Survey of Image Retrieval via Visual Attention & Saliency Feature

Prof. Anand K. Hase[#1], Prof. Baisa L. Gunjal[*2]

[#]*Assistant Professor, Department of Computer Engineering, Sahyadri Valley College of Engineering & Technology, Pune*
[*]*Assistant Professor, Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner*

Abstract— **In the real world applications such as landmark search, copy protection, fake image detection, partial duplicate image retrieval is very important. In the internet era users regularly upload images which are partially duplicate images on the various domains. The partial image is only part of whole image, and the various kind of transformation involves scale, resolution, illumination, viewpoint. This technique is demanded by various real world applications and thus has attracted towards this research. In object based image retrieval methods usually use the whole image as the query. This technique has been compared with the text retrieval system by using the bag of visual words (BOV)[3]. There is lots of background noise in the images and impossible to perform interaction operation on the large scale database of the images. Two observations are notable as a user point of view. First, people show various objects or region through the images which are shared on the web, we also expect that the returned result also focus on the major parts or objects. Regions of interest are only found in salient region of the retrieval. Second, and the similar region in the returned result also identical to the salient region of the images. To filter out the non salient region from the image, which able to eliminate the background noise we introduce visual attention analysis technique. We also want to generate saliency region which having the expected visual contents.**

Keywords— **Partial Duplicate Image, Bags of Visual Words (BOV), Visual Attention, Saliency Feature, Visually Salient & Rich Region.**

I. INTRODUCTION

In this paper, we compare a partial duplicate image retrieval scheme based on different saliency visual matching technique which abstract visually salient and rich region (VSRR) from the images. In these techniques we study the VSRR using a BOV model. To achieve a lower reconstruction error and obtaining a sparse representation at the region level we use a group sparse coding. We also compare our result of image retrieval performance with other image database and show the effectiveness and efficiency of our approach of image retrieval[7].

II. PARTIAL DUPLICATE IMAGE RETRIEVAL ARCHITECTURE

We propose partial duplicate image retrieval scheme based on visual attention and saliency feature[2]. We use the group sparse coding to abstract Visually salient and rich region (VSRR) in the images as retrieval units. BOV model used to represent VSRR. We also apply experiment on different database for partial duplicate image retrieval which shows the efficiency and effectiveness of our approach.

A. VSRR Generation

VSRR is a specific region from image which having rich visual contents and visual saliency[1]. The VSRR generation process is mainly divided into four different areas: Sensitive unit construction, generation of saliency map, VSRR generation and finally selection of ultimate VSRR. The resulted image decomposed into the VSRR sets.

B. Sensitive Unit Generation

Sensitive unit is defined as image patch that corresponds to the center of field which willing to accept new fields around it. For that in most cases graph base segmentation algorithm which merges smaller size patches with similar appearances and small minimum spanning tree weight.

C. Generation of Saliency Map

The specific region of the image which having strong contrast with their surrounding attracts human attention. This spatial relationship plays an important role in visual attention. The region having high attraction which is highly contrast with its near region than the high contrast with its far region. Saliency map is computer on the basis of spatial relationship with the contrast region. This technique is used to separate the object from their surroundings.

On the RGB color space the color histogram is built, then the sensitive unit, its saliency value is calculated by measuring color contrast of other perceptive units present in the image. The weight of spatial relationship is defined as increase the effect of closer region and decrease the effect of farther region.

D. Generation of VSRR

The specific region of the image which having strong contrast with their surrounding attracts human attention. This spatial relationship plays an important role in visual attention. The region having high attraction which is highly contrast with its near region than the high contrast with its far region. We compute saliency map on the basis of spatial relationship with the contrast region. This technique is used to separate the object from their surroundings.

Once the saliency map gets generated the next move is to compute the saliency region by saliency segmentation and then obtain the original VSRR by filtering the saliency[1][2]. After filtering we select the VSRR that contains large numbers of visual content.

By binarizing the saliency map using threshold we divide saliency map into background and initial saliency region. On the initial saliency region we apply grabcut. Grabcut is an interactive tool for foreground segmentation in still images using iterated graph cuts. Finally a group of region is obtained which is called as original VSRR.

After obtaining the VSRR, the popular image representation in image retrieval is nearest neighbor vector quantization (VQ) which is based on the BOV model. To improve the discriminative power of traditional BOV model we apply Group Sparse Coding (GSC) algorithm. This technique help us to improve our result as compare to traditional BOV model such as lower reconstruction error, lower storage space etc.

III. LITERATURE SURVEY

For evaluation metric we use mean average precision (MAP). We compute its average precision (AP) for each query image, then we take a mean value of all query images.

Advantages over traditional systems :
- Achieves lower reconstruction error rate.
- Economizes the storage space in encoding image descriptors.
- The sparse representation is simple and captures salient properties of images.
- Fast indexing and accurate retrieval.

Disadvantages in the system :
- Less MAP values causes defective output.
- Large dataset requires.
- Difficult to implement on smaller platform like android application.

IV. INDEXING AND RETRIEVAL TECHNIQUES

E. First Inverted File Structure

To retrieve images from large scale image retrieval system is a critical factor. For this inverted file index structure technique is used.
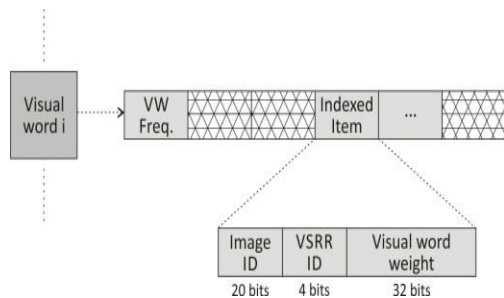

Fig. 1 First Inverted File Structure

To retrieve the result first search candidate VSRRs from the dataset and to refine the result via relative saliency ordering constraint. For efficiency this technique uses index structure with a bilayer inverted file. There are two inverted files, first preserves the VSRR information and second stores the saliency order of visual words in each VSSR. By executing the first inverted file we get the ID of candidate VSRR and image which is direct input to the second inverted file.

Figure 1 shows first inverted file structure. For each visual word in the dictionary D. this structure stores the list of VRSS which contains visual word occurs and its term weight. In figure "VW freq." is the sum of weight of visual word *i* which is obtain by the code of the visual word *i* by GSC for one VSRR. This file utilizes sparseness to index images and enables fast searching of candidate VSRRs.

F. Second Inverted File Structure

Figure 3.2 shows second inverted file structure. This structure stores the information of each VSRR. "VSRR area" is the pixel count of VSRR. "VW count" is number of visual words in VSRR. "VWi" is ID of the visual word I. These visual words are arranged according to their saliency value in ascending order. Dictionary D is different in first and secod inverted file. Dictionary D is obtained by a hierarchical K means clustering.
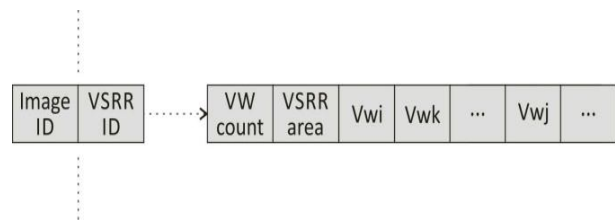

Fig. 2 Second Inverted File Structure

V. IMAGE DATASET

Sr. No.	Name of dataset	No. of images	No. of distractors
1	Public internet partial duplicate (PDID)	2000	30,000
2	Large Scale partial duplicate images (PDID1M)	10000	1 million
3	Mobile Dataset	300	10,200
4	Caltech256(50)	4988	-

TABLE I
IMAGE DATASET

For study and comparison between the datasets we compare it with four duplicate image retrieval schemes Soft BOV, Bundled feature, Bundled + and VLAD We take input images as Ganesha, KFC logo, Microsoft logo and Disney logo and by observation we can say that partial duplicate image retrieval using visual attention and saliency feature is better than other approaches. In case of Ganesha picture the results are not satisfactory for our system because many Ganesha images in the ground set dataset are detected as the VSRRs because of photoshop alteration these faces giffer and thus reduce the map[9].
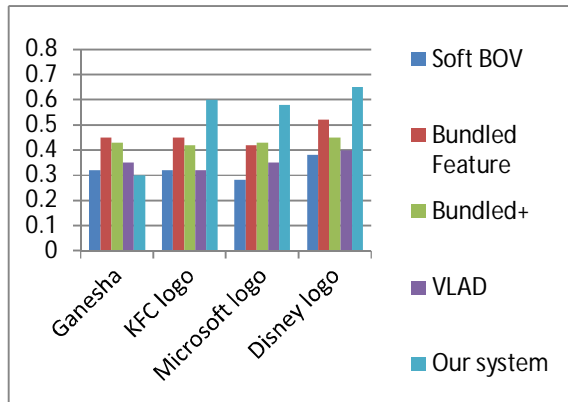
Fig. 3 MAP values comparison of Image dataset with existing system.

VI. SYSTEM FEATURES

G. Graphical User Interface : For the basic GUI of this application we create an android application which contains the button of selecting and uploading image in it. With the help of these buttons we can upload the image to the server.

H. Visual Attention Analysis : This selection appears to be implemented in the form of a spatially circumscribed region of the visual field, also called as focus of attention, which scans the scene both in a rapid, bottom-up, saliency driven and task independent manner as well as in a slower, top-down, volition-controlled and task dependent manner.

I. Group Sparse Coding : Group sparse coding is applied to visual contents of the file. The contents are separated by saliency regions. Visual content and sparse code is attached with the contents and it is inverted with first inverted filter with the matrix operation[7].

J. Byte to Pix Generator : This module is used to convert the bytes which are in the form of sparse code with image content and these contents has to be converted in to the pixels.

K. VSRR extraction and Group Sparse Coding : In this module we extract the image from the number of images and create a group of sparse coding with VSRR extracted images.

VII. CONCLUSION

We present a method for saliency computation based on an image abstraction by using contrast based saliency measures, Our filter based formulation allows for efficient computation and produces per-pixel saliency maps, with the currently best performance in a ground truth comparison.

ACKNOWLEDGMENT

REFERENCES

[1] Liang Li and Shuqiang Jiang, "Partial image retrieval via saliency guided visual matching," IEEE computer society, pp. 13_23

[2] Z. Wu et al., "Adding Affine Invariant GeometricConstraint for Partial-Duplicate Image Retrieval,"Proc. 20th IEEE Conf. Pattern Recognition (ICPR), IEEE CS, 2010, pp. 842_845.

[3] J. Sivic and A. Zisserman, "Video Google: A TextRetrieval Approach to Object Matching in Videos,"Proc. 2003 IEEE Conf. Computer Vision (ICCV),vol. 2, IEEE CS, 2003, pp. 1470_1477.

[4] M. Chen et al., "Global Contrast Based Salient Region Detection," Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS, 2011, pp. 409_416.

[5] H. Jegou et al., "Aggregating Local Descriptorsinto a Compact Image Representation,"Proc. 2010 IEEE Conf. Computer Vision andPattern Recognition (CVPR), IEEE CS, 2010,pp. 3304_3331.

[6] W. Zhou et al., "Spatial Coding for Large Scale Partial-Duplicate Web Image Search," Proc. Int'l Conf. Multimedia, ACM, 2010 pp. 510_520.

[7] D. Qin et al., "Hello Neighbor: Accurate Object Retrieval with k-Reciprocal Nearest Neighbors," Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS, 2010, pp. 777_784.

[8] Z. Wu et al., "Bundling Features for Large Scale Partial-Duplicate Web Image Search," Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS, 2009, pp.25_32.

[9] F. Perronnin et al., "Large-Scale Image Retrieval with Compressed Fisher Vectors," Proc. 2010 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS, 2010,pp. 3384_3391.

[10] J. Philbin et al., "Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 08), IEEE CS, 2008.

[11] H. Wu et al., "Resizing by Symmetry-Summarization," ACM Trans. Graphics, vol. 29, no. 6, 2010, article no. 159.

[12] P. Felzenszwalb and D. Huttenlocher, "Efficient Graph-Based Image Segmentation," Int'l J. Computer Vision, vol. 59, no. 2, 2004, pp. 197_181.

[13] D.G. Lowe, "Distinctive Image Features fromScale Invariant Keypoints," Int'l J. Computer Vision, vol. 60, no. 2, 2004, pp. 91_110.

[14] X. Wang et al., "Contextual Weighting for Vocabulary Tree Based Image Retrieval," Proc. 2011 IEEEConf. Computer Vision (ICCV), IEEE CS, 2011,pp. 209_216.

[15] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts," ACM Trans. Graphics, vol. 23, no. 3, 2004, pp. 309_314.