Classifying Personal Photo Collections: An Event-Based Approach

Ming Geng[1], Yukun Li[1,2(✉)], and Fenglian Liu[1,3]

[1] Tianjin University of Technology, Tianjin 300384, China
[2] Tianjin Key Laboratory of Intelligence Computing and Novel Software
Technology, Tianjin, China
{gengming1124,liyukun_tjut}@163.com
[3] Key Laboratory of Computer Vision and System, Ministry of Education
Tianjin, Tianjin, China
lflian@tjut.edu.cn

Abstract. Due to the increasing number of photos taken by mobile phone, how to improve the efficiency of re-finding personal photos becomes an important research issue. Based on the survey of the current methods to re-find expected photos in mobile phones, this paper proposes a photo classification strategy based on specific events. We firstly discover the relationship of shooting time and shooting location to specific events, and propose a method to classify personal photos based on them. Because there is no existing data set for experiment, we collect several representative photo collections of different people as experimental data sets, and the experiments show the efficiency of our method. We implement an album system based on our method and compare it with other existing classification ones, and the results show that our strategy can improve the efficiency of re-finding photos.

Keywords: Specific events · Photo classification · Re-finding personal photos

1 Introduction

How to manage personal information based on knowledge is an important research issue. Personal information shows different types like documents, emails, photos and so on. In this paper we studied the problem by focusing on how to classify personal photos to make re-finding easy. With the wide use of mobile phone, people tend to use it to take photos for different purposes. Taking photos is not only a way to record good moments, but also a method to record important information for reusing later. For example, a person sometimes takes a photo to record the username and password of a website for reusing laer. So the number of photos taken by mobile phones is increasing day by day. How to effectively manage the photos and improve the re-finding efficiency are valuable research issues, and the classification of personal photos is a basic topic.

There are some challenges for people to study the problem. People have some personalized features, and the number of each person's photos is different. It is not easy to get a general method to classify personal photos. Because some personal photos have some privacy information, how to get a data set for experiment is also a hard problem.

This paper studied the problem and proposed a solution. The main contributions can be summarized as follows:

(1) We conducted a survey on this problem, and observed some features about personal photo classifying and re-finding. We propose the idea of re-finding personal photos by classifying them based on specific events. We also propose and define some relevant concepts.

(2) We studied the classification method based on specific events, and propose a photo classification method based on shooting time and location with high accuracy and low computational cost.

(3) Because there are no relevant data sets and benchmark, we established the experimental data sets based on the real personal photo collections and tested the proposed method from two aspects of the classification accuracy and the efficiency of re-finding personal photos. The results verify the effectiveness of our method.

2   Related Work

The existing classification methods of photos are divided into four categories. The first is based on geographical information, the second is face recognition and identity recognition; the third is based on annotations; the last one is based on events.

There are some works based on geographical information. Lo et al. [1] designed a clustering algorithm that clusters geo-tagged photos in accordance to thresholds of different scales. Papadopoulos et al. [2] proposed a way to automatically detect landmarks and events using visual features and tag similarity in a set of images with tags.

About face recognition and identity recognition, Kumar et al. [3] proposed a semi-supervised framework for recognizing faces. Brenner et al. [4] used sparse Markov Random Fields for face recognition, Xu et al. [5] and O'Hare et al. [6] implemented techniques to organize photos based on people's identity. This classification method mainly depends on the accuracy of image recognition technology. The image matching algorithm's time complexity is very high.

There are some works on annotations, Andrade et al. [7] focused on people annotation, location annotation and event annotation. Kim et al. [8] developed a system collecting and storing annotations about photos to manage and search for photos. Ulges et al. [9] and Zigkolis et al. [10] both presented an annotation framework to facilitate event classification. The method of annotation requires the image recognition technology to recognize the image content with high computational cost and time complexity, and the added annotations may not accurately express the photo semantics.

There are also some works to classify personal photos based on events. Bacha et al. [11] and Cao et al. [12] both use the connection between scenes and events to identify events. Because activities and events in our lives are structural, Bosselut et al. [13] presented a data-driven approach to learning event knowledge. Wang et al. [14] and Tang et al. [15] utilized metadata and visual features for event recognition. Yuan et al. [16] mined some features from GPS and visual cues for event recognition. There are a lot of events going on, and even the same event contains many scenes and objects. It is difficult to complete the scenes and objects statistics of all events. The same scene or object often corresponds to different events, such as hiking and beach, which may include the sky. Therefore, it is not accurate to use scenes and objects for image event recognition, and the time complexity and computational cost are very high.

According to our findings, these kinds of classification methods are only applicable in specific conditions. People taking a lot of photos in their resident locations, the classification based on geographical information is not effective to deal with this situation. For the classification methods based on face recognition and identity recognition, they only work well to the photos with human face. The annotations for classification may not accurately express the photo semantics. It is inaccurate to judge events by the scenes and objects, and the classification based on events is likely to separate photos taken in continuous time. The classification methods of photos should not only take the features of mobile phones into account, but also have high efficiency. Therefore, we need a general and efficient method to classify personal photos for re-finding.

3   Conceptual Model

3.1   Survey and Findings

We made a survey on the problem referring to 200 users, and have the following observations: (1) Current mobile phone albums use a flat storage structure to store the photos made by mobile phones, displaying thumbnails of all photos. However, the screen of the mobile phone is very small, and many photo thumbnails of the same event are all displayed. Then it is inconvenient for people to re-find photos, so it is necessary to have a suitable method for mobile phones classification and re-finding. (2) Most of the photos are made when people go out to take part in activities, and they often make dozens or even hundreds of photos, but these photos are not often re-used by people. People often reuse the photos recording some important information. Such photos are usually taken a small number at a time, and are mixed into those photos taken when people go out for fun or to take part in activities. (3) The main clues used by people to re-find a photo are the approximate shooting time and the specific event related to it. (4) Some mobile phone albums already have the function of classification based on

image content. But this function destroys the memory info about time, and they often have a low performance for content analysis.

3.2   Definitions

Based on the survey and findings, this paper proposes a photo classification strategy based on specific events. To make our approach clear, we propose some concepts and give their definitions as below.

Definition 1. Specific event. A specific event can be taken as a series of actions taking place for the same target subject in a relatively continuous period of time and relatively neighboring location. For example, attending a wedding of a friend at noon one day is a specific event.

The concept of "specific event" is not the same as the concept of "event". An event may include two or more different specific events, such as the "wedding" event may include several different people's weddings. A specific person's wedding is regarded as a specific event.

Definition 2. Personal Photo Sequence. A personal photo sequence is list of personal photos ordered by their shooting time ascendingly.

Definition 3. Time Interval. Let P_i and P_j be any two photos, the time interval of them is the time difference between the shooting time of P_i and P_j. The unit is second (s). Because most mobile phones have the function of taking picture and the shooting time can be recorded automatically, to compute time interval of two given photos is not a hard thing.

Definition 4. Photo Distance. Let P_i and P_j be any two photos, the distance between the shooting location of P_i and P_j is the photo distance. The unit is meter (m). Now most mobile phones has GPS function and the location of shooting can be recorded easily, then to obtaining distance of two given photos is also not hard.

Definition 5. Rate of Location Change. For any two photos P_i and P_j, let their time intervals be T, their photo distance be L, and the rate of location change of P_i and P_j is denoted as V, then $V = L / T$. The unit of it is m/s.

4   The Classification Method

Considering the characteristics of specific events, we intend to classify photos by specific events based on shooting time and location. At present, photos taken by mobile phones contain some metadata such as shooting time and GPS (latitude, longitude, and altitude). It is not difficult to get the shooting time and location of the photos taken by mobile phone.

Table 1. Representative seven photos of personal photo sequence

Photos	Shooting time	Latitude(°)	Longitude(°)	Altitude(m)	Description of specific events
P_1	2017-01-29 08:58:34	39.13472	115.250954	0	A selfie on the train
P_2	2017-01-29 08:59:13	39.132011	115.239204	0	A selfie on the train
P_3	2017-02-02 09:31:59	39.060826	117.133378	53	Recording information
P_4	2017-02-02 11:33:34	39.060828	117.133378	54	The flowers on table
P_5	2017-02-02 11:33:40	39.060828	117.133378	54	The flowers on table
P_6	2017-02-04 22:23:05	39.051158	117.139953	0	The allergic eyes
P_7	2017-02-05 06:00:11	39.051156	117.139953	0	The allergic eyes

4.1   A Sample of Personal Photo Sequence

To make our method clear, seven photos taken continuously by one participant were selected as an example, as shown in Table 1. The P_1 and P_2 are the photos of two selfies made on the train. The P_3 is a photo for recording some information in the laboratory. The P_4 and P_5 are the photos of the flowers on the laboratory table. The P_6 and P_7 are two photos of eyes taken when the eyes were allergic. Although the time interval between P_6 and P_7 is long, the photos' content is same, and they belong to the same specific event. Therefore, the correct photo sets of the seven photos based on specific events are: $\{(P_1, P_2), P_3, (P_4, P_5), (P_6, P_7)\}$.

4.2   Specific Event-Based Classification Method

After manually classifying experimental training data sets, it was found that the shooting time and location of the same specific events were relatively continuous, and that the changes of shooting time and location within the same specific events were significantly different from the ones between different specific events. After observing some regulations about the relationship between specific events and the changing of shooting time and location, this paper proposes a method to classify photos based on shooting time and location. Algorithm 1 shows the process of our method.

Algorithm 1. The classification algorithm based on shooting time and location

---

Input: Personal photos, I_1, ... , I_n.
Output: Photo sets associated with specific event, S_1, ... , S_j.
1: Obtain the photos' shooting time and location
2: Arrange personal photos, I_1, ... , I_n, into personal photo sequence P_1, ... , P_n.
3: j = 1;
4: **For each** ($i \in [1,\text{n-1}]$)
5:     Calculate the time interval and the photo distance of two adjacent photos in the personal photo sequence T_i, as L_i;
6:     Calculate the rate of location change of two adjacent photos in the personal photo sequence V_i;
7:     If ($V_i > 2$) $L_i = L_i / V_i$;
8:     Else $L_i = L_i$;
9:     Normalize the time interval T_i and the photo distance L_i and the normalized values are T_i' and Li';
10:     Calculate the spatio-temporal distance of two adjacent photos in the personal photo sequence ST_D_i;
11:     If ($ST\_D_i \leq \text{BST\_D}$) P_i, $P_{i+1} \in S_j$;
12:     Else $P_i \in S_j$, $P_{i+1} \in S_{j+1}$; j = j + 1;
13: **End For**
14: **Return** photo sets associated with specific event, S_1, ... , S_j.

As Algorithm 1 shown, we firstly obtain the photos' shooting time and location, and the location information is represented by a 3-ary (W, J, H), where the three parameters' meaning are latitude, longitude, and altitude. The unit of latitude and longitude is degree (°), and the unit of altitude is meter (m). Then, arrange the photos as a personal photo sequence, and afterwards calculate the time interval T and the photo distance L of any two adjacent photos. For the method to compute time interval is easy and obvious, we don't detailed it here

(1) **Calculate photo distance**

It is more accurate to consider three elements (W, J, H) than to only considering longitude and latitude for calculating distance. For example, there might be two photos with the same longitude and latitude, but different altitude. The commonly used method to calculate the geographical space distance is the spherical model, which regards the earth as a standard sphere, and the distance between two points on the sphere is the arc length. Any position point can be converted into three-dimensional coordinates of the

sphere. The two locations are given (W_1, J_1, H_1) and (W_2, J_2, H_2). The steps to compute the distance of the two location is as below.

a. To translate H_1 and H_2 into the distance of the location to the core of the earth, as shown with the formula (1).

$$\begin{cases} h_1 = H_1 + 6371229m \, (the \; earth's \; radius) \\ h_2 = H_2 + 6371229m \, (the \; earth's \; radius) \end{cases} \tag{1}$$

b. Convert longitude J_1, J_2 and latitude W_1, W_2 into radians j_1, j_2, w_1 and w_2, as shown with the formula (2).

$$\begin{cases} w_1 = W_1 * \pi/180 \\ w_2 = W_2 * \pi/180 \end{cases}, \quad \begin{cases} j_1 = J_1 * \pi/180 \\ j_2 = J_2 * \pi/180 \end{cases} \tag{2}$$

c. Convert the position information of the two photos into three-dimensional coordinates of the sphere, as shown by formula (3).

$$\begin{cases} X_1 = h_1 * \cos(w_1) * \cos(j_1) \\ Y_1 = h_1 * \cos(w_1) * \sin(j_1) \\ Z_1 = h_1 * \sin(w_1) \end{cases}, \quad \begin{cases} X_2 = h_2 * \cos(w_2) * \cos(j_2) \\ Y_2 = h_2 * \cos(w_2) * \sin(j_2) \\ Z_2 = h_2 * \sin(w_2) \end{cases} \tag{3}$$

d. The formula of the distance L of the two locations is as formula (4), by which we can obtain the distance of two give locations.

$$L = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2} \tag{4}$$

(2) **The fusion of time and distance**

We consider the effect of time interval and photo distance on specific events. Before we do that, we first have to think about a situation. Some specific events happened by means of running or transportation, such as the $\{(P_1, P_2)\}$ in the sample of Sect. 4.1, who are made on the train, and the rate of location change is fast, so the photo distance is relatively large. For example, the time interval between P_1 and P_2 is just 39 s, the photo distance is 1057.3 m, and the rate of location change is 27 m/s. It may lead to wrong effect to the fusion of time and distance. Considering actual situations, when the speed is faster than 2 m/s, it can generally be concluded that the object has moved by means of running or vehicles. Therefore, after calculating the rate of location change V of two adjacent photos, if $V > 2$ m/s, the photo distance L will be reduced in proportion $L = L / V$.

Then, we consider the fusion of time and distance. Because the units are inconsistent, we must normalize them at first. The normalization method used is the Min-Max Normalization method. The formula is as follows:

$$X = \frac{x - \min}{\max - \min} \tag{5}$$

The x is the value before normalization, and the X is the value after normalization. Find the maximum and minimum values from the sequence of time intervals and photo distances. Then use the normalization method to convert the time interval T to T' and the photo distance L to L' between [0, 1].

After that we use time and distance to calculate spatio-temporal distance. Some photos are taken at the same location, but at large time intervals. These photos belong to different specific events, such as P_3 and P_4. Some photos with large changes in location but small changes in time belong to the same specific event, such as P_1 and P_2. So the weight of time and location for classification are likely to be different. We set weights to the time and location for calculating the spatio-temporal distance ST_D. We detail how to select the weight in Sect. 5.2.

$$ST\_D = \lambda * T' + (1 - \lambda) * L' \tag{6}$$

(3) **Choose the best spatio-temporal distance as a classifier**

When the best weights are selected, calculate the accuracy of classification by different spatio-temporal distances. When the accuracy is the highest, the corresponding spatio-temporal distance are the best spatio-temporal distance BST_D. Take the best spatio-temporal distance BST_D as a classifier.

Then calculate the spatio-temporal distance ST_D of two adjacent photos in the personal photo sequence. Compare ST_D and BST_D, if ST_D is not bigger than BST_D, these two photos belong to the same specific event; if ST_D is bigger than BST_D, these two photos belong to different specific events.

5   Evaluation

5.1   The Training Data Sets

The object of this study is to classify personal photo collections taken by mobile phones, but there are no relevant public data sets. The public photo data sets don't contain all photos taken by a user with the mobile phone, such as Flickr. Only the photos that the user wants to share out, and the time interval of photos may be long apart. During this time interval, users actually took a lot of other photos with their mobile phones, but didn't share them. In addition, photos shared by users may not be the photos taken by their own mobile phones. But our classification method is for personal photo collections taken by mobile phones, so the public photo data sets cannot be used for experiment and evaluation.

To ensure the validity of our classification method, we collected several personal photo collections. Considering time-length of photo collections may affect experimental results, six representative photo collections with different time-length were

selected as the training data sets. The six personal photo collections were manually classified based on specific events, and the information showing in Table 2.

Table 2. Representative users' information

Users	Sex	Age	Profession	Number of photos	Number of specific events	Time-length of photo cluster
User1	Female	24	Graduate student	491	148	10 months
User2	Female	25	Graduate student	634	178	12 months
User3	Male	24	Graduate student	646	220	7 months
User4	Male	25	Graduate student	209	53	3 months
User5	Female	23	Graduate student	886	265	16 months
User6	Male	25	Graduate student	1260	300	18 months

5.2   Experiments

(1) Determine the weight in the spatio-temporal distance ST_D

We experimented with six training data sets, that is, a total of 4126 photos. When we calculated the spatio-temporal distances of these photos, we found that most of spatio-temporal distances of the two adjacent photos in the same specific events are less than 0.001, and most of spatio-temporal distances between different specific events are greater than 0.01. Therefore, the range of spatio-temporal distances we experimentally selected for classification is $0.001 \sim 0.01$. We selected different values of the weight λ to calculate the spatio-temporal distance. The value of λ is chosen from 0.1 to 0.9 with a step size of 0.1. Then we delineate the line charts of accuracy that select different spatio-temporal distances as the standard to classify photos when the weight λ takes different values.

As can be seen from the Fig. 1, when the accuracy is the highest in training data set, the weight λ is 0.7. Therefore, the spatio-temporal distances is calculated as *ST_D = 0.7* T'+0.3* L'*.

(2) Determine the best spatio-temporal distance BST_D

From the Fig. 1, it can be seen that the data set has the highest sum of accuracy when classify photos by the spatio-temporal distance of 0.002. So the best spatio-temporal distance *BST_D* is selected as 0.002.
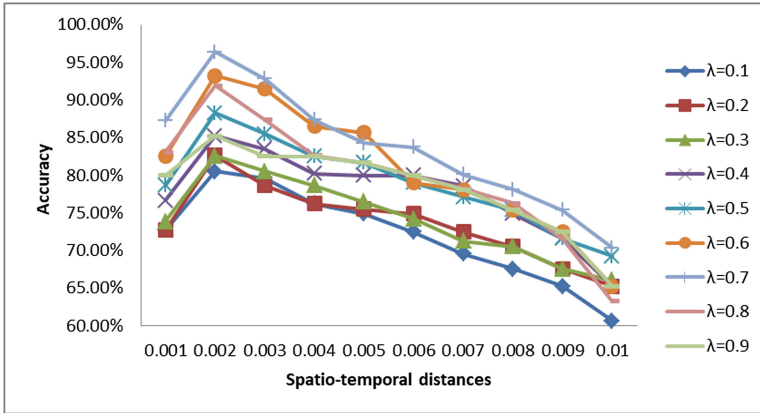
Fig. 1. The accuracy of classification by spatio-temporal distances with the weight λ taking different values

5.3 The Test Data Sets

Three users are selected randomly to test whether the method is effective, and it is also tested whether the album classifying photos based on this method can increase the speed of re-finding photos. The three users' information is as following Table 3.

Table 3. The three test users' information

Users	Sex	Age	Profession	Number of photos	Number of specific events	Time length of photo cluster
User7	Female	24	Graduate student	866	202	12 months
User8	Male	25	Graduate student	653	224	16 months
User9	Female	24	Graduate student	524	186	8 months

5.4 Evaluations of the Classification Method

In order to verify the accuracy of the method, we use the three test users' photo collections to classify photos with the spatio-temporal distance 0.002. Accuracy is a measure widely used in the field of information retrieval and statistics to evaluate the quality of results. We use the accuracy of the classification results to evaluate the methods. The accuracy is shown in Fig. 2. Since there is no relevant method of classifying based on specific events, we can't conduct comparative experiments with other methods in terms of classification accuracy.
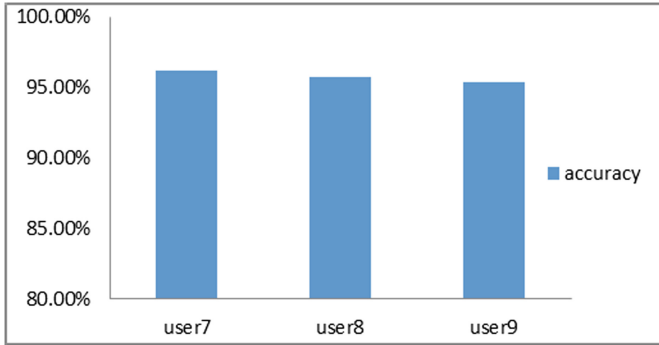
Fig. 2. The accuracy of classification by three users with the spatio-temporal distance 0.002

As shown by the Fig. 2, the accuracy of the classification based on shooting time and location is even over 95%. The biggest advantage of the proposed method is the time complexity is low, closing to O(n).

The reasons why the accuracy of the experimental results is less than 100% are that a few specific events' time interval is really small and the change in location is little. For example, one person takes a photo of the related knowledge of the papers, and then takes a selfie next minute in the laboratory. Or photos with long time intervals and little change in location, but still are the same specific event, such as $\{(P_6, P_7)\}$ in the sample of Sect. 4.1. Such situations can't be identified by this method. So this method has some limitations.

5.5   Evaluations of the Specific Event-Based Album

To verify whether the strategy classifying photos based on specific events can improve the efficiency of re-finding photos, we used the Android Studio software and the Recyclerview framework to complete the photo album based on this classification approach. In order to compare with the current photo albums in re-finding photos, the layout of design album imitated the current photo albums' structure. But the outermost layer of the album we designed is the representative photos of specific events, not all the photos. Each photo represents a specific event. Clicking the photo will jump to another page with all photos of this specific event. Currently, the characteristics of representative photos have not been studied yet, so we selected the first photos taken in specific events as the representative photos.

At present, there are a few mobile phones contains the photos classification by geographic information, events or face recognition. But there are some photo album applications including relatively complete photo classification, from which we have chosen one application with faster classification rate, and higher classification accuracy: ``time album''. We compared the speed of re-finding photos in four ways storing photos taken by mobile phones: the classification based on specific events, the classification based on geographic information, the classification based on events and the current photo album. Fig. 3. shows the albums in four ways storing photos taken by

mobile phone. Fig. 3.(a) shows the layout of the current photo album. Fig. 3.(b) and
(c) respectively show the classification based on geographic information and the
classification based on events in the time album. Fig. 3.(d) shows the album we
designed classifying photos based on specific events. We let three test users use the
four ways to re-find photos, and record the time spent on each searching. The re-finding
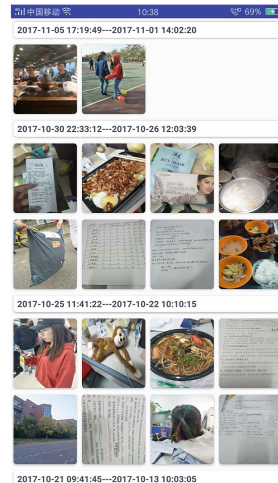results are shown in Table 4.

(a)

(b)

(c)

(d)

Fig. 3. (a) shows the layout of the current mobile phones' album; (b) shows the classification
based on geographic information in the time album; (c) shows the classification based on events
in the time album; (d) shows outermost layer in the album we designed with each photo
representing a specific event

Table 4. The users' re-finding time

Users	Photos	Current album	Based on geographic information	Based on events	Based on specific events
User7	Photo$_{a1}$	6.93 s	13.56 s	8.66 s	3.29 s
User7	Photo$_{a2}$	9.82 s	31.24 s	24.13 s	4.33 s
User7	Photo$_{a3}$	18.03 s	26.1 s	15.74 s	23.5 s
User7	Photo$_{a4}$	19 s	13.62 s	25.74 s	6.63 s
User7	Photo$_{a5}$	28.12 s	40.45 s	24.56 s	10.24 s
User8	Photo$_{b1}$	25.6 s	33.23 s	40.69 s	10.47 s
User8	Photo$_{b2}$	32.96 s	23.65 s	16.78 s	8.51 s
User8	Photo$_{b3}$	22 s	45.29 s	19.56 s	9.5 s
User8	Photo$_{b4}$	15.1 s	19.64 s	20.77 s	12.3 s
User8	Photo$_{b5}$	26.76 s	15.96 s	19.34 s	12.85 s
User9	Photo$_{c1}$	36.94 s	48.7 s	39.5 s	28.61 s
User9	Photo$_{c2}$	13.95 s	9.52 s	16.54 s	6.31 s
User9	Photo$_{c3}$	26.64 s	24.33 s	18.1 s	11.17 s
User9	Photo$_{c4}$	21.7 s	17.87 s	24.33 s	15.73 s
User9	Photo$_{c5}$	12.3 s	9 s	8.51 s	6.7 s

As can be seen from the results of finding photos, our classification method used to photo albums can greatly improve the speed of finding photos. But there are one photo's re-finding time with our method is longer than that of other methods. By analysis, we find there are three main reasons:

(1) During the experiment, each re-finding process used the classification based on specific events firstly. Because of the person's own memory function, they would have some vague memory about the photos' information. It would be favorable when users used the other three ways to re-find the same photos.
(2) This is a new classification method. Users are not accustomed to using it, but they are skilled at using their own mobile phone's original album or other existing photo classification methods.
(3) The outermost representative photos selected by this album are not representative enough, and the user does not recall the specific event by this photo.

Even though there are some unfavorable factors to using this new classification to re-find photos, the results show that the classification approach based on specific events can greatly improve the speed of re-finding photos.

Thus, this classification strategy for albums can greatly improve the efficiency of re-finding photos. Because this approach can greatly reduce the time taken for sliding the screen up and down to re-find a photo, and the photos people often reuse are those ones recording information, which almost are on the outermost layer of the album. What's more, this classification strategy conforms to people's memory habits. Thus it saves time for re-finding.

6   Conclusions

This paper proposes a photo classification strategy which classifies photos based on specific events and meets the human memory habits. We also propose a classification method based on shooting time and location with high accuracy and low computational cost. It can be used for classifying photos and greatly improve the efficiency of re-finding photos. However, the accuracy of classification results cannot reach 100%, and a few photos' re-finding time is longer than using the original photo album. The next step we plan to consider combining the shooting time, location and the photos' semantic content to find a more accurate method classifying photos based on specific events. Besides, we'll do more research about the classification strategy to improve the re-finding efficiency.

References

1. Lo, S.H.: Design and implementation the incremental clustering algorithm for geotagged photos on a map-enabled photo web service. In: Tenth International Conference on Mobile Data Management: Systems, Services and MIDDLEWARE, pp. 666–671. IEEE (2009). https://doi.org/10.1109/mdm.2009.114
2. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based landmark and event detection for tagged photo collections. IEEE Multimed. **18**(1), 52–63 (2011). https://doi.org/10.1109/MMUL.2010.68
3. Kumar, V., Namboodiri, A., Jawahar, C.V.: Semi-supervised annotation of faces in image collection. Signal Image Video Process. **12**(1), 1–9 (2017). https://doi.org/10.1007/s11760-017-1140-5
4. Brenner, M., Izquierdo, E.: Joint people recognition across photo collections using sparse Markov random fields. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014. LNCS, vol. 8325, pp. 340–352. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04114-8_29
5. Xu, Y., Peng, F., Yuan, Y., Wang, Y.: Face album: towards automatic photo management based on person identity on mobile phones. In: IEEE International Conference on Acoustics, Speech and Signal Processing (2017). https://doi.org/10.1109/icassp.2017.7952713
6. O'Hare, N., Smeaton, A.F.: Context-aware person identification in personal photo collections. IEEE Trans. Multimed. **11**(2), 220–228 (2009). https://doi.org/10.1109/tmm.2008.2009679
7. Andrade, D.O.S.D., Maia, L.F., Figueirêdo, H.F.D., Viana, W., Trinta, F., Baptista, C.D.S.: Photo annotation: a survey. Multimed. Tools Appl. 1–35 (2016)
8. Kim, J., Lee, S., Won, J.S., Moon, Y.S.: Photo cube: an automatic management and search for photos using mobile smartphones. In: IEEE Nineth International Conference on Dependable, Autonomic and Secure Computing, vol. 19, pp. 1228–1234. IEEE (2012). https://doi.org/10.1109/dasc.2011.199

9. Ulges, A., Worring, M., Breuel, T.: Learning visual contexts for image annotation from flickr groups. IEEE Trans. Multimed. **13**(2), 330–341 (2011). https://doi.org/10.1109/tmm.2010.2101051

10. Zigkolis, C., Papadopoulos, S., Filippou, G., Kompatsiaris, Y., Vakali, A.: Collaborative event annotation in tagged photo collections. Multimed. Tools Appl. **70**(1), 89–118 (2014). https://doi.org/10.1007/s11042-012-1154-5

11. Bacha, S., Allili, M.S., Benblidia, N.: Event recognition in photo albums using probabilistic graphical model and feature relevance. In: International Conference on Pattern Recognition, pp. 2819–2823. IEEE (2017). https://doi.org/10.1109/icpr.2016.7900063

12. Cao, L., Luo, J., Kautz, H., Huang, T.S.: Image annotation within the context of personal photo collections using hierarchical event and scene models. IEEE Trans. Multimed. **11**(2), 208–219 (2009). https://doi.org/10.1109/tmm.2008.2009693

13. Bosselut, A., Chen, J., Warren, D., Hajishirzi, H., Choi, Y.: Learning prototypical event structure from photo albums. In: Meeting of the Association for Computational Linguistics, pp. 1769–1779 (2016). https://doi.org/10.18653/v1/p16-1167

14. Wang, Y., Lin, Z., Shen, X., Mech, R., Miller, G., Cottrell, G.W.: Recognizing and curating photo albums via event-specific image importance (2017)

15. Tang, F., Tretter, D.R., Willis, C.: Event classification for personal photo collections. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 877–880. IEEE (2011). https://doi.org/10.1109/icassp.2011.5946544

16. Yuan, J., Luo, J., Wu, Y.: Mining compositional features from GPS and visual cues for event recognition in photo collections. IEEE Trans. Multimed. **12**(7), 705–716 (2010). https://doi.org/10.1109/tmm.2010.2051868