Relative Image Searching With Image As An Input

Atharva Bachha[1], Gaurav Malge[2], Onkar Nirhali[3], Shubham Pingale[4], Mr. Nikhil Dhavase[5]
[1234]BE Student, Department of Information Technology, Marathwada MitraMandal's College of EngineeringKarvenagar, Pune-411052, and Maharashtra, India
[5]Professor, Department of Information Technology, Marathwada MitraMandal's College of EngineeringKarvenagar, Pune-411052, and Maharashtra, India

Abstract- Storage requirements for visual and Text data have increased in recent years, following the appearance of many interactive multimedia services and applications for mobile devices in personal and business scenarios. Hash methods are useful for a variety of tasks and have attracted great attention in recent times. They have proposed different approaches to capture the similarities between text and images. However, most of the existing work use bag-of-words method is used to represent text information. Since words with different forms may have a similar meaning, the similarities of the semantic text cannot be well worked out in these methods. To address these challenges in this paper, a new method called semantic cross-media hashing, which uses the continuous representations of the proposed words capturing the semantic textual similarity level and the use of a deep conviction network to build the correlation between different modes. To demonstrate the effectiveness of the proposed method, it is necessary to consider three commonly used data sets that are considered basic. Experimental results show that the proposed method achieves significantly better results in addition, the effectiveness of the proposed method is similar or superior to other hashmethods.

Keywords- SCMH, SIFT Descriptor, Word Embedding, Ranking, Mapping

I. INTRODUCTION

With the fast development of internet and multimedia, information with various form has become enough smooth, simple and easier to access, modify and duplicate. Information with various forms may have semantic correlation for example a microblogs in Facebook often consist of tag, a video in YouTube is always associated with related description or tag as semantic information inherently consist of data with different modality provide an great emerging demand for the applications like cross media retrieval, image annotation and recommendation system. Therefore, the hash similarity methods which calculates or approximate search suggested and received a remarkable attention in last few years. The core problem of hash learning is how to formulate underlay co-relation between multiple modality and retain / protect the similarity relation in each respective modalities. Generally hashing method divided into 2 categories: matrix decomposition method and vector based method. Matrix decomposition based hashing method search low dimensional spaces to construct data and quantify the reconstruction coefficient to obtain binary codes. Such kind of methods avoid graph construction and Eigen decomposition. The drawback with such methods, causes large quantization errors which decorate such performance for large code length. We have design multi-modal hashing model SCMH which focuses on Image and Text type of data with binary representation Hashing. This method processed text data using Skip gram model and image data using SIFT Descriptor. After it generates hash code using Deep Neural network by avoiding duplicates.

II. LITERATURESURVEY

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.
In this section, we briefly review the related work on Tag Search and Image Search and their different techniques.

Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin: This paper addresses the problem of learning similarity-preserving binary codes for efficient similarity search in large-scale image collections. We formulate this problem in terms of finding a rotation of zero-cantered data so as to minimize the quantization error of mapping this data to the vertices of a zero-cantered binary hypercube, and propose a simple and efficient alternating minimization algorithm to accomplish this task[1].

Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui: we demonstrate in this paper that the above two fundamental challenges can be mitigated by jointly exploring the cross-view learning and the use of click-through data. The former aims to create a latent subspace with the ability in comparing information from the original incomparable views (i.e., textual and visual views),while the latter exploresthe largely available and freely accessible click-through data (i.e., –crowdsourced‖human intelligence) for understanding query[2].

D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao:

In this paper, we study HFL in the context of multimodal data for cross-view similarity search. We present a novel multimodal HFL method, called Parametric Local Multimodal Hashing (PLMH), which learns a set of hash functions to locally adapt to the data structure of each modality[3].

G. Ding, Y. Guo, and J. Zhou: In this paper, we study the problems of learning hash functions in the context of multimodal data for cross-view similarity search. We put forward a novel hashing method, which is referred to Collective Matrix Factorization Hashing (CMFH)[4].

H. J_egou, F. Perronnin, M. Douze, J. S_anchez, P. P_erez, and C. Schmid: This paper addresses the problem of large-scale image search. Three constraints have to be taken into account: search accuracy, efficiency, and memory usage. We first present and evaluate different ways of aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words approach for any given vector dimension[5].

J. Zhou, G. Ding, and Y. Guo: In this paper, we propose a novel Latent Semantic Sparse Hashing (LSSH) to perform cross- modal similarity search by employing Sparse Coding and Matrix Factorization. In particular, LSSH uses Sparse Coding to capture the salient structures of images, and Matrix Factorization to learn the latent concepts from text[6].

Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y. Zhuang: In DCDH, the coupled dictionary for each modality is learned with side information (e.g., categories). As a result, the coupled dictionaries not only preserve the intra-similarity and inter-correlation among multi-modal data, but also contain dictionary atoms that are semantically discriminative (i.e., the data from the same category is reconstructed by the similar dictionary atoms) [7].

H. Zhang, J. Yuan, X. Gao, and Z. Chen: In this paper, we propose a new cross-media retrieval method based on short-term and long-term relevance feedback. Our method mainly focuses on two typical types of media data, i.e. image and audio. First, we build multimodal representation via statistical canonical correlation between image and audio feature matrices, and define cross- media distance metric for similarity measure; then we propose optimization strategy based on relevance feedback, which fuses short-term learning results and long-term accumulated knowledge into the objective function [8].

A. Karpathy and L. Fei-Fei: We present a model that generates natural language descriptions of images and their regions. Our approach leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Our alignment model is based on a novel combination of Convolution Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding [9].

J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen: In this paper, we present a new multimedia retrieval paradigm to innovate large-scale search of heterogeneous multimedia data. It is able to return results of different media types from heterogeneous data sources, e.g., using a query image to retrieve relevant text documents or images from different data sources [10].

III. EXISTINGSYSTEM

Lot of work has been done in this field because of its extensive usage and applications. In this section, some of the approaches which have been implemented to achieve the same purpose are mentioned. These works are majorly differentiated by the algorithm for multimediaretrieval.

In another research, the training set images were divide into blobs. Each such blob has a keyword associated with it. For any input test image, first it is divided into blobs and then the probability of a label describing a blob is found out using the information that was used to annotate the blobs in the trainingset.

As my point of view when I studied the papers the issues are related to tag base search and image search. The challenge is to rank the top viewed images and making the diversity of that images is main task and the search has that diversity problem so the open issue is diversity.

IV. PROPOSEDSYSTEM

We propose a novel hashing method, called semantic cross-media hashing (SCMH), to perform the near-duplicate detection and cross media retrieval task. We propose to use a set of word embeddings to represent textual information. Fisher kernel framework is incorporated to represent both textual and visual information with fixed length vectors. For mapping the Fisher vectors of different modalities, a deep belief network is proposed to perform the task. We evaluate the proposed method SCMH on two commonly used data sets. SCMH achieves better results than state-of-the-art methods with different the lengths of hash codes and also display query results in rankedorder.

Advantages:

1. Improve accuracy usingSCMH.
2. Retrieve relevant images using SIFTdescriptor.

System Architecture:

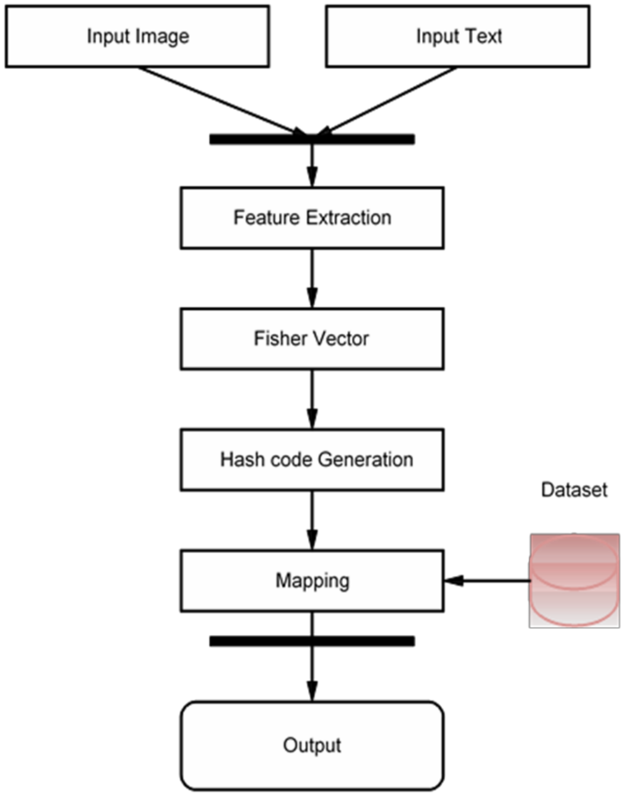

Fig.1: System Architecture

V. CONCLUSION

In this paper, propose a new SCMH novel hashing method for duplicate and cross-media retrieval. We are proposing to use a word embedding to represent textual information. The Fisher Framework Kernel used to represent both textual and visual information with fixed length vectors. To map the Fisher vectors of different modes, a network of deep beliefs intends to do the operation. We appreciate the proposed method SCMH on Mriflicker dataset. In the Mriflicker data set, SCMH over other hashing methods, which manages the best results in this data sets, are text to image & image to Text tasks, respectively. Experimental results demonstrate effectiveness proposed method in the cross-media recoverymethod.

VI. REFERENCES

[1].Y. Gong, S.Lazebnik, A. Gordo,andF. Perronnin, ―Iterativequantization: Aprocrusteanapproach to learning binarycodes for large-scale image retrieval,‖ IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pp. 2916–2929, Dec.2013.

[2].Y.Pan,T.Yao,T.Mei,H.Li,C.-W.Ngo,andY.Rui,―Clickthrough-basedcross-viewlearningforimagesearch,‖inProc. 37th Int.ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp.717–726.

[3].D.Zhai,H.Chang,Y.Zhen,X.Liu,X.Chen,andW.Gao,―Parametriclocalmultimodalhashingforcross-viewsimilarity search,‖ in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp.2754–2760.

[4].G.Ding,Y.Guo,andJ.Zhou,―Collectivematrix factorizationhashingformultimodaldata,‖inProc.IEEEConf.Comput. Vis. Pattern Recog., 2014, pp.2083–2090.

[5].H.J_egou,F.Perronnin,M.Douze, J.S_anchez,P.P_erez,andC.Schmid,―Aggregatinglocalimagedescriptorsintocompact codes, IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 1704–1716, Sep.2011.

[6].J.Zhou,G.Ding,andY.Guo,―Latentsemanticsparsehashingforcross-modalsimilaritysearch,‖inProc.37thInt.ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp.415–424.

[7].Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, andY. Zhuang, ―Discriminative coupled dictionaryhashingfor fast cross-media retrieval,‖ in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp.395–404.

[8].H.Zhang,J.Yuan,X.Gao,andZ.Chen,―Boostingcross-mediaretrievalviavisual-auditoryfeatureanalysisandrelevance feedback,‖ in Proc. ACM Int. Conf. Multimedia, 2014, pp.953–956.

[9].A.KarpathyandL. Fei-Fei, ―Deep visual-semantic alignments for generating image descriptions,‖in Proc. IEEE Conf.Comput. Vis. Pattern Recog., Boston, MA, USA, Jun. 2015, pp.3128–3137.

[10]. J.Song,Y.Yang,Y.Yang,Z.Huang,andH.T.Shen,―Inter-mediahashingforlarge-scaleretrievalfromheterogeneousdata sources,‖ in Proc. Int. Conf. Manage. Data, 2013, pp.785–796.