Journal of Mobile Multimedia, Vol. 6, No. 2 (2010) 158–169

AUTOMATIC IMAGE REPRESENTATION AND CLUSTERING ON MOBILE DEVICES

MARCO LA CASCIA MARCO MORANA
Dipartimento di Ingegneria Informatica, Università degli Studi di Palermo
viale delle Scienze ed. 6, Palermo, Italy
lacascia@unipa.it morana@dinfo.unipa.it

FILIPPO VELLA
Istituto di Calcolo e Reti ad Alte Prestazioni, Consiglio Nazionale delle Ricerche
viale delle Scienze ed. 11, Palermo, Italy
filippo.vella@cnr.it

Received August 17, 2009
Revised January 28, 2010

In this paper a novel approach for the automatic representation of pictures on mobile devices is proposed. With the wide diffusion of mobile digital image acquisition devices, the need for managing a large number of digital images is quickly increasing. In fact, the storage capacity of such devices allow users to store hundreds or even thousands, of pictures that, without a proper organization, become useless. Users may be interested in using (i.e., browsing, saving, printing and so on) a subset of stored data according to some particular picture properties. A content-based description of each picture is needed to perform on-board image indexing. In our work, the images are analyzed and described in three representation spaces, namely, faces, background and time of capture. Faces are automatically detected, and a face representation is produced by projecting the face itself in a common low dimensional eigenspace. Backgrounds are represented with low-level visual features based on RGB histogram and Gabor filter bank. Temporal data is obtained through the extraction of EXIF (Exchangeable Image File Format) data. Faces, background and time information of each image in the collection is automatically organized using a mean-shift clustering technique. Significance of clustering has been evaluated on a realistic set of about 1000 images and results are promising.

Keywords: CBIR - Content Based Image Retrieval, automatic image annotation, mobile devices

1 Introduction

The increasing number of digital image acquisition devices, equipped with large storage capacity, allows users for storing a quite large number of pictures, making the device itself a sort of digital photo album wallet. We addressed the scenario in which an user takes pictures in different sessions and different places, that is pictures belong to different *contexts*. In this case users may also be interested in using such devices to instantly manage (i.e., browse, save, print and so on) a subset of captured pictures according to some particular picture properties.

Usually, digital photo libraries are organized by keywords given by the user. This process has been observed to be inadequate since it requires users to manually associate keywords to pictures. Moreover users add few keywords for large set of images and, on the other side,

keywords tend to be ambiguous. This is even worse on mobile devices due to limited text input capabilities. Time of shooting is usually available for free since all the digital camera-phone attach a timestamp to the pictures they take, however its power in term of searching capabilities is quite limited [1].

Our point is that considering image collections stored on mobile devices, the user is mainly interested in *who* is in the picture (usually a relatively small number of different individuals) and *where* and *when* the picture was shot. *Who*, *where* and *when* are the fundamental aspects of photo information and input images can be intrinsecally split in three domains of interest[2].

In our system, the faces are extracted from images so that it is possible to identify *who* is in the picture while the remaning part of the image is considered as image context. Low-level features, based on color and texture, are used to identify different contexts (*where*) by analyzing the information stored in the image background. Nowadays more and more devices are equipped with GPS that allow to store camera position within the EXIF information, however GPS data is available only in outdoor environments so that visual analysis is required in any case. The *when* aspect is bound to when the picture was captured and is typically referred to temporal ranges or particular user events (e.g. *birthdays, weddings, travels*).

To automatically organize image data based on faces, background and time descriptors a mean-shift based approach is presented. Image features are automatically extracted and clustering parameters are automatically determined according to a proposed entropy based figure of merit.

The paper will show the following structure: an analysis of related work will be given (Sect. 2). The Sect. 3 will give an overview of the techniques we used to represent the images, while the three-domain clustering will be described in Sect. 4. Experimental results will be shown and discussed in Sect. 5. Conclusions will follow in Sect. 6.

2 Related Works

With the widespread diffusion of digital photography, personal photo collection management has become an active field of research. Several studies [3, 4, 5, 6, 7, 8] addressed the problem of semi-automatic personal photo collection management based on the observation that pictures often depict people and use this information for more effective searching and browsing. Most of the effort has been devoted to finding techniques to help the user in annotating the collection. For example Zhang et al. [9] developed a system where the user is allowed to select multiple images and assign them personal names. Then the system tries to propagate names from photograph level to face level exploiting face recognition and CBIR techniques. A similar approach has been implemeted in iPhoto 09 [10], recently proposed by Apple. iPhoto allows users to organize their libraries by using semi-automatic features detection (i.e., faces, places and events) and search them by person, location, title, album, event, or keyword. Face arrangement in photo [11] or clothes and nearby regions [12] have also been exploited to cluster the collection.

Several researchers address the problem of personal photo album management in an image clustering framework. For example in [13] the authors use color histogram and histogram intersection distance measure to perform hierarchical clustering. In Deng et al. [14] a self-organizing map is used to let the structure of the data emerge and then to browse the collection. Many techniques have been proposed to refine the automatic clustering approach with

human intervention to make cluster semantically homogeneous. Goldberger et al. [15] proposed a generalized version of the information bottleneck principle where images are clustered to maximally preserve the mutual information between the clusters and image contents. In other cases the presence of faces in an attempt to bridge the gap between visual and semantic content is exploited. For example in [16] face detection is performed on captioned images and clustering is used to associate automatically extracted names to the faces.

Previous work regarding CBIR on mobile devices has been mainly limited to particular problems like the generation of an initial query set [17], to energy efficiency [18] or to the development of a mobile front-end to traditional CBIR systems [19, 20, 21] in a client server framework. In our work we propose a fully automatic approach for image searching and browsing on mobile devices based on image clustering. Our goal is in fact the development of a system to improve user experience in managing photos on the mobile device itself without the need of trasferring and processing them on a host computer.

3 Image Representation

In this work we propose a novel approach for automatic image indexing on mobile devices. The key point is the representation of each image with multiple descriptors in a suitable form for clustering. An image can be represented in several spaces allowing to capture different aspects of input data [22]. In the proposed system, each image in the collection is represented with features related to the presence of faces in the image, features characterizing background and time information. A data oriented clustering allows to generate aggregation structures driven by statistical regularities in the represented data. The proposed process of image representation is shown in Fig. 1.

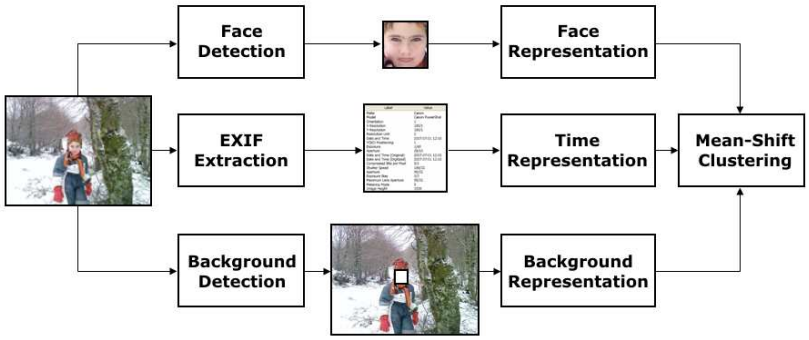

Fig. 1. Image representation for personal photo collections.

For each face in the personal album the global representation is given by:

$$\mathbf{x} = [\mathbf{x}^f, \mathbf{x}^b, \mathbf{x}^t] \tag{1}$$

where $\mathbf{x}^f \in \mathbf{R}^M$ is the representation of face in the eigenspace of detected faces, $\mathbf{x}^b \in \mathbf{R}^P$ is the background representation for the corresponding image, and $\mathbf{x}^t \in \mathbf{R}$ is the time of capture.

3.1 Face Representation

As first processing step, each image to be archived in the system is searched for faces. Some cameras are already able to perform real-time face detection to perform pre-processing tasks (e.g., autofocus). Our point is that in next future all cameras will be equipped with ad-hoc face detection circuits so that the same module could be used for indexing purposes. At this time, we chose to use the state of the art approach to face detection, that is the framework proposed by Viola and Jones [23]. A few detected faces are reported in Fig. 2.

Fig. 2. Examples of detected faces.

Several appearance-based approaches could be used for face representation. We used *eigenfaces* [24], i.e., a principal compenent analysis (PCA) technique, as it is one of the most mature and investigated method.

Given the set of all detected faces in the image collection, the mean face Ψ and the eigenvectors $\mathbf{e}$ are calculated. Each image in the data set is represented subtracting the mean image Ψ and projecting the face vector in the eigenspace. If Φ is the difference between the face and the average face, the representation in the eigenspace is $w_i = \mathbf{e}_i^T \Phi$.

The face space, as well as the average face, is learned off-line on a significant subset of the image collection and it is not updated. At any time, if most of the faces present in the image collection differ significantly from the training set, it is possible to build a new face space and recompute the projection of each detected face in the new face space.

3.2 Background Representation

Given the faces contained in the image, the remaining part of the image information is related to the context of the scene. Background information is represented in terms of color and texture features using a single, composite, vector for each image. Color information is captured through histograms in the RGB color space using 20-bin histograms of the R, G and B channels. The texture feature is composed by 90 elements obatined by applying Gabor technique [25] with 6 different filters (i.e., 3 orientations and 2 scales). For each filter the energy value is evaluated and represented as a 15 bins histogram. Thus, considering both color and texture information, for each image the background is represented by a single vector of 150 elements. Considering that features are distributed according to a gaussian density function, values near mean value are very common and - for this reason - less discriminative. To stretch values towards lower or higher values they are processed through a sigmoid.

Sigmoids are characterized by the parameters α and β as given in Equation (2)

$$f(\mathbf{x}) = \frac{1}{1 + e^{-(\alpha x - \beta)}} \tag{2}$$

The value of α is chosen to modulate the mapping of feature and get a softer or stronger stretching. The value of β is chosen to translate the sigmoid across the mean and is set to $\beta = -\mu\alpha$ where μ is the mean value.

3.3 Time Representation

Temporal data information is available through the extraction of Exif (Exchangeable image file format) data. This metadata is attached when picture is captured for storing information about camera model, exposure parameters, GPS coordinates and time (e.g., date and hour) of the image shot. Liu et al. [26] use part of this information to classify image in indoor or outdoor classes. Here we focus only on the time of capture, leaving out information referred to camera sensor and image exposure. The value stored in the time field as date and hour of capture is converted in an integer number counting seconds from a fixed data (i.e., Jan 1, 1970). Images are placed in the time line and organized according to time similarity and a parameter q is chosen to represent time scattering of samples in a coarser or finer representation.

$$t_q = floor\left(\frac{t}{q}\right) \tag{3}$$

The larger is q the more events will be mapped in the same t_q, the smaller is q the more the event temporal description is detailed.

4 Image Clustering

Faces, background and time information of each image in the collection is automatically organized using a mean-shift clustering technique. Once the input data is organized into clusters we obtain "new collections"of similar images that could be easily browsed and searched.

Mean shift is a technique for kernel density estimation that applies gradient climbing to probability distribution [27]. Given n data points $\mathbf{x}_i, i = 1, 2, ..., n$ in the d-dimensional space R^d, a multivariate kernel density estimator $\hat{f}(\mathbf{x})$ is calculated as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K(\frac{\mathbf{x} - \mathbf{x_i}}{h}) \tag{4}$$

where h is the bandwidth and the kernel $K(.)$ is the Epanechnikov kernel. Using a differentiable kernel, the estimate of the gradient density can be written as the gradient of the kernel density estimate(4):

$$\hat{\nabla} f(\mathbf{x}) \equiv \nabla \hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla K(\frac{\mathbf{x} - \mathbf{x}_i}{h}) \tag{5}$$

For the Epanechnikov kernel the density gradient estimate is:

$$\hat{\nabla} f(\mathbf{x}) = \frac{n_c}{nV_d} \frac{d+2}{h^d} \left(\frac{1}{n_c} \sum_{\mathbf{x} \in S(x)} (\mathbf{x}_c - \mathbf{x}) \right) \tag{6}$$

where $S(\mathbf{x})$ is the hyper-sphere of radius h, having volume $h^d V_d$, centered in $\mathbf{x}$ and containing L_c data points. The quantity $M_h(\mathbf{x})$ defined as

$$M_h(\mathbf{x}) \equiv \frac{1}{n_c} \sum_{\mathbf{x} \in S(\mathbf{x})} (\mathbf{x}_c - \mathbf{x}) \tag{7}$$

is called Mean Shift Vector that can be expressed as:

$$M_h(\mathbf{x}) = \frac{h^d}{d+2} \frac{\hat{\nabla} f(\mathbf{x})}{\hat{f}(\mathbf{x})} \tag{8}$$

The Mean Shift Vector at location $\mathbf{x}$ is aligned with the local density gradient estimate and is oriented towards the direction of maximum increase in density. For each point the Mean Shift Vector defines a path leading from the fixed point to a stationary point of estimated density where gradient is equal to zero.

Each picture is represented as a generic point in the feature space composed by representation in time, faces and backgrounds spaces. The Mean Shift Vector shown in Equation (7) describes a trajectory in the density space converging to points where the density is maximum. The set of all points converging to a local maximum is the *basin of attraction* for the found maximum density point. The procedure for the detection of modes in the data distribution is composed of two steps:

- Run mean shift to find stationary points for $\hat{f}(\mathbf{x})$
- Prune the found points retaining only the local maximum points

Clusters are refined through a merging procedure unifying adjacent clusters. Clusters are merged if:

$$\left\| \mathbf{y}_i - \mathbf{y}_j \right\| < \frac{h}{2} \tag{9}$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are two local maximum points, $i \neq j$, and h is the bandwidth used to estimate the distribution density.

The clustering process is driven by a set of parameters and, although the number of clusters is not fixed, the best bandwidth must be selected. To evaluate the best clustering parameters, a number of evaluation indexes have been proposed, from the older Partition Coefficient and Partition Entropy[28] to the newest as partition based on exponential separation [29]. All of them tend to capture the quality of the separation proposed by clustering. Typically these methods are oriented to fuzzy clustering more than to hard (crisp) clustering and they use an estimation of the density to evaluate the clustering performance (e.g. Parzen Windows). Since we adopted a density estimation in the mean-shift procedure, to avoid a biased clustering measure, we choose to evaluate clustering as function of scattering of hand assigned identifiers in the clusters. These identifiers are related to the image content and are the names of people in a picture - for faces domain - , the identified context - for background domain - and event

for time domain. These identifiers are usually referred as labels, indicating the ground truth for the given images. We define two indexes; the *Intra-Cluster Entropy* is defined as:

$$E_c = -\frac{1}{log(N_C) * log(N_L)} \sum_{i=1}^{N_L} \sum_{j=1}^{N_C} \frac{u_{ij}}{T_j} log \frac{u_{ij}}{T_j} \tag{10}$$

where N_C is the number of clusters, N_L is the number of *labels*, u_{ij} is the number of times the $i-th$ label is present in the $j-th$ cluster and T_j is the number of samples in the $j-th$ cluster. This index gives a measure of the entropy inside clusters. If many different labels are present in a cluster, the value of ratio u_{ij}/T_j is near the average and the value of *Intra-Cluster Entropy* is high. If a label is concentrated in few clusters and is absent in all the other the ratio u_{ij}/T_j is near 1 or near 0 and the entropy has a low value. This index measures the uncertainty of labels inside clusters.

The second index, the *Intra-Label Entropy* is defined as:

$$E_l = -\frac{1}{log(N_L) * log(N_C)} \sum_{i=1}^{N_L} \sum_{j=1}^{N_C} \frac{u_{ij}}{S_i} log \frac{u_{ij}}{S_i} \tag{11}$$

where N_C is the number of clusters, N_L is the number of labels, u_{ij} is the number of times the $i-th$ label is present in the $j-th$ cluster and S_i is the number of occurrence of the $i-th$ label. This function provides a measure of the distribution of a label across clusters. If a label is always present in a cluster, or in the opposite way always absent, the ratio u_{ij}/S_i is near 1, or near 0, and the entropy has a low value. On the other side if a label is generally present in many clusters, the more the value u_{ij}/S_i is near the average, the higher is the entropy.

To reduce the *Intra-Cluster Entropy* a lower bandwidth should be preferred, while to reduce *Intra-Label Entropy* a higher bandwidth should be chosen. To modulate this tradeoff, a measure depending on *Intra-Cluster Entropy* and *Intra-Label Entropy* is defined and is called *Global Clustering Entropy*:

$$E_G = \zeta \cdot \ E_c + (1 - \zeta) \cdot E_l \tag{12}$$

The value of the parameter ζ allows to modulate weight of *Intra-Cluster Entropy* and *Intra-Label Entropy* in the final clustering.

We used the Global Clustering Entropy to empirically evaluate the value of bandwith to use. In practice we computed the bandwith where entropy is minumum for several real datasets, noticing that the same value might be used while considering different image collections.

The clusterization of data through the mode seeking assumes the possibility to estimate distribution density with a single kernel being the data characterized by the same density distribution in all the vector space. In the case here considered, the samples in image collection can be split in multiple representation carrying orthogonal information composed together in a single data vector.

To cluster data represented in multiple domains, mean-shift algorithm is applied in a similar way to what is done in image segmentation by Comaniciu et al.[27]. Assuming that domains adopted to describe items of image collection, allow the Euclidean norm as metric,

a multivariate kernel is defined as product of three radially symmetric kernels:

$$K_{h_f,h_b,h_t}(\mathbf{x}) = \frac{C}{h_f^M h_b^P h_t} k\left(\left\|\frac{\mathbf{x}^f}{h_f}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^b}{h_b}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^t}{h_t}\right\|^2\right) \tag{13}$$

where $\mathbf{x}^f$ is the data represented in the first domain, $\mathbf{x}^b$ is the data referred to the second domain, $\mathbf{x}^t$ is the representation in the third domain, h_f,h_b and h_t are the corresponding kernel bandwidths, C is the normalization constant.

Information is described as a composition of face representation, time of capture and background representation. Faces information has a dimensionality m corresponding to the dimension of the eigenspace adopted. Background information has a dimensionality equal to p that is the sum of the dimensions of the chosen features (Sec. 3.2). Time information is represented with a scalar value. To cluster this composite information, a multivariate kernel is applied with mean shift procedure. Being the data intrinsically composed by three domain independent parts, a composition of three Epanechnikov kernels is applied. Instead of evaluating empirically the performance of multiple values of the bandwidth, the *Global Clustering Entropy* is used as performance measure. Driven by clustering results, the bandwidth value is automatically chosen. The process is run for the three domains, and ideally can be applied to all the set of orthogonal feature representing input samples.

5 Results

The proposed system has been tested on a real photo collection, i.e., 1000 images (VGA and double VGA images), captured in about two months by a mobile device. Tests have been performed on a traditional computer while taking into account the cost of each operation to be sure that the whole processing may be performed on a mobile device. Each image has been manually labeled to store information on the presence of faces, background characteristics and time of shooting. The face detection step brought to the extraction of 734 images of faces and four known people have been chosen so that each face is defined by an ID as reported in Table 1.

Table 1. Faces and background labels.

Faces	
ID	*id1, id2, id3, id4, unknown*
Background	
Type	*indoor, urban, green, snow*
Time	
Type	*birthday, christmas, christmas trip, winter 08/09, ski holiday*

The clustering parameters have been empirically evaluated according to the values of Global Clustering Entropy for face, background and time domains using a subset of the labelled image collection.

Background images are classified according to four categories (*indoor*, *urban*, *green*, *snow*) representing some typical contexts mainly present in the collection. The results for the clustering of background are shown in the Table 2. All clusters with a single element have been discarded and label distribution is shown for the remaining seven clusters.

Table 2. Percentage occurrence of labels in generated clusters

	indoor	urban	green	snow
Cl 1	-	64%	36%	-
Cl 2	12%	-	-	88%
Cl 3	-	58%	27%	15%
Cl 4	43%	36%	-	21%
Cl 5	-	55%	41%	4%
Cl 6	44%	-	56%	-
Cl 7	24%	-	27%	49%

Faces are clustered according the parameters of the *Global Clustering Entropy*. Discarding all the clusters with less than two elements, the number of remaining clusters is equal to 6 and the distribution is shown in Table 3. The id from 1 to 4 are the most recurrent in image repository, all the other faces are associated to a "unknown"label.

Table 3. Percentage occurrence of identities in generated clusters

	Pers 1	Pers 2	Pers 3	Pers 4	unknown
Cl 1	100%	-	-	-	-
Cl 2	3%	-	5%	78%	14%
Cl 3	-	32%	-	-	68%
Cl 4	-	-	74%	-	26%
Cl 5	77%	-	-	-	23%
Cl 6	-	67%	19%	14%	-

Time information is clusterized considering the found parameters and results evaluated according to manually given temporal labels (*birthday*, *christmas*, *christmas trip*, *winter 08/09*, *ski holiday*). The Mean Shift approach is used for clustering the personal album using information from the three considered domains. Clustering results are evaluated by calculating the Global Clusterization Entropy (Equation (12)) with labels given by 3-tuples (*identity*, *context label*, *time label*).

Table 4. Percentage occurrence of time labels (TL) in generated clusters

	birthday	*christmas*	*christmas trip*	*winter 08/09*	*ski holiday*
cl 1	-	23%	77%	-	-
cl 2	54%	-	-	26%	20%
cl 3	-	-	-	-	100%
cl 4	-	63%	11%	26%	-
cl 5	-	-	-	100%	-
cl 6	-	-	100%	-	-
cl 7	11%	7%	53%	-	29%
cl 8	-	-	-	64%	36%

The most frequent 3-tuple for each cluster are shown in Fig. 3 and reported in Table 5.

Table 5. The most frequent 3-tuple for each cluster.

	Cluster 1	*Cluster 2*	*Cluster 3*	*Cluster 4*
who	person 3	person 2	person 4	person 1
where	indoor	indoor	snow	green
when	winter 08/09	winter 08/09	ski holiday	christmas trip

Fig. 3. Image clusterization exploiting multi-domain representation.

6 Conclusions

A fully automatic approach for image searching and browsing on mobile devices has been presented. The goal of the proposed system is to improve user experience in managing (i.e., browsing, saving, printing and so on) photos on the mobile device itself without the need of trasferring and processing them on a host computer. We focused on the three most important aspects of image collections stored on mobile devices: who (faces) is in the picture, where (background) and when (time) the picture was shot. These three aspects (i.e., *picture context*) have been managed in a homogenous way using a mean-shift clustering technique. Moreover, the indexing process is transparent to the user since image features are automatically extracted and clustering parameters are automatically determined according to a proposed entropy based figure of merit. The proposed system has been tested on a real photo collection captured by a mobile device and experimental results are very interesting. The main contribution of the proposed work is to consider the mobile multimedia device as a standalone device that allows an user to instantly manage its own pictures collection. Thus, all image representation and clustering steps have been performed simulating the device constraints, that is taking into account the cost of each operation while optimizing the whole system. Some points (e.g., objective vs subjective clustering evaluation, system performances while using collections of different size, time of execution on different mobile devices) could be better investigated and this will be subject of future work.

References

1. A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2002.
2. Edoardo Ardizzone, Marco La Cascia, and Filippo Vella. Mean shift clustering for personal photo album organization. In *IEEE International Conference on Image Processing (ICIP)*, pages 85–88. IEEE, 2008.
3. H. Kang and B. Shneiderman. Visualization methods for personal photo collections: Browsing

and searching in the photofinder. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2000.
4. L. Zhang, L. Chen, M. Li, and H. Zhang. Automated annotation of human faces in family albums. In *Proceedings of ACM International Conference on Multimedia*, 2003.
5. M. Naaman, R. B. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2005.
6. B. N. Lee, W.-Y. Chen, and E. Y. Chang. A scalable service for photo annotation, sharing and search. In *Proceedings of ACM International Conference on Multimedia*, 2006.
7. J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang. Easyalbum: An interactive photo annotation system based on face clustering and re-ranking. In *Proceedings of ACM Special Interest Group on Computer-Human Interaction*, 2007.
8. A. Girgensohn, J. Adcock, and L. Wilcox. Leveraging face recognition technology to find and organize photos. In *Proceedings of ACM International Conference on Multimedia Information Retrieval (MIR)*, 2004.
9. L. Zhang, Y. Hu, M. Li, W. Ma, and H. Zhang. Efficient propagation for face annotation in family albums. In *Proceedings of ACM International Conference on Multimedia*, 2004.
10. Apple Inc., Iphoto'09. Avalaible at: http://www.apple.com/ilife/iphoto/.
11. M. Abdel-Mottaleb and L. Chen. Content-based photo album management using faces' arrangement. In *Proceedings IEEE International Conference on Multimedia and Expo (ICME)*, 2004.
12. C.-H. Li, C.-Y. Chiu, C.-R. Huang, C.-S. Chen, and Lee-Feng Chien. Image content clustering and summarization for photo collections. In *Proceedings of ICME*, pages 1033–1036, 2006.
13. S. Krishnamachari and M. Abdel-Mottaleb. Hierarchical clustering algorithm for fast image retrieval. In *IS&T SPIE Conference on Storage and Retrieval for Image and Video databases VII.*, 1999.
14. D. Deng. Content based comparison of image collection via distance measuring of self organized maps. In *Proceedings of 10th International Multimedia Modelling Conference*, 2004.
15. J. Goldberg, S. Gordon, and H. Greenspan. Unsupervised image-set clustering using an information theoretic framework. *IEEE Transaction on Image Processing*, 15(2):449–458, 2006.
16. T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Yee-Whye Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–848–II–854 Vol.2, 2004.
17. Deok-Hwan Kim, Chan Young Kim, and Yoon Ho Cho. Automatic generation of the initial query set for cbir on the mobile web. In *PCM (1)*, pages 957–968, 2005.
18. Karthik Kumar, Yamini Nimmagadda, Yu-Ju Hong, and Yung-Hsiang Lu. Energy conservation by adaptive feature loading for mobile content-based image retrieval. In *ISLPED '08: Proceeding of the thirteenth international symposium on Low power electronics and design*, pages 153–158, New York, NY, USA, 2008. ACM.
19. I. Ahmad, S. Abdullah, S. Kiranyaz, and M. Gabbouj. Content-based image retrieval on mobile devices. In *Proceedings of SPIE (Multimedia on Mobile Devices)*, 2005.
20. J. S. Hare and P.H. Lewis. Content-based image retrieval using a mobile device as a novel interface. In *Storage and Retrieval Methods and Applications for Multimedia*, 2005.
21. M. Gabbouj, I. Ahmad, Malik Y. Amin, and S. Kiranyaz. Content-based image retrieval for connected mobile devices. In *Proceedings of Second International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2006.
22. E. Ardizzone, M. La Cascia, and F. Vella. A novel approach to personal photo album representation and management. In *Proceedings of Multimedia Content Access: Algorithms and systems II. IS&T SPIE Symposium on Electronic Imaging*, volume 6820, 2008.
23. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

24. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
25. A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. In *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, 1990.
26. X.Z. Liu, L. Zhang, M.J. Li, H.J. Zhang, and D.X. Wang. Boosting image classification with lda-based feature combination for digital photograph management. *Pattern Recognition*, 38(6):887–901, June 2005.
27. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24:603–619, May 2002.
28. J.C. Bezdek. *Pattern Recognition with Fuzzy Object Function.* Plenum, 1981.
29. K.L. Wu and M.S. Yang. A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, 26:1275–1291, 2005.