

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

SCMH : Cross Media Retrieval

Mayuri D. Pawar, Prof. Sandhya S. Deore

Department of Computer Engineering, SRES COE, Kopargao , Pune, India

ABSTRACT: Hashing methods are important and widely useful technique in recent years. Several methods it introduced to find out the similarities between text, video-audio, and cross-media information. In existing, use a bag-of-words is a technique of representing text data and also used in natural language process-ing and information retrieval. Because information with various forms may have a same meaning, semantic text similarities are not well elaborated in these methods.

In this paper, proposed a new method called semantic cross media hashing (SCMH), which uses bag of words by capturing the semantic textual similarity level. In proposed method, two commonly used cross media datasets (i.e. mriflicker and label me).

KEYWORDS: Fisher vector, hashing method ,SIFT descriptor, word embedding.

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 a 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 detorate 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 represen-tation 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. REVIEW OF LITERATURE

In paper[7], author addresses the Problem of large scale image search.They have considered Three points for the performance such as: search accuracy, efficiency, and memory utilization. In this paper, author uses different ways for ag-gregating local image descriptors into vectors and evaluates it and also evaluates that method which shows that Fisher Kernal Achieves better performance than bag of words approach. In paper[6], focuses on problems of learning Hash functions for multi-modal data.author also introduce a new hashing method, Collective Matrix Factorization Hashing (CMFH) which gen-erate the unified hash code by collective matrix factorisation of various modalities of one sample, item.This method consider that each model of instance produce a unique hash code this constraints gives

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

limitations to the of pairwise data points and ignores relationship between different pairwise data. In paper[8] new technique of multimedia retrieval paradigm is presented to discover large scale search for heterogeneous multimedia data It is capable to return a retrieved results of various media categories from heterogeneous information sources / ways, e.g., considering a query image to recover or retrieved a significant content documents or pictures from distinctive information sources.

Y. Pan et al.[2] demonstrate a Latent Semantic Sparse Hashing (LSSH) to achieve cross-modal resemblance or close-ness search by utilizing Sparse Coding and Matrix Factorization. In specific, LSSH need Sparse Coding to capture the salient structures of images, and Matrix Factorization to learn the idle concepts from content. The first aims to construct a latent subspace with the capability in matching information from the unique exceptional views (i.e., textual and visual views), while the last mentioned, investigates the generally accessible and openly available click-through information (i.e., crowd sourced human intelligence) for analysing query. Paper[4] analyse the issue of learn resemblance, sameness preserving(retain) binary codes for powerful similarity search in large-scale image collections. author define this issue in terms of finding a rotation of zero-centered data so as to minimize the quantization error of mapping this data to the vertices of a zero-centered binary hypercube, and developed a easy simpler and efficient substitute minimization algorithm procedure to achieve a task.

Paper[6], demonstrate HFL in the context of multi-modal information for cross-view likeness search. It display HFL strategy, called Parametric Local Multimodal Hashing (PLMH), which usage a set of hash functions to locally prepare the data structure of various methodology. In DCDH, the coupled conjugate dictionary for each modality is determine with sidewise information (e.g., categories). As the result information contain dictionary atoms that are semantically dis-criminative (i.e., the information from the similar group,class is reproduced by comparable dictionary atoms).

In paper[5], author suggest a modern cross-media recovery strategy based on relevance score. The strategy basically targeted on two normal category, form of media data, i.e.image and sound. Initially,It construct multi-modal representation by means of statistical canonical relationship among picture and sound feature matrices, and define cross-media distance metric for similarity degree; at that point author explain optimization methodology it depends on relevance score, which results gathered information into the objective function. In paper[3], author demonstrate a method that produce natural language detail description of images and their regions. This technique jump on datasets of images(Flicker) and their sentence descriptions to learn about the inter-modal analogy between language and visual data. This arrangement demonstrate a combination of Convolution Neural Systems over image fields, bidirectional Repetitive Neural Systems over sentences, and a design objective that adjusts bi-modalities wrapped up a multimodal embedding.

III. SYSTEM OVERVIEW

The Semantic Cross Media Hashing(SCMH) model combines more data method in a unitary representation that can be used for classification and recovery. Image search is an essential method to find images shared by social, popular community users in such social websites. The challenging task is to find out the top rank relevant result with diversity. A fundamental problem is-rank the result of images or texts. In societal image retrieval is how to accurately and loyally solve these images or texts mismatch problems.

Use a novel SCMH technique, to implement the near-duplicate detection and cross media retrieval task. As shown in Fig.1 it uses a set of word embeddings to display text data. Fisher framework is organise to show, produce both text and optical information with fixed vectors. For checking the Fisher vectors of various techniques-modalities, a deep belief network is used to achieve the task.

SIFT 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

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

descriptor space represents each image.

Skip Gram Skip-gram algorithm is used for word embedding. After Skip-descriptor steps, a variable size

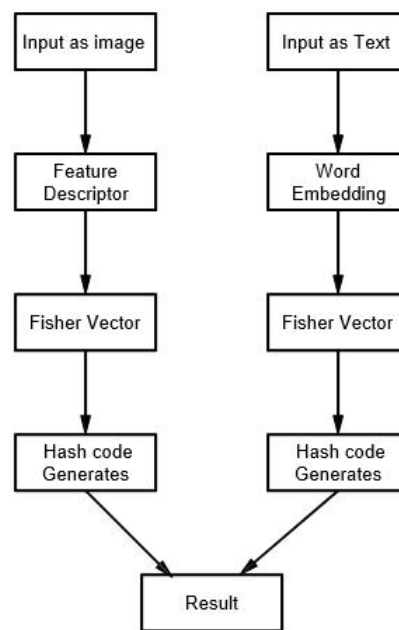


Fig. 1. System Architecture.

set of points in the embeddings space represents the text.

IV. SYSTEM ANALYSIS

Consider a collection of data, text and image data represent respectively,

– Text Representation : Skip Gram model

All the words in the given text information is trans-late or switch to distributed shared vectors reproduce by the word embedding task.

– Image Representation : SIFT Descriptor

SIFT Descriptor is used to detect and extract key points of images. By calculating / extracting image key points SIFT identifies the object or perform object Recognition.

– After these two steps, a variable size set of points representing their area.

– SIFT descriptor : Display each Image

– Word embedding : Display Textual Information.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

- For combining and representing all points of different modality to one a framework used called fisher kernel.
- Fisher Framework: Utilized to combine these points of distinct / particular space into fixed length vectors.
- All information are represented by a fixed length vectors. i.e. Images and Text are display with fixed length vectors. After this using mapping function to calculate or checking textual and visual fisher vectors (FV). We are using mapping method to convert FV of single procedure, technique / style to different one. Finally for converting fisher vector to binary vector we are using hashing technique.

Output: Result- most relevant media with higher score.

B. Dataset

We are using MRI-Flicker dataset. flicker dataset consists of 25000 images.

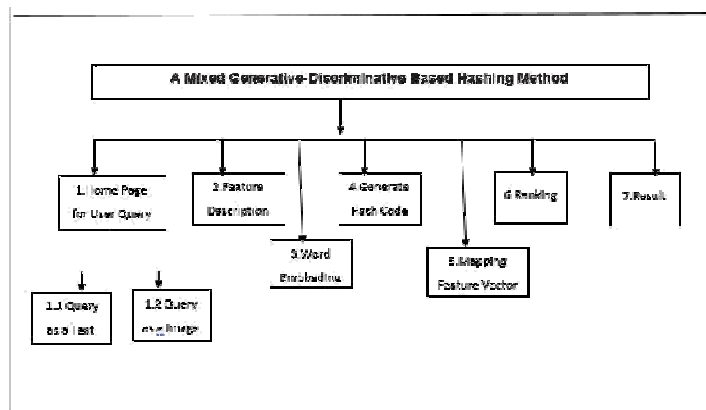


Fig. 2. Breakdown Structure.

Fig.2 shows the breakdown structure of system which contain following modules:

- T1: Home Page : After login user will get home page on that he/she will enter Query. Input Query is in the form of Text or Image
- T2: Features Extraction : System will extract Features of input query Image. SIFT descriptor Used for Extracting Key features of image.
- T3: Word Embedding : Perform word embedding of Textual information using skip gram model.
- T4: Generate Hash code : System Generates Hash code for query.
- T5: Mapping Feature Vector : The process of mapping used to perform the duplicate detection.
- T6: Ranking :Ranking are performed with relevance score of user query with retrieved document or media.
- T7:Result:most relevant media with higher score.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

A. System Steps

1. Input: User Query
2. Query Input as Image or Text: user enter any media (text or image) as an input
3. Perform feature description (Image)
4. Use SIFT detector to extract image key points
5. Perform Word embedding(Text)
6. Feature Vector: Represent fisher vector for both media (text and image).
7. Generate Hash Code-converting fisher vector to the binary vector
8. Mapping Feature Vector-after mapping vector most relevant media ranked or indexed accordingly.
9. Ranking-ranking are performed with relevance score of user query with retrieved document or media.

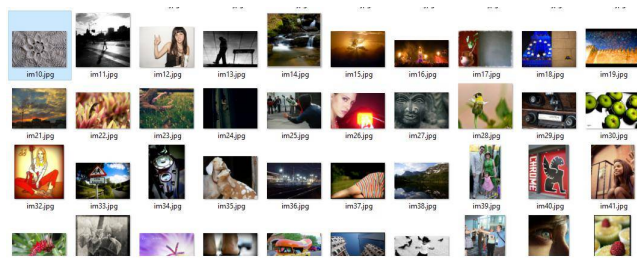


Fig. 3. mri-flicker dataset

C. Result

In the MRIFLICKER data set, which manages the best results.

Training:



Fig. 4. Admin Side

Admin module contains number of steps as shown in Fig.5.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

View all user and Authorised. Add images with Title.
View all added images.

View all images with similar hash code. View all images with different hash code.

User Side:

According to query which is entered by user further

steps are carried out.

Tag based search

For example: As shown in Fig.9 user enters query as 'girl' then user get results which is most relevant to that particular query i.e. Images of girls.

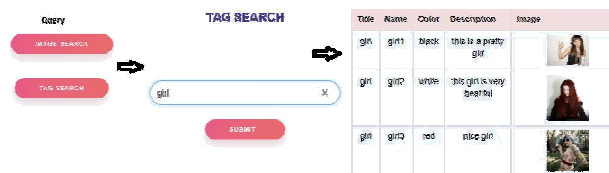


Fig. 5. Retrieved relevant Result

Image Search

For Example: User enter query as 'Nature' then user get retrieved result as - all relevant images which is related to particular query image i.e. Images of Nature.

V. CONCLUSION

The system uses a new hashing method, SCMH a duplicate and cross-media detection restoration activity. This system uses a series of words to represent text Data. The Fisher Kernel Framework built 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.

ACKNOWLEDGEMENT

I would like to thank my esteemed guide Prof. S. S. Deore, whose interest and guidance helped me to complete the work successfully. Also the valuable help of Dr.D.B.Kshirsagar (HOD Comp. Dept.) and Prof. P.N.Kalavadekar "PG Co-ordinator" who has provided facilities to explore the subject with more enthusiasm.

REFERENCES

- [1] Qi Zhang, Yang Wang, Jin Qian, and Xuanjing Huang, "A Mixed Generative-Discriminative Based Hashing Method," in IEEE Transaction on Knowledge and data Engineering, VOL. 28, NO. 4, APRIL 2016
- [2] G. Ding, Y. Guo, and J. Zhou, "Collective matrix factorization hashing for multimodal data," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014.
- [3] 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.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

- [4]Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui, "Clickthrough-based cross-view learning for image search," in Proc. 37th Int.ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 717726.
- [5]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. 785796.
- [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. 415424.
- [7]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. 31283137.
- [8]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. 27542760.
- [9]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. 785796.
- [10]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. 395404.
- [11] 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. 953956.