Robust RML Estimator – Fuzzy C-Means Clustering Algorithms for Noisy Image Segmentation

Dante Mújica-Vargas, Francisco Javier Gallegos-Funes,
Alberto J. Rosales-Silva, and Rene Cruz-Santiago

Mechanical and Electrical Engineering Higher School
National Polytechnic Institute of Mexico
Av. IPN s/n, Edificio Z, acceso 3, 3er piso; SEPI-Electronica,
Col. Lindavista, 07738, México D.F. México
adorinam@yahoo.com.mx, fcogf@hotmail.com

Abstract. Image segmentation is a key step for many images analysis applications. So far, there does not exist a general method to segment suitable all images, regardless if these are corrupted or noise free. In this paper, we propose to modify the Fuzzy C-means clustering algorithm and the FCM_S1 variant by using the RML-estimator. The idea to our method is to get robust clustering algorithms able to segment images with different type and levels of noises. The performance of the proposed algorithms is tested on synthetic and real images. Experimental results show that the proposed algorithms are more robust to the noise presence and more effective than the comparative algorithms.

Keywords: robust estimators, RML-estimator, Fuzzy C-Means, segmentation, noise.

1 Introduction

Image segmentation is the process of segmenting an image into several disjoint regions whose characteristics such as intensity, color, texture, etc., are significantly different with respect to the same characteristics. It is a key step in early vision problem and it has been widely investigated in the field of image processing [1]. Numerous segmentation techniques have been developed and reported in the literature. But, there does not exist a general algorithm that can excellently perform the segmentation task for all type of images.

Fuzzy clustering as a soft segmentation method has been widely studied and successfully applied to image segmentation. Among the fuzzy clustering methods, Fuzzy C-Means (FCM) [2] algorithm is the most popular method because it is simple, easy to program, and can retain much more information than hard methods. Although fuzzy clustering methods work well on the most of noise-free images, they have a serious limitation: they do not incorporate any information about spatial context, which cause them sensitivity to the noise or outliers data. Then it is necessary to modify the objective function to incorporate local information of the image to get better results.

I. Batyrshin and G. Sidorov (Eds.): MICAI 2011, Part II, LNAI 7095, pp. 474–486, 2011.

In this paper, we propose to modify the fuzzy c-means clustering algorithm, and the FCM_S1 [3], [4] using a RML estimator [5], [6], [7], [8]. The objective is to get robust algorithms to segment images under different noise conditions.

The experimental results show that the proposed algorithms are more effectives and robust to the noise than all reference algorithms.

2 Proposed Method

2.1 RML-Estimator

How it was proposed in [7], [9], the Median M-type (MM) – estimator can enhance the properties of the M- and R-estimators,

$$\theta_{medM} = med\{X_i\tilde{\psi}(X_i - med\{\vec{X}\}), i = 1, \dots, N\} \tag{1}$$

where, X_i is the data sample, $\vec{X}$ is the data vector, and $\tilde{\psi}$ is the influence function. The robustness of the L-filter is improved by means of the RM-estimate (1). The representation of the L-filter is,

$$\theta_L = \sum_{i=1}^{N} a_i X_{(i)} \quad \text{with} \quad a_i = \int_{i-1/N}^{i/N} h(\lambda)d\lambda \Big/ \int_0^1 h(\lambda)d\lambda \tag{2}$$

where, $X_{(i)}$, $i=1,\dots,n$ is the i-th order statistics of the data sample (ascending order), $a_i, i = 1, \dots, N$ are the weighted coefficients of filter, and $h(\lambda)$ is the noise probability distribution function $[0,1] \rightarrow R$ satisfying $\int_0^1 h(\lambda)d\lambda \neq 0$.

To introduce the RM-estimator [9] in the scheme of L-filter, the ordered data sample of L-filter should be presented as function of an influence function [9].

$$\theta_L = \sum_{i=1}^{N} a_i \psi(X_i) X_i \tag{3}$$
$$\psi(u) = \begin{cases} c & |u| \leq r \\ 0 & otherwise \end{cases}$$

where $N = (2L+1)^2$ is the filtering window size, $\psi(X_i)X_i$ is the ordered data sample, $\psi(u)$ is the influence function, c is a constant, and r is connected with the range of $\psi(u)$. Then, the RM L-filter can be obtained by merging the L-filter (3) and the RM-estimator (1),

$$\theta_{RML} = \frac{med\{a_i[X_i\psi(X_i - med\{\vec{X}\})]\}}{a_{med}} \tag{4}$$

where, $X_i\psi(X_i - med\{\vec{X}\})$ are the selected pixels in accordance with the influence function in the sliding filter window, the coefficients a_i are computed using the Laplacian, Uniform and Exponential distribution functions in $h(\lambda)$, and a_{med} is the median of coefficients a_i used as scale constant. Table 1 shows the influence function used in this paper.

Table 1. Influence function used in the RML-estimator

Influence function	Formulae
Hampel's three part redescending	$\psi_{\alpha,\beta(r)}(x) = \begin{cases} x & 0 \le \lvert x\rvert < \alpha \\ \alpha\, sgn(x) & \alpha \le \lvert x\rvert < \beta \\ \alpha \dfrac{r-\lvert x\rvert}{r-\beta} sgn(x) & \beta \le \lvert x\rvert < r \\ 0 & r \le \lvert x\rvert \end{cases}$

To improve the properties of impulsive noise suppression of the RML-estimator an impulse detector is used.

$$[(rank(x_{ij}) \le s) \vee (rank(x_{ij}) \ge N - s)] \wedge (|x_{ij} - med\{x_n\}| \ge U_2) \tag{5}$$

where, x_{ij} is the central pixel in the filtering window, $s > 0$ y $U_2 \ge 0$ are thresholds, N is the length of the data sample and $med\{x_n\}$ is the median of pixels into the filtering window.

2.1 Robust Fuzzy Clustering Algorithms

Fuzzy C-Means is a method for data classification, where each data belongs to a cluster to some degree, which is specified by a membership value. This algorithm performs the iteration of two indispensable conditions to minimize the following objective function.

$$J_f(U,V,X) = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^{m} \|x_i - v_j\|^2$$
$$subject\ to \sum_{i=1}^{N} u_{ij} > 0\ \forall_j\ and \sum_{j=1}^{c} u_{ij} = 1\quad \forall_i \tag{6}$$

where, $X = \{x_i | i = 1, \dots, N\}$ denotes the set of N feature vectors, c is the number of classes, $m \in [1, \infty)$ is a weighting exponent called the fuzzifier, $\|x_i - v_j\|^2$ is the square of the Euclidean distance from feature vector x_i to the center of the class v_j and $V = (v_1, \dots, v_j)$ is a vector with all center classes. $U = [u_{ij}]$ is a Nxc matrix denoting the constrained fuzzy c-partition. The value of u_{ij} denotes the degree of membership of x_i to the class v_j. To cover the noise sensitivity of the FCM algorithm, the RML-estimator is applied on the feature vector (intensity pixel) x_i. Based on the standard fuzzy c-means algorithm (6) and the RML-estimator (4). The objective function of the new algorithm can be written as:

$$J_{rf}(U,V,\Theta)=\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\theta_{RML\,i}-v_j\right\|^2 \tag{7}$$
$$subject\ to\sum_{i=1}^{N}u_{ij}>0\ \forall_j\ \ and\ \sum_{j=1}^{c}u_{ij}=1\ \ \forall_i$$

where, $\Theta=\{\theta_{RML\,i}|i=1,\dots,N\}$ is the RML-estimator applied on the intensity feature vector. Taking in account both constrains, the membership matrix and the cluster prototypes can be calculated with the following equations.

$$v_j=\frac{\sum_{i=1}^{N}u_{ij}^{m}\ \theta_{RML\,i}}{\sum_{i=1}^{N}u_{ij}^{m}} \tag{8}$$

$$u_{ij}=\frac{\left\|\theta_{RML\,i}-v_j\right\|^{-\frac{2}{m-1}}}{\sum_{l=1}^{c}\left\|\theta_{RML\,i}-v_l\right\|^{-\frac{2}{m-1}}} \tag{9}$$

A shortcoming of the FCM_S1 algorithm [3], [4] proposed by Chen and Zhang (10), is that the effect of the noisy image could be more than the mean-filtered image and the median-filtered image getting then an unsuitable segmentation.

$$J(U,V,X)=\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|x_i-v_j\right\|^2+\alpha_1\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\bar{x}_i-v_j\right\|^2+\alpha_2\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\tilde{x}_i-v_j\right\|^2$$
$$subject\ to\sum_{i=1}^{N}u_{ij}>0\ \forall_j\ and\sum_{j=1}^{c}u_{ij}=1\ \ \forall_i \tag{10}$$

where, $\bar{x}_i$ and $\tilde{x}_i$ are the mean and median of the neighboring pixels lying with a window around x_i, respectively. The parameters α_1 and α_2 control the effect of the mean and median of the neighboring pixels. So, our propose is to modify the objective function of this algorithm changing the x_i, $\bar{x}_i$ y $\tilde{x}_i$ terms by $\theta_{RML}-Uniform$, $\theta_{RML}-Exponential$, and $\theta_{RML}-Laplacian$ terms to get a more robust segmentation.

$$J_{rdf}(U,V,\Theta)=\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\theta_{RML-U\,i}-v_j\right\|^2+\alpha_1\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\theta_{RML-E\,i}-v_j\right\|^2$$
$$+\alpha_2\sum_{i=1}^{N}\sum_{j=1}^{c}u_{ij}^{m}\left\|\theta_{RML-L\,i}-v_j\right\|^2 \tag{11}$$
$$subject\ to\sum_{i=1}^{N}u_{ij}>0\ \forall_j\ \ and\ \sum_{j=1}^{c}u_{ij}=1\ \ \forall_i$$

where, θ_{RML-U}, θ_{RML-E} and θ_{RML-L} are the θ_{RML} estimators (4) computed using the Uniform, Exponential and Laplacian distributions respectively, α_1 and α_2 are parameters that control the effect of , θ_{RML-E} and θ_{RML-L} terms. By minimizing (11), the membership matrix and the cluster prototypes can be stated as.

$$v_j = \frac{\sum_{i=1}^{N} u_{ij}^m \ [\theta_{RML-U\,i} + \alpha_1 \theta_{RML-E\,i} + \alpha_2 \theta_{RML-L\,i}]}{(1+\alpha_1+\alpha_2)\sum_{i=1}^{N} u_{ij}^m} \tag{12}$$

$$u_{ij} = \frac{\left[\left\|\theta_{RML-U\,i} - v_j\right\|^2 + \alpha_1 \left\|\theta_{RML-E\,i} - v_j\right\|^2 + \alpha_2 \left\|\theta_{RML-L\,i} - v_j\right\|^2\right]^{-\frac{1}{m-1}}}{\sum_{l=1}^{c}[\|\theta_{RML-U\,i} - v_l\|^2 + \alpha_1\|\theta_{RML-E\,i} - v_l\|^2 + \alpha_2\|\theta_{RML-L\,i} - v_l\|^2]^{-\frac{1}{m-1}}} \tag{13}$$

3 Experimental Results

The performance of the proposed method was tested on synthetic and real images. In both cases the quantitative results were compared using the optimal segmentation accuracy (SA). SA was measured as the total number of correctly classified pixels divided by the total number of pixels (see Ahmed et al., 2002) [4].

$$SA = \frac{Total\ number\ of\ correctly\ classified\ pixels}{total\ numer\ of\ pixels}\ x100\% \tag{14}$$

3.1 Evaluation on a Synthetic Image

A synthetic image (8-bit pixels) containing 128x128 pixels is depicted in Fig.1(a). This one has *three clusters* with gray values of 85, 170, and 255 respectively. The original image was corrupted by different levels of Gaussian noise, Salt & Pepper noise and a mix of these ones. The results are compared with the FCM_S1, FCM_S2, Nyström, NCut, and NL_SSC algorithms [3], [4], [10]. For both proposed algorithms (7) and (11) the parameters used were stated as: c=3, m=2, ε=1e-6 in the clustering (initialized randomly); r=5, α=0.16r, β=0.8r in the RML-estimation, s=4 and U_2=5 in the impulse detection, α_1=0.02 and α_2=0.05 in the RDFCM algorithm (11). Table 2 shows the SAs values for all algorithms and Fig.2 depicts the visual results in the case of mixed noise.

Table 2. SA % on the synthetic image

Algorithm	Gaussian			Salt & Pepper			Mixed		
	σ=10	σ=20	σ=40	5%	10%	15%	σ=10 + 5%	σ=20 + 10%	σ=40 + 15%
FCM_S1	97.95	97.91	93.59	96.39	95.13	92.81	96.33	95.10	90.30
FCM_S2	97.98	97.95	93.48	96.39	95.13	92.83	96.34	95.25	91.30
Nyström	76.72	74.98	71.68	96.82	94.56	92.13	76.87	75.62	72.93
NCut	99.69	99.62	98.84	94.81	94.90	94.11	95.58	94.31	92.84
NL_SSC	99.80	99.71	99.47	98.42	98.01	97.26	98.71	98.09	97.37
RFCM	**99.95**	**99.85**	**99.66**	**99.96**	**99.92**	**99.89**	**99.87**	**99.83**	**99.49**
RDFCM	**99.93**	**99.85**	**99.61**	**99.97**	**99.92**	**99.90**	**99.85**	**99.80**	**98.46**

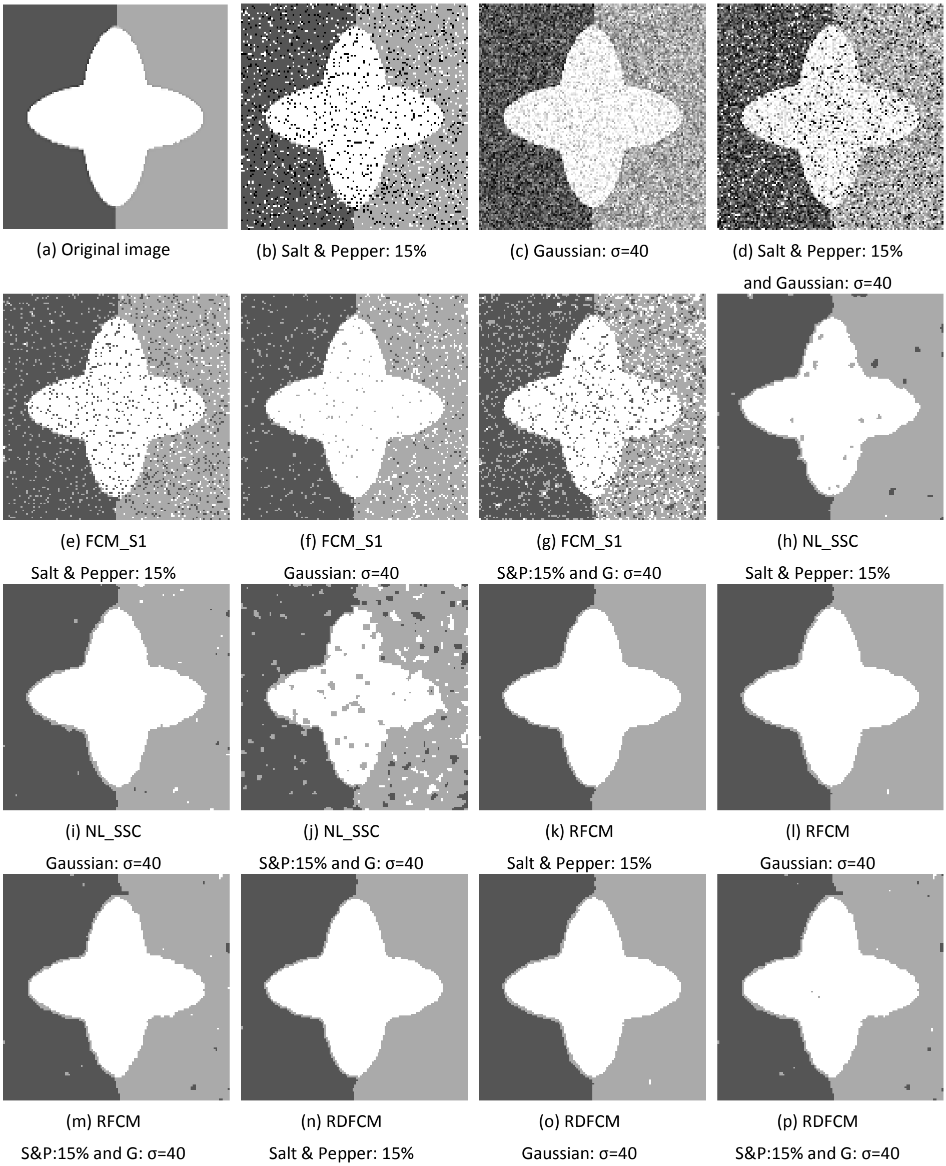


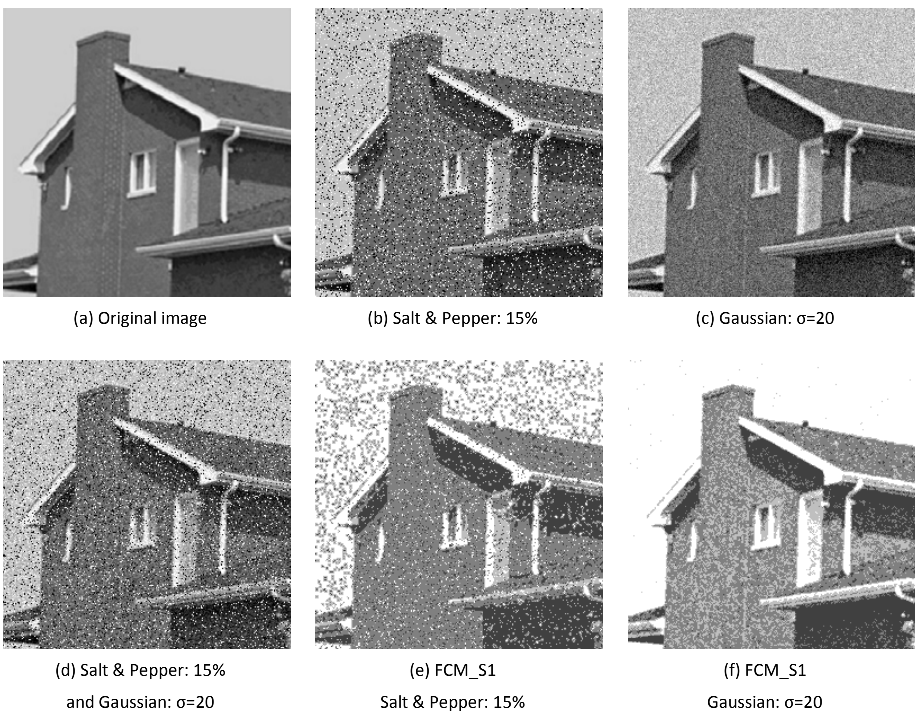
Fig. 1. Segmentation results in a synthetic image

3.2 Evaluation on a Real Image

The robustness on a real image (Fig.2(a)) with a format of 8-bit pixels was tested using the *home* real image with 256x256 pixels size corrupted by different levels of Gaussian noise, Salt & Pepper noise and a mix of these ones. This image was tested with *four clusters*. The results are compared with FCM_S1, FCM_S2, Nyström, NCut, and NL_SSC algorithms [3], [4], [10], [11]. For proposed algorithms the parameters used were stated as: c=4, m=2, ε=1e-6 in the clustering (initialized randomly); r=5, α=0.16r, β=0.8r in the RML-estimation, s=4 and U_2=5 in the impulse detection, α_1=0.8 and α_2=0.4 in the RDFCM algorithm. Table 3 shows the calculated SAs for all algorithms and Fig. 2 depicts the visual results.

Table 3. SA % on the real image

Algorithm	Gaussian			Salt & Pepper			Mixed		
	σ=5	σ=10	σ=20	5%	10%	15%	σ=5 + 5%	σ=10 + 10%	σ=20 + 15%
FCM_S1	97.51	97.56	70.98	96.72	82.22	71.17	96.63	77.15	66.11
FCM_S2	97.52	97.50	70.61	96.78	78.17	70.35	97.14	69.87	62.74
Nyström	80.27	76.39	70.28	93.66	91.76	88.43	77.28	73.69	68.75
NCut	97.44	95.83	92.48	91.36	88.90	84.33	92.65	89.95	86.75
NL_SSC	98.85	97.62	96.17	99.28	98.64	98.06	98.97	97.80	96.82
RFCM	**99.83**	**99.76**	**98.72**	**99.92**	**99.90**	**99.81**	**99.88**	**99.75**	**98.60**
RDFCM	**99.77**	**99.72**	**98.91**	**99.61**	**99.55**	**99.62**	**99.84**	**99.78**	**98.96**



(a) Original image

(b) Salt & Pepper: 15%

(c) Gaussian: σ=20

(d) Salt & Pepper: 15% and Gaussian: σ=20

(e) FCM_S1 Salt & Pepper: 15%

(f) FCM_S1 Gaussian: σ=20

Fig. 2. Segmentation results in a real image

(g) FCM_S1
S&P:15% and G: σ=20

(h) NL_SSC
Salt & Pepper: 15%

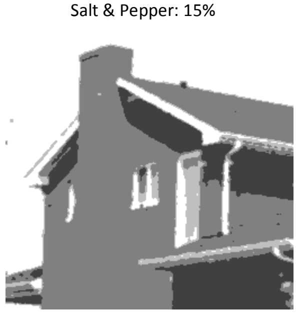
(i) NL_SSC
Gaussian: σ=20

(j) NL_SSC
S&P:15% and G: σ=20

(k) RFCM
Salt & Pepper: 15%

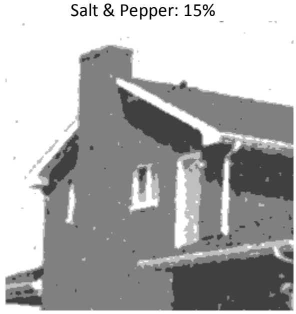
(l) RFCM
Gaussian: σ=20

(m) RFCM
S&P:15% and G: σ=20



(n) RDFCM
Salt & Pepper: 15%

(o) RDFCM
Gaussian: σ=20



(m) RDFCM
S&P:15% and G: σ=20

Fig. 2. (*continued*)

3.3 Evaluation with Berkeley Image Segmentation Data Set 500 (BSDS500)

In this experiment, we evaluate the image segmentation performance of the proposed algorithms and the results are compared with the FCM_S1 and NL_SSC algorithms, using a subset of the Berkeley image segmentation dataset. The images were corrupted by Salt & Pepper, Gaussian and a mix of them. For proposed algorithms the parameters used were stated as: m=2, ε=1e-6, c=3 for all images (except 42049-image, where c=2) in the clustering (initialized randomly); r=5, α=0.16r, β=0.8r in the RML-estimation, s=4 and U_2=5 in the impulse detection, α_1=0.8 and α_2=0.4 in the RDFCM algorithm. Table 4 shows the calculated SAs and Figs. 3, 4, 5, 6, and 7 depict the visual results.

Table 4. SA % on the real images

Algorithm	Noise	Image				
		42049	35010	24063	8068	302003
FCM_S1	Salt & Pepper: 15%	92.76	97.92	86.29	92.51	95.94
	Gaussian: σ=20	97.95	98.07	98.97	97.76	98.17
	Mixed: 15% + : σ=20	90.77	96.49	84.97	89.99	93.06
NL_SSC	Salt & Pepper: 15%	99.05	98.04	98.94	98.38	97.84
	Gaussian: σ=20	98.02	98.75	98.15	98.13	97.29
	Mixed: 15% + : σ=20	98.21	97.82	97.96	98.04	96.33
RFCM	Salt & Pepper: 15%	**99.95**	**99.40**	**99.97**	**99.95**	**99.52**
	Gaussian: σ=20	**99.92**	**99.48**	**99.77**	**99.42**	**97.98**
	Mixed: 15% + : σ=20	**98.30**	**98.75**	**99.05**	**99.58**	**97.48**
RDFCM	Salt & Pepper: 15%	**99.68**	**98.58**	**99.65**	**99.60**	**98.23**
	Gaussian: σ=20	**99.91**	**98.39**	**99.78**	**99.53**	**97.86**
	Mixed: 15% + : σ=20	**98.34**	**98.11**	**99.17**	**99.50**	**97.20**

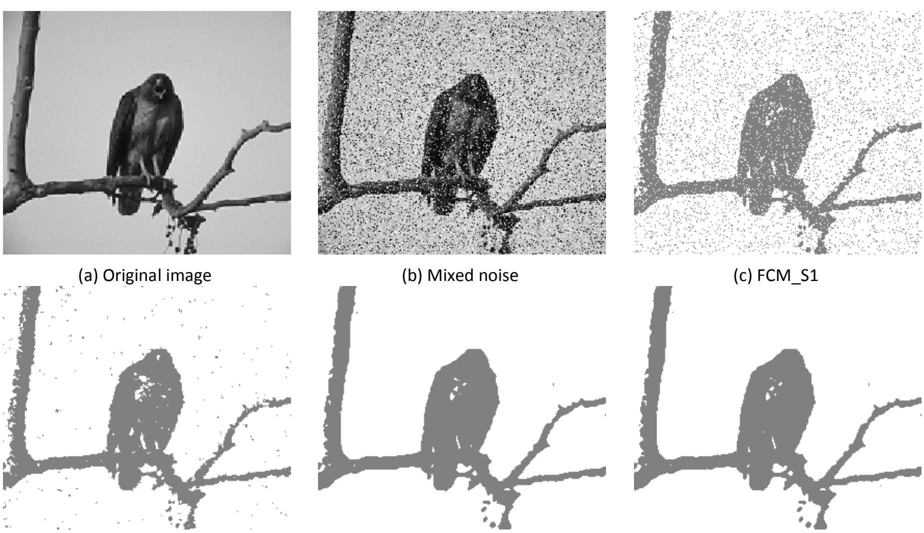


Fig. 3. Image 42049 with c=2

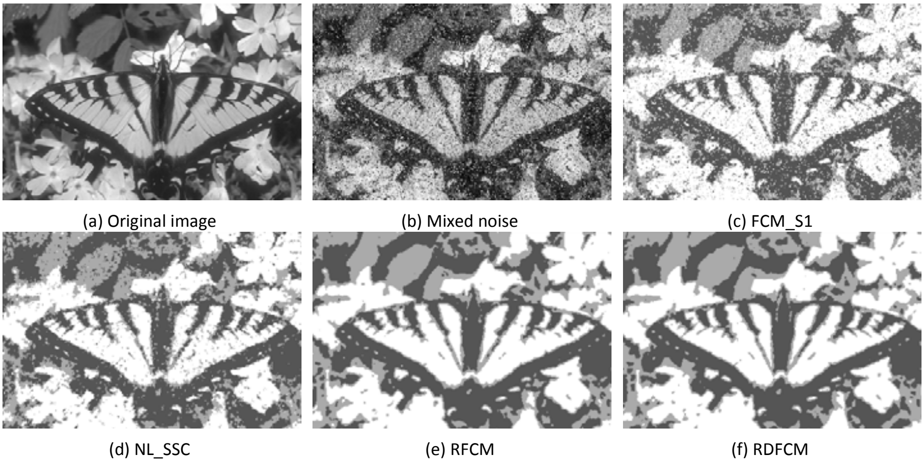


Fig. 4. Image 35010 with *c=3*

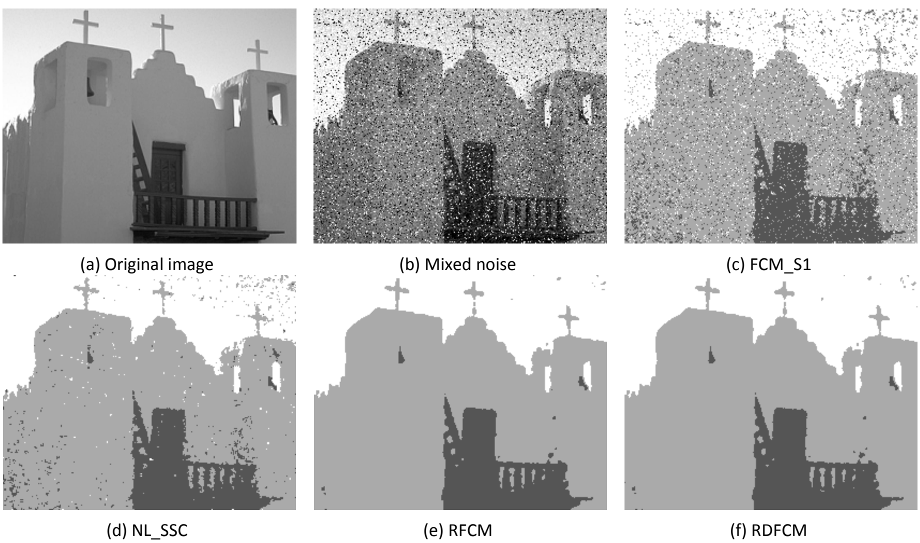


Fig. 5. Image 24063 with *c=3*

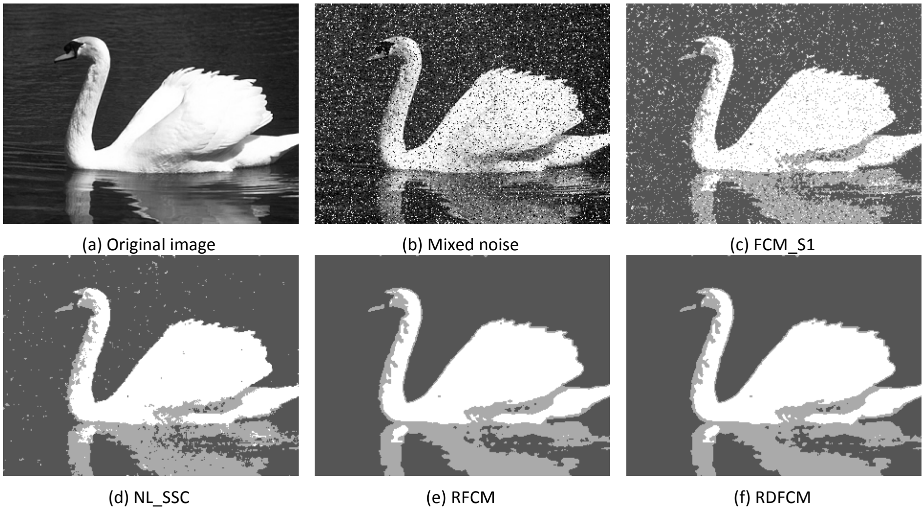


Fig. 6. Image 8068 with $c=3$

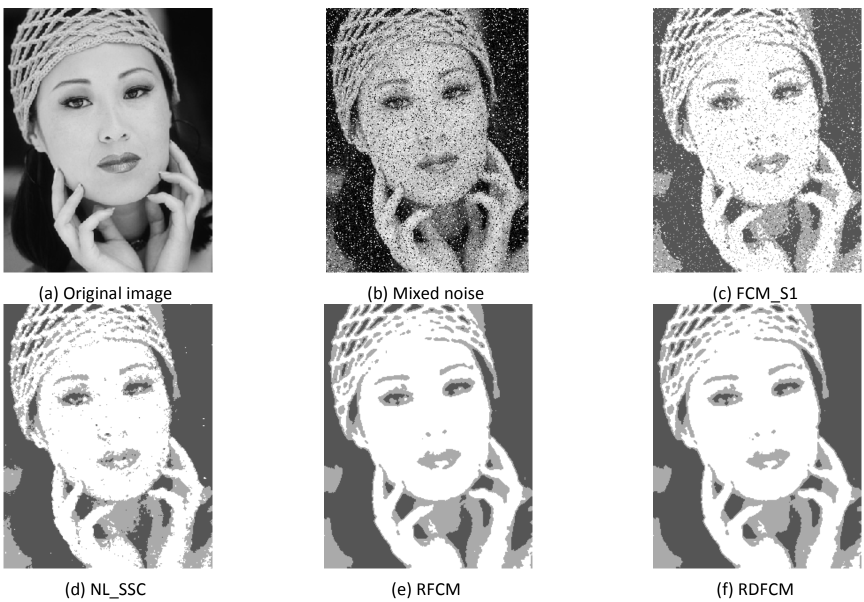


Fig. 7. Image 302003 with $c=3$

4 Discussion of Results

The experimental results on both test images shown that the proposed algorithms obtained better results than the comparative algorithms. Also, was observed that the incorporation of the RML-estimator into de FCM_S1 algorithm gave it greater ability so segment noisy images in comparison of its original version. The SA% values obtained by our algorithms in both experiments is close to 100%, so, this is appreciated in better conservation of the details in the images after segmentation.

As a future work another indexes will be evaluate to measure the segmentation quality, some of these are the Probabilistic Rank Index (PRI), Variation of Information (VOI), Global Consistency Error (GCE) and the Boundary Displacement Error (BDE) [12].

5 Conclusions

This paper presented a method for incorporating the RML-estimator into the Fuzzy C-means and its variant FCM_S1, making the more robust to segment noisy images. The performance of the proposed methods was better than the reference algorithms. However, to give them greater ability to segment color images as well as other types and levels of noise should be modified and optimized.

Acknowledgments. This work is supported by National Polytechnic Institute of Mexico and Conacyt.

References

1. Yang, F., Jiang, T.: Pixon-based Image Segmentation with Markov random fields. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China (2002)
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
3. Wang, Z., Lu, R.: A New Algorithm for Image Segmentation Based on Fast Fuzzy C-Means Clustering. School of Electronics and Information Engineering, Southwest University, Chongqing, China (2008)
4. Cai, W.L., Chen, S.C., Zhang, D.Q.: Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Department of Computer Science & Engineering. Nanjing University of Aeronautics & Astronautics, Nanjing, China (2007)
5. Gallegos-Funes, F., Varela-Benitez, J., Ponomaryov, V.: Rank M-Type L (RM L)-Filter for Image Denoising. IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences E91-A(12), 3817–3819 (2008)
6. Gallegos-Funes, F.J., Ponomaryov, V.: Real-time image filtering scheme based on robust estimators in presence de noise impulsive. Real Time Imaging 8(2), 78–90 (2004)
7. Gallegos-Funes, F.J., Varela-Benitez, J.L., Ponomaryov, V.: Real-time image processing based on robust linear combinations of order statistics. In: Proc. SPIE Real-Time Image Processing 2006, vol. 6063, pp. 177–187 (2006)
8. Varela-Benitez, J.L., Gallegos-Funes, F.J., Ponomaryov, V.: RML-filters for real rime imaging. In: Proc. IEEE 15th International Conference on Computing, CIC 2006, pp. 43–48 (2006)

9. Gallegos-Funes, F., Ponomaryov, V., De-La-Rosa, J.: ABST M-type K-nearest neighbor (ABSTM-KNN) for image denoising. IEICE Trans. Funds. Electronics Comms. Computer Science E88A(3), 798–799 (2005)
10. Liu, H.Q., Jiao, L.C., Zhao, F.: Non-local spatial spectral clustering for image segmentation. Neurocomputing 74, 461–471 (2010)
11. Liu, H.Q., Jiao, L.C., Zhao, F.: A novel fuzzy clustering with non-local adaptive spatial constraint for image segmentation. Signal Processing 91, 988–999 (2011)
12. Le Capitaine, H., Frélicot, C.: A fast fuzzy c-means algorithm for color image segmentation. Laboratoire Mathémathiques, Image et Applications, Université de La Rochelle, France (2011)