Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

INCORPORATING CONTEXT IN STOCHASTIC SEGMENTATION OF B&W IMAGERY USING TEXTURED OPERATORS

S. GOH, M.M. TRIVEDI, C.A. HARLOW

Louisiana State University/Department of Electrical and Computer Engineering
Baton Rouge, Louisiana

I. ABSTRACT

The objective of the research reported in this paper is to present the implementation and results of an adaptive image segmentation procedure based on Kalman filter approach. The primitive operators used for the segmentation are texture measures. The procedure and the technique for incorporating contextural information in segmentation are tested on a Black and White (B&W) digital image.

II. INTRODUCTION

Image segmentation is recognized as an important task in an image understanding system. Identification and proper labeling of regions in the scene, which correspond closely to the human perceptual interpretation is the primary objective of the segmentation procedure. The primitive operators used in procedure developed are texture measures derived from cooccurrence matrices. In particular, the spatial gray level dependance method (SGLDM) is utilized because theoretical studies have shown them to be superior and experimental studies on real world images including medical imagery and high resolution aircraft and satellite imagery have indicated their utility in characterizing complex scenes. Also, it has been shown that the SGLDM matches a level of human perception process.

An effective image segmentation methodology requires ability to incorporate contextual information in the processing of a scene element. The procedure developed in this research has the following attributes:

i) The method employs a split-and-merge type of approach.

ii) The procedure incorporates a multicategory Baysian classifier to label scene elements which do not require any further processing [1].

*This work is supported in part by U.S. Army Grant #DAAG23-82-K-0189.

iii) The procedure utilizes Kalman filter approach to realize a relaxation scheme which enables incorporation of the contextual information. Such an approach provides a statistical decision making strategy and effective termination criteria to indicate whether or not further processing at a higher level of segmentation is required for a scene element. Since Kalman filter provides a globally optimum solution, such an approach seems attractive.

The filtering approach treats two sets of information at a particular level: The class-conditional probabilities of scene elements, which can be termed as an intrinsic descriptor of the scene element and the joint-conditional probabilities, which can be termed as an extrinsic descriptor for a scene element since it depends on the neighborhood, *contextual*, information. These probabilities are assumed as independent random variables. Given the extrinsic description of a scene element the intrinsic descriptor can be estimated for each scene element, such that the estimation error is minimized. Kalman filter, therefore, seems like most appropriate method for the stochastic, context-based decision making.

iv) The last attribute of the segmentation is that the above implementation is completely adaptive, for general purpose applicability.

III. KALMAN FILTER APPROACH

MATHEMATICAL MODEL

The intrinsic information about the image elements can be derived by using a classification algorithm. For this study a multicategory classifier was used [1]. The external information is based on relational information between the elements of an image and the neighbors of them. This can be expressed as conditional probabilities of the image elements given the neighbors of them. The external information may be what is derived from the visual process of skilled human observer or what is calculated over the entire image under consideration. One such example of the former is well described in ref. 2 and of the latter will be described in this paper using the global statistics of the image which is useful for improving the local description in the later stage.

We have two probabilities of each element, both of which are corrupted by the noise. Since these probabilities can be treated as a random variable, the remaining problem is how to get an estimate which combines both probabilities such that the cost of the estimation will be minimal in a statistical sense. To solve this problem, we will use Kalman filtering. We will regard the intrinsic probabilities of image elements as measured quantities, while those calculated from the external information as quantities calculated through assumed stochastic processes among image elements.

It is assumed that the probability random process is adequately described by the state variable dynamic model.

$$P_i(w_k) = \sum_{j \varepsilon 0_i(w_k)} \sum_{w \varepsilon \Omega} [P_{ij}(w_k|w)P_j(w)+u_{ij}(w_k|w)]$$

$$q_i(w_k) = P_i(w_k) + v_i(w_k)$$

i is an index for an image element and is normally two dimensional coordinates, but it can be a node in a semantic graph; $0_i(w_k)$ is a collection of sets of neighbors which have relations with the element, i, when i has a class label w_k, j is an element of $0_i(w_k)$ and has m-1 elements if m'th order statistics is assumed; Ω is a collection of sets of class labels; w is an element of Ω, each element of which corresponds to the class label of the element of j. The random process, $P_i(w_k)$ is the probability with known statistics that i has a class label w_k; $P_{ij}(w_k|w)$ is the conditional probability that i belongs to class w_k, given a set of neighbors, j, which has a set of class label w; $P_j(w)$ is the joint probability with known statistics that the set j has a set of class label w; $q_i(w_k)$ is the measured quantity of $P_i(w_k)$. The noise random process, $u_{ij}(w_k|w)$ explains the inadequacy of the model of the external information and the noise process, $v_i(w_k)$ interprets the noisy component in the intrinsic information. Both noise processes are white noise processes with known statistics. The random processes, $P_i(w_k)$, $P_j(w)$, $u_{ij}(w_k|w)$, and $v_i(w_k)$ are statistically independent. It is also assumed P_{ik}, P_{ij}, and $P_j(w)$ are spatially invariant. Indices ij in $P^j_{ij}(w_k|w)$ is used for spatial relationship in the neighborhood but $P_{ij}(w_k|w)$ is the same over the image. It is worth noting that the equation can be expressed with the matrix if we regard k as a row index and an index of lexicographical ordering of w as a column index. In the real situation, it is difficult to measure $P_j(w)$ and its statistics. In addition to this reason, for notational simplicity, we will only consider the case of second order statistics. Extension of this result to the general case is straighforward. With the notations,

$$P_i = [P_i(w_1) \ldots . P_i(w_L)]^T$$

$$u_{ij} = [u_{ij}(w_1) \ldots . u_{ij}(w_L)]^T$$

$$q_i = [q_i(w_1) \ldots . q_i(w_L)]^T$$

$$v_i = [v_i(w_1) \ldots . v_i(w_L)]^T$$

$$P_{ij} = [P_{ij}(w_k|w)],$$

we have

$$P_i = \sum_{j \varepsilon 0_i(w_k)} [P_{ij}P_j + u_{ij}]$$

$$q_i = P_i + v_i.$$

P_i, q_i, u_{ij}, v_i are L-dimensional vectors, among which P_i and q_i are stochastic vectors; the sum of their components are equal to one. These vectors are mutually orthogonal in terms of inner product defined by

$$(y,z) = E[y^T z]$$

with known statistics. Particularly, the covariance matrices of the two white noise processes are

$$(u_{ij},u_{ij}) = Q_{ij}$$

$$(v_i,v_i) = R_i.$$

P_{ij} is a L x L matrix whose (k,l) element is represented by $P_{ij}(w_k|w)$.

LINEAR ESTIMATION

Given the preceding model, we can now determine an estimate $\hat{P}_i$ for P_i. Which is a linear combination of $\hat{P}_i$'s and q_i. In the following, we will treat the intrinsic probability in two different ways; if it is with the image element under consideration, it will be regarded as the measured quantity, q_i and if it is with a neighbor of this element, it will be regarded as the estimate available from the former processing, $\hat{P}_i$. The estimate should be optimal in the sense that the expected value of the sum of the squares of the errors in the estimate is a minimum. The objective function is naturally defined as

$$J[\hat{P}_i] = E[(\hat{P}_i - P_i)^T(\hat{P}_i - P_i)].$$

Our goal is to express $\hat{P}_i$ as a linear combination of $\hat{P}_j$

$$\hat{P}_i = \sum_{j \varepsilon 0_i(w_k)} P_{ij}\hat{P}_j + K_i[q_i - \sum_{j \varepsilon 0_i(w_k)} P_{ij}\hat{P}_j].$$

The unknown matrix K_i is L x L and is called the

Kalman gain matrix. For notational simplicity, we use the trace of the error covariance matrix, C_i, instead of $J[\hat{P}_i]$.

$$J[\hat{P}_i] = \text{trace}[C_i] = \text{trace}E[(\hat{P}_i - P_i)(\hat{P}_i - P_i)^T].$$

The above has its minimum

$$C_i = C_i' - K_i C_i', \text{when } K_i = C_i'(C_i' + R_i)^{-1}$$

where

$$C_i' = \sum_{j\varepsilon O_i(w_k)} [\hat{P}_i - P_i)(\hat{P} - P_i)^T P_{ij}^T - Q_{ij}].$$

We have not yet considered the constraint that $\hat{P}_i$ should be a stochastic vector, that is

$$\sum_{k=1}^{L} \hat{P}_i = 1, \text{ and } 0 \leq \hat{P}_i.$$

The first constraint requires that the solution vector lie on the coset of the subspace, $\sum_{k=1}^{L} \hat{P}_i = 0$ of the solution space of $\hat{P}_i$. We define the linear transformation, A_i from the solution space to the subspace. Then, the projection T_i of $\hat{P}_i$ into the coset is expressed as

$$T_i\hat{P}_i = A_i\hat{P}_i + a,$$

where a is L-dimensional vector, every element of which is $\frac{1}{L}$.

It can be shown [1] that if $\hat{P}_i$ has some negative components, by setting those components equal to 0 and subtracting these total increases divided by the number of the rest components from each of the rest components, we can get the minimum C_i. The above mentioned operation may be expressed by the diagonal matrix with 0 in the places corresponding to the negative components and the ratios of the new values to the old in the places corresponding to the rest ones. Let B_i represent such an transformation, then the final estimate and its error covariance becomes

$$\hat{P}_i = B_i T_i \sum_{j\varepsilon O_i(w_k)} P_{ij}\hat{P}_j + K_i[q_i - \sum_{j\varepsilon O_i(w_k)} P_{ij}\hat{P}_i$$

$$C_i = (C_i' - K_i C_i')A_i^T A_i B_i B_i^T.$$

TERMINATING CONDITION

Basically, our Kalman filter performs the estimation by processing the intrinsic and the external information just once. However, to the updated intrinsic information which hopefully contain less noisy components than the original, we may apply the filter iteratively to get the better estimate. But this iterative information can be modeled by m'th order statistics since from the second processing, we are using neighbors which have been estimated with the elements beyond the proposed neighbors.

A possible justification for the iterative processing is based on the fact that our external information is not complete. Especially, when we use global statistics over the image as external information, it is probable that we can get improved external information with the updated data. If we have a complete external description from the beginning, the process should be performed just once. Hence, it seems to be natural to terminate the processing when the possible error covariance with the perfect external information is achieved. In this case, $u_{ij}(w_k|w) = 0$ and as a bound for the error covariance matrix, we have

$$D_i = (D_i' - K_i D_i')A_i^T A_i B_i B_i^T$$

where

$$D_i' = \sum_{j\varepsilon O_j(w_k)} P_{ij}(\hat{P}_i - P_i)(\hat{P}_i - P_i)^T P_{ij}^T$$

We can also use the confidence level as a terminating criterion. If the error variance of the largest component of the stochastic vector is σ_1^2 and that of the second is σ_2^2, we may use some constant multiple of $(\sigma_1 + \sigma_2)$ as the criterion to guarantee the desired confidence level. In the experiment, we have used both criteria. The process terminates whenever

$$\text{diag}[C_i - k_1 D_i] \leq 0 \text{ and } c_{ii} - c_{jj} \geq k_2(\sigma_1 + \sigma_2),$$

where k_1 and k_2 are some constants which are determined considering the tradeoffs between the computational effort and the desired accuracy of the classification, and c_{ii} and c_{jj} are the variance of the first and the second largest components of the stochastic vector.

IV. IMPLEMENTATION OF THE KALMAN FILTER

With the estimated probabilities of image elements as the intrinsic information, we implemented the Kalman filtering on the digital computer to apply it to the real world imagery. Our image elements were produced by a set of equally spaced grids over the image. From the training stage, we collected the statistics of the intrinsic of v_i.

Since we feel that the classes of the image elements can be divided into the two major categories, enlongated and large homogeneous areas, we used different approaches for each category.

For the elongated area, we assume general N'th order statistics since we are interested in determining the location of the element which will fill a broken part of the area. For the large homogeneous area, on the other hand, we assume the second order statistics between the center element and the eight nearest neighbors under the stochastic vector assumption. The image elements determined as belonging to elongated areas are not considered any more in processing the large homogeneous area. The treatment of elongated areas are not presented here for the sake of brevity. Interested reader is referred to reference 1 for details. We use only the statistics of the training stage and the apriori knowledge, what category each class belongs. Both are normally available in a supervised learning. Other information is collected adaptively from the whole image.

Large homogeneous areas are dealt with as an estimation problem. To solve this problem, we need methods to calculate the conditional probability, P_{ij}, and the error covariance of u_{ij}. An external information characterizing this category may contain the convexity and the size of the area. A part of this information seems to be reflected in the average number of neighbors which have the same class label as that of the center image element in a fixed size neighborhood.

For each image element, we count the elements in its neighborhood which have the same class label as that of the element under consideration. Then, these counted numbers are averaged for each large homogeneous class over the whole image. The inverse of the closest integer to the average is used as the elements of P_{ij}.

$$P_{ij} = \mathrm{diag}[P_{11},\ldots,P_{LL}]$$

$$P_{ii}: 1/N_i$$

N_i: average number of elements in the neighborhood of the elements with class label i.

Note that we are using a diagonal matrix, P_{ij} to reduce the calculation; the external information between the elements with different class labels is not considered. To get more reliable conditional densities, we may use only the elements, which has relatively higher class conditional probability, as the center element. With these conditional probabilities and the class conditional probabilities of the neighbors, we estimate the class conditional probability of the center element. Then, we calculate the error covariance of this estimate over the whole image. Note that we are not considering the noise processes and this estimate is not based on the Kalman filtering. This process is just for obtaining the statistics of u_{ij}.

$$Q_{ij} = E[(P_i - P_{ij}P_j)(P_i - P_{ij}P_j)^T]$$

With P_{ij} and the covariance of u_{ij}, we apply the Kalman filter to each image element. To ensure the stochastic vector assumption, we use N_i neighbors for the class, w_i. These neighbors are selected to be evenly distributed around the center element and to have the highest class conditional probabilities among those evenly distributed sets of neighbors. Note that each class need not have the same number of neighbors. This approach keeps the small area, composed of the elements from the class whose area size is normally small, from disappearing on the image, while regarding the small area of the relatively large-size class as a noise. With this approach, the element in the interior of the large homogeneous area will tend to have a high class conditional probabilities; the class conditional probabilities of the element of the boundary will be changed to those which agrees more with the external information based on the statistics of the shapes of the competing classes. After the Kalman filtering, we project the $\hat{P}_i$ on the hyperplane defined by the stochastic vector assumption. The whole processes are iterated until the terminal condition is met.

V. EXPERIMENTAL RESULTS

Experiments have been performed on a B&W aerial image [3]. Training samples were taken from the same image of estimate the parameters of the multicategory classifier and to collect the statistics of the noise random process v_i. The same statistics was used as the initial statistics for the probability random process P_i. We divided the image into 34 x 34 polygons, the width and the height of which are 145 pixel spacing, each. With these parameters and statistics, the class conditional probabilities of each polygon were estimated. Kalman filter was applied to the class conditional probabilities of each polygon.

The dimension of the original observation space is 198. The construction of the observation space based on texture measures is explained in the reference [3]. Training results of the multicategory classifier are given in Table 1.

Table 1. Training results derived using the multicategory classifier.

Class#	Class Name	No. of Samples	Classification Accuracy (%)
1	Residential Area	83	77.1
2	Commercial Area	328	85.7
3	Mobil Home	41	95.1
4	Vehicle Parking	25	68.0
5	Dry Land	353	94.6
6	Water	85	98.8
7	Runway	78	87.2
8	Aircraft Parking	21	85.7
9	Highway	124	91.9
Total		1138	89.5

89.5 percent of samples are correctly classified, whereas the results by the pairwise linear discriminant function [3] shows 90.3 percent of correct classification.

Before applying the Kalman filter, we deleted the classes 4 and 8 and included them into class 2 and 7, respectively, since these classes contain too few training samples. It seems to be reasonable that vehicle parking is contained in commercial area and airplane parking is contained in airport. The new assignment of classes is presented in Table 2.

filter be applied only to those classes with the meaningful number of samples to acquire the statistics from. Improvement of contextual information is reflected by the decreasing error covariance as the filtering process progresses.

For the analysis we used the verified classes (ground truth) as of the polygons in the reference [2]. These classes, however, do not reflect any contextual information. Our segmentation is intended as a global one, leaving further detailed description of the image for further processing which can use apriori knowledges about the possible image elements in these global classes. One can argue, for example, why a small portion of dry land in a commercial area should be called as dry land. For commercial area normally includes some part of dry land. Because of similar reasons, we modified the verified classes based on the following rules:

1) Commercial area mixed with dry land is called as commercial area.

2) Residential, commercial, and dry land classes mixed with road classes are called residential, commercial, and dry land classes, respectively if they are not required to connect the road classes.

3) If polygons of mixed classes in the interior of a large homogeneous area include a part of the surrounding area they are assigned to this area.

Table 2. Classification results acquired with the multicategory classifier and by applying four iterations of Kalman filter.

Class No.	Class Name	No. of Samples	Classification (%) Correct Initial	Correct After Filtering	Incorrect Initial	Incorrect After Filtering	No label Initial	No label After Filtering
1	Mixed	136	63.2	85.3	36.8	14.7	0	0
2	Residential Area	50	50.0	48.0	50.0	28.0	0	24.0
3	Commercial Area	364	61.0	72.3	39.0	13.5	0	14.3
4	Mobil Home	9	88.9	44.4	11.1	0.0	0	55.6
5	Dryland	325	85.8	70.5	14.2	17.2	0	12.3
6	Water	104	84.6	88.5	15.4	6.7	0	4.8
7	Airport	81	84.0	93.8	16.0	0.0	0	6.2
8	Highway	87	80.0	87.4	19.5	4.6	0	8.0
Total		1156	73.2	76.1	26.8	13.0	0	10.9

To begin the processing the Kalman filter was applied to class 8, the highway class which is the only elongated area in the image. 75 percent of the samples which were classified as highway class and with larger class conditional probabilities were used. The noisy polygons were deleted in the selection process of 75% highway samples. Among 87 highway polygons, 18 polygons were connected in filtering process: 12 of them were in correct location, 5 of them were deviated one polygon spacing, and 1 of them was deviated by two polygons from the correct position.

Except for the polygons which were classified as the elongated area, others went under the large homogeneous area processing. We have used 1.5 and 1.0 for K_1 and K_2 (see section III). This roughly corresponds to improving the quality of the external information at least twice as good as that of the initial information. The conditional probabilities were calculated with 50% of samples. It was observed that the smallest class, the mobile home class, has the largest error covariance. We suggest that the Kalman

The results are presented in Table 2. For the mixed class. We counted as a correct classification polygons which were assigned to the mixed class or one of the classes mixed for comparison purpose. After terminating at the fourth stage of the Kalman filter, percentage of correct classification is improved from 73.2 to 76.1 percent. More importantly 26.8 percent of misclassification is reduced to 13.0 percent providing unclassified samples for further processing. The reduced classification accuracy of the dry land class needs to be explained. There are many small dry land areas with one or two polygons in the interior of other large homogeneous areas which disappear from the image after the filtering. However, note that these small dry lands can be assigned to the homogeneous area surrounding them, since dry land is one of the elements which most homogeneous areas are composed of. Furthermore, the dry land class suffers more from the boundary effect than any other class. Drastic change of shape of this class over the image is another reason fo the poor classification results. The results of each stage of the Kalman filtering reflects the inevitable trade-offs between increasing the classification accuracy and decreasing the misclassification.

VI. CONCLUSIONS

We have described a new approach to a relaxation algorithm based on the stochastic dynamic model and developed the adaptive Kalman filtering procedure under the stochastic vector assumption. This procedure has an advantage over the existing relaxation algorithm in that we can incorporate statistical apriori knowledge of the intrinsic and external information into the estimation process.

Results of applying Kalman filter to the image segmentation is generally good but in most cases seem to be limited by the boundary effect and the extreme changes of the statistical properties over the image. While the initial results are rather encouraging, further work remains to improve the efficacy of this approach.

VII. REFERENCES

1. Goh, S., "Developing an Adaptive Kalman Filter for Image Segmentation Using a Stochastic Dynamic Scene Model," M.S. Thesis, Electrical and Computer Engineering Department, Louisiana State University, Baton Rouge, April 1983.

2. O. D. Faugeras, M. Berthod, "Improving Consistency and Reducing Ambiguity in Stochastic Labeling; An Optimization Approach," IEEE Trans. PAMI-3, NO. 4, pp. 412-424, July 1981.

3. Conners, R. W., M. M. Trivedi, and C. A. Harlow, "Segmentation of a High Resolution Urban Scene Using Texture Operators," Computer Graphics and Image Processing, to appear 1983.

AUTHOR BIOGRAPHICAL DATA

Semoon Goh, received his B.S.E.E. degree from the Seoul National University, Korea, in 1972 and his M.S. in electrical engineering from the Louisiana State University, Baton Rouge, in 1983. Since January 1983 he is working on a Ph.D. program in electrical engineering at the Purdue University.

He worked for the Korean Agency for Defense Development from 1972-1980. His research interests are Image Analysis and Mathematical System Theory. He is a student member of the IEEE.

Mohan M. Trivedi, received his B.E. (Hons) in Electronics from the Birla Institute of Technology and Science, Pilani, India, in 1974 and M.E. and Ph.D. degrees in electrical engineering from Utah State University, Logan, in 1976 and 1979, respectively.

Since 1979 he has been with the Department of Electrical Engineering at Louisiana State University as an Assistant Professor. His research interests are pattern recognition and applications of remote sensing. He is a member of IEEE, Tau Beta Pi, Phi Kappa Phi, and the Pattern Recognition Society.

Charles A. Harlow, was born in New Boston, TX, on March 14, 1940. He received the B.S. and Ph.D. degrees from the University of Texas in 1963 and 1967, respectively, both in electrical engineering.

In 1967 he joined the staff of the Department of Electrical Engineering at the University of Missouri, Columbia, as Assistant Professor. In 1972 he was appointed Professor of Electrical Engineering. In 1974 he was visiting Professor of Electrical Engineering and Computer Science at the University of California, Berkeley. In 1978 he joined the faculty of Louisiana State University as Professor of Electrical Engineering. His principal areas of research interest are image analysis and software design.

Dr. Harlow is a member of Tau Beta Pi, Eta Kappa Nu, the Association for Computing Machinery and the American Society of Photogrammetry.