Efficient Cross Media Retrieval using Mixed Generative based Hashing Method

MrunmayiPati[1], Mr. Vinod Wadne[2]
[1]*ME Student,* [2]*Professor,*
[12]*Department of Computer Engineering, JSPM's Imperial College of Engineering and Research,Wagholi,Pune.*

Abstract- Hash methods are proven useful for a variety of tasks and have sparked great attention in recent years. They have proposed several approaches to capture the similarities between textual, visual, and cross-cultural hashing. However, most existing text bag methods used to represent textual information. Because words with different shapes can they have a similar meaning, semantic text similarities cannot be well elaborated in these methods. To address these challenges in this paper, a new method called semantic cross medial hash (SCMH), which uses continuous representations of proposed words by capturing the semantic textual similarity level and using a deep conviction network (DBN) to build correlation between different modes. In order to demonstrate the effectiveness of the proposed method, three methods commonly used are to be considered background set in this workbook is used. The experimental results show that the proposed method achieves significantly better results in addition, the effectiveness of the proposed method is similar or superior some other hashing methods.

Keywords- Fisher vector, 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 micro blogs 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 forthe applications like cross media retrieval, image annotation and recommendation system. Therefore, the hash similarity methods which calculate 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 avoids graph construction and Eigen decomposition. The drawback with such methods, causes large quantization errors which detonate such performance for large code length and 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.

A. *Problem Statement*

Image search is an important method to find images contributed by social users in such websites. However, how to make the result relevant and with diversity is challenging.

II. RELATED WORK

Literature survey is the most important step in any kind of research. Before start developing need to study the previous papers of our domain which 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, 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 binary codes that maintain similarity for efficient search for similarities in large-scale image collections. They have formulated this problem in terms of searching for a zero-angle data rotation to minimize the error of quantizing the mapping of these data to the vertices of a binary zero-angle hypercube and they have proposed a simple and efficient alternative minimization algorithm to perform this operation [1].

Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui: They demonstrate in this paper that the two fundamental challenges listed above can be mitigated by jointly exploring cross-view learning and the use of click data. The first one aims to create a latent subspace with the ability to compare information from the original incomparable views (ie visual and textual), while the second explores the click data widely available and freely accessible (ie human intelligence) "collective participation") to understand the query [2].

D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao: In this article, They study HFL in the context of multimodal data for cross-section similarity research. Theypresent a new multimodal method of HFL, called Hashing multimodal local parametric (PLMH), which learns a set of hash functions to adapt locally to the data structure of each mode [3].

G. Ding, Y. Guo, and J. Zhou: In this article, Theystudy the problems of learning hash functions in the context of multimode data for cross-look similarity research. Theypresent a new hash method, which refers to the factorial factoring hash (CMFH)[4].

H. J_egou, F. Perronnin, M. Douze, J. S_anchez, P. P_erez, and C. Schmid: This document addresses the problem of large-scale image research. Three restrictions must be considered: the accuracy of research, efficiency and use of memory. First we present and evaluate different ways to add local image descriptors to a vector and show that Fisher's kernel performs better than the reference image focus of the image bag for any vector dimension[5].

J. Zhou, G. Ding, and Y. Guo: In this paper, they propose a new latent latent semantic dispersion (LSSH) to perform an intermodal similarity search using scattering and matrix factorization. In particular, LSSH uses Sparse Coding to capture outgoing image structures and Matrix Factorization to learn latent text concepts [6].

Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y. Zhuang: In DCDH, the paired dictionary for each mode is learned with complementary information (for example, categories). Consequently, coupled dictionaries not only maintain the intra-similarity and correlation between multimode data, but also contain dictionary atoms that are semantically discriminating (that is, data in the same category are reconstructed from similar dictionary atoms) [7].

H. Zhang, J. Yuan, X. Gao, and Z. Chen: In this paper, we propose a new method of cross-media recovery based on short and long-term relevance feedback. Our method focuses mainly on two typical types of multimedia data, namely image and audio. First, we construct a multimodal representation through the canonical statistical correlation between the image matrices and the audio characteristics, and define the metric of the distance between the means for measuring similarity; therefore theyproposean optimization strategy based on the relevant feedback, which combines the results of short-term learning and the accumulated long-term knowledge in the objective function[8].

A. Karpathy and L. Fei-Fei: They present a model that generates descriptions in the natural language of images and their regions. Our approach takes advantage of the image data sets and their sentence descriptions to know the intermodal correspondences between language and visual data. Our alignment model is based on a new combination of convolution neural networks on image regions, bidirectional neuronal networks recurring on sentences and a structured goal that aligns the two modalities through a multimodal incorporation[9].

J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen: In this paper, Theypresent a new multimedia recovery paradigm to innovate large-scale research of heterogeneous multimedia data. You can return results from different types of media from heterogeneous data sources, for example, by using a query image to retrieve relevant text documents or images from different data sources [10].

III. EXISTING SYSTEM

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 multimedia retrieval.

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 training set.

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.

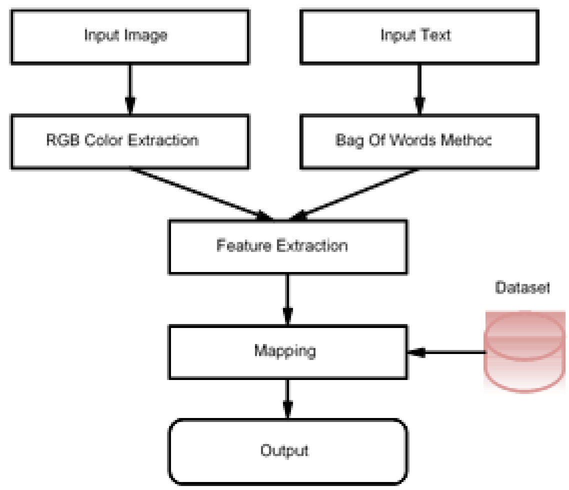

Fig.1: Existing System Architecture

IV. PROPOSED SYSTEM

Propose a new method of hashing, called semantic cross media hashing (SCMH), to perform the task of almost duplicate detection and recovery of cross media. We propose to use a set of words inlays to represent textual information. The Fisher kernel structure is incorporated to represent textual and visual information with fixed-length vectors. To map Fisher vectors in different ways, a network of deep beliefs is proposed to carry out the task and evaluated the proposed SCMH method in two commonly used data sets. SCMH achieves better results than cutting-edge methods with different lengths of hash code and displays query results in ranking order.

Advantages:

- Introduce a novel based method to construct the correlation between different modalities.

- The Proposed method can significantly outperform the state-of-the-art methods.
- Improve the searching accuracy.

System Architecture:

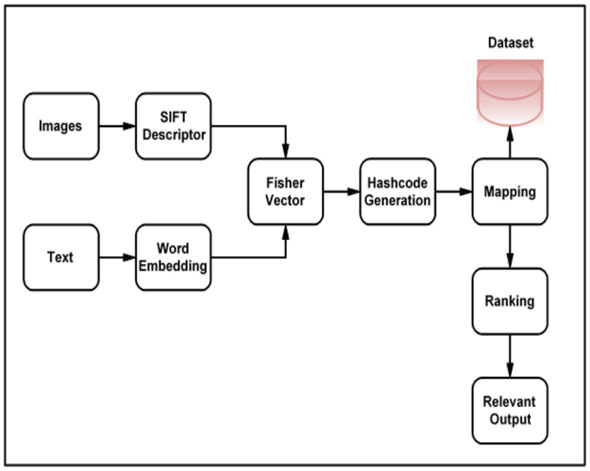

Fig.2: System Architecture

Algorithms:

A. FEATURE DESCRIPTOR

SIFT Descriptor is used for representing Images, we use SIFT detector to extract image key points. SIFT descriptor is used to calculate descriptors of the extracted key points and a variable size set of points in SIFT descriptor space represents each image.

1. The procedure to search in a repository R with query image Q.
2. The input for this operation on the user side is IDR, Q, repository key rkR, and parameter k (the number of most similar results to be returned).
3. User U starts by generating Q's searching trapdoor CQ, through IES-CBIR.
4. Then sends it to the cloud server, along with k and IDR, as parameters for the Search remote invocation.
5. The cloud starts by extracting CQ's feature-vector, stems it against CBR to determine its visual words vwCQ, and accesses IdxR with them to retrieve the respective posting lists PLvw.
6. Then, for each image referenced in each of the posting lists retrieved, the cloud calculates its scaled tf-idf score and adds it to the set of results for the query. In this set, scores for the same image but different visual word are summed.
7. Finally, the cloud sorts this set by descending score and returns the results to user.

B. WORD EMBEDDING

Skip-gram algorithm is used for word embedding. After Skip-descriptor steps, a variable size set of points in the embedding's space represents the text.

C. HASHCODE GENERATION

MD5 algorithm is used hash function producing a 128-bit hash value.

The MD5 message-digest algorithm is a widely used cryptographic hash function producing a 128-bit (16-byte) hash value, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity.

Steps:

- A message digest algorithm is a hash function that takes a bit sequence of any length and produces a bit sequence of a fixed small length.
- The output of a message digest is considered as a digital signature of the input data.
- MD5 is a message digest algorithm producing 128 bits of data.
- It uses constants derived to trigonometric Sine function.
- It loops through the original message in blocks of 512 bits, with 4 rounds of operations for each block, and 16 operations in each round.
- Most modern programming languages provides MD5 algorithm as built-in functions

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 recovery method.

VI. REFERENCES

[1]. Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes 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, and Y. Rui, "Clickthrough-based cross-view learning for image search," in Proc. 37thInt.ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 717–726.

[3]. D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao, "Parametric local multimodal hashing for cross-view similarity

search," in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp. 2754–2760.

[4]. G. Ding, Y. Guo, and J. Zhou, "Collective matrix factorization hashing for multimodal data," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 2083–2090.

[5]. H. J_egou, F. Perronnin, M. Douze, J. S_anchez, P. P_erez, and C. Schmid, "Aggregating local image descriptors into compact codes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 1704–1716, Sep. 2011.

[6]. J. Zhou, G. Ding, and Y. Guo, "Latent semantic sparse hashing for cross-modal similarity search," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 415–424.

[7]. Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y. Zhuang, "Discriminative coupled dictionary hashing for 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, and Z. Chen, "Boosting cross-media retrieval via visual-auditory feature analysis and relevance feedback," in Proc. ACM Int. Conf. Multimedia, 2014, pp. 953–956.

[9]. A. Karpathy and L. 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, and H. T. Shen, "Inter-media hashing for large-scale retrieval from heterogeneous data sources," in Proc. Int. Conf. Manage. Data, 2013, pp. 785–796.