Signature-Image-Based Event Analysis for Personal Photo Albums [*] [†]

Minh-Son Dao Duc-Tien Dang-Nguyen[‡] Francesco De Natale

mmLAB - Department of Information Engineering and Computer Science
University of Trento
Via Sommarive, 14I-38123 POVO(TN), Italy
(dao, dangnguyen, denatale)@disi.unitn.it

ABSTRACT

Quick reorganizing and draft annotating personal photo albums under event scheme is an emerging trend. In this research, a method has been developed to meet such requirements using the idea of *gist* and mosaic art so that viewers could understand the meaning of a whole scene without paying much attention in individual details. First, given a photo album, all chronologically ordered images are normalized to a smaller size, and then mosaicked side-by-side to create a signature image representing for that album. Next, by integrating the optimized linear programming with the color descriptor of the signature image, not only the event-type of the album but also all sub-event-types of the sub-sequence photos are decided. More than 19,000 images of five varied event-types have been used to evaluate the proposed method. Experimental results show that the proposed method could detect events towards annotation and re-organization of personal photo albums with high accuracy at a rapid speed.

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: Scene Analysis; H.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Algorithms

Keywords

Gist, Mosaic art, Optimization Linear Programming, Event Analysis, Personal Photo Album

1. INTRODUCTION

It has been said that a picture is worth a thousand words. Thus, there is no surprise when people prefer to use photos to index events they have been involved in. Studies in cognitive psychology have shown that when looking at photos, the first thing people recall is the event itself, then who was involved in, and where and when that event happened [9]. Therefore we could say that personal photo albums are indexes of the events in life.

We have witnessed a vast increase in uploading, sharing and tagging personal photo albums on social networks within a decade. Billion of photos can be found over the Internet, on social network websites like Facebook[1], or public photo-sharing websites such as FlickR[2] and Photobucket[3], to name a few. Nevertheless, the more images people share on social networks, the more difficult to retrieve what they need. This is because of the big difference between human vision system and computer vision w.r.t understanding well the semantic meaning conveyed by images. Although content-based media retrieval has been developed significantly these days, there is still lack of successfully automatic tools to support systems as well as users to control their data [5][7]. Therefore, we would like to propose a new method to meet such an emerging requirement. The proposed method is based on two observations:

- **Common Patterns**: One of the most interesting users' habits is taking a series of images around the places where events happened. That leads to the assumption that inside one event's episode, there might be some common patterns (e.g. colors, landmarks, perspective, or objects) that could help us understand events holistically.
- **Common Structures**: In social life, there are a lot of spoken and unspoken rules human beings must follow without questioning or thinking about it. Thus, it should be reasonable to assume that real-life events should have structures. Consequently, their photo albums should also have the same structures.

Based on those two observations, we propose a method to detect event-type, organize, and annotate personal photo albums into a set of subfolders associated with the structure of

[*] Area chair: Tat-Seng Chua

[†] This work is supported by GLOCAL project (www.glocal-project.eu)

[‡] co-author

[1] www.facebook.com

[2] www.flickr.com

[3] www.photobucket.com

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
Copyright 2011 ACM 978-1-4503-0616-4/11/11 ...$10.00.

event-types. To distinguish from previous approaches focusing on individuals of events, we process these individuals as a whole - a holistic event. That helps us bypass the complexity of existing problems in computer vision, such as object detection, image segmentation, concept detection.

The rest of the paper is organized as follows. In section 2 we overview the related work on this problem. Section 3 describes our methodology for personal photo album analysis. We also introduce *Signature Image*, a low-level feature representing the gist of an event. Experimental results are shown in section 4. In section 5 includes conclusions and ideas for future work.

2. RELATED WORK

In this section, we briefly review some existing literature related to Personal Photo Collections analysis.

In [3], a method for event detection of a single photo from specific sport events is presented. Compared to previous work, authorsmade a significant improvement by using a graphical model for labeling and adding SIFT features for matching. By using only a single photo, their approach can only apply to certain events with certain photo types.

In [10], Yu et al. presented an approach for photos grouping based on the similarity of photos among users on the network based on SimRank and spectral clustering. The results of real users' collections showed accuracy significantly improved (up to 63%). Another interesting method with the same approach was introduced in [1]. This research used the conditional random field models to calculate correlation among photos in a collection in terms of time, location, and scene of the collection.

In [4] an event detection method was proposed for home photos using a conceptual graph where concepts must be detected from photos. This method used weighted histogram-intersection-based measurements to match the concepts with the event models to estimate the event type. Experimental results on 2400 photos with an event taxonomy of 5 layers and 20 categories showed that the average precision is approximately 60%.

Nevertheless, none of all information used in these methods could be easily found in photo-sharing social networks due to lack of popular input devices (e.g. GPS) as well as high accuracy of user-generated-content data being generated and shared. Besides, the complexity of these methods is very high due to a high computational cost of each member's problem (e.g. object detection, image segmentation, concept detection). Moreover, most of these methods treat images as individuals instead of considering as a part of the whole event, though we know that "The whole is different from the sum of its parts" (Aristotle in the Metaphysics).

3. METHODOLOGY

In this section, a series of algorithms to create signature images (SIB), and to analyze photo albums by using SIB are introduced and discussed thoroughly. For clarification, before discussing our methodology in details, we clarify some terminologies and assumptions as follows: (1) a personal photo album or photo album is used to describe a set of images taken by a single digital image device during the time a certain real-life event happened. All images must have UTC time, and (2) Let E_k, P_i^k, and $I_j^{k,i}$ denote k^{th} event type, i^{th} photo album of event type E_k, and j^{th} image of photo album P_i^k, respectively. Let γ, λ, and T denote *similar* threshold, *updating* threshold, and size of normalized image, respectively.

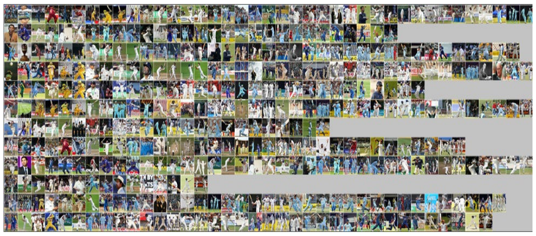


Figure 1: SIB of *Cricket* Event Type

3.1 SIB:Signature Image

Given an unknown event photo album P contained N images $\{I_j\}_{j=1..N}$, the SIB of this event is created by using Alg. 1.

Algorithm 1: *Creating Signature Image for a Given Event*

```
1. Init $SIB$ as 0x0 color pixels
2. FOR ALL $I_j$ DO
    (a) Scale all images $I_j$ to TxT color pixels
    (b) Sort time ascending these images. Let $\widehat{I}_j$ denotes images after sorting so that $\forall j$, $time(\widehat{I}_{j-1}) < time(\widehat{I}_j)$
    (c) Resize $SIB$ to $(T \times |P|) \times T$
    (d) Mosaic time ascending all images $\widehat{I}_j$ to $SIB$ at region (left, top, right, bottom)=$(0, 0, T \times |P|, T)$
3. END ALGORITHM
```

Torralba et al [8] proposed a research method which allows us to scale an image to a smaller one without losing significat information. They prove that the ability of gathering information is remarkably tolerant to degrations in image solution. Moreover, psychophysics has shown that with a single glance of an image, human can perceive the meaning of event conveyed in such an image [6]. Thus, mosaicking all images of an event into one big picture could help to understand that event better - in a gist manner, or even to create a new perspective of an event similar to the way the mosaic arts work.

Given N photo albums $\{P_i^k\}_{i=1..N}$ of event type E_k, a new set of *Signature Image* of this event type, $SIB^k = \{SIB_i\}_{i=1..N}$, is created. For visualization and storage, we mosaic all SIBs row by row to create an unique image SIB^k. Fig. 1 illustrates the image SIB^k of k^{th} event type, in this case E_k=*Cricket*, and N=12. It should be noted that each row in Fig. 1 denotes one *Cricket* photo album.

3.2 Event Type Classification

Now we focus on the issue of event type classification using SIB. Assume that there are K signature images $\{SIB^k\}_{k=1..K}$, each SIB^k represents an event type E_k. Given an unknown-event-type photo album P, the Alg. 2 is used to decide to which event type the photo album P belongs.

Algorithm 2: *Classifying event type using SIBs*

```
1. Calculate $SIB$ of photo album $P$ using Alg. 1
2. $D = 0$
```

```
3. FOR ALL $SIB^k$ DO
    (a) Calculate dissimilarity distance between $SIB$ and $SIB^k$: $d_k = d(SIB, SIB^k)$
    (b) $D = D \cup \{d_k\}$
4. Calculate $d_{min} = min(d_k)$, and $idx = argmin_k(d_k)$, where $d_k \in D$
5. IF $d_{min} \leq \gamma$ THEN
    (a) $SIB^{temp} = SIB^{idx}$
    (b) Update $SIB^{temp} = SIB^{idx} \cup SIB$
    (c) IF $d(SIB^{temp}, SIB^{idx} > \lambda)$ THEN $SIB^{idx} = SIB^{temp}$
    (d) Return $idx$ (the given photo album belongs to event type $E_{idx}$)
6. ELSE the given photo album is assigned as unknown event type
7. END ALGORITHM
```

In fact, the main idea of using SIB to detect event type of an unknown event is to measure the similarity between two SIBs: one identifies event type (SIB^k), and the other represents the unknown event SIB_i. The larger the distance similarity is, the better accuracy for event type classification. Therefore, using SIB can simplify the complexity of event type classification by turning the problem to color image retrieval with a small number of images (SIBs).

SIB event type classification is an unsupervised method. It also has an ability to enhance its knowledge throughout its life. The more photo albums the users update to a certain event type, the more characteristic information of such event type accrues by simply increasing its SIB's size. However, to prevent the oversize of SIB^k, when an incoming unknown event is classified successfully into a certain event type, whether or not its SIB image could be mosaicked to SIB^k is decided by updating threshold (see 5.c in Alg. 2).

Color feature is also popular in content-based image retrieval community due to its straightforward information and low computational cost [5]. There are several state-of-the-art methods using color feature as the main feature for comparing two images. Thus, we decided to use colors as the main descriptor to calculate $d(SIB, SIB^k)$.

3.3 Event Analysis

This section introduces a method by which a photo album could be divided into a set of subfolders sharing the same semantic meaning. As we mentioned earlier, most of real-life events have their own structure influenced by their culture, education, or social life level. For example, one western marriage often has a rehearsal party, a ceremony with a bride, a groom, and a priest, and a party at the end; a soccer match usually starts with the first half and is followed by the second half. Therefore, if we can define the structure of those events, we can reorganize and annotate photo albums by using SIB.

Assuming that there is a set of photo albums $P^k = \{P_i^k\}$, $i = 1..N$ belongs to event E_k where N is the number of photo albums. We can also assume that there is an event model of event E_k: $ME_k = \{SE_{k,j}\}$, $j = 1..M$ where M is the number of sub-event-types, and $SE_{k,j}$ is the j^{th} sub-event-type of E_k. Infact, event type and sub-event-type are equal with regards to semantic meaning. They are used to express the meaning of social activities or complex actions. Hence, we can use Alg. 2 to create SIBs for sub-event-types.

For each photo album P_i^k, based on event model, we manually group temporally successive images into different sub-photo-albums whose contents share the same semantics; each sub-photo-album represents for each sub-event-type SE. Then, we using Alg. 2 to create SIBs for these sub-photo-albums. Let $SIB^{k,j}$ denote the SIB of $SE_{k,j}$.

At this point, if we consider:(1) the event model ME_k as a *corpus*; (2) each $SIB^{k,j}$ as a *chunk of speech units* in the corpus ME_k; and (3) SIB of the photo album P as a *sentence* whose characters are I_j (see Alg. 1), our problem could be seen as *text to speech synthesis* problem. There are several techniques to solve this problem, such as HMM, Maximum Expectation, or Maximum Matching Algorithm [2]. We, however, choose the last one to create sub-photo-albums due to its simplicity and popularity.

4. EXPERIMENTAL RESULTS

The dataset used in our experiment is downloaded from Picasa[4]:207 photo albums (19101 images) of five various event-types covering sport, social, and natural landscape categories are downloaded from Picasa Web Albums to form the testbed: *Basketball* (30 albums, 2498 images), *Cricket* (35 albums, 3525 images), *SkiHoliday* (33 albums, 3165 images), *Graduation* (34 albums, 3634 images), and *Marriage* (75 albums, 3634 images). We have also obtained a thorough ground truth annotation for every photo album as well as UTC time information for each image.

We have also obtained a thorough ground truth annotation for every photo album as well as UTC time information for each image.

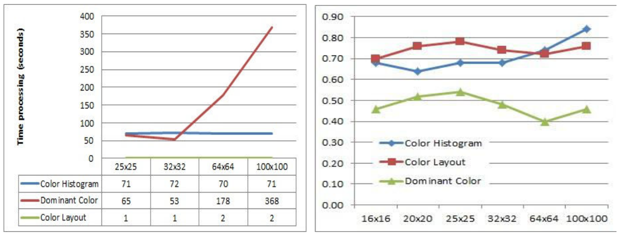


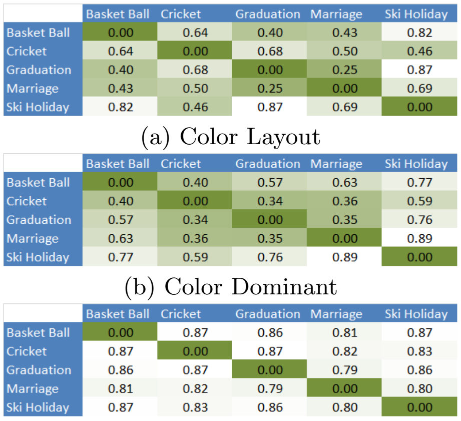
Figure 2: (left)Comparing Time Processing with Different Color Descriptors; (right)Average Recall of Event-type Classification with Different Color Descriptors and Image Size

First, we would like to discuss the accuracy and time processing of the proposed method. The results are illustrated in Fig. 2. In general, the accuracy of classification is very high, and the time processing is fast.

Second, we checked the misclassification rate among event types (see Fig. 3). Noted that each cell of diagram denotes the dissimilarity degree between two event types, and is normalized to [0,1]. It means that the higher the cell's value is, the smaller misclassification rate.

Third, we discuss the convergence of SIB's size and the method for choosing *updating* threshold. In Fig. 4, after 10-15 updating times, the distance similarity between the SIB before updating and the SIB after updating, $d(SIB_{i-1}^k, SIB_{i-1}^k \cup SIB)$, starts converging to a small value (see the

[4] www.picasa.com



	Basket Ball	Cricket	Graduation	Marriage	Ski Holiday
Basket Ball	0.00	0.64	0.40	0.43	0.82
Cricket	0.64	0.00	0.68	0.50	0.46
Graduation	0.40	0.68	0.00	0.25	0.87
Marriage	0.43	0.50	0.25	0.00	0.69
Ski Holiday	0.82	0.46	0.87	0.69	0.00

(a) Color Layout

	Basket Ball	Cricket	Graduation	Marriage	Ski Holiday
Basket Ball	0.00	0.40	0.57	0.63	0.77
Cricket	0.40	0.00	0.34	0.36	0.59
Graduation	0.57	0.34	0.00	0.35	0.76
Marriage	0.63	0.36	0.35	0.00	0.89
Ski Holiday	0.77	0.59	0.76	0.89	0.00

(b) Color Dominant

	Basket Ball	Cricket	Graduation	Marriage	Ski Holiday
Basket Ball	0.00	0.87	0.86	0.81	0.87
Cricket	0.87	0.00	0.87	0.82	0.83
Graduation	0.86	0.87	0.00	0.79	0.86
Marriage	0.81	0.82	0.79	0.00	0.80
Ski Holiday	0.87	0.83	0.86	0.80	0.00

(c) Color Histogram

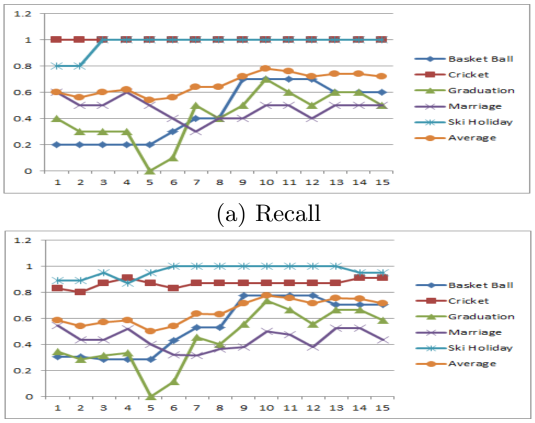
Figure 3: Distance matrices of Event-type Classification with Different Color Descriptors, T=25

average line in diagrams). Therefore, after a certain number of updating times, we easily detect the *updating* thresholds by calculating the average of the total range of $d(SIB^k_{i-1}, SIB^k_{i-1} \cup SIB)$ during the stabilizing period.

Finally, we compare our method with two existing methods focusing on photo album analysis: (a) Lim et al [4], and (b) Cao et al [1] (please refer to previous sections to have more information how these methods work). From the comparison (Fig. 5), we could conclude that our method works better than the others.

5. CONCLUSION

We have presented SIB method, and a novel event type analysis for personal photo collections. Our method shows the advantages of fast computing and is easy to implementation. Moreover, it is an unsupervised method and has the ability to enhance its knowledge throughout its life. This is the first time the mosaic art and the advantages of human visual system are integrated to classify event type towards organizing personal photo albums. With our method, there is no need to detect objects, landmarks or the meaning of tags to understand the global information of events. The experiments show that our method has very high accuracy and is extremely fast. Since event type classification plays an important role in many problems of event mining in multimedia, such asevent detection or event recognition, our method could be used not only in event type classification but also for other relevant problems. In future work, we will create better SIBs by integrating spatial information in terms of number of photos with geo-tag increasing, and also by finding the suitable distances as well as features for the improved SIBs qualifications.



(a) Recall

(b) F-scores

Figure 4: Convergence of SIB's size

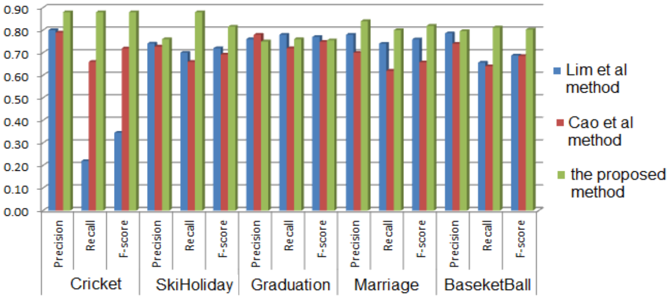


Figure 5: Photo album analysis; Comparing to other methods

6. REFERENCES

[1] L. Cao, J. Luo, H. Kautz, and T. Huang. Image annotation within the context of personal photo collections using hierarchical event and scene models. *IEEE Trans. on Multimedia*, 11(2):208–219, Feburary 2009.

[2] T. Dutoit. *An Introduction to Text-to-Speech Synthesis*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[3] L. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recogntion. In *ICCV*, pages 1–8, 2007.

[4] J. H. Lim, Q. Tian, and P. Mulhem. Home photo content modeling for personalized event-based retrieval. *IEEE Multimedia*, 10:28–37, 2003.

[5] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Journal of Pattern Recognition*, 40:262–282, 2007.

[6] A. Oliva and A. Torralba. Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research*, pages 23–36, 2006.

[7] P. Sandhaus and S. Boll. Semantic analysis and retrieval in personal and social photo collections. *Journal of Multimedia Tools and Application*, 51:5–33, 2011.

[8] A. Torralba, R. Fergus, and W. Freeman. Tiny images. *MIT Technical Report - MIT-CSAIL-TR-2007-24*, 2007.

[9] W. Wagenaar. My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, 18(2):225–252, 2004.

[10] J. Yu, X. Jin, J. Han, and J. Luo. Mining personal image collection for social group suggestion. In *IEEE ICDMW*, pages 202–207. IEEE, 2009.