Finding Image Semantics from a Hierarchical Image Database Based on Adaptively Combined Visual Features

Pritee Khanna[1], Shreelekha Pandey[1], and Haruo Yokota[2]

[1] PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Dumna Airport Road, Jabalpur 482-005 M.P. India
{pkhanna,shreelekha}@iiitdmj.ac.in
[2] Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-kuTokyo, 152-8552 Japan
yokota@cs.titech.ac.jp

Abstract. Correlating image semantics with its low level features is a challenging task. Although, humans are adept in distinguishing object categories, both in visual as well as in semantic space, but to accomplish this computationally is yet to be fully explored. The learning based techniques do minimize the semantic gap, but unlimited possible categorization of objects in real world is a major challenge to these techniques. This work analyzes and utilizes the strength of a semantically categorized image database to assign semantics to query images. Semantics based categorization of images would result in image hierarchy. The algorithms proposed in this work exploit visual image descriptors and similarity measures in the context of a semantically categorized image database. A novel 'Branch Selection Algorithm' is developed for a highly categorized and dense image database, which drastically reduces the search space. The search space so obtained is further reduced by applying any one of the four proposed 'Pruning Algorithms'. Pruning algorithms maintain accuracy while reducing the search space. These algorithms use an adaptive combination of multiple visual features of an image database to find semantics of query images. Branch Selection Algorithm tested on a subset of 'ImageNet' database reduces search space by 75%. The best pruning algorithm further reduces this search space by 26% while maintaining 95% accuracy.

1 Introduction

Cognitive psychology defines categories by grouping "similar objects" and super-categories by grouping "similar categories". Semantic categories form clusters in visual space, and visual similarity is correlated to semantic similarity [1]. Humans can easily correlate these similarities which gives them enormous power to distinguish a large number of objects. Content Based Image Retrieval (CBIR) systems use only visual similarity obtained in terms of low level image features to interpret images [2-4]. The lack of coincidence between the high level semantic and the low-level features of an image is known as semantic gap [5]. In an attempt to reduce semantic gap, proposed work aims to correlate visual similarity and semantics of images in a

H. Decker et al. (Eds.): DEXA 2013, Part I, LNCS 8055, pp. 103–117, 2013.

semantically categorized large image database. Semantic based categorization of an image database would result in categories and subcategories of images. Visual features of images in such a database forms a semantics based hierarchical search space. This tree is searched to assign semantics to query images. For efficient search, it is not advisable to traverse the entire tree or even an entire branch. A novel 'Branch Selection Algorithm' effectively traverses this hierarchical search space and selects a few subtrees to search. Pruning Algorithm further reduces this search space, while maintaining the accuracy. An adaptive combination of multiple visual features and similarity measures are used to design branch selection and pruning algorithms. To ensure the applicability of the proposed algorithms, their performance has been tested on a subset of ImageNet database.

The paper is organized as follows. A review of the related research is given in Section 2. Section 3 emphasizes on correlating visual and semantic similarity. Section 4 gives an insight of related databases. Proposed system is explained in Section 5. Experimental setup is given in Section 6. Section 7 summarizes results and discussion on related issues. Finally, Section 8 concludes the work.

2 Related Work

Computer Vision and Machine Learning approaches use learning based systems to reduce semantic gap [6-7]. It has already been recognized that learning accompanied by object extraction produces good results [8]. In [9], semantic templates are automatically generated during the process of relevance feedback. WordNet is used to construct a network of such semantic templates, which helps in retrieving images based on semantic. The system works on 500 images from categories like human, animal, car, etc. A statistical modeling approach for automatic linguistic indexing of pictures is introduced in [6]. Each of the 600 concepts is represented by a two-dimensional multi-resolution hidden Markov model and is trained using categorized images. A likelihood function measures the extent of the association between an image and the textual description of a concept. The model given in [10] learns visual recognition from semantic segmentation of photographs. For efficient labeling of object classes, a combination of integral image processing and feature sharing is employed. The developed classifier reports 70.5% region-based recognition accuracy on a 21-class database. The work presented in [7] focuses on using a few training images for quick learning. Generative probabilistic models of object categories are learned using a Bayesian incremental algorithm. The system quoted a feasible real-time learning rate for 101 object categories. A region-based image retrieval system with high-level semantic learning is given in [11]. The system uses a decision tree based image semantic learning algorithm but learns natural scenery image semantics only.

Besides in literature, one can find a few more learning based techniques to minimize the semantic gap [5]. Unlimited number of concepts in the real world is a major hindrance for learning based approaches. Most of the works have considered non-hierarchical image database with thousands of images. The proposed work uses a hierarchical image database to correlate visual similarity with semantic similarity. Such a correlation would be an asset to people working in image processing and computer vision. The main aim of this study is to efficiently assign semantics to images through such correlations. Instead of using any of the available learning techniques, this work exploits the inherited features of a hierarchical image database.

3 Correlating Visual Similarity with Semantics of Images

Humans have natural instinct in distinguishing object categories, both in visual as well as in semantic space, but to accomplish this computationally is yet to be fully explored. The semantic based categorization of images would give a hierarchical tree structure having images of different categories at various levels. The focus of this work is to explore whether semantic categories (e.g. dog, flower, mountains etc.) can also be visually segregated.

A semantically categorized database may contain images belonging to a domain or spread over multiple domains. It becomes difficult for a common user to search such database if the nature of its classification or the exact semantics required for the search is unknown. For example, medical terminology is an obvious choice for categorizing medical images but it is very difficult for a common user to understand semantics of these categories and hence finding proper keywords to search the database. In another scenario, a categorized database may have ten categories corresponding to dog based on their breeds, tail, coat etc. Looking at the image of a white dog with black spots, a user may not exactly know its breed name i.e. 'Dalmatian'. The user can derive such knowledge (semantics) by our approach.

Proposed approach utilizes visual features of a categorized image database of any depth and height to determine semantics of images. Huge size of the search space demands algorithms which keep only the desired categories/subcategories in consideration during search. A novel 'Branch Selection Algorithm' has been designed and tested on a large hierarchical image database.

4 Related Databases and Database Used for Experimentation

The nature and scope of image data influences the performance of retrieval algorithms. For decades, in the absence of standard test data, researchers used self-collected images to show their results. Many domain specific and uncategorized databases came into existence lately for example, WANG, UW, IRMA 10000, ZuBuD, and UCID [3]. Some more challenging datasets are Caltech 101/256 [7], Coral Image, Tiny Image, ESP, LabelMe, Lotus Hill, and ImageNet [12].

A publicly available, densely populated, and semantically organized hierarchical image database covering a wide range of domains was required for experimentation. With large number of images for nearly all object classes, ImageNet serves the purpose. Built upon the backbone of the WordNet structure, a subset of ImageNet 2011 Winter Release given in Table 1 is used for experimentation. A category in ImageNet corresponds to a synonym set (synset) in WordNet. Fig. 1 shows some representative images of ImageNet.

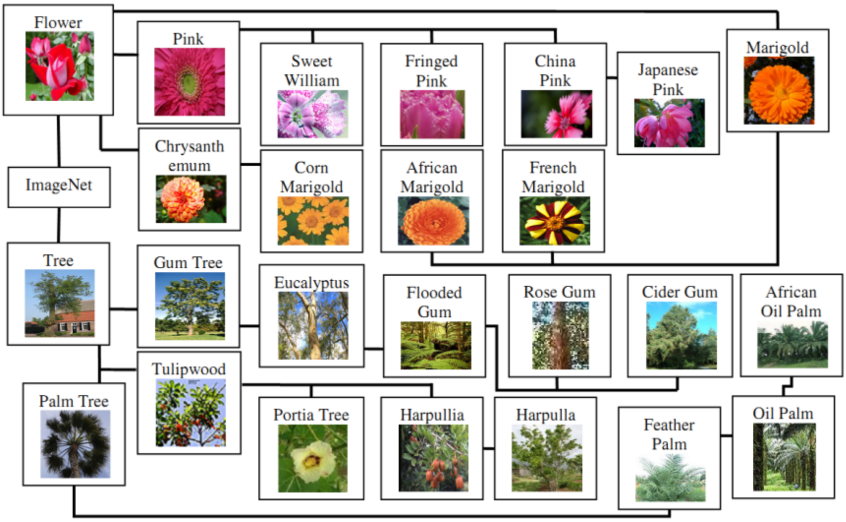


Fig. 1. A snapshot of Flower and Tree subtrees of ImageNet 2011 Winter Release

Table 1. Subset of ImageNet database used for experimentation

Subtree	Width	Depth	# of Synsets	# of Images (K)
Animal	9	9	32	38
Appliance	4	4	29	32
Fabric	2	5	12	11.5
Flower	9	3	24	26
Fruit	6	5	42	30.5
Geological Formation	5	5	50	55
Person	12	4	34	16.5
Sport, Athletic	5	4	23	30.5
Structure	6	6	36	33
Tree	7	6	42	24
Vegetable	6	5	41	35
Total (on an average 910 images per synset)			**365**	**332 K**

5 Methodology

The work visualizes categories/subcategories of a semantically categorized image database as nodes in the image tree. The flow of execution shown in Fig. 2 starts with an offline extraction of visual features of images. Visual features of images belonging to a node form visual signatures of that node. On the basis of the distance between query image and visual signatures of nodes, the Branch Selection Algorithm selects some subtrees to search. This search space is further reduced by pruning algorithms. Retrieval module assigns semantics of the nodes at lower distances to the query image. The proposed system supports both types of searches, i.e. aimed search to get a specific semantic; and category search to find a group of similar semantics.

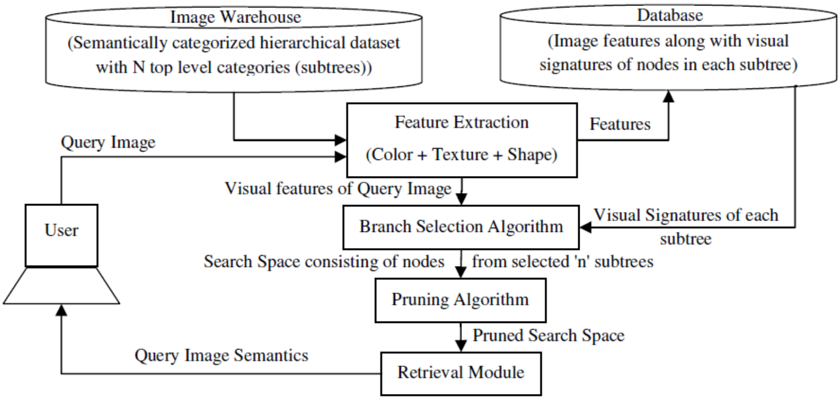


Fig. 2. Work flow of the proposed system

5.1 Feature Extraction Techniques

Conventionally, color, texture, and shape features are used to measure visual similarity of images. The combination of these features gives better results [13].

Color Features. Color is one of the most widely used low-level visual features. It is invariant to size and orientation of image [2]. It shows the strongest similarity to human eye [14]. Color histogram is the most commonly used representation. Various versions of histogram e.g. cumulative histograms, quantized color space histograms have been proposed [3, 15-16].

This work uses a histogram with perceptually smooth color transition in HSV color space [17]. When applied on an image as a whole, the Global Color Histogram (GCH) feature is obtained. Five color histograms corresponding to five regions (central ellipsoidal region and four surrounding regions) are concatenated to form a Local Color Histogram (LCH) feature. In general, GCH and LCH are represented as (1).

$$F_{GCH} = (h_1, h_2, \dots, h_{51}, i_{52}, \dots, i_{68}); h, i \textit{ represent hue and intensity resp .}$$
$$F_{LCH} = (h_1^1, h_2^1, \dots, h_{51}^1, i_{52}^1, i_{53}^1, \dots, i_{68}^1, \dots h_1^5, h_2^5, \dots, h_{51}^5, i_{52}^5, i_{53}^5, \dots, i_{68}^5) . \quad (1)$$

Another popular color feature is a statistical model of color representation [18-19]. Color distribution of each channel of an image is uniquely characterized by its three central moments i.e. average (E_i), variance (σ_i) and skewness (s_i) as given in (2).

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} , \sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2\right) \textit{ and } s_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^{\frac{1}{3}}\right) . \quad (2)$$

p_{ij} = value of i[th] color channel at j[th] image pixel and N = number of image pixels. Images are compared by taking a weighted sum of differences of corresponding color moments. Similarity between two images with r color channels and color moments (E_{i1}, σ_{i1}, s_{i1}) and (E_{i2}, σ_{i2}, s_{i2}) is given in (3).

$$d_{mom} = \sum_{i=1}^{r} w_{i1}|E_{i1} - E_{i2}| + w_{i2}|\sigma_{i1} - \sigma_{i2}| + w_{i3}|s_{i1} - s_{i2}| . \quad \text{where, } w_{kl} \geq 0 \text{ is specified by the user [19].} \tag{3}$$

Similar to histogram, Global Color Moment (GCM) and Local Color Moment (LCM) features of an image in HSV color space are obtained as shown in (4).

$$F_{GCM} = (E_1, \sigma_1, s_1, E_2, \sigma_2, s_2, E_3, \sigma_3, s_3);\ F_{LCM} = (E_1^1, \dots, s_3^1, \dots, E_1^5, \dots, s_3^5) . \tag{4}$$

Texture Features. Texture captures the information of patterns lying in an image. An image may contain textures of different degrees of detail. Grey level co-occurrence matrices (GLCM) and Tamura Features are popular single scale texture features. Multi-resolution texture features include Pyramidal Wavelet Transform (PWT), Tree-Structured Wavelet Transform (TSWT), Discrete Cosine Transform (DCT), Gabor filters, and ICA Filters [14]. The most frequently used Gabor filter is given by (5).

$$\begin{gathered} g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_y^2} + \frac{y^2}{\sigma_y^2}\right)\right] + 2\pi jWx . \\ g_{mn}(x,y) = a^{-m} g(x', y');\ m, n = int, m = 0,1, \dots, S-1, \\ x' = a^{-m}(x\ cos\theta\ + \mathrm{y}\ \sin\theta), y' = a^{-m}(-x\ sin\theta\ + \mathrm{y}\ \cos\theta) . \end{gathered} \tag{5}$$

Suitable dilations and rotations of the Gabor function g(x,y) through the generating function g_{mn} give a self-similar filter dictionary. Here $\theta = n\pi/K$, K = total number of orientations, S = number of scale, a = (U_h/U_l)-1/(S-1). U_h and U_l are upper and lower centre frequencies of interest [20]. This work uses Gabor filter with four scales and six orientations. For retrieval purposes the most commonly used measures are mean μ_{mn} and standard deviation σ_{mn} of the magnitude of the wavelet transform coefficients. The resulting Gabor Texture (GT) feature vector is given in (6).

$$F_{GT} = (\mu_i, \sigma_i), \quad i = 1,2,3, \dots, 24 \tag{6}$$

Shape Features. Shape features are powerful descriptors in image retrieval. Generic Fourier Descriptors, Zernike and Pseudo Zernike Moments, and Wavelet Descriptors are some popular representations [14]. Recent researches focus on computationally efficient local image descriptors. Scale Invariant Feature Transform (SIFT) extract large number of keypoints from image that leads to robustness in extracting small objects among clutter [8, 21]. This work uses SIFT with 4 octaves and 5 levels. K-means clustering forms 32 clusters per image [3]. For each cluster, count, mean and variance form a SIFT Shape (SS) feature vector given in (7).

$$F_{SS} = (CV_1, CV_2, \dots, CV_3), \left(MV_{1,1}, MV_{1,2}, \dots MV_{32,128}\right), (VV_{1,1}, VV_{1,2}, \dots VV_{32,128}) \,. \quad (7)$$

5.2 Construction of Visual Signature of a Node/Category

To correlate low-level visual features and high level semantics of images belonging to a node, a visual signature is attached to each node. Feature vector of an image is a combination of GCH, LCH, GCM, LCM, GT and SS. Mean feature vectors of all the images in a node, GCHmean, LCHmean, GCMmean, and LCMmean, GTmean, and SSmean form its visual signature. To get the semantics of an image, Branch Selection and Pruning algorithms make use of the similarity measures summarized in Table 2.

Table 2. Visual signatures and similarity measures

Visual Signature	Similarity Measure
GCHmean, LCHmean	Vector Cosine Distance
GCMmean, LCMmean	City Block Distance
GTmean	Euclidean Distance
SSmean	Earth Mover's Distance

5.3 Branch Selection Algorithm

The work proposes a novel Branch Selection Algorithm given in Fig. 3.

Steps to find the subtrees, semantically similar to query image, at each level are as follows:

Step 1: Calculate feature vector of the query image.

Step 2: Let there are N nodes at this level. Calculate the distance of query image with N nodes. For each feature, select n subtrees ($n \leq N$) having minimum distance from the query image. This results in three lists, one corresponding to each feature, containing n entries. It gives rise to any of the three possibilities:

a. If subtree X is 1st choice in all the three lists, then select only this subtree for search. As root of this subtree X has the closest distance with query image with respect to all the feature vectors considered. Go to Step 3.

b. If subtree X is 1st choice for any two lists, then select this subtree X for search. In addition,

   i. Select (n-1) more subtrees having maximum frequency of appearance in the two lists where X is 1st choice, and go to Step 3. In case, subtrees have same frequency then go to Step (ii).

   ii. Select one/more subtrees which have minimum sum of distances based on all 3 features.

   iii. Go to Step 3, if (n-1) subtrees are selected by now, otherwise go to Step (i).

c. If 1st choice of subtrees for all 3 lists is different, then select top n subtrees based on the maximum frequency of their appearance in these 3 lists. In case of a tie, select one/more subtrees which have minimum sum of distances based on all 3 features.

Step 3: Repeat Step 2 for subtrees at every level.

Fig. 3. Branch Selection Algorithm

Branch Selection Algorithm selects a few subtrees (n) out of ‘N’ available at the first level of the image tree. Only limited nodes that belong to these n subtrees are searched to find semantics of the query image. The algorithm aims to reduce the search space as much as possible, without compromising the accuracy of the system. The sum of distances based on GCH, LCH, GCM, and LCM is color distance. Distance based on GT is texture distance, and sum of distances based on SIFT Mean and Variance is shape distance. The algorithm prepares three lists corresponding to these distances and ‘adaptively’ selects a branch.

Performance of the system greatly depends on the value of n chosen. Experimental results for n=N/4, allows 75% pruning of the actual search space in terms of subtrees. Initial pruning for more than this results in rejection of the target subtree most of the time and therefore it is not fruitful to generate further results on its output.

An output of this algorithm for n=3 is shown in Fig. 4, where query image “n00450866_898” has been taken from “pony-trekking” synset. At the first level 11 subtrees are used for experimentation. The algorithm selects 3 subtrees (concepts) i.e. Geological Formation, Tree, and Sport, Athletic. At the subsequent levels, synsets of these three high level semantics are chosen to get the complete search space for this query image. In this case, algorithm selects only 51 synsets out of the total 365 synsets in the image tree. Thus search space is reduced by 86% w.r.t. number of synsets to be searched, still keeping the desired subtree in consideration.

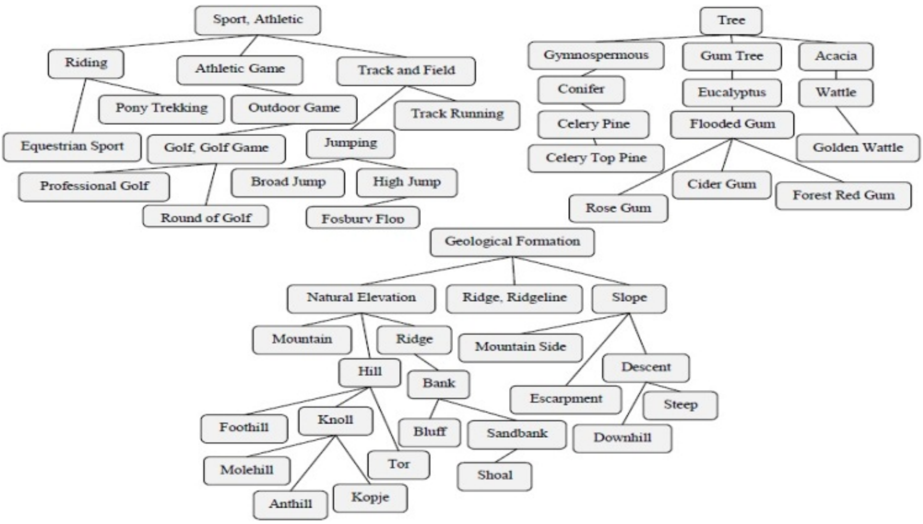


Fig. 4. Output of the Branch Selection Algorithm (n=3) for query image “n00450866_898”

5.4 Pruning Algorithms

Branch Selection Algorithm applied on an image database results in any number of nodes depending on the height and width of the n subtrees chosen in its step 2. Pruning of this search space would further improve the performance. This pruning helps in retaining good nodes while discarding the bad nodes of a selected subtree. A good node is the one that lies on the path leading to the node containing images semantically similar to the query image, while a bad node leads to either a different path in the same subtree or a different subtree. Ideally, bad nodes and their subtrees are to be pruned.

Goodness of a node is tested in terms of distances explained later in this section. In "strict pruning", the whole subtree is pruned if its root fails to prove itself good. While developing pruning approaches, it is observed that often a particular node on the path does not fulfills the criteria of being good one but the query image belongs to some lower level node of that path. Based on this observation, a "soft pruning" is proposed, which removes only the so-called bad node from the path and not the entire subtree following it. The children of this bad node become the children of its parent.

Fig. 5 explains these approaches with the same query image n00450866_898. Strict pruning shown in Fig. 5(a) loses the target synset "pony-trekking" because its parent "riding" fails to prove itself good. A less restricted soft pruning approach shown in Fig. 5(b) preserves the target synset even if its parent is being neglected. This less restrictive approach for pruning is followed in this work.

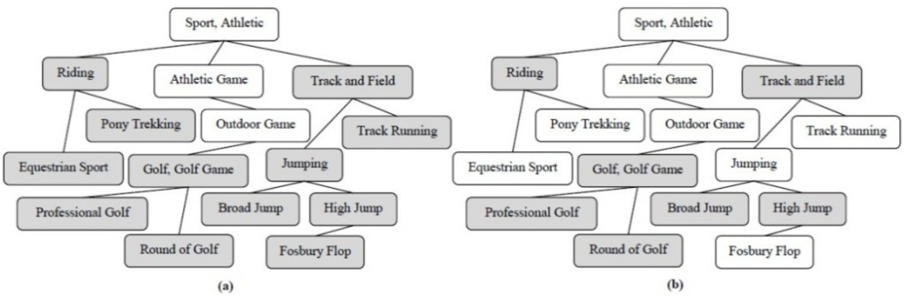


Fig. 5. (a) Strict Pruning. (b) Soft Pruning Approaches (gray nodes are the pruned ones).

Fig. 6 shows proposed pruning algorithms working on the distances corresponding to the dominant visual feature. Dominant feature of a subtree is the feature which gives top rank to this subtree. Dqs_i is the distance (already calculated) between query image and i^{th} node of the subtree (having N_s nodes) w.r.t. dominant feature. Dmean and Dmed are mean and median of Dqs. If a subtree having 10 nodes is given top ranking by texture feature, then mean and median of the GT based distances between query image and each of these 10 nodes are calculated. Additionally, extended mean distance (Dmeanx) and extended median distance (Dmedx) are calculated as shown in (8). Dmeanx is the sum of Average Absolute Deviation (AAD) in Dmean. Dmedx is the sum of Median Absolute Deviation (MAD) in Dmed. AAD and MAD are less affected by extreme observations than are variance and standard deviation [22].

$$Dmeanx = Dmean + \sum_{i=1}^{N_s}(|Dqs_i - Dmean|/N_s) \; .$$
$$Dmedx = Dmed + \; median(|Dqs_i - Dmed|) \qquad (8)$$

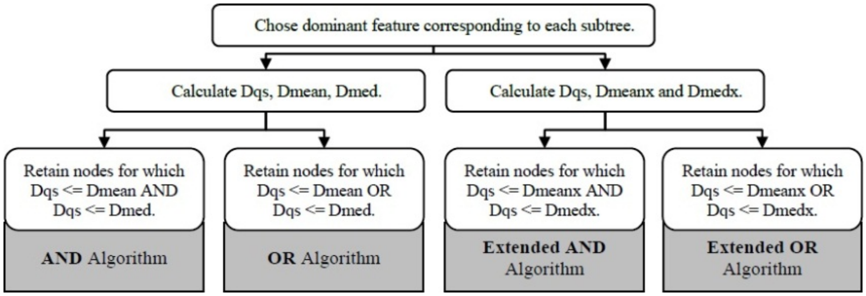


Fig. 6. Pruning Algorithms

In the quest of good pruning algorithms, the four possible combinations shown in Fig. 6 are exhaustively tested. The performance of pruning algorithms is judged on two parameters: The number of nodes retained to be searched in retrieval module and the ability to preserve the nodes appearing in the path ending at the target synset. The nodes retained by pruning algorithm are used to assign semantics to the query image.

6 Experimental Setup

The most common computing facility consisting of a PC with Intel Core 2 Quad processor, 8GB RAM and 500GB hard disk is used to get a fair idea about the performance of proposed algorithms. All experiments are performed on a subset of ImageNet shown in the Table 1. A set of query images is formed by automatically and randomly selecting 5% of images from each synset. Query images are taken from ImageNet database only because the attached semantic hierarchy with images helps to automate performance analysis of the branch selection and pruning algorithms. No manual intervention is required as human subjectivity may affect the understanding of correlation between visual and semantic similarity.

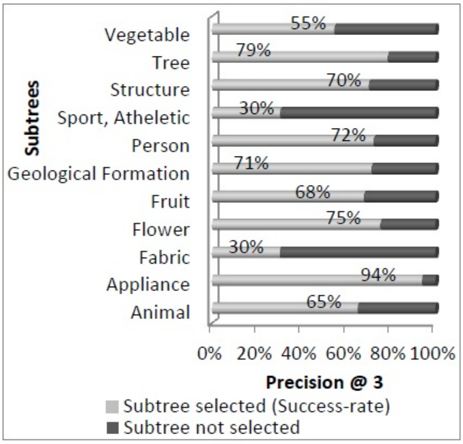
Feature vector of images and visual signatures of nodes in the database are generated through an offline procedure. During experimentation new images are not inserted in the database. In real life scenario, database may be kept in the updated mode. This insertion intensiveness can be easily handled in the online version. Insertion of an image requires its feature extraction and re-computation of the visual signature of the node to which this image is added.

7 Results and Discussion

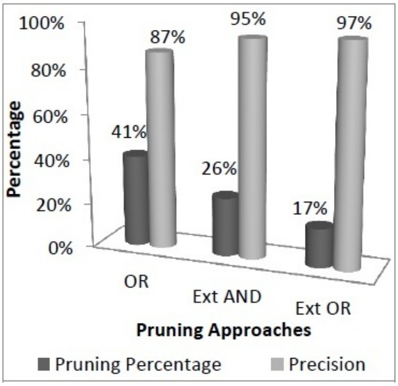
The following discussion establishes a correspondence between visual similarities and semantic similarity in a semantically categorized hierarchical image database.

7.1 Performance of Branch Selection Algorithm

Branch Selection Algorithm prunes the search space but the precise selection of target subtree based on the query image is to be ensured. This selection is expressed in terms of 'Precision' that denotes the selection of target subtree in terms of percentage. The graph in Fig. 7(a) shows the performance of Branch Selection Algorithm on 11 subtrees of ImageNet (Table 1) for n=3. This prunes the search space by 75%. Out of 11 hierarchies tested, 9 give more than 50% precision, while precision of 70% or more is achieved for 6 hierarchies. The algorithm out-performs if a query image is from 'Appliance' (94%), but opposite is the case if it is 'Fabric' or 'Sports, Athletic' (30%). This happens due to the nature of the images that constitutes these categories. Fig. 8 gives a glimpse of some of the images at the top level categories for these synsets. Appliance synset contains visually as well as semantically closer images, while 'Fabric' or 'Sports, Athletic' consist of images poorly related on the semantics. Visual signatures of nodes having dissimilar images do not represent these nodes well. This greatly affects the performance of the algorithm. Average precision @ 3 over all 11 branches is 65.36%, but on removing the two outliers i.e. 'Fabric' and 'Sports, Athletic', it becomes 73.22%, which is fairly acceptable. The target subtree is selected for approximately 75%, while pruning the search space by the same amount.



(a)



(b)

Fig. 7. Performance of (a) Branch Selection Algorithm. (b) Pruning Algorithm.

Fig. 8. Some representative images of three categories

The execution time of the algorithm is significantly affected by the width and depth of the subtrees selected at level 1. If the width of subtrees is more, then it would lead to selection of more subtrees and searching of; while higher depth means more iteration. On an average, execution time of the algorithm is 70 sec. For an online search this time is high but considering the computational facility used and the absence of an appropriate indexing of image features with this size of database, results are encouraging. In the real time environment algorithm would be executed at the server end with indexed feature vectors, which will significantly reduce the time.

7.2 Performance of Pruning Algorithms

It is desirable to have pruning algorithms with high pruning percentage and high precision. It is difficult to achieve high precision with high pruning percentage as they are inversely proportional to each other. The best algorithm would be the one that gives the highest pruning percentage with the desired precision. Performance of only three pruning algorithms is shown in Fig. 7(b) as they maintain good precision. AND operator is more restrictive and reduces search space significantly but results in poor precision. OR operator is less restrictive but improves precision. Ext AND algorithm seems to be the best with 95% precision and 26% pruning. The time required for Pruning Algorithms depend totally on the output of the Branch Selection Algorithm.

7.3 Semantics Assigned to Query Images

The query image is given the semantic of the nodes that are closer to it. Table 3 shows the output of branch selection and Ext AND pruning algorithm on query images.

Table 3. Images along with the semantics assigned to them by proposed approach

ImageNet Semantic	Query Image	Proposed Semantics		Query Image	Proposed Semantics	
		General	Specific		General	Specific
Animal		Animal Vegetable Geological Formation	Live stock, Ravine, Insectivore, Draw		Animal Tree Sports	Ungulate, Pachyderm, Animal, Ming tree
Appliance		Appliance Structure Person	Clothes dryer, Refrigerator, Coffee maker, Electric range		Appliance	Deep freeze, Clothes dryer, Oven, Wringer
Fabric		Fabric Fruit Structure	Hand towel, Viscos rayon, Towel, Honeydew		Fabric Appliance Sports	Rayon, Fabric, Towel, Pony-trekking
Person		Person Appliance Vegetable	Optimist, Personification, Neutral, Refrigerator		Person Fruit Sports	Neutral, Master of ceremonies, Entertainer, Person

The output shows the top four semantics assigned to the query image. A query image from any category; say Person, retrieves not only a general semantic 'Person' but also a number of specific semantics like 'Optimist, Personification, Neutral', etc. Presence of misclassified images in the database adversely affects the performance of proposed algorithms.

Table 4 lists some conflicting images in ImageNet. For example, the first image in the Table 4 belongs to 'animal' category while visually it seems to be a 'structure'. Proposed approach keeps it closer to the 'structure' semantics. The proposed approach also helps to identify such cases and reclassification of these images will further improve the performance.

Table 4. Some misclassified images and their correct classification by the proposed approach

Image	ImageNet Semantic	Proposed Semantics	
		General	Specific
	Animal	Structure, Animal, Vegetable	Parapet, Otter shrew, Support, Elephant
	Fruit	Tree, Sports, Flower	Gum tree, Gymnospermous, Conifer, Eucalyptus
	Fruit	Tree	Gymnospermous, Gum tree, Rose gum, Tree
	Flower	Tree, Geological Formation, Vegetable	Ravine, Forest red gum, Rose gum, Eucalyptus

7.4 Other Issues

Size of the Visual Signature of a Node. The size of the visual signature of a node although large for an online application, but the algorithms assign efficient semantics to the images. In future, efforts would be made to obtain compact visual signatures.

Lack of Comparative Evaluations. As most of the available image databases are flat in nature, the performance of proposed algorithms cannot be compared. Due to lack of hierarchy, subtrees selected by the Branch Selection Algorithm contain only a single node, which serves as both the root and the leaf. Pruning algorithms are also insignificant for flat structures. Further, most of the work done in this field is based on the personal databases and thus, it is not possible to get the results of the proposed algorithms on those databases.

In the present work, for the purpose of comparison, WANG database is categorized at the top level. Table 5 shows the performance of the proposed Branch Selection Algorithm on WANG and compares it with other related work. It gives an overall precision of 94.2% with 75% reduction in the search space. As a result the retrieval process is much faster in comparison to other approaches.

Table 5. A comparison on WANG database using average precision values

Category	Proposed Approach	F. Malik et al. [23]	R. Gali et al. [24]	P. Kinnaree et al. [25]
Reduction in search space	75%	0%	0%	0%
Africa	0.93	1	0.76	1
Beach	0.9	0.58	0.587	1
Bus	0.96	0.61	0.963	1
Dinosaur	1	0.71	1	1
Elephant	0.96	0.49	0.741	1
Flower	0.97	0.58	0.945	1
Food	0.9	0.48	0.733	1
Horse	0.95	0.72	0.941	1
Monument	0.9	0.57	0.714	1
Mountain	0.95	0.47	0.457	1
Average	0.942	0.621	0.7841	1

8 Conclusion and Future Scope

The paper discusses an open ended problem of semantic gap and proposes some algorithms to correlate visual and semantic similarity. The algorithms are developed for semantically categorized image database. The experiments show that visual features based on the adaptive combination of multiple low level features of image may serve well for a semantically categorized large image database. It shows that if categorized properly, low level features of the images can be combined with their semantics. The selection of good nodes by proposed algorithms ensures better performance of the system. Derived semantics can be used for effective image retrieval as a future research. Proper indexing of visual signatures can significantly reduce the time required for Branch Selection Algorithm. Inclusion of user feedback will also enhance the performance of retrieval system.

Acknowledgements. The authors acknowledge the support provided by JSPS to carry out this work under JSPS Invitation Fellowship for Research in Japan (Long-Term).

References

1. Sternberg, R.J.: Cognitive Psychology, 5th edn. Wadsworth Cencage Learning, Belmont (2008)
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Comput. Surv. 40(2), 5–60 (2008)
3. Deselaers, T., Keysers, D., Ney, H.: Features for Image Retrieval: An Experimental Comparison. Inf. Retr. 11(2), 77–107 (2008)
4. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. Pattern Recogn. 40(1), 262–282 (2007)
5. Wang, H.H., Mohamad, D., Ismail, N.A.: Semantic Gap in CBIR: Automatic Objects Spatial Relationships Semantic Extraction and Representation. IJIP International Journal of Image Processing 4(3), 192–204 (2010)

6. Li, J., Wang, J.Z.: Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. IEEE Trans. Pattern Anal. Mach. Intell. 25(9), 1075–1088 (2003)
7. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. 106(1), 59–70 (2007)
8. Lowe, D.G.: Object recognition from local scale-invariant features. In: International Conference on Computer Vision, ICCV 1999, vol. 2, pp. 1150–1157 (1999)
9. Zhuang, Y., Liu, X., Pan, Y.: Apply Semantic Template to Support Content-based Image Retrieval. In: Proceeding of IS&T and SPIE Storage and Retrieval for Media Databases, San Jose, California, USA, January 23-28, pp. 442–449 (2000)
10. Shotton, J., Winn, J., Rother, C., Criminisi, A.: *TextonBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
11. Liu, Y., Zhang, D., Lu, G.: Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recogn. 41(8), 2554–2570 (2008)
12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (June 2009)
13. Wang, X.Y., Yu, Y.J., Yang, H.Y.: An effective image retrieval scheme using color, texture and shape features. Computer Standards & Interfaces 33(1), 59–68 (2011)
14. Vassilieva, N.S.: Content-based Image Retrieval Methods. Program. Comput. Softw. 35(3), 158–180 (2009)
15. Liu, G.H., Yang, J.Y.: Content-based image retrieval using color difference histogram. Pattern Recogn. 46(1), 188–198 (2013)
16. Pandey, K.K., Mishra, N., Sharma, H.K.: Enhanced of color matching algorithm for image retrieval. International Journal of Computer Science Issues 8(3), 529–532 (2011)
17. Sural, S., Qian, G., Pramanik, S.: A Histogram with Perceptually Smooth Color Transition for Image Retrieval. In: 4th International Conference on Computer Vision, Pattern Recognition and Image Processing, Durham, North Carolina, pp. 664–667 (2002)
18. Shih, J.L., Chen, L.H.: Colour image retrieval based on primitives of colour moments. IEEE Proceedings on Vision, Image and Signal Processing 149(6), 370–376 (2002)
19. Stricker, M., Orengo, M.: Similarity of Color Images. In: SPIE Conference on Storage and Retrieval for Image and Video Databases III, San Jose, CA, USA, vol. 2420, pp. 381–392 (1995)
20. Manjunath, B.S., Ma, W.Y.: Texture Features for Browsing and Retrieval of Image Data. IEEE Trans. Pattern Anal. Mach. Intell. 18(8), 837–842 (1996)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
22. NIST/SEMATECH: e-handbook of statistical methods (2012), `http://www.itl.nist.gov/div898/handbook/eda/section3/eda356.htm`
23. Malik, F., Baharudin, B.: Quantized histogram color features analysis for image retrieval based on median and Laplacian filters in DCT domain. In: International Conference on Innovation Management and Technology Research (ICIMTR), Malacca, Malaysia, May 21-22, pp. 624–629 (2012)
24. Gali, R., Dewal, M.L., Anand, R.S.: Genetic Algorithm for Content Based Image Retrieval. In: International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), Phuket, Thailand, July 24-26, pp. 243–247 (2012)
25. Kinnaree, P., Pattanasethanon, S., Thanaputtiwirot, S., Boontho, S.: RGB Color Correlation Index for Image Retrieval. Procedia Engineering (8), 36–41 (2011)