

Dempster-Shafer's Basic Probability Assignment Based on Fuzzy Membership Functions

Abdel-Ouahab Boudraa*[†], Ayachi Bentabet^{‡†} Fabien Salzenstein⁺ and Laurent Guillon*

* *IRENav, Ecole Navale, 29200 Brest—Armées, France.*

[‡] *Department of Mathematics and Computers Sciences, Sherbrooke University, Canada.*

⁺ *Laboratoire Phase, ULP Strasbourg I, 67000 Strasbourg, France.*

[†] *CREATIS, CNRS UMR 5515, INSA 502, 69621 Villeurbanne, France.*

Received 14 May 2003; revised 14 Mars 2004; accepted 16 April 2004

Abstract

In this paper, an image segmentation method based on Dempster-Shafer evidence theory is proposed. *Basic probability assignment (BPA)* is estimated in unsupervised way using pixels fuzzy membership degrees derived from image histogram. No assumption is made about the images data distribution. *BPA* is estimated at pixel level. The effectiveness of the method is demonstrated on synthetic and real images.

Key Words: Data fusion, Basic probability assignment, Image segmentation, Dempster-Shafer evidence theory, Fuzzy membership functions.

1 Introduction

Multisensor data fusion is an evolving technology that is analogous to the ongoing cognitive process used by human to integrate data from their senses continually and make inferences about the external world [1]. The information provided by one sensor is usually limited and sometimes of low accuracy. The use of multiple sensors is an alternative to improve accuracy and provide the user with additional information of increased reliability about the environment in which the sensors operates. Applications of data fusion range from medical imaging, scene analysis, Robotics, non destructive evaluation, target tracking to airborne surveillance. Data fusion can be done at different levels of representation: signal, pixel, feature and symbolic levels. In this work we address the problem of pixel-level fusion. Different strategies have been developed for data fusion. The frameworks used for data management are Bayesian inference, Dempster-Shafer (DS) theory [2],[3] and fuzzy logic inference. DS theory makes inferences from incomplete and uncertain knowledge, provided by different independent knowledge sources. A first advantage of DS theory is its ability to deal with ignorance and missing information. In particular, it provides explicit estimation of imprecision and conflict between information from different sources and can deal with any unions of hypotheses (clusters) [4]. This is particularly useful to represent "mixed" pixels in image segmentation problems. The main limitation of Bayesian inference is that it cannot model imprecision about uncertainty measurement. The degree of belief we have on a union of

Correspondence to: boudra@ecole-navale.fr

Recommended for acceptance by Enric Marti

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

clusters (without being able to discriminate between them) should be shared by all the simple hypotheses, thus penalizing the good one. DS theory handles uncertain and incomplete information through the definition of two dual non additive measures: plausibility and belief. These measures are derived from a density function, m , called *basic probability assignment (BPA)* or *mass function*. This probability assigns evidence to a proposition (hypothesis). The derivation of the *BPA* is the most crucial step since it represents the knowledge about the application as well as the uncertainty incorporated in the selected information source. *BPA* definition remains a difficult problem to apply DS theory to practical applications such in image processing. For example, *BPA* may be derived, at pixel level, from probabilities [5]-[7] or from the distance to cluster centers [8]. In this work *BPA* is estimated in an unsupervised way and using fuzzy membership functions to take into account the ambiguity within pixels. This ambiguity is due to the possible multi-valued levels of brightness in the image. This indeterminacy is due to inherent vagueness rather than randomness. The number of the clusters of the image is supposed known. In [7] the *BPA* estimation is based on the assumption that the probability distribution of the gray level values (image histogram) is Gaussian model. Our estimation approach does not make any assumption about the probability distribution of the gray level histogram and is not limited to only two sources.

2 Dempster-Shafer theory

In DS theory, there is a fixed set of q mutually exclusive and exhaustive elements, called the frame of discernment, which is symbolized by:

$$\Theta = \{H_1, H_2, \dots, H_q\}$$

The representation scheme, Θ , defines the working space for the desired application since it consists of all propositions for which the information sources can provide evidence. Information sources can distribute mass values on subsets of the frame of discernment, $A_i \in 2^\Theta$ (1). An information source assigns mass values only to those hypotheses, for which it has direct evidence.

$$0 \leq m(A_i) \leq 1 \quad (1)$$

BPA has to fulfill the conditions: $m(\emptyset) = 0$ and $\sum_{A_i \in 2^\Theta} m(A_i) = 1$. If an information source can not distinguish between two propositions, H_i and H_j , it assigns a mass value to their union ($H_i \cup H_j$). Mass distribution from different information sources, $m_j (j = 1, \dots, d)$, are combined with Dempster's orthogonal rule (2). The result is a new distribution, $m(A_k) = (m_1 \oplus m_2 \oplus \dots \oplus m_d)(A_k)$, which incorporates the joint information provided by the sources.

$$m(A_k) = (1 - K)^{-1} \times \sum_{A_1 \cap A_2 \dots A_d = A_k} \left(\prod_{1 \leq j \leq d} m_j(A_j) \right) \quad (2)$$

$$K = \sum_{A_1 \cap A_2 \dots A_d = \emptyset} \left(\prod_{1 \leq j \leq d} m_j(A_j) \right) \quad (3)$$

K is often interpreted as a measure of conflict between the different sources (3) and is introduced as a normalization factor (2). The larger K is the more the sources are conflicting and the less sense has their combination. The factor K indicates the amount of evidential conflict. If $K = 0$, this shows complete compatibility, and if $0 < K < 1$, it shows partial compatibility. Finally, the orthogonal sum does not exist when $K = 1$. In this case, the sources are totally contradictory, and it is no longer possible to combine them. In the cases of sources highly conflicting, the normalization used in the Dempster combination rule can be mis-taking, since it artificially increases the masses of the compromise hypotheses [9]. One may suggest as in [9] that the conflict come from the "true" assumption has been forgotten (in the set of hypotheses). However, this cannot occur under closed-world assumption, which is our case, and thus the high conflict level is rather due to the fact that one of

the sources is erroneous. In such case, conflict problems should not occur provided that the source information modelling was correctly done, in particular including, when necessary, an ignorance or error term (by affecting non null masses to compound hypotheses and Θ) [10]. Finally, we find a normalization process is necessary to satisfy the relations $m(\emptyset) = 0$ and $\sum_{A_i \in 2^\Theta} m(A_i) = 1$ and to preserve the associative properties of the combination rule. From a mass distribution, numerical values can be calculated that characterize the uncertainty and the support of certain hypotheses. Belief (4) measures the minimum or necessary support whereas plausibility (5) reflects the maximum or potential support for that hypothesis. These two measures, derived from mass values, are respectively defined from 2^Θ to $[0, 1]$:

$$Bel(A_i) = \sum_{A_j \subseteq A_i} m(A_j) \quad (4)$$

$$Pls(A_i) = \sum_{A_j \cap A_i \neq \emptyset} m(A_j) \quad (5)$$

The equations (4) and (5) imply that $Bel(\cdot)$ and $Pls(\cdot)$ are dual measures related by

$$Pls(A_i) = 1 - Bel(\neg A_i) \quad (6)$$

3 Fuzzy approach

Modeling real problems typically involves processing uncertainty of three types. Uncertainty of probabilistic nature, uncertainty due to the lack of specification and fuzziness. Traditionally nonfuzzy uncertainties are handled by probabilistic methods such as Bayesian networks and DS theory while fuzziness uncertainty is modeled by fuzzy set theory. Fuzzy uncertainty deals with situations where boundaries of the sets (clusters) under consideration are not sharply defined (partial occurrence of an event). On the other hand, for nonfuzzy uncertainties there is no ambiguity about set boundaries, but rather, about the belongingness of elements or events to crisp sets. Real data are often imprecise and contain some ambiguity caused by the way they have been obtained. Origins of this kind of ambiguity may be inaccuracy of the used devices involving an error of measurement of fuzzy nature. In image processing, images which are mappings of natural scenes are always accompanied by an amount of fuzziness due to imprecision of gray values and ambiguity created by the mapping mechanism. There are many situations where we often face at the same time fuzzy and nonfuzzy uncertainties. This suggests to combine DS and fuzzy sets frameworks. Thus, the goal of this work is to estimate *BPA*s using fuzzy membership functions which capture vagueness.

Let $X = \{x_1, x_2, \dots, x_{MN}\}$ be an image set of size $M \times N$ with L levels, $g = 0, 1, 2, \dots, L - 1$, and x_{mn} is the gray level of a (m, n) th pixel in X . Let $\mu(X) = \{\mu(x_1), \mu(x_2), \dots, \mu(x_{MN})\}$ be the corresponding fuzzy membership degrees derived from X . $\mu(\cdot)$ is obtained by operating a fuzzifier on X . This fuzzifier performs a mapping from crisp data values (X) into a fuzzy set represented by $\mu(X)$. We denote by $\mu_i(x_{mn})$ the fuzzy membership degree of pixel x_{mn} to fuzzy subset (cluster) i of X . In this work, Fuzzy C-means (FCM) algorithm [11] is used as fuzzifier. FCM has an advantage of clustering data without the need for a statistical model of the data. For image fuzzification we use an histogram based gray-level fuzzification [12]. Thus, we use gray level g instead of the intensity of (m, n) th pixel x_{mn} (Figure 1). The FCM only operates on the histogram and consequently is faster than the conventional version [11], which processes the whole data set X .

4 Basic probability assignment

In image segmentation problem, Θ is the set of all the clusters of the image, $|\Theta| = C$ is the number of clusters and 2^C contains all the possible unions of clusters. The hypotheses considered in DS formulation are: \emptyset (whose

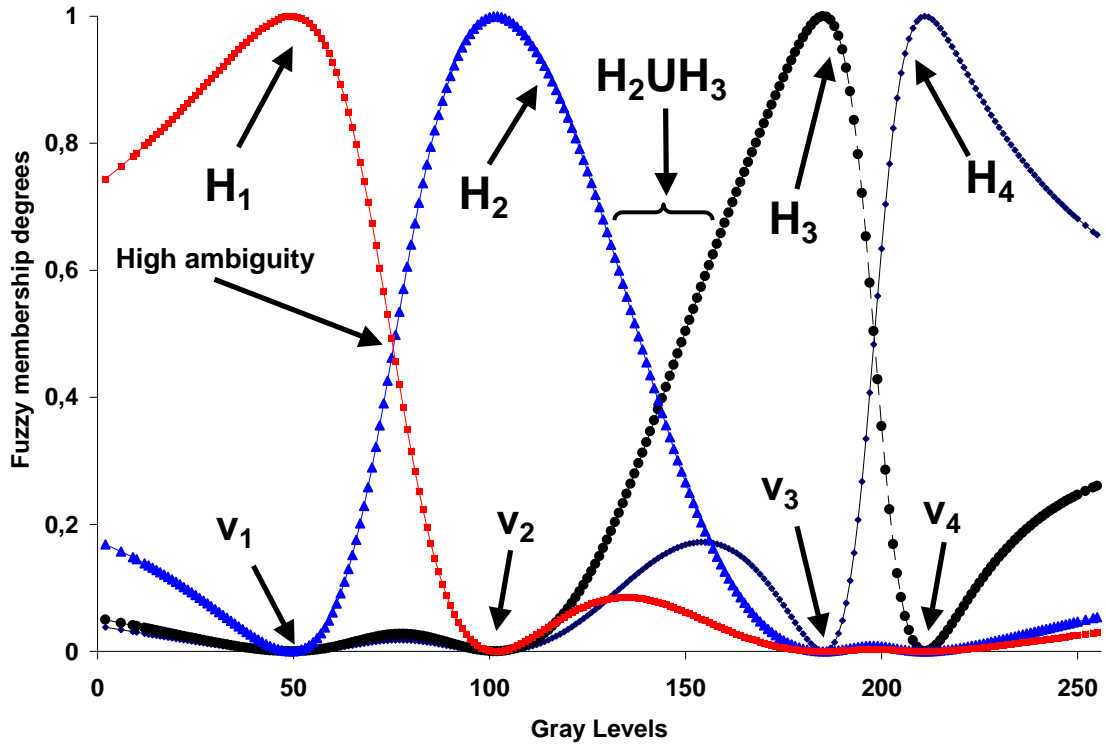


Figure 1: Plot of fuzzy membership functions generated by FCM algorithm (RX image). V_i stands for the centroid of the i^{th} cluster.

mass is null), simple hypothesis H_i and compound hypotheses $H_j \cup \dots \cup H_l$. For the choice of the BPA of H_k and H_l , the following strategy is used :

1. Affecting a non null mass to $H_k \cup H_l$ if H_k and H_l are not discriminated on the image (not distinguishable by the sensor) (Figure 1). There is an ambiguity between H_k and H_l . In this case affecting a pixel with gray level g to cluster k or l using of fuzzy membership rule is not valuable ($\mu_k(g) \approx \mu_l(g)$).
2. Affecting a null mass to $H_k \cup H_l$ if H_k and H_l are discriminated on the image. There is no or less ignorance about clusters k and l .

In performing a fuzzy clustering on image histogram [12], the intersection between two fuzzy membership degree curves $\mu_k(g)$ and $\mu_l(g)$ to two consecutive centroids V_k and V_l (Figure 1), occurs in one and only one point. This point corresponds to a high degree of ambiguity and then to maximum value of $m(H_k \cup H_l)$. For example, at pixel with gray level $g = 139$ information source can not distinguish between between clusters H_2 and H_3 , $m(H_2 \cup H_3)(g = 139) \neq 0$, while at pixel with gray level $g = 50$ there is no ambiguity to affect $g = 50$ to cluster H_1 and thus $m(H_2 \cup H_3)(g = 50) = 0$, (Figure 1). BPAs are normalized such that $\sum_{H_i \in 2^C} m(H_i) = 1$. Using image histogram, for each level g , and according the C values, different cases are

distinguished. For more convenience, we use the following notations:

$$\beta = \max_{1 \leq i \leq C} (\mu_i(g)) \quad (7)$$

$$\alpha = \beta - \min_{1 \leq i \leq C} (\mu_i(g)) \quad (8)$$

$$I = \{1, \dots, C\}$$

where

$$\sum_{i=1}^C \mu_i(g) = 1$$

I is the set of cluster indices and its cardinal is the number of clusters, C . $\arg(\beta) = \arg(\max_{1 \leq i \leq C} \mu_i(g))$ is the maximum fuzzy membership defuzzification rule. The pixel with gray level g is affected to cluster $\arg(\beta)$.

In the proposed fusion scheme for all C values, both simples and compound hypotheses are taken into account. In the framework of histogram based segmentation and for $C \geq 3$, ambiguity can not occur between all the C classes. Thus, a null mass is affected to the union of hypotheses ((12),(16),(22),(25)). For $C = 2$, in general there is at least one pixel where the two classes (hypotheses) are not sufficiently distinguishable from each other so that a new compound hypotheses is created with a non null mass (9). However, if the two hypotheses are well distinguishable from each other, the mass value of their union is null (20). For all C values, and for all cases (with less or high ambiguity), the mass value affected to single hypothesis proportional to the corresponding fuzzy membership degree ((10),(13),(17),(18),(19),(23),(26)). The mass value affected to compound hypotheses is proportional to the sum of their fuzzy membership degrees ((11),(14),(15),(21),(24)). In each case, the normalization condition must be verified.

- If there is high ambiguity to affect a pixel with gray level g to cluster k or l : $|\mu_k(g) - \mu_l(g)| \leq \xi$ then

1. For $C = 2$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = \alpha \quad (9)$$

$$m(H_i)(g) = [1 - \alpha] \times \mu_i(g) \quad i \in I \quad (10)$$

2. For $C = 3$

$$m(H_k \cup H_l)(g) = \alpha \times [\mu_k(g) + \mu_l(g)] \quad (k, l)_{k \neq l} \in I \quad (11)$$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = 0 \quad (12)$$

$$m(H_i)(g) = [1 - m(H_k \cup H_l)(g)] \times \mu_i(g) \quad i, (k, l)_{k \neq l} \in I \quad (13)$$

3. For $C > 3$

$$m(H_k \cup H_l)(g) = \alpha \times [\mu_k(g) + \mu_l(g)] \quad (k, l)_{k \neq l} \in I \quad (14)$$

$$m\left(\bigcup_{\substack{i=1 \\ i \neq k, i \neq l}}^C H_i\right)(g) = \alpha \times \sum_{\substack{i=1 \\ i \neq k, i \neq l}}^C \mu_i(g) \quad (15)$$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = 0 \quad (16)$$

$$m(H_t)(g) = [1 - m\left(\bigcup_{\substack{i=1 \\ i \neq k, i \neq l}}^C H_i\right)(g) - m(H_k \cup H_l)(g)] \times \mu_t(g) \quad t, (k, l)_{k \neq l} \in I \quad (17)$$

- If there is less or no ambiguity to affect a pixel with gray level g to cluster k : $|\mu_k(g) - \mu_l(g)| > \xi$ then

1. For $C = 2$

$$m(H_l)(g) = \mu_l(g) \times (\mu_k(g) - \mu_l(g)) \quad (k, l)_{k \neq l} \in I \quad (18)$$

$$m(H_k)(g) = 1 - \mu_l(g) \times (\mu_k(g) - \mu_l(g)) \quad (k, l)_{k \neq l} \in I \quad (19)$$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = 0 \quad (20)$$

2. For $C = 3$

$$m(H_k \cup H_l)(g) = \alpha \times [\mu_k(g) + \mu_l(g)] \quad (k, l)_{k \neq l} \in I \quad (21)$$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = 0 \quad (22)$$

$$m(H_i)(g) = [1 - m(H_k \cup H_l)(g)] \times \mu_i(g) \quad i, (k, l)_{k \neq l} \in I \quad (23)$$

3. For $C > 3$

$$m\left(\bigcup_{\substack{i=1 \\ i \neq k}}^C H_i\right)(g) = \sum_{\substack{i=1 \\ i \neq k}}^C \mu_i(g) \times (\beta - \mu_i(g)) \quad (24)$$

$$m\left(\bigcup_{i=1}^C H_i\right)(g) = 0 \quad (25)$$

$$m(H_t)(g) = \left(1 - m\left(\bigcup_{\substack{i=1 \\ i \neq k}}^C H_i\right)(g)\right) \times \mu_t(g) \quad t \in I \quad (26)$$

ξ is a threshold value. We make assumption that the images are well registered. Since, images are clustered separately then a spatial correspondence between the labels of clusters of different images is necessary so that pixels representing the same physical object of the scene may be superimposed and thus to be able to correctly combine the different information sources (2). The label-to-label mapping strategy is described in [13]. The use of image histogram loose spatial information about pixels arrangement and the spatial correlation between adjacent pixels. Furthermore, the membership degrees resulted from the FCM algorithm are considerably troublesome in a very noisy environment. To reduce noise effect and to improve the classification results contextual processing is performed. Thus, before *BPA*s estimation, membership value of each pixel is updated by using its neighborhood contextual membership values. In this work, a 3×3 neighborhood mean and median filters are used [13].

5 Results

The proposed data fusion method is first tested on synthetic images. Two images, corrupted by Gaussian noise, simulating US and RX acquisitions are shown in Figure 2. Each image contains four clusters ($C=4$). In the US image (Fig. 2(a)), one region (smallest thickness) is confused with the background and in the RX image (Fig. 2(b)) the greatest thickness is under-exposed and the thicker regions are not well distinguished. The aim here is to exploit, through using the proposed data fusion technique, the redundant and complementary information of the two images in order to correctly segment the image in four clusters. The maximum of plausibility is used as a decision rule. Figures 2(e) and 2(f) show the DS fusion result obtained using median and average

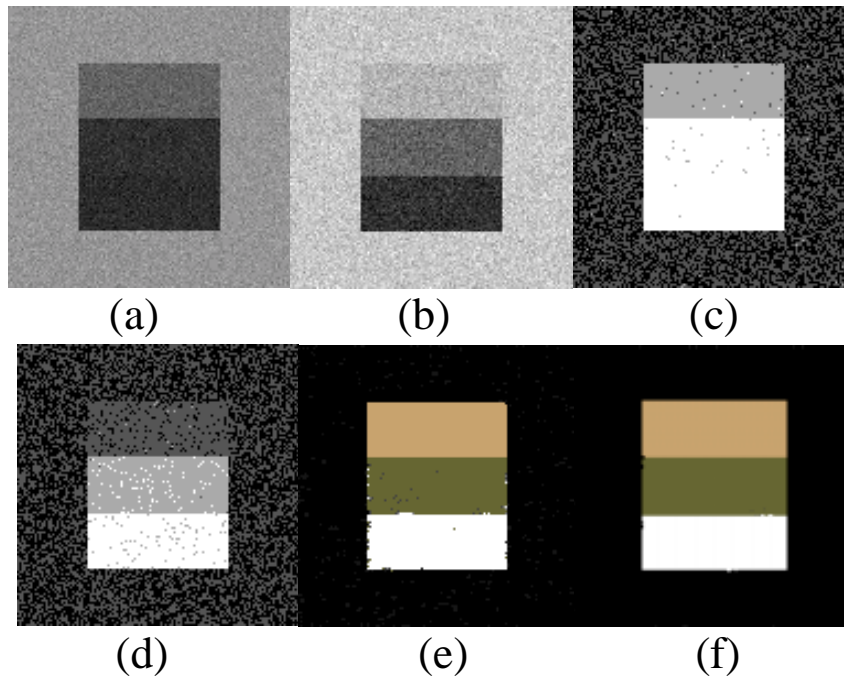


Figure 2: (a) US image. (b) RX image. (c) Fuzzy segmentation of US image. (d) Fuzzy segmentation of RX image. Fused image obtained using median (e) and averaging filters (f).

filters respectively. ξ is set 0.05. Note that within the segmented regions, some artefacts are present (Figs. 2(e)-(f)), reflecting the influence of noise present in the initial images (Figs. 2(a)-(b)) on final segmentation. Both filters give a good results but the regions given by the averaging filter are more homogeneous than in the median case. The four regions are well brought out and this shows that informations provided by two images are well exploited by the fusion scheme. This result also shows that the estimated *BPA*s are a good modeling of the information associated to simple and compound hypotheses. This also shows the interest of taking into account the contextual information in *BPA*s estimation. The proposed fusion based segmentation is compared to a conventional segmentation method where only information derived from one image is exploited. The segmentation results in Figs 2(c) and 2(d) have been obtained using fuzzy clustering. These results (Figs. 2(c)-(d)) show that 23,94 % and 34,94% of pixels have been mis-segmented for RX and US images respectively. Segmentation errors are largely reduced when using simultaneously the two images through the use of DS fusion scheme including spatial information. Indeed, in the latter case, only 0,95 % of pixels have been mis-segmented. This good performance difference between these two types of segmentation approaches can also be easily assessed by visually comparing the segmentation results.

Figure 3 illustrates the application of the proposed fusion scheme to human brain Magnetic Resonance (MR) of three patients with Multiple Sclerosis (MS) lesions (Figures 3(a)-(f)). Figures 3(a)-(c) represent T_2 -weighted images and Figures 3(d)-(f) the corresponding Proton Density (PD) weighted images. Each pair of images (T_2 ,PD) are strongly correlated and also spatially registered, and show the MS lesions as hypersignal regions. The fused images are shown in Figures 3(g)-(i). In each patient, regions such as white matter, grey matter, cerebrospinal fluid (CSF), background are correctly segmented. This is of great interest in medical applications in particular the estimation of size and volume of the brain tissues. However, the proposed scheme is not able to separate the MS lesions region from CSF (Fig. 3(g)-(i)). This is due essentially to the fact that pixels of CSF and MS lesions share the same intensities.

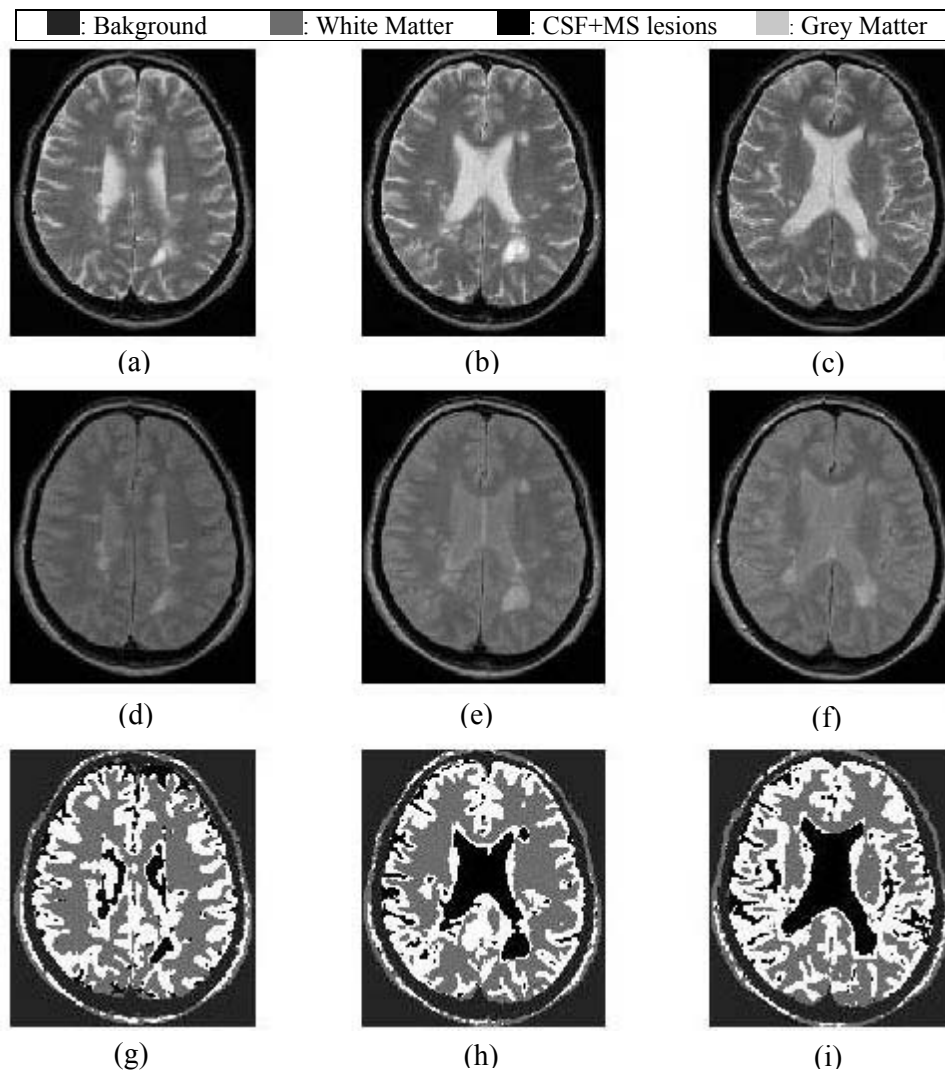


Figure 3: Segmentation result of MR images obtained in 3 patients with MS lesions. (a), (b), (c) T_2 weighted images. (d), (e), (f) PD weighted images. (g), (h), (i) DS fusion result of the three patients using average operation.

6 Conclusion

In this paper, an original data fusion scheme to multisensor images segmentation based on the DS and fuzzy logic theories to take into account nonfuzzy and fuzzy uncertainties is proposed. This methodology consists in estimating basic probability assignment using fuzzy membership degrees derived from gray-level image histogram. A contextual processing is introduced to integrate the spatial correlation between adjacent pixels. The obtained results on synthetic and medical images are encouraging. In this work, we make assumption that images are well registered. The presented results are limited to only one modality. Extensive tests on real data and analysis of several decision rules are necessary in order to evaluate the robustness of the method. Only filters with 3×3 size are used. Thus, different window sizes must be tested to show their effect on the fusion results.

References

- [1] D.L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, MA, 1992.
- [2] A. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Annals of Mathematical Statistics* (38): 325-339, 1967.
- [3] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [4] S. Le Hegarat-Masclé, I. Bloch and D. Vidal-Madjar, "Application of DS evidence theory to unsupervised classification in multiple remote sensing", *IEEE Trans. Geosci. Remote Sensing* (35): 1018-1031, 1997.
- [5] T. Lee, J.A. Richards and R.H. Swain, "Probabilistic and evidential approaches for multisource data analysis", *IEEE Trans. Geosci. Remote Sensing* (25): 283-293, 1987.
- [6] F. Salzenstein and A.O. Boudraa, "Unsupervised multisensor data fusion approach", *IEEE Proc. ISSPA*, Kuala Lumpur, 1:152-155, 2001.
- [7] M. Rombaut and Y. M. Zhu Study of Dempster-Shafer theory for image segmentation applications *Image and vision computing*, 20(1):15-23,2002.
- [8] I. Bloch, "Some aspect of Dempster-Shafer evidence theory for classification of multimodality medical images taking partial volume effect into account", *Pattern Recognition Letters* (17): 905-916, 1996.
- [9] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. Patt. Anal. Mach. Intell.* (12): 447-458, 1990.
- [10] S. Le Hegarat-Masclé, D. Richard and C. Otle Multi-scale data fusion using Dempster-Shafer evidence theory *Integrated Computer-Aided Engineering*, (10):9-22, 2003.
- [11] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, NY, 1981.
- [12] A.O. Boudraa and P. Clarysse, "Fast fuzzy gray level image segmentation method", *Medical Biological Engineering Computing* 35:686, 1997.
- [13] A. Bentabet, *Détermination des fonctions de masses dans le cas de la théorie de l'évidence par coalescence floue*, MSc. Dissertation, Institut National des Sciences Appliquées de Lyon, France, 1999.