Noisy data reduction by Using Tensor and Fuzzy C-means Algorithm.

Mongkol Hunkrajok and Wanrudee Skulpakdee
Mathematics and Statistics Program Faculty of Science
Chandrakasem Rajabhat University
39/1 Ratchadapisek Road , Khet Chatuchak , Bangkok , 10900
Thailand

Abstract: - Classification of image (both 2D and 3D) and noisy data using eigenvalues of tensor as features is found to be simple, but effective method for reducing noise. The features constitute a systematic structure that can be segmented one from another. We propose the segmentation of class clustering by fuzzy c-mean algorithm which can be applied to classify image and noisy data; thus, unnecessary data from the systems can be removed.

Key-Words: - tensor, eigenvalue, clustering, fuzzy c-mean.

1 Introduction

When collecting 2D or 3D data using tools like sensors or scanners [1], [2], there usually is unnecessary data found, called “noise” – which takes computer memory and reduces the quality and clarity of the images. There are various approaches in Image Processing aiming to remove noise, reduce data and preserve the shape and fine details, including: Applying edge-preserving filtering on scanned points [1]; Applying Fuzzy c-mean algorithm to search for clusters that can be described by circular arcs [3]; or Using Tensor Voting for the robust inference of features from noisy data [4],[5],[6],[7]. And from the [4] studies, we find the interesting distinction of image data and noisy data from tensor.

Tensor shows relation between data using eigenvectors and eigenvalues. The [4] uses various tensor voting fields to obtain saliency features. Different shape description data contains different eigenvalues, also noisy data contains different eigenvalues, and therefore, we can use eigenvalues as feature to separate noisy data from image data.

The fuzzy c-means (FCM) clustering algorithm defined by Dunn [10] and generated by Bezdek [11] is the best-known and most powerful method in cluster analysis. We use FCM to separate data sets into 2 groups: image data and noisy data.

The remaining of this paper will discuss in details as follows: Section 2 Feature capturing from Tensor; Section 3 Fuzzy c-means clustering; Section 4 Experimental result, and the final section is conclusion and future work.

2 Feature capturing from Tensor

3D local image features are encoded into a tensor field $F:\Omega \to T_3\left(\mathbb{R}^3\right)$, where Ω is the image domain, and $T_3\left(\mathbb{R}^3\right)$ denotes the set of symmetric positive semidefinite tensor on $\mathbb{R}^3$.

Let $A \in T_3\left(\mathbb{R}^3\right)$, then A is a symmetric positive semidefinite 3x3 matrix, representing tensor. We can decompose such matrix into its eigenvectors and eigenvalues

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \lambda_3 e_3 e_3^T \tag{1}$$

where λ_1 , λ_2 and λ_3 are nonnegative eigenvalues $\left(\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge 0\right)$

e_1 , e_2 and e_3 are orthogonal eigenvectors

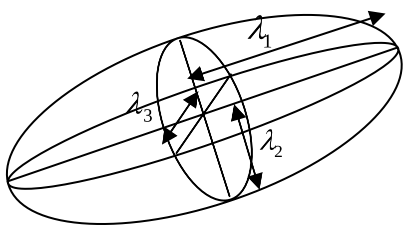


Fig.1 Graphical representation of symmetric positive semidefinite tensor on $\mathbb{R}^3$

For 2D data, tensor can be described in the same way [8]

In this paper, we search for features in arbitrary 3D images that show relation of local points by

Let x_i arbitrary point using to create tensor matrix. Let B_i a set of local points surrounding x_i which can be written as

$$B_i = \left\{ x \in \Omega \,\middle|\, \|x - x_i\| \le r \right\} \;,\; i = 1,2,......., N \qquad (2)$$

when Ω is a set of whole data
N is number of all data in Ω
r is radius constant

we consider variation along direction v among all point in B_i

$$\begin{aligned}\mathrm{var}(v) &= \sum_{x\in B_i} \left\| (x-x_i)^T \cdot v \right\|^2 \\ &= \sum_{x\in B_i} v^T (x-x_i)(x-x_i)^T v \\ &= v^T \left[\sum_{x\in B_i} (x-x_i)(x-x_i)^T \right] v \\ &= v^T A_i v \end{aligned}$$

where
$$A_i = \sum_{x\in B_i} (x-x_i)(x-x_i)^T \qquad (3)$$

A_i is tensor matrix shown feature of x_i

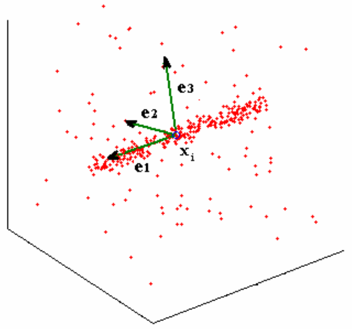


Fig.2 Eigenvectors and eigenvalues of point in 3 dimensional data.

Figure 2 illustrates eigenvectors with largest eigenvalues capture the most variation (e_1) and eigenvectors with smallest eigenvalues has the least variation (e_3) among all of vector $(x - x_i)$

One factor affecting eigenvalues of A_i matrix is a norm of vector $(x - x_i)$, which sometime affects the segmentation and lead to error clustering. To solve this problem, we normalize vector $(x - x_i)$, then equation (3) can be re-written as follows**:**

$$A_i = \sum_{x\in B_i} \frac{(x-x_i)(x-x_i)^T}{\|x-x_i\|^2} \qquad \textbf{(4)}$$

Feature capturing algorithm:
1. Initialize r to create B_i as shown in equation (2)
2. Calculate A_i matrix from equation (4)
3. Calculate eigenvalues of A_i matrix to use as features in clustering

3 Fuzzy c-means clustering

Fuzzy c-means clustering is the method for partitioning data that has an objective function

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad (5)$$

where m is real number greater than 1,
u_{ij} is degree of membership of x_i in the cluster j
x_i is any point
c_j is center of cluster j

This method used iteration for optimized an objective function by update u_{ij} and c_j from below

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \qquad (6)$$

where
$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (7)$$

Iteration will stop when

$$\max_{ij} \left\{ \left\| u_{ij}^{k+1} - u_{ij}^{k} \right\| \right\} < \varepsilon \qquad (8)$$

where ε is the termination criterion and k are iteration step.

The algorithm is the following step:

1. Initialize U^0 where $U = [u_{ij}]$ matrix
2. Calculating center in k^{th} iteration $C^k = [c_{ij}]$
3. Update U^k to be U^{k+1}
4. Return iteration until the condition $\max_{ij}\{\|u_{ij}^{k+1} - u_{ij}^{k}\|\} < \varepsilon$ is true.

4 Experimental Results

In this research, we propose noised reduction from data sets by capturing features from tensor, then using FCM to classify them into 2 clusters: image data and noisy data. We can summarize the approach as Fig. 3

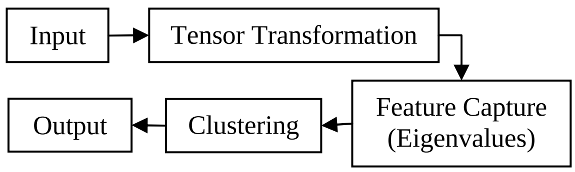


Fig.3 Complete block-diagram of the proposed noised reducing algorithm

For this experiment, we test with the synthetic 2D ellipse, 3D plane, cylinder, sphere and hyperboloid; then add random noisy data at 50%, 100% and 200% of the image data.

The 2D ellipse has been generated between 100 and 200 feature points, 3D plane and hyperboloid generated between 1,000 and 2,100 points, while cylinder generated between 3,000 and 6,000 points, and sphere generated between 15,000 and 31,000 points.

From these 3D surface sampled data sets -- after calculating eigenvalues to capture features and FCM clustering, the corresponding result is depicted in Fig.4

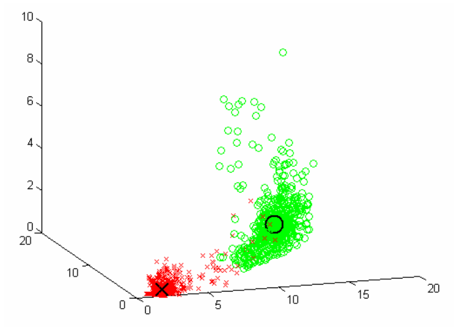
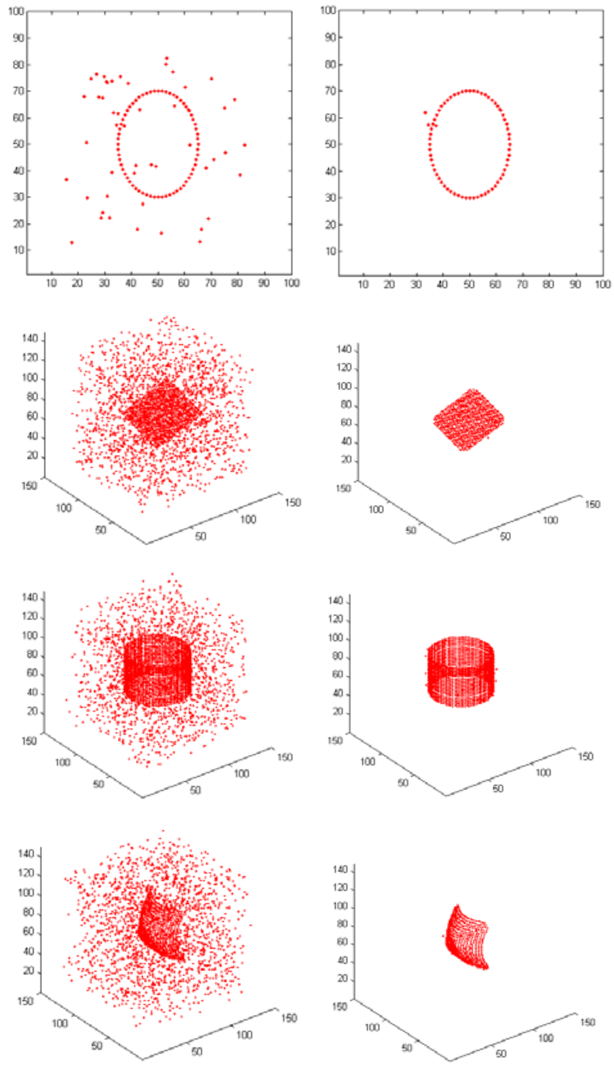


Fig.4 FCM Clustering using eigenvalues

Fig.5 (a) illustrates the samples of 2D and 3D images with noisy data, comparing with Fig.5 (b) which illustrates the sampled output after noised reducing algorithm.

Table 1 demonstrates percentages of irremovable noisy data and percentages of lost image data after clustering algorithm. From Table 1, it can be seen clearly that data size does not affect the efficiency of the clustering algorithm; but what affects it is the proportion of noisy data comparing to image data. If the data sets have low percentages of noise, the algorithm could remove most of noisy data from the sampled images. However, some feature points of the images also get lost. In contradiction, the higher percentages of noise in the date sets are, the higher percentages of irremovable noise remain, but the percentages of lost image data also decrease.



(a) (b)

Fig.5 Example data before and after noise reducing Algorithm

The irremovable noisy data and lost image data are normally found near the edges of images. The principal reason is that eigenvalues of noisy data and edging data is almost equivalent.

Table 1: Percentages of irremovable noisy data and percentages of lost image data after clustering algorithm

SHAPE	DATA SIZE	% NOISE	% IRREMOVABLE NOISE	% LOST DATA
2D ellipse	98	50%	3.08	0.00
	130	100%	23.08	0.00
	195	200%	38.46	0.00
3D plane	1014	50%	0.00	18.93
	1352	100%	0.30	15.38
	2028	200%	2.07	2.07
3D cylinder	3000	50%	2.05	0.00
	4000	100%	4.00	0.00
	6000	200%	8.35	0.00
3D hyperboloid	1014	50%	0.30	17.60
	1352	100%	1.33	15.53
	2028	200%	3.85	1.78
3D sphere	15396	50%	3.21	0.00
	20528	100%	7.26	0.00
	30792	200%	15.34	0.00

5 Conclusion

We propose the noisy data reduction by using Tensor's eigenvalues and Fuzzy C-means Algorithm, the result shows that using eigenvalues as feature for clustering can give the effective output. We find the error of this algorithm to be losing data of images when percentages of noisy data in the set are low. In the other hand, when the percentages of noise in data sets have increased, the percentages of the error have decreased.

However, this research is studied and based on synthetic data, so we would experiment our future work with the real data.

References:

[1] A.Miropolsky and A.Fischer , Reconstruction with 3D Geometric Bilateral Filter , *ACM Symposium on Solid Modeling and Applications*, 2004 , pp.225-229.

[2] Xiaoyou Ying , John Koivukangas , Jarkko Oikarinen , Jyrki Alakuijala and Yrjo Louhisalmi , Strategy for Removal of surrounding noise in 3D Head Tomographic Images without changing Brain image Data , *IEEE* , pp.1874-1876.

[3] Raghu Krishnapuram , The Fuzzy C Spherical Shells Algorithm : A New Approach , *IEEE Transactions on neural networks* , Vol.3 , No.5 , 1992 , pp.663-671.

[4] Gerard Medioni, Chi-Keung Tang and Mi-Suen Lee,Tensor Voting:Theory and Applications, *in 12eme Congres Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, Feb 2000.

[5] Mi -Suen Lee and Gerard Medioni , Grouping .,- ,→, 0,into Regions , Curves and Junctions , *Computer Vision and Image Understanding* , Vol.76 , No.1 , 1999 , pp.54-69.

[6] Chi-Keung Tang and Gerard Medioni , Robust Estimation of Curverture Information from Noisy 3D Data for Shape Description , *National Science Foundation* , No.9811883.

[7] Gerard Medioni , Chi-Keung Tang and Mi -Suen Lee , *A Computer Framework for Segmentation and Grouping* , Elsevier , 2000.

[8] Erik Franken, Markus van Almsick , Peter Rongen , Luc Florack and Haar Romeny , An Efficient Method for Tensor Voting Using Steerable Filters , *Springer-Velag Berlin Heidelberg* , ECCV2006 , partIV , LNCS3954 , 2006 , pp.228-240.

[9] Carl-Fredrik Westin , Tensor Framework for Multidimensional Signal Processing , *Linkoping Studies in Science and Technology* , Dissertations , No.348 , 1994.

[10] J.C.Dunn, A Fuzzy relative of the ISODATA process and its use in detecting compact well-seperated clusters , *J. Cybernet* , Vol.3 , 1974 , pp.32-57.

[11] J.C.Bezdek , Pattern Recognition with Fuzzy Objective Function Algorithm , *New York : Plenum Press* , 1981.