

A Robust Clustering Algorithm for Video Shots Using Haar Wavelet Transformation

Jia Liao

College of Information Science
Northeastern University, China

liaojia_email@yahoo.com

Bo Zhang

College of Information Science
Northeastern University, China

crzhang_bo_1030@yahoo.com.cn

ABSTRACT

Automatic clustering of video shots is an important issue of video abstraction, browsing and retrieval. Most of the existing shot clustering algorithms need some prior domain knowledge or thresholds to obtain good clustering results, and they also have to face the difficult task of choosing proper initial cluster centers. To resolve the discommodious problems for users, this article proposes a robust unsupervised shot clustering algorithm which is called CAVS (Clustering Algorithm for Video Shots). In CAVS, multi-resolution analysis and Haar wavelet transformations are first applied as a dimensionality reduction approach for the high-dimensional feature vectors of shots. Then CAVS performs on the remained subspace and merges the most similar shots into one cluster by the iterative merging procedures. The iterative merging procedures are repeated until a novel stop criterion based on the theory of Fisher Discriminant Function is satisfied, and the clustering results and the number of clusters are obtained without any parameters.

1. INTRODUCTION

Video shot is the basic structure unit and content unit of videos, and it comprises a sequence of interrelated consecutive frames taken continuously by a single camera and represents a continuous action in time and space [6]. As a bridge of low-level features and high-level contents of videos, shot clustering becomes an important issue in video abstraction, browsing and retrieval. The results of shot clustering directly influence the further content analysis of videos.

Most of the existing shot clustering algorithms are designed inheriting the idea of the traditional algorithm *k-means* and some hierarchical clustering algorithms. These methods, however, require some prior domain knowledge, some parameters or some thresholds to obtain good clustering results which bring so much troubles for the users. [5] needs two pre-defined thresholds for the splitting and merging procedures of clustering. The number of clusters and a threshold ϵ are needed in the shot clustering processing by

k-means for the further indexing of videos in [4]. *X-means* algorithm which is a reformative one of *k-means* for shots was proposed in [2] and it applied Bayesian information criterion to estimate the number of clusters. [1] introduced a two-level hierarchical clustering by making use of both color and motion features of shots. Among the algorithms above, although [2] and [1] do not require parameters, they both have to face the problem of choosing proper initial cluster centers and it is well known that the decision of good cluster centers has been a longstanding problem in cluster analysis. But automatic shot clustering is actually an important tache in the field of video content analysis.

Generally, video shots are usually represented as high-dimensional feature vectors. But, the dilemma of "dimensionality curse" directly influence the clustering results of most existing algorithms, dimensionality reduction appears to be the most promising method to solve this problem. Wavelet transformation is a useful approximation scheme and multi-resolution analysis is an efficient approach for the processing and analysis of digital signals. In the paper, they work together as an efficient dimensionality reduction method to process high-dimensional data.

As a solution of the problems of most existing algorithms and the "dimensionality curse", we propose a robust and efficient clustering algorithm called CAVS for video shots. For CAVS, Haar wavelet and multi-resolution analysis is first used to achieve the goal of dimensionality reduction. Then CAVS which is an unsupervised algorithm performs on the remaining subspace of dimensions. By a novel stop criterion which is taking Fisher Discriminant Function for reference, the merging procedures of CAVS can stop automatically and estimate the number of clusters at the same time. CAVS does not need any parameters to give beforehand, and avoids the problem of choosing proper initial cluster centers.

2. DIMENSIONALITY REDUCTION

Haar wavelets are one of the most elementary example of wavelets and they are effective for most applications. Multi-resolution analysis is an efficient approach for the processing and analysis of digital signals and it always works with some approximate transformations. In our work, they just work together as an efficient dimensionality reduction method.

By Haar wavelet transformation, each video shot which is a high-dimensional feature vector can be mapped into the Haar coefficients space. Because the reconstruction procedures of Haar wavelets are hierarchical, Haar coefficients can be divided into two parts, the principal part which includes the first few coefficients contains an overall approximation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Proceedings of SIGMOD2007 Ph.D. Workshop on Innovative Database Research 2007(IDAR2007), June 10, 2007, Beijing, China.

of the data and the detailed part which contains the additional coefficients. Actually the Euclidean distance of two high-dimensional feature vectors of video shots in their Haar coefficients space is equal to their original distance (in the interest of space, the proof is not given). Thus, the relationships among the original data are preserved.

With multi-resolution analysis, we can conclude that the principal part of Haar coefficients are enough to deliver the intrinsic characteristic of the original video shot. And for the application of clustering, the intrinsic characteristic of data which represents the distribution of data is the most important part to be taken into account. Thus, with the multi-resolution analysis, in Haar coefficients space, the first N_c coefficients which corresponds to the principal part are remained for approximating the original data. That achieves the goal of dimensionality reduction.

3. THE ROBUST SHOT CLUSTERING ALGORITHM (CAVS)

CAVS is an unsupervised merging clustering algorithm. By the iterative merging procedures, similar shots or clusters are merged into one cluster, and a novel stop criterion for the merging iterations make the merging procedure stop without any parameters. When the merging procedures stop, the clustering results are obtained and the number of clusters is estimated automatically at the same time.

At the beginning of CAVS, each video shot is initialized as one cluster, then the merging procedures begin to perform on them. In each iterative merging procedure, a cluster is represented by its centroid, and two clusters whose distance is the smallest are merged into one. The merging procedures repeat until the novel stop criterion for the algorithm is satisfied.

The stop criterion for the iterations is the most critical technique of unsupervised clustering algorithm. It directly determines the results of clustering. In the paper, we devise a novel stop criterion which uses Fisher Discriminant Function [3] for reference for the merging procedures of CAVS, and it synthetically considers the relationship of the intra-distance of each cluster and the inter-distance among different clusters. In addition, with the stop criterion, CAVS can estimate the number of clusters without any parameters.

For CAVS, we use two factors for the stop criterion to evaluate the clustering results. One is r_l which is used to measure the relationship of the intra-distance and the inter-distance of clusters, the other is n_l which can express the ratio of merging degree and indirectly estimate the number of clusters.

Let r_l denote the ratio of the intra-distance of one cluster over the inter-distances among clusters when the number of clusters is N_l , and the best clustering results we want are the results with smallest value of r_l . The value of r_l can be calculated by the Equation (1) below:

$$r_l = \frac{\sum_{c=0}^{N_l} d_w^c}{d_t} = \frac{\sum_{c=0}^{N_l} \sum_{i=0}^{m_c} |S_i^c - S_r^c|_2}{\sum_{j=0}^N |S_j - S_r|_2} \quad (1)$$

where d_t is the initial distance among clusters, d_w^c is the intra-cluster distance of cluster c . N is the initial number of clusters at the beginning, while m_c is the number of shots

in cluster c . $|\bullet|_2$ denotes the Euclidean distance. S_i^c and S_r^c represent the i th shot and the representative vector of cluster c respectively, while S_j and S_r are used for denoting the same concept of the initial clusters.

Apart from r_l , the other important factor n_l which is the statistic information of the number of clusters is considered in CAVS. n_l can be defined as the following Equation(2), it is the ratio of the cluster number N_l over the initial total number of shots N and indicates the merging degree of the algorithm.

$$n_l = \frac{N_l}{N} \quad (2)$$

At the beginning of CAVS, each shot is initialized as one cluster, the value of r_l is 0, and the value of n_l is 1. Then as the merging procedures proceed, the value of r_l is increasing while n_l is descending. When all the shots are merged into one cluster, the value of r_l reaches 1, and n_l reaches its smallest value. Since the encouraging clustering results should have both smaller r_l and n_l , we make a tradeoff of the two factors and choose $sp = \min(r_l + n_l)$ as the stop criterion for CAVS. When $r_l + n_l$ reaches its smallest value sp , the iterations of merging stop.

4. CONCLUSIONS

In the paper, we have introduced a robust clustering algorithm for video shots called CAVS. It is an unsupervised clustering algorithm, no special domain knowledge and parameters are needed and it avoids the problem of choosing proper initial cluster centers. By the novel stop criterion, it can also estimate the number of clusters automatically.

Acknowledgments This work is supported by National Basic Research Program of China(973) under Grant No.2006 CB303103, and partially supported by National Natural Science Foundation of China under grant No.60573089 and 60273079. We thank our advisor Professor Guoren Wang for his instructions in our research topic.

5. REFERENCES

- [1] C. W. Ngo, T. C. Pong, and H. J. Zhang. On clustering and retrieval of video shots. In *Proceedings of ACM Multimedia Workshops*, pages 51–60, 2001.
- [2] D. Pelleg and A. W. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. In *Proceedings of ICML*, pages 727–734, 2000.
- [3] M. Sever, J. Lajovic, and B. R. Robustness. Robustness of the fisher's discriminant function to skew-curved normal distribution. In *Proceedings of Int. Conf. of Applied Statistics*, pages 231–242, 2005.
- [4] H. T. Shen, B. C. Ooi, and X. F. Zhou. Towards effective indexing for very large video sequence database. In *Proceedings of SIGMOD*, pages 730–741, 2005.
- [5] X. Q. Zhu, J. P. Fan, and A. K. Elmagarmid. Hierarchical video content description and summarization using unified semantic and visual similarity. *Journal of Multimedia Systems*, 9(1):31–53, 2003.
- [6] X. Q. Zhu, X. D. Wu, and A. K. Elmagarmid. Video data mining: Semantic indexing and event detection from the association perspective. In *Proceedings of TKDE*, pages 665–677, 2005.