

Image retrieval with hand drawn draft based on saliency detection

Xiaofei Sun ^a, Wenwen Pan ^b, Xia Wang ^c

College of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China

^a34740935@qq.com, ^bpanwenwen@vip.qq.com, ^cWangxia1026@163.com

Keywords: saliency map; image retrieval; image segmentation; color sparse histogram.

Abstract. An image retrieval algorithm with hand drawn draft based on saliency detection was proposed and improved. First, salient regions were obtained by an image segmentation method, next region color sparse histograms were computed, and then the histogram was used to compute each region color contrast. The saliency map was improved by an enhancement and suppression method. Salient object was obtained by eliminating low contrast regions and was divided into blocks equally. The average gray value of each block and spatial relationships were feature data for searching. Search process was filling hand draw contour and extracting the feature data and comparing the feature data on disk, then the difference was combined with the spatial weights. Finally the results are sorted, and the pictures are similar if the difference is small. This method can extract good salient objects and obtain accurate retrieval results.

Introduction

With the development of science and the living demand growth, it becomes more convenient to create massive image data. Pictures contain a large amount of information which is various and complete. Image retrieval is the premise of effective use of image data. Image retrieval on smart devices is mainly based on text, such as name, keywords, description text. This approach cannot express the complete information of the image^[1]. Hand-drawn sketches image retrieval is more imaginative and creative than text mode on smart devices, and it has a special expression way to meet human's demand. Hand-drawn sketches using Fourier shape is more efficient than the stroke description way and insensitive about input sequence, so it can obtain better retrieval results^[2]. Retrieval method based on structure diagram reduces the dimension of the image contour effectively with less image shape losing, and provides rotation, translation and scaling invariance.

The essence of hand-drawn sketches image retrieval is a kind of content based image retrieval. Existing content based image retrieval methods express the image semantics mostly through the extraction of image low-level feature (color, texture and shape)^[4], and don't allow computers to understand images in human's cognitive way. There are mainly two kinds of the latest research, one is the use of the weighted sum approach to comprehensive utilization of various features^[5, 6, 7], another is to construct mixed feature^[8, 9, 10, 11, 12]. Color image segmentation with improved watershed algorithm calculates saliency of each sub region with the regional spatial attention model. This method tends to comprehend salient object as a whole, so it causes loss of edge details^[13]. If a color distributes widely or dispersedly in an image, it's unlikely included in the salient object. The specific color's global spatial distribution can be used to describe space saliency of the salient object. The preliminary classification of colors in the image through the K-means clustering algorithm is performed, then model color probability distribution using multivariate normal distribution^[14].

Existing search method based on content can overcome some shortcomings of text based search, but there are still problems such as high computational complexity, large storage quantity is large, low efficiency. This paper presents an algorithm which first gets a salient region and then extracts the salient region's similar characteristics. Image semantic contents focus on salient region, so it can be used to represent the image in image retrieval.

Saliency map based on image segmentation

First, the original image is segmented into regions. Then, each region's color histogram is calculated using algorithm mentioned in literature[4], the RGB component pixel color in each region of the original image were quantified for 12 values. Next, color histogram of each region is calculated. Colors are ranked according to frequency of occurrence. Low-frequency colors are replaced by similar high-frequency colors, and colors which frequency is 0 are removed, we obtain sparse histogram lastly. So images can keep basically the same visual effect with the original image, but greatly reduce the number of colors.

Then each color histogram is converted to Lab space, and color distances among different regions are calculated. The formula of color distance between regions r_m and r_n is as following:

$$D_r(r_m, r_n) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(c_{m,i}) f(c_{n,j}) D(c_{m,i}, c_{n,j}) \quad (1)$$

In particular, $f(c_{m,i})$ is the proportion of color i in region r_m , and the proportion is calculated in the whole picture not the region. This help saliency map better. $D(c_{m,i}, c_{n,j})$ is the Euclidean distance between color i in region r_m and color j in region r_n . N_1 represents color number of region r_m , and N_2 is similar.

$$S(r_k) = \sum_{r_i \neq r_k} \frac{w(r_i) D_r(r_k, r_i)}{\delta D_s(r_k, r_i)} \quad (2)$$

The farther between regions, the less effect about saliency. $S(r_k)$ is the saliency of region r_k , and $D_s(r_k, r_i)$ is the distance between region r_k and region r_i . $S(r_k)$ is inversely proportional to $D_s(r_k, r_i)$, so $1/\delta D_s(r_k, r_i)$ is used as the distance weight. δ is a constant used for regulating. The value of δ is 180, the effect is the best. $D_s(r_k, r_i)$ is the city block distance between centroids of regions which is described as following.

$$D_s(r_k, r_i) = |x_k - x_i| + |y_k - y_i| \quad (3)$$

x_k is the X coordinate value of region r_k centroid, and y_k is the Y coordinate value of region r_k . City block distance is more efficient. At the same distance, $w(r_i)$ effects on region r_k more if $w(r_i)$ has more pixels, so $S(r_k)$ is directly proportional to $w(r_i)$. At last the following formula is used to improve the saliency map.

$$S'(r_k) = \lambda S(r_k) G_{(x,y)} \quad (4)$$

$$G_{(x,y)} = \text{average}(r_{(x,y)}, g_{(x,y)}, b_{(x,y)}) \quad (5)$$

Our saliency maps perform as well as literature [4]. Without color spatial smoothing, our algorithm is more efficient. Some of our saliency maps are as Figure 1.

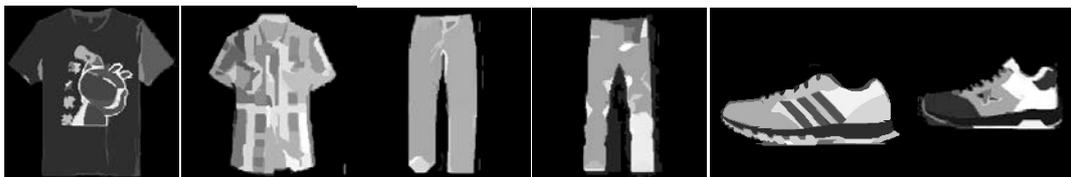


Figure 1 saliency map

Hand-drawn sketches image retrieval based on saliency map

Saliency $S'(r_k), k=1, 2, 3 \dots N$ are sorted, then the outermost regions are excluded, and the remaining regions $r_1, r_2, r_3 \dots r_m$ are shoes, blouses or trousers. Next, the remaining regions are treated as a whole, $R = r_1 \cup r_2 \cup r_3 \dots \cup r_m$. Minimum exterior rectangle of R can be found and proportionally divided

into 10×10 blocks $blk_i, i = 1, 2, 3 \dots 100$, and $R = blk_1 \cup blk_2 \cup blk_3 \dots \cup blk_{100}$. In this way, salient object information can be represented in proportion. If salient objects in different images are similar but have different size, the algorithm gets similar features. So the algorithm can be used to get perfect salient regions and completely retain salient object information.

All the pixel's gray value in block blk_i are added up to $gray_i$ which represents the gray value of block blk_i . Then $gray_i$ is decided by pixel number n_i , and the average gray value is kept in matrix m_i . When retrieving images, simply compare m_k with m_i .

$$D(m_k, m_i) = \sum_{blk_k=blk_i, t_{blk}} \frac{i}{t_{blk}} D_{blk}(gray_k, gray_i) \quad (6)$$

$$gray_i = \frac{c_i}{n_i}, i = 1, 2, 3 \dots m \quad (7)$$

$D_{blk}(gray_k, gray_i)$ is the average gray distance between block blk_k and blk_i . Considering the spatial relationship, the linear distance weight i/t_{blk} is used. t_{blk} is block number which value is 100, and i is block's identifier (1, 2, 3...100). The smaller the $D(m_k, m_i)$, the more similar the images.

Experiment and analysis

This paper selected 90 images from the major shopping websites. These images include shoes, blouses and trousers. In another word, there are 30 images per category. Images are retrieved using our algorithm. Some examples are as following. The three images above are the hand-drawn sketches, and the images below are retrieval results.

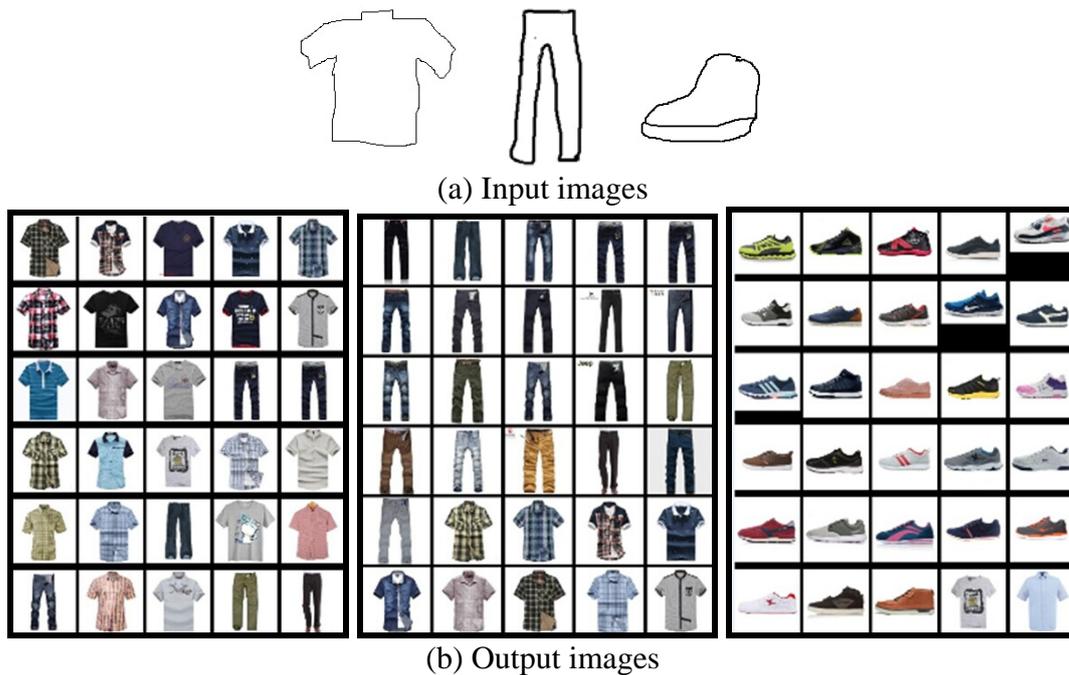


Figure 2 three retrieval examples

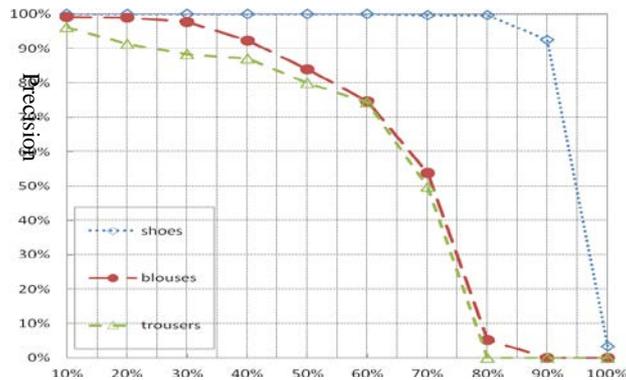
In order to evaluate the feasibility of this method more objectively, we calculate the precision in two ways. One is precision based on recall, the other is precision based on number of retrieval results. Precision is defined as right results divided by all retrieval results, and recall is defined as right results divided by all right results.

$$Precision = \frac{Result_{right}}{Result_{all}} \quad (8)$$

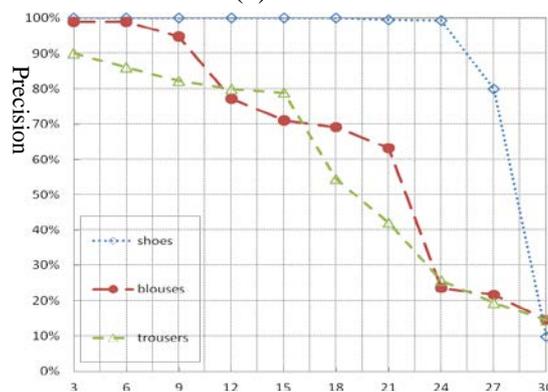
$$Recall = \frac{Result_{right}}{Right_{all}} \quad (9)$$

As shown in figure 3(a), with the increase of recall, the precisions of shoes and trousers always remain over 90%, and the precision of blouses remains over 50% in most cases. The precision of

blouses has a obvious decline, the reason is that blouses in our database are various and not given explicit categorizations and sketches describe shapes accurately, so blouses which shapes are quite different from sketches are excluded. As shown in figure 3(b), when retrieval image number is less than 20, the precision is good.



(a) recall



(b) recall

Figure 3 precision in two ways

Conclusion

A hand drawn draft image retrieval algorithm based on saliency detection was proposed, and then the saliency map was improved by an enhancement and suppression method. The salient regions are treated as a whole, and divided into 10×10 blocks. Each image's features are composed of blocks' average gray value and spatial information. The database contains images from shopping websites. The experiment provides good saliency map using our algorithm and retrieves images accurately. Precisions of each image in 10 cases are calculated, so 900 statistics are made to test the retrieval precision of this algorithm. Experiments show that similar images can be retrieved by this algorithm using hand-drawn sketches. The algorithm will be widely used in Electronic Business, costume designing and artistic designing.

Acknowledgments

Our work is supported by a project of Shandong province higher educational science and technology program (grant No. J12LN53) and a project supported by youth foundation of Zaozhuang University (grant No. 2011QN43).

Reference

[1] Huang Xianglin, Shen Lansun. Research on content based image retrieval technology. Chinese Journal of Electronics [J]. 2002, 30(7): 1065-1071.

- [2]Li Manwu,Sun Zhengxing,Gu Qingdong. Research on contour based hand-drawn sketches image retrieval [J]. Computer application.2005, 25(2):344-347.
- [3]Gao Zhuhong,Tang Jin,Luo Bin. Sketches image retrieval based on structural graph [J]. Computer technology and development. 2008, 18(3): 32-35.
- [4] Ming-Ming Cheng, Guo-Xin Zhang,et al.Global Contrast based Salient Region Detection[C]. IEEE International Conference on Computer Vision and Pattern Recognition, 2011.409-416.
- [5]Piamsa-nga P, Alexandridis N A, Srakaew S, et al. Multi-feature content based image retrieval [C]//Proceedings of International Conference on Computer Graphics and Imaging. Halifax, Canada: IASTED,1998.
- [6]Sun Junding,Cui Jiangtao,Wu Xiaoxing. Color image retrieval method based on color and shape features [J]. Chinese Journal of image and graphics, 2004, 9(7):820-827.
- [7]Deng Chengqiang,Feng Gang. A multi feature integrated image retrieval based on content [J].Computer application, 2003, 23(7):100-102.
- [8]Wu Jie,Qing Zhengding. Progress in research on fusion underlying content features in image retrieval [J]. Chinese Journal of image and graphics, 2008, 13(2):189-197.
- [9]Hu M. Visual pattern recognition by moment invariants [J].IRE Transactions on Information Theory, 1962,(8) :179-187.
- [10]Huang Yiping,Huang Xinsheng,Li Xiaolei. A new invariant of gray variation [J]. Journal of Dalian Maritime University, 2009, 34(4):23-27..
- [11]Mingdru F, Tuytelaars T, Gool L V, et al. Moment invariants for recognition under changing viewpoint and illumination[J].Computer Vision and Image Understanding,2004,94(1):3-27.
- [12]Lower D G. Distinctive image features from scale-invariant key points [J].International Journal of Computer Vision 2004, 60(2):91-110.
- [13]Meng Lu. Color image salient region extraction based on visual attention mechanism [J]. Computer Application Research, 2013, 30(10):3159-3161.
- [14]Yang Jun,Lin Tusheng,Xiao Yingwang. Salient object detection based on the global color contrast [J]. Computer Application Research, 2014.31(1):292-298.