FACIAL IMAGE CLUSTERING IN SINGLE CHANNEL AND STEREO VIDEO CONTENT

G. Orfanidis N. Nikolaidis I. Pitas
Aristotle University of Thessaloniki, Department of Informatics
Box 451, 54124 Thessaloniki, Greece

email: nikolaid@aiia.csd.auth.gr, pitas@aiia.csd.auth.gr

ABSTRACT

In this paper spectral clustering techniques are implemented and evaluated on image clustering for single channel and 3D video analysis. The main idea is to use mutual information to create a similarity matrix for image pairs and then apply spectral clustering. Then, spectral clustering techniques can be used for image clustering. Such clustering techniques are then extended to stereo video. The application at hand includes facial image clustering on single view and stereo videos facial images.

Index Terms— Spectral clustering, Mutual Information, Stereo video

1. INTRODUCTION

Clustering is one of the fundamental topics in computer science. It has been studied at a great degree but nevertheless even at present day it is not considered solved. The importance of clustering is based on the desire to analyze existing information and to group it into categories making easier the processing and manipulation of the information.

Clustering is the division of an existing set of objects C (in our case this is typically images) into a number of subsets (clusters) $C_i = \{C_i \mid \subseteq P\}$. Those subsets have to fulfill the following conditions: $\bigcup_{C_i \in C} C_i = P$ and $\forall C_i, C_j,\ i \neq j \in C$: $C_i \bigcap C_j = \emptyset$.
So clustering means the assignment of each sample of a data set to a non overlapping set of groups, although there are exceptions to this like in the fuzzy c-means algorithm.

Image clustering is connected to other topics like classification and label propagation. Clustering is an unsupervised technique while classification belongs to the supervised techniques and label propagation to the semi-supervised ones. Classification is a general term used to describe the process of assigning an unknown sample to some known categories. This is preceded by a training step. On the other hand label propagation is a semi-supervised technique used to spread the labels of an already labeled data set to an unlabeled one thus creating a new labeled data set.

There are several approaches to implement clustering. One of the first approaches was to consider each sample as a vector and to try to organize these samples-vectors at geographically separated groups using appropriate distance or similarity measures as k-means, k-medians and fuzzy c-means are some based on this methodology. Another more recent approach tries to represent the data set through a graph. This has major advantages since the graph theory tools can be utilized. Spectral clustering [3,7] uses this approach, i.e. it considers a graph representing the relation between the samples along with matrix calculus and eigenanalysis. This method will be explained with more details in next section.

The main application of this paper is clustering and more precisely facial image clustering, where the goal is to separate face images into groups, for which within-cluster similarity is high whereas between-cluster similarity is smaller. This problem has been the subject of several previous works [13,12,11,2]. Ideally each cluster should contain only instances of a single person and if possible all instances of that person. This means that if a person has many appearances all these appearances should be gathered in one cluster or at least clusters should contain only images from one person.

Although considerable has been performed on clustering of facial images from single channel video, facial image clustering in stereo videos is a novel field of research. Research in this field can follow two different and sometimes supplementary approaches: the first approach is to apply already well studied algorithms designed for single channel video like [13,12,11] modified specifically for use with stereo video. The second approach is to specifically develop algorithms from scratch for stereo video. In this paper we consider the first approach. Stereo image clustering has certain differences from its single channel counterpart. Indeed for each sample, two images from the left and right channel exist.

978-1-4673-4989-5/13/$31.00 ©2013 IEEE

In this paper we follow an approach for solving the clustering problem stated before which utilizes spectral graph theory tools. In order to use these tools we represent our facial images as graph nodes and also make use of mutual information theory as a similarity measure between facial images. This similarity measure is utilized as graph edge so as to create our fully connected graph. After this step a variation of the well-studied Normalized Cuts (N-Cuts) algorithm is presented and used for the actual clustering step.

The rest of this work is organized as follows: Section 2 contains a short presentation of Mutual Information and its normalized version used in the proposed approach. In section 3 spectral graph clustering is presented. Section 4 presents the application of spectral clustering in facial images obtained from single view and stereo videos.

2. MUTUAL INFORMATION

Mutual Information (MI) is used as the similarity measure among our samples. It is defined as the common information of two distributions. Let us first provide some relevant definitions. Joint entropy of two random variables X and Y is defined as:

$$H(X,Y) = -\sum p(x,y)\log(p(x,y)$$

where $p(x,y)$ is the joint probability density function of variables X and Y. Entropy for each X and Y is defined as:

$$H(X) = -\sum p(x)\log(p(x))$$

where $p(x)$ is the probability density function of the variable X. Mutual information is defined as:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) =$$

$$-\sum p(x,y)\log(\frac{p(x,y)}{p(x)p(y)}) \quad (1)$$

We choose use a normalized variant of MI so that the values of MI span in the range [0,1]. In [9] Studholme et al. have shown that this version presents some advantages as robustness to the size of the overlapping image regions in image registration. We get the normalized mutual information [4]:

$$MI(X,Y) = \frac{H(X)+H(Y)}{2H(X,Y)}$$

In this work, the HSV color space was used to evaluate similarities among image samples as it has been shown to be more robust in illumination changes as compared to RGB color space. More specifically, we will follow the approach of [13] where only Hue (H) and Saturation (S) channels of HSV are used. According to this approach, 4D normalized MI is given by:

$$MI(X,Y) = \frac{H(H_1)+H(S_1)+H(H_2)+H(S_2)}{2H(H_1,S_1,H_2,S_2)}$$

where H_i is the Hue and S_i is the Saturation of each image respectively. Of the two Hue is considered more informative. For the 4D joint histogram a similar approach to [12] was used.

3. SPECTRAL GRAPH CLUSTERING

We represent our image samples using a graph. More specifically each image sample is considered as a node of the graph connected with edges w to all other nodes. Between two nodes i, j the edges is represented as w_{ij}. By this way a similarity matrix $\boldsymbol{W}$, having elements $w_{ij}, i, j = 1,..,N$ can be created. As edge weight we use the mutual information, defined before, between the images i and j.

Having defined our graph we can utilize the spectral clustering [3,7] approach in order to cluster the data set. First we define the Laplacian matrix:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$$

where $\boldsymbol{D}$ is defined as the diagonal matrix $D_{ii} = \sum_j w_{ij}$. Matrix $\boldsymbol{D}$ is proved to be positive semi definite so all its eigenvalues are non negative. In order to find the optimum separation point we use Normalized Cuts method described below.

5.1. Normalized Cuts

Normalized Cuts (Ncuts) is a method used to retrieve the optimal cut of a connected graph. For retrieving the optimal solution two simultaneous criteria are being searched: maximize the within-cluster similarity and minimize the between-cluster similarity. The method was presented in [6] and it has been proven that the solution we are looking for is the solution of the generalized eigenvector problem defined as:

$$\text{argmin}_{\boldsymbol{y}^T\boldsymbol{D1},\, y(i)\in\{1,-b\}} \frac{\boldsymbol{y}^T(\boldsymbol{D}-\boldsymbol{L})\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}} \quad (2)$$

The second constraint, $y(i) \in \{1, -b\}$ i.e. the discrete nature of $\boldsymbol{y}$, is typically relaxed so as to consider $\boldsymbol{y}$ to be continuous and to get an approximate solution to our problem. This gives us the advantage of using classical eigenvector analysis to solve our problem. We solve the generalized eigenvector $\boldsymbol{D}\boldsymbol{y} = \lambda(\boldsymbol{D}-\boldsymbol{L})\boldsymbol{y}$ problem and keep the eigenvector corresponding to the smallest eigenvalue since the eigenvalue is zero and its eigenvector is a vector consisting of ones: $\boldsymbol{1}$. But from the first constraint, $\boldsymbol{y}^T\boldsymbol{D1}$, we can see that this solution is rejected. So the solution to our problem is the eigenvector corresponding to the second smaller eigenvalue. but since the smallest eigenvalue since the smallest one is rejected by the constraint. One more observation is that the mean value of every other

eigenvector is always zero as a result of being orthogonal to the first eigenvector, $\mathbf{1}$:

$$\forall \boldsymbol{y}_i : \boldsymbol{y}_i^T \mathbf{1} = \sum_{N}^{j=1} y_i = 0, i = 2, .., N$$

where $\boldsymbol{y}_i$ are the eigenvectors corresponding to the solution of the previous eigenvector problem. Thus, using the mean value of each vector as a threshold is equivalent to the samples being separated by their signs.

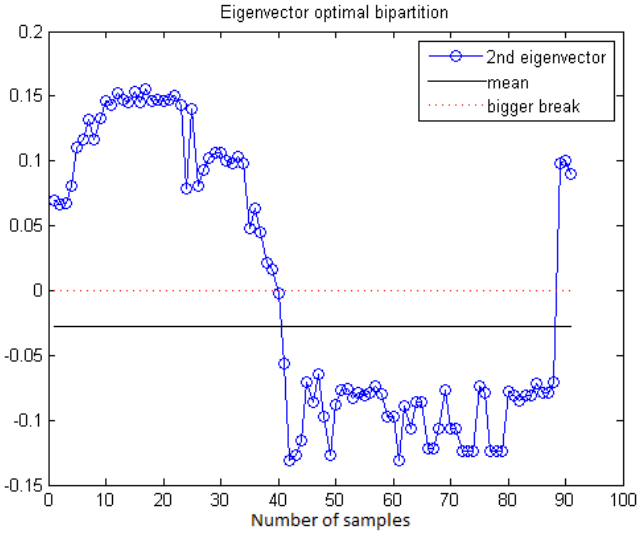

Fig. 1. Optimal eigenvector bipartition point

In this paper we used the general approach of Ncuts but we also introduce some variation that proved to perform better. Thus, in order to bipartition the samples we did not consider the sign of the second smallest eigenvector but we search for the biggest gap in the eigenvector's range. The solution achieved this way follows the human notion of partition. The optimal separation can be seen in Figure 1. In this example, we can seen that the first separation criterion assigns sample #40 in the first cluster, while the second to the second cluster. Our approach has also proven to achieve better performance compared to typical Ncuts approach in most cases. This could be considered a byproduct of the relaxed version of our problem: variable $\boldsymbol{y}$ takes now continuous values and so searching for the greater gap makes sense. The theoretical base on this choice can be found in [10].

4. EXPERIMENTAL RESULTS

In this section we present the results of experiments conducted in real video data. The first set of experiments dealt with face clustering in single view videos and was conducted in the "Hollywood Human Actions dataset" [5]. This is a database consisting of 23 movies, each movie containing a part of a Hollywood movie. In order to evaluate the performance of the algorithm F-measure [8] was used for all 23 movies. Results are shown in Table 1. Overall performance reached a mean F-measure of 91.43%. Facial images were retrieved by using a face detector and tracker in each video. We used the approach presented in [1] for choosing the threshold for defining the clusters. It must be noted here that F-measure punishes more clusters containing samples from different ground truth clusters than splitting a ground truth cluster into two clusters which nevertheless contains only samples of a ground truth cluster.

Hollywood database clustering performance			
Movie title	Performance	Movie title	Performance
American Beauty	100%	The Graduate	96.53%
As Good As It Gets	75.18%	I Am Sam	94.24%
Being John Malkovich	93.54%	Indiana Jones And The Last Crusade	98.38%
Big Fish	82.88%	Kids	93.46%
Big Lebowski	100%	LOR-Fellowship Of The Ring	94.66%
Bringing Out The Dead	100%	Lost Highway	84.0%
Butterfly Effect	88.27%	Mission To Mars	85.1%
Crying Game	81.33%	The Pianist	93.65%
Dead Poets Society	97.92%	Pulp Fiction	100%
Erin Brockovitch	90.47%	The Godfather	100%
Forest Gump	95.64%	Two Week Notice	71.07%
Gandhi	86.59%		

Table 1. Performance on Hollywood Human Actions dataset

The second set of experiments was conducted on a stereo video depicting a concert, see Figure 2. We consider using all images from both channels in order to achieve a better performance. Ncuts has been proven to fulfill the consistency criterion [10], which is, if more and more samples are being added to the dataset a convergence can be expected, so our intention has theoretical foundations.

As mentioned before HSV color space was used as similarity measure. Since Hue is proven to be more informative [13] as compared to Saturation we use a weighting factor of 360/256 for those two channels respectively.

Fig. 2. Facial image samples from stereo video.

The performance of the algorithm on the stereo video was quite promising. This video was taken from a concert and 3 clusters existed to this video, each corresponding to a different person. The total number of facial images was 270. The F-measure performance of the algorithm was at 81.79% which is quite well considering the various illuminations and postures of each person.

5. CONCLUSIONS

In this paper we have modified and evaluated a clustering algorithm using Ncuts and spectral graph clustering techniques. The results are quite promising and encourage us to continue our research in this field. We are now orienting our efforts in devising a method that would improve the results and also could cope with more demanding clustering problems aiming at facial image clustering in stereo videos that would more effectively benefit from extra information in stereo video.

Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTVS). This publication reflects only the author's views. The European Union is not liable for any use that may be made of the information contained therein.

6. REFERENCES

[1] Chrysouli, C., Vretos N., and Pitas I.. "Face clustering in videos based on spectral clustering techniques." First Asian Conference on Pattern Recognition (ACPR), 2011.

[2] Foucher, S., and Langis G.. "Automatic detection and clustering of actor faces based on spectral clustering techniques." Fourth Canadian Conference on Computer and Robot Vision, CRV'07, 2007.

[3] Jordan M. I. and Bach F. R. "Learning spectral clustering." Proceedings of the 2003 Conference in Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference. Vol. 16. MIT Press, 2004

[4] Zengyou H., Xiaofei X., and Shengchun D. "k-ANMI: A mutual information based clustering algorithm for categorical data." Information Fusion, vol. 9(2), pp. 223-233, 2008.

[5] Laptev, I., et al. "Learning realistic human actions from movies." IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR. 2008.

[6] Jianbo S, and Malik J. "Normalized cuts and image segmentation." , IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 22(8), pp. 888-905, 2000.

[7] Schaeffer, S. E. "Graph clustering." Computer Science Review vol 1(1) , pp. 27-64, 2007.

[8] Stein, B., zu Eissen M., and Wissbrock F. "On cluster validity and the information need of users." Artificial Intelligence and Applications. ACTA Press, 2003.

[9] C. Studholme, “Measures of 3D Medical Image Alignment”, Doctoral dissertation, King’s College London, University of London, 1997.

[10] Von Luxburg, U. "A tutorial on spectral clustering." Statistics and Computing, pp. 395-416, 17.4.2007.

[11] Vretos, N., Solachildis V., and Pitas I. "A mutual information based face clustering algorithm for movies." IEEE International Conference on Multimedia and Expo, 2006.

[12] Vretos, N., Solachildis V., and Pitas I. "A Face Tracker Trajectories Clustering Using Mutual Information." Multimedia Signal Processing,. MMSP 2007. IEEE 9th Workshop on. IEEE, 2007.

[13] Vretos, N., Solachildis V., and Pitas I. "A mutual information based face clustering algorithm for movie content analysis." Image and Vision Computing, pp. 693-705, 29.10.2011