Representative Picture Selection from Albums

Gábor Szűcs, Tamás Leposa, Sándor Turbucz
Dept. of Telecommunications and Media Informatics
Budapest University of Technology and Economics
Budapest, Hungary, 2nd Magyar Tudósok Krt., H-1117
e-mail: szucs@tmit.bme.hu, lepi_t@sch.bme.hu, sandor.turbucz.work@gmail.com

Abstract—**The paper is concerned with managing image album, where the picture set (containing very similar or different images) in each album is given. The goal has been to select the most representative pictures from the album. Our solution is based on the clustering of the images. The developed clustering procedure takes the large variety of the pictures and different type of image features into account. We have solved the incomplete feature value problem as well. The central pictures of the largest clusters are selected for representing the album.**

Keywords - representative picture; k-means++ clustering; qualitative and quantitative features; content features; EXIF data;

I. Introduction

There are lot of solutions and systems (e.g., FotoFile [3]) at the multimedia organization and retrieval; and demands are growing with new functionalities in the future too. A large part of these deals with personal photograph retrieval [2].

For a good organization the human persons usually put the pictures into albums as they wish (may be based on users' feeling). But there is a problem at the large set of pictures and albums: it is not easy to give representing issues (e.g., titles) for these albums. A representative image and its thumbnail is an ideal solution for this problem. The goal of our work has been to select the most representative picture from each album automatically without human interaction.

We consider very realistic situations, where the images come from different sources (camera, edited or created by software), the resolutions are various, the qualities are also different, so the variety of them is large.

For the above mentioned problem with realistic situations we present a solution in this paper. The structure of the paper is the following: Section II describes the background, our solution with clustering is detailed in Section III, brief conclusion and future works can be found in Section IV.

II. Background

Finding the interesting photos from collections is a similar task to our goal, but the selection of them is always based on user feedbacks. (i) Commercial systems such as Flickr use an interaction mechanism for sampling the collection, it relies on social activity analysis for determining the notion of interestingness. Photo album creation can benefit from leveraging information learned from many users in regard of the album's content, structure, and semantics [4]. (ii) An alternative technique [1] is based on content analysis, the solution uses the combination of visual attention models and an interactive feedback mechanism to compute interestingness.

For representative photo selection and smart thumbnailing an other solution [5] uses the results of near-duplicate detection. Near-duplicate photo pairs are first determined, and the relationships between them are modeled by a graph. The most typical one is then automatically selected by examining the mutual relation between them. For smart thumbnailing, the region-of-interest of the selected representative photo is determined based on locally matched feature points, which is a view different from conventional saliency-based approaches [6][7].

The related works in the topic of representative images have solved the problem in three different ways: textually interesting [4], mutually distinct [5] and presence of faces in the image [12] or combination of them [1]. These works have used content features of the images. Only one or two papers have mentioned a few EXIF data, but these have been the *time* and the *camera name* [12] only. The works have not dealt with all EXIF data from camera; these metadata could be equal to content features.

III. Solution with Clustering

In picture selection procedure different types of features can be considered. If we consider only content features of the images, then the search space for the most representative picture is narrow. If we take both the content and the metadata features (from the camera) into account, then the search space will be wider. In this wide space the search procedure may find easier the most representative picture. The consequence of the narrow space is the possibility to take a bad decision in selection of the most representative picture. E.g., if all photos – except one or few – are taken by flash, then users probably will not consider a *photo without flash* as a representative picture. Another example is about the focal length: if all photos – except one or few – are taken with ordinary focal length (tableau), then a picture with small focal length (portrait) will not representative. Thus our solution considers both the content and the metadata features (EXIF data from the camera).

A. Overview of the picture selection procedure

The first idea for selecting the most representative picture in an album is choosing the central picture in the place of the pictures, where the place of the pictures is a vector space, and each picture is transformed into a point in this space. In this space the central picture can represent the whole set in an album.

But it occurs many times, that the album consists of different larger groups of images, where the pictures are similar in a group and far away between groups. In this case a strange situation can occur, where the distance between the central picture and the others is large (this image is alone), and the central picture will not be representative.

In order to avoid this often occurring situation we suggest a solution using clustering. After the clustering the procedure suggests the nearest picture of central point of the largest cluster. The solution contains some phases:

- Content feature values are extracted from the pixel data of the picture.
- Metadata features are the EXIF (Exchangeable Image File Format [9]) data.
- Clustering algorithm calculates the clusters of the pictures.
- Central point is determined of each cluster.
- The closest picture to the central point – namely the central picture – is marked as candidate for selection.
- The central picture of the largest cluster is selected for representative picture.
- If it is necessary more than 1 picture for representing the whole album, then central pictures of the second largest, third largest, etc. cluster are selected.

B. Features for clustering

In our picture selection solution the challenge has been taking both the content and the metadata features into account (correlation may occur between features). The content features are based on the statistics of RGB values of the picture points: mean, variance, mode, range, quartiles. There are 3 features for each statistical type: a feature related to red (R), one related to green (G) and one related to blue (B). These content features characterize the image and the values of them are indifferent from the orientation and size of the pictures.

The metadata features are the EXIF data of the pictures made by cameras. These data are not always available because an album can contain not only photos, but drawn, animated, edited pictures as well. (The absent metadata features naturally may influence the goodness of the result.) The used metadata have been the all accessible EXIF data: exposure program, contrast, flash, light source, metering mode, saturation, scene capture type, sharpness, white balance, image orientation, exposure time, F number, focal length in 35 mm film, ISO speed ratings. Some of these features are qualitative, others are quantitative. E.g., exposure program may be portrait, landscape, sea, mountain, etc. (so this is a qualitative feature), flash is a binary feature with two values: yes or no.

C. Clustering algorithm with content and metadata features

The content features are always available, nevertheless metadata features may be partly or totally deficient, which leads to problem in the comparison of pictures.

A distance (similarity) value needs for every picture pair for the clustering. The difficulties come from the different feature types (content and the metadata features), the different scales (qualitative and quantitative), and the deficient metadata.

The quantitative values of the pictures can be presented in a vector space, where each coordinate axis is a quantitative feature; and each picture is a point in this vector space. The distance between two pictures is calculated by the Euclidean distance, the number of all features gives the dimension.

If one of the values (of two pictures) at some features is missing, then the Euclidean distance formula can not be used. Let us omit the squared differences in the sum, where one of the two values is missing. The number of the rest of features is *k*. Instead of the Euclidean distance formula we have used normalized version of the distance (related to 1 dimension) for the missing value problem.

We have used normalized values (between 0 and 1) in x_i and y_i for the quantitative features, but we should have solved the other problem – distance calculation for qualitative features – as well. In the clustering the most frequent work is calculation the distance between a point and a cluster. For this we have used an idea about the histogram of the given cluster. Let us denote the mode of the histogram by *mode*, the frequency of the given x_i feature by $f(x_i)$. The distance only in the examined coordinate axis is defined in (1), where C is the examined cluster.

$$d_i(x_i, C) = \frac{f(\mathrm{mode}) - f(x_i)}{f(\mathrm{mode})} \qquad (1)$$

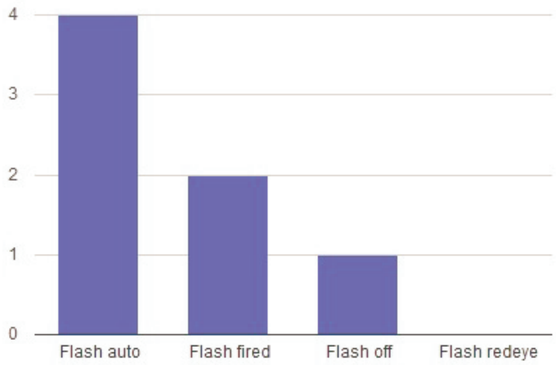

Figure 1. Example histogram for a qualitative feature

If the cluster contains only one element, then this will be zero or one. Fig. 1. shows an example histogram for a flash qualitative feature, where the values can be "flash auto", "flash fired", "flash off" or "flash redeye". The mode is the "flash auto" and the frequency of this is 4. In the comparison of this cluster and an image four different results can be: if at the examined picture the feature is "flash auto" ("flash fired", "flash off", "flash redeye"), then the distance is 0 (1/2, 3/4, 1 respectively).

In order to tune the relative importance of the features, particularly the balance between the content and the metadata features, we have introduced weights for each feature, and the distance between two pictures is modified as can be seen in (2). In the current phase of our implemented system the w_i weights have been determined manually (balanced between content and metadata features) based on the results of thousand pictures (there was a fine tuning), but we have intend to estimate the weights by automatically using supervised learning, where albums and their most representative pictures are given as training set.

$$d(x,y)=\sqrt{\frac{1}{k}\sum_{i=1}^{k}w_i(x_i-y_i)^2} \qquad (2)$$

D. *The k-means++ clustering algorithm*

The k-means method is a widely used clustering technique that seeks to minimize the average squared distance between points in the same cluster. Although it offers no accuracy guarantees, its simplicity and speed are very appealing in practice (it is standard practice to choose the initial centers uniformly at random from *X* space). By augmenting k-means with a simple, randomized seeding technique, a new algorithm, so called k-means++ [10] has been outlined with the optimal clustering. Preliminary experiments show that the augmentation improves both the speed and the accuracy of k-means.

The k-means algorithm begins with an arbitrary set of cluster centers, but k-means++ algorithm uses a specific way of choosing these centers. At any given time, let D(x) denote the shortest distance from a data point x to the closest center we have already chosen; so k-means++ algorithm is the following:

- 1a. Choose an initial center c_1 uniformly at random from *X*.
- 1b. Choose the next center c_i, selecting $c_i = x' \in X$ with probability p, where p can be calculated by (3).

$$p=\frac{D(x')^2}{\sum_{x\in X}D(x)^2} \qquad (3)$$

- 1c. Repeat Step 1b until we have chosen a total of *k* centers.
- 2. For each $i \in \{1, \ldots, k\}$, set the cluster C_i to be the set of points in *X* that are closer to c_i than they are to c_j for all $j \neq i$.
- 3. For each $i \in \{1, \ldots, k\}$, set c_i to be the center of mass of all points in $C_{i,}$, as can be seen in (4), where ${}_mc_i$ and ${}_mx$ is the m^th coordinate of the c_i point and x point respectively.

$${}_mc_i=\frac{\sum_{x\in C_i}{}_mx}{|C_i|} \qquad (4)$$

- 4. Repeat Steps 2 and 3 until C no longer changes [10].

Choosing the number of the clusters in k-means++ algorithm is a sensitive parameter for the goodness of the results. We have used the rule of thumb formulated in (5) for the determination of the clusters.

$$k=\sqrt{n/2} \qquad (5)$$

After the clustering the closest picture to the cluster central point of the largest cluster is selected for the most representative image. Our solution is able to select more than 1 picture for representing the whole album with choosing central pictures of the second largest, third largest, etc. clusters. This will be very useful at characterization of large image sets, where not only one picture characterize the all images. At this case similar pictures at the selection would be a wrong result, which is avoided in our solution because of very different pictures.

E. *Results*

We have implemented our ideas and solution described above in Python programming language. The Python Imaging Library (PIL) [11] has been used for the image handling. This library contains useful functions for basic content features. The extraction of EXIF features has been solved also in Python.

The method has been just now implemented, the evaluation can be subjective (users' decisions may be based on emotion anyway). There is subjective evaluation of images in other works (e.g. consumer photography [8]) as well.

We have used the implemented program for personal images. In Fig. 2. a little part of the album can be seen: two rows are the results of the clustering and the pictures with different border are the central images. The cluster of the pictures in the bottom row is largest cluster, so the 6th image is the most representative picture in the album.

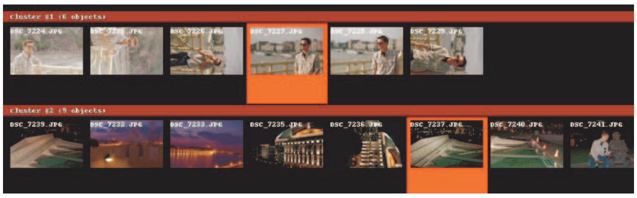

Figure 2. Examples for clustering and picture selection

F. Experimental evaluation

We have collected 600 pictures from different sources (camera with EXIF data, other sources without EXIF data), and we have organized them in 20 albums with different topics (party, holiday, town, or unified mood, etc.). Three human evaluators have selected the most representative pictures (as first in the order), then second ones, etc., so they have sorted the pictures in each album. Our implemented solution has also selected a picture (as most representative one) in each album. These machine results have been compared with the aggregated order of three human decisions (the aggregation is based on Borda method). The machine results are not always the first in the humans' order, but at 25% of albums they are in the best 5 representative pictures (denoting by p_5=25%). Furthermore we have summarized how many cases, where the machine results are in the best 10 representative pictures, we have counted 14 cases in 20 albums, so this is 70% (denoting by p_{10}=70%). These figures are not excellent, but good enough. We have investigated the humans' order, and we have concluded that humans' decisions are dispersing. With cross-validation only two humans' order were considered and aggregated, then were compared with the most representative pictures of third person. The comparison results of cross-validation for the first person: p_5=30%, p_{10}=85%, for the second person: p_5=40%, p_{10}=75%, for the third person: p_5=45%, p_{10}=70%. These figures present that our automatic solution is almost good as humans' decisions.

CONCLUSION AND FUTURE WORK

This paper presents a description of a work in progress with new idea. The aim is to find the best way to automatically choose a picture from an album in order to be the best representation of it. The new idea is the consideration (in clustering) of different type of image features (content and EXIF data) with incomplete feature value possibilities. After clustering the central pictures of the largest clusters will be selected for representing the album. We have implemented this idea in Python and

In calculation of distances for clustering many features are considered, but the set of features can be expanded. We are at the beginning of this research, we intent to take more features – like texture, local features, time-based features – into account. Further development will be the automatic calculation of weights in distance formula.

ACKNOWLEDGEMENT

The research was supported by ETOCOM project (TÁMOP-4.2.2-08/1/KMR-2008-0007) through the Hungarian National Development Agency in the framework of Social Renewal Operative Programme supported by EU and co-financed by the European Social Fund.

REFERENCES

[1] K. Vaiapury and M. S. Kankanhalli, "Finding interesting images in albums using attention", Journal of Multimedia, Vol. 3., Num. 4., October 2008, pp. 2-13.

[2] P. D. B. Bujac and J. Kerins, "Developing and implementing a sparse ontology with a visual index for personal photograph retrieval" AI & Society, 2009, Vol. 24, Num. 4, pp. 383–392, DOI: 10.1007/s00146-009-0221-6.

[3] A. Kudhinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gvvizdka, "FotoFile: a consumer multimedia organization and retrieval system", CHI '99 Proceedings of the ACM SIGCHI conference on Human factors in computing systems: the CHI is the limit, May 15-20, 1999, Pittsburgh, Pennsylvania, USA , pp. 496-503, DOI: 10.1145/302979.303143.

[4] S. Boll, P. Sandhaus, and U. Westermann, "Semantics, content, and structure of many for the creation of personal photo albums", ACM Multimedia '07, Proceedings of the 15th international conference on Multimedia, Augsburg, Germany, September 24-29, 2007, pp. 641-650, DOI: 10.1145/1291233.1291385

[5] W.-T. Chu and C.-H. Lin, "Automatic selection of representative photo and smart thumbnailing using near-duplicate detection" MM '08, Proceeding of the 16th ACM international conference on Multimedia, Vancouver, Canada, October 26-31, 2008, pp. 829-832, DOI: 10.1145/1459359.1459498.

[6] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention" Vision Research 40, 2000, pp. 1489–1506

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for Rapid Scene Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, Issue 11, 1998, pp. 1254-1259, DOI: 10.1109/34.730558.

[8] A. E. Savakis, S. P. Etz, and A. C. Loui, "Evaluation of image appeal in consumer photography", In Proceedings SPIE Human Vision and Electronic Imaging San Jose, CA, 2000, pp. 111–120.

[9] Japan Electronic Industry Development Association, Digital Still Camera Image File Format Standard, (Exchangeable image file format for Digital Still Camera:Exif) Version 2.1, Dec. 1998, http://www.exif.org/dcf-exif.PDF (last access date: 2011-01-24).

[10] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding", SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 2007, pp. 1027–1035.

[11] http://www.pythonware.com/library/pil/handbook/ (last access date: 2011-01-24).

[12] E. Potapova, M. Egorova, and I. Safonov, "Automatic Photo Selection for Media and Entertainment Applications", GraphiCon'2009, Proceedings of The 19th International Conference on Computer Graphics and Vision, October 5-9, 2009, Moscow, Russia, pp. 117-124.