ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

Enhanced and Efficient Image Retrieval via Saliency Feature and Visual Attention

Anand K. Hase, Baisa L. Gunjal

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 domains like facebook, instagram and whatsapp etc. 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 retrieval technique is similar to object based image retrieval. This technique is compare with text retrieval system by using the bag of visual words (BOV). Typically no any spatial information is used to retrieve image, so this approach is not workable in background noise. Typically 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.

***Index Terms*—Partial duplicate image, Bags of visual words (BOV), Image retrieval, Visual Attention, Saliency feature, Visually Salient and rich region (VSRR).**

I. INTRODUCTION

In this paper, we propose a partial duplicate image retrieval scheme based on nearest saliency visual matching. We abstract visually salient and rich region (VSRR) from the images. We represent 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.

Fig. 1 Set of partially duplicate web images

II. PARTIAL DUPLICATE IMAGE RETRIEVAL ARCHITECTURE

We propose partial duplicate image retrieval scheme based on visual attention and saliency feature. 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. To accelerate the process of retrieval we proposed an efficient algorithms which minimizes storage space and computation time. We also apply experiment on different database for partial duplicate image retrieval which shows the efficiency and effectiveness of our approach.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

A. VSRR Generation

VSRR is a specific region from image which having rich visual contents and visual saliency. 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 Construction

Sensitive unit is defined as image patch that corresponds to the center of field which willing to accept new fields around it. For that we used a 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. 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.

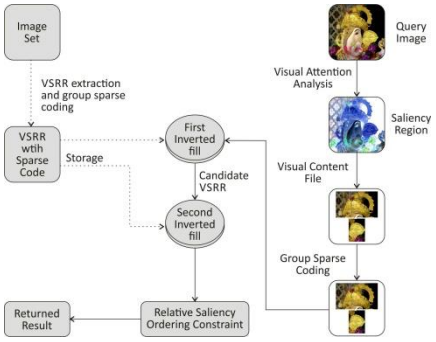

Fig. 2 Partial duplicate image retrieval scheme

On the L*a*b color space the color histogram is built, then the sensitive unit r_k, 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. This will defined as the term

$$S(r_k) = \sum_{r_k \neq r_i} \exp\left(\frac{D_s(r_k, r_i)}{\sigma_s^2}\right) w(r_i) D_r(r_k, r_i) \quad \ldots.(1)$$

Where, $\exp\left(\frac{D_s(r_k, r_i)}{\sigma_s^2}\right)$ is the spatial weight and $\left(\frac{D_s(r_k, r_i)}{\sigma_s^2}\right)$ is the spatial distance between the centroid of perceptive unit r_k and r_i and $w(r_i)$ is the number of pixels in the region r_i. σ_s controls the spatial weight. Values of σ_s are affects spatial weight. $(D_s(r_k, r_i))$is the color distance between r_k and r_i.

D. Generation of VSRR

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. 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 grab cut. Grab cut 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.The amount of visual content in the VSRR is measured as

$Score = \sum_{i=1}^{K} \frac{1}{N} \times ni$Where *K* is the size of dictionary and N and *ni*,are the numbers of visual word I in this database and VSRR respectively. 1/N represents the informative ness of the visual word and *ni* represents the repeated structure in the VSRR. After obtaining the VSRR, the popular image representation in image retrieval is

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

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. RELATIVE SALIENCY ORDER CONSTRAINTS

If we ignore geometric relationship it limits the discrimination power of the BOV model. To propose a relative saliency ordering constraints we have to first find the matching pairs between VSSRs. Suppose query VSRR *q* and candidate VSRR *c* have n numbers of matching visual words.VSRR(q) = $\{V_{q1}, \ldots, V_{qn}\}$, and VSRR($c$) = $\{V_{c1}, \ldots, V_{cn}\}$ and V_{qi} and V_{ci} are the visual matching words. So $S(q) = \{\alpha_{q1} \ldots \alpha_{qn}\}$ and $S(c) = \{\alpha_{c1} \ldots \alpha_{cn}\}$ represents the saliency value in q and c respectively. We construct the relative matrix called saliency relative matrix (SRM)

$$SRM = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix} \qquad r_{ij} = \begin{cases} 0 & \alpha_i > \alpha_j \\ 1 & otherwise \end{cases}$$

r_{ij} is defined by comparing the saliency values α_i and α_j of v_i and v_j in VSRR. Each visual word is compared with other visual words. The inconsistency of query SRM and the candidate SRM is measure by the Hamming distance

$$Dis = \left| SRM_q \oplus SRM_c \right|_0 \qquad \ldots.(2)$$

Where $|.|_0$ (I_0 norm) is the total number of nonzero elements.

IV. INDEXING AND RETRIEVAL

To retrieve images from large scale image retrieval system is a critical factor. We introduced inverted file index structure.

Index structure - To retrieve the result we have to first search candidate VSRRs from the dataset and to refine the result via relative saliency ordering constraint. For efficiency we use 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.

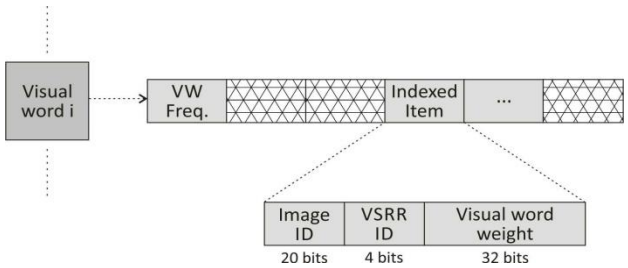

Fig. 3 First inverted file structure

1. *First inverted file structure*

Figure 3 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.

2. *Second inverted file structure*

Figure 4 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 second inverted file. Dictionary D is obtained by a hierarchical K means clustering.

V. RETRIEVAL SCHEME

We first have to find candidate VSRR. The process is similar like voting scheme. The score of all VSRR in the database is initialized to 0, then for each visual word j, we retrieve the list of VSRR from first inverted file. We

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

increase the VSRR score by the weight of visual word score L $(i) = \frac{\text{Weight of visual wo}}{\text{VW j freq.}}$ ….(3)

After processing of all the visual words the final score of VSRR *i* is product between the vectors of VSRR *i* and the query *q*. Now we have to compute the SRM inconsistency between the query and candidate VSRR.We define the total matching score *M(q,c)* $M(q,c) = M_v(q,c) + \lambda M_r(q,c)$

Where $M_v(q,$ is a visual similarity, and $M_r(q,$ is the consistency relative saliency constraint, which is equal to $1 - \left(inconsistency\left(SRM(q,c)\right)\right)$ is a weight parameter. After obtaining similarity we define similarity between query image I_q and the candidate image I_m as

$Sim\left(I_q, I_m\right) = \sum_{qi \in I_q} \frac{2 \times sqrt\left(R_{area}(q_i)\right)}{1+R_{area}(q_i)} \times M(q_i, n$ ….(4)Where q_i is the i[th] VSRR of image I_q.

VI. SYSTEM FEATURES (MODULES)

Module 1: 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.

Module 2: 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.

Module 3: 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.

Module 4: 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.

Module 5: 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. IMAGE DATASETS

Sr. No.	Name of dataset	Number of images	Number 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	-

For study and comparison between the datasets we compare it with four duplicate image retrieval schemes

1. Soft BOV – enhanced BOV method with softness where the number of nearest neighbor is set to 3, and σ^2=6,250.
2. Bundled feature – in this scheme SIFT and maximally stable extreme regions (MSER) are extracted from images and bundled into groups.
3. Bundled+ – In this scheme bundled feature scheme is improved by adding an geometric constraints which shares common SIFT vocabulary of 5,000 visual words.
4. VLAD – it stands for state-of-the-art vector of locally aggregated descriptor (VLAD) that is derived from both BOV and the fisher kernel.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

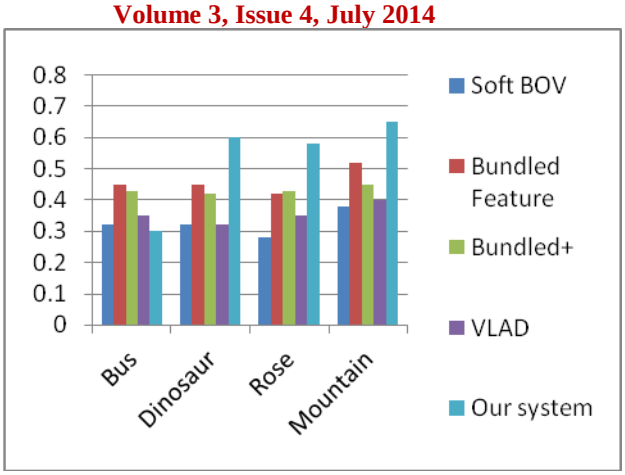

Fig. 6 Comparison of five approaches with mean precision for partial duplicate image retrieval

We take input images as Bus, Horse, Dinosaur, Rose, and Mountain 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 facesgiffer and thus reduce the map.

ALGORITHM OF LOCAL SELF SIMILARITY DESCRIPTOR (LSSD)

Input: Image I, interest points $K_{i,}$ i=1, K
Output: the set of LSSD
1: for each interest point *Ki*do
2: Get the circular region centered at *Ki* with redius.
3: Get LBP histogram H' of a square centered at k_i
4: for each point $P_{i,j}$in the circular region do
5: Compute LBP histogram $H_{i,jl}$of a square centered at $P_{i,j}$
6: Get the similarity by comparing H' with $H_{i,j}$
7: end for
8: Get the circular self-similarity map
9: Transform the map into polar coordinator
10: Quantize the polar coordinator into $S_i$with bins
11: end for

VIII. 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.

IX. FUTURE SCOPE

In the future we plan to develop an android application for the partial duplicate image retrieval system. We also try to develop the application that capture the image from mobile and upload that image, after uploading we getting detail information about uploaded image product or location or any related information etc. In existing system we try to develop more sophisticated techniques for image abstraction, including robust color or structure distance measures will be beneficial.

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 Geometric Constraint for Partial-Duplicate Image Retrieval,’’Proc. 20th IEEE Conf. Pattern Recognition (ICPR), IEEE CS, 2010, pp. 842_845.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

[3] J. Sivic and A. Zisserman, ‘‘Video Google: A Text Retrieval 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 Descriptors into a Compact Image Representation,’’Proc. 2010 IEEE Conf. Computer Vision and Pattern 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. Hutten ocher, ‘‘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 from Scale Invariant Key points,’’ 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 IEEE Conf. Computer Vision (ICCV), IEEE CS, 2011, pp. 209_216.

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

[16] X. Shen et al., ‘‘Object Retrieval and Localization with Spatially-Constrained Similarity Measure and k-NN Re-ranking,’’ Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS, 2012, pp. 3013_3020.