

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Text and Image Retrieval via Saliency Feature and Visual Attention

Kadam Vijay Murlidhar¹

Student

Department Of Computer Science and Engineering,
Alwar,(RTU), India

Sunil Gupta²

Associate Professor

Siddhi Vinayak Collage of Science and Higher Education
Alwar (Rajasthan), India

Abstract: In the real world applications such as copy protection, fake image detection, landmark search, partial duplicate image retrieval is very needful to any user. Regular users of the internet regularly upload images, photograph which are partially duplicate on the domains like facebook, instagram and whatsapp etc. The partial image is the part of whole image, and the various kind of transformation involves scale, resolution, illumination, and viewpoint. Similar technique is required 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, 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), 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.

A. 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 algorithm 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.

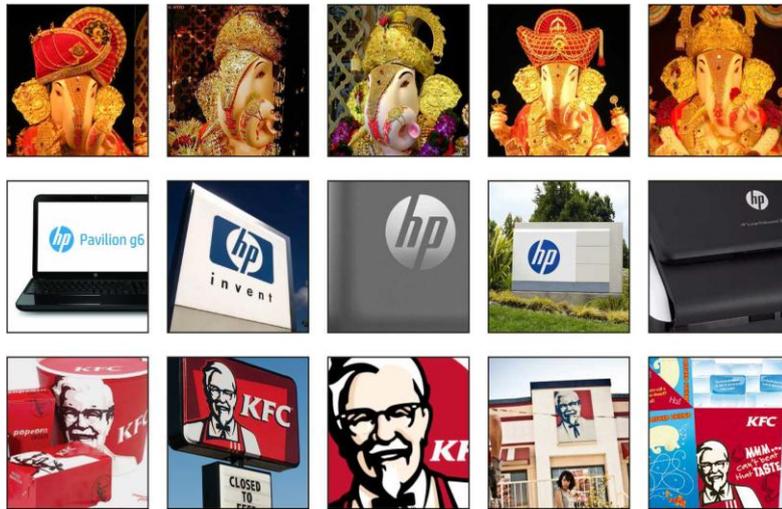


Fig. 1 Set of partially duplicate web images

1. 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.

2. 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.

3. 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.

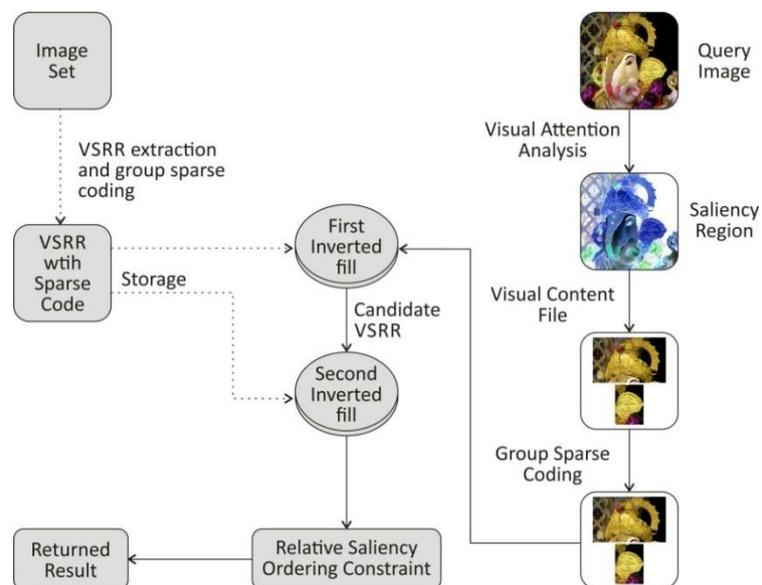


Fig. 2 Partial duplicate image retrieval scheme

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

$$S(r_k) = \sum_{r_k \neq r_i} \exp\left(\frac{D_S(r_k, r_i)}{\sigma_s^2}\right) w(r_i) D_r(r_k, r_i) \dots(1)$$

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 .

4. 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 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.

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 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.

B. 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 have n numbers of matching visual words.

$VSRR(q) = \{V_{q1}, \dots, V_{qn}\}$, and $VSRR(c) = \{V_{c1}, \dots, V_{cn}\}$ and V_{qi} and V_{ci} are the visual matching words. So $S(q) = \{\alpha_{q1} \dots \alpha_{qn}\}$ and $S(c) = \{\alpha_{c1} \dots \alpha_{cn}\}$ represents the saliency value in q and c respectively. We construct the relative matrix called saliency relative matrix (SRM)

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

$$r_{ij} = \begin{cases} 0 & \alpha_i > \alpha_j \\ 1 & otherwise \end{cases}$$

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

$$Dis = |SRM_q \oplus SRM_c|_0 \dots(2)$$

Where $|\cdot|_0$ (l_0 norm) is the total number of nonzero elements.

II. 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.

III. INDEXING AND RETRIEVAL

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

A. 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 VSRR. By executing the first inverted file we get the ID of candidate VSRR and image which is direct input to the second inverted file.

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.

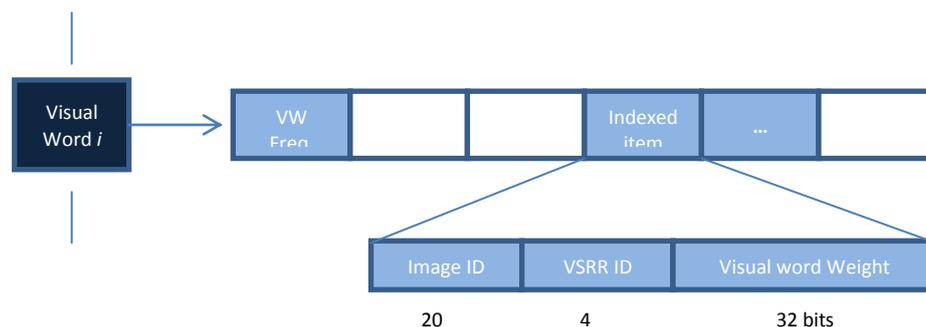


Fig. 3 First inverted file structure

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.

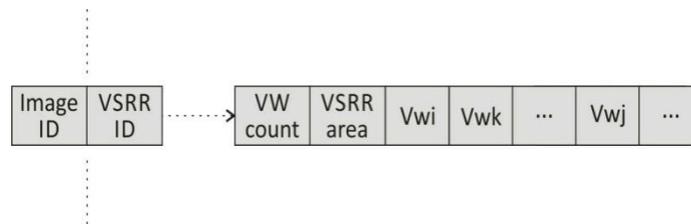


Fig. 4 Second inverted file structure

IV. 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 increase the VSRR score by the weight of visual word score L

$$(i) = \frac{\text{Weight of visual word } j}{\text{VW } j \text{ freq.}} \quad \dots(3)$$

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

$$\text{Sim}(I_q, I_m) = \sum_{q_i \in I_q} \frac{2 \times \text{sqrt}(R_{\text{area}}(q_i))}{1 + R_{\text{area}}(q_i)} \times M(q_i, m_i) \quad \dots(4)$$

Where q_i is the i^{th} VSRR of image I_q .

V. SYSTEM FUTURE (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.

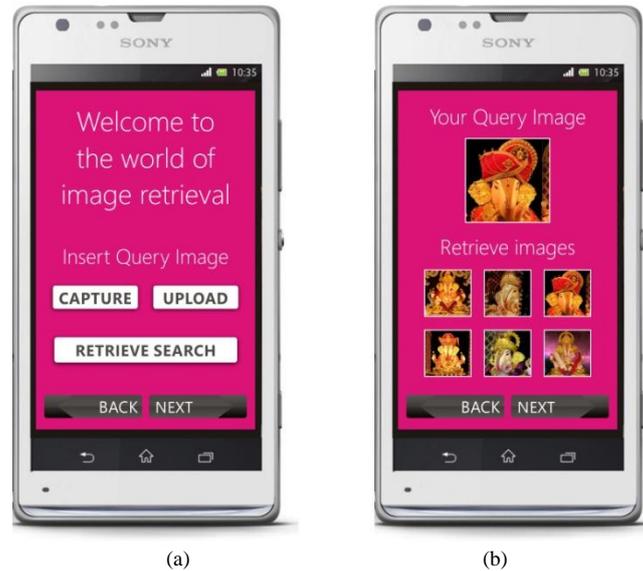


Fig. 5. Android application GUI a) Insert image screen b) Retrieve image screen

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.

B. Segmentation Algorithm

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

Output : Segmentation of V into components $S = (C_1, \dots, C_r)$

0: Sort E into $\pi = (O_1, \dots, O_m)$, 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 $q=1, \dots, m$.

3: Construct S^q given S^{q-1} as follows. Let v_i and v_j denote the vertices connected by the q -th edge 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 v_j . If c_i^{q-1} and c_j^{q-1} . Otherwise $S^q = S^{q-1}$

4: Return $S = S^m$.

C. 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 K_i
- 2: Get the circular region centered at K_i with radius.
- 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 $H_{i,j}$ 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_j with bins
- 11: end for

VI. 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.

VII. FUTURE WORK

In the future we plan to develop an android application for the partial duplicate image retrieval system. We also try to develop the application who 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. S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
2. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
3. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.
4. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.
5. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.