountains behind and fields on left...”. This suggests that use of underlying semantic knowledge in a qualitative representation language may provide a way to model the human context and is a natural way to bridge semantic gap for better image understanding, categorization and retrieval capabilities.

This paper thus proposes a qualitative knowledge-driven semantic modelling approach for IR. Qualitative representation of the local semantic contents of an image allows for representation and reasoning of content structures at a higher abstraction level than low level features. In earlier work [13], we showed how category descriptions for a set of images could be learned using qualitative spatial representations (QSR) over a set of local semantic concepts (LSC) such as sky, grass. There were six global categories (e.g. coasts, forest etc) [19] and we used three kinds of QSR techniques to demonstrate that supervised learning using QSR of semantic image concepts can rival a non qualitative approach for image categorization [19,13], and moreover result in a more intuitive and more human understandable image description.

Our hypothesis in this paper is that the qualitative representations which were able to effectively support categorization may also provide an effective and natural way to support content-oriented querying. A query can either be directly described in the qualitative representation, or in the evaluation of our approach described below, a query can be given as a sample image (i.e. query by example: QBE) – the system then forms a qualitative description of it by a conjunction of qualitative relations between the semantic concepts. In both cases the system then compares the query qualitative description with qualitative descriptions of images in the database of images, and uses a *qualitative similarity measure* to retrieve qualitatively similar images, and show how retrieved images can be ordered accordingly. We do not assume that images have already been assigned categories/classes. The qualitative similarity measure is based on the notion of a *conceptual neighbourhood* (CN) [10] – see §4.

In experiments, using this technique on the different QSRs, we observed that the various representations had different levels of performance for different categories of images; this lead us to investigate the use of voting schemes in order to combine the different QSR to enhance the performance of the retrieval system overall.

A quantitative metric based evaluation of approaches based on qualitative representations has always been difficult. In order to evaluate the performance of this approach to IR, we take advantage of manually assigned categories for the image DB in our experiments. Although we are not performing image categorization, and the retrieval algorithm does not use the category information, success of retrieval is evaluated by counting the number of highly ranked images in the same category as the query.

The experimental data set is a collection of 700 natural scenes images, provided and hand labelled with categories by Vogel et al, who developed a semantic modelling framework for image categorisation and retrieval [19]. Our approach builds on her work, an overview of which is presented in §3.

The rest of the paper is structured as follows. Related work is briefly discussed in §2. §3 describes our approach to image description using QSR. A qualitative similarity based IR approach is presented in §4. §5 presents the results and evaluation of the approach, while §6 presents our conclusions and suggestions for future work.

2. Related Work

In the IR literature, image description and better understanding of underlying semantic content play important roles as the nature and structure of the query depends on the underlying image description. In this section, we first we describe the most relevant work from allied disciplines of content-based IR and then briefly survey the field of QSR.

CBIR systems have become an active research area in computer vision. [7,9,15,18] review the state of the art in segmentation, indexing and retrieval techniques in a number of CBIR systems. Despite increased work in aspects related to high level semantics of image features, the gap between low level image features and high level semantic expressions is a bottleneck in accessing multimedia data from databases. These surveys reveal that almost all existing approaches rely on using low level image features for image description, categorization and retrieval. Since image understanding is key to all content-based image categorisation and retrieval systems, so a human understandable image description may yield more robust systems since humans normally tend to use semantic and qualitative terms to describe a situation/image. Therefore, a retrieval system based on qualitative description of underlying semantic knowledge may help a non-expert user query such systems more effectively. Research has already been done focusing on the use of labelling the image regions with semantic concepts and carrying out key-word based IR. One such probabilistic approach [4] is to assign small image areas labels such as “man-made” and “natural”, and global labels such as “inside”, “outside” to whole images using class likelihoods from colour-texture features of images for semantic IR. Local regions of images have been annotated with 11 and 10 semantic

categories respectively [17,20]; in [17] a global label is not assigned to images, so retrieval is based on local semantic concepts only. An IR approach based on semantically labelled image regions is demonstrated in [1]. These image regions have been hierarchically classified based on their semantics using low level image features. Retrieval is based on these semantic keywords attached to particular images.

In an approach [21] for semantic retrieval based on content and context of image regions and which supports both keyword and QBE queries, images are segmented using a semantic codebook based on colour and texture classification. The content and context describe a region's low level features and their relationships respectively. It uses only dominant semantic categories of an image and the most typical images in that category are selected manually from an image database which can best model the codebook representing colour and texture classification for that particular semantic category. Another query by semantic example (QBSE) approach is based on posterior concept probabilities of each concept in an image [14]. QBSE is accomplished by comparing the probability simplexes of the query image and all database images to find the closest neighbours. The perceptual segmentation approach in [8] has not been applied in their work for image categorization and retrieval, but the relative effectiveness of their approach to image segmentation and labelling can be used to perform keyword based IR. The VISENGINE system [16] relies on segmenting image regions by clustering visual features like colour, texture, shape etc and differentiating them into foreground and background regions. The approach is largely user-centred, and therefore results may vary depending on human perception and context. Since only large regions are identified during segmentation, small image areas do not contribute towards the retrieval process which may inhibit a true semantic similarity in the retrieved images. Progress can also be made algorithmically, e.g. it has been shown that classification and retrieval accuracy can be boosted by combining different approaches [11]. The use of ontologies and metadata representation languages is another recent trend for annotating and retrieving images [12]. A prerequisite for this approach is the construction of generic and possibly domain specific ontologies from which the detailed annotations are constructed.

One crucial research question for QBE systems is how to measure the level of similarity, and assess the accuracy of such a technique. Defining a notion of similarity is difficult since context may play a pivotal role. Moreover, when using a qualitative representation, where feature descriptions do not take quantitative values, the very notion of a metric becomes problematic; approaches to qualitative similarity are discussed in [3]. In computer vision and image processing, metric approaches have generally been used to compute scene similarity, e.g. a measure based on normalised distance for a semantic ordering of natural scenes in categories such as forest and mountains, mountains and rivers/lakes [19].

The field of QSR has become increasingly more active within AI as it arguably provides cognitively or intuitively relevant representations for spatial information – typical spatial expressions in natural language are qualitative rather than quantitative. Moreover, qualitative representations abstract away from noise and uncertainty in perceptual data. It has increasingly been used in different application domains like GIS, NLP, robotics, computer vision etc, see [6] for a review. There are many QSR, covering aspects such as topology, distance, orientation, and shape. Rather than attempt an exhaustive analysis of the utility of all these calculi, we concentrate on a small set of QSR here; we do not claim these are necessarily the best calculi for image description, or even for the particular kinds of images in the database we use here, but leave that for further work. Our aim is simply to illustrate the use of qualitative calculi for IR and to demonstrate their potential applicability and suitability for CBIR.

In the qualitative framework, in which images are described using a small finite set of relations or qualitative values, similarity can be computed by using the distance in the CN graph. The notion of a CN was first put forward [10] in the context of a set of 13 pairwise and disjoint relations between temporal intervals and was defined as "two spatial

or temporal relations are conceptual neighbours if one can be transformed into the other by a single [continuous] transformation/transition". Given two such qualitative image descriptions, their similarity is proportional to the number of such transformations required to turn one into the other [5].

3. Qualitative Image Description

Our approach builds on Vogel et al's work [19] in which images from a 700 image corpus were divided into a grid of 10x10 regions (instead of using segmentation techniques) and nine local[1] and discriminating semantic concepts were identified: sky, water, grass, foliage, flowers, field, mountain, snow, trunks and sand. Vogel et al manually annotated 99.5% of the images with these concepts, and used this as input to supervised learning techniques to annotate image patches automatically. A label "rest" is used for unidentified patches or occurrences of other semantic categories. Images were represented by frequency histograms of local semantic concepts and based on a semantic typicality measure; images were categorized into one of the six semantically meaningful categories sky_clouds (34), coasts (143), landscapes_with_mountains (lwm) (178), fields (128), forests (103), waterscapes (114). (The numbers in brackets show total number of images for the respective category.) This approach is partially spatial through its division of the image into horizontal bands (e.g. top (T), middle (M) and bottom (B)) but is mainly based on the metric value of the percentages of discriminant semantic concepts.

We use the hand labelled data set in the experiments reported here in order to evaluate using the "gold standard" rather than be affected by the particular model learned for annotation. The images are described using the following QSRs:

1) The relative size (measured in grid squares) for all possible pairwise combinations of the semantic labels. Each may be regarded as an attribute of the image with possible values of 'Greater than' (>), 'Less than' (<) and 'Approximately Equal to' (≈) – we allow a ±10% tolerance for ≈.

2) Allen relations [2] (measured on vertical axis between the intervals representing the maximum vertical extent of each concept occurrence). The 13 relations are: 'before' (<), 'meets' (m), 'overlaps' (o), 'during' (d), 'starts' (s) and their inverses 'after' (>), 'met-by' (mi), 'overlapped-by' (oi), 'contains' (di), 'started-by' (si), 'finished-by' (fi) respectively, and 'equal' (=). A 14th relation 'no' is used if neither attribute is present.

3) Chord patterns [15] of semantic concepts applied to each grid row. Each semantic feature is a 'tone' and each row forms a 'chord' of tones. The 10x10 grid generates 10 chords, one for each row, such as "foliage sky" or "grass sky sand water"[2] etc.

4) A binary 'Touching' relationship (additional to the above 3 representations already used in our work [13]), which records whether one patch type is spatially in contact with another in the image. Note that, although apparently similar, the Allen 'meets' relation is not equivalent since the 2 patches may be at different sides of the picture.

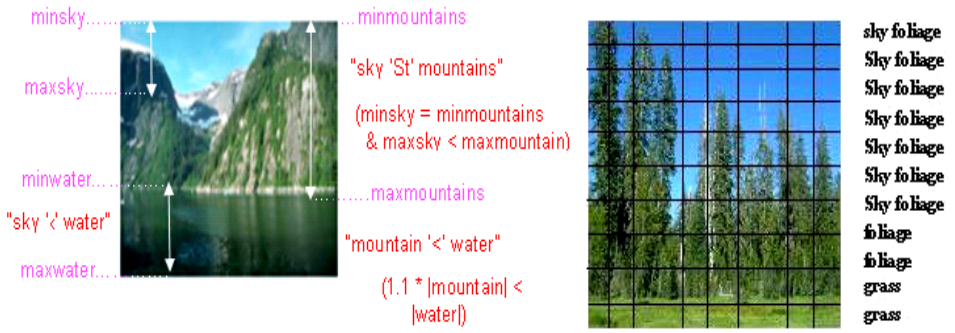
For comparison purposes, we also ran experiments with a purely quantitative metric based retrieval scheme based on the respective percentages of each of the semantic concepts in each image in the style of [19]. This representation is labelled as "Percentages" in Table 1. Similarity is computed using the sum of absolute differences in percentage values for each attribute in a pair of images.

Fig. 1(b) illustrates the chord representation while Fig. 1(a) the relative size and Allen relationships. Several variants of the above QSRs were also investigated; we report

[1] There are 9 semantic concepts in [19], while the data set provided and which has been used in our experiments contains 2 extra ones (mountain and snow) – however these occur infrequently and the basis for comparison will be thus essentially unaffected.

[2] This representation can be regarded as an abstraction of the relation used by [19] – whereas they record the percentage of each attribute in each horizontal band, in the chord representation it is only the presence or absence which is recorded.

on just one here where the relative size representation is recorded separately within 3 image areas: Top (T: top 3 rows), Middle (M: rows 4-7), Bottom (B: rows 8-10).



(a) (b)

Fig. 1. QSR using (a) relative size and Allen's calculus (b) chord representation

4. IR Based on Qualitative Similarity

We envisage a CBIR system in which a query is specified either by giving an example image or by a symbolic query expressed in terms of the qualitative relations defined above, e.g. "retrieve images with rocks touching water and more water than foliage". In the former case, we can compute a qualitative description of the image using one more of our qualitative schemes, but in this case it is more likely that no image will exactly match – this could also happen in the latter case. It would clearly be convenient to be able to retrieve images which nearly match the query (which ever way it is specified). The problem is to define what "nearly matches" means, since in a qualitative representation we do not have raw numbers available. In the remainder of this section we define notions of qualitative similarity for each the qualitative representations.

The CN of Allen relations is presented in Fig 2(a) below. The links connect neighbouring relations – ones which are most similar – as one traverses more links from a particular relation, the relations become progressively less similar. Thus if in image 1 sky **<** grass, and also in image 2, then they are identical (in this comparison); if in image 3 sky **m** grass, then image 3 is similar to image 1, whilst if image 4 has sky **o** grass, then image 4 is also similar to image 1 but not as similar as image 3, and so forth. Since there are many attributes in each description of an image (e.g. 66 in Allen representation), we have to find a way to combine the similarities of each pairwise comparison. The CN for the Allen relations is already a partial order, and it is clear that the cross product across all the attributes is even more so. To achieve a total ordering we assign a weight of 1 to each arc in the CN, and sum the number of arcs traversed across all the attributes in order to transform one description into another (using the shortest route). Clearly we could assign non uniform weights to the different arcs but in the absence of any particular reason to do this, a uniform weighting appears to be the obvious choice. The situation where one of the relations from a particular pair of images for a pair of attributes is "no" whilst the other is not, deserves some discussion – what should be the weight in this case (since "no" does not appear in the CN)? One possibility is to choose a weight of 7 (one more than the maximum weight otherwise in the Allen CN), though other choices could clearly also be used, and indeed we also experimented with the choice of zero[3] and values greater than and less than 7. In an implementation for an end user, this could be a parameter (perhaps a slider in the interface).

[3] This was particularly motivated by classes such as "lwm" where the set of concepts present can vary considerably, and penalizing image with a different set of concepts to the query image had a great effect on the results. A penalty weight of 0 implies that the similarity of images is determined only by the relationship between common semantic concepts in the query and database images, and missing concepts do not contribute towards total penalty weight.

The CN for the relative size representation is much simpler with just three nodes, one for each of the three relations, with ≈ neighbouring each of $<$ and $>$ and the maximum weight is 2. For missing patch types we do not need a 'no' relation in this representation since their size is zero and the existing three relationships are still applicable.

For the case of the chord representation, we can think of the CN as being equivalent to a complete lattice generated by the power set of the set of patch types; effectively this means that the similarity is directly proportional to the number of insertions and deletions required to transform one chord into another.

For the representation of spatial touching, there are just 2 nodes in the CND (touching and not-touching) and a single link connecting them. We experimented with this representation, however eventually used a similarity measure which also takes account of the degree of touching. Each patch in the rectangular grid can touch up to 8 other patches. For a pair of given patch types p1 and p2, we compute how many patches of type p1 touch a patch of type p2, and vice-versa for p2 and p1; the maximum of these 2 values is then recorded as one of the attributes in this representation of an image. To compute the degree of similarity between two images using this representation we simply take the sum of the absolute differences in each of the corresponding attribute values for each image. This representation thus combines a very qualitative representation, touching, which is a purely topological relationship, with a metric measurement of its applicability to a particular image. Thus, for example, for an image with extended sky-grass spatial connection will be more similar than ones with small amount of spatial connection between the two concepts.

Thus given a representation "R" with attributes $\mathrm{A}_1^{\mathrm{R}} \ldots\ldots \mathrm{A}_{|\mathrm{R}|}^{\mathrm{R}}$, and a function $f^R(u,v)$ which gives the similarity between two attribute values u and v then the overall similarity S^R (x,y) between two images *x* and *y* in representation 'R' is given by:

$$S^R(x,y) = \sum_{i=1}^{i=|R|} f^R(A_i^R(x), A_i^R(y)) \qquad (1)$$

We then can compute rank of an image *y* in the database for query image *x* as:

$$Rank^R(x,y) = |\{z : S^R(x,z) < S^R(x,y)\}| \qquad (2)$$

5. Results and Evaluation

We have conducted experiments with each of the representations above individually and also in various combinations. To illustrate the results obtained, we first present (fig. 2(b)) a sample query image and the top 5 results according to the qualitative similarity measures described in §4 for Allen representation. This does not give any quantitative evaluation of the quality of the retrieval and we next turn to this question. To provide a more thorough quantitative analysis of the performance of the various representations, we used the following experimental setup. Each of the 700 images in the database was used as a query image in turn, and a similarity ordering computed for all the other 699 images. However this does not tell us whether images high in the ordering really are intuitively similar to the query image. As a proxy for an extensive user evaluation of each of these rank orderings, we use the hand assigned category labels used for previous work on this dataset for supervised learning of category descriptions [19,13].

Given a query image in category *c*, we can evaluate the number and hence the percentage of images in the same category in the top *k* images in the rank ordering. For cases where the number of images of a particular category in the DB is less than *k* clearly 100% scores cannot be achieved.

The number *k* may be user defined, or be determined by conditions such as how many images of a certain size fit on a user's screen, or could be determined by analysis of the actual similarity values. Table 1 shows, for each class, the number of retrieved images of that class in the top ranked 20 and the top *k* images (where *k* is the number of images in the respective class, e.g. *k*=34 for sky_clouds), each row giving the values for a different representation. The last two rows in Table 1 shows the statistics when using the percentage of each semantic attribute as the representation for comparison with the quantitative techniques of [19]. The results reveal the following interesting conclusions:

- The recall rate clearly validates the measures of similarity used, since as the number of images retrieved increases, the accuracy of retrieved images goes down (measured by successive retrieved images of the same category).
- the recall percentages are well above the baseline statistical likelihood of each category of images in the population.
- The chord representation performs relatively well. Arguably this is because it closely resembles the human cognition of similarity because a human may describe or compare an image in terms such as "having sky in the top, foliage and water in the middle, water and sand at the bottom of image" – remembering that the semantic categories were assigned by a human (though without being aware of the possibility of subsequently using the chord representation (or indeed any other).

(a) (b)

Fig. 2. (a) CN for Interval Calculus [42] **(b)** Query & top 5 retrievals using Allen's Rep

- The representation 'relative size' performs surprisingly well, given the low information content. Moreover, the relative size on TMB regions of image representation performs at least as well if not even better in overall compared to the purely metric representation (Percentages and Percentages on TMB).
- The touch based representation does not perform particularly well – arguably it does not encode sufficient information to be able to adequately distinguish cognitive similarity in the image dataset.

Table 1 only considers individual representations. Since the performance of representations varies across categories (and bearing in mind that we assume we do not know the category of an image – we are using this information here purely for evaluation purposes), we also experimented with similarity measures based on combinations of four different qualitative representations[4] – Allen, relative size, chord and touching.

There have been a number of approaches in image categorization research involving bagging/boosting while in IR, multiple query processing or use of low level and semantic labels has been used to improve the retrieval accuracy. We investigated voting approaches based on combining the respective penalty weights of images in individual representations, and on combining the ranks of retrieved images in each selected QSR.

[4] Of course each representation might itself be viewed as a hybrid representation with the 66 attributes (or whatever number of attributes used in the particular representation) combining together to assign an overall similarity to an image pair.

In order to count the accumulative effect of penalty weights in all of the 4 selected representations and also the overall ranking of an image in the list of database images, several other kinds of weighted voting schemes ($V_1 - V_4$) were investigated (Table 2):

V_1- Compute:

$$S^{V_1}(x,y) = \sum_{r=1}^{r=4} S^r(x,y) \quad (3)$$

for each image in the DB for a query *x* and then sort in ascending order.

V_2- Compute:

$$S^{V_2}(x,y) = \underset{r=1}{\overset{r=4}{Min}}\ S^r(x,y) \quad (4)$$

for each image in the DB for a query *x* and then sort in ascending order: (variant of V1).

Although the weights within in each representation may be regarded as comparable, it is arguable as to whether this also holds with respect to the weights in other representations. We thus investigated schemes based solely on the rank within each of the four representations.

V_3- Compute:

$$S^{V_3}(x,y) = \sum_{r=1}^{r=4} rank^r(x,y) \quad (5)$$

for each image in the DB for a query *x* and then sort in ascending order.

V_4- Compute:

$$S^{V_4}(x,y) = \underset{r=1}{\overset{r=4}{Max}}(rank^r(x,y)) + \underset{r=1}{\overset{r=4}{Max2}}(rank^r(x,y)) \quad (6)$$

where "Max" and "Max2" compute the maximum and 2nd highest values respectively.

The results suggest the following conclusions:

- The purely qualitative approaches perform comparably or even slightly better in some cases to the quantitative ones. The former have added advantage that they also allow retrieval based on simple linguistic descriptions using qualitative descriptions over the semantic attributes.
- The voting schemes based on accumulative weighted votes and weighted rank votes ($V_1 - V_4$) perform better than the approaches using a single representation only.
- The overall accuracy of the retrieval process compared with the actual class labels is somewhat problematic due to the fact that many images may be categorized as either "lwm" or "coast" – i.e. most of the images in the DB have some aspects of "lwm" or "coast", and arguably it is a matter of degree or personal preference when an lwm with sky above becomes a "sky_clouds". Similarly, there is lot of potential confusion in images categorised in classes like "fields" and "sky_clouds". This fact was also established in [13, 19] while learning the class descriptions.
- The voting schemeV_1 performs much better in the top 20 and the top *k* experiments as it is based on accumulative row weights of an image corresponding to 4 representations chosen. Its performance is comparable to the quantitative approach. Furthermore, both of the basic voting schemes, V_1 and V_3, are better than the individual representations in terms of accuracy of IR using the "ground truth" of the hand assigned labels.
- It can be seen that coasts and waterscapes do relatively badly compared to the other categories, and this is also true about sky_clouds and fields categories in some of the representations, which is not altogether surprising from a semantic/intuitive viewpoint. If these two categories are combined into a single category then the rate of accuracy improves significantly. This fact has also been observed in the confusion matrices of different learning schemes in [13].

Categories / QSRs	Coasts Out of		Field Out of		Forest Out of		LWM Out of		Sky_Clouds Out of		wscapes Out of		Overall	
	20	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**
Allen only	56	33	38	26	66	41	84	48	49	35	46	26	59	36
Touch	57	33	40	27	73	51	85	52	51	40	42	22	61	38
Chord	56	41	66	34	91	68	82	59	91	**89**	47	36	70	50
Size only	63	**46**	57	34	86	66	**88**	61	60	44	**51**	**37**	70	49
Size on TMB	**67**	45	**68**	**38**	**92**	**75**	88	**65**	**93**	82	47	34	**74**	**53**
Percentages-%s	62	47	**70**	36	92	69	**84**	61	93	91	47	**36**	**73**	52
%s on TMB	**64**	**48**	69	**36**	**93**	**72**	84	**62**	**94**	**92**	**48**	35	73	**53**

Table 1. Recall percentages on per category and overall basis in top 20 & number of images in each category (k) for all representations used.[5]

Categories / QSRs	Coasts Out of		Field Out of		Forest Out of		LWM Out of		Sky_Clouds Out of		Wscapes Out of		Overall	
	20	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**	**20**	**k**
V_1	**67**	**45**	**69**	**35**	**95**	**78**	92	**69**	**88**	**78**	**51**	**35**	**76**	**54**
V_2	55	33	37	26	65	42	83	48	50	35	47	27	59	36
V_3	66	44	60	33	93	72	**93**	65	79	63	50	33	**74**	**51**
V_4	66	42	60	34	87	64	90	60	69	48	**51**	33	72	47

Table 2. Recall percentages on per category and overall basis in top 20 & number of images in each category (k) for weighted voting schemes.

6. Conclusions And Further Work

We have presented an approach to CBIR based on semantic knowledge and QSR. The approach does not rely either on segmentation techniques applied directly or on low level image features for an image description. We have presented similarity measures of the qualitative spaces based on the conceptual neighbourhoods that typically accompany qualitative calculi and experimental results for IR using a variety of qualitative description languages and several combinations of these. We are not necessarily arguing that these are the best languages either for this particular data set or in general. It is the overall approach we present which we believe is the most important result of this research, which shows that qualitative representations can rival metric ones, whilst providing more intuitive descriptions. We have also presented a variety of voting schemes for combining representations and evaluated their success on the image dataset. The evaluation was based on a hand labelled categorization which although it has some disadvantages, does provide a cognitive basis for evaluating the retrieval results. It may be noted that in all cases, the recall percentages are well above the baseline statistical likelihood of each category of images in the population.

A variety of further work suggests itself including the evaluation on other data sets, using actual user analysis to evaluate the results (cf the psychophysical experiments in [19]), experimentation with other qualitative calculi, and combining qualitative and quantitative representations. We already have a prototype user interface to an IR system based on the ideas presented here; this could be further improved to provide a flexible interface based on query by image or by qualitative description, or a combination of the two, with the user free to select the kinds of descriptions, similarity measures and voting

[5] Bold figures in Table 1 and Table 2 indicate best ones in qualitative and quantitative representations, while k=143,128,103,178,34 and 114 for above mentioned six classes – in order as these appear in table.

schemes most appropriate to their needs. The analysis here provides the basis for reasonable default choices.

Acknowledgements: We thank Julia Vogel for providing the labelled data set and helpful discussions and acknowledge financial support provided by National University of Sciences & Technology, Rawalpindi – Pakistan, and EPSRC grant EP/DO61334/1 to Zia Ul Qayyum and A.G. Cohn, respectively.

References

[1] Aghrabi, Z. , Makinouchi, A. "Semantic Approach to Image Database Classification & Retrieval". NII journal,. 7 (.9), 2003.
[2] Allen, J. F. "Maintaining knowledge about temporal intervals". C of ACM, 26(11), 1983.
[3] Bonan, Li, and Fonsesca, F. "TDD: A Comprehensive Model for Qualitative Spatial Similarity Assessment". In J. of Spatial Cognition and Computation, 6(1), pp.31-62, 2006.
[4] Bradshaw, B. "Semantic Based Image Retrieval: A Probabilistic Approach." ACM Multimedia, October 2000.
[5] Burns, H.T., Egenhofer, M.J. "Similarity of Spatial Scenes". J.-M. Kraak and M.Molenaar (Eds), 7th Int Symp on Spatial data Handling, Taylor & Francis, London, pp. 173-184, 1996.
[6] Cohn, A G and Hazarika, S M. "Qualitative Spatial Representation and Reasoning: An Overview". Fundamenta Informaticae, 46(1-2), pp. 1--29, 2001.
[7] Deb, S., & Zhang, Y. "An overview of Content-based Image Retrieval Techniques". Proc 18th Int. Conf on Advanced Information Networking & Application , 2004, pp. 59-64.
[8] Depalov, D, Pappas, T, Li, D, & Gandhi, B, "Perceptually Based Techniques for Semantic Image Classification & Retrieval". Human Vision & Electronic Imaging, XI (B.E. Rogowitz, T.N. Pappas, & S.J. Daly Eds.), Proc. SPIE,. 6057, CA, 2006.
[9] Enser, P., Sandom, C., "Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval." Springer LNCS, Vol. 27-28, pp. 279-287, 2003.
[10] Freksa, C. "Temporal Reasoning Based on Semi-intervals." Art. Int., 54(1-2), 199-227.
[11] Howe, N. "A Closer Look at Boosted Image Retrieval". Int. Conf on Image & Video Retrieval, Vol. 2728, LNCS, pp. 61-70, 2003.
[12] Hyvonen, E., Styrman, A., and Saarela, S. "Ontology-Based Image Retrieval". HIIT publications, No. 2002-03, pp. 15-27, Helsinki Institute of IT, Helsinki, Finland, 2002.
[13] Qayyum, Z.U. and Cohn, A.G. "Qualitative Approaches to Semantic Scene Modelling and Retrieval". In Proc. Of SGAI-AI'06, Research and Development in Intelligent Systems XXIII, Springer-Verlag, 2006.
[14] Rasiwasia, N., Vasconcelos, N. and Moreno, P.J. "Query by Semantic Example". H. Sundaram et al (Eds.): CIVR 2006, LNCS 4071, pp. 51-60, 2006.
[15] Sebe, N., Lew, M.S., Zhou, X., Huang, T.S., and Bakker, E.M. "The State of the Art in Image and Video Retrieval". Springer LNCS, Vol. 2728, pp. 1-8, 2003.
[16] Sun, J.Y., Sun, Z.X., Zhou, R.H., and Wang, H.F. "A Semantic-Based Image Retrieval System: VISENGINE". Proc. 1st Int. Conf on Machine Learning and Cybernetics, 2002.
[17] Town, C., and Sinclair, D. "Content-based image retrieval using semantic visual categories". Tech. Report 2000.14, AT&T Laboratories Cambridge, 2000.
[18] Veltkamp, R.C., and Tanase, M. "Content-Based Image Retrieval Systems: A Survey". Univ. Utrecht, Utrecht, The Netherlands, Tech. Rep. UU-CS-2000-34.
[19] Vogel, J., and Schiele, B. "Semantic Modelling of Natural Scenes for Content-Based Image Retrieval". Int J of CV, Springer , 10.1007/s 11263-006-8614-1, 2006.
[20] Wang, W., ,Song,Y., Zhang, A. "Semantics-Based Image Retrieval By Region Saliency". LNCS, Vol. 2383, Proc. Int. Conf on Image and Video Retrieval, pp. 29 – 37, 2002.
[21] Wang, W., Song, Y., Zhang, A. "Semantic Retrieval by Content and Context of Image Regions". Proc 15th Int. Conf on Vision Interface (VI'02), 2002.