

A penalized fuzzy clustering algorithm

Miin-Shen Yang^a, Wen-Liang Hung^b and Chia-Hsuan Chang^a

^aDepartment of Applied Mathematics,
Chung Yuan Christian University,
Chung-Li 32023, Taiwan, ROC

^bDepartment of Applied Mathematics,
National Hsinchu University of Education,
Hsin-Chu, Taiwan, ROC

Abstract: In this paper, we propose a penalized inter-cluster separation (PICS) fuzzy clustering algorithm by adding a penalty term to the inter-cluster separation (ICS) algorithm. Numerical comparisons are made for several fuzzy clustering algorithms according to criteria of accuracy and computational efficiency. The results show that the PICS has better accuracy and efficiency. Image segmentation is an important step in any image analysis system. Existing various segmentation methods for magnetic resonance image (MRI) have been used to differentiate abnormal and normal tissues. We apply the PICS algorithm to the MRI segmentation of an ophthalmic patient. In these MRI segmentation results, we find that PICS provides useful information as an aid to diagnosis in ophthalmology.

Key-Words: Fuzzy clustering; fuzzy c-means (FCM); inter-cluster separation (ICS); penalized ICS; image segmentation; Magnetic resonance image (MRI).

1 Introduction

Cluster analysis is a method of grouping data with similar characteristics into larger units of analysis. Since Zadeh [17] first articulated fuzzy set theory which gave rise to the concept of partial membership, based on a membership function, fuzziness has received increasing attention. Fuzzy clustering, which produces overlapping cluster partitions, has been widely studied and applied in various areas. In fuzzy clustering, the fuzzy *c*-means (FCM) clustering algorithm is the best known and most powerful methods used in cluster analysis ([1]).

The idea of penalization is important in statistical learning. For example, ridge regression shrinks the regression coefficients by imposing a penalty on their size. Based on penalty idea, we added a penalty term to the inter-cluster separation (ICS) clustering algorithm ([6]) and then proposed the penalized ICS (PICS). Numerical comparisons are made with several fuzzy clusterings according to criteria of accuracy and computational efficiency.

MRI segmentation provides important information for detecting a variety of tumors, lesions, and abnormalities in clinical diagnosis. As described by Yang et al. ([12]), most medical images often present overlapping gray-scale intensities for different tissues. MRI medical imaging uncertainty is widely presented in data because of the noise and blur in acquisition and

the partial volume effects originating from the low resolution of the sensors. In particular, borders between tissues are not clearly defined and memberships in the boundary regions are intrinsically. Therefore, fuzzy clustering methods are suitable for the MRI segmentation (see [4],[7], [8], [10],[12]). In this paper, the PICS algorithm is applied to the segmentation of magnetic resonance image (MRI) of an ophthalmic patient. In these MRI segmentation results, we find that PICS provides useful information as an aid to diagnosis in ophthalmology.

2 A PICS fuzzy clustering algorithm

Let $X = \{x_1, \dots, x_n\} \subset R^s$ be a data set and let c be a positive integer greater than one. A partition of X into c clusters is represented by mutually disjoint sets X_1, \dots, X_c such that $X_1 \cup \dots \cup X_c = X$ or equivalently by the indicator functions u_1, \dots, u_c such that $u_i(x) = 1$ if x is in X_i and $u_i(x) = 0$ if x is not in X_i for all $i = 1, \dots, c$. This is known as clustering X into c clusters X_1, \dots, X_c by a hard c -partition $\{u_1, \dots, u_c\}$. A fuzzy extension allows $u_i(x)$ to take on values in the interval $[0, 1]$ such that $\sum_{i=1}^c u_i(x) = 1$ for all x in X . In this case, $\{u_1, \dots, u_c\}$ is called a fuzzy c -partition of X ([?]). Thus, the FCM objective function J_{FCM} is defined as

([1])

$$J_{FCM}(u, a) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - a_i\|^2$$

where $u = \{u_1, \dots, u_c\}$ is a fuzzy c -partition with $u_{ij} = u_i(x_j)$ being the membership of the data point x_j in cluster i , $a = \{a_1, \dots, a_c\}$ is the cluster centers, the weighting exponent m is a fixed number greater than one establishing the degree of fuzziness and the notation $\|x_j - a_i\|$ denotes the Euclidean distance between the data point x_j and the cluster center a_i . Thus, the FCM clustering algorithm is an iteration through the necessary conditions for minimizing J_{FCM} with the following update equations:

$$a_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m},$$

and

$$u_{ij} = \frac{\|x_j - a_i\|^{-2/(m-1)}}{\sum_{k=1}^c \|x_j - a_k\|^{-2/(m-1)}}.$$

The FCM algorithm is a well-known and powerful method in clustering analysis. One of important parameters in the FCM is the weighting exponent m . When m is close to one, the FCM approaches the hard c -means algorithm. When m approaches infinity, the only solution of the FCM will be the mass center of the data set. Therefore, the weighting exponent m plays an important role in the FCM algorithm. Recently, Yu and Yang [15] provided the theoretical analysis for selecting the parameters in some generalized FCM algorithms that including FCM.

The mixture maximum likelihood approach to clustering is a remarkable model-based clustering method. Scott and Symons [9] proposed the so-called classification maximum likelihood (CML) procedure, named first in Bryant and Williamson [2], that many of the commonly used clustering procedures correspond to applications of the maximum likelihood approach for normal groups with various restrictions on the covariance matrices and with the indicator classification variables of group membership associated with the data treated as unknown parameters. Yang [11] made the fuzzy extension of the CML procedure in conjunction with fuzzy c -partitions and called it a class of fuzzy CML procedures. On the other hand, the idea of penalization is important in statistical learning. For example, ridge regression shrinks the regression coefficients by imposing a penalty on their size. Combining the CML procedure and penalty idea, Yang ([11]) added a penalty term to the FCM objective function J_{FCM} and then extended the FCM to the

so-called penalized FCM (PFCM). Thus, the PFCM objective function is defined as follows:

$$J_{PFCM}(u, a) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - a_i\|^2 - w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i$$

where $w \geq 0$, $\forall i, \alpha_i \geq 0$ and $\sum_{i=1}^c \alpha_i = 1$. The necessary conditions for a minimum of $J_{PFCM}(u, a)$ are

$$a_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad \alpha_i = \frac{\sum_{j=1}^n u_{ij}^m}{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m},$$

and

$$u_{ij} = \frac{(\|x_j - a_i\|^2 - w \ln \alpha_i)^{-1/(m-1)}}{\sum_{k=1}^c (\|x_j - a_k\|^2 - w \ln \alpha_k)^{-1/(m-1)}}.$$

Based on the numerical results of Yang and Su [13], the PFCM is more accurate than FCM. Furthermore, the PFCM has been applied in various areas (cf. [4], [5], [14]).

On the other hand, by minimizing the FCM objective function and simultaneously maximizing the inter-cluster separation (ICS) measure, Özdemir and Akarun [6] proposed the ICS clustering algorithm with the objective function

$$J_{ICS}(u, a) = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^c \left(\mu_{ij}^m \|x_j - a_i\|^2 - \frac{\gamma}{c} \sum_{t=1}^c \|a_i - a_t\|^2 \right),$$

where the parameter $\gamma \geq 0$. Thus, the update equations for the ICS algorithm are as follows (see [6] and [16]):

$$a_i = \frac{\frac{1}{n} \sum_{j=1}^n \mu_{ij}^m x_j - \frac{2\gamma}{c} \sum_{t=1}^c a_t}{\frac{1}{n} \sum_{j=1}^n \mu_{ij}^m - 2\gamma},$$

$$\mu_{ij} = \frac{\|x_j - a_i\|^{-2/(m-1)}}{\sum_{k=1}^c \|x_j - a_k\|^{-2/(m-1)}}.$$

We see that the PFCM algorithm has more accuracy than the FCM method. It means that the penalty term can improve the performance of FCM. To improve the performance of ICS, we consider adding the penalty term $(-w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i)$ to ICS and call the penalized ICS (PICS). The PICS objective function is given by

$$J_{PICS}(u, a) = J_{ICS}(u, a) - w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i.$$

The update equations for minimizers of $J_{PICS}(u, a)$ are

$$\alpha_i = \frac{\sum_{j=1}^n u_{ij}^m}{\sum_{k=1}^c \sum_{j=1}^n u_{kj}^m}; \quad i = 1, 2, \dots, c,$$

$$a_i = \frac{\frac{1}{n} \sum_{j=1}^n u_{ij}^m x_j - \frac{2\gamma}{c} \sum_{t=1}^c a_t}{\frac{1}{n} \sum_{j=1}^n u_{ij}^m - 2\gamma};$$

$$i = 1, 2, \dots, c,$$

$$u_{ij} = \frac{\|x_j - a_i\|^2 - w \ln \alpha_i)^{-1/(m-1)}}{\sum_{k=1}^c (\|x_j - a_k\|^2 - w \ln \alpha_k)^{-1/(m-1)}};$$

$$i = 1, 2, \dots, c; j = 1, 2, \dots, n.$$

Thus, the PICS algorithm can be summarized as follows:

PICS Algorithm

Set the iteration counter $\ell = 0$ and choose the initial values $\alpha_i^{(0)}$, $i = 1, \dots, c$ and the initial values $\mu_{ij}^{(0)}$, $i = 1, \dots, c; j = 1, \dots, n$.

Step 1. Find $\alpha_i^{(\ell+1)}$ using (2);

Step 2. Find $a_i^{(\ell+1)}$ using (3);

Step 3. Find $\mu_{ij}^{(\ell+1)}$ using (4);

IF $\max_i \|z_i^{(\ell+1)} - z_i^{(\ell)}\| < \varepsilon$, THEN stop;

ELSE $l = l + 1$ and go to step1.

3 Numerical comparisons and application to MRI segmentation

In this section, we make a comparison of four different algorithms: FCM, PFCM, ICS and PICS, according to the bivariate normal mixtures of two classes under the accuracy and computational efficiency criteria. The accuracy of an algorithm is measured by the mean squared error (MSE) that is the average sum of squared error between the true parameter and its estimate in N repeated trials. The computational efficiency of an algorithm is measured by the average numbers of iterations (NI) in N repeated trials.

Let $N_2(\mathbf{a}, \Sigma)$ represent the bivariate normal with mean vector \mathbf{a} and covariance matrix Σ . As the separation between subpopulation is determined by varying the parameters of subpopulations, without loss of generality we give that one subpopulation bivariate normal is mean vector $\mathbf{a}_1 = \mathbf{0}$ and identity covariance matrix \mathbf{I} and the other is mean vector \mathbf{a}_2 and identity covariance matrix \mathbf{I} . That is, we consider the random sample of data drawn from $\alpha_1 N_2(\mathbf{0}, \mathbf{I}) + \alpha_2 N_2(\mathbf{a}_2, \mathbf{I})$ with $\alpha_2 = 1 - \alpha_1$. We also design various bivariate normal mixture distributions shown in Table 1. In Tests A1 and A2, we consider a well-known clustering

problem [3] where there is an inordinate difference in the number of members in each cluster. But Test A3 has almost equal size in each cluster and Test A4 has well-separated clusters.

Table 1. Various bivariate normal mixture distributions for the numerical tests

Test	mixture model
A1	$0.1N_2(\mathbf{0}, \mathbf{I}) + 0.9N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{I}\right)$
A2	$0.3N_2(\mathbf{0}, \mathbf{I}) + 0.7N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{I}\right)$
A3	$0.5N_2(\mathbf{0}, \mathbf{I}) + 0.5N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{I}\right)$
A4	$0.1N_2(\mathbf{0}, \mathbf{I}) + 0.5N_2\left(\begin{pmatrix} 3 \\ 0 \end{pmatrix}, \mathbf{I}\right)$

In each test, we consider the sample size $n = 100$, $\epsilon = 0.0001$ and $N = 500$. The MSE is calculated by

$$\frac{1}{2N} \sum_{k=1}^N \sum_{i=1}^2 \|\hat{\mathbf{a}}_i^{(k)} - \mathbf{a}_i\|^2$$

where $\hat{\mathbf{a}}_i^{(k)}$ is the estimated mean vector for the k th trial and \mathbf{a}_i is the true mean vector. How to select m depends on the user. Because most researchers have used $m = 2$, we also choose $m = 2$ in this section. Next, we choose $w = 0.5, 1$ and 2 in PFCM, and $\gamma = 0.003, 0.0005$ in ICS.

The numerical results are shown in Table 2. Compared PFCM with FCM, we see that PFCM with $w = 1.0$ can lead to a MSE reduction of 27.6% in Test A1, 27.6% in Test A1, 44.0% in Test A2, 49.4% in Test A3 and 4.9% in Test A4, respectively. These results illustrate the penalty term $(-w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i)$ to the FCM objective function can improve the accuracy of FCM. This is why we add the penalty term $(-w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i)$ to the ICS objective function. Compared PICS with ICS, we see that: (i) PICS ($\gamma = 0.003$) with $w = 1.0$ can lead to a MSE reduction of 31% in Test A1, 47.9% in Test A2, 48.8% in Test A3 and 19.3% in Test A4, respectively; (ii) PICS ($\gamma = 0.0005$) with $w = 1.0$ can also lead to a MSE reduction of 33.0% in Test A1, 54.9% in Test A2, 55.1% in Test A3 and 19.9% in Test A4, respectively. Based on the above results, we find that the reduction percentage of PICS is greater than PFCM. It illustrates that the effect of the penalty term $(-w \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \ln \alpha_i)$ on ICS algorithm is significant.

Moreover, we also find that: (i) PICS ($\gamma = 0.003$) with $w = 1.0$ has the smallest MSE in Test A1; (ii)

PICS ($\gamma = 0.0005$) with $w = 1.0$ has the smallest MSE in Tests A2 and A3; (iii) PFCM and PICS ($\gamma = 0.003$) with $w = 1.0$ have good accuracy in Tests A2 and A3; (iv) As we expected, five different algorithms have good accuracy in Test A4.

Table 2. Accuracy and computational efficiency for different clustering algorithms

Test		FCM	PFCM			ICS	
			$w = 0.5$	$w = 1.0$	$w = 2.0$	$\gamma = 0.003$	$\gamma = 0.0005$
A1	MSE	0.3557	0.3154	0.2576	0.3902	0.3560	0.3948
	NI	84.5	96.8	109.4	35.6	84.6	118.4
A2	MSE	0.1763	0.1346	0.0987	0.2615	0.1918	0.1885
	NI	49.2	46.1	64.5	50.1	64.9	64.1
A3	MSE	0.1424	0.0958	0.0720	0.1893	0.1362	0.1400
	NI	39.9	37.1	65.9	65.6	39.7	36.3
A4	MSE	0.0142	0.0128	0.0135	0.0118	0.0150	0.0156
	NI	8.2	9.1	10.4	16.2	8.5	8.2

Table 2. (Continued)

Test		PICS ($\gamma = 0.003$)			PICS ($\gamma = 0.0005$)		
		$w = 0.5$	$w = 1.0$	$w = 2.0$	$w = 0.5$	$w = 1.0$	$w = 2.0$
A1	MSE	0.3241	0.2455*	0.3457	0.2986	0.2647	0.3832
	NI	76.3	113.8	35.4	83.3	109.8	36.8
A2	MSE	0.1369	0.0999	0.2547	0.1407	0.0850*	0.2441
	NI	51.4	74.6	53.7	44.3	67.5	48.6
A3	MSE	0.1009	0.0697	0.1869	0.1023	0.0628*	0.1706
	NI	37.2	56.1	70.9	42.5	51.3	60.6
A4	MSE	0.0126	0.0121	0.0126	0.0128	0.0125	0.0112*
	NI	9.1	10.6	16.1	8.9	10.3	16.3

* represents the smallest value.

Next, we use PICS ($\gamma = 0.003, 0.0005$ and $w = 1.0$) in a real case study of MRI segmentation to differentiate between normal and abnormal tissues in ophthalmology. The MRI data sets are from a 2-yr old female patient that had been analyzed by Yang et al. [12]. She was diagnosed with retinoblastoma of her left eye, an inborn malignant neoplasm of the retina with frequent metastasis beyond the lacrimal cribrosa. The MRI images showed an intra-muscle cone tumor mass with high T1-weight signal images and low T2-weight signal images in the left eyeball. The tumor measured 20 mm in diameter and occupied nearly the entire vitreous cavity. There was a shady signal abnormality all along the optic nerve reaching as far as the optic chiasma near the brain. Here we analyzed two MRI data sets. The first MRI data set is illustrated in Figs. 1 & 2. The second MRI data set is shown in Fig. 3.

We first attempt to cluster the full size images (Figs. 1 & 2) into the same five clusters as used by Yang et al. [12]. The categories are as follows: Muscle tissue, connective tissue, nervous tissue, the lens, and tumor tissue. According to Yang et al. [12], a window segmentation (Fig. 3 for the second MRI data set) can be used to enhance areas of the tumor to better detect small tumors. We also apply PICS ($\gamma = 0.003, 0.0005$ and $w = 1.0$) to a window segmentation illustrated in Fig. 3. The lens and muscle



Fig.1 Original MR image

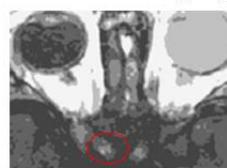


Fig. 1.1 Segmentation results of PICS($\gamma=0.003,w=1.0$)

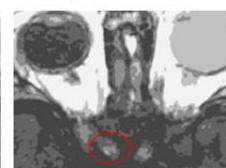


Fig. 1.2 Segmentation results of PICS($\gamma=0.0005,w=1.0$)



Fig. 2 Distorted MR image



Fig. 2.1 Segmentation results of PICS($\gamma=0.003,w=1.0$)

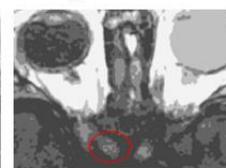


Fig. 2.2 Segmentation results of PICS($\gamma=0.0005,w=1.0$)



Fig. 3 Original MR image and its window selection

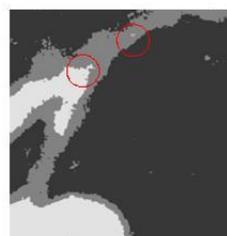


Fig. 3.1 Segmentation results of PICS($\gamma=0.003,w=1.0$)



Fig. 3.2 Segmentation results of PICS($\gamma=0.0005,w=1.0$)

tissue are excluded from the window so that the original five categories are reduced to three; connective tis-

sue, nervous tissue and tumor tissue. A gray scale histogram comparison shows that there are actually three peaks appearing in the segmentation window image.

The two pictures (Figs. 1 & 2) were processed at 400×286 pixels. The pictures are clustered into four tissue classes and one tumor class. From the red circle on the full size two dimensional MRI in Fig. 1, we can clearly detect white tumor tissue at the chiasma. PICS with $\gamma = 0.003, w = 1.0$ (see Fig. 1.1) and $\gamma = 0.0005, w = 1.0$ (see Fig. 1.2) are able to distinguish the tumor from the healthy tissue using five clusters. MRI medical imaging uncertainty is widely presented in the collected data because of noise in the partial volume effects originating from the low resolution of the sensors. Another factor causing uncertainty is the fact that the eyeball moves during the imaging and it is difficult to control this movement, especially in younger patients. A distorted MR image, shown in Fig. 2, is used here to illustrate how PICS with $\gamma = 0.003, w = 1.0$ and $\gamma = 0.0005, w = 1.0$ are able to detect tumorous tissue, despite uncertainty.

Fig. 3 in the second MRI data set was processed at 283×292 pixels. From this picture, one lesion was clearly seen in the MR image. However, some fuzzy shadows of lesions were suspected of tumor invasion. These suspected abnormalities are not easily ascertained to be tumorous. For the purpose of detecting these abnormal tissues, a window of the area around the chiasma is selected from the original MR images as shown in Fig. 3. We then applied PICS with $\gamma = 0.003, w = 1.0$ and $\gamma = 0.0005, w = 1.0$ to the window selection as illustrated in Figs. 3. We can see occult lesions (red circles) clearly enhanced with Figs. 3.1 and 3.2. This shows that PICS with $\gamma = 0.003, w = 1.0$ and $\gamma = 0.0005, w = 1.0$ give good results in MRI segmentation.

4 Conclusions

In this paper we added a penalty term to the ICS algorithm [6] and then extended the ICS to the so-called penalized ICS (PICS). Numerical comparisons are made for several fuzzy clusterings according to criteria of accuracy and computational efficiency. The results show that the PICS is better. Finally, the PICS algorithms are applied in the segmentation of the magnetic resonance image (MRI) of an ophthalmic patient. In these MRI segmentation results, we find that PICS provides useful information as an aid to diagnosis in ophthalmology.

References:

- [1] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981).
- [2] P.G. Bryant and J.A. Williamson, Maximum likelihood and classification: a comparison of three approaches, in: W. Gaul and M. Schader, Eds., *Classification as a Tool of Research* (North-Holland, Amsterdam, 1986) 35-45.
- [3] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
- [4] J.S. Lin, K.S. Cheng and C.W. Mao, Segmentation of multispectral magnetic resonance image using penalized fuzzy competitive learning network, *Computers and Biomedical Research* **29** (1996) 314-326.
- [5] S.H. Liu and J.S. Lin, Vector quantization in DCT domain using fuzzy possibilistic c -means based on penalized and compensated constraints, *Pattern Recognition* **35** (2002) 2201-2211.
- [6] Özdemir and Akarun, Fuzzy algorithm for combined quantization and dithering, *IEEE Trans. Image Processing* **10** (2001) 923-931.
- [7] D.L., Pham and J.L., Prince, Adaptive fuzzy segmentation of magnetic resonance images, *IEEE Trans. Med. Imaging* **18** (1999) 737-752.
- [8] W.E., Phillips, R.P., Velthuisen, S., Phuphanich, L.O., Hall, L.P., Clarke, M.L., Application of fuzzy c -means segmentation technique for differentiation in MR images of a hemorrhagic glioblastoma multiforme, *Magnetic Resonance Imaging* **13** (1995) 277-290.
- [9] A.J. Scott and M.J. Symons, Clustering methods based on likelihood ratio criteria, *Biometrics* **27** (1971) 387-397.
- [10] J., Suckling, T., Sigmundsson, K., Greenwood and E.T., Bullmore, A modified fuzzy clustering algorithm for operator independent brain tissue classification of dual echo MR images. *Magnetic Resonance Imaging* **17** (1999) 1065-1076.
- [11] M.S. Yang, On a class of fuzzy classification maximum likelihood procedures, *Fuzzy Sets and Systems* **57** (1993) 365-375.
- [12] M.S., Yang, Y.J., Hu, K.C.R., Lin and C.C.L., Lin, Segmentation techniques for tissue differentiation in MRI of Ophthalmology using fuzzy clustering algorithms, *Magnetic Resonance Imaging* **20** (2002) 173-179.
- [13] M.S. Yang and C.F. Su, On parameter estimation for normal mixtures based on fuzzy clustering algorithms, *Fuzzy Sets and Systems* **68** (1994) 13-28.

- [14] M.S. Yang and N.Y. Yu, Estimation of parameters in latent class models using fuzzy clustering algorithms, *European Journal of Operational Research* **160** (2005) 515-531.
- [15] J. Yu and M.S. Yang, Optimality test for generalized FCM and its application to parameter selection, *IEEE Trans. Fuzzy Systems* **13** (2005) 164-176.
- [16] J. Yu and M.S. Yang, A note on the ICS algorithm with correction and theoretical analysis, *IEEE Trans. Image Processing* **14** (2005) 973-978.
- [17] L.A. Zadeh, Fuzzy sets, *Inform. Control* **8** (1965) 338-353.