

Multimodal Information Retrieval by using Visual and Textual Query

Ankita Soni

M.Tech. Scholar

Department of Electronics and Communication
Oriental Institute of Science and Technology,
Bhopal (M.P.)

Richa Chouhan

Assistant Professor

Department of Electronics and Communication
Oriental institute of Science and Technology,
Bhopal (M.P.)

ABSTRACT

As the number of internet users are increasing day by day. So amount of data also increases, so fast response is desired by different users. So most of the researchers are working in relevant information retrieval. Proposed work has focus on image retrieval by the combination of textual and visual features. In this work annotations of the image are use for the initial filtering of image from the dataset. After first filtration selected images are sent for visual feature base comparison. On the basis of the visual feature combination final ranking of the images are done as per the input query. Results shows that combination of both type of feature give more effective results as compare to single one.

Keywords

Digital Image Processing, Information Extraction, feature extraction, Re-ranking.

1. INTRODUCTION

Search engines are playing very important role in today's era, as number of internet users are increasing day by day. Hundreds of services are available on the internet for different purposes, now search engine help in finding those services as per user requirement in form of query.

Image retrieval or searching also plays important role in searching. So researchers are working on this field from last few decades. Out of these many are working on image indexing or runtime indexing on the basis of visual features. As in popular search engine like Google, Yahoo, Bing, etc. use this concept of the searching the image as user type query for the same but the ranking is totally base on the textual data attached with the image.

Another way of searching the image is by passing the image as a query as in above user query in form of text [3, 13]. In this visual features of the image is extract with the help of various feature extraction algorithm. Here same type of feature are extract from the query image and compare with the dataset images. So base on the similarity of the visual features final ranking is done.



Fig1: Image retrieval by text query.

One more approach for this is done by the utilization of the textual feature which are name or any other content with the visual feature of the image [4]. So a combination of both are use here results are much better as compare to the single feature query algorithm. But here the execution time is comparatively very large so time complexity is more because of this although results are much effective but time duration make in impractical for common use.




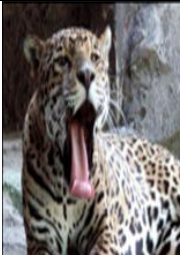
Dataset	Query By Example	Relevant images
		
		



Fig. 2 Image retrieval by visual query.

This paper is focus on the visual as well as textual feature of the image and reranked images on the basis of the query passed by the user. Some of the important feature of the image which is used in the images are color, CCM, GLCM. This paper give a detailed study of the different feature and there combination for the result.

2. RELATED WORK

In [1] image are categorize into two class first is relanvent and other is irrelevant. So re-ranking of the database is done on the basis of the visual features. So image harvesting is done where relevant image are rank as top filtered image.

In [7] a method that uses user intervention to re-rank the results is proposed. In their approach, a small subset of correct images is manually constructed and this subset is then used to filter the noise in the data. There have been some works exploiting multiple modalities for image search applications. In [8] perform joint clustering in a space built from visual and tag descriptors to find iconic summaries of abstract concepts.

In [10] fetching of image are done from the web by reading the surrounding image. Similarly In [9], textual web content is utilized for the web page ranking, where support vector machine is trained for ranking this textual information.

Few of the image reranking algorithm utilize the web page ranking method as in [11] where important pages are ranked first base on their importance. It builds a graph representing the link structure of the web. As the web page get impotant depend on the number of hyperlink present in it.

As [2, 12] shows the approach for page reranking by the multi-modal where it utilize both the visual as well as textual feature. Here it jointly fetch both the feature and implement it in the many applications. One more approach is use here that include autoannotation where text data is automatically updated.

Background

As Image is collection of pixel and each pixel is treat as single value which is a kind cell in a matrices. In order to identify an object in that image some features need to be maintained as different object have different feature to identify them which are explain as follows:

Algorithm for Re-Ranking by color and CCM Feature:

Coloure feature: By storing the pixel value in an image is nothing accept an representation of the color intensity at that

point in an image. Because of the various image format pixel values are different for the same image. This can be

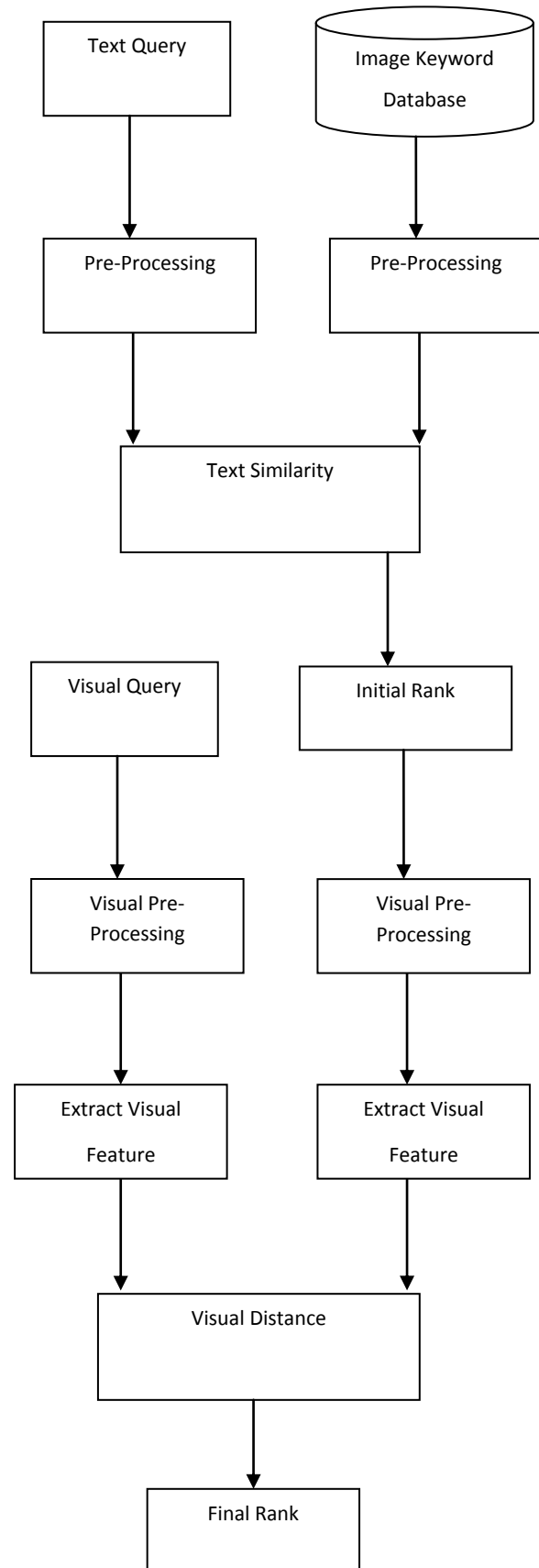


Fig 2: Block diagram of proposed work.

understand as in RGB three matrix of same image are prepare for Red, Green, Blue matrix range of any pixel value is between 0-1. While in the case of the gray format it is of single matrix range of any pixel value is between 0-255.

3. PROPOSED WORK

Pre-Processing

By pre-processing of the dataset overall size of the data get reduce which of the help in increasing the efficiency work. As some stop words are present in the query for the framing of sentence which can be remove. Now read the file which contains stop words then remove similar words from the vector. Once the data is pre-process then it will be the collection of the words that may be in the vectn be remove from the query. For example let one query is taken and its text vector is $Rd[] = \{a1, f1, s1, a2, s2, a3, a4, f2, \dots, an\}$ and let the stop words collection is $S[] = \{a1, a2, a3, \dots, am\}$. Then the vector obtain after the Pre-Processing is $D[] = \{f1, s1, s2, f2, \dots, fx\}$.

$$D[] = Rd[] - S[]$$

For Example: $Rd[] = \{\text{'Every', ' morning', 'Ram', ' study', ' for', ' two', ' hour', ' and', ' during', ' this', ' time', ' his', ' mother', ' give', ' him', ' one', ' glass', ' milk', ' with', ' bread', ' jam', ' in', ' breakfast'}\}$

After pre-processing

Now $D[] = \{\text{'Ram', ' hour', ' time', ' glass', ' milk', ' bread', ' jam', ' breakfast'}\}$

Text Similarity:

Here similarity between the query image and database image is calculated. In order to find the distance between text information, calculate the similar words. In other words text query and database image annotations are taken as two sets where number of common element is consider as the similarity value. This can be understand as let query $Q = \{a, b, c\}$ and database annotation $I_1 = \{p, a, b, w\}$ then common element $|Q \cap I_1| = 2$. In similar fashion other image annotations of the image are compared with query Q.

Initial Rank

So rank of the database image is initial done on the base of the text similarity. Here all similarity is sort in decreasing order then top t images such as (t=20) is analyzed for the image re-ranking. This step is also considered as the filtration of the databases on the basis of the image query by the user. Due to this filtration of the database overall searching time get reduced.

Visual Pre-Processing

In this step image is resize in fix dimension. As different image have different dimension. So conversion of each is done in this step. One more work is to convert all image in gray format. AS different image has RGB, HSV, etc. format so working on single format is required.

Extraction of Visual Feature

This work has evaluate two visual features first is edge and second is corner.

Edge feature: In case of edge feature canny algorithm is apply.

- Smooth the Image with Gaussian Filter.
- Compute the Gradient Magnitude and Orientation using finite-difference approximations for the partial derivatives.
- Apply non-maxima suppression to the gradient magnitude.
- Use the double thresholding algorithm to detect and link edges.

Corner Detection:

This feature is calculate to detect the corner feature of the image. As corner feature find the image in all case of object detection. Here when image get zoom-in or zoom out then similarity between these images is identified by corner feature. Corner points are calculate by the Harris algorithm in [13].

Visual Distance

In case of corner feature distance between the images are obtained by calculating the similar points between the images. Similarity is obtained by the surf feature algorithm [14].

Eludician Distance:

In case of edge feature Eludician distance formula is used. This can be understand as Let X be a query image matrix and Y be the dataset image matrix. Then distance between them is calculate by:

$$D = \sqrt{\text{sum}((X - Y)^2)}$$

Base on the minimum distance value between query and dataset image rank is assigned to the image. This is consider as final rank of the work.

4. EXPERIMENT AND RESULT

In this section, first introduce experimental settings, and then present the experimental results that validate the effectiveness of the approach. The experiments actually contain two parts. This work is compare with other several existing methods.

We adopt NDCG [6, 12] as the performance evaluation measure.

The NDCG measure is computed as

$$NDCG@P = Z_P \sum_{i=1}^P \frac{2^{l(i)} - 1}{\log(i + 1)} \quad (9)$$

where P is the considered depth, $l(i)$ is the relevance level of the i -th image and Z_P is a normalization constant that is chosen to let the optimal ranking's NDCG score to be 1.

Data Sets and Preprocessing

In order to conduct the experiment an artificial dataset which is a collection of images from different category are utilize. As images are of different format so first it is necessary to make it in readable format for experiment tool MATLAB. Now this collection of images of different category is shown in table 1 for which one can make some important keyword collection for different images. In this way each image have one more feature to identify that is the keys of the images.

Table 1 Dataset of Different category.

Category	Examples
Objects	Ipod, map
animal	Butterfly, Gorilla
scene	Taj Mahal, Hotel Taj
Person	Barack Obama, Lena

By entering the query and search the desired image it was obtained that they can be categorize into few levels such as relevant or not. It can be further categorize into most relevant, relevant, less relevant, irrelevant.



Fig 2: Above image are generate from two category relevant and irrelevant for two query 'Taj Mahal', 'Barack Obama'.



Fig 3: Results obtained by visual,



Fig 4: Results obtained by Textual,



Fig 5: Results obtained by Visual +Textual.

Table 2. Average Values of NDCG@10 by Different features and there combination.

Methods		NDCG Values		
		Person	Animal	Scene
1	Color [1]	0.623846	0.371394	0.225686
2	CCM [1]	0.377296	0.253081	0.319241
3	Color+CCM [1]	0.366881	0.349068	0.360709
4	Textual	0.641046	0.504601	0.27267
5	Edge	0.47577	0.327928	0.38786
6	Corner	0.124454	0.4495	0.233562
7	Edge+Corner+Text	0.928603	0.860581	0.936379

From the above table it is find that the including of the new feature for query has increase the efficiency of image re-ranking. In different categories of the images one can find that results are improved.

Table 3. Average Values of NDCG@7 by Different features and there combination.

Methods		NDCG Values		
		Person	Animal	Scene
1	Color [1]	0.546217	0.481413	0.213262
2	CCM [1]	0.377296	0.29258	0.319241
3	Color+CCM [1]	0.337398	0.318737	0.337398
4	Textual	0.295875	0.701131	0.295875
5	Edge	0.469738	0.431281	0.469738
6	Corner	0.155433	0.431281	0.155433
7	Edge+Corner+Text	0.954188	0.934013	0.954188

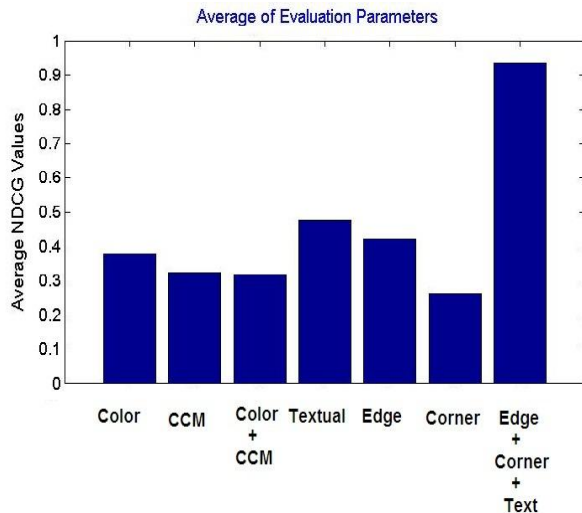


Fig 6: Average NDCG values for different environment.

It is clear from the above bar graph that the values at different NDCG levels is better then the individual feature method. So the utilization of the visual features has increase the efficiency of the work then the single features.

5. CONCLUSION

As the users are diverted to automatic tools where efficiency of the result plays an important role. So proposed work has focus to increase the image retrieval work by utilizing multimodal. Web image re-ranking has been widely used to reduce the user searching time on the internet; its success mainly depend on the accuracy of image features similarities. This paper present utilizing of the text as well as visual features for ranking the image as both make the re-ranking process more powerful, which is shown in results.

In future in order to improve the efficiency more features of images will be include. As to reduce the server time for re-ranking one filter need to be inserted so that it will filter relevant and irrelevant images at the initial stage.

6. REFERENCES

[1] Ms. Apurva N. Ganar, C. S. Gode, Sacine M. Jambhulkr. "Enhancement of Image Retrieval By using Color, Texture and Shape Features". International Conference on Electronic System, singnal Processing and Computing Technology. IEEE computing Society 2014

[2] K. Barnard, P. Duygulu, N. De Freitas, D. Forsyth, D. Blei, And M. Jordan. Matching Words And Pictures. J. Machine Learning Research, 3:1107–1135, Feb 2003.

[3] G. Iyengar, H. J. Nock, And C. Neti, "Discriminative Model Fusion For Semantic Concept Detection And Annotation In Video," In *Proc. Acm Multimedia*, 2003, Pp. 255–258.

[4] R. Yan And A. Hauptmann, "The Combination Limit In Multimedia Retrieval," In *Proc. Acm Multimedia*, 2003, Pp. 339–342.

[5] Msapthagiri.K, Manickam.L. Based On Color, Texture (GlcM & Ccm) Features, And Genetic-Algorithm. International Journal Of Merging Technology And Advanced Research In Computing. Issn: 2320-1363

[6] K. Jarvelin And J. Kekalainen, "Cumulated Gain-Based Evaluation Of Ir Techniques," *Acm Trans. Inf. Syst.*, Vol. 20, No. 4, Pp. 422–446, 2002.

[7] N. Morsillo, C. Pal, And R. Nelson. Mining The Web For Visual Concepts. In *9th Kdd Multimedia Data Mining Workshop*, 2008.

[8] R. Raguram And S. Lazebnik. Computing Iconic Summaries Of General Visual Concepts. *Computer Vision And Pattern Recognition Workshop*, 0:1{8,2008.

[9] F. Schro@, A. Criminisi, And A. Zisserman. Harvesting Image Databases From The Web. In *Computer Vision, 2007. Iccv 2007. Ieee 11th International Conference On*, Pages 1{8, Oct. 2007.

[10] G. Wang And D. Forsyth. Object Image Retrieval By Exploiting Online Knowledge Resources. In *Ieeeconference On Computer Vision And Pattern Recognition*, Pages 1{8, 2008.

[11] Y. Jing And S. Baluja. Visualrank: Applying Pagerank To Large-Scale Image Search. *Ieee Trans. Pattern Anal. Mach. Intell.* , 30(11):1877{1890, 2008.

[12] Meng Wang, Hao Li, Dacheng Tao, Ke Lu, And Xindong Wu "Multimodal Graph-Based Reranking For Web Image Search. *Ieee Transaction On Image Processing* Vol. 21, No. 11, November 2012.

[13] G. Bradski And A. Kaehler, *Learning Opencv : Computer Vision With The Opencv Library*, O'reilly, Sebastopol, Ca, 2008.

[14] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, Surf: Speeded Up Robust Features", *Computer Vision And Image Understanding (Cviu)*, Vol. 110, No. 3, Pp. 346--359, 2008.