Multiclass VisualRank: Image Ranking Method in Clustered Subsets Based on Visual Features

Mitsuru Ambai
Denso IT Laboratory, Inc.
Shibuya CROSS TOWER 28th Floor 2-15-1
Shibuya Shibuya-ku Tokyo, 150-0002 Japan
manbai@d-itlab.co.jp

Yuichi Yoshida
Denso IT Laboratory, Inc.
Shibuya CROSS TOWER 28th Floor 2-15-1
Shibuya Shibuya-ku Tokyo, 150-0002 Japan
yyoshida@d-itlab.co.jp

ABSTRACT

This paper proposes Multiclass VisualRank, a method that expands the idea of VisualRank into more than one category of images. Multiclass VisualRank divides images retrieved from search engines into several categories based on distinctive patterns of visual features, and gives ranking within the category. Experimental results show that our method can extract several different image categories relevant to given keyword and gives good ranking scores to retrieved images.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval; I.4 [**Image Processing and Computer Vision**]: Miscellaneous

General Terms

Algorithms, Theory

Keywords

Ranking, Clustering, Visual Feature

1. INTRODUCTION

Image search engines widely available on the Web retrieve images sorted in descending order of ranking scores that are calculated from text information around the images. However, the search engines sometimes give high scores to images irrelevant to queried keywords, or vice versa. Image ranking often fails because the meaning of text information does not always correspond to the meaning of images.

Jing et al.[2] has proposed VisualRank that uses visual features instead of the text information to refine ranking scores of images retrieved from an image search engine. While VisualRank achieves high retrieval precision, the top results tend to be occupied by similar images as shown in figure 1(B). There is not always one representative image for a queried keyword. It is preferable that the user can obtain a diverse set of images.

In this paper, we propose Multiclass VisualRank, a method that expands the idea of VisualRank into more than one category of images. Multiclass VisualRank divides images retrieved from search engines into several categories based on distinctive patterns of visual features, and gives ranking within the category. This method displays the images in multiple sequences. Each of the sequences contains categorized images that are sorted by their ranking scores as shown in figure 1(C). Our method works as a post-filtering for existing image search engines. This helps users to grasp the entire results retrieved from image search engines.

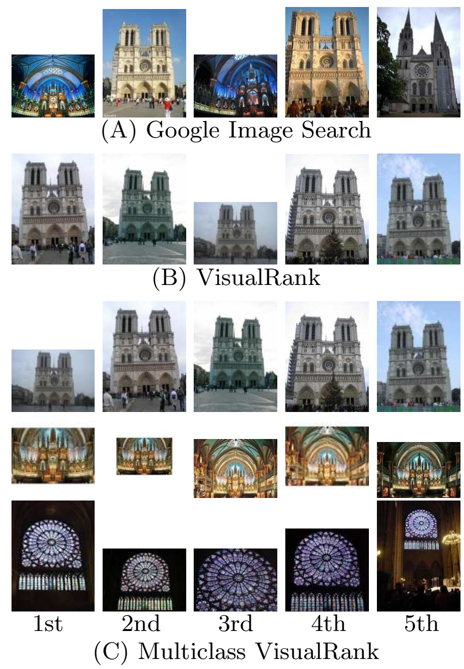

Figure 1: Top 5 results of a keyword, 'Notre Dame', in each ranking method.

Section 2 describes our algorithm in detail. Section 3 shows our experimental results using retrieved images from Google Image Search.

2. MULTICLASS VISUALRANK

Multiclass VisualRank is composed of following three steps: obtaining visual similarity, clustering and ranking. As well as VisualRank, SIFT key points [3] and PageRank [1] are used in the step of obtaining visual similarity and the step of ranking, respectively. The principal contribution of this paper is that the clustering algorithm is incorporated to the framework of VisualRank in order to extract different image categories related to given keywords.

2.1 Obtaining Visual Similarity

The visual similarity w_{ij} between two images I_i, I_j is calculated by using SIFT key points. An ratio C_{ij} is defined as the number of sharing key points between I_i and I_j devided by the mean number of key points extracted from I_i, I_j.

While original VisualRank uses the ratio C_{ij} as the visual similarity, in our method a sigmoid function is applied to C_{ij} to weaken a large value. Because the image search engines sometimes retrieve exactly the same images for a given keyword. In those cases, the value C_{ij} becomes too large compared to the other visual similarities, and probably worsens the performance of clustering. The insertion of the sigmoid function helps to avoid this issue.

2.2 Clustering

The images connected with their visual similarities can be regarded as a weighted graph. In particular, similar images are mutually connected with high visual similarity. The graph contains several clusters that correspond to different image categories.

Normalized cuts [4], that is a representative method of spectral clustering, is useful to extract each cluster from the graph. Normalized cuts is formulated as generalized eigenvalue problem as follows:

$$(D - W)v = \lambda Dv \tag{1}$$

where W is an adjacency matrix whose elements are w_{ij}, D is a degree matrix, λ is the eigenvalue and v is the eigenvector. The eigenvector corresponding to the second least eigenvalue provides optimal two-way partitioning that minimizes normalized cuts criteria $NCut$ defined in [4]. This two-way partitioning is recursively repeated until the value $NCut$ exceeds a predefined threshold N_{th}. The number of clusters is automatically determined depending on N_{th}. In our experiment, N_{th} is set to 0.4.

2.3 Ranking

According to [2], VisualRank inspired by PageRank is formulated as follows:

$$\mathbf{r} \leftarrow (1 - \alpha)W\mathbf{r} + \alpha\mathbf{p} \tag{2}$$

where $\mathbf{r} = (r_1, \cdots, r_N)^\top$ is a vector of the ranking scores, $\mathbf{p}$ is a uniform vector that models random walk of Web browsing and α is a balancing factor that is set to 0.15 in our experiment. The ranking score vector $\mathbf{r}$ is updated by the procedure in (2) repeatedly.

In the case of multiclass, the adjacency matrix W is modified as follows:

$$w'_{ij} = \begin{cases} w_{ij} & \text{if } I_i \text{ and } I_j \text{ belong to the same category} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Instead of W, the modified adjacency matrix W' is used to calculate ranking scores. The equation (3) means that the visual similarities between different categories are ignored. It is preferable that a image does not receive ranking scores from images belonging to different categories. In this way, the more similar to the canonical appearance of each category an image is, the higher ranking score it obtains.

3. EXPERIMENTAL RESULTS

We tested three sets of keywords. The part of keywords is listed below[1]:

(a) **Sightseeing spots**: Tokyo-tower, Notre Dame, etc.
(b) **Artists**: Rembrandt, Leonardo da Vinci, Klimt, etc.
(c) **Product names**: Wii, Xbox, iPhone, Gameboy, etc.

Each set includes 10 keywords. For each keyword in each set, top 250 images were downloaded from Google Image Search. Large images were resized to 300K pixels keeping their aspect ratio.

Table 1: Evaluation results of extracting categories

Keyword	ANC	Relevant	Irrelevant
Sightseeing spots	1.9	1.8 (R:0.6, U:1.2)	0.1
Artists	3.2	2.9 (R:0.6, U:2.3)	0.3
Product names	2.3	2.3 (R:0.2, U:2.1)	0.0

ANC: Average Number of Categories per query
Relevant (or Irrelevant): Relevant (or irrelevant) categories per query
R: Redundant categories per query
U: Uniquely identified (non-redundant) categories per query

The precision in the top 10 re-ranked images in all of the extracted categories was 0.949. According to [2], the precision of original VisualRank is 0.953. As well as VisualRank, our method achieved high retrieval precision.

Table 1 shows the evaluation results of extracting image categories. Our method provided 2.3 relevant categories per query. The keyword set of artists tended to give more categories than the other sets. In the case of artists, several representative paintings by the artists were extracted as image categories. For instance, the keyword, 'Leonardo da Vinci', provided 5 categories. This would help users to grasp their representative paintings.

The same objects taken from different view points or under different lighting conditions were appeared in the retrieved images. These images were occasionally divided into two or more categories. If these categories were visually similar, they were regarded as redundant categories. The average numbers of redundant categories of the sightseeing spots and artists were relatively high compared to product names.

An average number of irrelevant categories per query proved to be small. The precision of obtaining relevant categories among all the keywords was 0.95. The extracted categories were mostly related to the queries.

In conlcusion, the experimental results revealed that relevant yet various categories can be automatically extracted, and the images belonging to each of the categories were sorted by their ranking score at high precision. This method would provide better usability for image search engines.

4. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, Apr. 1998.

[2] Y. Jing and S. Baluja. VisualRank: Applying PageRank to large-scale image search. *PAMI*, 30(11):1877–1890, Nov. 2008.

[3] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[4] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, Aug. 2000.

[1] All keywords were queried in Japanese. They are translated into English in this paper.