INITIALIZATION-INDEPENDENT SPECTRAL CLUSTERING WITH APPLICATIONS TO AUTOMATIC VIDEO ANALYSIS

Ahmet Ekin[a], Sharath Pankanti[b], Arun Hampapur[b]*

[a]Video Proc. and Visual Perception Group, Philips Research, Eindhoven, The Netherlands
[b]Exploratory Computer Vision Group, IBM T. J. Watson Research Lab, Hawthorne, NY, USA
ahmet.ekin@philips.com,{sharat,arunh}@us.ibm.com

ABSTRACT

Popular clustering algorithms, such as K-Means (KM) and Expectation Maximization (EM), are sensitive to the initialization of cluster centers. In contrast, recently proposed K-Harmonic Means (KHM) algorithm is more robust to the randomness of the initialization. However, KHM works best when the dimensionality of the data (N) is small (usually less than 8). Because the dimensionality of features that are used for many clustering problems in image/video and speech processing is large, the benefits of KHM cannot be exploited. Based upon this observation, this paper proposes a novel method to employ KHM for high-dimensional data so as to realize *initialization-independent clustering.* The proposed method employs efficient spectral clustering techniques whereby the affinity matrix of the data is decomposed into its eigenvectors and k (total number of clusters) eigenvectors corresponding to the k largest eigenvalues are retained. That is, we represent N-D data by k-D transformed data when $k < N$ and propose to employ KHM over this k-D transformed data. The assumption of $k < N$ indeed encompasses a large number of significant video processing and computer vision problems where the use of KHM was not beneficial before. We demonstrate the effectiveness and the efficiency of the proposed algorithm for *face clustering* in the domains where the number of persons (k) is not large, such as anchorperson grouping, video-based speaker clustering in videoconferencing, and identity-based tracking in small office environments.

1. INTRODUCTION

Clustering is an unsupervised grouping of data [1] and has many applications in speech/video processing, computer vision, data mining, and statistics. Clustering is different from discriminant analysis (supervised classification) which involves a training stage with a set of labeled data (hence supervised) so that discriminating features of the patterns can be identified beforehand. Because there is no such pre-labeled data available for clustering algorithms, the representative samples (cluster centers) have to be determined automatically. It is desirable that the final set of clusters be invariant to a particular selection of cluster centers. However, a common drawback of some popular clustering algorithms, such as K-Means (KM) and Expectation Maximization (EM), is their dependency on the initialization of cluster centers that results in undesirable fluctuations in the system performance.

Recently, K-Harmonic means (KHM) algorithm [2] has been introduced to remedy the inconsistencies of clustering results due to different initializations. However, KHM is independent of the quality of initialization iff the dimensionality of the data is small (usually less than 8) [3]. Otherwise, the performance of KHM is similar to that of KM and EM. As a result, the benefits of KHM cannot be exploited for high-dimensional data clustering problems. Because the dimensionality of many features that are employed for image/video analysis is large, a robust clustering algorithm for high-dimensional data is extremely desirable. To this effect, this paper proposes a novel *initialization-independent clustering algorithm for high-dimensional data.* We also demonstrate the effectiveness of the proposed algorithm in a *face clustering* application.

One of the appealing features of the proposed algorithm is its representation of the data in a transformed domain that is computed by eigenvalue decomposition of the data affinity matrix. This type of clustering is referred to as *spectral clustering* and it is shown to be able to deal with many types of distributions that would be problematic to cluster directly in the data domain [4]. Another favorable property of spectral clustering approach that we are going to exploit is its transformation of N-dimensional data to k (total number of clusters) dimensions before employing a clustering algorithm. In this paper, we propose to use k eigenvectors of the affinity matrix instead of N-dimensional data as an input to KHM when *k is smaller than N* and *in the range of KHM requirements.* This transformation effectively makes it possible to use initialization-independent KHM clustering algorithm for high-dimensional data as long as k is small.

We observe that a large number of video/speech processing and computer vision problems fit nicely into the above assumption, i.e., N being large and k being small. Examples include *automatic image/video classification* where the number of classes is usually limited whereas a large number of low-level spatio-temporal features are used, *object grouping for video surveillance, audio-based speaker recognition* when the number of speakers is not large as it is the case for news and sports broadcasts. As a demonstration of the proposed clustering method, this paper also introduces *a novel face clustering algorithm* that is particularly useful for video-based anchorperson grouping, video-based speaker clustering in videoconferencing, and identity-based tracking in small office environments.

In the next section, we describe the proposed clustering algorithm in detail. In the same section, we also present important aspects of spectral clustering techniques and KHM. In Section 3, the proposed face clustering algorithm, which is based on widely-known eigenface approach [5], is explained. Section 4 presents an

This work was performed when the first author was a summer intern at IBM T.J. Watson Research Lab, Hawthorne, NY.

extensive set of experiments while Section 5 concludes the paper.

2. PROPOSED CLUSTERING ALGORITHM

This section first gives general information on spectral methods, describes one of the efficient spectral clustering algorithms commonly used in the literature, and then explains the KHM algorithm. Finally, we introduce the proposed clustering algorithm.

2.1. Spectral Methods for Clustering

Spectral methods have recently emerged as powerful tools for clustering. In general, spectral clustering involves the use of top eigenvectors of data affinity matrix for transformation of the actual data points. After that, transformed data can be clustered by any of the clustering methods, such as KM and EM. As we also demonstrate with an example, this transformation-based approach was shown to be superior to clustering directly in the data domain [4].

In this paper, we obtain the spectral representation of the data by the algorithm proposed in [4]. Given a set of M points $S = \{s_1, s_2, ..., s_m\}$ in R^N that are to be represented by k clusters:

- Compute the MxM data affinity matrix A defined by
$$A_{ij} = exp(-\parallel s_i - s_j \parallel^2 / 2\sigma^2).$$
- Define D as the diagonal matrix whose i^{th} diagonal is the sum of i^{th} row of A, i.e., $D_{ii} = \sum_{j=1}^{M} A_{ij}$.
- Compute the Laplacian $L = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$.
- Find k largest eigenvectors of L and form the matrix V_{Mxk} whose columns are k selected eigenvectors.
- Normalize each row of V so that each row has a unit length, i.e., $Y_{ij} = V_{ij} / (\sum_{j=1}^{k} Vij^2)^{\frac{1}{2}}$.
- Cluster the M rows of the Y matrix treating them as M points (In [4], KM is used, but we will use KHM because of the aforementioned reasons).
- Assign the original point s_i to the cluster of the i^{th} row of Y.

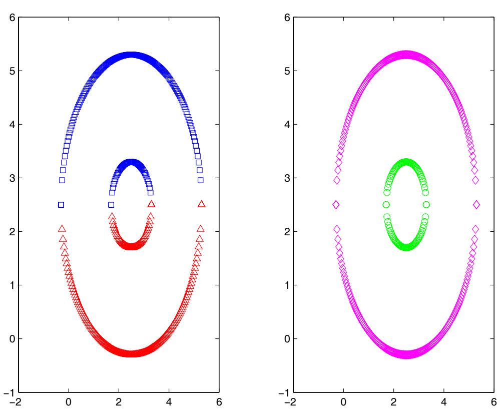
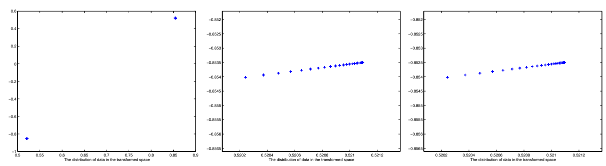


Fig. 1. Clustering results after running KM in the data domain (left) and in the transformed domain (right)

An important parameter above is σ whose value is critical in the efficiency of the algorithm. There is not yet a formal method in the literature for its automatic computation. The common practise is to try several alternatives and pick the best fitting one to the data. In spite of this, we have found out that a fixed σ value for a certain application, such as face clustering, usually gives robust results. Therefore, in the remainder of the paper, we will not be concerned about the selection of σ.

In Figure 1, an example case where the data-domain clustering fails whereas the spectral method yields correct clustering is demonstrated. The reason of the success of the spectral method is graphically shown in Figure 2, where the transformed points form linearly separable clusters.



(a) (b) (c)

Fig. 2. a)The distribution of the points in the transformed domain, b) zoomed view around the first cluster, c) zoomed view around the second cluster

2.2. K-Harmonic Means (KHM) Algorithm

K-Harmonic means is a clustering algorithm that assigns soft membership to each data point. In that sense, it resembles Fuzzy clustering methods. The performance function that is optimized in KHM is given in Equation 2 and that of KM is defined in Equation 1 for comparison. As seen in Equation 2, KHM uses harmonic averages of data points in the performance function, which also gives the algorithm its name. The implementation requires some measures against numerical instabilities [2]. KHM is reported to be robust to variations in initialization when the dimensionality of the data is low; hence, the clustering results are reproducible. It is shown that KHM demonstrates similar problems to KM and EM when the dimensionality of the data is greater than or equal to 8 [3].

$$\sum_{l=1}^{K} \sum_{x \epsilon S} \|x - m_l\|^2 \tag{1}$$

$$\sum_{i=1}^{M} \frac{K}{\sum_{i=1}^{K} \frac{1}{\|x - m_l\|^2}} \tag{2}$$

2.3. Proposed Clustering Algorithm

We propose a clustering algorithm that integrates spectral clustering methods and KHM in order to achieve initialization-independent clustering for high-dimensional data. The proposed method is particularly effective for N-D data, where $N >= 8$, that will be clustered into k clusters where $k < 8$, which can be effectively clustered by KHM. Given the same set of data in Section 2.1, i.e., M points $S = \{s_1, s_2, ..., s_m\}$ in R^N that are to be clustered into k clusters:

- Find Mxk spectral representation matrix Y of S by using the algorithm in Section 2.1.
- Define the new transformed set S^T in R^k as the M rows of Y.

- Apply KHM algorithm over the new transformed data S^T.

3. FACE CLUSTERING

In this section, we propose a novel automatic *frontal face clustering algorithm.* The proposed method uses robust face detection algorithm developed by Viola and Jones [6], that is extended and made publicly available by Lienhart and Maydt [7]. We also automatically extract facial features, mainly eyes and nose, by the algorithm in [8]. After that the faces are mapped to a standard space by appropriate translation, rotation, and scaling coefficients. The face images are also intensity-normalized in order to mitigate the effects of the intensity variations between the face images of the same person.

In order to cluster face images, a set of features has to be extracted. We use the projections of each normalized face image to eigenface space as our features. This is the same set of features used for recognition in [5]. In general, because 30 or more eigenfaces are used for generic data sets, the dimensionality of each face feature is very large; hence, direct application of KHM is not beneficial. Instead, when the number of persons in a specific application is limited, we can use the algorithm proposed in Section 2.3 to achieve initialization-independent clusters. Although the condition of having limited number of people looks too restrictive at first sight, a number of important problems has these properties. For example, clustering faces of a small office occupants, anchorpersons in the news video, face-based identity tracking in retail industry are some of the applications of the proposed face clustering approach.

The face clustering algorithm we are proposing is as follows:

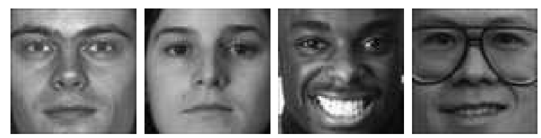
- Given a set of automatically detected and normalized M_1 intensity face images, $F_C = \{f_1, f_2, ..., f_{M_1}\}$, belonging to k individuals (Clusters) and a fixed M_2 face images $F_{DB} = \{f_1, f_2, ..., f_{M_2}\}$, construct the new set $F = F_C U F_{DB}$ (We use F_{DB} as well as F because images in F_C may not have enough variations to be useful for eigenface analysis.).
- Find eigenfaces of the whole set F and retain N eigenfaces that correspond to N largest eigenvectors (N can be determined automatically to satisfy an error criterion or can be fixed).
- Project each of the M_1 faces in F_C to N eigenfaces to construct $W = \{w_1, w_2, ..., w_{M_1}\}$ in R^N.
- Apply the the proposed clustering algorithm explained in Section 2.3 on W to find k clusters of faces by KHM.

3.1. Related Work

In the literature, although many algorithms for face recognition exist, there are only few works about face-based person clustering. In [9], a face clustering algorithm is presented. The proposed algorithm requires training of HMM for each iteration and is computational. Furthermore, because it uses KM, the clustering results is dependent on the selection of initial cluster centers. In [10], an algorithm for video-based person-of-interest retrieval is presented. This algorithm is not based on clusters, however, it is different from the mainstream face recognition algorithms. The proposed method requires manual specification of different face poses of each queried person. Name-It [11] recognizes faces in the news by visual and textual cues. Similar to [10], Name-It is not concerned with face clustering but identity descriptor extraction.

4. RESULTS

The proposed algorithm has been tested over some portion of FERET database and a small set of frontal face images captured at IBM Exploratory Computer Vision Lab. Independent of the specific application, we use 100 frontal images from FERET *fa* and *fb* data sets, some of which are shown in Figure 3, as F_{DB} to introduce some variations to the face images that will be clustered, which make up F_C. F_{DB} and F_C are used to compute eigenfaces. In all of the experiments, we use 30 eigenfaces that correspond to 30 largest eigenvalues, i.e., the dimensionality of the features we would like to cluster is $N = 30$. The first 15 eigenfaces for an example set are shown in Figure 4.



(a) (b) (c) (d)

Fig. 3. Some of the normalized frontal faces from FERET data sets of *fa* and *fb*

Fig. 4. Top 15 eigenfaces for a specific clustering example

In order to test the effectiveness of the eigenface projection features in clustering, we performed a clustering experiment over a different set of FERET images that are not included to F_{DB}. We selected 90 images of 45 individuals where each person is represented by two images, one in *fa* and one in *fb.* In that experiment, 41 pairs were correctly clustered, making the accuracy of the overall system 91.1%. However, the result corresponds to the best result of the KM algorithm. Reruns over the same set usually do not generate the same result. The proposed clustering algorithm in this paper is not applicable to this example because of the large k (45).

The effectiveness of the proposed algorithm is demonstrated over a set of 39 frontal face images of 6 people. Each normalized face is represented by 30-dimensional eigenface projection vector. Top 6 eigenvectors are retained in spectral representation (we used

$\sigma = 25$ in all of the experiments). The best result we obtained in this data set is given in Figure 5, where five of the labels are incorrect making the total clustering accuracy 87.8%.

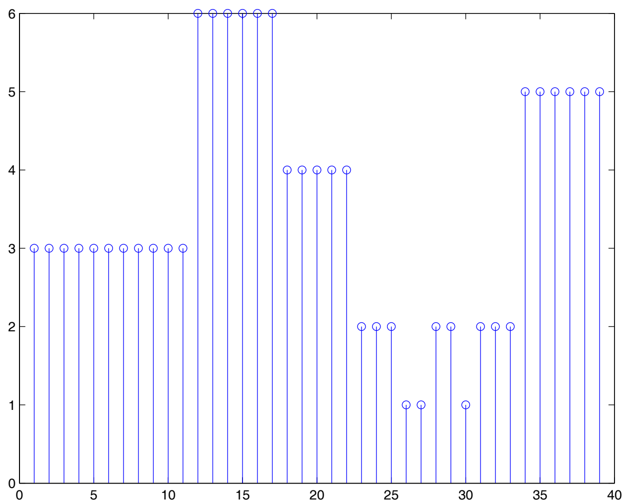


Fig. 5. The best result of clustering 39 images to 6 individuals (the images labeled 26-33 belong to the same individual).

The result above is defined as the ground truth and the consistency of KHM and KM is compared. Starting with randomly initialized cluster centers, KHM resulted in the result in Figure 5 47 times out of 50. In contrast, KM provided 9 different clusters in 50 runs with lower accuracy than KHM.

The consistency of KHM is further tested in the worst cases, where we instantiated the cluster centers as the face images of the same person. The algorithm resulted in the same cluster in Figure 5 in all 20 trials that include starting with the clusters on four of the persons alone, and two of them together.

5. CONCLUSIONS

In this paper, we presented an initialization-independent clustering algorithm for high-dimensional data that can be represented by a few clusters. We applied powerful spectral clustering approach to overcome the limitations of KHM for high-dimensional data. As an example to the proposed clustering algorithm, we also introduced a face clustering method that can be used in video surveillance, tracking, and video indexing and retrieval applications.

The main limitation of the proposed algorithm is the requirement of having small number of clusters. We are currently working on applying hierarchical spectral clustering techniques to deal with this limitation.

6. REFERENCES

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, pp. 264–323, Sept. 1999.

[2] B. Zhang, M. Hsu, and U. Dayal, "K-harmonic means - a data clustering algorithm," HP Technical Report, 1999.

[3] B. Zhang, "Comparison of the performance of center-based clustering algorithms," in *PAKDD-03*, May 2003.

[4] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Advances in Neural Information Processing*, 2001.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71–86, 1991.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE CVPR*, 2001.

[7] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *IEEE ICME*, 2002.

[8] A.W.Senior, "Face and feature finding for a face recognition system," in *Proc. Audio- and Video- based Biometric Person Authentication*, March 1999, pp. 154–159.

[9] S. Eickeler, F. Wallhoff, U. Iurgel, and G. Rigoll, "Content-based indexing of images and video using face detection and recognition methods," in *IEEE ICASSP*, 2001.

[10] L. Torres and J. Vila, "Automatic face recognition for video indexing applications," *Pattern Recognition*, vol. 35, pp. 615–625, Dec. 2001.

[11] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: naming and detecting faces in news videos," *IEEE Multimedia*, vol. 6, no. 1, pp. 22–35, Jan-Mar. 1999.