4Q-5

A Research on Automatic Classification and Image Retrieval based on Locations and Faces

Wafaa Eid,† Tsuyoshi Saitoh†
Tokyo Denki University

Abstract

This paper presents an approach for classifying photos based on the image visual contents. Photos can be classified into faces and places. A face detector is used to extract faces from the photos, while a face recognizer is used to classify the detected faces. GPS information which is used to detect places is extracted from geo-tagged photos in a database by image matching. The reverse geo-tag technique enables to find keywords related to the place, and the keywords help to retrieve photos.

Introduction

Nowadays the development of technology is so fast that the market of digital cameras is getting bigger as the user demands are getting higher. The number of people who have cameras, either digital one or one included on the mobile phone, is increasing. The minimum price is getting lower, while the memory space is getting bigger. People document their life by taking a lot of photos per day. Storing and organizing such a huge amount of photos can be a challenging task. It will be hard to retrieve a desired photo taken in a certain place with a certain person. Most of the current available image classification systems use manual annotation to tag the GPS data though it's a time consuming task. For these reasons, it's necessary to have an automated classification and retrieval system that can classify photos and suggest keywords for easy retrieval process. Place detection part was presented in [1]. This paper will focus on face detection and recognition effect to this research.

Automated Classification and Retrieval System

System Overview

The system allows users to upload photos and then it will detect whether the photos contain GPS information or not. If they contain GPS information, keywords will be extracted and added to the photos. If not, the system will search for a similar photo in a geo-tagged database, extract the GPS information from the matched geo-tagged photo and add the keywords to the users' photos. The system also enables users to organize photos by faces. Detected faces are stored as a training set and used for recognition purpose.

System Features

1. GPS Extraction

The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites [2]. GPS data can be recorded live in Exchangeable Image File (EXIF) [3] attached to the photo. The system can extract the GPS data by reading the EXIF file attached to the photo.

† Graduate School of Information System and multimedia Design, Tokyo Denki University

The extracted GPS data can help to get the name of the place using reverse geocoding. Reverse geocoding is the process of translating a point from a map into a human-readable address [4]. Using this feature available from google developer, the system can extract keywords which are related to the place. Such keywords will be helpful in retrieving photos. However, not all the photos can have GPS data in their EXIF. In such cases the system will find a match to the user's photo from a geo-tagged database. The image matching process is explained in details in section 2.3. The system will be able to detect and recognize faces. Hence face detection and recognition have not been developed in the system yet.

2. Keywords Extraction

Keywords are terms used for retrieval purposes. The result of reverse geocoding is the XML file with address elements. These elements can be also used as a keywords for image retrieval purpose. However, some elements are repeated in the file which makes it necessary to filter the keywords before saving them in the database. The system filters the keywords and save it the database. An example of the keywords after filtering: ２; ３; １丁目; Asakusa; Taito; Tokyo; Japan; JP;浅草雷門（バス;111-0034;Kaminarimon; Tawaramachi Station; １; Nishiasakusa; 111-0032; Asakusa Station; ４; Hanakawado; ２５

3. Image Matching

Image matching is to find similar visual contents between 2 photos (a user's photo and a geo-tagged photo in the database). The process of image matching is as follows. First, photos from the user are compared with photos from the database using SURF detection. SURF (Speeded Up Robust Features) is a robust local feature detector which is mainly used in object recognition. It works by finding a set of distinctive interest points. These points are detected by calculating the Hessian-Laplace detector in different scale spaces. The descriptor will be extracted from the normalized region around the interest points. The matching interest points are calculated by comparing the contrast. If they have the same type of contrast, it's matched [5]. Color filtering is applied to the output of SURF detection process to eliminate all the colors of the image except what's specified. Filtering the image by one color can help in detecting shapes more easily and faster. SURF detection is applied to photos imported by users and geo-tagged photos in the database until a rectangle is detected. To make the detection easier, the system will filter the resulted image by color to avoid wrong detection. After that, it will apply an edge detector [6] to detect the rectangle, and the GPS data will be extracted from the geo-tagged photo.

4. Face Detection and Recognition

Face detection is a computer vision technology that determines the locations and sizes of human faces in arbitrary images [2].The method used in this research is Viola-Jones

method [7]. Viola-Jones method uses a Haar wavelet feature. To determine wither a feature is presence or not in every location in the image, integral images are used. Each feature is presented by a single value which is calculated by subtracting the sum of the white and black rectangle. A cascade classifier is composed of stages and is used to determine if the image contain faces or not. If a sub-window is considered as non-face it will be discarded. However the ones which are considered as faces will presume to the next stage and so on. Figure shows the result of face detection. The method used for face recognition is PCA (principal component analysis) based face recognition [8]. This is done by creating a training set of images. Then train the recognizer by converting face images to face vectors. To have a unique face vectors a normalization process is needed. The normalization process will remove the average face which have a common features from face vectors. The eigenfaces are calculated using the covariance matric of reduced dimensionality. The weight vector is calculated for each face image. This is used to represent how much percentage this eigenface contributes to this face. When a new face s detected it will be converted to a face vector and the normalized process will be applied to it. The eigenfaces will be calculated as well as the weight vector for that image. To recognize the face, the distance is calculated between the weight vector of the new image with the weight vectors of the training set. If the distance is less than the threshold then the face is recognized and the name will be returned. But, if the distance is higher than the threshold it's considered as unknown face. Figure shows the flow chart of face recognition process.

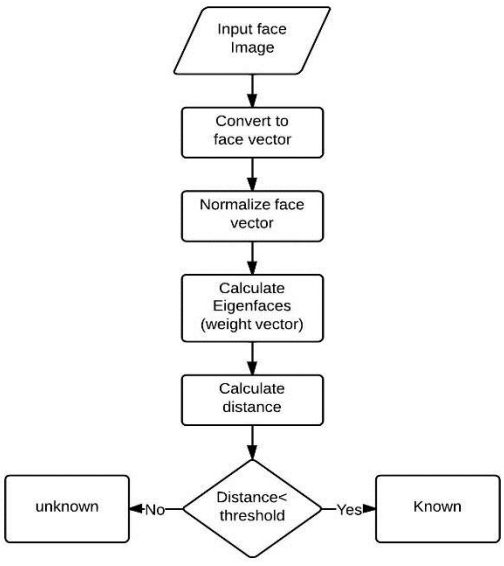
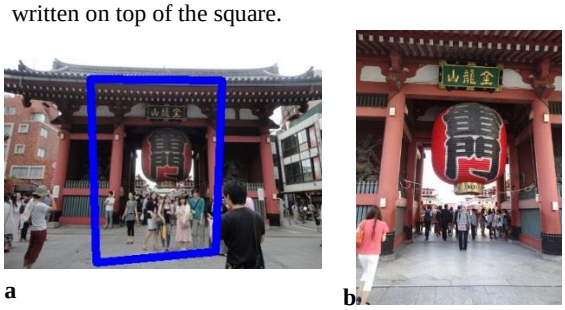


Figure 1.Face recognition flowchart.

Experiment Results

The system classify images to faces or places based on the place detection method and face recognition method. Sample of the result is shown in figure 2 and 3. Results shows that some faces couldn't be detected as the used method detect only frontal view faces. Another problem is the lightning of the photo can effect place detection. The used method for place detection cannot be used with places that does not have a unique feature in it.

Figure 2. Face detection and recognition result with the name written on top of the square.



a **b**

Figure 3. Result of place detection a. user's photo after place detectio b. matched photo in geo-taged database.

Conclusion

This paper introduces how photos can be classified into faces and places and how visual contents of photos can be valuable resource for organizing and retrieving purposes. The fast image matching technique enables to add GPS data from geo-tagged photos in a database to user's photos. Based on the results, this method can detect similar geo-tagged photos. It helps to retrieve place information and provide keywords for fast text-based image retrieval. Face detection and recognition organize photos based on who is in the photo. This technique helps in retrieving images using the label of the photo. More development will be pursue in our feature work.

References

[1]Wafaa Eid. And Tsuyoshi Saitoh "A Research on Automatic Picture Classification and Image Retrieva", presented at the 12th Conf. Forum on Information Technology, Sendai, Japan, 2013.
[2] Wikipedia, http://en.wikipedia.org/.
[3] EXIF: Exchangeable image file format for digital still cameras: EXIF version 2.2. Technical report, Japan Electronics and Information Technology Industries Association (2002).
[4] Google Developers, https://developers.google.com.
[5] Herbert Bay , Andreas Ess , Tinne Tuytelaars , Luc Van Gool, Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding, v.110 n.3, p.346-359, June, 2008.
[6] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. lntell. PAMI-8, 679- 698 (November 1986).
[7]P. Viola and M. Jones "Robust Real-Time Face Detection", Int\'l J. Computer Vision, vol. 57, no. 2, pp.137 -154 2004
[8] M. Turk and A. Pentland (1991). "Face recognition using eigenfaces". Proc. IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–591.