

# Applying Voting to Segmentation of MR Images

Lasse Riis Østergaard and Ole Vilhelm Larsen

Dept. of Medical Informatics and Image Analysis, Aalborg University,  
Fredrik Bajers Vej 7D, 9220 Aalborg, Denmark  
lasse@vision.auc.dk  
<http://www.vision.auc.dk/~lasse/index.html>

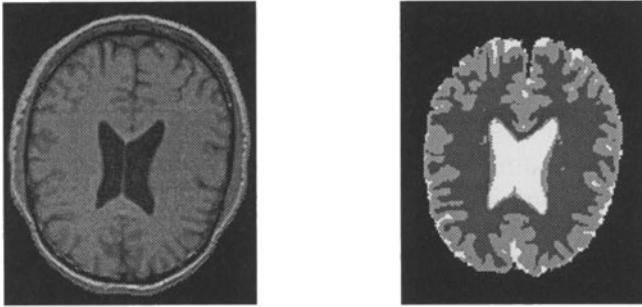
**Abstract.** The performance of applying voting to MR segmentation is investigated. Three different segmentation methods (fuzzy c-means, Bayes, and k-nearest neighbour) are used as input to the voting algorithm. Using human expert segmented images as a reference an error rate of 7.1% is obtained when applying voting. When comparing to the other methods it is seen that the results of applying the voting algorithm are slightly improved in terms of the error rate, minimum and maximum error.

## 1 Introduction

Since the mid-eighties medical images have gained increasing importance in the health care industry world-wide. Traditionally, the images are used for diagnostic purposes. A medical expert gives a qualitative evaluation of the images and through the "view inside the patient" provided by the images, the diagnose is made.

In recent years, sparked by the rapid development in computer power, new applications using medical images have emerged. From a sequence of images, 3D computer models of the patients organs can be produced [9]. One of the areas where such models are utilised is within the development of surgery simulators [2][6]. The general idea of such simulators is to allow the surgeon to practice the operation on a computer model of the patient prior to performing the real operation. Such applications rely on a very accurate localisation of the anatomical structures, e.g. when estimating the volume of a tumour or when the surgical pathway to the tumour is decided. As high accuracy is needed for the entire model the segmentation methods should be robust in order to assure high accuracy whenever applied to MR (Magnetic Resonance) images of the organs.

The construction of 3D models requires such a large amount of images that it is not feasible to have an expert segment the individual images manually. Automatic computerised segmentation of medical images is therefore highly required. Figure 1 shows a Magnetic Resonance image (MR image) of a patients brain and a medical experts segmentation of the image. The image is segmented into the different main tissue types found in the human brain, white matter, grey matter and Cerebro Spinal Fluid (CSF).



**Fig. 1.** Left) MR image of the human brain. Right) Segmented version of the MR image. White represents CSF, grey represents grey matter, and dark grey represents white matter.

*This paper focuses on the development of methods for automatic segmentation of MR images of the human brain.* The work is related to the Virtual Brain project<sup>1</sup> which aims to develop a surgical simulator for neurosurgery.

Automatic segmentation of MR images has been of interest for years and as of today several algorithms have been developed, including some based on the principles of statistical pattern recognition. For a survey of the methods see [1]. No single algorithm solves the problem in general. The individual algorithms seem to have their own pros and cons.

The idea behind this paper is not to apply a specific algorithm but to have several algorithms “running in parallel” and to have each of them cast a vote on the segmentation of each element in the picture. Based on the individual votes a segmentation is performed. Such a technique, called voting, has been utilised in other application areas with success [8], but has not to our knowledge previously been applied to MR image segmentation. Voting has the ability to combine decisions from a number of mechanisms and dependent on the strategy it can improve performance. In [7] it is stated that the judgement of a group is superior to those of individuals provided the individuals have reasonable competence. This brings us to expect that voting will increase segmentation performance given that the segmentation methods used will give reasonable segmentation results. *The objective of this paper is to investigate the performance of voting applied to MR segmentation in terms of accuracy and robustness.*

In section 2 the segmentation methods and voting algorithm used in this study are described and the image data used for the experiments is described. In section 3 the experiments for evaluating the performance of using voting in the context of MR image segmentation are described. A discussion and conclusion is given in section 4.

<sup>1</sup> A formal collaboration between Aalborg Hospital and Aalborg University, Denmark

## 2 Methods and Material

To enable an investigation of the performance when applying voting to MR segmentation a number of segmentation methods must be selected to provide inputs to the voting algorithm based on sequences of MR images.

### 2.1 Segmentation Methods

No thorough analysis of different segmentation methods has been performed in order to make a choice of which to use. It has been chosen to use three different multispectral segmentation methods. These are fuzzy c-means, Bayes, and k-nearest neighbour. All three methods are well-known pattern recognition techniques which have been used with success in different applications [1][10]. The latter two methods require representative training sets for the segmentation process. All methods are based on image intensities from a number of images. In the following a brief description of each of the methods will be given.

**Fuzzy c-means** In the literature fuzzy c-means has been reported as being one of the most promising MR segmentation methods [3]. Fuzzy c-means is an iterative clustering approach. It resembles the maybe more well-known technique c-means but it uses fuzzy membership functions instead of hard values. Fuzzy c-means partitions the data set  $X = x_1, x_2, x_3, \dots, x_n$  into  $c$  fuzzy subsets  $u_i$  where the value  $u_i(x_k)$  is the membership of  $x_k$  in class  $i$ . The values of  $u_i(x_k)$  are arranged as a  $c \times n$  matrix  $U$ . The method approximately minimises the sum of squared error function defined as

$$J_m(U, V : X) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \|x_k - v_i\| \quad (1)$$

where  $V = v_1, v_2, \dots, v_c$  is a set of cluster centers and  $m > 1$  is a weighting exponent affecting the fuzziness of  $u$ . The parameters  $(U, V)$  may minimise  $J_m$  only if  $u_{ik}$  and  $v_i$  are defined as

$$u_{ik} = \left[ \sum_{j=1}^c \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{2/(m-1)} \right]^{-1} \quad \text{for all } i, k \quad (2)$$

and

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \quad \text{for all } i. \quad (3)$$

The stop criteria for the method is determined by  $E_t \leq \varepsilon$  where

$$E_t = \sum_{i=1}^c \|v_{i,t+1} - v_{i,t}\| \quad \text{for all } t \quad (4)$$

The method contains the following iteration steps:

1. Initialize  $U_0$ . Choose  $c$ ,  $m$  and  $\varepsilon$ .
2. Compute all  $c$  cluster centers  $v_{i,0}$ .
3. Compute all  $c \times n$  memberships  $u_{ik,t}$  and update all  $c$  cluster centers  $v_{i,t+1}$ .
4. Compute  $E_t$ .
5. If  $E_t \leq \varepsilon$  then stop; else return to 3.

When the stop criteria is fulfilled a fuzzy output is available. However, in the further processing a hard valued output is necessary. To produce a hard valued output of the fuzzy c-means algorithm maximum memberships are found, that is, if  $\max(u_{ik}) = u_{1k}$  then  $x_k$  is assigned the label associated with class 1.

**Bayes** Bayes' optimal decision is a purely statistical approach to segmentation. The method is most commonly based on assumptions about the distribution of the data set  $X = x_1, x_2, x_3, \dots, x_n$ . For multispectral segmentation the multivariate normal distribution is assumed. The parameters which describe this distribution are the data means  $\mu$  and the covariance matrix  $\Sigma$ . These parameters are estimated from representative training sets of the different tissue types. In equation (5) Bayes theorem is expressed.

$$P(\omega_i|x_k) = \frac{p(x_k|\omega_i)P(\omega_i)}{p(x_k)} \text{ for all } i, k \quad (5)$$

where

$$p(x_k) = \sum_{j=1}^c p(x_k|\omega_j)P(\omega_j) \text{ for all } k \quad (6)$$

Bayes theorem gives the probability of a class  $\omega_i$  being present when a sample  $x_k$  is observed (a posteriori probability), provided we know the probability of the sample being observed when the class is present, the a priori probability of the class, and the probability of that sample. Assuming multivariate normal distributed data  $p(x_k|\omega_i)$  can be estimated directly from the density function

$$f_i(x_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{(x_k - \mu_i)^T \Sigma_i^{-1} (x_k - \mu_i)}{2}\right) \text{ for all } i, k \quad (7)$$

where  $d$  is the number of features. To estimate  $P(\omega_i)$  information about the distribution of the different tissue types must be used. By using a discriminant function  $g_i(x_k) = P(\omega_i|x_k)$  the sample  $x_k$  is assigned the label associated with the highest a posteriori probability.

**k-Nearest Neighbour** The k-nearest neighbour rule is a non-parametric technique, that is, no assumptions are made about the distribution of the data. The decisions are only based on representative training sets. However, it is still possible to estimate  $P(\omega_i|x_k)$  from the training sets containing  $n$  labeled samples by using the samples to estimate the densities involved. If a cell of volume  $V$  is

placed around  $X$  so that  $k$  samples are captured  $k_i$  of these will be labeled  $\omega_i$ . An estimate of the joint probability  $p(x_k, \omega_i)$  is

$$p_n(x_k, \omega_i) = \frac{k_i/n}{V} \quad (8)$$

A reasonable estimate of  $P(\omega_i|x_k)$  is then

$$P_n(\omega_i|x_k) = \frac{p_n(x_k, \omega_i)}{\sum_{j=1}^c p_n(x_k, \omega_j)} = \frac{k_i}{k} \quad (9)$$

This means that the a posteriori probability that  $\omega_i$  is the class, is the fraction of the samples within the cell that are labeled  $\omega_i$ . Using the discriminant function  $g_i(x_k) = P(\omega_i|x_k)$ ,  $x_k$  is assigned the label most frequently represented among the  $k$  nearest samples in the training sets.  $k = \sqrt{n}$  has been used as suggested in [4].

## 2.2 Voting Algorithm

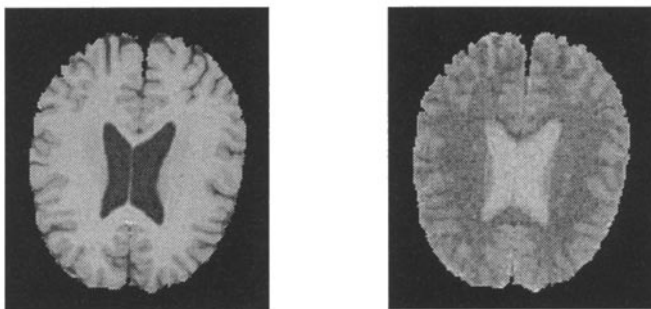
Voting is a technique used in a variety of applications [7][8] which combines information from a number of sources and produces outputs which reflects the consensus or a compromise of the information. In the context of MR image segmentation it is feasible to produce a consensus output in order to obtain a reliable segmentation. The reliability of the result is much dependent on the competence of the inputs (voters) and the number of inputs as the performance of the combined decision is an increasing function of the number of inputs [7]. Assuming that the inputs to the voting algorithm are statistically independent it can be shown theoretically, that it is possible to gain an increase in performance by using voting. In general, two subschemes exist related to consensus voting, that is, threshold voting and plurality voting. Threshold voting requires that the number of inputs which agree must exceed a certain preset threshold in order to make a decision. Using segmented images from  $m$  different methods as input a decision can be made when a number (decided by the preset threshold) of the segmentation methods have produced the same output. In plurality voting a decision is made based on the maximum support from the inputs. It is decided to use a voting algorithm from the threshold voting subscheme called majority voting which has been used with success in other applications [7]. In majority voting a sample is assigned the class for which there is a consensus or when at least  $l$  of the  $m$  inputs are agreed on the identity, where

$$l = \begin{cases} \frac{m}{2} + 1 & \text{if } m \text{ is even} \\ \frac{m+1}{2} & \text{if } m \text{ is odd} \end{cases} \quad (10)$$

Having input from three different segmentation methods a decision is made when at least two of the methods agree otherwise it is called a non-decision. Non-decision situations will theoretically still occur and therefore it must be dealt with. This is done by letting the most accurate of the three segmentation methods make decisions in such situations.

## 2.3 Image Data

The inputs to the three segmentation methods are two MR image sequences each consisting of 51 images obtained from the same person. The images are co-registered meaning that they are spatially aligned. The images have been acquired using a 1 Tesla MR scanner. The sequences are a MP-RAGE sequence [5] and a FISP sequence [5] yielding different discriminations between the tissue types. The MP-RAGE sequence and the FISP sequence have been processed by manually extracting the brain and putting it on a uniform black background. In that way only the relevant soft tissue structures are present. The images contain no pathological tissue. In figure 2 an example of spatially corresponding images from each sequence is shown. The different intensities in the images are due to



**Fig. 2.** Left) MP-RAGE image. Right) FISP image.

different scanner settings enabling enhancement of certain tissue structures. A reference segmentation exists for all the images from a manual segmentation by a human expert. The spatial resolution of the images is  $256 \times 256$  pixels with a depth resolution of 8 bit (256 grey levels). Each pixel represents a volume of size  $(1.168)^3 mm^3$ .

## 3 Experimental Evaluation

### 3.1 Objective

The objective of the experimental evaluation is twofold:

- Analyse the accuracy and robustness of MR image segmentation when applying voting.
- Compare the results with those of the individual voters (fuzzy c-means, Bayes, and k-nearest neighbour).

The accuracy should reflect the overall performance of a given segmentation method and the robustness should reflect the variations of the accuracy in a number of segmentations using that segmentation method.

### 3.2 Design

If counting the number of correct classified pixels and misclassified pixel for each image in a multispectral MR image sequence (MP-RAGE and FISP) it is possible to extract the accuracy and robustness of a given segmentation method. To decide whether a pixel has been correct classified or misclassified a human expert segmentation of the multispectral MR image sequence is used as a reference. The error rate of the segmentation is used as a measure of the accuracy and is calculated as:

$$\text{error rate} = \frac{\text{total number of misclassified pixels}}{\text{total number of pixels}} \quad (11)$$

The total number of misclassified pixels and the total number of pixels are found from a misclassification table. This table has entries  $n_{ij}$  corresponding to the number of pixels from class  $\omega_i$  classified as belonging to class  $\omega_j$ , thus the error rate is calculated as  $\sum_i \sum_{j \neq i} n_{ij} / N$ . Calculating the error rate for each image instead of for the entire image sequence, the variability of the error rate can be calculated and used as a measure of the robustness.

The protocol for the experiment can be made as:

1. Perform segmentations using fuzzy c-means, Bayes, and k-nearest neighbour.
2. Analyse the results of the segmentations.
3. Use the segmented image sequences from the three segmentations as input to the voting algorithm (majority voting). The most accurate (experimentally finding) method of the three segmentation methods decides the output whenever a non-decision situation occurs.
4. Analyse the result of the voting algorithm.
5. Compare the results obtained.

As a sub-experiment, the error rate is calculated using unanimity voting, that is, all segmentation methods must agree. This is done to indicate the lowest obtainable error rate when using voting. By disregarding non-decisions only the “ultrareliable” results are included in the calculation of the error rate.

The training sets used for Bayes and k-nearest neighbour have been extracted from a set of images (MP-RAGE and FISP) where CSF, white matter, and grey matter were well-represented. For this purpose the corresponding human expert segmented image has been used to decide which pixels belongs to each of the tissue types. From the histograms of the extracted training sets it seems reasonable to assume that the data is multivariate normal distributed. This is in accordance with the assumption made in Bayes.

Four parameters must be initialized, when using fuzzy c-means.  $U_0$  is initialized by first manually selecting three cluster centers (CSF, grey matter, and white matter) from a scatter plot of the first set of images and from these all memberships are computed. For the rest of the images the cluster centers found in the previous scatter plot is used as an initial estimate.  $m$  (weighting exponent) and  $\varepsilon$  (threshold value) are empirically chosen to 2 and 0.5 respectively. By removing the uniform black background of the images prior to segmentation  $c$  (the number of classes) can be set to 3. This is done for all three methods.

### 3.3 Results

In order to calculate the error rate of the three segmentation methods misclassification tables are made for each method. The values in table 1 are the number of pixels and the corresponding percentages which are shown in parentheses.

**Table 1.** Misclassification table (percentages are given in parentheses). Columns: Correct class. Rows: Classified as.

Segmentation method		CSF	grey matter	white matter
fuzzy c-means	CSF	56000 (98.41)	13074 (5.89)	0 (0)
	grey matter	891 (1.57)	180298 (81.29)	15944 (6.72)
	white matter	12 (0.02)	28414 (12.81)	221493 (93.29)
Bayes	CSF	53994 (94.89)	3049 (1.37)	0 (0)
	grey matter	2909 (5.11)	208278 (93.91)	21037 (8.86)
	white matter	0 (0)	10459 (4.72)	216400 (91.14)
k-nearest neighbour	CSF	54732 (96.11)	4233 (1.88)	0 (0)
	grey matter	2171 (3.89)	214752 (96.92)	43914 (19.97)
	white matter	0 (0)	2801 (1.19)	193523 (80.03)
voting	CSF	55004 (96.66)	4417 (1.99)	0 (0)
	grey matter	1899 (3.34)	206745 (93.21)	19750 (8.32)
	white matter	0 (0)	10624 (4.79)	217687 (91.68)

It seems from table 1 that all methods provide reasonable segmentations as all percentages of pixels classified correct are above 80%. It should be noted however, that Bayes provide percentages of pixels classified correct which are all above 90%. Not surprisingly it is seen that all methods have the largest difficulties in distinguishing grey and white matter. In table 2 the error rate for each of the segmentation methods is shown in the first column as a mean value of all the images. It is seen from table 2 that the smallest error rate of

**Table 2.** Error rate.

	Mean [%]	Std. dev. [%]	Min [%]	Max [%]
fuzzy c-means	11.30	3.63	6.65	18.89
Bayes	7.26	2.26	4.27	11.58
k-nearest neighbour	10.94	1.78	7.49	13.55
voting	7.11	2.26	4.05	11.27

the tree segmentation methods is obtained when using Bayes. In terms of the error rate, Bayes is the most accurate of the three segmentation methods, hence



it must make the decisions in non-decision situations when using voting. Using the segmentation results obtained by the three segmentation methods as inputs to the voting algorithm yields a segmented sequence of images which can be analysed in the same way as for the output of the three segmentation methods.

The misclassification table in table 1 indicates that voting yields a reasonable segmentation result since all values are above 90%. The error rate when using voting has been found to be 7.1% (see table 2 which is a slight improvement compared to Bayes. In table 2 the variations of the error rate in terms of the standard deviation and the minimum and maximum error are shown. It is seen from table 2 that by using k-nearest neighbour the lowest standard deviation is obtained. However, the maximum and minimum error found when using voting are lower than for the other methods.

The misclassification table when using unanimity voting is shown in table 3. It is seen from table 3 that the percentages of pixels classified correct are above

**Table 3.** Misclassification table (percentages are given in parentheses). Columns: Correct class. Rows: Classified as.

	CSF	grey matter	white matter
CSF	53333 (99.07)	2455 (1.33)	0 (0)
grey matter	502 (0.93)	179684 (97.35)	13501 (6.64)
white matter	0 (0)	2432 (1.32)	189793 (93.36)

93%. The error rate when using unanimity voting is 4.28% which is an indication of the lowest obtainable error rate.

## 4 Conclusion

The performance of a voting algorithm applied to MR segmentation has been investigated. Using three segmentation methods (fuzzy c-means, Bayes, and k-nearest neighbour) to provide inputs to the voting algorithm the accuracy and the robustness of the voting algorithm has been analysed and compared to those of the three segmentation methods. It was found that Bayes was the most accurate of the three segmentation method however a slight improvement of the accuracy was obtained when using voting. Regarding the robustness it was found that k-nearest neighbour was more constant with its error rates than the other methods. This must be seen in the light of the fact that k-nearest neighbour generated an overall error which was 3.8% larger than that when using voting. Both the maximum and minimum error were lowest when using voting. In general there were no larger deviation between the methods though fuzzy c-means showed a maximum error which was more than 5% larger than the others. To show the potential of voting an experiment was made using unanimity voting.

The experiment showed that the lowest obtainable error rate when using voting was 4.28%. Based on this study it must be concluded that improvements of the performance when applying voting to MR segmentation is possible. These improvement might not be significant improvements, however it is much dependent on the competence of the voters and the number of voters. Also, it can be argued whether the three segmentation methods produce statistically independent segmentations since all methods are intensity based. One possible source of error in this investigation has been the fact that the human expert segmentations is partly based on what he sees in the images and partly based on his knowledge about neuroanatomy. In future work this problem will be addressed by incorporating anatomical knowledge into the segmentation process.

## 5 Acknowledgements

The authors would like to thank Anders Lassen of the Dept. of Clinical Physiology and Nuclear Medicine, Bispebjerg University Hospital, Denmark for providing the human expert segmented images. Also the authors would like to thank Paolo Pirjanian of the Laboratory of Image Analysis, Aalborg University, Denmark for discussions related to voting.

## References

1. J.C. Bezdek, L.O. Hall, and L.P. Clarke: Review of MR image segmentation techniques using pattern recognition. *Medical Physics*, 20(4):1033-1048, 1993.
2. M. Bro-Nielsen and S. Cotin: Real-time Volumetric Deformable Models for Surgery Simulation using Finite Elements and Condensation. Institute of Mathematical Modelling, Technical University, Denmark, 1995.
3. L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger: MRI Segmentation: Methods and Applications. *Magnetic Resonance Imaging*, 13(3):343-368, 1995.
4. R.O. Duda and P.E. Hart: *Pattern Classification and Scene Analysis*. Wiley-Interscience, 1973.
5. P.T.English and C. Moore: *MRI for Radiographers*. Springer-Verlag, 1995.
6. U. Kuhn, Kühnapfel, H.-G. Krumm, and B. Neisius: *The Karlsruhe Endoscopic Surgery Trainer - A "Virtual Reality" based Training System for Minimal Invasive Surgery*. CAR, 1996.
7. L. Lam and C.Y. Suen: Application of Majority Voting to Pattern recognition: An Analysis of its Behaviour and Performance. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(5):553-568, 1997.
8. B. Parhami: Voting Algorithms. *IEEE Transactions on Reliability*, 43(4):617-629, 1994.
9. A. Pommert, M. Riemer, T. Schiemann, R. Schubert, U. Tiede, and K.H. Höhne: Three-Dimensional Imaging in Medicine: Methods and Applications. In *Computer-Integrated Surgery*, pages 155-174. The MIT Press, 1996.
10. M. Vaidyanathan, L.P. Clarke, C. Heidtman, R.P. Velthuizen, and L.O. Hall: Normal Brain Volume Measurement using Multispectral MRI Segmentation. *Magnetic Resonance Imaging*, 15(1):87-97, 1997.