ISSN: 2321-7782 (Online)
Volume 4, Issue 2, February 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Image Retrieval Using Vector Of Locally Aggregated Descriptors

Vijay Vitthal Thitme[1]
Master Of Engineering Scholar
Indore Institute of Science and Technology
Indore - India

Prof. Akanksha Varghese[2]
Assistant Professor
Indore Institute of Science and Technology
Indore - India

Abstract: Partial duplicate image retrieval is very powerful and important task in the real world applications such as landmark search, copyright protection, fake image identification. In the internet applications users continuously upload images which may be partially duplicate images on the domains like social sites orkut, facebook, and related applications etc. The partial image is nothing but segment of whole image, and the various methods of transformations are scaling, resolution, illumination, rotation and viewpoint. This method is considered as of much more valuable by different real world aspects and has motivated towards this study. The method of retrieving images which is based on the object methods generally uses the whole image as the query image. In object based image retrieval methods usually use the whole image as the query. This method is compared with text system by using the bag of visual words (BOV) Generally there may be lots of noise on the images so it is impossible to perform operations on large scale dataset of images. This approach is not much more used because no any spatial data is used to retrieve the efficient images.The art of image retrieval methods represent image with a large dimensional vector of visual words by making quantization of local features, such as Scale Invariant Feature Transform, solely on the descriptor space. Quantization of the Local features to visual words are done firstly in descriptor space and then in orientation space. Local Self-Similarity Descriptor (LSSD) value is used which is used to captures the internal geometric layouts in the local text self-similar regions near interest points.

Keywords: Partial duplicate image, Bags of visual words (BOV), Image retrieval, Visual Attention, Saliency regions, Visually Salient and rich region (VSRR) large scale quantization, spatial embedding, Local Self Similarity Descriptor value, Semi-Relative Entropy, Scale Invariant Feature Transform, Content Based Image Retrieval (CBIR), SIFT,VLAD.

I. INTRODUCTION

The popularity of smart mobile phones and text and information sharing Web sites have produced large number of partial duplicate Web images and texts, makes the need for efficiently and effectively detecting partial duplicate images from large amount of datasets. CBIR methods have achieved greatest improvement in retrieving duplicate images from the large datasets of images. In this paper we proposed the partial duplicate images retrieval method based on the saliency region mapping using different region mapping methods and also using visually matching. In this we are using VSRR to abstract Visually Salient and Rich Region from the images. The Bags-Of-Visual words model is used to represent the Visually Salient and Rich Region. Group sparse coding are used to obtain lower reconstruction errors and sparse representation at regional level. We can also compare our result of image retrieval performance with other image database and show the robustness and efficiency of our approach of image retrieval. To effectively retrieve the partial duplicate images from the images of large set that have very small portion of duplicate regions with various transformations operations. Given query image and our target is to search for its nearly duplicated or partially duplicated images.

II. Duplicate Image Retrieval Architecture

We are presenting partial duplicate image detection technique based on visual matching using VSRR and saliency region mapping. Bag –of-visual words model and group sparse coding is used to represent visually Salient Rich Region to find the region which is visually rich. To understand the process of images retrieval we present an efficient and very effective algorithm which requires less processing, less computational effort, less time and less storage space to save it into the memory. We also tried to implement our approach with different database for partial duplicate image retrieval which shows the more accurate and effective results of achieving the partial duplicate images. This approach also achieves the low reconstruction errors and more powerful results. Saliency Matrix of candidate image and query images are used to calculate X-OR results of images. The first inverted file uses the VSRR information. Second inverted file is used to store the saliency ordering of the visual words in each VSRR. Fig. 1 shows sample set of duplicate images.

Fig. 1 Set of duplicate images

1. Producing VSRR

VSRR is retrieval unit in this approach of duplicate image detection method. VSRR is the region in which it has capability to calculate region of the images which has visually rich regions. VSRR producing process is mainly divided into four different dependent parts: Perspective unit construction, saliency map creation, VSRR generation and finally selection of ultimate VSRR. The final image decomposed into the sets of VSRR. In VSRR creation both candidate and query images are used as to generate VSRR information of both. After map is generated the saliency ordering is calculated using ordering constraints.

2. Defining Related Perspective Unit

Perspective unit is defined as image patch that is related to the field centroid point which accepts new fields around this field. For this purpose we used a graphical base partition algorithm which combines the smaller size patches with the likely appearances and small minimum spanning tree weights. Sensitive unit is important element for saliency map generation method and visual matching methods to find the duplicate images.

3. Creation of Saliency Map

The typical image region which having high and strong contrast with their surrounding images takes human attention towards that particular point .when visual attention is used at that point saliency region mapping play significant role. The region having more attraction which is highly contrast with its nearest region than the high contrast with its far region. Saliency map are computed on the basis of spatial regions relationship with the contrast region. This saliency region map generation technique is used to separate the object from their surroundings.

The L*a*b color space method is coordinate based method in which L, a, b are the co-ordinates. The the L*a*b color space method to obtain the color histogram is built, then the perspective unit r_k, its saliency value is calculated by making the measurement of color contrast of other perceptive units which are available in the image. The weight of spatial region relationship is nothing but as increase the effect of nearest regions and decrease the effect of far regions.

This can also be defined as the term

$$S(r_k) = \sum_{r_k \neq r_i} \exp\left(\frac{D_S(r_k, r_i)}{\sigma_S^2}\right) w(r_i) D_r(r_k, r_i)$$

Where, $\exp\left(\frac{D_S(r_k,r_i)}{\sigma_S^2}\right)$ are the weights of spatial information and $\left(\frac{D_S(r_k,r_i)}{\sigma_S^2}\right)$ is the spatial distances between the centre point of perceptive unit $r_k and\ r_i$ and $w(r_i)$ is the quantity 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 $r_k and\ r_i$.

4. Construction of VSRR

After successfully generating saliency map we can move to next step of calculating the salient regions this can be done with the help of saliency segmentation. Then original VSRR is obtained and select the VSRR that will contain large amount of visual content After filtering we select the VSRR that contains large numbers of visual content.

Binarization of map is done means using thresh-hold value we can divide the saliency map into background and initial regions. 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 bunch 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 dictionary sizes 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 nearest neighbour 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, less calculation effort, less processing ,lower storage space and more perfect and accurate results within short period of times etc.

5. Ordering using Visual words in VSRR

To find ordering in the visual words in VSRR we have to find saliency matrix of the candidate and query images and then by finding matching pair between visual words

This matrix is called as Salient Relative Matrix

$$SRM = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix}$$

Then we can obtain the results by making XOR of the two matrices

$r_{ij}$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

$$Dis = |SRM_q \oplus SRM_c|_o$$

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

6. Algorithm to retrieve duplicate images

Input – I_Q

Output - $I_i = \{ I_1, I_2, I_3, \ldots, I_n \}$

1. Clear all previous analysis.
2. Insert dataset of I_n images for analysis.
3. Retrieve SRM contains the visual saliency values of Dataset images.
4. input the query image I_Q.
5. Analyse the query image and retrieve its saliency values.
6. *SRM* for query matrix and *SRM+* for candidate matrix is generated.
7. First inverted & Second inverted file structure gives saliency order of *VWi.*
8. *SRMq XOR SRMc* generates XOR result of two Saliency related matrix.
9. Calculate Precision, Recall, *AveP(q), & MAP values.*
10. Generates partial duplicate images.
11. Stop

III. Index Structure and Access Scheme

To retrieving images and text from large data set of images is complex and critical method. We used first and second inverted file index structure to achieve this scheme. To achieve the result we have to search VSRR information about candidate Saliency Relative Matrix and result is refined via saliency ordering constraints. There are two inverted files, in the first inverted file VSRR information and second file stores the order of visual word in each and every VSRR

1. First inverted file index internal format

By using the first inverted file we get the image ID and VSRR ID. The first file is direct input or index to the second inverted file. Figure indicates first inverted file structure. For each visual word in the dictionary 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. Fig. 2 shows first inverted file and Fig. 3 shows second inverted file respectively.

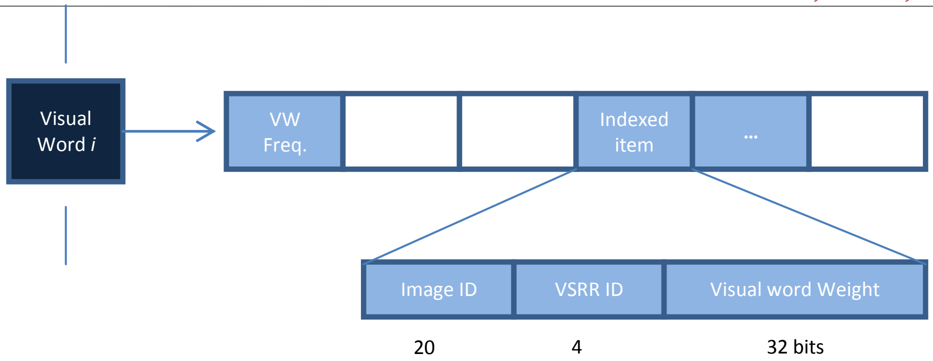

Fig.2 First inverted file internal details

2. Second inverted file index format

This second file uses input as the first file and then generate ordering information about visual words in every VSRR Figure 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.

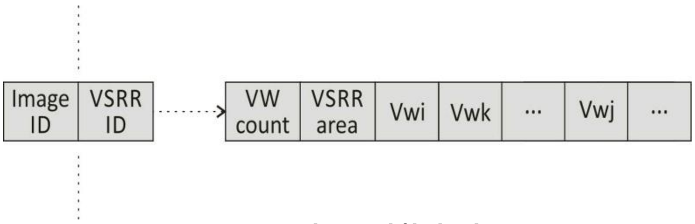

Fig. 3 Second inverted file details

IV. Visual Saliency

Visual saliency at lowest level of scene can be generally defined as by using local contrast of luminance opposing colours or more focusing on edges of the images. Information carrying capacity of region visually salient carry more information than physically observed salient regions. Physical observed features keeps more attention in bottom up fashion. This is early stage used in image processing. In this method biology based computational model is used to construct the saliency map of visual scenes. This means concentration is on the nearest locations.

V. Eye Movement Tracking, Recording and Analysis

The eye-tracking process was divided into10 number of blocks with break of self-inserted type in between. Six full sequences block of images were presented for eye movement and tracking system. Eye position was saved with an infrared remote Eyelink structure (SR Research) eye tracker, sampling position at 500 Hz. Both eyes locations were recorded but only that data from better calibrated eye was analysed and saved. For every experimental blocks calibration and validation procedures are used at the beginning. Eyelink 1000 algorithm was used to detect the fixations Average calibration error of the analysed eye was 0.398.

VI. Vector Representation of Image

This section gives the overview of the vector representation of the images there are different approaches of this representation like Bag-Of-Features, Fisher Kernel and VLAD. We will see detailed description about the VLAD (vector of locally aggregated descriptors).

1. VLAD (vector of locally aggregated descriptors)

We present the images using vector representation in which aggregates descriptors based on a locality criterion in feature space. This method can be treated as simplified version of the Fisher kernel. As for BOF, we first learn a codebook C = {c1,

...ck} of k visual words with k-means. Every local descriptor x is related to its nearest visual word ci = NN(x). The concept VLAD descriptor is to accumulate, for each and every visual word ci, the differences x−ci of the vectors x assigned to ci. This characterizes the distribution of the vectors with respect to the centre point.

Fig. 4 Images and their corresponding VLAD

Assuming the local descriptor to be d-dimensional, the dimension D of our representation is D = k × d. In the following, we represent the descriptor by vi,j , where the indices i = 1 . . . k and j = 1 . . . d respectively index then visual word and the local descriptor component. Hence, a component of v is produced a sum over all the image local descriptors:

$$v_{i,j} = \sum_{x \text{ such that } \mathrm{NN}(x)=c_i} x_j - c_{i,j}$$

where xj and ci,j respectively denote the jth component of the descriptor x considered and of its corresponding visual word ci. The vector v is subsequently L2-normalized by v := v/||v||2 . Fig. 4 shows images with their VLAD.

VII. LITERATURE SURVEY

Following are features of existing system:

- In SIFT (scale invariant feature transform) it uses multi description of images.
- VLAD (Vector of locally aggregated descriptor) significantly outperform for BOV of the same size.
- In saliency guided visual matching author use L*a*b color scheme to measure its contrast.
- In fisher vector technique memory and computational cost is not feasible.
- BOV (Bags of Visual Words) is not useful with large scale database.
- Emotional contents like eye moment is attracts attention more than visually salient feature.

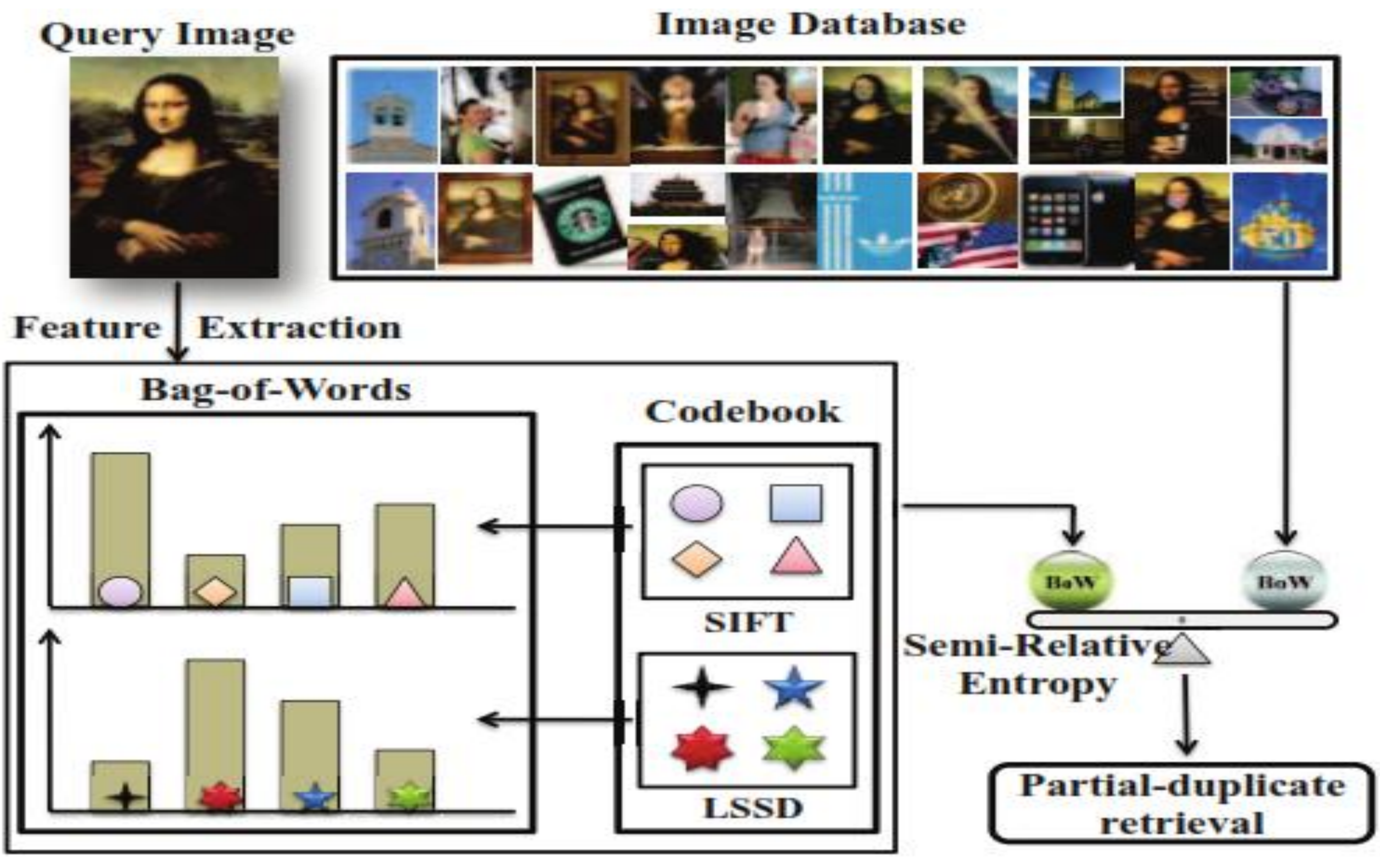

Fig. 5 : Multi Description of local interest point for partial duplicate image retrieval

VIII. SYSTEM COMPONENTS

MODULE 1 : SALIENCY MAP GENERATION

MODULE 2: SALIENCY RELATIVE MATRIX

MODULE 3 : FIRST INVERTED FILE STRUCTURE

MODULE 4: SECOND INVERTED FILE STRUCTURE

Module 5: Result generation using VSRR

IX. IMAGE SEGMENTATION PROCEDURE

Input : Graph $G = (V,E)$, with n vertices and m edges.

Output : Segmentation of V into components $S = (C1,\ldots,Cr)$

0:Sort Einto $\pi = (O_1, \ldots, O_m)$, by non-decreasing edge weight.

1: Start with a segmentation S^0, where each vertex $v_i$is in its own component.

2: Repeat step 3 for $q=1,\ldots,m$.

3: Construct S^qgivenS^{q-1} as follows. Let $v_i$and$v_j$denote the vertices connected by the q-thedge in the ordering, i.e., $o_q=(v_i,v_j)$.If v_i and $v_j$are in disjoint components of $S^{q-1}$and $w(o_q)$ is small compared to the internal difference of both those components, then merge the two components otherwise do nothing. More formally, let c_i^{q-1} be the component of $S^{q=1}$ containing v_i and c_j^{q-1} the components containing vj. If $c_i^{q-1}$and c_j^{q+-1}. Otherwise $S^q= S^{q-1}$

4: Return $S = S^m$.

X. Conclusion

We present the contrast based saliency computation methods, which generates spatially consistent high quality saliency maps at the cost of reduce computational efficiency. We evaluate our methods on the largest publicly available data set and compared our system with other methods. We present a method for saliency computation on image abstraction into contrast based saliency. It allows efficient computation with the best performance system.

XI. Future Work

We also try to retrieve the result that is based on the shape of the segmented portion. We also try to develop it for android operating system by which any person who use android device can easily use this application on large scale. We will also try to find duplicate text using this scheme.

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.
16. X. Shen et al., ''Object Retrieval and Localization with Spatially-Constrained Similarity Measure and k-NN Re-ranking,'' Proc. 2012 IEEE Conf. ComputerVision and Pattern Recognition (CVPR), IEEE CS,2012, pp. 3013_3020.
17. Torralba, R. Fergus, and Y. Weiss. Small codes and large databases for recognition. In CVPR, June 2008.
18. L. Torresani, M. Szummer, and A. Fitzgibbon. Learning query-dependent prefilters for scalable image retrieval. In CVPR, June 2009.
19. J. van Gemert, C. Veenman, A. Smeulders, and J. Geusebroek. Visual word ambiguity. PAMI. To appear.
20. Y. Weiss, A. Torralba, and R. Fergus. Spectral hashing. In NIPS, December 2008.
21. S. Winder and M. Brown. Learning local image descriptors. In CVPR, June 2007.
22. S. Winder, G. Hua, and M. Brown. Picking the best Daisy. In CVPR, June 2009.