A comprehensive survey on the reduction of the semantic gap in content-based image retrieval

Jayant Jagtap*

Symbiosis Institute of Technology (SIT),
Symbiosis International (SIU) (Deemed University),
Lavale, Pune, Maharashtra, India
Email: jayant.jagtap@sitpune.edu.in
*Corresponding author

Nilesh Bhosle

Department of Electronics and Telecommunication Engineering,
D.Y. Patil College of Engineering,
Ambi, Pune, Maharashtra, India
Email: bhoslenp@gmail.com

Abstract: In the last few decades, content-based image retrieval is considered as one of the most vivid research topics in the field of information retrieval. The limitation of current content-based image retrieval systems is that low-level features are highly ineffective to represent the semantic contents of the image. Most of the research work in content-based image retrieval is focused on bridging the semantic gap between the low-level features and high-level semantic concepts of image. This paper presents a thorough study of different techniques for the reduction of semantic gap. The existing techniques are broadly categorised as: 1) image annotation techniques to define the high-level concepts in image; 2) relevance feedback techniques to integrate user's perception; 3) machine learning and deep learning techniques to associate low-level features with high-level concepts. In addition, the general architecture of semantic-based image retrieval system has been discussed in this survey. This paper also highlights the current and future applications of content-based image retrieval. The paper concludes with promising future research directions.

Keywords: content-based image retrieval; CBIR; semantic gap; deep learning; image annotation; image retrieval; information retrieval; relevance feedback; survey.

Biographical notes: Jayant Jagtap received his BE in Electronics and Telecommunication Engineering at Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India, in 2008, MTech in Electronics Engineering at Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra, India, in 2011, and PhD in Electronics and Telecommunication Engineering at Swami Ramanand Tirth Marathwada University, Nanded, Maharashtra, India, in 2017. He is presently working as an

Assistant Professor in the Department of Electronics and Telecommunication Engineering, Symbiosis Institute of Technology, Pune, Maharashtra, India. His current research interests include digital image processing, pattern recognition, computer vision, medical image processing and machine learning.

Nilesh Bhosle received his BE in Electronics Engineering and MTech in Electronics Engineering from Shri Guru Gobind Singhji Institute of Engineering and Technology Nanded, Maharashtra, India in 2003 and 2006, respectively. He completed his PhD degree from Swami Ramanand Teerth Marathwada University Nanded, India in March 2018. From August 2003 to July 2004, he worked in the industry. In October 2007, he began his career in academics as a Lecturer in the Department of Electronics and Telecommunication Engineering, ICFAI Tech Hyderabad, India. He is presently an Associate Professor in the Department of Electronics and Telecommunication Engineering, D.Y. Patil College of Engineering, Ambi, Pune, India. His research interest includes image processing, content-based image retrieval, and pattern recognition. He is a life member of Indian Society for Technical Education.

1 Introduction

The expeditious advancement in technology for computation and communication of multimedia has prompted expanded requests for interactive multimedia data. This in turn created a demand for efficient tools for retrieval of multimedia information such as images and video retrieval from large databases. These tools are required by users from various domains like medicine, crime prevention, and remote sensing. Hence the need of effective image retrieval tools is turning into an essential part of the present frontline innovation.

To fulfil this requirement, many researchers have contributed by proposing efficient image retrieval systems. These systems are broadly categorised into two main frameworks Lakdashti et al. (2008): text-based image retrieval (TBIR) and content-based image retrieval (CBIR). In TBIR method set of keywords are assigned to images manually and used for retrieval. This method suffers from two drawbacks. Firstly, the manual annotation takes more time. Secondly, there are chances of human error. These drawbacks of TBIR were dealt by proposing CBIR approach. In CBIR approach, images are annotated with the global observable features like texture and shape (Faloutsos et al., 1994; Pentland et al., 1996; Gupta and Jain, 1997). The performance of CBIR systems, based on low-level features is not up to the mark because of two fundamental problems. One of the problems is semantic gap. This can be considered as a difference between the information extracted from the image and the interpretation of the image by the human (Smeulders et al., 2000). Humans use high-level features to understand the images. The features extracted using computer vision methods are low-level features. This introduces the gap between high-level and low-level features, known as semantic gap. Hence, the performance of CBIR systems is not up to the mark. Another problem with CBIR system is the variation in the human perception. The human perception varies with persons and circumstances.

To overcome these problems, the CBIR system must be developed by considering high-level querying and browsing. Initially, the CBIR methods proposed by researchers were focusing on extracting the efficient feature for image representation. Now in recent years the researchers are contributing to reduce the semantic gap (Smeulders et al., 2000) between low-level features and high-level image concept. A detailed survey of existing CBIR systems can be found in Smeulders et al. (2000), Kokare et al., (2002) and Oussalah (2008). A comprehensive survey of different techniques proposed by researchers to reduce the semantic gap in CBIR is presented by Liu et al. (2007).

A survey on state-of-the-art techniques for the reduction of the semantic gap in CBIR is presented in this paper. The existing techniques are mainly divided into three broad categories, namely image annotation techniques, relevance feedback (RF) techniques and machine learning and deep learning techniques. This paper also discusses the general architecture of semantic-based image retrieval system and some current and future applications of CBIR. The paper concludes with future research directions to encourage the emerging researchers in the field of CBIR.

The rest of the paper is organised as follows. The architecture of semantic-based image retrieval system is discussed in Section 2. Section 3 explains the different techniques available in the literature for reduction of semantic gap. Section 5 deals with the discussion and future research directions. Finally, the paper is concluded in Section 6.

2   System architecture

The general architecture of semantic-based image retrieval system is shown in Figure 1. The architecture is divided into two categories, namely feature extraction and query processing.

1   Feature extraction

> The feature extraction section concerns with the extraction of semantic features of the database images. It is further divided into sub-blocks such as database, low-level feature extraction, object/region recognition and semantic feature extraction.
>
> a   Database: The first step in system realisation is database collection. It consists of images among which the query image will be searched.
>
> b   Low-level feature extraction: The visible information in the input images is extracted in the form of low-level features. The extracted features are then given as input to the object/region recognition section.
>
> c   Object/region recognition: To derive the semantic features of the image it is first segmented into meaningful regions. The meaningful regions can be identified using the similarity characteristics of the low-level features.
>
> d   Semantic feature extraction: In this section the identified regions/objects are transformed into semantic features. The semantic features can be extracted by associating the semantic words to the identified regions using the concept mapping and image annotation process. The extracted semantic features are then stored into the database.

2   Query processing

In semantic-based image retrieval systems, the user specifies the query using the keywords or textual words. When the user submits the query using the keywords or texts, the system will go through two main steps that are semantic feature translator and semantic image mapping.

a   Semantic feature translator: The keywords submitted by the user will be converted into semantic features using a semantic feature translator. The semantic feature translator will first identify the meaningful concept (regions/objects) from the query words and assigns the list of possible low-level features to these concepts. The identified regions/objects are then transformed into semantic features by associating the semantic words using the concept mapping and image annotation process.

b   Semantic image mapping: In this step the semantic features of the query will be matched with the semantic features of the database images. After the mapping of the semantic features, the database images matched to the query will be retrieved and submitted to the user.

Figure 1   Architecture of semantic-based image retrieval system
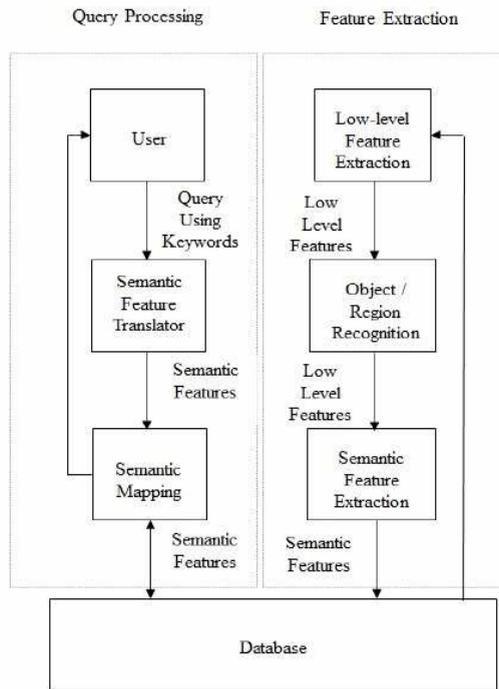
3   **Reduction of semantic gap**

The various techniques proposed in the literature for reduction of the semantic gap in CBIR are divided into three categories, namely Image annotation techniques to define the

high-level concepts in image, RF techniques to integrate user's perception and machine learning and deep learning techniques to associate low-level features with high-level concepts.

3.1   Image annotation techniques to define the high-level concepts in image

Assigning the keywords to the images to capture their semantic contents is known as image annotation. Image annotation is of prime importance for reduction of semantic gap. In image annotation, labels are assigned to the images based on some information about the image contents. There are three approaches to assign the labels to the images viz. manual annotation, semi-automatic image annotation (Djordjevic and Izquierdo, 2007; Xuelong et al., 2007) and automatic image annotation (Lu et al., 2005; Jiang et al., 2009). Manual annotation approach, in which the annotations are assigned to the images manually, is considered to be tedious and time consuming process. In automatic annotation approach the labels are assigned to the images automatically. In such approach manually annotated images are considered as the training dataset and the annotations of these images are then propagated to the unlabelled images using the machine learning methods. In semi-automatic approach the user feedback will be considered to annotate the images.

There exists two ways of image annotation namely image level and sub-image (object or region) level. Image level annotation (Vailaya et al., 1998; Han et al., 2005; Fan et al., 2005) is accomplished through image classification approaches. In image classification approach, the system classifies the images according to the semantic groups using the global features of the image. Region-based annotation (Wang et al., 2006; Mezaris et al., 2003; Liu et al., 2005; Asghar and Rao, 2008) uses the regional features to annotate the images. Some researchers also used both the regional and global features to annotate or classify the images (Nguyen et al., 2009; Fan et al., 2008).

3.1.1   Image level annotation

Image level annotations can be achieved through image classification approaches. Image level annotation systems are easy to implement and most suitable for images having few semantic concepts. Vailaya et al. (1998) discussed a system to learn high-level concepts from low-level features to classify the images into the city (containing street, buildings, and bridges) and landscape (containing mountains, forests, beaches and farmlands). The input to the system is an image, which is then compared with labelled images (training set) and the output is the confidence level belonging to the particular classes using weighted k-NN classifier. They evaluated their system with a database of 2,716 images and got an accuracy of 93.9%.

An approach to annotate the images by transferring the annotation from labelled images to unlabelled images is proposed by Han et al. (2005). Initially the training images are labelled manually by only one keyword (category label). Then, using this training data each classifier is trained for a particular semantic category. Then these trained classifiers are used to find probabilities for each unannotated image. After the fifth iteration of the RF the keywords are assigned to the unannotated images, according to the calculated probabilities. Thus, category keywords are assigned to the unannotated images using the semantic correlation between the annotated and unannotated images. According to authors the proposed algorithm possesses high precision as compared to

existing algorithms for image annotation. Semantic image retrieval by using keywords at different semantic levels can be achieved by using multi-level image annotation. A multi-level image annotation by using presiding image components and the relevant image semantic concept is presented by Fan et al. (2005).

3.1.2 Region-based annotation

The region-based techniques are mostly used to search objects by using local features of segmented regions. In such technique the images are described at the region or object level and the search results solely depend on the segmentation algorithm. Many approaches in the literature consider image annotation as assigning the list of words by using co-occurrence between visual features and textual descriptors (Wang et al., 2006). The success of annotation technique also depends on the model used to associate low-level features with textual descriptors. The co-occurrence model for such technique is initially proposed by Mori et al. (1999). This model associates the low-level features with the keyword based on co-occurrence knowledge. The annotation process begins by dividing the images into rectangular regions of the same size. Then for each rectangular region a feature descriptor was calculated using colour and texture features. All such descriptors are then divided into different clusters and in the last step the probability of a label for each cluster is calculated by using co-occurrence of label and rectangular region within the cluster.

An object-based approach using intermediate level descriptors (object ontology) to describe the region attributes of the image is discussed in Mezaris et al. (2003). Before inserting the image into the database, it undergoes a segmentation process to divide it into a number of regions. After that, from each region, a set of indexing features are extracted. These features are then subsequently translated to intermediate-level descriptors. When the user specifies the query in terms of the keywords, the first step in executing the query is to associate the keywords with image regions by comparing their intermediate-level descriptors. The output of this step is a set of potentially relevant images, which are presented to the user at random order for the RF. The user submits feedback on the relevancy of each image to the query. This RF is learned by SVM and finally, the top-ranked images are presented to the user.

Liu et al. (2005) presented the region-based annotations using colour and texture features. The images are first segmented into different regions and then from each region, they have defined a colour name. For colour feature, they have used the average HSV value of a region, which is then converted into semantic colour names such as light sky blue, grass green and, pale yellow. In the retrieval phase, the matching images from the database with colour names as that of the query image are selected and further ranked based on similarity metrics. The system has been tested with 5,000 Corel database images and achieved promising results.

The region-based annotation techniques such as presented in Mezaris et al. (2003), and Liu et al. (2005) may suffer from the problem of meaningful segmentation as the segmentation technique may not be able to find the important regions from the image. To overcome the above-mentioned problems, Asghar and Rao (2008) presented a technique using the adaptive medoidshift (AMS) algorithm and normalised-cut (N-cut) algorithm in combination with the SVM classifier. The overall approach helps to reduce the semantic gap problem.

3.2   *RF techniques to consider the user's view on retrieval results*

The continuous refinement of a query using user feedback is called as RF in information retrieval (Giacinto and Roli, 2004). In the last few decades, RF is used as an effective technique for query modification in information retrieval. RF is a three-step process (Shoaie and Jinni, 2010). First, the system shows matched images with the query image. Second, the relevant and irrelevant images are marked by the user. Finally, the system changes its behaviour using the feedback submitted in the second step and refines the results as per the user requirement (Shoaie and Jinni, 2010). Query shifting, feature relevance weighting, and probabilistic methods are proposed to implement the feedback loop. Query shifting methods involve shifting the query towards relevant images (Ciocca and Schettini, 1999; Wichterich et al., 2008; Traina et al., 2006). Feature relevance weighting techniques aims to update the weights of features for weighted similarity metric to integrate user's retrieval intention (Rui et al., 1998; Lu et al., 2003; Yu et al., 2007). The Bayesian method (Qian et al., 2002; Xin and Jin, 2003; Jiang et al., 2004) calculates the probability for all the relevant images in the database. This method is computationally inefficient. Some systems incorporated a combination of these techniques (Li and Yuan, 2004; Giacinto et al., 2001).

3.2.1   *RF using query point movement*

Query shifting technique is more useful in a situation where the first retrieval iteration retrieves a few relevant images (Giacinto et al., 2001). This may happen if the user has queried the database images using a query sample that is situated around the border of the relevant images. This problem can be overcome by shifting the query more towards the relevant image space. Ciocca and Schettini (1999) proposed a technique using query point movement. They proposed an algorithm to compute a new query for retrieving relevant images. The new query point is computed using the standard deviation of the relevant image features. Query point shifting using Bayesian decision theory is presented by Giacinto and Roli (2004).

CBIR system based on RF suffers from marking a large number of relevant images for real time applications. This problem is solved by Wichterich et al. (2008). They proposed a method to reshape the query by utilising user feedback. They provided an efficient approach to find the history-based region. Using the same framework, they extended the RF system to recognise that an RF approach has two stages namely finding a good query point and finding the correlation of query features with features database images.

Traina et al. (2006) proposed two RF techniques for query point movement named RF projection and RF using multiple query centre. They used a weighted similarity function to rank the database images. The authors claimed improvement in precision up to 42% in five iterations of user feedback. They also claimed that the proposed technique requires less than a second to process the queries for every feedback iteration.

3.2.2   *RF using feature re-weighting*

Feature re-weighting is useful to refine the query rather than moving the query point away (Giacinto et al., 2001). This technique uses a weighted similarity metric in which RF helps in updating the weights of features (Shoaie and Jinni, 2010). Rui et al. (1998)

proposed a feature re-weighting method with user query and user perception subjectivity. They represented images by a vector of weights, which represented the importance of components within the vectors and across the vectors. The system updates the queries by placing more weights on relevant features and less on irrelevant features. Experimentally it has been observed that the proposed approach increases the convergence speed and captures the user's need more precisely.

Most of the existing CBIR methods perform RF on low-level image features. High-level semantic and low-level features are used in an RF framework proposed by Lu et al. (2003). They have also presented a system for web image retrieval named 'iFind' based on the proposed framework. This system deals with three types of retrieval namely retrieval using keywords, retrieval using examples, and retrieval using predefined category.

A novel interactive boosting framework for semantic gap reduction is proposed by Yu et al. (2007). They achieved the improved performance of the proposed system by using RF in the boosting scheme. A new method proposed in Nezamabadi-pour and Kabir (2009) uses a combination of the visual and semantic features for CBIR. In this method, images are labelled using a fuzzy k-NN classifier. The label updating is performed using RF from the user.

3.2.3   *RF using a probabilistic (Bayesian) approaches*

The probability of relevant images to the query images can be estimated by using the Bayesian method. The estimated probability updates with iteration (Giacinto and Roli, 2004). The Bayesian method works well with a small amount of training data as compared to existing learning methods. Qian et al. (2002) used the Gaussian mixture model (GMM) for the representation of the distribution of images. They improved the performance of the CBIR system using GMM-based RF over single Gaussian model. Xin and Jin (2003) proposed an RF technique using GMM and expectation maximisation (EM) to address the user inconsistency issue. Multi-layer semantic representation (MSR) using long-term RF learning for image database is proposed by Jiang et al. (2004). Wang et al. (2003a) improved the performance of CBIR system by proposing a model named iSearch based on the EM algorithm.

Most of the classification techniques suffer from the problem of an imbalanced dataset, as the dataset consists of more number of irrelevant images as compared to the relevant images. This may cause the instability and poor retrieval results. To tackle this problem, Peng and King (2009) proposed a modified minimax probability machine called a biased minimax probability machine (BMPM), and achieved the improved performance over existing methods. Some methods available in the literature have incorporated the combination of the above three techniques of the RF (Li and Yuan, 2004; Giacinto et al., 2001) to further improve the performance of the system.

3.3   *Machine learning and deep learning techniques to associate low-level features with high-level concepts*

The recent development in the CBIR system is based on machine learning and deep learning framework. These methods address the semantic gap issue using supervised or unsupervised techniques.

3.3.1 *Supervised learning to associate low-level features with high-level concepts*

In supervised learning techniques, an image is classified using user feedback and pre-labelled images. Images classification is considered as a pre-processing step in image retrieval process and image annotation technique. SVM (Vapnik and Chervonenkis, 1974, 1998) and Bayesian classifier (Nguyen et al., 2009) are the commonly used classifiers in the pre-processing step of image retrieval. Tian et al. (2000) updated the weights of relevant images in the RF step using SVM, in turn, relaxes the user from the manual update of weights. Qi (2009) fine-tuned the retrieval results after the initial user feedback using SVM. Wang et al. (2003b) solved the problem of the imbalanced datasets by using SVM and achieved improved performance by relaxing the user from manual labelling. Tsai et al. (2003) proposed a two-stage CBIR system using SVM in the first stage followed by a fuzzy inference system in the last stage. Li et al. (2008) system based on fuzzy SVM (FSVM) for content-based semantic indexing of images. Yang et al. (2007) proposed a method using SVM as a concept detector for semantic categorisation of images. SVM with different kernel functions like fuzzy function (Spyrou et al., 2005), principle component analysis (Shi et al., 2008, and graph (Rui-zhe et al., 2009) is proposed by researchers for the betterment of the CBIR system.

3.3.2 *Unsupervised learning to classify the images into different groups*

Image clustering is one of the unsupervised learning methods used in CBIR systems for representing the high-level concepts within the images. The search space for image retrieval can be reduced by using image clustering (Liu et al., 2006). Widely used clustering algorithms for CBIR are k-means and N-cut. Han et al. (2003) proposed a memorisation learning model to bridge the semantic gap. They used an image link network and k-means algorithm. Maheshwari et al. (2009) proposed a method based on Gabor filters and k-means clustering. They compared the results of k-means clustering with hierarchical clustering. Experimentally it has been observed that the precision and recall values of hierarchical clustering are under-performing than the precision and recall values of k-means clustering. Chen et al. (2003) proposed a CBIR system named CLUster-based rEtrieval (CLUE) of images based on unsupervised learning. The resemblance between the query and target is determined using similarity metrics followed by a clustering algorithm in the retrieval step. The proposed system has experimented with 60,000 images for testing the efficacy.

Liu et al. (2006) proposed a system for a user having an interest in a particular region of an image having some semantic meaning. In this approach images are segmented into various semantic regions and each region is represented by eight feature vectors. After image segmentation, initial clustering is performed using a genetic algorithm. Further the results of clustering are improved by using the maximum flow/minimum cut theorem from graph theory. The retrieval process works at two levels, in first level candidate image segments are defined using the user-specified query. In the second level, the user gives the feedback to the initial retrieved images, which will be learned by one-class SVM to refine the retrieved results.

Table 1 Summary of existing CBIR methods

Author	Method	Features	Database	Precision (%)	MAP
Han et al. (2005)	Memory learning approach	Colour correlogram, semantic annotation, 144	Corel	78	-
Qi (2009)	Feedback based SVM classifier	Texture histogram 256	3lian.com	85.91	-
Shi et al. (2008)	Image semantic classification using SVM	Semantic features	Corel	89.13	-
Shoaie and Jinni (2010)	Semantic image retrieval reinforcement learning	-	Corel	90	-
Nezamabadi-pour and Kabir (2009)	Image retrieval using fuzzy k-NN and RF	138	Corel	90	-
Rui et al. (1998)	Image retrieval using RF	Colour histogram, colour moments, co-occurrence matrix	Corel	91.9	-
Giacinto and Roli (2004)	Bayesian relevance feedback image retrieval	Gabor features	MIT	93	-
Lu et al. (2003)	Joint semantics and feature based image retrieval	Keyword	Corel	95	-
Tzelepi and Tefas (2018)	CNN	256	Paris6k	-	0.83 (unsupervised), 0.98 (supervised)
Alzu'bi et al. (2017)	CNN	16	Oxford5K, Oxford105K	-	0.957 (Oxford5K), 0.886 (Oxford105K)
Ng et al. (2015)	CNN and VLAD	128	Paris6k, Oxford5K, Holidays	-	0.694 (Paris6k), 0.649 (Oxford5K), 0.838 (Holidays)
Yu et al. (2017)	CNN	128	Oxford5K, Holidays	-	0.615 (Oxford5K), 0.914 (Holidays)
Razavian et al. (2014)	CNN and PCA	500	Paris6k, Oxford5K, Holidays, UKbench	MOP-CNN and PCA	0.68 (Oxford5K), 0.79 (Paris6k), 0.843 (Holidays), 0.911 (UKbench)
Gong et al. (2014)	-	2,048	Holidays	-	0.802

Bhattacharya et al. (2006) proposed a framework using both supervised (probabilistic multi-class SVM) as well as unsupervised (fuzzy c-means clustering) learning-based techniques. In this method, the images are represented into semantic regions based on membership scores obtained from the learning algorithm.

3.3.3 Deep learning techniques to associate low-level features with high-level concepts

The CBIR techniques discussed in Sections 3.3.1 and 3.3.2 are mainly based on extracting visual features using traditional feature extraction methods. In the last few years, deep learning techniques attracted the attention of the industry and academic researchers and replaced the handcrafted feature extraction approach in many applications including CBIR. Tzelepi and Tefas (2018) used a deep convolutional neural network (CNN) for representing an input image with 256 features. They obtained a mean average precision (MAP) of 0.83 for unsupervised learning and 0.98 for supervised learning on the Paris6k database (Philbin et al., 2008). Alzu'bi et al. (2017) proposed the CBIR method based on CNN with 16 features and achieved a MAP of 95.7% and 88.6% on Oxford5K and Oxford105K database (Philbin et al., 2007) respectively.

Ng et al. (2015) extracted convolutional features from CNN layers and used vector locally aggregated descriptors (VLAD) for encoding into a 128-dimension feature vector. They achieved a MAP of 0.694 on Paris6k database, 0.649 on Oxford5K database, and 0.838 on Holidays database (Jegou et al., 2008). Yu et al. (2017) extracted the convolutional features from CNN layers and achieved a MAP of 0.615 and 0.914 on Oxford5K, and Holidays database, respectively. Razavian et al. (2014) used CNN to extract convolutional features of size 4,096 and reduced the dimension to 500 features using principal component analysis (PCA). They obtained MAP of 0.68, 0.79, 0.843 and 0.911 on Oxford5K, Paris6k, Holidays and UKbench database (Nister and Stewenius, 2006) respectively. Gong et al. (2014) used multi-scale orderless pooling convolutional neural network (MOP-CNN) in combination with PCA and whitening to represent the image with a feature dimension of 2,048. They obtained a MAP of 0.802 on the Holidays database. The survey is summarised in Table 1.

4   Practical applications of CBIR

CBIR technology has been used in several applications such as crime prevention, military, education, etc. The current and future applications of CBIR are discussed in this section.

1   Crime prevention

   Crime investigation agencies maintain large collections of visual evidence. During the investigation of crime, they use to check the similarity of evidences with the evidences in the database. Most of the police forces around the world are using CBIR systems like fingerprint matching and face recognition systems for the investigation of crime.

2 Fashion and interior design

CBIR finds application in fashion and interior design. Here the designer has to make use of external constraints like the choice of materials. CBIR system may be useful in retrieving a collection of fabric with different combinations of colour and texture. The system presented in Li et al. (2007) can be used to retrieve the images having a particular texture.

3 Journalism and advertising

Advertisements and advertising campaigns inherently use CBIR technology for improving their business. For example, newspaper agencies maintain a database of new articles or advertisements. Such database is generally too large and impressively expensive to maintain if detailed keyword indexing is provided. So, the CBIR is the best solution to handle such huge data and retrieve the images with increasing speed and accuracy.

4 Medical diagnosis

The main requirement of a medical imaging system is to retrieve images stored by patient name and to retrieve similar previous cases. A physician may need to retrieve images with precise information on patients with certain anomalies. In such cases the required images can be retrieved by submitting the query like – 'Retrieve all lung X-rays of Mr. David where disjoint lesions are found'.

5 The military

The important requirement of military forces is to identify the possible threats it may be called upon to face. The CBIR system is useful in searching the images in the database with a particular target image.

6 Remote sensing and geographical information system

With the development of remote sensing techniques a large number of digital images are being generated everyday. Information retrieval from remote sensing images has numerous real-life applications (Manjunath and Ma, 1996; Zhixiao et al., 2008) in the field of business and research. For example, retrieving the land and its types such as agricultural land or non-agricultural land from the remote sensing geographic images. Eklund and Kirkby (1998) classified the soil types from the remote sensing geographic images using inductive-learning techniques and artificial neural networks.

7 Historical research

Researchers from various disciplines such as art, sociology use visual information to carry out their research activities. For example, Archaeologists may need visual comparisons of their archaeological findings at some stage of their research. Access to the visual content of a museum or an art gallery might prove their research hypothesis. Thus, the system having the ability to retrieve the objects sharing some visual similarity can be helpful to historical researchers.

8   Education and training

CBIR technology can be used to search the good teaching material on a particular topic from a large database. This will benefit in teaching and learning process. Hence the CBIR technology can be used to augment the educational quality.

9   Home entertainment

Much of home entertainment data consist of images or videos collected on different occasions. Due to the advancements in digital cameras, individuals' photos or videos are being accumulated in the personal database in huge quantities. However, peoples do not arrange them in semantically meaningful groups. Therefore, dividing the images into different meaningful classes automatically is an essential application for a general image retrieval system. The method proposed in Yang et al. (2007) is useful to classify the general home photos in the semantic groups so that large databases will be divided into meaningful small groups. Thus, the CBIR can be used for the management of such a home entertainment collection.

10   Web searching

Internet is becoming the most popular source of knowledge and entertainment. And for this several search engines are being used to locate the required information from the web. Some of the commercial search engines such as Yahoo! Image Surfer (http://isurf.yahoo.com) and AltaVista's AV Photo Finder (http://image.altavista.com/cgi-bin/avncgi) are based on CBIR technology.

5   Discussion and future research directions

With many potential multimedia applications, CBIR is widely used for image management and web search. To address the problem of the semantic gap most of the research efforts has been made to develop the CBIR systems which support high-level query. But still, there are some problems open for researchers to put effort. To provide future research directions for the emerging researchers in the field of CBIR, emerging challenges and open issues are listed and discussed in the following section.

1   The researchers are unable to define the semantic meanings of an image specifically. Only they are able to classify the images in certain groups. But, designing a CBIR system capable of image understanding and efficient retrieval from larger databases is still a challenging task.

2   The image-based annotation method faces difficulty in identifying the local visual features within the image in turn fails to define the salient semantic features.

3   The region-based annotation method faces difficulty with the underlying image segmentation algorithm. Thus, there is a need to design a semantically balanced annotation process by considering both local and global features of the image.

4   A semi-automatic annotation system can be introduced for improving the retrieval accuracy by combining the manual training with feedback to intelligently annotate multimedia.

5   There exists scope for further improvement in the RF performance. As pointed in Xin and Jin (2003), the research issue on which less attention is given is user inconsistency during the feedback process. If the user is not consistent in providing the feedback, there will be degradation in retrieval results. Also, the convergence speed of RF is one of the urgent problems that need to be solved.

6   A new learning scheme for CBIR is still in demand. Learning-based image retrieval faces difficulty because of the small sample size. In some cases, the size of the training data is smaller as compared to the feature dimension. This affects the performance of classifiers like SVM classifiers used for multi-class classification. A solution to this problem is to increase the training sample size. But this solution suffers from manual labelling of data and computational cost significantly. Also, there is a need for standard methods to evaluate and compare different active learning algorithms.

7   Because of the large size of images and feature vectors, high dimensionality is one of the key issues in CBIR. Hence, efficient dimensionality reduction techniques are required to improve the performance of CBIR task.

8   The retrieval speed and accuracy of existing CBIR systems are not up to the mark. Hence, more efforts can be put in the direction to improve retrieval speed and accuracy.

6   Conclusions

This paper presents a review of existing CBIR systems, especially CBIR systems based on reduction of semantic gap. The existing CBIR methods are divided into three categories namely image annotation techniques, RF techniques, and machine learning and deep learning techniques. Based on the architectures of current CBIR systems, the general architecture of the semantic-based image retrieval system has been proposed in this paper. The comparison of existing CBIR systems based on features, classifiers, database, and performance metrics is provided as a summary at a glance. Some current and future applications of CBIR are also discussed in this paper.

   The challenges faced and unsolved issues by the researchers in the field of CBIR are listed and discussed in the survey. This provides the motivation and future research directions for emerging researchers working on the design and development of efficient CBIR systems. Each technique discussed in this paper has its own advantages and disadvantages. The techniques which contribute to increasing the acceptance of CBIR systems in real applications are in demand. Still, the use of the CBIR system is very limited, specifically in image-based search engines. Hence, there exists a wide scope for putting efforts in proposing novel and efficient CBIR systems for practical applications.

References

Alzu'bi, A., Amira, A. and Ramzan, N. (2017) 'Content-based image retrieval with compact deep convolutional features', *Neurocomputing*, Vol. 249, No. C, pp.95–105.

Asghar, A. and Rao, N. (2008) 'Semantics sensitive segmentation and annotation of natural images', *IEEE International Conference on Signal Image Technology and Internet Based Systems*, pp.387–394.

Bhattacharya, P., Rahman, M. and Desai B. (2006) 'Image representation and retrieval using support vector machine and fuzzy C-means clustering based semantical spaces', *Eighteenth International Conference on Pattern Recognition (ICPR'06)*, Vol. 1, pp.929–935.

Chen, Y., Wang, J., and Krovetz, R. (2003) 'An unsupervised learning approach to contentbased image retrieval', *2003 Seventh International Symposium on Signal Processing and Its Applications*, Vol. 1, pp.197–200.

Ciocca, G. and Schettini, R. (1999) 'A relevance feedback mechanism for content-based image retrieval', *Information Processing & Management*, Vol. 35, No. 5, pp.605–632.

Djordjevic, D. and Izquierdo, E. (2007) 'An object-and user-driven system for semanticbased image annotation and retrieval', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 3, pp.313–323.

Eklund, P. and Kirkby, S. (1998) 'Data mining and soil salinity analysis', *International Journal of Geographical Information Science*, Vol. 12, No. 6, pp.247–268.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D. and Equitz, W. (1994) 'Efficient and effective querying by image content', *Journal of Intelligent Information Systems*, Vol. 3, Nos. 3–4, pp.231–262.

Fan, J., Gao, Y. and Luo, H. (2008) 'Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation', *IEEE Transactions on Image Processing*, Vol. 17, No. 3, pp.407–426.

Fan, J., Gao, Y., Luo, H. and Xu, G. (2005) 'Statistical modeling and conceptualization of natural images', *Pattern Recognition*, Vol. 38, No. 6, pp.865–885.

Giacinto, G. and Roli, F. (2004) 'Bayesian relevance feedback for content-based image retrieval', *Pattern Recognition*, Vol. 37, No. 7, pp.1499–1508.

Giacinto, G., Roli, F. and Fumera, G. (2001) 'Comparison and combination of adaptive query shifting and feature relevance learning for content-based image retrieval', *Eleventh International Conference on Image Analysis and Processing*, Vol. 2, pp.422–427.

Gong, Y., Wang, L., Guo, R. and Lazebnik, S. (2014) 'Multi-scale orderless pooling of deep convolutional activation features', *Computer Vision – ECCV 2014*, pp.392–407.

Gupta, A. and Jain, R.(1997) 'Visual information retrieval', *Communications of the ACM*, Vol. 40, No. 5, pp.70–79.

Han, J., Li, M., Zhang, H. and Guo, L. (2003) 'A memorization learning model for image retrieval', *2003 International Conference on Image Processing*, Cat. No. 03CH37429, Vol. 3, pp.280–283.

Han, J., Ngan, K., Li, M. and Zhang, H. (2005) 'A memory learning framework for effective image retrieval', *IEEE Transactions on Image Processing*, Vol.14, No. 4, pp.511–524.

Jegou, H., Douze, M. and Schmid, C. (2008) 'Hamming embedding and weak geometric consistency for large scale image search', *Lecture Notes in Computer Science*, pp.304–317.

Jiang, L., Hou, J., Chen, Z. and Zhang, D. (2009) 'Automatic image annotation based on decision tree machine learning', *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp.170–175.

Jiang, W., Er, G. and Dai, Q. (2004) 'Multilayer semantic representation learning for image retrieval', *2004 International Conference on Image Processing, ICIP'04*, Vol. 4, pp.2215–2218.

Kokare, M., Chatterji, B. and Biswas, P. (2002) 'A survey on current content based image retrieval methods', *IETE Journal of Research*, Vol. 48, Nos. 3–4, pp.261–271.

Lakdashti, A., Moin, S. and Badie, K. (2008) 'Semantic-based image retrieval: a fuzzy modeling approach', *IEEE/ACS International Conference on Computer Systems and Applications*, pp.575–581.

Li, B. and Yuan, S. (2004) 'A novel relevance feedback method in content-based image retrieval', *International Conference on Information Technology: Coding and Computing*, Vol. 2, pp.120–123.

Li, J., Huang, S., He, R. and Qian, K. (2008) 'Content-based semantic indexing of image using fuzzy support vector machines', *2008 Chinese Conference on Pattern Recognition*, pp.1–6.

Li, Q., Shi, Z. and Luo, S. (2007) 'A neural network approach for bridging the semantic gap in texture image retrieval', *2007 International Joint Conference on Neural Networks*, pp.581–585.

Liu, Y., Chen, X., Zhang, C. and Sprague, A. (2006) 'An interactive region-based image clustering and retrieval platform', *2006 IEEE International Conference on Multimedia and Expo*, pp.929–932.

Liu, Y., Zhang, D., Lu, G. and Ma, W. (2005) 'Region-based image retrieval with high-level semantic color names', *Eleventh International Multimedia Modelling Conference*, Vol. 2, pp.180–187.

Liu, Y., Zhang, D., Lu, G. and Ma, W. (2007) 'A survey of content-based image retrieval with high-level semantics', *Pattern Recognition*, Vol. 40, No. 1, pp.262–282.

Lu, J., Ma, S. and Zhang, M. (2005) 'Automatic image annotation based-on model space', *International Conference on Natural Language Processing and Knowledge Engineering*, pp.455–460.

Lu, Y., Zhang, H., Wenyin, L. and Hu, C. (2003) 'Joint semantics and feature based image retrieval using relevance feedback', *IEEE Transactions on Multimedia*, Vol. 5, No. 3, pp.339–347.

Maheshwari, M., Silakari, S. and Motwani, M. (2009) 'Image clustering using color and texture', *2009 First International Conference on Computational Intelligence, Communication Systems and Networks*, pp.403–408.

Manjunath, B. and Ma, W. (1996) 'Texture features for browsing and retrieval of image data', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp.837–842.

Mezaris, V., Kompatsiaris, I. and Strintzis, M. (2003) 'An ontology approach to object-based image retrieval', *International Conference on Image Processing*, Vol.2, pp.2–511.

Mori, Y., Takahashi, H. and Oka, R. (1999) 'Image-to-word transformation based on dividing and vector quantizing images with words', *First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, pp.1–9.

Nezamabadi-pour, H. and Kabir, E. (2009) 'Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval', *Expert Systems with Applications*, Vol. 36, No. 3, pp.5948–5954.

Ng, J., Yang, F. and Davis, L. (2015) 'Exploiting local features from deep networks for image retrieval', *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.53–61.

Nguyen, L., Yap, G., Liu, Y., Tan, A., Chia, L. and Lim, J. (2009) 'A Bayesian approach integrating regional and global features for image semantic learning', *IEEE International Conference on Multimedia and Expo*, pp.546–549.

Nister, D. and Stewenius, H. (2006) 'Scalable recognition with a vocabulary tree', *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.2161–2168.

Oussalah, M. (2008) 'Content based image retrieval: review of state of art and future directions', *First Workshops on Image Processing Theory, Tools and Applications*, pp.1–10.

Peng, X. and King, I. (2009) 'A biased minimax probability machine-based scheme for relevance feedback in image retrieval', *Neurocomputing*, Vol. 72, Nos. 7–9, pp.2046–2051.

Pentland, A., Picard, R. and Sclaroff, S. (1996) 'Photobook: content-based manipulation of image databases', *International Journal of Computer Vision*, Vol. 18, No. 3, pp.233–254.

Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A. (2008) 'Lost in quantization: improving particular object retrieval in large scale image databases', *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8.

Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A. (2007) 'Object retrieval with large vocabularies and fast spatial matching', *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8.

Qi, Y. (2009) 'Relevance feedback based on texture histogram and SVM', *2009 International Workshop on Intelligent Systems and Applications*, pp.1–4.

Qian, F., Li, M., Zhang, L., Zhang, H. and Zhang, B. (2002) 'Gaussian mixture model for relevance feedback in image retrieval', *IEEE International Conference on Multimedia and Expo*, Vol. 1, pp.292–232.

Razavian, A., Azizpour, H., Sullivan, J. and Carlsson, S. (2014) 'CNN features off-the-shelf: an astounding baseline for recognition', *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.512–519.

Rui, Y., Huang, T., Ortega, M. and Mehrotra, S. (1998) 'Relevance feedback: a power tool for interactive content-based image retrieval', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp.644–655.

Rui-zhe, Z., Jia-zheng, Y., Jing-hua, H., Yu-jian, W. and Hong, B. (2009) 'A novel generalized SVM algorithm with application to region-based image retrieval', *2009 International Forum on Information Technology and Applications*, Vol. 2, pp.280–283.

Shi, L., Gu, G., Liu, H. and Shen, J. (2008) 'Image semantic classification algorithm research on kernel PCA support vector machine', *2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, pp.422–424.

Shoaie, Z. and Jinni, S. (2010) 'Semantic image retrieval using relevance feedback and reinforcement learning algorithm', *Fifth International Symposium on I/V Communications and Mobile Network*, pp.1–4.

Smeulders, A., Worring, M., Santini, S., Gupta, A. and Jain, R. (2000) 'Content-based image retrieval at the end of the early years', *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 22, No. 12, pp.1349–1380.

Spyrou, E., Stamou, G., Avrithis, Y. and Kollias, S. (2005) 'Fuzzy support vector machines for image classification fusing MPEG-7 visual descriptors', *The Second European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 2005*, Ref. No. 2005/11099, pp.23–30.

Tian, Q., Hong, P. and Huang, T. (2000) 'Update relevant image weights for content-based image retrieval using support vector machines', *2000 IEEE International Conference on Multimedia and Expo, ICME2000, Proceedings on Latest Advances in the Fast Changing World of Multimedia*, Cat. No. 00TH8532, Vol. 2, pp.1199–1202.

Traina, A., Marques, J. and Traina, C. (2006) 'Fighting the semantic gap on CBIR systems through new relevance feedback techniques', *9th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)*, pp.881–886.

Tsai, C., McGarry, K. and Tait, J. (2003) 'Using neuro-fuzzy techniques based on a twostage mapping model for concept-based image database indexing', *Fifth International Symposium on Multimedia Software Engineering 2003*, pp.6–12.

Tzelepi, M. and Tefas, A. (2018) 'Deep convolutional learning for content based image retrieval', *Neurocomputing*, Vol. 275, No. C, pp.2467–2478.

Vailaya, A., Jain, A. and Zhang, H. (1998) 'On image classification: city vs. landscape', *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp.3–8.

Vapnik, V. and Chervonenkis, A. (1974) *Theory of Pattern Recognition*, Nauka, Moscow.

Vapnik, V. and Chervonenkis, A. (1998) *Statistical Learning Theory*, Wiley New York.

Wang, H., Ooi, B. and Tung, A. (2003a) 'iSearch: mining retrieval history for content-based image retrieval', *Eighth International Conference on Database Systems for Advanced Applications, 2003(DASFAA 2003)*, pp.275–282.

Wang, L., Chan, K. and Zhang, Z. (2003b) 'Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval', *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp.629–634.

Wang, Z., Feng, D., Chi, Z. and Xia, T. (2006) 'Annotating image regions using spatial context', *IEEE International Symposium on Multimedia (ISM'06)*, pp.55–61.

Wichterich, M., Beecks, C. and Seidl, T. (2008) 'Ranking multimedia databases via relevance feedback with history and foresight support', *IEEE International Conference on Data Engineering Workshop*, pp.596–599.

Xin, J. and Jin, J. (2003) 'Learning from user feedback for image retrieval', *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia, Proceedings of the 2003 Joint*, Vol. 3, pp.1792–1795.

Xuelong, H., Yuhui, Z. and Li, Y. (2007) 'A new method for semi-automatic image annotation', *Eight International Conference on Electronic Measurement and Instruments*, pp.2-866–2-869.

Yang, S., Kim, S., Seo, K. Ro, Y., Kim, J. and Seo, Y. (2007) 'Semantic categorization of digital home photo using photographic region templates', *Information Processing & Management*, Vol. 43, No. 2, pp.503–514.

Yu, J., Lu, Y., Xu, Y., Sebe, N. and Tian, Q. (2007) 'Integrating relevance feedback in boosting for content-based image retrieval', *IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 1, pp.I-965–I-968.

Yu, W., Yang, K., Yao, H., Sun, X. and Xu, P. (2017) 'Exploiting the complementary strengths of multi-layer CNN features for image retrieval', *Neurocomputing*, Vol. 237, No. C, pp.235–241.

Zhixiao, X., Charles, R. and Brian, J. (2008) 'Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian Pine in south Florida', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 63, No. 6, pp.647–660.