Revising Knowledge Discovery for Object Representation with Spatio-Semantic Feature Integration

Madhuri B. Dhas
PG Student
Department Of Computer Engineering
VPCOE, Baramati, Savitribai Phule University
Pune, India
madhuri.dhas13@gmail.com

Prof. S. A. Shinde
Assistant Professor
Department Of Computer Engineering
VPCOE, Baramati, Savitribai Phule University
Pune, India
shinde.meetsan@gmail.com

Abstract— In large social networks, web objects become increasingly popular. Multimedia object classification and representation is a necessary step of multimedia information retrieval. Indexing and organizing these web objects for the purpose of convenient browsing and search of the objects, and to effectively reveal interesting patterns from the objects. For all these tasks, classifying the web objects into manipulable semantic categories is an essential procedure. One important issue for classification of objects is the representation of images. To perform supervised classification tasks, the knowledge is extracted from unlabeled objects through unsupervised learning. In order to represent the images in a more meaningful and effective way rather than using the basic Bag-of-words (BoW) model, a novel image representation model called Bag-of-visual phrases(BoP) is used. In this model visual words are obtained using hierarchical clustering and visual phrases are generated by vector classifier of visual words. To obtain the Spatio-semantic correlation knowledge the frequently co-occurring pairs are calculated from visual vocabulary. After the successful object representation, the tags, comments, and descriptions of web objects are separated by using most likelihood method. The spatial and semantic differentiation power of image features can be enhanced via this BoP model and likelihood method**.**

Keywords*- Knowledge Discovery, Bag-of-Visual Phrases Model, Correlation Knowledge, Spatio- Semantic Feature Integration ;*

*****

I. INTRODUCTION

The social media repositories like Flicker[1], and Zoomr allow their users to interpret their images with their own tags chosen by users and these tags are used as indexing keywords for faster image search and other applications [13].Billions of photos are uploaded to and collected by Flickr (www.flickr.com) and Facebook (www.facebook.com), millions of videos are being uploaded to YouTube (www.youtube.com); millions of products are being sold on Amazon (www.amazon.com);millions of research papers are referenced on CiteULike (www.citeulike.com).

For these increased multimedia objects, web users provide the textual information about these web objects with tags, comments and description of that objects, but true labels of web images for classification are difficult to obtain and expensive also. Traditional learning techniques[9]have the assumption that training and test data are drawn from the same data distribution and that's why they are not suitable for dealing with the situation where new unlabeled data are obtained from fast evolving , related but different information sources. Number of applications of supervised learning requires good generalization from the limited labeled data. For all these tasks, classification of these web objects into manipulable semantic categories is an essential preprocess for browsing, searching, indexing, and mining these web object. Advances in computer and multimedia technologies allow for the production of images and large repositories for image storage with little cost.

Due to the explosive growth of heterogeneous web objects [13] especially non-textual objects such as products, pictures, and videos, have made the problem of web classification [1][14] increasingly challenging. Such objects often suffer from lack of easy-extractable features with semantic information, interconnections between each other, as well as training examples with category labels. The main goal is to create, manage, and query image databases in an efficient and effective way that is in accurate manner.

The remainder of this paper is organized as follows: Section II describes old methods for Multimedia Object Representation and Classification. Section III describes implementation details of proposed method. Section IV describes datasets required and Section V describes conclusion.

II. RELATED WORK

There have been some works attempting to seek visual word combinations to capture the spatial information among visual words. Previously used BoW [1] [2][10] model is useful for image domain but containing very limited semantic information.

[1] http://www.flicker.com/

Each visual word is represented by region/patches which are come from different parts of multimedia objects. To deal with semantic information among visual words, number of methods has been proposed like frequently co-occurring patterns[10], frequent adjacent regions/patches pairs, defining meaningful phrases[1][10][11].Using BoW model with more visual words improves the system performance but when the number of visual words reaches specific level then performance degrades.

Existing studies have focused on representing web multimedia objects and classify them for indexing, browsing and searching purpose[14].Classic supervised learning is suitable only for labeled data samples but fails where the true labels of the images are missing and difficult to obtain. It integrates different types of features from different information sources for classification purpose. Feature Integration [3], Kernel Integration, and Semantic Integration these three methods of multi-view learning [3] works at three levels. Feature integration works at Feature level. In feature level, feature extraction and representation is more informative but difficulty is as the data dimension increases, its learning complexity increases. In semantic integration, results on different feature space are combined at semantic level but the proper correlation structure between different features is somewhat difficult to obtain [3].

Classification can be performed with a heterogeneous transfer learning framework [7] for knowledge transfer between text and images. Some annotated images can be found on many social Web sites, a target-domain classification problem occurs. [13] Proposed the method of web object classification as an optimization problem on a graph of objects and tags. An efficient algorithm is used to enrich the semantic features for the objects and classify the web objects into different categories of unlabeled objects from both homogeneous and heterogeneous labeled objects, through the implicit connection of social tags.

To improve the classification accuracy by considering both labeled and labeled data, co-updating method [4] is used previously which mainly focus on unlabeled data classification. A new method self –taught learning [5]is used to handle large number of unlabeled images (or audio samples, or text documents) randomly downloaded from the Internet to improve performance on a given image (or audio, or text) classification task. A self-taught learning approach uses sparse coding to construct higher level features using the unlabeled data. Largely unsupervised learning algorithms were presented for improving performance on supervised classification tasks.

Self-taught learning process consists of two stages: First learn a representation using only unlabeled data and then apply this representation to the labeled data, and use it for the classification task involving a Self-taught learning algorithm. This large unlabeled data is easily handle by Self-taught learning known as Transfer learning but this unlabeled data cannot be assigned to the supervised learning task's class labels.

Multimodal Integration contributes the combination of multiple classifiers for the semantic fusion process [3]. Multimodal systems which uses semantic fusion including semantic recognizers and a sequential integration process. These systems integrate at the feature level and at a semantic level. By evaluating the multimodal recognition probabilities, identification of the factors that affects the multimodal recognition performance was carried out.

Our classification and representation method find out the correlation between different information sources as knowledge and applicable to new learning tasks. This method involves both the image domain and text domain with which deals with not only the Spatio information but also semantic information and integrated features of it.

III. PROPOSED SYSTEM

The proposed framework is organized into four main steps:

(i) Data Preprocessing
(ii) Feature Extraction
(iii) Creating visual vocabulary Model (BoP Model)
(iv) Creating textual vocabulary model using vector classifier
(v) Spatio -Semantic Object Representation Model using Most Likelihood method

The system architecture is as shown in fig.1. Flicker Dataset is used for this system in which collection of web images with their textual descriptions is given. Collected Flicker dataset is passed through the Feature Extraction phase to generate vocabulary tree and Feature Integration phase to find out the frequently co-occurring pairs.

A. Data Preprocessing

Flicker dataset contains web images with their corresponding textual descriptions. This data should be passed to the feature extraction phase with some preprocessing. Three operations are performed on the text domain of this dataset. They are

1. Stopwords Removal
2. Nonwords Removal
3. Stemming

B. Feature Extraction

As Flicker dataset contains the web images with their corresponding textual description, we have to extract the features of the images that are present in our dataset. We are having I as a set of images and D as a set of textual descriptions. For image the SIFT (Scale Invariant Feature Transform)[17] is used to extract the features from image and representing each image as a Bag-of-Words (BoW) which gives limited semantic information. For text, the tf-idf[15] is used to extract the features from the corresponding textual descriptions that are assigned to each image.

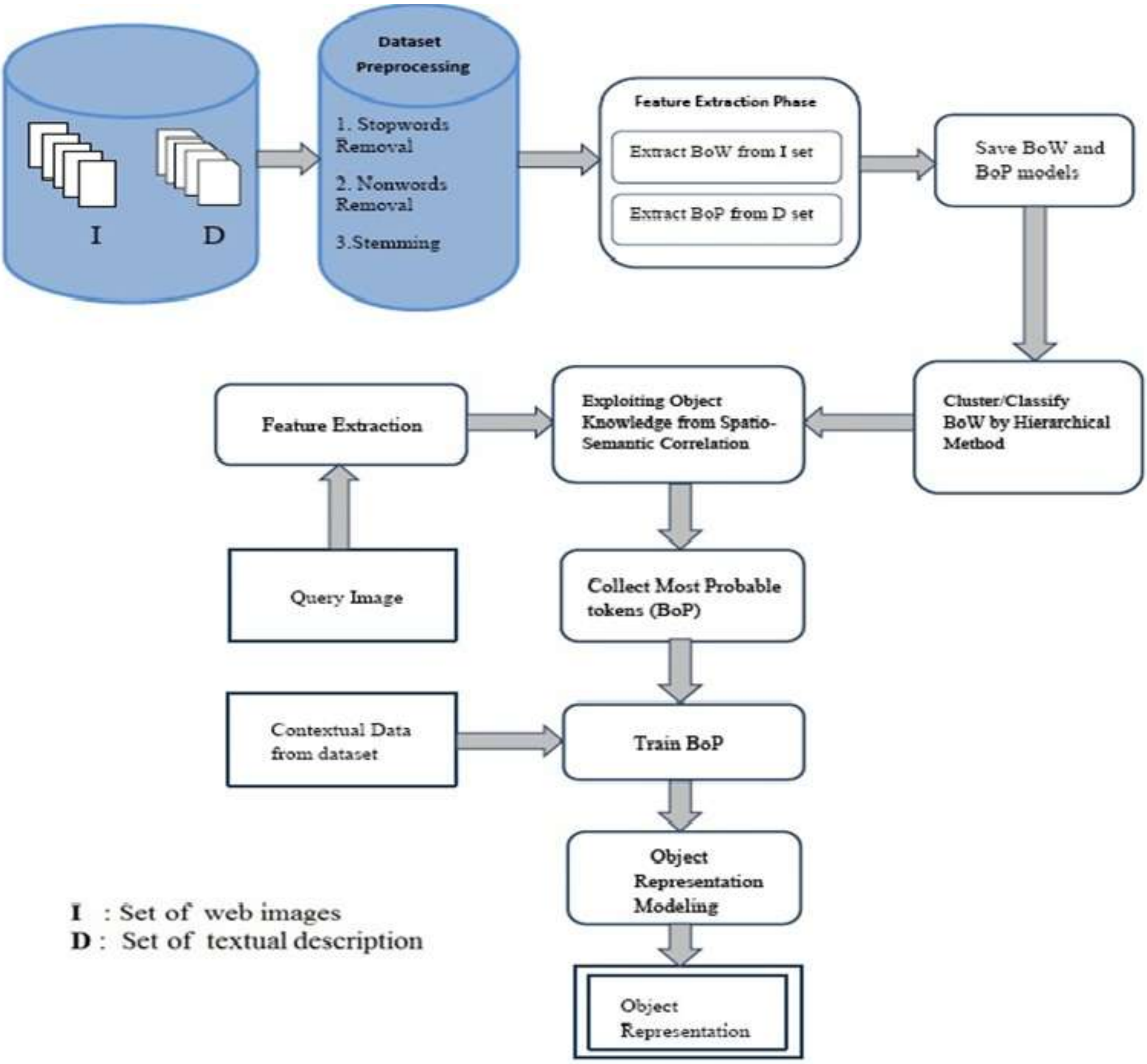


Fig: Flow of System

Given the collections of all such terms called textual words, T = {t_1,t_2,..,t_w} as each textual descriptor, Vi, is represented as a vector of terms(text) along with their weights, Vi= {$f_{i,1}$,$f_{i,2}$,..,$f_{i,w}$} where fij is a term(textual word) weight representing the number of occurrences of term(textual word) tj in the textual descriptor fi. These term weights are then transformed using a standard (tf-idf) approach [15].

C. Creating Visual Vocabulary Model

After Feature Extraction we get visual and textual words. Each feature extracted provides us the description of patch/region. From that extracted features vocabulary of visual words is generated using Hierarchical clustering. SIFT feature extraction provides us all the keypoints that are extracted from dataset images.

These Keypoints are grouped together in such a way that any keypoint that belongs to a specific cluster that should not be assigned to another cluster. To find out the semantic relationship between visual words SIFT features along with color histogram is applied and from that Vocabulary dictionary of visual words is generated which shows most matching cluster for finding the correlation. Here we get the tag and appropriate title for the given object.

Output of this vocabulary is forwarded to vector quantization mechanism to generate the visual vector in which each image is represented by words and phrases.

$$\mathit{AvgSim}\ (F_i , C_k) = \sum \mathrm{Sim}(F_i , F_j) / | F_j - F_i | \ldots\ldots\ldots(1)$$

and

$$\exists\ F \mid F\ \mathrm{AvgSim}\ (F_i , C_k) \geq \mathrm{AvgSim}(F_i , C_j)\ \ldots\ldots\ldots(2)$$

Where,

F = set of features
C = set of clusters

D. Creating Textual Vocabulary Model using vector classifier

Knowledge Based method find out the correlation knowledge for classification purpose from one feature space to another feature space. We are dealing with both image domain and text domain. Basic BoW model is ignoring some spatial information and lacking of semantic relationship information. Our method discovers the semantic correlation between different information sources as knowledge. Then calculating a set of strongly correlated groups (frequent tokens from dictionary) fig. 2 to form a Spatio-Semantic correlation knowledge.

E. Spatio-Semantic Object Representation Model using Most Likelihood method

This framework takes the user query image and extracts the features from that query image. After query image feature extraction it exploits object knowledge from Spatio-Semantic Correlation Knowledge which is formed by vocabulary of visual components and textual words. The object representation model is based on the consistency between “visual similarity” and “semantic similarity” in social images where visually similar images frequently behave in a particular way to have similar semantic descriptors and vice versa.

From the dictionary of frequent tokens vector classifier is applied to train the web object. Then collecting all the sentences from given textual descriptions and finding out the probability score of each sentence.According to that probability score sorting the sentences using most likelihood method to get proper textual description for the query image. These frequently co-occurring tokens are then used to construct the correlation with which the Spatio-Semantic Correlation Knowledge is mined.

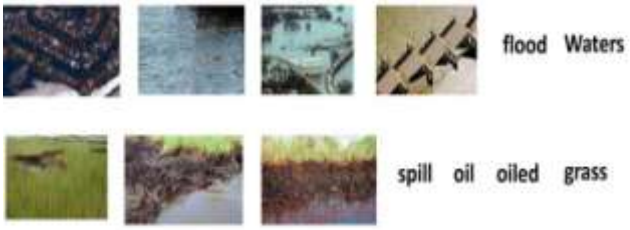


Fig 2: Correlated groups

A new technique called most likelihood method is used to separate out the tags, comments, and textual descriptions of that web objects after representing it with Spatio-Semantic Feature Integration.

IV. DATASET

The new method can be evaluated in the context of different real world datasets for web object classification and representation collected from Flicker[1], CNN news[2], Google+[3], Twitter[4], LinkedIn[5]. Flicker dataset is used where social media users assign the textual descriptions to the images for image retrieval and indexing purpose.

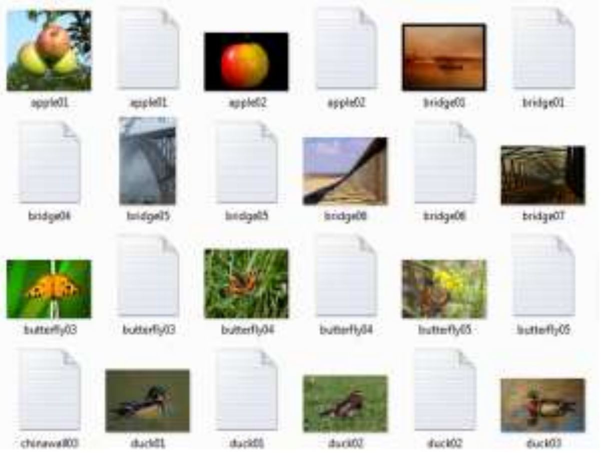


Fig : Web images and their corresponding textual description

[1] http://www.flicker.com/

[2] http://www.cnn.com/

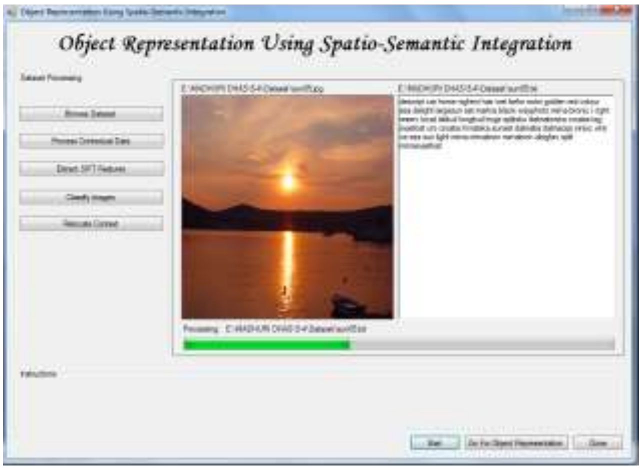
[3] http://plus.google.com/

[4] http://twitter.com/

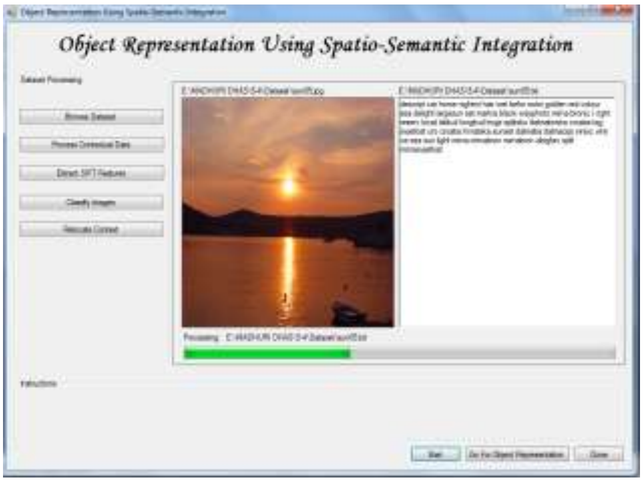
[5] http://www. Linkdedin.com/

V. Results

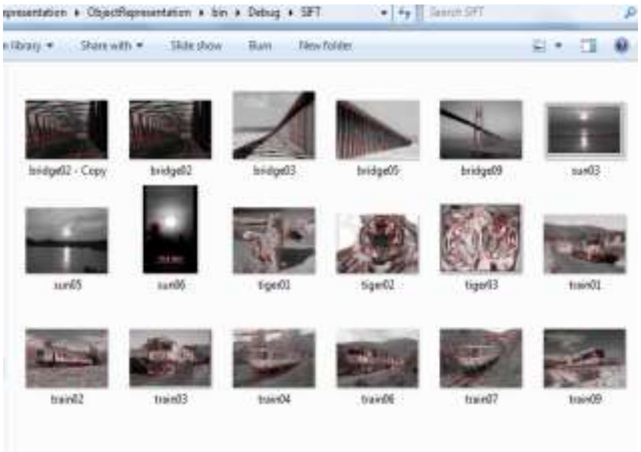
1. Dataset Preprocessing



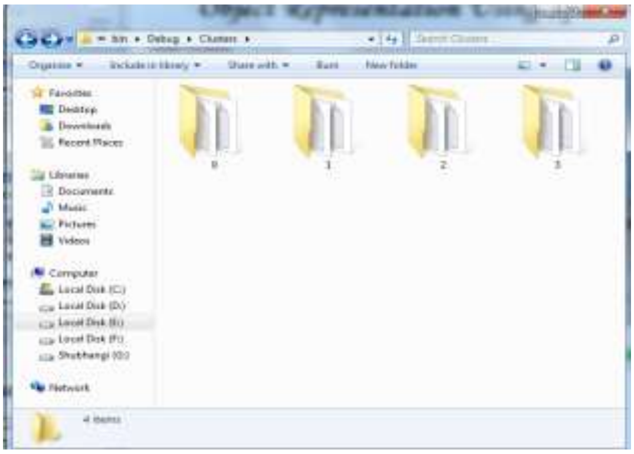
2. SIFT feature extraction



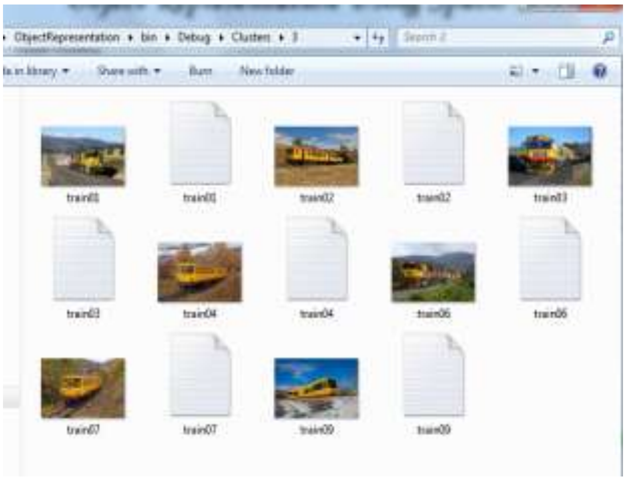
3. SIFT features saved



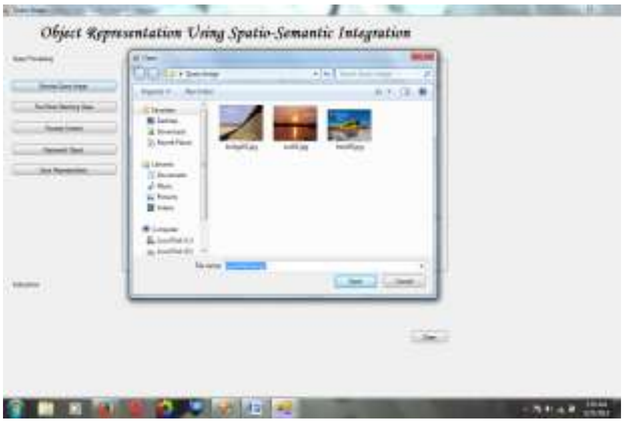
4. Classification using Hierarchical clustering



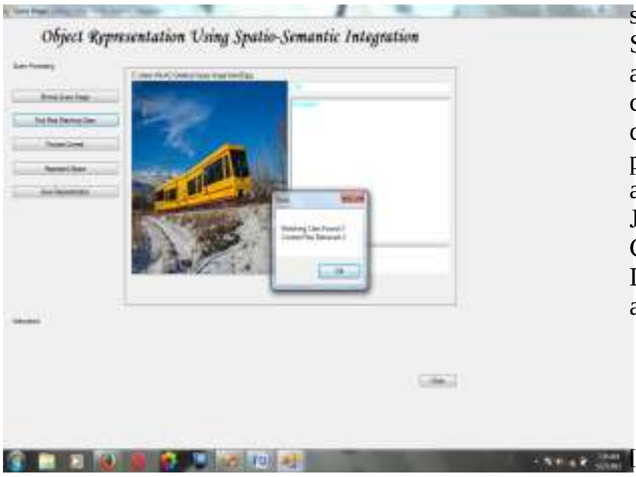
5. Classified Images



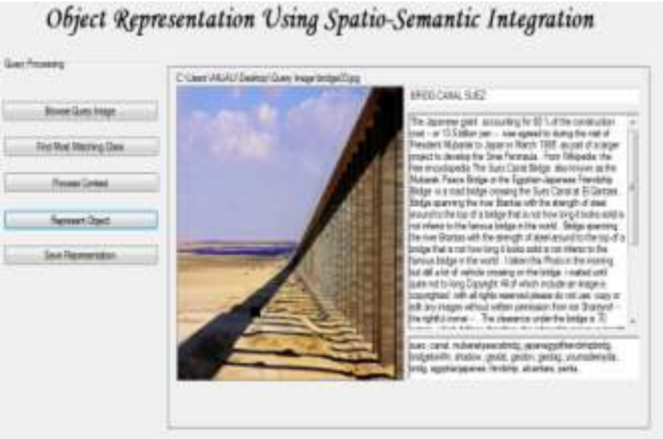
6. Query image



7. Most matching class found for given query image



8. Object Representation



VI. CONCLUSION

The new approach provides both the Spatio-Semantic information with which web multimedia object classification and representation performance get increased. This framework exploits the correlation knowledge among the visual components and textual words (visual similarity and textual similarity) which is useful in browsing, searching, indexing the web multimedia objects. Applied BoP model with Most Likelihood method on a large-scale Flicker dataset shows that this framework greatly outperforms.

ACKNOWLEDGMENT

This is to acknowledge and thank all the individuals who played defining role in shaping this paper. Without their coordination, guidance and reviewing this task could not be completed alone. I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide Prof. S. A. Shinde sir for giving his valuable guidance, inspiration and encouragement to embark this paper. This project being conceptual one needed a lot of support from my guide so that I could achieve what I was set out to get. I also take great pleasure in thanking Prof. S. S. Nandgaonkar for providing appropriate guidance. I would personally like to thank Prof. G. J. Chhajed, Head of Computer Dept. at Vidya Pratishthan's Collage of Engineering (VPCOE), and our Hon'ble principal Dr. S. B. Deosarkar sir who creates a healthy environment for all of us to learn in best possible way.

REFERENCES

[1] Wenting Lu, Jingxuan Li, Tao Li, Weidong Guo, Honggang Zhang, and Jun Guo, "Web Multimedia Object Classification Using Cross-Domain Correlation Knowledge," *IEEE transactions on multimedia*, vol. 15, No. 8, December 2013.

[2] Madhuri Dhas, Prof. S. A. Shinde "Survey of Knowlwdge Discovery for Object Representation using Spatio-Semantic Feature Integration," *International Journal of Engineering Research and General Science* Volume 2, Issue 6, October-November, 2014 ,ISSN 2091-2730.

[3] L. Wu, S. Oviatt, and P. Cohen, "Multimodal integration-a statistical view," *IEEE Trans. Multimedia*, vol. 1, no. 4, pp. 334–341, 1999.

[4] L. Wu, S. Oviatt, and P. Cohen, "Multimodal integration-a statistical view," *IEEE Trans. Multimedia*, vol. 1, no. 4, pp. 334–341, 1999.

[5] T. Li and M. Ogihara, "Semisupervised learning from different information sources," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 289309, 2005.

[6] R. Raina, A. Battle, H. Lee, B. Packer, and A. Ng, "Self-taught learning: Transfer learning from unlabeled data," in Proc. ICML, 2007, pp. 759766.

[7] Wei Jiang, Eric Zavesky, Shih-Fu Chang,Alex Loui, "Cross-Domain learning methods for high-level visual concept classification," *IEEE* ICIP-2008.

[8] Y. Zhu, Y. Chen, Z. Lu, S. Pan, G.Xue, Y.Yu, and Q.Yang, "Heterogeneous transfer learning for image classification," in Proc. AAAI, 2011.

[9] S. Roy, T. Mei, W. Zeng, and S. Li, "Socialtransfer: Cross-domain transfer learning from social streams for media applications," in *Proc.* pp.649658 *ACM Multimedia*, 2012.

[10] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann, 2006.

[11] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, "Descriptive visual words and visual phrases for image applications," in *Proc*. pp. 7584, *ACM Multimedia*, 2009.

[12] M. Sadeghi and A. Farhadi, "Recognition using visual phrases," in *Proc. IEEE CVPR*, pp. 1745–1752, 2011.

[13] J. Hare and P. Lewis, "Automatically annotating themir flickr dataset," in *Proc. Multimedia Information Retrieval*, 2010.'

[14] Z. Yin, R. Li, Q.Mei, and J. Han, "Exploring social tagging graph for web object classification," in *Proc. ACM SIGKDD*, 2009, pp. 957–966.

[15] Pedro R.Kalva, Fabricio Enembreck, and Alessandro L. Koerich, "Web image classification based on th fusion of image and text classification," In ICDAR, pages 561-568,2007.

[16] C.D. Manning, P. Raghavan, and H. Schutze, Introduction to Information Retrieval. Cambridge Univ. Press, 2008.

[17] C. Bishop, *Pattern Recognition and Machine Learning*. NewYork, NY, USA: Springer, 2006.

[18] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.