Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 11; April-June, 2015 pp. 79-84

http://www.krishisanskriti.org/acsit.html

# A Survey of Semantic Gap Reduction Techniques in Image Retrieval Systems

**Navreen Kaur Boparai[1] and Amit Chhabra[2]**

*[1,2]Dept. of Computer Engineering & Technology Guru Nanak Dev University Amritsar, Punjab, India*
*E-mail: [1]navreen.boparai@yahoo.in, [2]amit.cse@gndu.ac.in*

**Abstract—***An efficient and accurate image retrieval system is required to handle the increased usage of images. Low-level image features – color, shape and texture give the machine description of an image by describing its visual content which doesn't match exactly with the high-level semantics of image. This mismatch corresponds to the semantic gap in image retrieval systems focusing only on low-level image features. They can satisfy the user's demand of finding similar or relevant images only to some extent due to the semantic gap problem. Research focus needs to be shifted towards minimizing this semantic gap between the machine description and the human semantics of images. This paper presents a comprehensive review of the work done in reducing the semantic gap while retrieving images from the image database. The five state-of-the-art semantic gap reduction approaches: ontology, relevance feedback, machine learning, semantic template generation, and web based image retrieval are discussed in the paper. All the five techniques are evaluated on the basis of attributes- user involvement, offline/online processing, accuracy, iterative nature, time consumed, and search space reduction. These techniques can be used either individually or in hybrid manner to boost the performance and the relevance in image retrieval systems. The paper also highlights various ways of integrating these techniques together.*

## 1. INTRODUCTION

In view of the rapid progress and advancement of internet and digital technologies, a lot of multimedia data such as images, audio and videos, is used today as a part of our daily life. This data needs to be stored and retrieved in an efficient and effective manner. Many image retrieval systems are developed to serve this ongoing demand of retrieving images in image databases. They are used in many areas like medical, fashion, architectural designs, advertising, crime prevention, digital forensics, surveillance system and many more. The traditional approach for indexing, searching and retrieving images is based on manual text annotations, called Text Based Image Retrieval (TBIR) systems, where the images are first annotated manually by text or keywords. The commonly used TBIR system is Google Images. However, it becomes very difficult to express the whole visual content of images in keywords which may give irrelevant results. This technique requires vast amount of manual labor and effort. In addition, manual indexing of images is not always correct and is time-consuming. Content-Based Image Retrieval (CBIR) systems solve the above problems of TBIR systems as images are indexed by their visual contents rather than keywords.

A CBIR system describes visual contents of the images in terms of low level features-color, texture, shape and spatial locations to represent the images in the databases. Thus, images are indexed automatically by their feature vector. CBIR system first extracts the features of the query image resulting in a feature vector, which is then compared to that of the images in the database using similarity measures like Euclidean distance, Minkowski-Form distance, etc. Finally, the similar images having the least distance are shown as the retrieval results. The performance of such system depends on choosing the most effective visual features and similarity metrics.

In CBIR systems, database images are not annotated with any keyword either manually or automatically. Traditional CBIR system has no understanding of the image's semantics and cannot meet the user's needs due to the "semantic gap" [1] between the low-level content descriptors (features) and the high-level human perception of concepts. For example, a fish may be considered as the same as an airplane, and a red flower the same as a rising sun, and etc. Thus, semantic gap refers to the limitations of low-level image features in describing human semantics. This leads to development of Semantic Based Image Retrieval (SBIR) which is more user-oriented as it supports query by keywords i.e. high-level concepts. The SBIR tries to minimize this 'semantic gap'. SBIR is also called automatic image annotation system as images are annotated with keywords obtained by automatically learning the semantics of the images. This learning is done after low-level feature extraction. Therefore, semantic image retrieval depends on both low-level features and high-level keywords of the images.

This paper gives a detailed survey of approaches that can be used for minimizing this semantic gap problem and is organized in the following sections: section 1 gives a detailed explanation of semantic gap reduction techniques, section 2 compares all the mentioned techniques and section 3 concludes the paper.

## 2. SEMANTIC GAP REDUCTION TECHNIQUES

Nowadays, image retrieval systems are becoming more and more interesting as lot of work has been done in this field and still many more advancements are waiting. But, main research problem is minimizing the semantic gap between low-level feature descriptors and high-level user semantics in CBIR systems. This section discusses the approaches being suggested in various papers to bridge this semantic gap problem.

Human perception should be considered at every step of image retrieval including initial preprocessing steps such as image segmentation. Choras [2] defined filtering, segmentation, and object identification as the preprocessing steps, resulting in a number of significant regions and objects. Image segmentation separates the desired objects from the background [3] and then extracting features from these objects. These methods represents image at object level and thus, overcoming the limitations of global features. This causes retrieval to be close to the human visual perception. Even low-level feature descriptors should also be chosen keeping in mind this human way of seeing things. For instance, RGB color space image is converted into its HSV components [4] and wavelet transform is used as texture descriptor. Chaudhari et al. [5] focused the integration of low level features by assigning them weights according to user's choice.

Techniques for reducing the semantic gap and deriving high-level semantics can be classified into five categories:

(1) Object ontology,
(2) Machine learning tools,
(3) Relevance feedback (RF),
(4) Generating semantic template (ST),
(5) Web image retrieval.

### 2.1 Object Ontology

Object ontology [6, 7, 8] defines high-level concepts and their relationships forming a hierarchical structure. It defines an image using its semantics. The intermediate level descriptors representing high-level concepts are described for low level features of images [9]. Such descriptors form a vocabulary. For example, grass is described as a region of "green" color and "lower" spatial location. Further, levels of green color can be "light green", "medium green", and "dark green". An ontology-based system is described in [10, 11], in which the regions of an image are represented with their color in appropriate color space, their position in both horizontal and vertical axis, their shape and size. This representation is done using intermediate-level descriptors, which are mapped to the values of corresponding low-level features. Thus, high-level concepts are associated with the relevant image regions.

Ontology can also represent the hierarchical relationship between the objects/concepts defined in the system. A high-level image-semantic ontology tree representing the hierarchical relationship between the concepts is first constructed in the system defined in [12], for traversing and finding the semantic domains relevant to the input image. Then a visual comparison is to be done for finding similarity between the images in the reduced domains and the input image. This results in retrieving target images with high precision. Thus, few numbers of comparisons are required that too only in reduced domains for capturing the category semantics associated with an input image.

To support semantic-based image retrieval, color naming can be used effectively to quantize color information. Berk, Brownston and Kaufman developed a color naming system CNS[13] that quantizes the hue values to a group of basic colors such as red, blue, green, yellow, orange, purple, brown, white, black and grey. But no proper texture naming system is available yet.

### 2.2 Machine Learning Tools

Second technique is using machine learning tools [6, 8] like supervised and unsupervised learning, for associating low-level features to query concepts. Supervised learning is often used to predict high-level semantic concepts on the basis of low-level image features. Unsupervised learning such as k-means clustering, organizes the input features to form clusters without having any outcome measures.

**2.2.1. Supervised Learning.** In supervised machine intelligence, low level features from a number of images are extracted and fed to a binary classifier like Bayesian classifier, support vector machine (SVM), Decision tree, Neural network which are trained to detect semantic category label.

Support vector machine (SVM) [14] works efficiently for small training datasets as it needs only those training samples called support vectors, closest to the optimal separating hyper plane. If each sample or image in the training data set is represented by a feature vector and corresponding class label, the SVM classifier finds an optimal hyper plane from the training samples and separates the images by calculating the maximum distance amongst different classes. Being a binary classifier, SVM can learn only one class at a time. However, multiple binary classifiers can be used to learn more than one class.

Decision tree techniques can also be used to learn semantic concepts. A decision tree (DT) [6] is a decision making tool that classifies the images or decides the class labels of images using a set of human readable "if–then-else" rules. During training, a DT is constructed by partitioning the labeled training images recursively into separate sets depending on the value of attributes (such as color, texture, shape) of samples at each internal node. This partitioning is repeated till all the images of a set belong to the same class or no such attribute is there to divide them. At the end, leaf nodes denote the class of the samples remained in that node. Each individual path from the root node to the leaf node corresponds to a separate decision rule. After training, class of a new sample/query image can be learnt by traversing the tree from the root node

to a leaf node using the sample's attribute values and class of the query image is that of the leaf node where the search ends. Various DT algorithms [14] used are ID3, C4.5, and CART which differs by the attributes being used and their selection criteria, etc. ID3 and C4.5 are often used in relevance feedback loop. The CART algorithm maps global color distribution to textual description of images using a set of decision rules.

An Artificial neural network (ANN) is a learning network which is trained using low-level features of an image to learn its semantic class. Nagathan and Manimozhi [4] used a feed forward back-propagation neural network (FFBP) in their proposed system, which consists of multiple layers of interconnected neurons- input layer, hidden layer and output layer. In the training phase, FFBP computes output in forward direction and error in backward direction as the difference between that obtained output and the required output. Weights of the layers are changed till the error becomes zero. Training ends and weights are fixed when correct results are obtained. In the testing phase, network is tested with a query image to learn its class and return similar images. ANN is more suitable as a classifier due to its dynamic adjustment of weights. In spite of being computationally intensive, ANN shows high accuracy even in large image databases.

Of all the supervised learning methods, the DT learning is comparatively simpler to understand and insensitive to incomplete and noisy image data.

**2.2.2. Unsupervised Learning.** The task in unsupervised learning is to group together similar regions in an image without any measurements of outcome as in supervised learning. Image clustering is one such type of learning which combines together similar pixels forming a cluster. Clustering algorithms such as k-means[14,6], combine image pixels into different groups by partitioning an image into blocks of size 4*4 pixels and extracting color and texture features from these blocks. Then blocks with similar feature vectors are grouped together to form a cluster.

### 2.3 Relevance Feedback

Relevance feedback (RF) [6, 8] bridges the 'semantic gap' between low level features and what the user thinks by bringing the user in the retrieval process. RF boosts the CBIR systems through iterative and interactive learning. The steps for RF method in CBIR are given below:

(1) The initially retrieved images for the query image are shown to the users.

(2) User indication of above results as relevant and non-relevant images.

(3) Using Query refinement methods or Machine learning tools to know the needs of the user. Then repeat step (2).

Steps (2) and (3) are executed iteratively until the user gets satisfied with the retrieved images. The common approach in step (3) is to allow the user to refine the query representation or to use machine learning methods to categorize the relevant and the non-relevant images. The query refinement mechanism moves the query closer to the more relevant images, starting from the user desired objects in the query image and followed by the modification of query representation in feedback iterations. Query modification [15] can be done by using either of two techniques: query point movement and updating weight vector.

1) Query Point Movement: It moves the query point towards the relevant images (points) and distant from the irrelevant images to improve its estimate. Rocchio's formula [5, 6] mostly used to enhance this estimation iteratively, is given in (1),

$$Q' = \alpha Q + \beta \left( \frac{1}{N_R'} \sum_{i \in D_R'} D_i \right) - \gamma \left( \frac{1}{N_N'} \sum_{i \in D_N'} D_i \right) \quad (1)$$

where $Q$ and $Q'$ are the previous query and refined query, respectively, $D'_R$ and $D'_N$ are the sets of relevant and irrelevant images indicated by the user, $N'_R$, $N'_N$ are the number of images in $D'_R$ and $D'_N$, respectively, and $\alpha, \beta, \gamma$ are constants, also known as weight parameters [with$(\alpha + \beta + \gamma = 1)$ ] responsible for the relative significance of the previous query, the relevant images and the irrelevant images, respectively.

2) Updating Weight Vector (Re-weighting): Query weighting dynamically adjusts the relative weights of different low-level features in the query representation to represent the high-level concepts. If each image represented as an n-dimensional feature vector is seen as a point in an n dimensional space, then the main idea is to assign larger weights to those dimensions of a feature which are more important in retrieving relevant images and smaller weights to those which are less important in doing so.

Machine learning techniques such as decision trees, SVMs can also be used in step 3 of RF loop to learn the user's feedback. Hong et al. [16] proposed an approach to apply SVMs to separate the relevant images and the non-relevant images using the distance from the optimal hyperplane. In [17], a SVM active learner, $\text{SVM}_{\text{Active}}$ is proposed which combines active learning with SVMs for accurate and fast learning of a concept.

Lai and Chen [18] reduce the semantic gap by providing relevance feedback through interactive genetic algorithm (IGA), which evolves the image retrieval results. The fitness function in IGA is constructed by considering the user's evaluation rather than using the predefined mathematical formula as in genetic algorithm. The chromosome representation of the solutions of the problem represents the three features-color, texture, and edge of an image. In the beginning of the IGA process, first retrieval results of a query image are used as initial population. Evaluation function ranks

all the chromosomes in the population with respect to their "fitness" by considering the similarity between images and impact factor of user's preferences. The fitness function which evaluates the quality of the chromosome C corresponding to the query q, is given in (2):

$$F(q, C) = w_1 . sim(q, C) + w_2 . \delta \quad (2)$$

where sim(*q*,*C*) corresponds to the extent of similarity between images, $\delta$ represents the impact factor of user's preferences, $w_1$ and $w_2$ are the coefficients determining their respective importance in calculating the fitness, and $\sum w_i = 1$.

Then, chromosomes having better fitness are chosen to produce the new off springs for the next generation using genetic operators. Crossover operator produces new chromosomes by swapping genetic parts of the randomly paired chromosomes. Mutation operation can be skipped to speed up the process as it creates a new chromosome. Thus, the system iteratively generates the new population of images using the user's relevance information until the user is satisfied. Arevalillo-Herráez et al. [19] solved the problem of IGA having only smaller number of positive selections by including adjoining or nearby individuals of the positive selections, and assigning them a lower fitness value.

### 2.4 Semantic Templates

Fourth technique is generating semantic template (ST) as an association between low level features and high level concepts. Chang et al. [20] represented high-level concepts using templates consisting of a number of objects, icons or example scenes such as sunrise, waterfall, etc. Initially, user specifies the concept by giving details of objects, their attributes (color, shape, texture, etc.) and weights to be assigned to the features of these objects. Using relevance feedback, system then finally converges to a template that best matches the user's concept. But the user needs to have detailed knowledge of feature representation which is not suitable for normal user. Compared to this, Zhuang et al. [21] integrated the template generation process with the interactive relevance feedback without requiring the user to have any knowledge of feature representation. The system calculates the centroid of images which are returned after several interactions with user and are highly relevant to the query image. This centroid vector represents the high-level concept. The ST is described by a triplet, ST={C, F, W}, where C represents the user's concept, F is the centroid feature vector, W corresponds to the weight of feature vector. Moreover, pre-existing semantic templates can be associated by forming a semantic network via WordNet [22]. Then the system returns relevant images by using the F and W of a ST in the network, which corresponds to the user's query. Smith and Li [23] presented a method to decode image semantics using composite region templates (CRTs) which describe prototypal spatial arrangements of regions and features in the images. CRTs are generated by combining the segmented region strings of images. The CRTs of each semantic class are grouped together to form a CRT library. The system learns the semantics of query images by matching their region strings with the values in the CRT library.

### 2.5 Web Image Retrieval

In web based image retrieval, both the visual and textual features needs to be extracted from the images as web images have huge metadata like URL, filename and surrounding text. Ren [24] presented a Web image retrieval model transforming web images to their annotation keywords or semantics by integrating their text features(keywords), visual features (color, shape and texture) and hyperlinks. The HTML documents are fetched using a web crawler and keywords are extracted from: (1) descriptive tags containing the file name, ALT attributes of image tags, and image anchor tags, (2) Meta tags containing the HTML document title and (3) text passages; of HTML documents as informative sources of web image contents. These keywords are used to match the best images in the retrieval system. Vadivu et al. [25] analyzed the attributes of the <img src> TAG of the HTML documents to retrieve relevant images from the web and assign them weights according to their importance in expressing the semantics of the image.

## 3. COMPARATIVE ANALYSIS

All the above five methods of reducing semantic gap between machine description and human semantics of images are compared in table I shown below, on the basis of attributes: user involvement, offline/online processing, accuracy, iterative nature, time consumed, search space reduction. The user is involved in the feedback loop to give feedback regarding relevance of results. This leads to iterative nature of relevance feedback (RF) method as retrieval results are repeatedly refined until the user is not satisfied. This also corresponds to higher accuracy and performance boost in RF technique. Accuracy here means relevance of images returned. But, at the same time, it seems to be time-consuming and tiring for user. On the other hand, Ontology has a great advantage of reducing search space, which means the images are searched for similarity only in the reduced domain of the ontology tree resulting in reduced time. But, here accuracy is being compromised as intermediate-level descriptors are not available for texture naming. Mostly, there is a trade-off between accuracy and time consumed. Best way is to integrate one or more of these five techniques to improve precision of the CBIR systems. Semantic template (ST) is often combined with RF to make the generation of STs [20, 21] user-friendly. As machine learning method can learn complex semantics of an image, it can be combined with RF [16, 17] to enhance the performance. RF can also be integrated with both object-ontology and machine learning [26]. Web image retrieval systems have the lowest accuracy, but they often use one or more of the other four techniques to extract semantics of an image.

**Table 1: Comparison of Semantic Gap Reduction Techniques**

| Technique | User Involvement | Type of processing | Accuracy | Iterative | Time-consuming | Search Space Reduction |
|---|---|---|---|---|---|---|
| Ontology | No | Offline | High | No | No | Yes |
| Machine Learning | No | Offline | High | No | No | No |
| Relevance Feedback (RF) | Yes | Online | Very High | Yes | Yes | No |
| Semantic Template (ST) | Yes | Online, often combined with relevance feedback for template generation | High | May or may not be, depending on the template generation method | May be , if combined with relevance feedback | No |
| Web Image Retrieval | No | Offline | Low | No | Yes | No |

## 4. CONCLUSION

This paper analyses the performance of different semantic gap reduction methods that can be applied to minimize the semantic gap between machine and human understanding. This Semantic imbalance needs to be balanced by considering human way of seeing things at each and every step of the image retrieval process. The five basic semantic gap reduction techniques being discussed are: ontology, relevance feedback, machine learning, semantic template generation and web based image retrieval. These techniques can be integrated together to benefit from each other resulting in better performance than using either of them alone. It can be concluded by analyzing the various attributes of these techniques that integration of RF with machine learning results in higher efficiency as it takes advantage of higher accuracy of RF and complex semantics understanding of machine learning method. Selection of a particular method also depends upon the tradeoff between time and accuracy according to the requirements.

## REFERENCES


[1] Riad, A. M., Elminir, H. K., & Abd-Elghany, S. (2012). A Literature Review of Image Retrieval based On Semantic Concept. *International Journal of Computer Applications*. *40*(11), 12-19.

[2] Choras, R. S. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *International journal of biology and biomedical engineering*. 1(1), 6 -16.

[3] Ning, J., Zhang, L., Zhang, D., & Wu, C. (2010). Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*. *43*(2), 445-456.

[4] Nagathan, A., & Manimozhi, M. I. (2013). Content-Based Image Retrieval System Using Feed-Forward Backpropagation Neural Network. *International Journal of Computer Science Engineering*. 2(4), 143-151.

[5] Chaudhari, S., Chilveri, R., Nanda, A., & Borse, R. (2012, August). Efficient Implementation of CBIR System and Framework of Fuzzy Semantics. In *Proceedings of the IEEE International Conference on Advances in Mobile Network, Communication and its Applications (MNCAPPS)*. 111-114.

[6] Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*. *40*(1), 262-282.

[7] Kumar, K. V., Rao. R. R., Ramaiah, V. S., & Kaka, J. R. (2012). Content Based Image Retrieval System Consume Semantic Gap. *International Journal of Computer Science and Information Technologies*. *3*(5), 5231-5235.

[8] Goyal, N., & Singh, N. (2014). A Review on Different Content Based Image Retrieval Techniques Using High Level Semantic Features. *International Journal of Innovative Research in Computer and Communication Engineering*. *2*(7), 4933-4938.

[9] Khodaskar, A., & Ladke, S. A. (2012, June). Content Based Image Retrieval with Semantic Features using Object Ontology. In *International Journal of Engineering Research and Technology*. 1(4). ESRSA Publications.

[10] Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2003, September). An ontology approach to object-based image retrieval. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. 2, 511-514.

[11] Sheelavant, S. S., & Dharani, A. (2013). Region-based image retrieval with high-level semantics–A Comprehensive Investigation. *International Journal of Engineering and Technical Research.* 1(7), 17-20.

[12] Li, X., Shou, L., Chen, G., & Ou, L. (2006). A latent image semantic indexing scheme for image retrieval on the web. In *Web Information Systems–WISE 2006*. 315-326. Springer Berlin Heidelberg.

[13] Berk, T., Brownston, L., & Kaufman, A. (1982). A new color-naming system for graphics languages. *IEEE Computer Graphics and Applications*. *2*(3), 37-44.

[14] Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*. *45*(1), 346-362.

[15] Patil, P. B., & Kokare, M. B. (2011). Relevance Feedback in Content Based Image Retrieval: A Review. *Journal of Applied Computer Science & Mathematics.* 10 (5), 41-47.

[16] Hong, P., Tian, Q., & Huang, T. S. (2000). Incorporate support vector machines to content-based image retrieval with relevance feedback. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2000. 3, 750-753.

[17] Tong, S., & Chang, E. (2001, October). Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia*, 2001. 107-118.

[18] Lai, C. C., & Chen, Y. C. (2011). A user-oriented image retrieval system based on interactive genetic algorithm. In *IEEE Transactions on Instrumentation and Measurement. 60*(10), 3318-3325.

[19] Arevalillo-Herráez, M., Ferri, F. J., & Moreno-Picot, S. (2009, October). An interactive evolutionary approach for content based image retrieval. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2009. 120-125.

[20] Cheng, S. F., Chen, W., & Sundaram, H. (1998, October). Semantic visual templates: linking visual features to semantics. In *Proceedings of the IEEE International Conference on Image Processing (ICIP), 1998.* 531-535.

[21] Zhuang, Y., Liu, X., & Pan, Y. (1999, December). Apply semantic template to support content-based image retrieval. In *Electronic Imaging*. 442-449. International Society for Optics and Photonics.

[22] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, *3*(4), 235-244.

[23] Smith, J. R., & Li, C. S. (1998, June). Decoding image semantics using composite region templates. In Proceedings of the *IEEE Workshop on Content-Based Access of Image and Video Libraries, 1998.* 9-13.

[24] Ren, L. (2010, May). Web image retrieval in web pages. In Proceedings of the *2nd IEEE International Conference on Future Computer and Communication (ICFCC)*, 2010. 1, 26-31.

[25] Vadivu, P. S., Sumathy, P., & Vadivel, A. (2012). Image Retrieval From WWW Using Attributes in HTML TAGs. *Procedia Technology*, 6, 509-516.

[26] Zarchi, M. S., Monadjemi, A., & Jamshidi, K. (2014). A semantic model for general purpose content-based image retrieval systems. Computers & Electrical Engineering. 40(7), 2062–2071.