Unsupervised Clustering in Personal Photo Collections

Edoardo Ardizzone[1], Marco La Cascia[1], and Filippo Vella[2]

[1] DINFO - Dipartimento di Ingegneria Informatica,
University of Palermo, Palermo, Italy
{ardizzon,lacascia}@unipa.it
[2] ICAR - Istituto di Calcolo e Reti ad Alte Prestazioni
Italian National Research Council, Palermo, Italy
filippo.vella@pa.icar.cnr.it

Abstract. In this paper we propose a probabilistic approach for the automatic organization of collected pictures aiming at more effective representation in personal photo albums. Images are analyzed and described in two representation spaces, namely, faces and background. Faces are automatically detected, rectified and represented 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 energy. Face and background information of each image in the collection is automatically organized by mean-shift clustering technique. Given the particular domain of personal photo libraries, where most of the pictures contain faces of a relatively small number of different individuals, clusters tend to be semantically significant beyond containing visually similar data. We report experimental results based on a dataset of about 1000 images where automatic detection and rectification of faces lead to approximately 300 faces. Significance of clustering has been evaluated and results are very encouraging.

1 Introduction

With the widespread diffusion of digital cameras the cost of taking hundreds of digital pictures and storing them on personal computer is quickly approaching zero. People are then encouraged to take more and more pictures but the consequent risk is that they end up with tens of thousand of pictures stored on their PCs that, without a proper organization, become useless. Currently, the main way to search digital photo libraries is by mean of time of shooting and/or keywords given by the user. This modality of access to the library is definitely unsatisfactory, moreover it requires the user manually associates keywords to each picture. This process has been observed to be inadequate since users usually add few keywords for large set of images and, on the other side, keywords tend to be ambiguous. Time of shooting is a much more reliable cue and it is available for free as digital cameras attach a timestamp, in the EXIF data, to each pictures. However its power in term of searching capabilities is quite limited. An ideal

M. Detyniecki, U. Leiner, and A. Nürnberger (Eds.): AMR 2008, LNCS 5811, pp. 140–154, 2010.

system for image browsing should allow an automatic organization of pictures based on the semantic of photos. Our point is that personal photo libraries show peculiar characteristics compared to general image collection, namely the presence of people in most of the images and a relatively small number of different individuals across the whole library that allow to achieve reliable results with automatic approaches[3]. In particular, in personal photo collection 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. When the picture was shot is an information that comes for free as all the digital cameras attach a timestamp to the pictures they take. Even though this temporal information is a very useful piece of information [14] we focused in this work on the other aspects while this issue is object of further investigation. *Who* and *where* are the fundamental aspects of photo information and input images can be intrinsecally split in two domains of interest. Faces are extracted from the images and referred to person identity; the remaning part of the image is considered as image context. Known techniques[4] are used to detected and rectify faces from the data set allowing to project all the samples in a common low dimensional *face space*. Images' background are characterized by mean of low-level features based on color and texture characterizing different contexts (*where*). Also for this aspect, as the typical user is interested to a limited number of different contexts, the link between low-level features and context semantic content can be reasonably established.

To automatically organize image data based on faces and background descriptors we use a mean-shift based approach[7]. Organization of data does not need any human intervention as image features are automatically extracted and parameters of the clustering method are automatically determined according to a proposed entropy based figure of merit. The paper is organized as follows: Section 2 describes the known techniques for organization, storage and content based retrieval for personal photo collection; in Section 3 the proposed description of data composing personal album is given. The details of the image processing and analysis are described in 3.1 and in 3.2. The clustering process is presented in 4. Finally in Sections 5 and 6 are given the results of the experiments and the conclusions.

2 Related Works

One of the first personal photo collection browser has been reported by Kang and Shneiderman[15]. The goal of this system was to enable non-technical users of personal photo collection to browse and search efficiently for particular images. The authors proposed a very powerful user interface but implemented very limited CBIR capabilities. Moreover the search was heavily based on manual annotation of the data. As in personal photos the objects of interest are often people Zhang et al. [25] addressed the problem of automated annotation of human faces in family album. CBIR techniques and face recognition are integrated in a probabilistic framework. Based on initial training data models of each person are built and faces in images are often recognized correctly even in presence of some

occlusions. User interaction during the annotation process is also possible to reinforce the classifier. Experimental results on a family album of a few thousands photos showed the effectiveness of the approach. In a subsequent work [26] some of the authors 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. Abdel-Mottaleb and Chen [2] also studied the use of faces arrangement in photo album browsing and retrieval. In particular they defined a similarity measure based on face arrangement that can be computed automatically and is used to define clusters of photos and finally to browse the collection. A photo management application leveraging face recognition technology has also been proposed by Girgensohn et al.[12]. The authors implemented a user interface that greatly helps the users in face labelling. Other semi-automatic annotation tecniques for personal photo libraries have also been proposed recently[19,17,9]. Other researcher address the problem of personal photo album management in an image clustering framework. For example hierarchical clustering enable the users to navigate up and down the levels to find images. Navigating the collection is also useful in query-by-example systems to find the initial image. In any case the clusters prototype are a compact representation of classes of similar images and then can be used in browsing or searching the library. The efficacy of the clustering approach, as well as any CBIR system, is obviously affected by the goodness of the image features used to describe the images and the similarity metrics defined over these features. As similarity metrics may not reflect semantic similarity between images, sometimes clusters are not semantically homogeneous. Many techniques have been proposed to refine the automatic clustering approach with human intervention to make cluster semantically homogeneous. Several techniques have been proposed for the clustering of images. For example in[16] the authors use color histogram and histogram intersection distance measure to perform hierarchical clustering. Similarly, Chen et al. [6] used global color, texture and edge histogram and the L_1 distance to define an hierarchical browsing environment. In [11] a self-organizing map is used to let the structure of the data emerge and then to browse the collection. In [20] the authors use a clustering strategy based on Markov Model Mediators to improve efficiency and efficacy of retrieval in distributed image databases. Recently, Goldberger et al. [13] 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. Experimental results using different image representation are also reported. In other cases the presence of faces in an attempt to bridge the gap between visual and semantic content is exploited. For example in [4] face detection is performed on captioned images and clustering is used to associate automatically extracted names to the faces. In [18] the authors detect faces and describe clothes and nearby regions with color histogram. A similarity matrix of a photo collection is then generated according to temporal and content features and hierarchical clustering is performed based on this matrix. Song and Leung [21] aims at clustering the dataset such that each cluster contains images

of a particular individual. They use face and clothing descriptors to contruct an affinity matrix over the identities of individuals and perform clustering using a normalized-cut approach. In [10] a semi automatic photo annotation system based on enhanced spectral clustering is proposed. They use time, global color correlogram for location/event clustering and local facial features and color correlogram from human body area for face clustering. As automatic techniques cannot guarantee that all the faces in a cluster are related to the same individuals or that an individual is not spread across several clusters, the final validation of the clustering is done by hand.

3 Personal Photo Album Indexing

The focal point of the clustering is the representation of each image in a form suited for clustering. An image can be represented in several spaces allowing to capture different aspects of input data. In the proposed system, each image in the collection is represented with features related to the presence of faces in the image and features characterizing the background[3]. A data oriented clustering allows to generate aggregation structures driven by the regularities in the represented data. In the following sections the processing of visual information in the two chosen representation spaces is described. Faces are preprocessed to reduce the variation of the appearance and are mapped in an auto emerging space employing eigenfaces. The information from background is managed representing parts not associated in a vector space representing low-level features.

3.1 Face Representation

Finding faces in general images is a very challenging task due to variations in pose and illumination. Berg et al.[4] analyzed hundred of thousands of images taken from the Internet to detect faces *in the wild*. In a similar way in our approach each image to be archived in the system is searched for faces. Detected faces are then validated and rectified to a canonical pose and size. The face detector we adopted[23] is usually successful in detecting faces in a quite large range of pose, expression and illumination conditions. In some different application, such as detection of faces in video sequences, a SVM based face detector is used instead [1],[22] for the improved capability in face detection when images are blurred. For the detected faces, we try to detect five features per face *left eye extern corner*, *right eye extern corner*, *tip of the nose*, *mouth left corner*, and *mouth right corner* through SVM detectors and, if detection is successful, we estimate an affine transformation to rescale and align the face to canonical position. Fiducial points detectors have been trained with hundreds of positive and negative examples and Radial Basis Function have been employed for training. Test on detection are very good ($\tilde{9}0\%$ of true positive) A final crop to 100×100 pixels brings each face to a common reference system.

Faces where detection process produces features with low level of confidence were rejected. For successfully detected and rectified faces,a reduced dimension

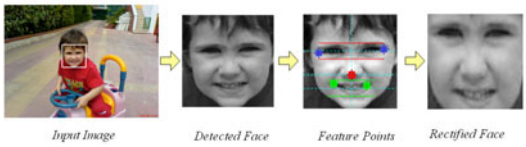

Fig. 1. Example of detected face and corresponding rectified image

face descriptor is computed. The descriptor is a vector $\mathbf{w}$ containing the projection of the rectified and cropped face in a subspace of the global face space. In practice the average face Ψ is subtracted from the 100×100 cropped and rectified face Γ_i and the obtained image Φ is then projected on the eigenspace to obtain $w_i = \mathbf{e}_i^T \Phi$. 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 effortlessly recompute the projection of each detected, rectified and cropped face in the new face space.

3.2 Background Representation

The largest part of semantic information in personal photo is conveyed in areas where faces appear, the remaining part of the image information is attributed the context of the scene. As described above, each picture is processed with the face detector selecting areas containing faces. These areas are approximated with bounding boxes and are dealt as seen in previous section. The remaining part, not representing a face, is then processed as background. Note that the partition of images in multiple areas of interest can be extended using additional detectors (e.g. a detector for entire body ccould be easily integrated in the system). Background information can be represented with a composition of color and texture features. Features are globally evaluated and a single vector for each image is produced. Color information is captured through histograms in the RGB color space. The 60-dimensional global descriptor is computed as the concatenation of the 20-bin histograms of the R, G and B channels. Texture is evaluated through Gabor filters considering 6 different filters, varying 3 orientation and 2 scales. For each filter the energy value is evaluated and represented as a 15 bins histogram. The total feature will be composed by a total of 15x6 components. Since mean values of features in the data space distribution are less indicative, image features are filtered with a sigmoid to stretch the values towards a low or high values according to a parameter α.

4 Image Clustering

4.1 The Mean Shift Algorithm

Mean shift is a technique for kernel density estimation that applies gradient climbing to probability distribution[8]. 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

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

where h is the bandwidth and the kernel $K(.)$ is the Epanechnikov kernel defined as:

$$K(x) = \begin{cases} \frac{1}{2V_d}(d+2)(1 - \|\mathbf{x}\|^2) \text{ if } \|\mathbf{x}\|^2 < 1 \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{2}$$

with V_d representing a volume of a unit d-dimensional sphere. Using a differentiable kernel, the estimate of the gradient density can be written as the gradient of the kernel density estimate(1):

$$\hat{\nabla} f(\mathbf{x}) \equiv \nabla \hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla K(\frac{\mathbf{x} - \mathbf{x}_i}{h}) \tag{3}$$

For the Epanechnikov kernel, shown in (2), the density gradient estimate is:

$$\hat{\nabla} f(\mathbf{x}) = \frac{n_c}{nV_d} \frac{d+2}{h^d} \left(\frac{1}{n_c} \sum_{\mathbf{x} \in S(x)} (\mathbf{x}_c - \mathbf{x}) \right) \tag{4}$$

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

$$M_h(\mathbf{x}) \equiv \frac{1}{n_c} \sum_{\mathbf{x} \in S(\mathbf{x})} (\mathbf{x}_c - \mathbf{x}) \tag{5}$$

is called Mean Shift Vector that can be expressed,using (4) as :

$$M_h(\mathbf{x}) = \frac{h^d}{d+2} \frac{\hat{\nabla} f(\mathbf{x})}{\hat{f}(\mathbf{x})} \tag{6}$$

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.

4.2 Mean Shift Clustering for Personal Album

Given a generic point in the feature space(faces, backgrounds), the Mean Shift Vector in equation (5) 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:

– 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:

$$\left\| \mathbf{y}_i - \mathbf{y}_j \right\| < \frac{h}{2} \tag{7}$$

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.

4.3 Entropy Based Clustering Measure

A number of evaluation indexes have been proposed to evaluate clustering methods, from the older Partition Coefficient and Partition Entropy[5] to the newest as partition based on exponential separation [24]. All of them tend to capture the goodness of the separation proposed by clustering. Typically these methods are oriented to fuzzy clustering more than to hard (crisp) clustering and they use a estimation of the density to evaluate the clustering performance (e.g. Parzen Windows). Since we already adopt a density estimation in the mean-shift procedure, to avoid a biased clustering measure, we choose to evaluate clustering from a set of hand annotated images that allow to evaluate the quality of label spreading in the set of samples. We define two indexes able to capture the clustering capability. The *Intra-Cluster Entropy* is defined as:

$$E_c = -\frac{1}{N_C * log(N_L)} \sum_{i=1}^{N_L} \sum_{j=1}^{N_C} \frac{u_{ij}}{T_j} log \frac{u_{ij}}{T_j} \tag{8}$$

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 labelled samples in the $j-th$ cluster. This function gives a measure of the entropy inside clusters. If many labels are present in a cluster the value u_{ij}/T_j is near the average and the *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.

In figure 2 is shown the values of *Intra-Cluster Entropy* for the clustering with mean shift of a set of eigen-faces when the value of bandwidth is chosen among 4000 and 8000. For lower value of the bandwidth, the kernel covers a reduced volume and the number of modes is over-estimated. In this case the number of sample inside each cluster is reduced and the disorder is limited. With a higher bandwidth, the number of clusters decreases until all the samples are merged in a single cluster(see figure 4). In this case the Intra-Cluster Entropy reaches a maximum and will remain constant for higher values of the bandwidth. The second index, the *Intra-Label Entropy* is defined as:

$$E_l = -\frac{1}{N_L * log(N_C)} \sum_{i=1}^{N_L} \sum_{j=1}^{N_C} \frac{u_{ij}}{S_i} log \frac{u_{ij}}{S_i} \tag{9}$$

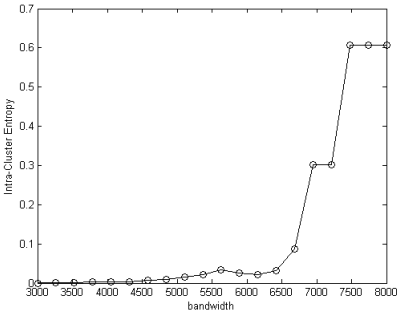

Fig. 2. Plot of *Intra-Cluster Entropy* for values of bandwidth from 4000 to 8000 for the clustering of eigenfaces

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 gives 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.

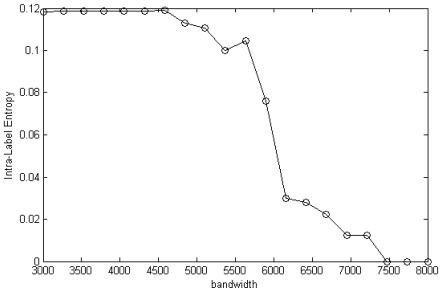

Fig. 3. Plot of *Intra-Label Entropy* for values of bandwidth from 4000 to 8000 for the Mean Shift clustering of eigenfaces

Similar consideration to the previous ones for *Intra-Label Entropy* can be drawn for this index.

Ideally, each label describing a set of samples, should be referred to a single cluster containing all uniform samples. In real cases this distribution is very rare due to the intrinsic variability of data and errors affecting sampling. Usually a tradeoff in the number of cluster must be fixed.

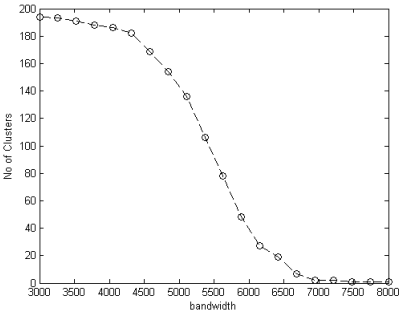

Fig. 4. Number of clusters determined with Mean Shift procedure on eigenfaces with values of bandwidth from 4000 to 8000

To modulate this tradeoff, a measure depending on *Intra-Cluster Entropy* and *Intra-Cluster Entropy* is defined and is called *Global Clustering Entropy*

$$E_G = \zeta \cdot \ E_c + (1 - \zeta) \cdot E_l \tag{10}$$

The value of the parameter ζ allows to modulate weight of *Intra-Cluster Entropy* and *Intra-Label Entropy* in the final clustering. The measure of *Global Clustering Entropy* referred to figure 2 and figure 3, with needed scaling, is shown in 5 considering ζ equal to 0.5.

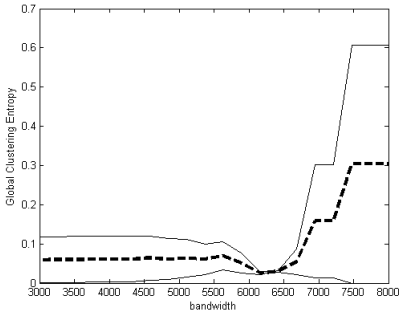

Fig. 5. Plot of the *Global Clustering Entropy*

Mean Shift Clustering for Composite Data. 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 space. In the case here considered, the sample in personal photo album can be split in multiple representation carrying orthogonal information composed together in a single data vector. Image data are represented with composite vectors merging multiple information that are processed with Mean Shift procedure. This technique has never been used to cluster data through

multiple information while in [8] a similar technique is used to segment images in homogeneous areas. Assuming that both domains (i.e. feature spaces) used to describe items of personal photo album allow the Euclidean norm as metric, a multivariate kernel is defined as product of two radially symmetric kernels:

$$K_{h_f,h_b}(\mathbf{x}) = \frac{C}{h_f^M h_b^P} k\left(\left\|\frac{\mathbf{x}^f}{h_f}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^b}{h_b}\right\|^2\right) \tag{11}$$

where $\mathbf{x}^f$ is the data in the first domain, $\mathbf{x}^b$ is the data referred to the second domain, h_f and h_b are the corresponding kernel bandwidths, C is the normalization constant. For personal photo album information is described, as shown in Section 3, as a composition of face representation and background representation and a multivariate kernel is applied to cluster information with Mean Shift procedure. The first part is referred to faces information and has a dimensionality f corresponding to the number of eigenfaces used. The second is background information with a dimensionality equal to b of the chosen features as shown in 3.2. The adopted kernels are two Epanechnikov kernel (2) each with a chosen value of bandwidth. Instead of evaluating empirically the performance of multiple values of the bandwidth, the *Global Clustering Entropy* introduced in the section 4.3 is used as performance measure. Driven by clustering results, the bandwidth value is automatically chosen. The process is run for both the domains, and ideally can be applied to all the set of orthogonal feature representing input samples, then the merging of clusters among multiple domains is performed (as in [8]):

- Run the Mean Shift procedure for the chosen domains fixing a value of bandwidth for each of them. The information about the convergence points are stored.
- Delineate in the joint domain the clusters by grouping the convergence points that are closer than the value of bandwidth in the corresponding domain. That is the basins of attraction of the corresponding convergence points are concatenated.
- Assign each point in the space to a cluster in the joint domain.

5 Experimental Results

To evaluate the performances of the proposed system we ran a set of experiments on a real photo collection. The digital album used is a subset of a real personal collection of 1008 images taken in the last three years. The presented process for face detection and rectification brought to the extraction of 331 images of rectified faces. The experiments have been aimed to the evaluation of the retrieval capability of the proposed system in terms of faces and background labeling but an entropy based analysis of the clustering process has also been performed to better understand the process itself.

We evaluated entropy as function of the bandwidth used in the clustering process and of parameters used in processing visual data, namely the α coefficient

of the sigmoid when processing the background data and the dimension of the eigenspace for face data. Since the variation of clusters composition according these parameters is smooth, the evaluation for a reduced set of combination of parameters produce significative results.

The background data analysis showed that the clustering providing the best value of Global Clustering Entropy was with a value of α equal to 0.95 and a bandwidth of 3.93. From the face data analysis we observed the best results for a dimension of eigenspace equal to 131 and a bandwidth of 5756. We also evaluated as the dimension of the eigenspace affect the number of clusters (see Fig. 8) and results where coherent with hypothesis made with our model.

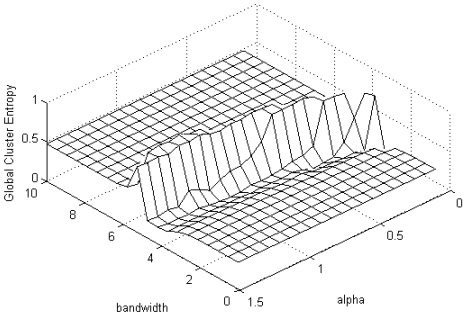

Fig. 6. Plot of the *Global Entropy* for the clustering of background data as function of the bandwidth and the α coefficient of the sigmoid

To evaluate our approach with respect to the semantic meaning of clusters we divided all the images of the test collection in six categories (*beach*, *indoor*, *nature*, *public garden*, *snow*, *urban*) representing six typical contexts mainly present

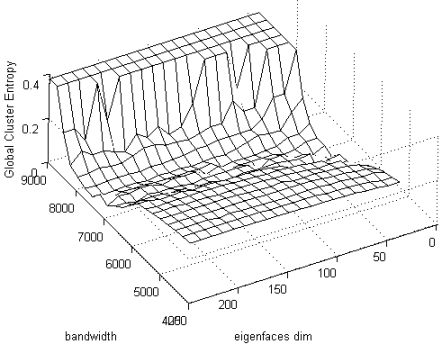

Fig. 7. Plot of the *Global Entropy* for the clustering of faces data as function of the bandwidth and the dimension of the eigenspace

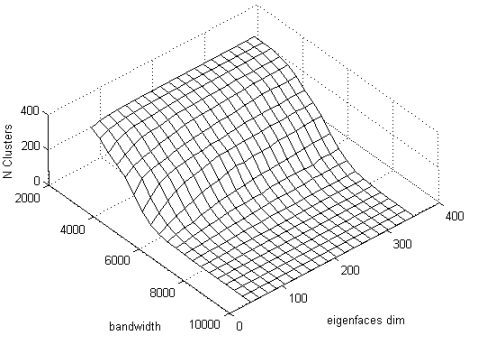

Fig. 8. Plot of the number of clusters as function of the bandwidth and dimension of the eigenspace

in the collection and assigned an identifier to the four people present in most of the photos.For the first experiments, data are clustered using a single domain information. For each domain the optimal point according to the Global Clustering Entropy (eq. 10) is chosen. The results for the clustering of background are shown in the table 1.

Table 1. Percentage occurrence of labels in generated clusters

	beach	indoor	nature	public garden	snow	urban
Cl 1	11%	32%	2%	40%		15%
Cl 2		89%		11%		
Cl 3		96%		4%		
Cl 4		100%				
Cl 5		13%		63%		25%
Cl 6	6%	42%		52%		
Cl 7		100%				
Cl 8		100%				
Cl 9				67%		33%

For faces data, the optimal point according the entropy measure is found for a bandwidth of 5756 and a dimension of the eigenspace equal to 131. Discarding all the cluster with a single element the number of remaining clusters is equal to 6 and the distribution is shown in table 2. The id from 1 to 4 are the most recurrent in image repository.

The clusters for the personal album using information from both domains are created using the procedure described in 4.3. An evaluation of this clusterization is achieved calculating the Global Clusterization Entropy (eq. 10) using labels given by couples (*identity*, *context label*). In the table 3 values of Entropy are shown when are varied values of clusterization bandwidths.

Table 2. Percentage occurrence of identities in generated clusters

	Id 1	Id 2	Id 3	Id 4	Other
Cl 1		77%	9%	5%	9%
Cl 2	3%	22%	53%	6%	17%
Cl 3	22%	33%	33%		11%
Cl 4	11%	11%	44%		33%
Cl 5			100%		
Cl 6			100%		

Table 3. Value of global entropy for clusterization

	3.0	3.5	4.0	4.5	5.0
4756	2.67	2.70	2.78	2.87	3.10
5256	2.70	2.72	2.82	2.92	3.11
5756	3.03	3.10	3.20	3.22	3.19
6256	4.31	4.87	5.89	5.71	5.45
6756	6.41	8.31	10.09	9.78	8.07

Along the columns of the table the value of the bandwidth for clustering of faces varies from 4756 to 6756 with even intervals. Along the columns the bandwidth varies from 3 to 5 with even intervals.

6 Conclusions

A novel approach to cluster composite data driven by a clusterization measure has been presented. The approach has been demonstrated on the very interesting problem of automatic organization of photos in personal album. To demonstrate the approach we developed techniques to represent images as composite data corresponding to foreground and background. In our experiments data in collection of photos are represented in two spaces retaining information about people in the picture and low level features related to the context of the picture itself. Due to the domain constraint we were able to use computer vision technique to extract information automatically and with reasonable correctness. Results of experiments on a real set of a thousand pictures based both on entropy measures of the cluster and on comparison to manually provided ground truth are very promising.

Future work include exploiting more image information automatically computed from image data. Other information as clothing characterization of people captured, more detailed information about backgrounds (e.g. multiple feature and color descriptors), time of capture and other hints from EXIF data could be easily added in the proposed framework and we expect even better results.

References

1. Caminati, L., Don, A., et al.: Detection of visual dialog scenes in video content based on structural and semantic features. In: Proc. of International Workshop on Content-based Multimedia Indexing, CBMI (2005)
2. Abdel-Mottaleb, M., Chen, L.: Content-based photo album management using faces' arrangement. In: IEEE International Conference on Multimedia and Expo., ICME (2004)
3. Ardizzone, E., La Cascia, M., Vella, F.: A novel approach to personal photo album representation and management. In: SPIE, vol. 6820 (2008)
4. Berg, T.L., Berg, A.C., Edwards, J., Maire, M., White, R., Teh, Y.W., Learned-Miller, E., Forsyth, D.A.: Names and faces in the news. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, CVPR (1994)
5. Bezdek, J.C.: Pattern Recognition with Fuzzy Object Function. Plenum (1981)
6. Chen, J.Y., Bouman, C.A., Dalton, J.C.: Hierarchical browsing and search of large image databases. IEEE Transaction on Image Processing 9(3), 442–455 (2000)
7. Cheng, Y.: Mean shift, mode seeking and clustering. IEEE Transaction on Pattern Analysis and Machine Intelligence, 790–799 (August 1995)
8. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, 603–619 (May 2002)
9. Cui, J., Wen, F., Xiao, R., Tian, Y., Tang, X.: Easyalbum: An interactive photo annotation system based on face clustering and re-ranking. In: Proc. of ACM CHI (2007)
10. Cuiy, J., Wenz, F., Xiaoz, R., Tianx, Y., Tang, X.: Easyalbum: An interactive photo annotation system based on face clustering and re-ranking. In: Proc. of CHI (2007)
11. Deng, D.: Content based comparison of image collection via distance measuring of self organized maps. In: Proceedings of 10th INternational Multimedia Modelling Conference (2004)
12. Girgensohn, A., Adcock, J., Wilcox, L.: Leveraging face recognition technology to find and organize photos. In: Proc. of ACM MIR (2004)
13. Goldberg, J., Gordon, S., Greenspan, H.: Unsupervised image-set clustering using an information theoretic framework. IEEE Transaction on Image Processing (2), 449–458 (2006)
14. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proc. of ACM JCDL (2002)
15. Kang, H., Shneiderman, B.: Visualization methods for personal photo collections: Browsing and searching in the photofinder. In: Proc. of IEEE International Conference on Multimedia and Expo., ICME (2000)
16. Krishnamachari, S., Abdel-Mottaleb, M.: Hierarchical clustering algorithm for fast image retrieval
17. Lee, B.N., Chen, W.-Y., Chang, E.Y.: A scalable service for photo annotation, sharing and search. In: Proc. of ACM International Conference on Multimedia (2006)
18. Li, C.-H., Chiu, C.-Y., Huang, C.-R., Chen, C.-S., Chien, L.-F.: Image content clustering and summarization for photo collections. In: Proceedings of ICME, pp. 1033–1036 (2006)
19. Naaman, M., Yeh, R.B., Garcia-Molina, H., Paepcke, A.: Leveraging context to resolve identity in photo albums. In: Proc. of ACM JCDL (2005)

20. Shyu, M.-L., Chen, S.-H., Chen, M., Zhang, C.: A unified framework for image database clustering and content-based retrieval. In: ACM International Workshop On Multimedia Databases archive Proceedings of the 2nd ACM international workshop on Multimedia databases, pp. 19–27 (2004)
21. Song, Y., Leung, T.: Context-aided human recognition clustering. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 382–395. Springer, Heidelberg (2006)
22. Spyrou, E., Kapsalas, P., Tolias, G., Mylonas, P., Avrithis, Y., et al.: The cost292 experimental framework for trecvid 2007. In: Proc. of 5th TRECVID Workshop (2007)
23. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, CVPR (2001)
24. Wu, K.L., Yang, M.S.: A cluster validity index for fuzzy clustering. Pattern Recognition Letters, 1275–1291 (2005)
25. Zhang, L., Chen, L., Li, M., Zhang, H.: Automated annotation of human faces in family albums. In: Proc. of ACM International Conference on Multimedia (2003)
26. Zhang, L., Hu, Y., Li, M., Ma, W., Zhang, H.: Efficient propagation for face annotation in family albums. In: Proc. of ACM International Conference on Multimedia (2004)