XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)
CIMNE, Barcelona 2010

REPRESENTATION OF MULTIPLE HYDROSTRATIGRAPHIC REGIONS FOR GROUND WATER REMEDIATION DESIGN SUBJECT TO UNCERTAINTY

Karen L. Ricciardi*

* University of Massachusetts in Boston
100 Morrissey Blvd., Boston, Massachusetts, 02184, USA
e-mail: Karen.Ricciardi@umb.edu, web page: http://www.umb.edu

Key words: Stochastic methods, uncertainty, remediation, optimization, heterogeneous, ground water.

Summary. This research involves the analysis of the effects of uncertainty in the boundary of heterogeneous regions on the determination of least-cost pump-and-treat remediation designs for contaminant containment problems.

1 INTRODUCTION

Predictive groundwater models are used for many different reasons in the field of hydrology. Decisions made in water resources that are based upon ground water models require that models are accurate and reliable. Two key parameters for ground water models that are often inaccurate and compromise the accuracy of the models are the spatial variability in the hydraulic conductivity of the aquifer and the location of the boundaries of hydrostratigraphic regions within the model[6]. This research is an effort to understand the impact these uncertainties have on the determination of reliable least-cost pump-and-treat ground water remediation designs.

A common approach taken to including spatial variability of hydraulic conductivity fields into optimal ground water management problems called a multi-scenario approach is one that involves analyzing ground water flow and contaminant transport on multiple hydraulic conductivity fields, or scenarios, that are possible realizations of the true field[8]. These fields are generated using the geospatial statistical parameters that are ascertained from data. Using a multi-scenario approach to representing the uncertainty, a number of different methods have been used to determine reliable remediation designs that are efficient[1].

When the hydraulic conductivity field is heterogeneous and involves hydrostratigraphic fields that have distinct hydrologic properties characterized by distinct hydraulic conductivity values, the locations and orientations of the boundaries of the hydrostratigraphic fields can have a significant impact on the reliability of management designs. These effects have been shown to exist in the determination of reliable pump-and-treat remediation designs for containment where each field is spatial variability but the boundaries of the fields are assumed to be fixed[7].

In this work, the uncertainty in the locations of distinct hydrostratigraphic regions is analyzed with respect to a pump-and-treat ground water remediation design. The determination of a set of scenarios that represent the uncertainty in these boundaries uses a novel approach that involves techniques in image processing to segment the hydraulic conductivity field called Delineation of Boundaries using Image Segmentation (DoBIS). Coupling these methods in image processing with statistical techniques, a set of realizations of different heterogeneous hydraulic conductivity fields is generated. Optimal cost pump-and-treat ground water remediation designs for contaminant containment are determined for each scenario, and the variability of these designs are analyzed. The results are compared to the distribution of results obtained with fixed boundaries but spatially variable fields. Through this comparison one draws conclusions about the relative importance of the uncertainty in the boundary of the hydrostratigraphic regions versus the spatial variability in the fields with respect to determination of reliable remediation designs. To better understand how the relative importance is impacted by the locations and orientations of the hydrostratigraphic fields to controls in the remediation design, namely the hydraulic head constraints necessary for containment and the possible well locations, four different heterogeneous fields of hydraulic conductivity are examined.

2 METHODS

2.1 Ground water flow model

A common pump-and-treat ground water remediation designs is one that contains the flow of a contaminant plume. Determining a least-cost design that meets the imposed flow constraints for containment requires solving the partial differential equation that describes the dynamics of ground water flow in a fully saturated, three-dimensional, porous media:

$$\nabla(\boldsymbol{K} \cdot \nabla h) = S_s \cdot \frac{\partial h}{\partial t} + f \qquad (1)$$

where ∇ is the spatial differential operator, $\boldsymbol{K}$ is the hydraulic conductivity tensor (L/T), h is the hydraulic head (L), S_s is the specific elastic storage (1/L), t is time (T) and f represents a source/sink (per unit volume) term (1/T). It is within the source/sink term, f, that the pumping rates are prescribed.

To solve flow equation for a given $\boldsymbol{K}$ field and pumping design in space and time, the ground water flow simulator MODFLOW-2000 is applied[4]. This simulator takes a finite difference approach to solve the flow equation. The problems examined in this work seek steady-state solutions, and the aquifers in the model are assumed to be confined. Under these conditions the head response to pumping rates is linear allowing for a response matrix approach to be utilized[3].

2.2 Optimal cost pump-and-treat design

In this model the cost associated with the remediation design is assumed to be directly proportional to the total sum of pumping from all the wells. A fixed number of wells are considered at specified locations. Flow constraints are placed upon the system so that the gradient of the hydraulic heads at specified locations is towards the wells. The optimization problem that minimizes the cost associated with this remediation plan is as follows:

Objective: $$\min \sum_{i=1}^{n} q_i \quad (2)$$

Subject to: $$g_j > \max g, \;\; j = 1..m \quad (3)$$
$$0 \le q_i \le \max q, \quad i = 1..n$$

where q_i is the pumping rate at well i, n is the total number of wells considered, max g is the maximum allowable gradient so that flow is towards the well at any of the m constraint locations; g_j is the gradient at the constraint location j given a pumping design of $\boldsymbol{q} = (q_1, q_2, \ldots, q_n)$; and max q is the maximum amount of pumping from any of the wells. Here pumping at any specific well, q_i, must be positive. This condition implies that only extraction wells are considered (no injection wells are considered) in the resulting remediation plan.

The gradient constraint is defined in terms of the finite difference mesh utilized in the ground water simulator. At any of the constraint locations, j, the gradient, g_j, is modeled by the difference in hydraulic head values at adjacent nodes oriented such that flow from the outer node, k, to the inner node, $k-1$, would be towards the well. Then $g_j = (h_{j,k} - h_{j,k-1})/\Delta x$, where $h_{j,k}$ and $h_{j,k-1}$ represent the hydraulic head values at adjacent nodes of the mesh associated with constraint j, and Δx represent the spatial distance between the adjacent nodes.

2.4 Genetic Algorithm

A genetic algorithm is employed to solve the optimization problem. A discrete and finite set of permissible pumping rates for each well is defined. The permissible rates are between 0 and max q. Each member of each generation of the population in the genetic algorithm represents a possible remediation design plan, i.e. combination of pumping rates at the proposed well locations. Each member of the initial population is determined by randomly selecting the pumping rates for each well from the set of permissible pumping rates. A fitness value is determined for each member of the population. The fitness value is equal to the sum total of the pumping rates at the wells, plus a scalar multiple of the sum of the violations of the gradient constraints at the constraint locations should the given pumping design fail to meet the constraints. This fitness value is thus defined by the following equation:

$$\sum_{i=1}^{n} q_i + \omega \sum_{j=1}^{m} \max\{0, \max g - g_j\} \quad (4)$$

where all variables are as stated previously, and ω represents the scalar multiple of the violations of the constraints called a penalty weight. Subsequent generations are determined by employing the rules of elitism, mating priority, crossover, mutation and random selection used in a genetic algorithm. The genetic algorithm continues to create new generations of populations of pumping combinations until the best pumping combinations in sequential generations repeat a specified number of times.

2.5 Representation of uncertainty

Two types of uncertainty are considered in this work: the spatial variability of hydraulic conductivity and the uncertainty in the locations of the boundaries of distinct hydrostratigraphic units in a heterogeneous model. Both types of the uncertainty are represented using a multi-scenario approach, whereby multiple realizations of the ***K*** field are generated using assumed known information that characterizes the uncertain parameter.

2.5.1 Spatial variability

To generate those fields representative of the spatial variability of ***K***, variograms are generated using the SGSIM package of the geostatistical software package GSLIB[2]. Simple Kriging is applied to obtain a field that follows a Gaussian distribution that is related to the lognormal distribution of the ***K*** field in each of the hydrostratigraphic regions. The boundaries of the regions are the same for each realization.

2.5.2 Boundary uncertainty

The fields representative of the uncertainty in the boundaries of the hydrostratigraphic regions were generated using the newly developed data driven approach called Delineation of Boundaries using Image Segmentation (DoBIS).

Segmentation methods refer to those methods in image processing whereby the colors assigned to each pixel in an image are categorized by particular groups[9]. To apply segmentation methods to determine the boundaries of the hydrostratigraphic regions, the number of regions in a given model is first determined. Data points, representative of measured ***K*** values at fixed observation locations in a heterogeneous aquifer are each assigned an integer value based upon the region in which the data is assumed to be sampled. The order of the value assigned is consistent with the ordering of the mean ***K*** values for each unit. For example, if there are 4 units with mean ***K*** values of 0.001, 0.01, 0.1 and 1.0 *m/d*, the unit with mean ***K*** of 0.001 *m/d* would be assigned the category value of 1, with mean of 0.01 *m/d* the value of 2, etc. If the hydrostratigraphic fields are representative of distinct geologic regions, such as a sand unit and a silty-sand unit, this process of categorizing the data is not difficult because the ***K*** values will differ by orders of magnitude and membership in a particular field, sand or silty sand, will be easy to determine. If the hydrostratigraphic units are not distinct, then techniques for determining the modes of a multi-modal distribution must be employed and membership of the data to a given distribution must be based upon likelihood measures[5].

To determine the location of the boundaries of the hydrostratigraphic units in the modeled region, a category value is interpolated for all nodes of the finite difference mesh where data does not exist. The method used to interpolate the data is that of universal kriging which utilizes a regression model to obtain a surrogate for the ***K*** category field. Here a linear regression model is utilized with a Gaussian correlation function. The correlation factor for the Gaussian model is 1 with an upper bound on the correlation distance of 5 nodes of the finite difference mesh. Through kriging in this manner, each node of the finite difference mesh is assigned a numeric value.

Different contour curves of the kriged category field are used to generate multiple realizations of heterogeneous ***K*** fields that represent the uncertainty of the boundary locations of the distinct hydrostratigraphic units. The locations of those contours associated with values that are midway between two sequential category values are said to represent the boundary locations that are most likely. For example, if two category values are 1 and 2, then this contour would be associated with the value of 1.5 in the kriged field. To generate a set of realizations representative of the uncertainty in the boundary locations, a truncated normal distribution with a mean of zero and standard deviation of one is randomly sampled. This distribution is truncated below at -1.7 and above at 1.7 so that the sampled values account for 80% of the normal distribution. The contour level, c_{bnd}, that determines the boundary between the hydrostratigraphic fields associated with the two categories is given by the following:

$$c_{bnd} = \frac{1}{3.4}\left(c_{\propto} - c_{\beta}\right)(r - c_{\propto}) + 1.7 \qquad (5)$$

where c_α and c_β are the integer category values of adjacent regions, α and β, in space.

3 EXAMPLES

3.1 Ground water flow model

Four hypothetical ground water flow models are examined. These models represent four different geologic settings. They are single layer models representative of a 1000*m* by 1000*m* region that is 100*m* thick. No flow boundaries exist on the northern and southern boundaries of the model, while constant head boundaries of -20*m* and 25*m* are set at the western and eastern boundaries respectively. The models represent confined aquifers run to simulate a steady-state condition. A uniform finite difference mesh of 50 nodes by 50 nodes is defined over the given area.

Sample	*i*	*j*	$K_A(m/d)$	$K_B(m/d)$	$K_C(m/d)$	$K_D(m/d)$
1	5	40	0.0103	0.95	0.0100	0.0097
2	7	6	0.96	0.92	1.05	0.0095
3	8	16	1.03	1.09	1.13	0.0096
4	12	33	0.0105	0.85	0.0098	0.0102
5	13	44	0.0095	1.02	0.0111	0.0095
6	18	11	0.95	1.01	0.95	0.0106
7	23	28	0.0076	1.08	0.0092	0.0104
8	30	6	0.94	0.0076	0.0102	0.98
9	31	41	0.0099	0.0089	0.0103	0.0102
10	38	17	1.12	0.0111	0.0094	0.93
11	39	44	0.0116	0.0095	0.0092	0.0106
12	45	33	0.0092	0.0097	0.0106	0.0103
13	46	4	1.18	0.0113	0.0093	0.99

Table 1. Measured *K* values at the given *i, j* index of the nodes in models A, B, C and D.

A hypothetical data set is said to exist for each of the four models consisting of 13 measured data values (Table 1). Two hydrostratigraphic units are defined for each model representative of a sand unit with a mean ***K*** value of 1.0*m/d* and a silty-sand unit with a mean ***K*** value of 0.001*m/d*. The locations of the measurement are the same for all the models, but the values of ***K*** observed differ. Utilizing the segmentation algorithm, the most likely location of the boundaries of the hydrostratigraphic units for these models is illustrated in Figure 1.

To represent the spatial variability in the hydraulic conductivity fields, 100 lognormal, spatially variable fields are generated using the SGSIM package of GSLIB. The associated normal distributions for the ***K*** fields of the distinct hydrostratigraphic units both have a standard deviation of one. The mean values of the lognormal distributions, however, differ in accordance with the mean values of the measured data of 1.0 and 0.01 *m/d*. The boundaries of the hydrostratigraphic units are the same for the multiple scenarios generated to represent the spatial variability for each of the models. An example of three different spatially variable fields for one of the Model D is illustrated in Figure 2.

A set of 100 realizations of the ***K*** field are generated using the segmentation algorithm for each model. These sets represent the uncertainty in the location of the boundaries of the units. A sample of three such fields for one of the models is illustrated in Figure 3.

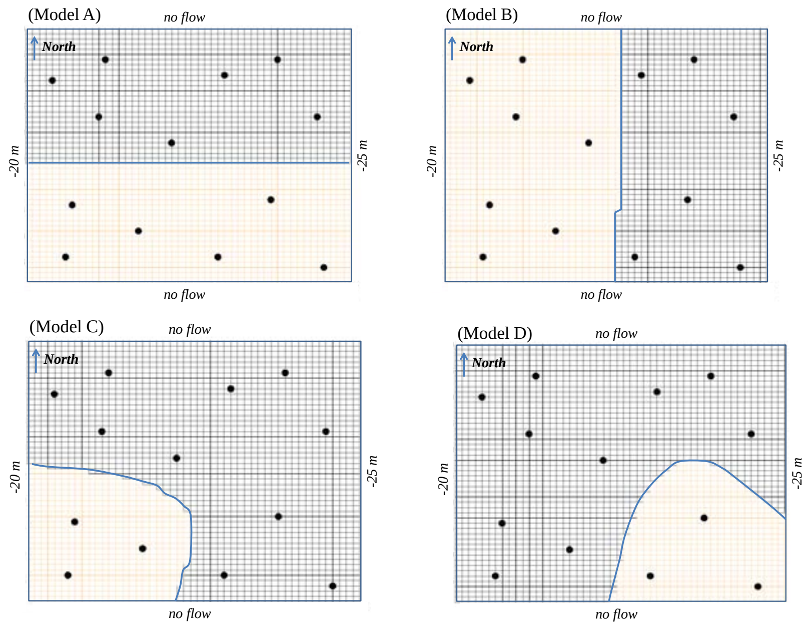

Figure 1. Models A, B, C and D with sample value locations, dots, and most likely boundary locations of perfectly homogeneous hydrostratigraphic ***K*** fields. The lighter and darker regions represent hydrostratigraphic regions with a mean ***K*** value of 1.0 and 0.01 *m/d* respectively.

3.2 Remediation design

A least-cost pump-and-treat remediation design is sought for each of the models that will contain contaminated ground water that is flowing from a point source along the western boundary of the model towards the east. To meet this goal, 14 gradient constraint locations and four possible pumping wells are positioned within the models as illustrated in Figures 2 and 3. The pumping rates are bounded between 0 and 300 m^3/d. Contaminant plume containment is achieved when max g is set equal to 0.01m.

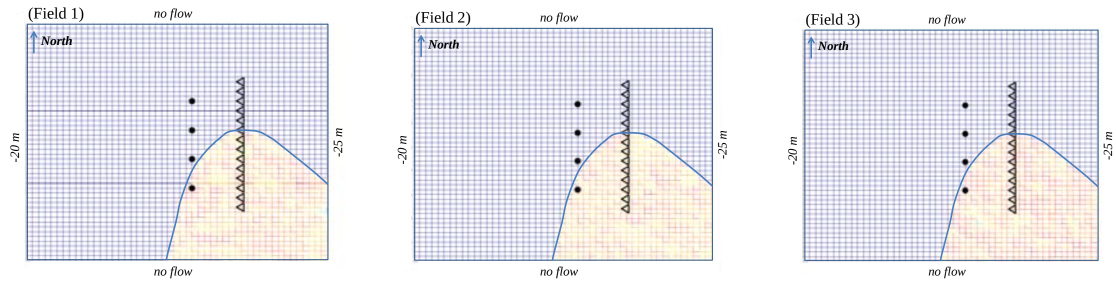

Figure 2. Three possible spatially variable fields for Model D with remediation plan superimposed. The dots represent possible pumping well locations and triangles represent gradient constraints. The lighter and darker regions represent hydrostratigraphic regions with a mean ***K*** value of 1.0 and 0.01 *m/d* respectively.

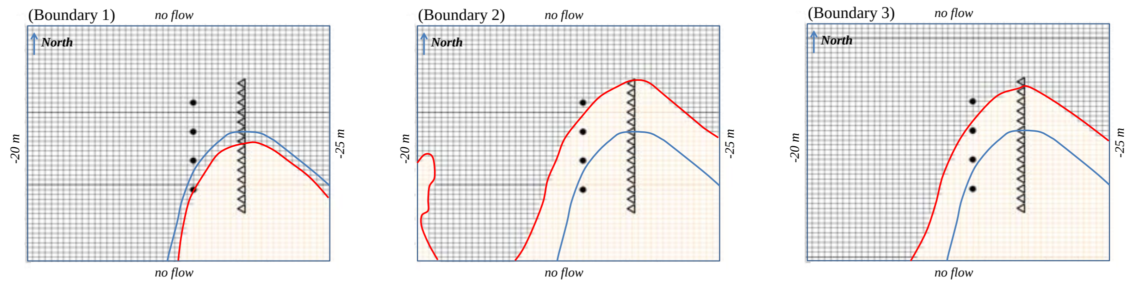

Figure 3. Three possible boundary locations for Model D with remediation plan superimposed. The dots represent possible pumping well locations and triangles represent gradient constraints. The blue contour line represents the most likely contour, while the red contour line represents that which was determined by randomly sampling a normal distribution. The lighter and darker regions represent hydrostratigraphic regions with a mean ***K*** value of 1.0 and 0.01 *m/d* respectively.

3.3 Genetic algorithm

The application of the genetic algorithm necessitates that q_i assume a fixed set of values. Here this set is: {0, 5, 10, 15,…, 295, 300m^3/d}. The penalty weight, ω, is 10^5. The population size for the genetic algorithm is 1000. The members of the population carried to the next generation through elitism consist of the top 10% of best fit members of the given population. The mating population consists of the top 50%, including the elite. Mating is conducted by applying one cross-over event with subsequent mutation occurring in 0.05% of the resulting pumping rates. Mating accounts for 70% of the population of the next generation. The remaining 20% of the next generation of pumping rates are created by randomly sampling the set of possible pumping rates for each well. The genetic algorithm creates new generations of populations the best fit pumping design for all of the members of the population repeats in 25 consecutive generations.

4 RESULTS

Assuming the hydrostratigraphic fields of a heterogeneous aquifer are perfectly homogeneous with hydraulic conductivity values of 0.01 and 1.0*m/d*. And assume the boundaries of these fields are fixed at the most likely locations. The least-cost pump-and-treat remediation solutions for containment are given in Table 2.

No uncertainty						
Model	Pumping(m^3/d)	Well 1	Well 2	Well 3	Well 4	Total
A		20	5	5	20	50
B		300	240	20	5	565
C		25	0	0	10	35
D		5	0	10	95	110
Spatially variable fields with fixed boundaries.						
A	Mean	20.10	4.80	4.20	22.20	51.30
	St. Deviation	0.70	2.92	3.87	3.04	2.20
B	Mean	298.15	246.45	28.05	5.95	578.80
	St. Deviation	4.59	26.77	27.72	4.80	29.72
C	Mean	24.25	1.50	0.35	9.05	35.15
	St. Deviation	1.79	3.59	1.63	1.97	0.86
D	Mean	5.18	3.59	2.41	98.29	109.47
	St. Deviation	0.93	2.26	4.27	4.47	3.94
Boundaries are uncertainty with homogeneous ***K*** fields.						
A	Mean	289.30	184.40	6.55	3.60	484.30
	St. Deviation	9.61	43.40	7.81	3.34	47.74
B	Mean	128.55	78.30	87.25	146.30	495.30
	St. Deviation	112.24	94.86	105.80	108.62	605.29
C	Mean	41.72	8.33	2.61	6.72	59.39
	St. Deviation	49.44	36.27	3.91	3.11	79.20
D	Mean	11.55	15.65	11.45	5.65	44.30
	St. Deviation	6.69	12.75	9.52	10.96	6.40

Table 2: Least-cost remediation designs where uncertainty is not considered; statistical means and variance of least-cost pump-and-treat remediation solutions for 100 heterogeneous models with spatial variability K fields; and statistical means and variance of least-cost pump-and-treat remediation solutions for 100 heterogeneous models with uncertainty boundaries of the hydrostratigraphic regions. Well 1 is the southern most well in the models while well 4 is the northern most. All values are pumping rates measured in m^3/d.

Taking into account spatial variability in the model, but keeping the boundaries of the hydrostratigraphic regions fixed, the mean pumping rates and standard deviations for the 100 scenarios considered are given in Table 2. The mean pumping rates are similar to the solutions for the scenarios where uncertainty is not considered. Only in Model D do the results vary at well 3. But the mean combined pumping from wells 3 and 4 in Model D in the spatially variable sets is the same as the combined pumping at in the case where uncertainty is not considered. The standard deviations of the pumping rates at the wells can be significant with respect the mean pumping rate when the boundary of the spatially variable fields is close to the well, as is the case in Models B, C and D.

Taking into account the uncertainty associated with the boundary of the hydrostratigraphic fields, but assuming the ***K*** fields over each region are perfectly homogeneous, the mean pumping rates and standard deviations for the 100 scenarios considered are given in Table 2. These results are drastically different than those of the case where uncertainty is not considered. As is observed in Figure 3, the uncertainty in the boundary of the fields can vary significantly over the sampled values. Some scenario will be such that the pumping wells are located in fields that are much different than when the most likely boundary is considered. These results suggest that from a planning perspective, it is extremely important to identify the boundary of the heterogeneous fields near the locations of the proposed well sites. Not obtaining this information can lead to highly unreliable predictions of the effectiveness of a remediation design and highly inaccurate expectations for pumping rates needed to meet the gradient constraints.

5 CONCLUSIONS

- The uncertainty of the boundaries of distinct hydrostratigraphic fields in a heterogeneous aquifer can be represented in a least-cost pump-and-treat remediation design plan through a multi-scenario approach.
- The scenarios used to represent the uncertainty in the boundaries of distinct hydrostratigraphic fields can be determined from the data by using the method called Delineation of Boundaries using Image Segmentation (DoBIS) herein described.
- Optimal pumping designs determined for the set of scenarios representing the uncertainty in the boundary of hydrostratigraphic fields are highly variable and suggest that the importance of reducing the uncertainty in the determining the location of these boundaries is extremely important from a management perspective.

REFERENCES

[1] Bau, D.A. and A.S. Mayer, "Optimal design of pump-and-treat systems under uncertain hydraulic and plume distribution", J. Cont. Hydrol., **100**(1-2), 30-46 (2008).

[2] Deutsch, C.V., and A.G. Journel, *GSLIB: Geostatistical Software Library*, Oxford University Press (1998).

[3] Gorelick, S.M., "A review of distributed parameter groundwater management modeling methods", Wat. Resour. Res., 19(2), 305-319, (1983).

[4] Harbaugh, A.W., E.R. Banta, M.C. Hill and M.G. McDonald, "MODFLOW – 2000, the U.S. Geological Survey modular groundwater model – User guide to modularization concepts and the Ground-Water Flow Process," *U.S. Geological Survey Open-File Report 00-92*, 121p. (2000).

[5] Le Gall, F., "Determination of the modes of a Multinomial distribution", *Statistic & Probability Letters*, **62**, 325-333 (2003).

[6] Nilsson, B., A.L. Hojberg, J.C. Refsgaard and L. Troldborg, "Uncertainty in geological and hydrogeological data," *Hydrology and Earth System Sciences*, **11**(1), 1551-1561 (2007).
[7] Ricciardi, K.L., "Remediation of Heterogeneous Aquifers Subject to Uncertainty," *Ground Water*, **47**(5), 675-685 (2009).
[8] Wagner, B.J., "Evaluating data worth for groundwater managment under uncertainty", J. of Wat. Res. Pl. and Management, 125(5), 281-288 (1999).
[9] Wilson, R., *Image Segmentation and Uncertainty (Pattern Recognition & Image Processing),* Research Studies Press, 1987.