Commerce Image Retrieval with Combination of Descriptors

He Zhang[†] and Zijun Sha[††]

Graduate School of Advanced Mathematics and Human Mechanisms, University of Toyama
Gofuku 930-8555, Toyama-shi, Japan

Summary
It is a critical problem to find the commerce quickly and accuracy. Information from different sources can improve the retrieval performance. In this paper, we proposed a combination algorithm of color descriptor, LBP texture descriptor and HOG shape descriptor for commerce image retrieval. The commerce image retrieval experiments on commerce image dataset PI 100 indicate the combination can boost the performance of retrieval system significantly.
Key words: *Commerce image retrieval, descriptor combination, PI 100*

1. Introduction

It is a critical problem to find the commerce quickly and accuracy. For the contend-based image retrieval, the image is submitted directly to the system, rather the description words for the product. The algorithm compares the features of submitted image with that of the images in the database, the most similar image will be returned to the users. The schematic process is illustrated as blow:

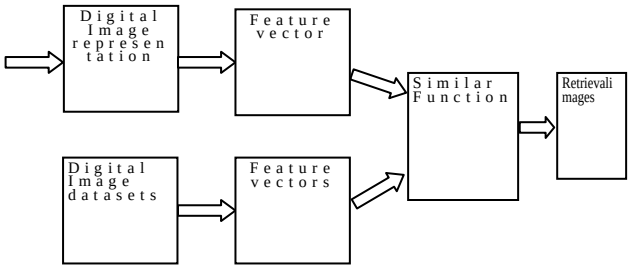

Figure 1 The structure of contend based retrieval system

2. Feature choice and combination

Information from different sources can improve the retrieval performance. The combination of descriptors associated with different visual properties, such as color, texture and shape.

2.1 Color

The color signature is extracted from the whole images to produce a global descripto. The color dictionary $\{C_1,C_2\ldots C_k\}$ is built by quantizing each components separately, and regularly in 4 or 8 bins, leading to 64 or 512 colors in total r[1]. In order to reduce the impact of large uniform colored areas, we collect typical and unique colors to build the color dictionary.

1) Collect some random images;
2) Resize each image to 100*100 pixels and convert to HSV, then split it in some 8*8 blocks;
3) Find the most occurring color of each block.
4) Cluster the colors from all imags with k-means algrithms, producing k color palette.

HSV color space is employed to the system as its mode is closer to the vision of human compare to the RGB color space. The conversion from default RGB space to the HSV space is as follows:
Assuming (*r*, *g*, *b*) are the three components (red, green, blue)of one specific color, the values of which range from 0 to 1。 And assuming *max* as the maximum of the three real number of *r*, *g* and *b*, *min* is as the minimum of the three components. (*h*, *s*, *v*) are the three components in HSV space, in which $h \in [0, 360)$ is the hue of the color , $s, v \in [0,1]$ are the value and saturation of the color, respectively. The transformation formula are as follow r[2] :

$$h=\begin{cases} 0^0 & if\ \max=\min \\ 60^0\times\dfrac{g-b}{\max-\min}+0^0, & if\ \max=r\ and\ g\geq b \\ 60^0\times\dfrac{g-b}{\max-\min}+360^0, & if\ \max=r\ and\ g<b \\ 60^0\times\dfrac{g-b}{\max-\min}+120^0, & if\ \max=g; \\ 60^0\times\dfrac{g-b}{\max-\min}+240^0, & if\ \max=b; \end{cases} \quad (1)$$

$$l=\frac{1}{2}(\max+\min) \quad (2)$$

Manuscript received June 5, 2013
Manuscript revised June 20, 2013

The color space quantization.:

$$H=\begin{cases}0, h\in(345,15]\\1, h\in(15,25]\\2, h\in(25,45]\\3, h\in(45,55]\\4, h\in(55,80]\\5, h\in(80,108]\\6, h\in(108,140]\\7, h\in(140,165]\\8, h\in(165,190]\\9, h\in(190,220]\\10, h\in(220,255]\\11, h\in(255,275]\\12, h\in(275,290]\\13, h\in(290,316]\\14, h\in(316,330]\\15, h\in(330,345]\end{cases} S=\begin{cases}0, s=(0,0.15]\\1, s=(0.15,0.4]\\2, s=(0.4,0.75]\\3, s=(0.75,1]\end{cases} V=\begin{cases}0, v=(0,0.15]\\1, v=(0.15,0.4]\\2, v=(0.4,0.75]\\3, v=(0.75,1]\end{cases} \quad (3)$$

For the component Hue of HSV space, 0 represents for red color, 120 for green color, 240 for blue color. The visible spectrum covers the hue region ranged from 0 to 240. In this paper, the HSV space are quantized into 256 bins, which is computed as illustrated in (3).

Color signature compute the histogram of colors in the image for the fixed codebook, firstly, resize the image to 100*100, for each pixel, select the closest color in the color codebook according to the Euclidean distance. The k-dimension histogram is built with the distribution of the color in the codebook.

Inverse document frequency down weights the impact of the common visual words and increases the importance of the rare visual words. The histogram is updated by applying idf weighting terms.The power-law method regularizes the contribution of each color in the final descriptor.

The L1 vector normalization is performed to make the vector comparable.Signatures are compared to find the nearest neighbour of the query images in a signature database, in which the choice of the metric is therefore critical.

2.2. Texture

The text features extraction is based on either statistics or structure. Maenpaa T, Ojala T, Pietikainen M and Soriano M presented LBP(Local Binary Pattern) algorithm[3,4], which analyzes the fix window features with structure methods and extract global feature with statistic methods. For RGB color images, convert the color space to the gray space:

$$G(i)=R(i)*0.3+G(i)*0.59+B(i)*0.11 \quad (4)$$

LBP operate on 3*3 windows, binary process for the pixels in the window according to the central pixel, the LBP value is obtained by the weighted sum of the pixels in the window.

$$\begin{array}{ccc}18 & 12 & 11\\11 & 15 & 22\\35 & 3 & 54\end{array}\quad\begin{array}{ccc}1 & 1 & 0\\0 & x & 1\\1 & 0 & 1\end{array}$$

(a) (b)

$$LBP=1*2^0+0*2^1+1*2^2+0*2^3+0*2^4+1*2^5+1*2^6+1*2^7 \quad (5)$$

Figure 2. Basic LBP operator

2.3 Shape

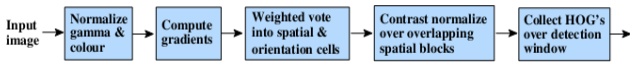

Figure 3. HOG operator

Dalal N and Triggs D firstly describe Histogram of Oriented Gradien (HOG) descriptors[5]. HOG is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

The first step of calculation is the computation of the gradient values, which simply applies the 1-D centered, and points discrete derivative mask in one or both of the horizontal and vertical directions. Specifically, this method requires filtering the color or intensity data of the image with the following filter kernels:

$$[-1,0,1] \; and \; [-1,0,1]^T$$

The second step of calculation involves creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The cells themselves can either be rectangular or radial in shape, and the histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is “unsigned” or “signed”. unsigned gradients used in conjunction with 9 histogram channels performed best in their human detection experiments. As for the vote weight, pixel contribution can be the gradient magnitude itself,

In order to account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is then the vector of the components of the normalized cell histograms from all of the block regions.

These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor.

The optimal parameters are found to be 3x3 cell blocks of 6x6 pixel cells with 9 histogram channels. Moreover, they find that some minor improvement in performance can be gained by applying a Gaussian spatial window within each

block before tabulating histogram votes in order to weight pixels around the edge of the blocks less.
Since the HOG descriptor operates on localized cells, the method upholds invariance to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions.

2.4. Descriptor combination

The color feature histogram has the advantages of simple, invariant of image rotation, scale and translation; however, the color feature histogram loses the space distribution information. The LBP and HOG features provide texture and shape information of an object within an image, respectively. In addition, they are robust to the noise, so it is beneficial to combine the three features.
In this paper, the color feature , LBP texture feature and HOG feature are normally weighted to get the retrieve result. For the query image Q and any image I in the database, the distance d(Q,I) is computed as:

$$d(Q,I) = W_{\text{color}}\, d_{\text{color}} + W_{\text{texture}}\, d_{\text{texture}} + W_{\text{shape}}\, d_{\text{shape}} \qquad (6)$$

The local feature is extracted from local regions using a scale-invariant detector. The descriptors are coded and pooled to a single vector.

3 Experiment

3.1 Experiment set

For the commerce retrieval tests, we employed the microsoft product image set PI 100[6] which contains 10000 commerce images from Amazon website, each image is adjusted to 100*100 resolution. The samples of PI 100 are illustrated as Fig 4.1

3.2 Result and analysis

The recall and precision are two common methods to scale the performance of the retrieval system.

Table 1 the component for recall and precision

	related	No-related
Returned	A	B
No returned	C	D

A: The number of returned positive images;

B: The number of returned negative images;

C: The number of loss-returned positive images

D: The number of losss-returned negative images;

Fig. 4 The samples of PI 100 product image database

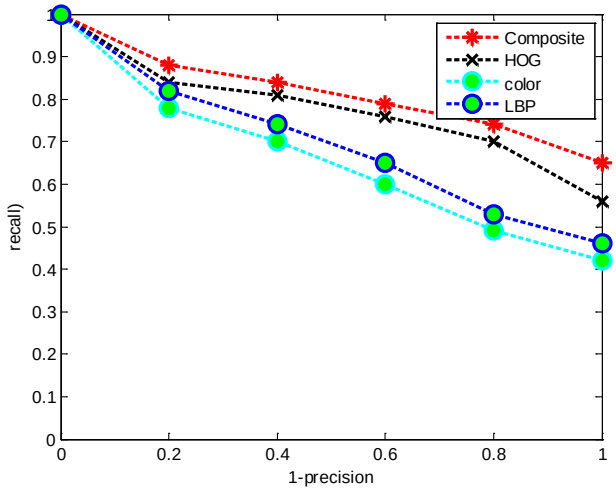

Figure 5 The experiment result

Recall is defined as “the number of returned positive imags /The number of all the positive images”, while “precision” is “the number of returned positive imags /The number of all the returned images”,that is:

Recall=A/(A+C);

Precision=A/(A+B) (6)

From the result curve, we can conclude that: HOG performs best among the three descriptors, the next is LBP descriptor, the employed color descriptor is the worst of all, which indicates that, compared to the shape and the texture, the color feature is less important to discriminate product

categories. The composite feature performs better than any single descriptor. Through the process of cross validation, the weigh parameters for W_{color}, W_{texture}, W_{shape} are set to 0.2, 0.3, 0.5, respectively. However, the experiments also indicate that the performance with uniform combination is very similar to the optimum parameter after validation, and is better than any single descriptors.

4 Conclusion

In this paper, we proposed a combination algorithm of color descriptor, LBP texture descriptor and HOG shape descriptor for commerce image retrieval. The commerce image retrieval experiments indicate the combination can boost the performance of retrieval system significantly. Future study will focus on the selection more efficient descriptors set and design more effective combination algorithms to improve the overall performance of commerce retrieval system.

References

[1] Wengert C, Douze M, Jégou H. Bag-of-colors for improved image search. Proceedings of the 19th ACM Multimedia.2011:1437-1440
[2] http://en.wikipedia.org/wiki/HSL_and_HSV
[3] Maenpaa T, Ojala T, Pietikainen M, Soriano M. Robust texture classification by subsets of local binary patterns, 15th International Conference on Pattern Recognition.vol 3,2000:935-938
[4] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns:Application to face recognition,IEEE Transaction on Pattern Analysis and Machine Intelligence 28(12) 2037-204.
[5] Dalal N and Triggs D.Histograms of oriented gradients for human detection. In: Anon.Proceedings of Conference on Computer Vision and Pattern Recognition. San Diego, California, USA. 2005. New York: IEEE Computer Society Press,2005.556~893
[6] Xing Xie, Lie Lu, Menglei Jia, Hua Li, Frank Seide, Wei-Ying Ma, Mobile Search with Multimodal Queries, Proceedings of the IEEE, Vol. 96, No. 4, Apr. 2008.

He Zhang received the B.S. degree from ChangChun University, Ji Lin, China, in 2005 and the M.S. degree from University of Toyama, Toyama, Japan in 2011 Now, he is working toward the D.E. degree at University of Toyama, Toyama, Japan. His main research interests are multiple-valued logic, image recognition.

Zijun Sha received the B.S. degree from Hunan University of science and technology, Hunan, China in 2008 and the M.S. degree from University of Toyama, Toyama, Japan in 2012. Now, he is working toward the D.E. degree at University of Toyama, Toyama, Japan. His main research interests are multiple-valued logic, artificial neural networks and information security