

3-D Contextual Bayesian Classifiers

Rasmus Larsen

Abstract

We extend a series of multivariate Bayesian 2-D contextual classifiers to 3-D by specifying a simultaneous Gaussian distribution for the feature vectors as well as a prior distribution of the class variables of a pixel and its 6 nearest 3-D neighbours.

Keywords

Classification, Segmentation, 3-D, Contextual methods

I. INTRODUCTION

WHEN In [1], [2], [3], [4] algorithms for 2-D images that base the classification of a pixel on the feature vectors of the pixel itself and those of the 4 nearest neighbours are introduced. In [3] it is assumed that the classes of the nearest neighbours of a pixel are conditionally independent given the class of the center pixel, whereas in [1], [2] it is assumed that the pixel size is small relative to the grains of the pattern under study. In this article we will extend these algorithms to 3-D images, and carry out a series of tests on two simulated 3-D images, one that expresses low spatial frequency of the signal and one with high frequency signal. Furthermore, we will illustrate the use of the 3-D contextual classification for tissue classification in a 3-D magnetic resonance image of a human brain.

II. METHODS

In this Section we develop a 3-D contextual classification rule, specify a Gaussian observation model and specify a prior distribution for the class variable. Additionally, we describe an algorithm based on the Potts model.

A. Construction of a Contextual Classification Rule

Suppose that a pixel belongs to one of the classes $\pi_1, \pi_2, \dots, \pi_k$. For each pixel we observe a vector of features $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$. The augmented feature vector consisting of the features vectors of the (north, south, east, west, top, and bottom) neighbours of a pixel is denoted $\mathbf{D}_\Delta = (\mathbf{X}_N^T, \mathbf{X}_S^T, \mathbf{X}_E^T, \mathbf{X}_W^T, \mathbf{X}_T^T, \mathbf{X}_B^T)^T$, and the augmented feature vector consisting of the feature vector of a pixel itself and those of its neighbours is denoted $\mathbf{D} = (\mathbf{X}^T, \mathbf{D}_\Delta^T)^T$. We assume knowledge of the prior probabilities, $P(C = \pi_i) = p_i, i = 0, 1, \dots, k$ where C is the class variable.

The Bayes solution to the classification problem for the case of equal losses is obtained by setting the discriminant score to be maximised across classes equal to the maximum a posteriori probability. The posterior distribution is

$$\begin{aligned} f(\pi_\nu | \mathbf{d}) &= P(C = \pi_\nu | \mathbf{D} = \mathbf{d}) = \frac{P(C = \pi_\nu) P(\mathbf{D} = \mathbf{d} | C = \pi_\nu)}{\sum_{i=1}^k P(C = \pi_i) P(\mathbf{D} = \mathbf{d} | C = \pi_i)} \\ &= \frac{\sum_{a,b,c,d,e,f} p_\nu P(\mathbf{D} = \mathbf{d} | C = (\pi_\nu, \pi_a, \pi_b, \pi_c, \pi_d, \pi_e, \pi_f)) g(\pi_a, \pi_b, \pi_c, \pi_d, \pi_e, \pi_f | \pi_\nu)}{h(\mathbf{d})} \end{aligned} \quad (1)$$

where $h(\mathbf{d})$ is the unconditional density of the augmented feature vector, (a, b, c, d, e, f) is one of the possible k^6 configurations of the class variables of the neighbouring pixels, C is the class configuration corresponding to the augmented feature vector \mathbf{D} , and $g(\pi_a, \pi_b, \pi_c, \pi_d, \pi_e, \pi_f | \pi_\nu)$ is the probability of the configuration of the class variables of the neighbouring pixels given that the center pixel has class π_ν . The spatial nature of the data is used in the modelling of the spatial dependence of the feature vectors and in the prior distribution of the classes of a pixel and its neighbours.

B. Specification of the Gaussian observation model

Following Hjort et al. [2] we assume that each feature vector may be written as a sum of two terms, i.e. $\mathbf{X} = \mathbf{Y} + \boldsymbol{\epsilon}$, where the \mathbf{Y} terms are independent given the classes and model the class dependency of the feature vectors, i.e. $(\mathbf{Y} | C = \pi_i) \in N(\boldsymbol{\mu}_i, (1-\theta)\boldsymbol{\Sigma})$ and $(\boldsymbol{\epsilon}_{s(1)}^T, \dots, \boldsymbol{\epsilon}_{s(N)}^T)$ is zero mean, multinormal and model an autocorrelated noise term with $E\{\boldsymbol{\epsilon}_{j_1} \boldsymbol{\epsilon}_{j_2}^T\} = \rho^{\|\mathbf{s}(j_1) - \mathbf{s}(j_2)\|_2} \theta \boldsymbol{\Sigma}$. ρ is the autocorrelation between first-order neighbours, and θ is the proportion of the covariance matrix $\boldsymbol{\Sigma}$ that is due to autocorrelated noise. $\mathbf{s}(j)$ refers to the spatial position of pixel number j , and N is the number of pixels. Here we use a correlogram model based on the 2-norm, alternative models include using the 1-norm (i.e. the city-block or Manhattan distance), or the ∞ -norm. In Fig. 2 realisations of autocorrelated noise patterns corresponding to using the 2-norm as well as the 1-norm and the ∞ -norm for the

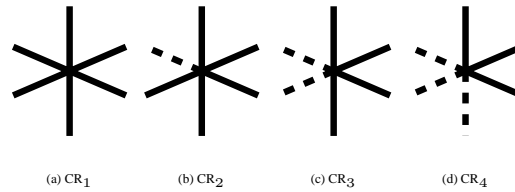


Fig. 1. Patterns in the model. Within the 'cross', that represents the neighbourhood of a pixel, i.e. the six nearest neighbours, it is assumed that at most two classes are present, and that the only possible configurations are these four types of 'crosses'.

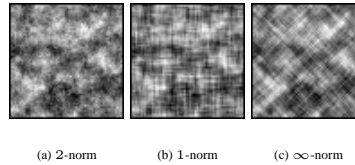


Fig. 2. 2-D Noise patterns corresponding to autocorrelation functions using (a) the 2-norm (Euclidean), (b) the 1-norm (Manhattan), and (c) the ∞ -norm. All three realizations have an autocorrelation of 0.8 in for first-order neighbours.

correlogram are shown. All these models are valid corellogram models [5], thus ensuring positive definiteness of the covariance matrix below.

Let \otimes denote the Kronecker product, and α , β , and γ be as given in Table I for the different autocorrelation models mentioned above. Then the conditional distribution of the augmented feature vector given the classes is

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{X}_N \\ \mathbf{X}_S \\ \mathbf{X}_E \\ \mathbf{X}_W \\ \mathbf{X}_T \\ \mathbf{X}_B \end{bmatrix} \mid \mathbf{C} = (\pi_\nu, \pi_a, \pi_b, \pi_c, \pi_d, \pi_e, \pi_f) \in N_{7p} \left[\begin{bmatrix} \mu_\nu \\ \mu_a \\ \mu_b \\ \mu_c \\ \mu_d \\ \mu_e \\ \mu_f \end{bmatrix}, \begin{bmatrix} 1 & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & 1 & \gamma & \beta & \beta & \beta \\ \alpha & \gamma & 1 & \beta & \beta & \beta \\ \alpha & \beta & \beta & 1 & \gamma & \beta \\ \alpha & \beta & \beta & \gamma & 1 & \beta \\ \alpha & \beta & \beta & \beta & \beta & 1 \\ \alpha & \beta & \beta & \beta & \beta & \gamma \end{bmatrix} \otimes \Sigma \right] \quad (2)$$

C. The OHM model for the prior distribution of the class variable

Following [1], [2] - the Owen-Hjort-Mohn model (OHM) - we assume that the regions of the image are large compared to the pixel size. Therefore, patterns other than CR_2 , CR_3 , and CR_4 shown in Fig. 1 will occur with very small probability on the region borders. Let the probability of a pixel being an interior point (i.e. CR_1) be p . Furthermore, let the probability of a pixel being on a border parallel to two of the coordinate axes (i.e. CR_2) be q , on a border parallel to only one of the coordinate axes (i.e. CR_3) be r , and on a plane that is not parallel to any of the coordinate axes (i.e. CR_4) be $s = 1 - p - q - r$. All other configurations are assumed to occur with probability 0.

By rotation we obtain six, twelve, and eight different CR_2 , CR_3 , and CR_4 patterns, respectively. Note, that from the assumption of the regions being larger than the pixel size, we have that the pixels within the 'cross' in the CR_2 , CR_3 , and CR_4 cases that are different from the center pixel, all have the same class. In all, given the center pixel class we have $1 + 6(k-1) + 12(k-1) + 8(k-1) = 26k - 25$ different configurations. These patterns are assigned positive a priori probabilities, while all other patterns are assigned the probability zero.

Under these assumptions the possible patterns in the prior model have the probabilities shown in Equation (3). For each cross an example of a configuration is shown. For the CR_2 example only the north pixel has a class different from the center pixel, for CR_3 the north and east pixel have a different class than the center pixel, and for CR_4 the north, east and top pixel belong to a different

TABLE I
AUTOCORRELATIONS BETWEEN FIRST-ORDER (α), SECOND-ORDER (β), AND THIRD-ORDER (γ) NEIGHBOURS CORRESPONDING TO THREE DIFFERENT NORMS USED IN THE DEFINITION OF THE AUTOCORRELATION MODEL.

	α	β	γ
∞ -norm	$\rho\theta$	$\rho\theta$	$\rho^2\theta$
2-norm	$\rho\theta$	$\rho\sqrt{2}\theta$	$\rho^2\theta$
1-norm	$\rho\theta$	$\rho^2\theta$	$\rho^2\theta$

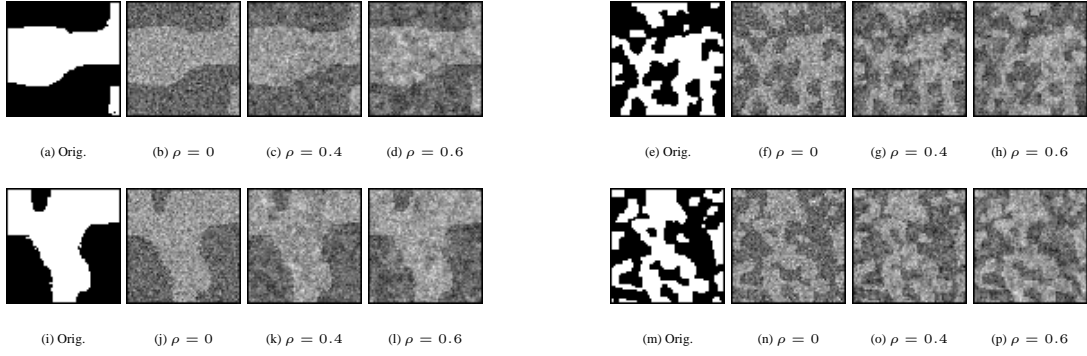


Fig. 3. Horizontal slices 32 (top row, (a)-(h)) and vertical slices 32 (bottom row, (i)-(q)) of the two original simulated data volumes and the six degraded sequences. The low frequency signal is on the left and the high frequency signal is on the right.

class than the center pixel.

$$\begin{aligned}
 \text{CR}_1 &: g(\pi_\nu, \pi_\nu, \pi_\nu, \pi_\nu \pi_\nu, \pi_\nu | \pi_\nu) = p + (q + r + s) \cdot p_\nu \\
 \text{CR}_2 &: g(\pi_i, \pi_\nu, \pi_\nu, \pi_\nu, \pi_\nu, \pi_\nu | \pi_\nu) = \frac{1}{6} q p_i \\
 \text{CR}_3 &: g(\pi_i, \pi_\nu, \pi_i, \pi_\nu, \pi_\nu, \pi_\nu | \pi_\nu) = \frac{1}{12} r p_i \\
 \text{CR}_4 &: g(\pi_i, \pi_\nu, \pi_i, \pi_\nu, \pi_i, \pi_\nu | \pi_\nu) = \frac{1}{8} s p_i
 \end{aligned} \tag{3}$$

where $\nu \neq i$, and $\nu, i = 1, \dots, k$. In this way we have obtained a huge reduction in the number of non-zero terms in the contextual classification rule (Equation (1)).

D. The WSH model for the prior distribution of the class variable

Alternatively, following [3], [4] - the Welch-Salter-Haslett (WSH) model - we assume independence between the class variables of the neighbours given the center pixel class, i.e.

$$g(\pi_a, \pi_b, \pi_c, \pi_d, \pi_e, \pi_f | \pi_\nu) = \phi(\pi_a | \pi_\nu) \phi(\pi_b | \pi_\nu) \phi(\pi_c | \pi_\nu) \phi(\pi_d | \pi_\nu) \phi(\pi_e | \pi_\nu) \phi(\pi_f | \pi_\nu).$$

Here $\phi(\pi_i | \pi_j) = P(C_A = \pi_i | C_B = \pi_j)$, where A and B are immediate neighbours. The model leads to a considerable simplification of Equation (1) in the case of conditional independence of the feature vectors given the class variables, i.e. $\theta =$

TABLE II

MISCLASSIFICATION RATES FOR EACH OF THE COMBINATIONS BETWEEN CLASSIFIER AND NOISE LEVEL. OHM AND WSH REFERS TO THE OWEN-HJORT-MOHN AND THE WELCH-SALTER-HASLETT METHODS, A PREFIX CAPITAL A DENOTES USE OF AN AUTOCORRELATED NOISE MODEL (I.E. $\theta \neq 0$), WHEREAS A MISSING CAPITAL A DENOTES THE USE OF A WHITE NOISE MODEL ONLY (I.E. $\theta = 0$). FINALLY, THE POSTFIX n -D INDICATES THE SIZE OF THE CONTEXT CONSIDERED IN THE ALGORITHM.

	Low frequency image			High frequency image		
	White noise	Autocorrelated noise		White noise	Autocorrelated noise	
		$\rho = 0.4$	$\rho = 0.6$		$\rho = 0.4$	$\rho = 0.6$
Non-context	15.8	15.8	16.0	15.9	15.8	16.2
AOHM 3-D	-	3.6	5.9	-	6.2	8.4
OHM 3-D	1.1	3.7	6.0	3.0	6.3	8.8
AWSH 3-D	-	5.0	7.0	-	7.7	9.6
WSH 3-D	2.2	4.9	7.0	4.6	7.8	10.0
AOHM 2-D	-	4.9	7.0	-	7.8	9.6
OHM 2-D	2.2	4.9	7.0	4.6	7.8	9.8
AWSH 2-D	-	4.9	7.1	-	7.7	9.9
WSH 2-D	2.2	4.9	7.1	4.6	7.8	10.0
AOHM 1-D	-	7.5	9.1	-	10.7	11.9
OHM 1-D	5.1	7.4	9.0	8.1	10.2	11.7
AWSH 1-D	-	7.3	8.9	-	10.3	11.9
WSH 1-D	4.9	7.3	9.0	8.1	10.4	12.1
Potts, $\beta = 0.1$	11.2	11.8	12.5	12.0	12.6	13.3
Potts, $\beta = 0.5$	0.4	0.8	3.9	3.1	4.7	6.9
Potts, $\beta = 1.0$	1.3	2.3	2.0	1.9	3.3	5.5

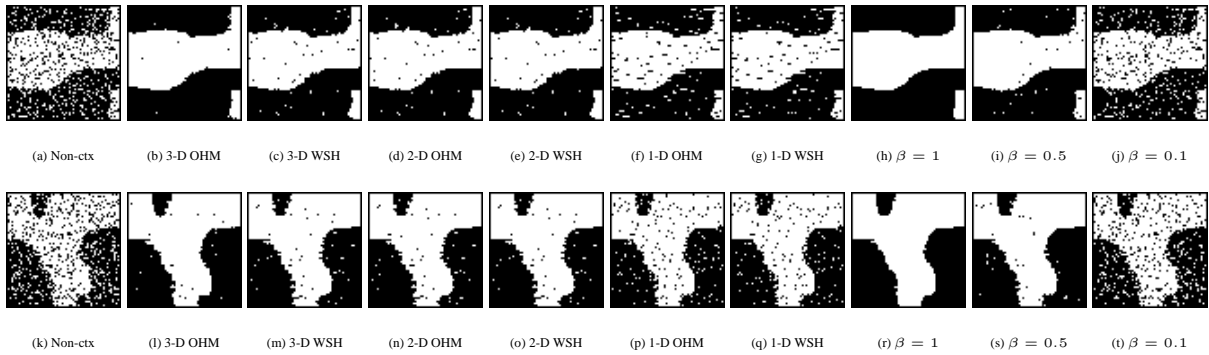


Fig. 4. Classification of the low frequency data volume degraded with white noise (signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

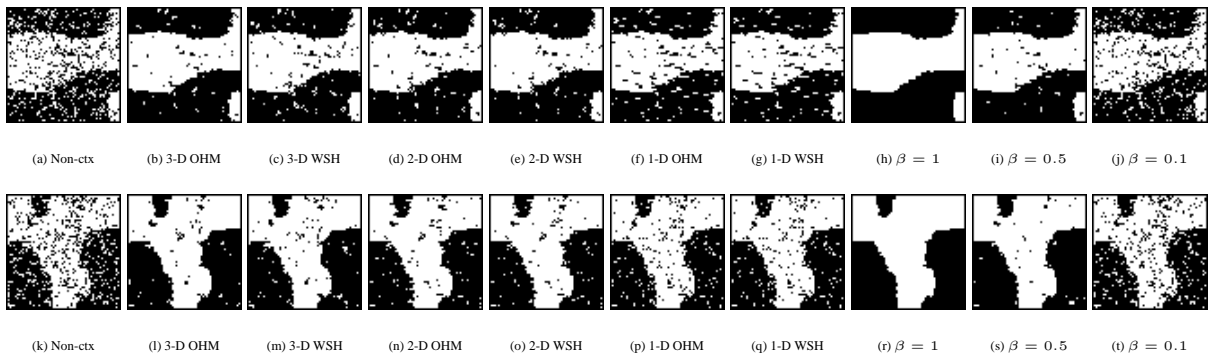


Fig. 5. Classification of the low frequency data volume degraded with autocorrelated noise ($\rho = 0.4$, signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

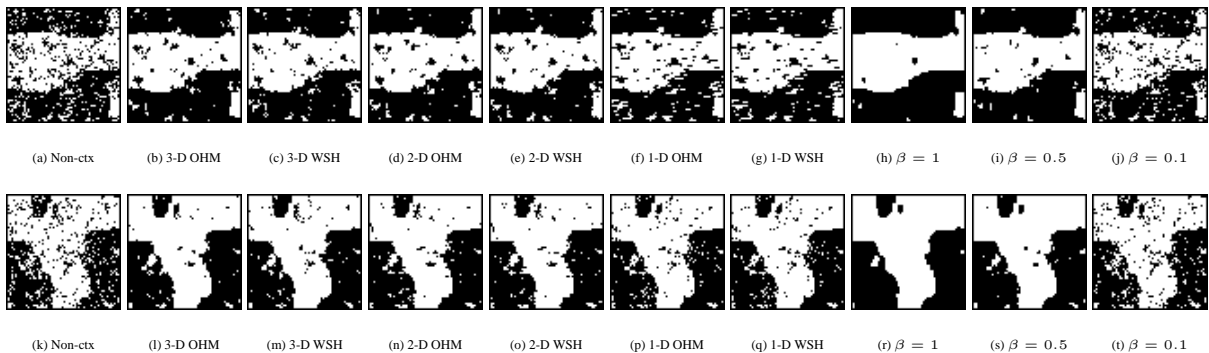


Fig. 6. Classification of the low frequency data volume degraded with autocorrelated noise ($\rho = 0.6$, signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

$0 \vee \rho = 0 \Leftrightarrow \alpha = \beta = \gamma = 0$ in Equation (2). In the case of autocorrelated noise, however, an approximation is necessary (for computational reasons). In [2] it is suggested to approximate the covariance matrix of \mathbf{D}_Δ by a diagonal matrix with equal diagonal elements having the same determinant as the original matrix.

E. Standard method for comparison

Finally, we include results obtained by performing the classification using a Potts model for the prior. For reviews of the Potts model see [7], [8]. The conditional probabilities are given by

$$P(C = \pi_\nu | \mathbf{C}_\Delta = \mathbf{c}_\Delta) = \frac{\exp(\beta u(\nu))}{\sum_{i=1}^k \exp(\beta u(i))} \quad (4)$$

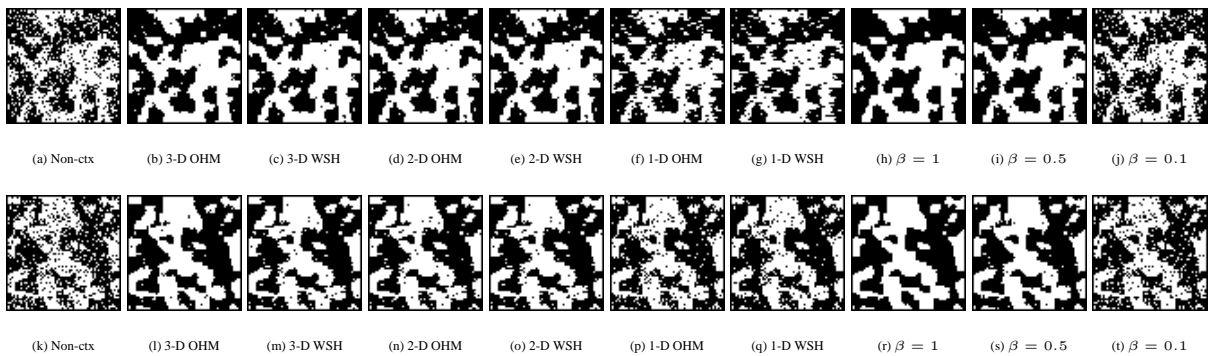


Fig. 7. Classification of the high frequency data volume degraded with white noise (signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

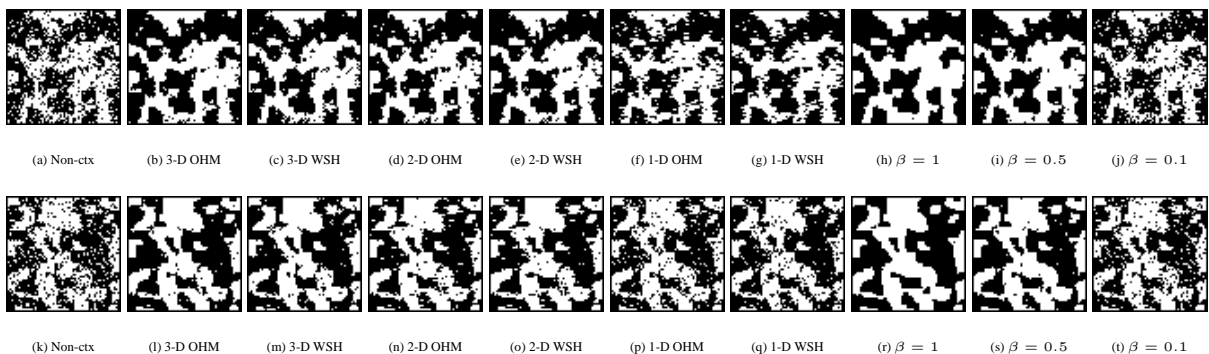


Fig. 8. Classification of the high frequency data volume degraded with autocorrelated noise ($\rho = 0.4$, signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

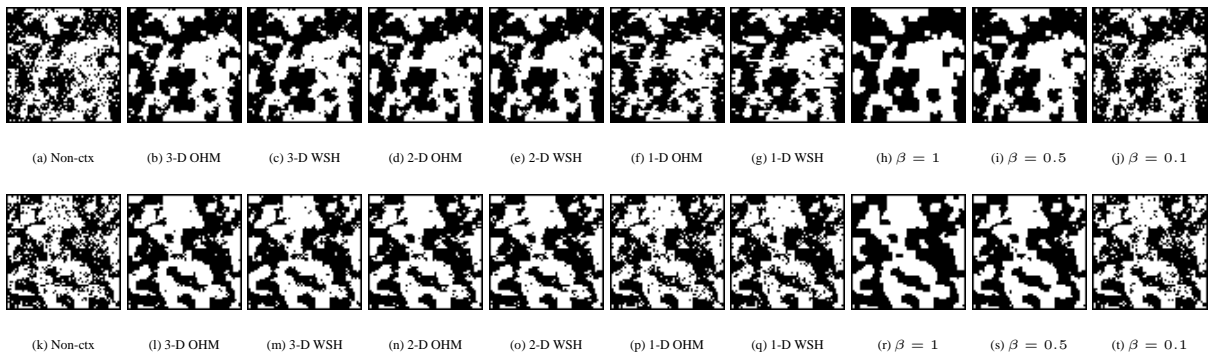


Fig. 9. Classification of the high frequency data volume degraded with autocorrelated noise ($\rho = 0.6$, signal-to-noise ratio 0 dB). Horizontal slice 32 (top row) and vertical slice 32 (bottom row) of the classified volumes using the contextual Bayesian algorithms ((a)-(g) and (k)-(q)), and a Potts model with three parameter settings ((h)-(j) and (r)-(t)).

where C_{Δ} is the classes of the neighbours, and $u(\nu)$ is the number of the neighbours that have class π_{ν} . If this Potts model has two states (i.e. $k = 2$) it is equivalent to an isotropic zero-field Ising model (e.g. see [9]).

We use a Gaussian observation model, i.e. $(\mathbf{X} | C = \pi_i) \in N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. In the posterior distribution the faith in the observations vs. the faith in the model is controlled by a parameter γ . Only one of β and γ can be varied freely. The maximum a posteriori estimate is found using the Gibbs sampler and simulated annealing [9].

III. DATA

The procedures described above are tested on Monte Carlo simulated data and a magnetic resonance image of a human brain.

The two simulated datasets consist of $64 \times 64 \times 64$ data volumes with one feature observed at every pixel and express low and high spatial frequency, respectively. In Figs. 3(a) and 3(i), and Figs. 3(e) and 3(m) horizontal slice 32 and vertical slice 32 of the volumes are shown.

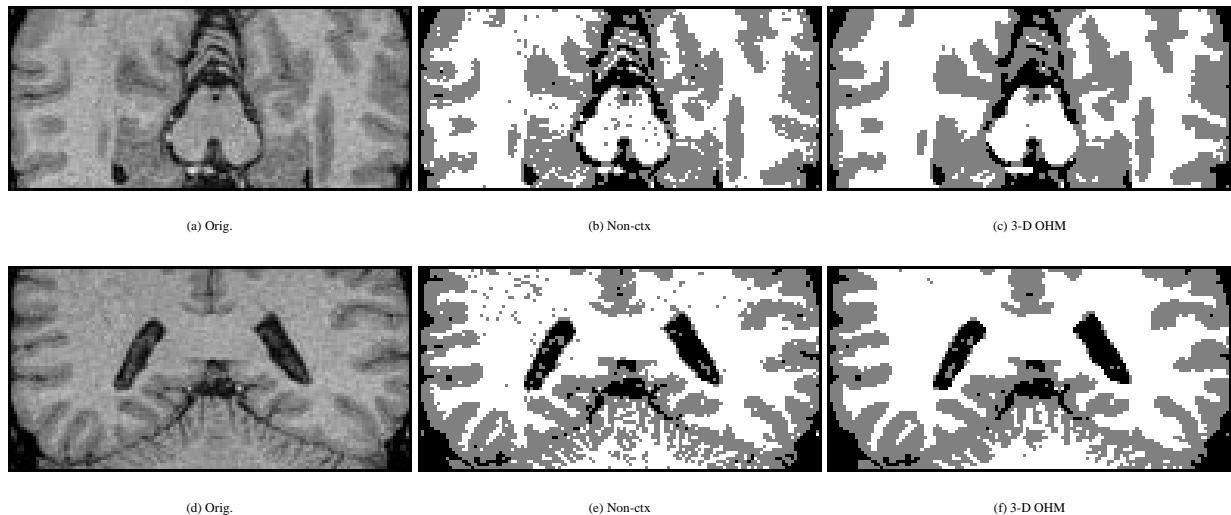


Fig. 10. Horizontal slice 44 ((a)-(c)) and vertical slice 16 ((d)-(f)) of the image of the human brain under study. The original image is shown on the left ((a), (d)). Classification results using non-contextual classification in the middle ((b), (e)), and classification using 3-D OHM on the right ((c), (f)). White matter is shown as white, grey matter as grey and cerebro-spinal fluid as black.

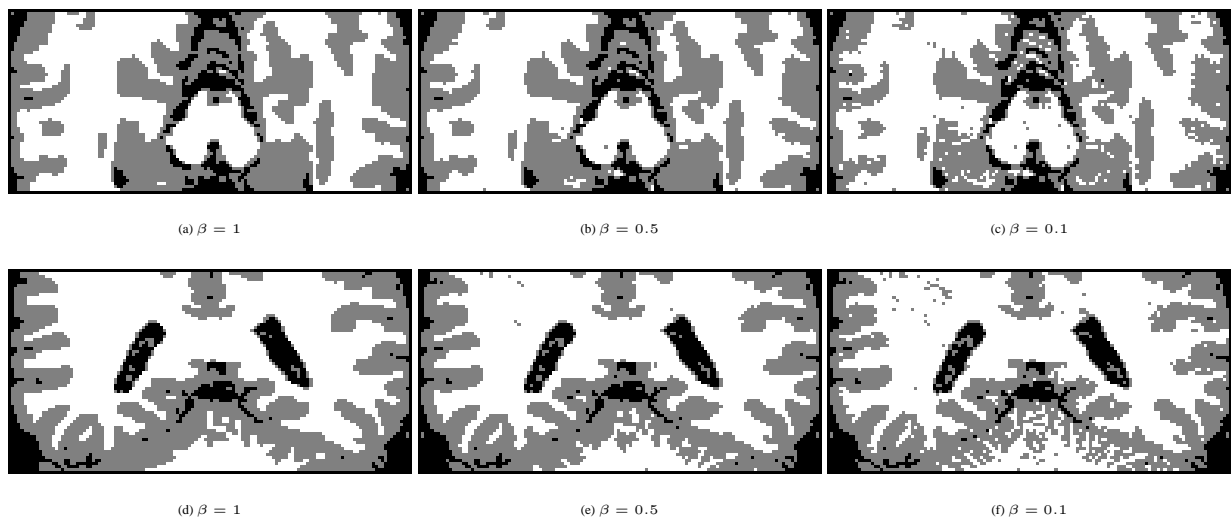


Fig. 11. Three classification results using stochastic relaxation with a Potts model. The parameter β controls the number of neighbours that have the same class. White matter is shown as white, grey matter as grey and cerebro-spinal fluid as black.

We consider two cases of degradation. First, the case of pure white noise, and second, the case of a mixture of white and autocorrelated noise. We use a Gaussian noise level with a signal-to-noise ratio of 0 dB. In both cases we compare the contextual classifiers with a classical pixelwise linear classifier (e.g. [6]). In addition to this we make comparisons between the classifications using the 3-D algorithms with implementations where contextual information is drawn only from 2-D (corresponding to the algorithms in [2], [3]), as well as implementations where only 1-D context is used.

The T1 weighted magnetic resonance image was scanned on a Magnetom Vision scanner from Siemens. We apply our algorithm to a $130 \times 68 \times 61$ subimage. A vertical and a horizontal slice is shown in Figs. 10(a) and 10(d). We want to discriminate between grey matter, white matter, and cerebro-spinal fluid.

IV. CLASSIFICATION RESULTS

All classifications of the simulated datasets are performed using the true parameters for mean values, variances, and autocorrelations. The transition probabilities of the WSH models and the prior distribution of the neighbourhood configurations of the OHM models are estimated by their relative occurrences in the simulated data volumes.

For each class the parameters of the algorithm are estimated from a training set. The training set is generated by a 3-D version of the 2-D seed algorithm described in [10]. For each class a single pixel (seed) is identified. An initial training set is generated by expanding this pixel to a small neighbourhood. From this neighbourhood initial mean value and dispersion matrix of the class

are estimated. These estimates are used to include new pixels in the training set. The inclusion of a new pixel in the training set is conditioned on the pixel being connected with the original seed through other pixels in the training set and on the estimated Mahalanobis distance from the feature vector of the pixel to the class centre being less than a specified quantile in the approximate $\chi^2(p)$ distribution of the Mahalanobis distance.

A. Case 1: White noise

In Figs. 3(b) and 3(j), and in Figs. 3(f) and 3(n) degraded slices are shown. The misclassification rates for the classifications are shown in Table II. For the Potts model based algorithm we apply the classifier with three different settings of the parameter β in Equation (4), and the weight of the observation model energy fixed to $\gamma = 1$.

For the non-contextual classifier the theoretical misclassification rate is 15.866%. The obtained results agree well with this. When compared with the contextual OHM 3-D classifier, we see that the inclusion of spatial information results in a misclassification rate that is a factor 15 lower for the low frequency image. For the WSH 3-D model the misclassification rate is also lower, though not as good as for the OHM model. It is noteworthy that whereas the OHM models increase their performance as more spatial dimensions are included, the misclassification rate does not decrease for the WSH model when going from 2-D to 3-D. Also, where OHM 2-D and WSH 2-D performs equally well, the OHM model is superior in the 3-D case.

Apart from the contextual methods performing significantly better in terms of misclassification rates the original patterns are clearly discernible when comparing with the non-contextual method, as is shown in Fig. 4. It should also be noted that the errors tend to occur on the edges, and that the errors also tend to lump together in the directions where contextual information is included (i.e. for the 1-D algorithms the errors frequently occur in east-west line segments, whereas in the other directions they seem to occur more randomly).

With respect to the spatial frequency of the signal the performance of the classifications decrease with spatial frequency. This is what would be expected. The number of edge pixels are significantly higher and errors seem to occur predominantly at edges.

In comparison with the Potts model restoration we see that in terms of misclassification rates the Potts model (with the right choice of parameter) performs a little better than the OHM 3-D method. However, it should be taken into account that no tuning is necessary for the OHM method.

B. Case 2: Autocorrelated and white noise

In Figs. 3(c)-3(d) and 3(k)-3(l), and in Figs. 3(g)-3(h) and 3(o)-3(p) degraded slices are shown. With respect to the contextual methods we see the same pattern as for the white noise only situation: The 2-D algorithms works equally well, whereas the extension to 3-D increase the performance only for the OHM model. The lowest misclassification rates are obtained for the (A)OHM 3-D classifiers. It should also be noted that for the WSH models the inclusion of spatial autocorrelation in the noise model does not have an effect. For the OHM models the effect of including spatial autocorrelation in the noise model is hardly discernible. Examples of the classification results on the slices shown in Fig. 3 are shown in Figs. 5 and 6.

With respect to the autocorrelated noise in the images it is clear that the relative improvement of including the context in the classification is less. The neighbours hold less extra information as is also noted in [3].

For the autocorrelated noise situation the advantage of the Potts model restoration over the OHM 3-D method is even more explicit. The Potts model has the advantage of drawing information from a larger neighbourhood (i.e. the entire image) when classifying a pixel.

C. MRI image of human brain

The human brain MRI image is classified using a non-contextual and the 3-D OHM algorithm as well as Potts model restoration using three parameter settings. For the non-contextual classification the result is shown in Figs. 10(b) and 10(e). For the 3-D OHM algorithm the results are shown in Figs. 10(c) and 10(f), and for the Potts model restoration classification examples are shown in Fig. 11. The non-contextual method gives a classification that expresses some salt-and-pepper noise. This noise is not present in the 3-D OHM result. For the parameter setting $\beta = 1$ the Potts model restoration also gives salt-and-pepper noise free results. However, this is at the cost of overly smoothing of some fine patterns. An example of which can be seen in the middle bottom of the vertical slice (Fig. 11(d)). Lowering β to e.g. $\beta = 0.1$ allows for a more reasonable classification of this pattern but this results in salt-and-pepper noise in other parts of the image.

For this real world example it seems that the 3-D OHM algorithm has some advantages over the Potts model.

V. CONCLUSION

We have described extensions of 2-D contextual classification algorithms by Owen, Hjort & Mohn (OHM) and Welch, Salter & Haslett (WSH) based on the simultaneous distribution of a pixel and its nearest neighbours to the 3-D case. The algorithms include contextual information for each pixel by including the feature vector of that pixel as well as the feature vectors of the 6 nearest neighbouring pixels in the decision. A joint Gaussian distribution for these feature vectors given the classes of the pixels has been specified. It is assumed that the noise can be modelled as a sum of white noise and autocorrelated noise, where the autocorrelation function is exponentially decaying with (Euclidean) distance. Furthermore, joint prior distributions of the class variables of a pixel and its 6 nearest neighbours have been specified.

Compared to a restoration based on a Potts model for the underlying images the 3-D OHM model performs almost as well in terms of misclassification rates for the case of white noise. When autocorrelated noise is added the Potts model has the advantage. For a case of classification of a magnetic resonance image of a human brain with respect to tissue type the 3-D OHM model is more robust than the Potts model. At the same time the 3-D OHM algorithm avoids salt-and-pepper noise and correctly classifies areas of fine structure.

VI. ACKNOWLEDGEMENTS

The MR image was recorded at the Copenhagen University Hospital by Shumei Murakami, D. D. S., Ph.D., Department of Oral Maxillofacial Radiology, Faculty of Dentistry, Osaka University, and put at my disposal by M. Sc. Per Rønsholt Andresen, Department of Mathematical Modelling, Technical University of Denmark, whose brain it is.

REFERENCES

- [1] A. Owen, "A neighbourhood-based classifier for LANDSAT data," *The Canadian Journal of Statistics*, vol. 12, pp. 191–200, 1984.
- [2] N. L. Hjort, E. Mohn, and G. Storvik, "Contextual classification of remotely sensed data, based on an auto-correlated model," in *Contextual classification of remotely sensed data: Statistical methods and development of a system*, H. V. Sæbø, K. Bråten, N. L. Hjort, B. Llewellyn, and E. Mohn, Eds. Norwegian Computing Center, 1985, Technical report No. 768.
- [3] J. Haslett, "Maximum likelihood discriminant analysis on the plane using a markovian model of spatial context," *Pattern Recognition*, vol. 18, no. 3, pp. 287–296, 1985.
- [4] J. R. Welch and K. G. Salter, "A context algorithm for pattern recognition and image interpretation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 1, pp. 24–30, 1971.
- [5] N. E. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, second edition, 1993, 900 pp.
- [6] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, second edition, 1984, 675 pp.
- [7] F. Y. Wu, "The potts model," *Reviews of Modern Physics*, vol. 54, no. 1, pp. 235–268, 1982.
- [8] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.
- [9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [10] A. A. Nielsen, H. Flesche, R. Larsen, J. M. Rykkje, and M. Ramm, "Semi-automatic supervised classification of minerals from x-ray mapping images," in *Proceedings of the Fourth Annual Conference of the International Association for Mathematical Geology (IAMG'98)*, A. Buccianti, G. Nardi, and R. Potenza, Eds., Ischia, Naples, Italy, Oct. 1998, pp. 473–478.



Rasmus Larsen was born in Brørup, Denmark, on March 14, 1966. He received his M.Sc.(Eng) from the Technical University of Denmark in 1991, and his Ph.D. in statistical image analysis from the Technical University of Denmark in 1994. Until 1997 he was an Assistant Professor (research) at the Department of Mathematical Modelling, the Technical University of Denmark. He is now a statistician with Novo Nordisk A/S.