J-049

An Analysis of Image Similarity Estimation based on Eye Information

Mario Saputra† Mutsumi Suganuma‡ Wataru Kameyama‡

1. Introduction

Image similarity is very subjective and different from person to person. To study about image similarity based on human subjectivity, we have utilized eye information to make image similarity map based on the implicit relevance feedback[1]. In this paper, more features from eye information are proposed to be used, and new experiments' results with singular value decomposition (SVD) as well as PCA are shown aiming to obtain better image similarity map.

2. Experiment Setup and Data Collection

2.1 Experiment Overview

We use Ogama[2] for recording and analyzing eye information to design the experiment. Information from user eye movement is collected by an eye tracker "EYE TRIBE" with the sampling rate of 30[Hz]. For the experimental task, first, a page containing one image called query image is shown to a subject. After that, the subject proceeds to the next page which contains 2 images by pressing a key on the keyboard. On this page, subject is asked to choose an image which is more similar to the query image. After five pages comparing two images, the query image is shown again to make the subject remember that. A query page and a comparing task page are shown in Fig.1.

Fig.1 Example of Experimental Pages

Two experiments have been done as follows.

2.2 First Experiment

There are totally 204 pages including 170 comparing task pages and 34 query pages. The number of images for the comparison task is 27 images, where 15 images are dummy and not used for the ranking. They are included in order to reduce the memory effect of ever-seen images in the task pages and to make a subject carefully focus on the image comparison. Each image is taken from LabelMe 8 Categories dataset[3]. To get similarity ranking, pairwise comparison method on 12 images is applied.

2.3 Second Experiment

In the first experiment, there was a possibility that the subjects may take different criterion to judge the image similarity, and the number of dummy images may not be large enough to reduce the memory effect. With this reason, we conduct the second experiment which is divided into two. The first one is to ask subjects to judge the similarity based on semantic feature, and the second one is to ask subjects to judge the similarity based on visual feature. Semantic feature means objects or scene context of an image, while visual feature means color, texture, shape, or spatial layout of an image. A clear definition and instruction about those similarities is given to the subjects before the experiment.

The second experiment uses 6 task images and 24 dummy images. And, there are totally 90 pages including 75 comparing task pages and 15 query pages. Each image is taken from LabelMe 8 Categories dataset differently from the first experiment. To get the ranking, pairwise comparison method on 6 images is applied.

2.4 Features used in Experiment

Fifteen features have been utilized as shown in Table 1 with new proposed features as of No. 6, No. 7, No. 14 and No.15.

Table 1 Features used in the Comparison

No	Feature Description	Proposed in
1	Total number of fixations in an image	[4], [5]
2	Total duration of fixations in an image	[4], [5]
3	Average duration of fixations in an image	[4], [5]
4	Maximum value of fixation duration in an image	[4]
5	Average fixation duration when an image first gazed	[4]
6	SD of fixation position in x-coordinate	New Prop.
7	SD of fixation position in y-coordinate	New Prop.
8	Number of measurements	[4]
9	Maximum pupil diameter	[4]
10	Average value of pupil diameter	[4]
11	Rank of Average Fixation Duration compared to other image in same page	[4]
12	Rank of Total Fixation Duration compared to other image in same page	[4]
13	Rank of Maximum Fixation Duration compared to other image in same page	[4]
14	SD of gaze position in x-coordinate	New Prop.
15	SD of gaze position in y-coordinate	New Prop.

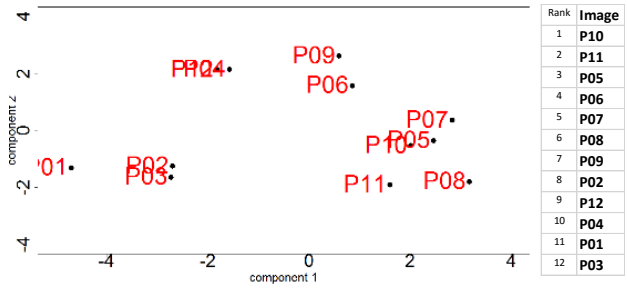
3. Result and Analysis

3.1 First Experiment Results

PCA and SVD algorithms have been utilized to get the similarity map result using the features in Table 1. For PCA, we choose two principal components because they cover more than 60% variance of data. And for SVD, we use one singular vector because it covers more than 90% variance of data. Data from 7

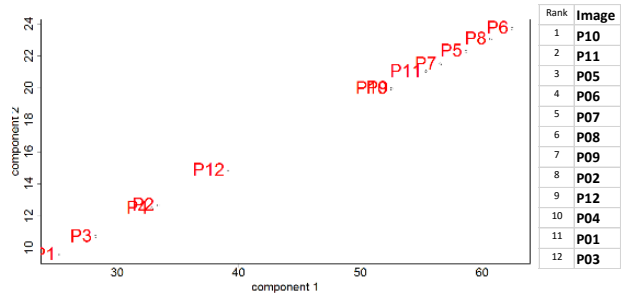
† Graduate School of Fundamental Sci. and Eng., Waseda University
‡ Faculty of Sci. and Eng., Waseda University

subjects (4 males, 3 females, 23.57 average age with 2.3 SD) have been analyzed. The distances between plots in the map represent the user subjectivity ranking. The best result obtained from Subject 5 as show in Fig. 2 and Fig. 3.



Rank	Image
1	P10
2	P11
3	P05
4	P06
5	P07
6	P08
7	P09
8	P02
9	P12
10	P04
11	P01
12	P03

Fig.2 Comparison between PCA Map and Ranking by Subject 5

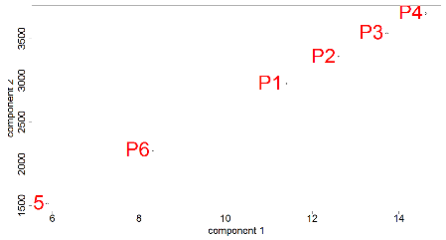


Rank	Image
1	P10
2	P11
3	P05
4	P06
5	P07
6	P08
7	P09
8	P02
9	P12
10	P04
11	P01
12	P03

Fig.3 Comparison between SVD Map and Ranking by Subject 5

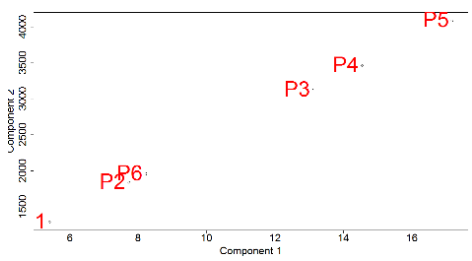
3.2 Second Experiment Result

Data from 4 subjects (3 males, 1 female, 22.25 average age with 1.5 SD) has been analyzed. The best result obtained from SVD result of Subject 3 as shown in Fig.4 and 5.



Rank	Pic
1	**P3**
2	**P4**
3	**P1**
4	**P2**
5	**P5**
6	**P6**

Fig.4 Semantic Comparison SVD Map and Ranking by Subject 3



Rank	Pic
1	**P3**
2	**P4**
3	**P1**
4	**P2**
5	**P5**
6	**P6**

Fig.5 Visual Comparison SVD Map and Ranking by Subject 3

3.3 Analysis

The Pearson correlation between the points in the similarity maps and the corresponding ranking distances in the subjective ranking tables are calculated for the accuracy analysis. Table 2 shows the Pearson correlation values from all 7 subjects in the first experiment. Table 3 shows them from all 3 subjects in the second experiment.

Table 2 Correlations in the First Experiment

Subject	PCA Accuracy	SVD Accuracy
Subject 1	0.18	0.05
Subject 2	0.12	0.08
Subject 3	0.27	0.20
Subject 4	0.12	0.28
Subject 5	0.50	0.65
Subject 6	0.05	0.06
Subject 7	0.06	0.09

Table 3 Correlations in the Second Experiment

Subject	Semantic Similarity		Visual Similarity	
	PCA Accuracy	SVD Accuracy	PCA Accuracy	SVD Accuracy
Subject 1	0.21	0.08	0.06	0.19
Subject 2	0.08	0.02	0.20	0.35
Subject 3	0.07	0.49	0.12	0.41
Subject 4	0.46	0.07	0.05	0.17

4. Consideration

On Subject 5 in the first experiment, a strong correlation (> 0.6) is observed in the SVD result. And Subject 3 in the both of the second experiment also shows moderate and week correlations for the SVD result. From the subjects' answers of the second experiment, more resembling ranking results between the users are obtained, which means it seemed easier for the subjects to make the comparison. However, the accuracy of the map is still not so different with first experiment. According to the results, SVD seems to produce better similarity mapping than PCA in most of the cases. In general, a similarity map sometimes reflects the subjective ranking, but sometimes not.

5. Conclusion and Future Works

We have shown image similarity analysis based on eye information. The results vary from subject to subject.

To improve the result, an optimized and efficient set of features have to be found, as well as better algorithm to classify them.

References

[1] Mario Saputra, Mutsumi Suganuma, Wataru Kameyama, "Preliminary Experimental Result on Image Similarity Estimation based on Eye Information", IEICE General Conference 2016, H-2-1, 2016
[2] http://www.ogama.net/ (last visited on 2015-12-23)
[3] http://people.csail.mit.edu/torralba/code/spatialenvelope/ (last visited on 2015-12-23)
[4] K. Pasupa, C. J. Saunders, S. Szedmak, A. Klami, S.Kaski and S.R. Gunn, "Learning to Rank Images from Eye movements", Proc. IEEE 12th International Conference on Computer Vision Workshops, pp.2009-2016, 2009
[5] S.N. Hajimirza, M. Proulx and E.Izquierdo, "Reading Users' Minds from Their Eyes: A Method for Implicit Image Annotations", IEEE Trans. Multimedia, Vol.14, No.3, pp.805-815, 2012