M. TECH (Computer Science) Dissertation Series

Discrimination and Classification of Multispectral Remote Sensing Data

Dissertation submitted in partial fulfilment of the requirements for the M.Tech (Computer Science) degree of the Indian Statistical Institute

by

Subhasis Ray

under the supervision of

Dr. Swapan Kumar Parui

Indian Statistical Institute
203, Barrackpore Trunk Road
Calcutta - 700 035.

ABSTRACT

This is a study of mixture model approach to clustering. In mixture model approach, each observation is assumed to come from a mixture of a finite number of populations. If we can separate out these populations, each will represent a cluster. Here we learn the underlying parameters of the model and then classify accordingly.

Contents

1 Introduction

With the increasing interest on a model based approach to clustering, the use of finite mixture models for this purpose has been the subject of close scrutiny. Attention is concentrated on the fitting of mixture models (on synthetic as well as real-life data) by a likelihood based approach. This approach would appear in general to be superior to other methods of fitting mixture models, and the *EM Algorithm* [7] provides a convenient way for the iterative computation of solutions of the likelihood equation. The emphasis here is on mixtures of normal distributions as these models are most frequently employed in practice, as well as most widely studied.

Before discussing this approach in detail in Section 3, a brief review of some of the classical clustering principles (e.g. *K-means clustering* etc.) is presented in Section 2.

2 Classical Clustering Algorithm

In Cluster Analysis there is no prior information regarding the underlying group structure. Still the intention is to partition the data into relatively homogeneous groups called clusters. So the ability to determine characteristic prototypes or group representatives in a given set of data plays a central role in clustering. Once the clusters are formed, classification of unlabelled samples is done on the basis of minimum distance (e.g. Euclidean distance, Mahalanobis' distance etc.) criteria.

2.1 Cluster Formation By K-Means Algorithm

This algorithm is based on the minimisation of a performance index which is defined as the sum of squared distances from all points in a cluster domain to the cluster center.

Step 1: Choose k (a constant fixed beforehand) initial cluster centers $C_1(1), C_2(1), \ldots, C_k(1)$. These are selected as the first k samples of the given sample set.

Step 2: At the l^{th} iterative step, distribute the samples $\{x\}$ among k clusters using the relation,

$$x \in S_j(l) \text{ if } \|x - C_j(l)\| < \|x - C_i(l)\|, \quad \forall i = 1(1)k, \;\; i \neq j,$$

where $S_j(l)$ denotes the set of samples whose cluster center is $C_j(l)$. Ties are resolved arbitrarily.

Step 3: From the results of **Step 2**, calculate the new cluster centers $C_j(l+1), \quad \forall j = 1(1)k$ as follows,

$$C_j(l+1) = \frac{1}{N_j} \sum_{x \in S_j(l)} x, \quad \forall j = 1(1)k,$$

where N_j is the number of samples in $S_j(l)$.

Step 4: If $C_j(l+1) = C_j(l), \forall j = 1(1)k$ then stop else go to **Step 2**.

2.2 Algorithm for classification of a data set x when the cluster centers are known.

Step 1: Initialise N_j to 0, $\forall j = 1(1)k$, where N_j is the number of samples in the j^{th} cluster S_j.

Step 2: For $i = 1$ to n do

if $\|x(i) - C_j\| < \|x(i) - C_l\|, \forall l = 1(1)k, l \neq j$ then

begin

$x(i) \in S_j$;

$N_j = N_j + 1$;

end.

where C_j = center of S_j, $\forall j = 1(1)k$

$x(i) = i^{th}$ sample, $\forall i = 1(1)n$

n = number of samples.

Step 3: Compute mean vectors and dispersion matrices for each cluster, if needed.

2.3 Application

The above algorithms were implemented on the 512×512 gray level image of Calcutta which has got 4 spectral bands - blue, green, red and infrared. Clearly the size of the sample set $= 512 \times 512 = 262144$.

Number of clusters, k, is taken to be 7 corresponding to concrete, asphalt, river, semi-concrete, soil, water, vegetation parts of Calcutta.

The following statistics are obtained when the algorithms 2.1 and 2.2 are run.

```
Statistics for the class :- concrete

The frequency is 11587

Mean vector :-

        47.048157 32.069906 35.579011 40.562441


Dispersion matrix :-

        6.075009 4.200914  2.881756  1.796423

        4.200914 4.610458  4.710451  3.026359

        2.881756 4.710451 13.333082  3.359332

        1.796423 3.026359  3.359332 15.386862


Statistics for the class :- asphalt

The frequency is 74065

Mean vector :-

        41.163397 25.668035 27.231796 30.275191

Dispersion matrix :-

        1.323467 0.589151  0.436722 -0.160702

        0.589151 0.836360  0.598598  0.456789

        0.436722 0.598598  2.832924 -0.002788

       -0.160702 0.456789 -0.002788  6.976967
```

Statistics for the class :- river_water

The frequency is 12288

Mean vector :-

44.106283 29.757650 31.278727 23.208089

Dispersion matrix :-

0.960221	0.408570	0.163247	0.657571
0.408570	1.021344	0.436365	-1.709091
0.163247	0.436365	1.269235	-0.074927
0.657571	-1.709091	-0.074927	12.781487

Statistics for the class :- semi_concrete

The frequency is 42227

Mean vector :-

42.910460 28.052952 29.953111 37.666019

Dispersion matrix :-

1.831730	0.748626	0.134978	-0.772107
0.748626	1.064428	0.632601	0.634802
0.134978	0.632601	4.153531	-0.559578
-0.772107	0.634802	-0.559578	14.463563

Statistics for the class :- soil

The frequency is 44780

Mean vector :-

39.751943 25.357034 25.450893 39.804779

Dispersion matrix :-

1.463479	0.746738	0.671293	-1.586166
0.746738	0.992937	0.846519	-0.590482
0.671293	0.846519	4.235485	-3.570574
-1.586166	-0.590482	-3.570574	20.737815

Statistics for the class :- clean_water

The frequency is 13578

Mean vector :-

36.719694 21.829283 19.883441 23.093534

Dispersion matrix :-

2.015993	1.365242	1.087568	0.197156
1.365242	1.550764	1.133568	0.494757
1.087568	1.133568	3.228252	2.493678
0.197156	0.494757	2.493678	18.053853

Statistics for the class :- vegetation

The frequency is 63619

Mean vector :-

38.150034 23.447272 23.150160 33.350241

Dispersion matrix :-

1.368505	0.670300	0.673537	-0.998508
0.670300	0.863074	0.684909	-0.276979
0.673537	0.684909	3.259255	-1.375452
-0.998508	-0.276979	-1.375452	10.091387

2.4 Disadvantages

1. K-means clustering algorithm is very much influenced by the number of cluster centers specified, the choice of cluster centers, the order in which samples were taken.

2. Cluster centers or variances are not true means or variances respectively i.e. if cluster centers corresponding to Calcutta image are used to classify an image of Sunderban forest (where vegetation cluster is predominant), a huge number of misclassifications yield i.e. estimation of cluster center is not robust.

3. Outliers are not detected i.e. outliers are forcibly classified by the minimum distance classifier even if it belongs to a completely new cluster.

3 Model based clustering

In this section we first formulate likelihood estimation for finite mixture models and then define the mixture likelihood approach to clustering and discuss the estimation procedure. Then algorithms for estimating the cluster parameters and for classifying non-outliers follows. Ultimately we critically judge its performance and present a number of applications.

3.1 Mixture model approach to clustering

Let $x_1, x_2, \ldots, x_n$ be a set of observations of dimension p, based on which a number of groups have to be formed. Under the finite mixture model, each x_j can be viewed as arising from a superpopulation G which is a mixture of finite number, say g, of populations (corresponding to each group) $G_1, G_2, \ldots, G_g$ in some proportions $\pi_1, \pi_2, \ldots, \pi_g$ respectively where $\sum_{i=1}^{g} \pi_i = 1$ and $\pi_i \geq 0 \quad \forall i = 1(1)c$.

The probability density function (p.d.f.) of an observation x in G can therefore be represented in the finite mixture form,

$$f(x;\phi) = \sum_{i=1}^{g} \pi_i g_i(x;\theta) \tag{1}$$

where $g_i(x;\theta)$ is the p.d.f. corresponding to G_i and θ denotes the vector of all unknown parameters associated with the parametric forms adopted for these g component densities. The vector $\phi = (\pi', \theta')'$ of all unknown parameters belongs to some parameter space Ω. Throughout this study we shall refer to G_i as either population, group, cluster or class regardless of the underlying situation.

Once $\hat{\phi}$ has been obtained, estimates of the posterior probabilities of population membership can be formed for each x_j to give a probabilistic clustering. The posterior probability that x_j belongs to G_i is given by

$$\tau_{ij} = \frac{\pi_i f_i(x_j;\theta)}{\sum_{l=1}^{g} \pi_l f_l(x_j;\theta)} \quad i = 1, 2, \ldots, g. \tag{2}$$

3.2 Likelihood Estimation For Mixture Models Via *EM Algorithm*

Corresponding to the formulation of finite mixture models in Subsection 3.1, we let the vector of indicator variables $z_j = (z_{1j}, \ldots, z_{gj})'$ be defined by

$$\begin{aligned} z_{ij} &= 1, \ \ x_j \in G_i, \\ &= 0, \ \ x_j \notin G_i, \end{aligned}$$

where $z_1, \ldots, z_n$ are identically and independently distributed (iid) according to a multinomial distribution consisting of one draw on g categories with probabilities $\pi_1, \ldots, \pi_g$ respectively. We write $z_1, \ldots, z_n \overset{iid}{\sim} \text{Mult}_g(1, \pi)$.

In accordance with the mixture model (1), it is further assumed that $x_1, \ldots, x_n$ given $z_1, \ldots, z_n$ respectively are conditionally independent, and x_j given z_j has log density

$$\sum_{i=1}^{g} z_{ij} \log f_i(x_j;\theta), \ \ \forall j = 1(1)n$$

Hence the log likelihood for the complete data,

$$X = (x_1', \ldots, x_n')' \ \text{ and } \ Z = (z_1', \ldots, z_n')',$$

is given by

$$L_c(\phi) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \{\log \pi_i + \log f_i(x_j;\theta)\}$$

The *EM algorithm* is applied to the mixture model (1) by treating Z as missing data. It proceeds in two steps, $\mathcal{E}$ (for expectation) and $\mathcal{M}$ (for maximization). Using some initial value for ϕ, say $\phi^{(0)}$, the $\mathcal{E}$ step requires the calculation of

$$Q(\phi, \phi^{(0)}) = \mathcal{E}\{L_c(\phi) \mid X; \phi^{(0)}\},$$

the expectation of the complete data log likelihood, $L_c(\phi)$, conditional on the observed data and the initial fit $\phi^{(0)}$ for ϕ.

On the $\mathcal{M}$ step first time through, the intent is to choose the value of ϕ, say $\phi^{(1)}$, that maximizes $Q(\phi, \phi^{(0)})$. One nice feature of the *EM algorithm* is that the solution to the $\mathcal{M}$ step often exists in closed form, as is to be demonstrated for mixture of normals in the following subsection.

3.3 Closed Form Estimates Under Normality

Under the assumption of nomality for each of the populations, x_j follows $N(\mu_i, \Sigma_i)$ in G_i with probability π_i $(i = 1, \ldots, g)$. The likelihood estimates of π_i, μ_i, and Σ_i satisfy

$$\hat{\pi}_i = \frac{\sum_{j=1}^{n} \hat{\tau}_{ij}}{n}, \tag{3}$$

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} \hat{\tau}_{ij} x_j}{n\hat{\pi}_i}, \tag{4}$$

and

$$\hat{\Sigma}_i = \frac{\sum_{j=1}^{n} \hat{\tau}_{ij}(x_j - \hat{\mu}_i)(x_j - \hat{\mu}_i)'}{n\hat{\pi}_i}, \tag{5}$$

for $i = 1, \ldots, g$.

In these equations, the posterior probability that x_j belongs to G_i, is given by appropriate version of 2, namely

$$\hat{\tau}_{ij} = \frac{\pi_i \mid \Sigma_i \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x_j - \mu_j)'\Sigma_i^{-1}(x_j - \mu_j)\}}{\sum_{t=1}^{g} \pi_t \mid \Sigma_t \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x_t - \mu_t)'\Sigma_t^{-1}(x_t - \mu_t)\}} \tag{6}$$

Initial estimates of the π_i, μ_i, Σ_i are obtained by one of a variety of methods (e.g. seeds returned by the k-means algorithm can be taken as a estimates of μ_i's) and these are then used to obtain first estimates of τ_{ij}; these are then inserted into Equations (3) - (5) to give revised parameter estimates and the process is continued until convergence.

3.4 Algorithm For Estimating Mixture Model Approach

Algorithm: It takes as input

1. number of groups (g).
2. data dimension (p).
3. number of observations (n).
4. entire data set ($X(i,j)$, $i = 1(1)n$, $j = 1(1)p$).

5. initial estimates for

a) mixing proportion ($\pi_k^{(0)}$, $k = 1(1)q$ such that $\sum_{k=1}^{q} \pi_k^{(0)} = 1$)

b) mean vectors for each group ($\mu_k^{(0)}$, $k = 1(1)q$)

c) dispersion matrices for each group ($\Sigma_k^{(0)}$, $k = 1(1)q$) or common dispersion matrix ($\Sigma^{(0)}$) (here, $\Sigma_k^{(0)} = \Sigma^{(0)}$ $\forall k = 1(1)q$).

If an initial partition is given for all n observations then **5a** to **5c** are obtained as follows:

$$\pi_k^{(0)} = \frac{\text{number of observations falling in the } k^{th} \text{ group}}{n} \tag{7}$$

$$= f_k/n, \quad \forall k = 1(1)q, \tag{8}$$

$$\mu_k^{(0)} = \sum_{r \in k^{th} \text{group}} \underline{X}_r / f_k, \quad \forall k = 1(1)q, \tag{9}$$

$$\Sigma_k^{(0)} = \sum_{r \in k \text{th group}} (\underline{X}_r - \mu_k^{(0)})'(\underline{X}_r - \mu_k^{(0)})/f_k \tag{10}$$

$$= T_k/f_k, \quad \forall k = 1(1)q. \tag{11}$$

$$\text{or} \quad \Sigma^{(0)} = \sum_{k=1}^{q} T_k f_k / n \tag{12}$$

This algorithm proceeds iteratively. At the t^{th} step of iteration ($t = 1,2,3,\dots$) it does the following things:-

Step 1 Likelihood corresponding to the i^{th} observation $= GUM^{(t)}(i)$

$$= \sum_{k=1}^{q} \frac{\exp[-1/2(X_i - \mu_k^{(t-1)})'\Sigma_k^{(t-1)-1}(X_i - \mu_k^{(t-1)})]}{(2\pi)^{p/2}\sqrt{|\Sigma_k^{(t-1)}|}} \,.\, \pi^{(t-1)}(k)$$

$$= \sum_{k=1}^{q} AL^{(t)}(k) \,.\, \pi^{(t-1)}(k)$$

where X_i = ith row vector of $X(i,j)$.

Step 2 Posterior probability for the i^{th} obsevation to fall in the kth group

$$= W^{(t)}(i,k) = \frac{\pi^{(t-1)}(k) \,.\, AL^{(t)}(k)}{GUM^{(t)}(i)}$$

Step 3 Log-likelihood of $X_1, \ldots X_n = XLOGL^{(t)} = \sum_{i=1}^{n} \log(GUM^{(t)}(i))$

Step 4 If $|\, XLOGL^{(t)} - XLOGL^{(t-1)} \,| < \varepsilon$ (ε is any +ve small quantity), output $\pi^{(t-1)}$, $\mu_k^{(t-1)}$, $\Sigma_k^{(t-1)} \quad \forall k = 1(1)q$ and go to **Step 9.**

(Assume $XLOGL^{(0)} = 0.0$)

Step 5 New estimate of mixing proportion $= \pi^{(t)}(k) = WSUM^{(t)}(k)/n \;\; \forall k = 1(1)q$, where $WSUM^{(t)}(k) = \sum_{i=1}^{n} W^{(t)}(i,k)$

Step 6 New estimate of mean for k^{th} group and j^{th} attribute

$$= f^{(t)}(k,j) = \frac{\sum_{i=1}^{n} X(i,j) \,.\, W(t)(i,k)}{WSUM(t)(k)}, \quad \forall k = 1(1)q, \quad \forall j = 1(1)p$$

Step 7 New estimate of dispersion matrix for k^{th} group $= \Sigma^{(t)}(k)$ where $(r,s)^{th}$ element of $\Sigma^{(t)}(k) = \Sigma^{(t)}(k,r,s)$

$$= \frac{\sum_{i=1}^{n} (X(i,r) - \mu^{(t)}(k,r)) \,.\, (X(i,s) - \mu^{(t)}(k,s))}{WSUM^{(t)}(k)}$$

$$= \frac{V^{(t)}(k,r,s)}{WSUM^{(t)}(k)} \quad \forall r = s = 1(1)p, \;\; \forall k = 1(1)q$$

Step 8 If it is known that all the groups have the same dispersion matrix then common covariance matrix $= \Sigma^{(t)}$, where $(r,s)^{th}$ element of $\Sigma^{(t)} = \Sigma^{(t)}(r,s)$

$$= \sum_{k=1}^{q} (WSUM^{(t)}(k) \,.\, V^{(t)}(k,r,s))/n \quad \forall r = s = 1(1)p$$

Step 9 Classify i^{th} observation into the k^{th} group if

$$W^{(t)}(i,k) = \max_{1 \le r \le q} W^{(t)}(i,r) \quad \forall i = 1(1)n$$

Step 10 Correct allocation rate corresponding to the k^{th} group $= XCC(k)$

$$= \frac{\sum_{i \in [1,n] \ \& \ i^{th} \text{ obsn. is classified in the } k^{th} \text{ group}} W^{(t)}(i,k)}{(n \, . \, \pi^{(t-1)}(k))}, \quad \forall k = 1(1)q$$

Step 11 Overall correct allocation rate $= CC = \sum_{k=1}^{q} XCC(k) \, . \, \pi^{(t-1)}(k)$

Step 12 End.

3.5 Algorithm for classification of a data set after learning mixture model parameters

For a sample pattern x the following steps are done :

Step 1 For $i = 1$ to g (i.e. for all groups) calculate the discriminant function,

$$g_i(x) = \ln \pi_i - \tfrac{1}{2} \ln |\, \Sigma_i \,| - \tfrac{1}{2}(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)$$

and the threshold T_i at some level of significance, say 95% as follows,

$$T_i = 4.744 - \tfrac{1}{2} \ln |\, \Sigma_i \,| + \ln \pi_i$$

Step 2 If $g_i(x) \le T_i \quad \forall i = 1(1)g$ then

infer that x is an outlier and hence is not classified

else

get hold of all those groups for which $g_i(x) > T_i$.

Let them be $i_1, \ldots, i_l \quad (l < g)$.

step 3 Classify x in i_k th group $(k \le g)$ if $g_{i_k}(x) > g_{i_j}(x) \quad \forall j \ne k$.

3.6 Critical remarks

The algorithm 3.4 is very simple to implement on a computer and in the case of well seperated classes and good initial estimates (in our implementation they are either supplied by the user or randomly chosen or the output of k-means algoritm) appears to convarge quite rapidly in both univariate and multivariate situations. However, a poor choice of initial estimates can exacerbate the situation making the rate of convergence very slow. Indeed, this algorithm converges provided the likelihood is continuous having a unique maxima and is bounded on the edge of the parameter space. The possibility of multiple solutions is present here, as will be shown by one of our applications.

Since we explicitly assume normality of each class, all the class parameters, thus estimated, represent true statistics. Also before classifying a sample pattern, we can test whether it's an outlier or not since $\pm 1.96\sigma$ limits contain probabilistically 95% of a population.

3.7 Applications

1. The first example involves Fisher's IRIS data. These data consist of 150 four-dimensional

labelled samples made up of 50 obsevations from each of three species of iris. The four measurements taken on each plant are sepal length, sepal width, petal length, petal length.

True Statistics:-

Mean vector for each group

65.8800	29.7400	55.5200	20.2600
59.3600	27.7000	42.6000	13.2600
50.0600	34.2800	14.6200	2.4600

Covariance matrix for group 1

40.4343			
9.3763	10.4004		
30.3290	7.1380	30.4588	
4.9094	4.7629	4.8824	7.5433

Covariance matrix for group 2

26.6433			
8.5184	9.8469		
18.2898	8.2653	22.0816	
5.5780	4.1204	7.3102	3.9106

Covariance matrix for group 3

12.4249			
9.9216	14.3690		
1.6355	1.1698	3.0159	
1.0331	0.9298	0.6069	1.1106

Results obtained from algorithm 3.4:-

Number of iterations : 22

Estimate of mixing proportion for each group

0.333 0.299 0.367

Number assigned to each group

50 45 55

Estimates of correct allocation rates for each group

1.000 0.985 0.986

Estimate of overall correct allocation rate 0.990

Estimates of Mean vector for each group

50.060001	34.279999	14.620000	2.460000
59.145519	27.776814	42.011272	12.967881
65.449150	29.488003	54.799564	19.847801

Estimate of Covariance matrix for group 1

12.176401			
9.723202	14.081599		
1.602800	1.146400	2.955600	
1.012400	0.911200	0.594800	1.088400

Estimates of Covariance matrix for group 2

27.480822			
9.670334	9.252395		
18.403929	9.087162	19.991385	
5.418780	4.292072	6.074602	3.192841

```
Estimates of Covariance matrix for group  3
        38.702080
         9.225613    11.039165
        30.263767     8.427150    32.742870
         6.147326     5.595409     7.425745     8.563549
```

Performance of algorithm 3.4:-

Original group	# observations classified into group		
	1	2	3
1	50	0	0
2	0	45	5
3	0	0	50

2. The second example is based on two sets of generated data.First we describe algorithms for drawing such data.

   (a) Algorithm for drawing a random sample from a mixture of normal populations:

      i. Read the dimension, number of populations, mean vectors and dispersion matrices for each of them. Read the mixing proportion and the size of the set of samples.

      ii. Choose a random number between 0 and 1. Select a population i if the random number is greater than the cumulated mixing proportion of population $i - 1$, but less than that of population i.

      iii. Once the class is fixed, select a random sample from it.

      iv. Repeat **Steps 2-3** as many times as the size of the data set.

(b) Algorithm for drawing a random sample in two dimension from a mixture of a number of normal populations with identical dispersion matrices and means lieing on the circumference of a circle:

i. Read the center(c_1, c_2) and radius(r) of the circle. Also read the size of the data set and the common dispersion matrix(σ).

ii. Select a a random number(i) between 0 and 1. Get $\theta = 2 \times \pi \times i$

iii. Now select a random number from bivariate normal population with mean vector $= (c_1 + r\cos(\theta),\ c_2 + r\sin(\theta))$ and dispersion matrix $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$.

iv. Repeat **Steps 2-3** as many times as the size of the data set.

Results of data generated from algorithm (2a):-

Here, a number of sets of observations are selected from the mixture of three normal populations with mean vectors:- (0, 100), (0, 0), (100, 0).

The second population has dispersion matrix:- $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$

while the other two have the identical dispersion matrix:- $\begin{pmatrix} \sigma^2 & -\sigma^2/2 \\ -\sigma^2/2 & \sigma^2 \end{pmatrix}$

Cases are considered for different mixing proportions and the value of σ

Each set of observations contains 2000 data.

Case 1. $\sigma = 20$

The mixing proportions = 0.1, 0.3, 0.6

```
Number of iterations :  15
```

```
Estimate of mixing proportion for each group :

    0.101    0.313    0.586

Number assigned to each group :

    202      625      1173

Estimates of correct allocation rates for each group :

    0.983    0.997    0.990

Estimate of overall correct allocation rate :   0.993

Estimates of Mean vector of each group :

    -0.392211     100.123306
     0.228428       0.764145
   100.884407      -0.492164


Estimate of Covariance Matrix for group :   1
   404.424072
  -193.554001    435.488525

Estimate of Covariance Matrix for group :   2
   375.200043
    17.729874    408.668701

Estimate of Covariance Matrix for group :   3
   389.550110
  -201.019379    436.056671
```

Case 2. $\sigma = 20$

The mixing proportions = 0.3, 0.1, 0.6

```
Number of iterations :  16
```

Estimate of mixing proportion for each group :

0.312 0.092 0.596

Number assigned to each group :

625 182 1193

Estimates of correct allocation rates for each group :

0.997 0.996 0.983

Estimate of overall correct allocation rate : 0.996

Estimates of Mean vector of each group :

-0.435669	99.979187
0.056349	-0.723473
98.659424	0.497254

Estimate of Covariance Matrix for group : 1

372.037170	
-177.024796	370.494812

Estimate of Covariance Matrix for group : 2

335.930908	
-14.639151	391.148956

Estimate of Covariance Matrix for group : 3

417.927490	
-213.981903	421.690735

Case 3. $\sigma = 20$

The mixing proportions = 0.3, 0.6, 0.1

Number of iterations : 16

Estimate of mixing proportion for each group :

0.295 0.613 0.091

Number assigned to each group :

591 1226 183

Estimates of correct allocation rates for each group :

0.989 0.993 0.995

Estimate of overall correct allocation rate : 0.994

Estimates of Mean vector of each group :

-0.917666	100.580521
0.091537	0.345071
99.372871	0.320097

Estimate of Covariance Matrix for group : 1

404.284637	
-215.043091	390.589569

Estimate of Covariance Matrix for group : 2

421.422119	
10.557711	455.181091

Estimate of Covariance Matrix for group : 3

310.371399	
-174.862152	411.657990

Case 4. $\sigma = 25$

The mixing proportions = 0.1, 0.3, 0.6

Number of iterations : 16

Estimate of mixing proportion for each group :

0.101 0.316 0.583

Number assigned to each group :

198 630 1172

Estimates of correct allocation rates for each group :

0.987 0.943 0.964

Estimate of overall correct allocation rate : 0.975

Estimate of Mean vector of each group :

-0.234051	100.846672
0.724739	0.793344
101.291077	-0.615607

Estimate of Covariance Matrix for group : 1

640.773315	
-290.710632	627.766113

Estimate of Covariance Matrix for group : 2

608.793396	
16.674536	630.282654

Estimate of Covariance Matrix for group : 3

605.389343	
-315.970703	675.939148

Case 5. $\sigma = 25$

The mixing proportions = 0.3, 0.1, 0.6

Number of iterations : 16

Estimate of mixing proportion for each group :

0.286 0.105 0.609

Number assigned to each group :

574 207 1219

Estimates of correct allocation rates for each group :

0.991 0.948 0.983

Estimate of overall correct allocation rate : 0.984

Estimates of Mean vector of each group :

-0.458357	99.735809
-5.299500	1.558519
99.715591	0.471394

Estimates of Covariance matrix for group : 1

583.770813	
-267.066162	593.437378

Estimates of Covariance matrix for group : 2

555.652954	
-0.124598	522.771118

Estimates of Covariance matrix for group : 3

622.366333	
-304.373169	617.530823

Case 6. $\sigma = 25$

The mixing proportions = 0.3, 0.6, 0.1

```
Number of iterations :  15

Estimate of mixing proportion for each group :

    0.290    0.608    0.101

Number assigned to each group :

    582 1216  202

Estimates of correct allocation rates for each group :

    0.974     0.983     0.955

Estimate of overall correct allocation rate :   0.977

Estimate of Mean vector of each group :

    1.148690       98.737877

    0.766814       -1.403253

  104.056717        0.647550


Estimate of Covariance matrix for group :   1

   657.415039

  -326.124573     621.708557

Estimate of Covariance matrix for group :   2

   663.444946

    10.881446     583.504089

Estimate of Covariance matrix for group :   3

   530.152588

  -263.069946     463.371246
```

Case 7. $\sigma = 30$

The mixing proportions = 0.1, 0.3, 0.6

Number of iterations : 18

Estimate of mixing proportion for each group :

0.098 0.302 0.599

Number assigned to each group :

200 584 1216

Estimates of correct allocation rates for each group :

0.974 0.912 0.913

Estimate of overall correct allocation rate : 0.950

Estimate of Mean vectors of each group :

2.667619	96.906075
0.008582	-2.463465
100.948265	-0.836262

Estimate of Covariance matrix for group : 1

1048.167603	
-428.572113	754.855164

Estimate of Covariance matrix for group : 2

887.045898	
37.990944	942.120605

Estimate of Covariance matrix for group : 3

885.614197	
-380.053345	804.028259

Case 8. $\sigma = 30$

The mixing proportions = 0.3, 0.1, 0.6

Number of iterations : 28

Estimate of mixing proportion for each group :

0.288 0.107 0.605

Number assigned to each group :

577 186 1237

Estimates of correct allocation rates for each group :

0.943 0.798 0.979

Estimate of overall correct allocation rate : 0.949

Estimate of Mean vectors of each group :

-2.579146	101.691238
2.902576	5.266840
99.752296	0.225007

Estimate of Covariance matrix for group : 1

858.447510	
-450.879486	866.791138

Estimate of Covariance matrix for group : 2

986.552734	
6.376415	909.992615

Estimate of Covariance matrix for group : 3

942.493774	
-542.169312	1053.305054

Case 9. $\sigma = 30$

The mixing proportions = 0.3, 0.6, 0.1

Number of iterations : 21

Estimate of mixing proportion for each group :

0.288 0.612 0.100

Number assigned to each group :

575 1227 198

Estimates of correct allocation rates for each group :

0.936 0.963 0.896

Estimate of overall correct allocation rate : 0.949

Estimate of Mean vectors of each group :

1.647508	98.720840
0.690766	-1.215889
104.921638	-0.240215

Estimate of Covariance matrix for group : 1

957.963806	
-481.344116	893.229553

Estimate of Covariance matrix for group : 2

933.206726	
15.572630	862.264465

Estimate of Covariance matrix for group : 3

781.993774	
-344.267365	633.120605

Case 10. $\sigma = 40$

The mixing proportions = 0.1, 0.3, 0.6

Number of iterations : 75

Estimate of mixing proportion for each group :

0.150 0.259 0.591

Number assigned to each group :

298 470 1232

Estimates of correct allocation rates for each group :

0.746 0.642 0.931

Estimate of overall correct allocation rate : 0.829

Estimates of Mean vector of each group As Row Vector :

11.117209	47.730080
-12.478016	-12.280910
100.800171	0.558255

Estimate of Covariance matrix for group : 1

1543.272217	
-960.350403	2856.628906

Estimates of Covariance matrix for group : 2

1495.782227	
-107.998611	1126.963135

Estimated Covariance Matrix for Group : 3

1642.506714	
-952.988159	1834.486328

Case 11. $\sigma = 40$

The mixing proportions = 0.3, 0.1, 0.6

Number of iterations : 32

Estimate of mixing proportion for each group :

0.277 0.172 0.551

Number assigned to each group :

580 239 1181

Estimates of correct allocation rates for each group :

0.880 0.553 0.944

Estimate of overall correct allocation rate : 0.859

Estimates of Mean vector of each group :

0.916070	99.372116
19.551199	10.426476
104.839165	-3.601208

Estimate of Covariance matrix for group : 1

1730.173828	
-852.904053	1557.038574

Estimate of Covariance matrix for group : 2

2024.004150	
543.490601	2037.905151

Estimate of Covariance matrix for group : 3

1453.925781	
-625.075439	1369.367554

Case 12. $\sigma = 40$

The mixing proportions = 0.3, 0.6, 0.1

```
Number of iterations :  19

Estimate of mixing proportion for each group :

    0.325    0.576   0.099

Number assigned to each group :

    635   1175   190

Estimates of correct allocation rates for each group :

    0.834     0.911     0.757

Estimate of overall correct allocation rate :    0.871

Estimates of Mean vector of Each Group :

     7.058172       91.178551

    -2.066078       -2.911061

   105.777245       -7.592432


Estimate of Covariance matrix for group :    1

  1860.651123

 -1039.864136    1880.863892

Estimate of Covariance matrix for group :    2

  1524.218140

    12.881596    1492.673584

Estimate of Covariance matrix for group :    3

  1459.214844

  -392.160400     914.777527
```

Case 13. $\sigma = 50$

The mixing proportions = 0.1, 0.3, 0.6

Number of iterations : 25

Estimate of mixing proportion for each group :

0.300 0.270 0.430

Number assigned to each group :

575 549 876

Estimates of correct allocation rates for each group :

0.794 0.593 0.849

Estimate of overall correct allocation rate : 0.763

Estimates of Mean vectors of each group :

-14.162048	17.562771
53.277687	26.593187
119.194534	-11.774804

Estimate of Covariance matrix for group : 1

2062.750244	
-282.452454	4724.073242

Estimate of Covariance matrix for group : 2

1198.721924	
-313.782104	2323.318848

Estimate of Covariance matrix for group : 3

1824.493164	
-668.068237	1980.807739

Case 14. $\sigma = 50$

The mixing proportions = 0.3, 0.1, 0.6

```
Number of iterations :  76

Estimate of mixing proportion for each group :

    0.202   0.207  0.591

Number assigned to each group :

    443  259   1298

Estimates of correct allocation rates for each group :

    0.808   0.451   0.925

Estimate of overall correct allocation rate :   0.803

Estimate of Mean vector of each group :

   -10.585078    111.386551

    17.625593     22.715710

   101.941689     -0.505170


Estimate of Covariance matrix for group :   1

  2366.805176

 -1084.525269   2050.948975

Estimate of Covariance matrix for group :   2

  2551.372314

   450.094482   3532.600098

Estimate of Covariance matrix for group :   3

  2440.798584

 -1145.285522   2293.508301
```

Case 15. $\sigma = 50$

The mixing proportions = 0.3, 0.6, 0.1

Number of iterations : 30

Estimate of mixing proportion for each group :

0.326 0.433 0.241

Number assigned to each group :

612 889 499

Estimates of correct allocation rates for each group :

0.780 0.822 0.618

Estimate of overall correct allocation rate : 0.746

Estimates of Mean vector of each group :

1.697739	105.763588
-12.214709	-11.593668
51.258606	19.037796

Estimate of Covariance matrix for group : 1

2621.964355	
-1232.882813	2200.416504

Estimate of Covariance matrix for group : 2

3859.423584	
-1501.965454	2175.956543

Estimate of Covariance matrix for group : 3

2033.690430	
-64.381004	2210.310547

Case 16. $\sigma = 50$

The mixing proportions = 0.33, 0.34, 0.33

Number of iterations : 24

Estimate of mixing proportion for each group :

0.261 0.321 0.419

Number assigned to each group :

536 613 851

Estimates of correct allocation rates for each group :

0.781 0.752 0.801

Estimate of overall correct allocation rate : 0.780

Estimates of Mean vector of each group :

-1.830849	106.771919
-4.462615	0.612914
88.615784	5.186072

Estimate of Covariance matrix for group : 1

2756.456299	
-1193.381226	2252.962891

Estimate of Covariance Matrix for group : 2

2144.407471	
137.807632	2309.318604

Estimate of Covariance matrix for group : 3

3087.415527	
-1477.998169	2506.156006

Results of data generated from algorithm (2b):-

Center of the Circle :- (0, 0)

Radius of the circle :- 10.0

Number of samples chosen :- 2000

Common Dispersion matrix :- $\begin{pmatrix} 10.0 & 0.0 \\ 0.0 & 10.0 \end{pmatrix}$

Case 1) Number of classes :- 2 and

Initial estimates of mean vectors :- (11.947064 0.031447)
(1.776658 11.548863)

```
Output:-

Number of iterations : 16

Estimate of mixing proportion for each group :

    0.484   0.516

Number assigned to each group :

    972 1028

Estimates of correct allocation rates for each group :

    0.964    0.963

Estimate of overall correct allocation rate :   0.963

Estimates of Mean vectors of each group  :

    4.359812      -4.727197

   -4.196358       4.698853
```

Estimate of the common Covariance matrix :

32.985062

20.305113 29.168730

Case 2) Number of classes :- 2 and Initial estimates of means are the seeds of k-means algorithm

Output:-

Number of iterations : 25

Estimate of mixing proportion for each group :

0.504 0.496

Number assigned to each group :

1012 988

Estimates of correct allocation rates for each group :

0.964 0.960

Estimate of overall correct allocation rate : 0.962

Estimates of Mean vector of each group :

-3.445137 -5.194888

3.390703 5.551369

Estimate of the common Covariance Matrix :

39.587063

-18.200884 22.490250

Case 3) Number of classes :- 4 and Initial estimates of means are the seeds of k-means algorithm

Output:-

```
Number of iterations :  39
Estimate of mixing proportion for each group :
    0.363   0.324   0.143   0.170
Number assigned to each group :
    735  641  287  337
Estimates of correct allocation rates for each group :
    0.973    0.953    0.913    0.917
Estimate of overall correct allocation rate :   0.948
Estimates of Mean vector of each group :
     1.157035      8.031337
    -0.713467     -8.296684
     9.604152     -1.496966
    -9.539909      0.751386
Estimate of the common Covariance matrix :
    21.935797
    -1.853388      5.258699
```

See that in all the above cases the class means are symmetrically placed with respect to the origin and the two axes.

3. The third appplication is done on the satellite image of Calcutta. Only the green and infrared components are used for learning parameters and classification.

i) The following statistics are obtained when the algorithms 3.4 is run. They are based on the data set containing every odd alternate pixel.

```
Number of iterations :  11
Estimate of mixing proportion for each group :
    0.068   0.301   0.034   0.136   0.165   0.045   0.251
Number assigned to each group :
    3326 19959 2257 7827 10801 2380 18986
Estimates of correct allocation rates for each group :
    0.585   0.860   0.994   0.531   0.674   0.749   0.808
Estimate of overall correct allocation rate :   0.752

Estimates of Mean vectors of each group :
Concrete :        28.984013     43.025143
Asphalt :         25.531445     29.564581
River-water :     30.024694     21.184803
Semi-concrete :   27.628201     37.334278
Soil :            25.534307     39.295780
Clean-water :     22.202835     22.218414
Vegetation :      24.241961     33.804958
```

Estimate of Covariance matrix for group : Concrete

14.717172	
-7.999398	42.856339

Estimate of Covariance matrix for group : Asphalt

2.561580	
1.098870	5.074855

Estimate of Covariance matrix for group : River-water

0.658370	
-0.232172	0.646024

Estimate of Covariance matrix for group : Semi-concrete

4.431402	
1.638282	7.782211

Estimate of Covariance matrix for group : Soil

3.418793	
1.521983	7.366956

Estimate of Covariance matrix for group : Clean-water

2.617187	
2.667316	15.507127

Estimate of Covariance matrix for group : Vegetation

2.107797	
0.748398	4.141095

ii) The following statistics are based on the data set containing every even alternate pixel.

Number of iterations : 16

Estimate of mixing proportion for each group :

0.266 0.044 0.203 0.246 0.034 0.167 0.039

Number assigned to each group :

17260 1661 14585 17913 2222 9534 2361

Estimates of correct allocation rates for each group :

0.830 0.470 0.801 0.784 0.996 0.454 0.857

Estimate of overall correct allocation rate : 0.741

Estimate of Mean vector of each group :

Asphalt :	25.394594	29.245977
Concrete :	28.662436	43.270638
Soil :	25.252880	39.932442
Vegetation :	24.005415	33.779476
River-water :	30.018211	21.121212
Semi-concrete :	26.911153	37.279152
Clean-water :	21.879190	21.443333

Estimate of Covariance matrix for group : Asphalt

2.090126

0.910376 4.695774

Estimate of Covariance matrix for group : Concrete

16.957106

```
-9.630012     66.457535

Estimate of Covariance matrix for group :   Soil

 6.495150

 0.313451      7.837987

Estimate of Covariance matrix for group :   Vegetation

 1.917037

 0.914022      5.678341

Estimate of Covariance matrix for group :   River-water

 0.605753

-0.213953      0.534497

Estimate of Covariance matrix for group :   Semi-concrete

 3.121879

-1.429364      8.103225

Estimate of Covariance matrix for group :   Clean-water

 2.051969

 1.125721     11.567801
```

Classification of the Calcutta image using algorithm 3.5 for both the cases i) and ii) are shown in figures 1-8.

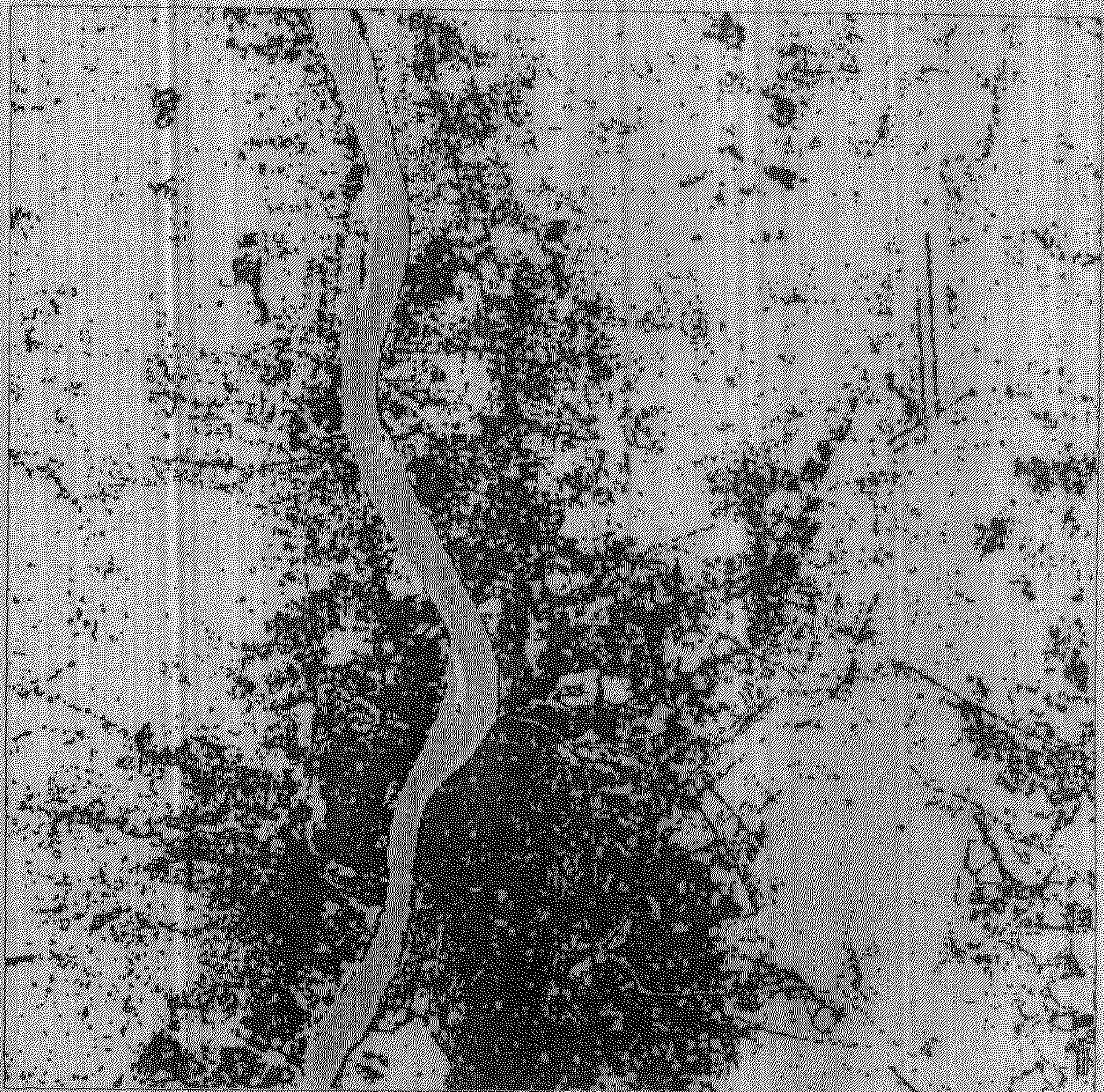

Fig.1 Asphalt(dark) and River_water(gray)

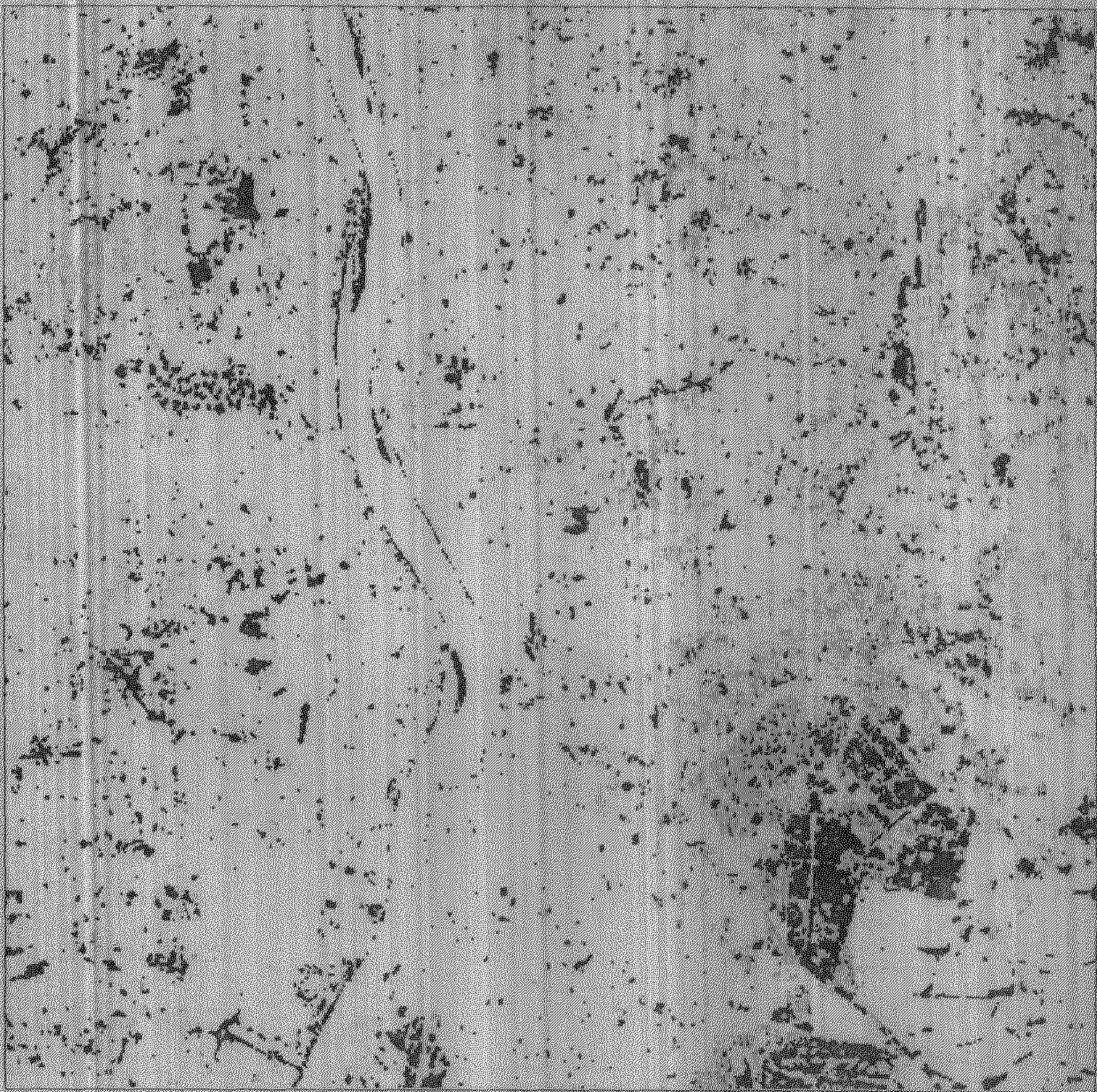

Fig.2 Concrete(dark) and Semiconcrete(gray)

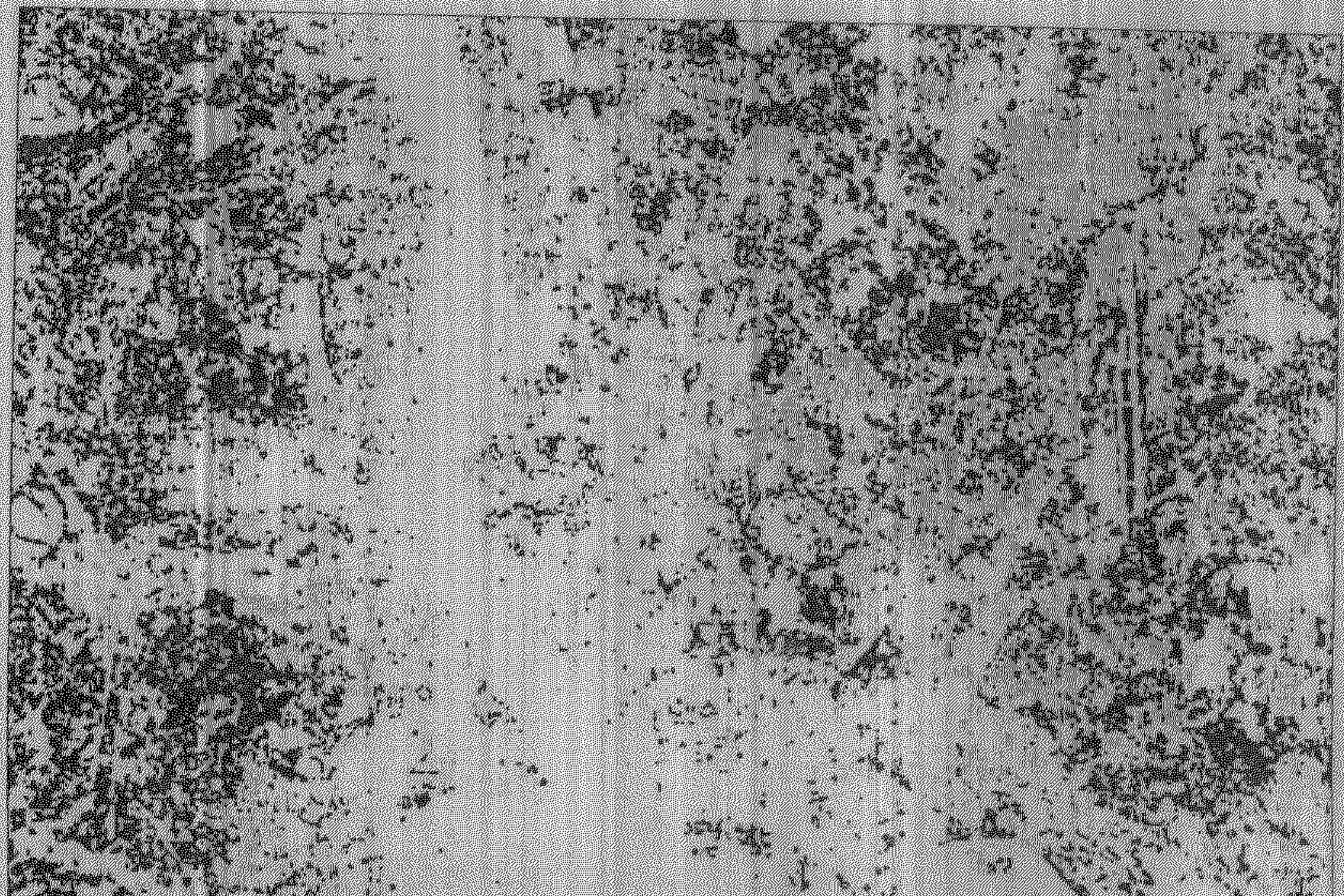

Fig.3 Soil(dark) and Vegetation(gray)

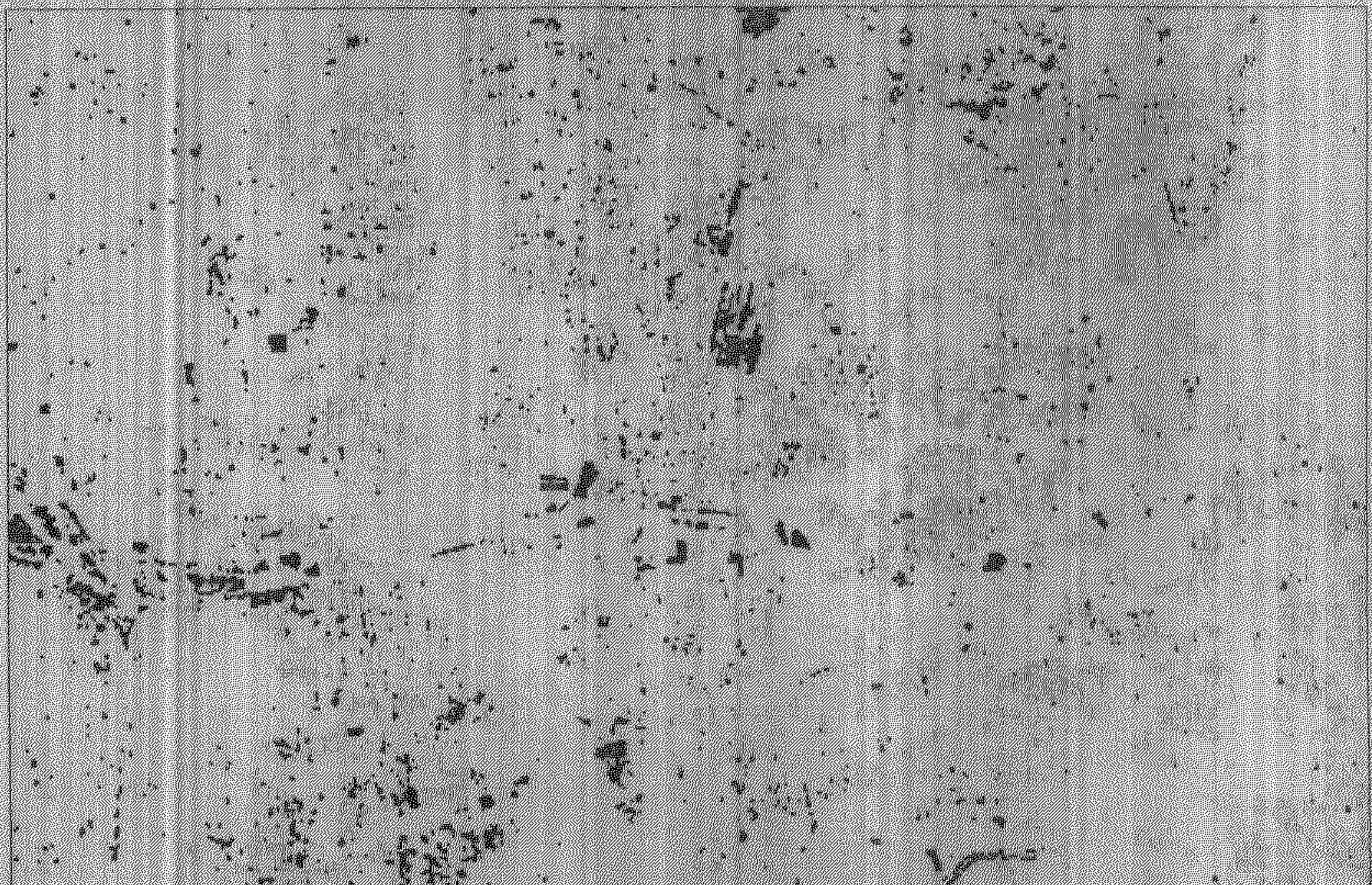

Fig.4 Water(dark) and Vegetaion(gray)

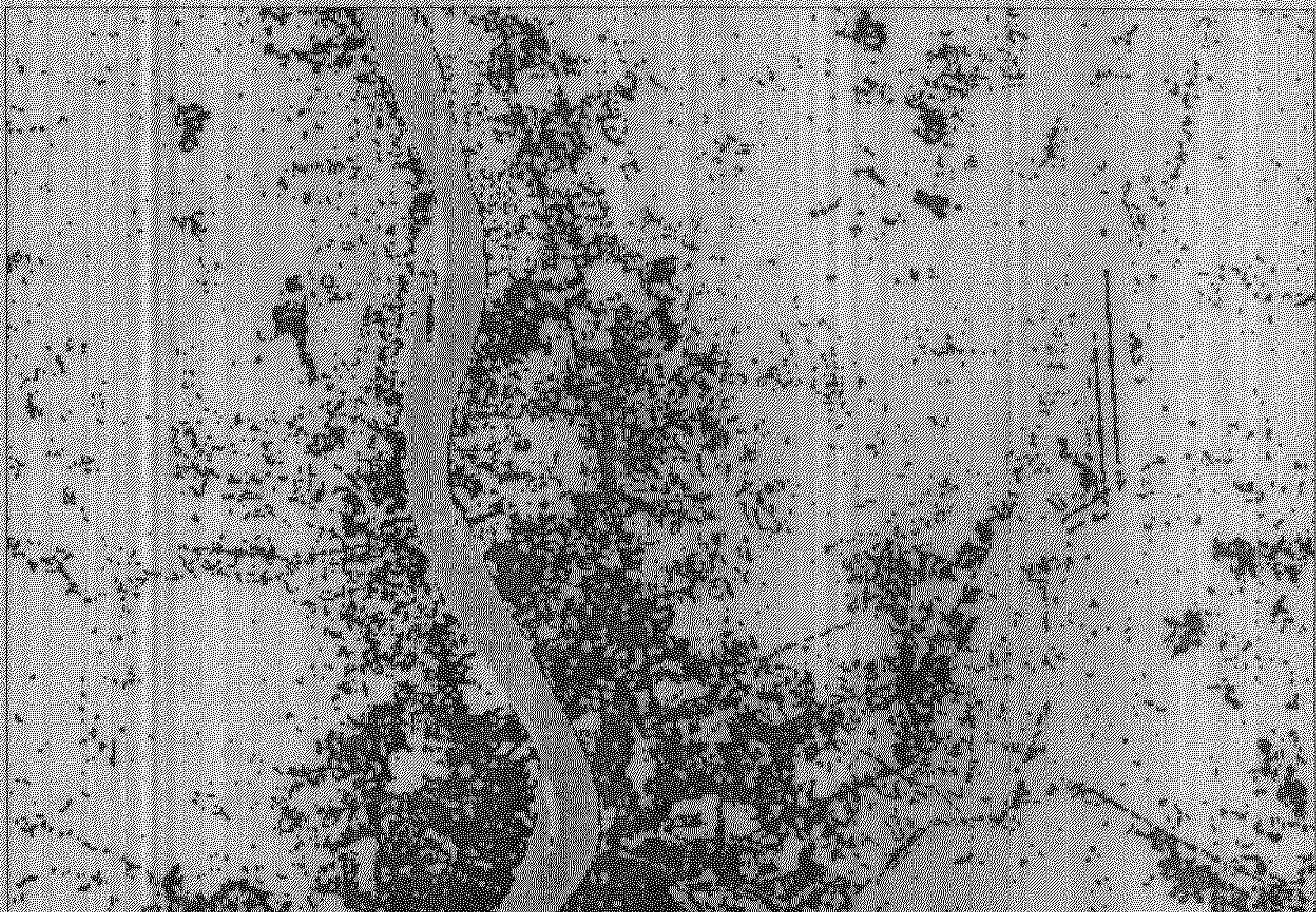

Fig.5 Asphalt(dark) and river_water(gray)

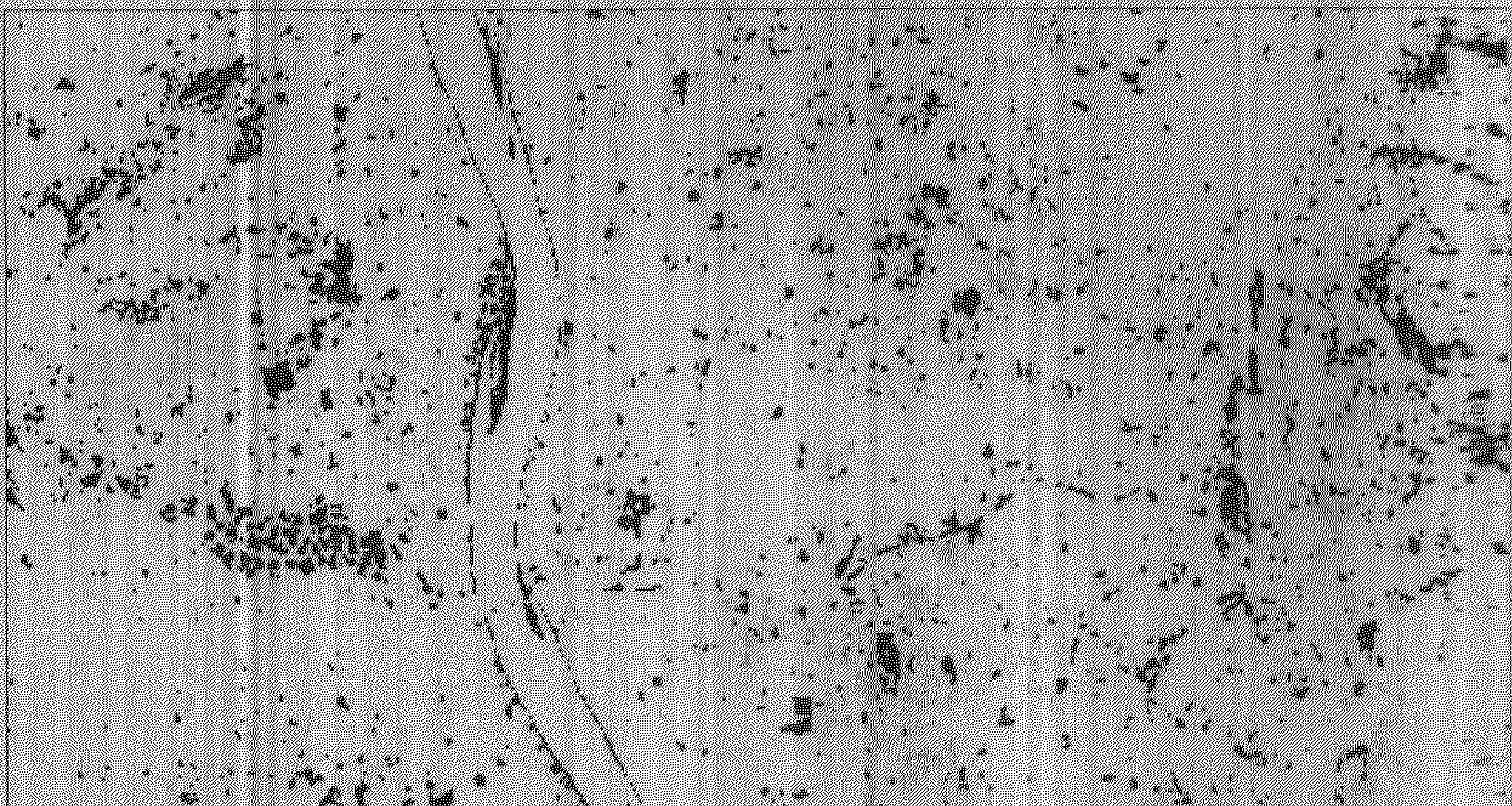

Fig. 6 concrete(dark) and semiconcrete(gray)

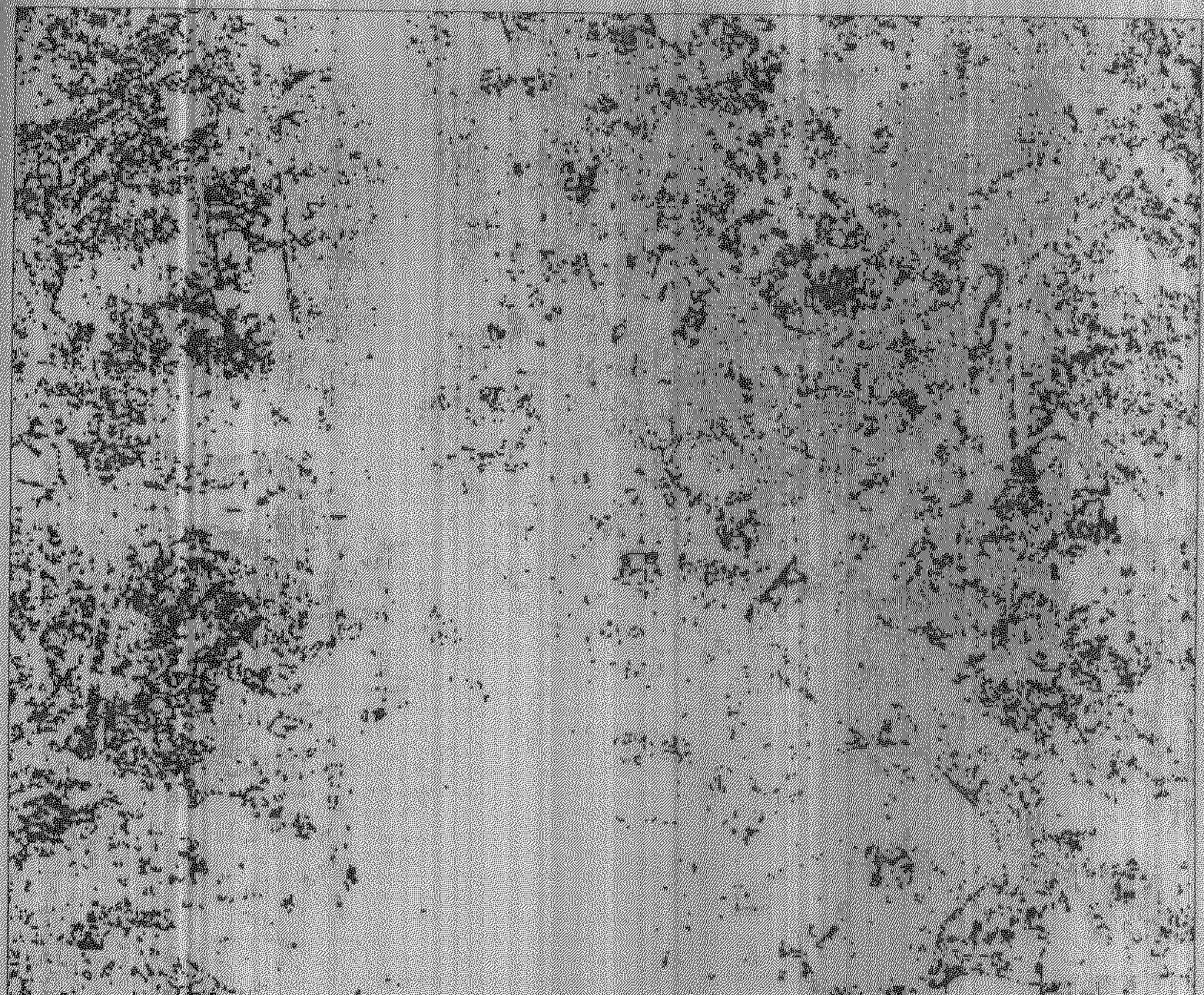

Fig.7 Soil(dark) and Vegetation(gray)

Fig.8 Water(dark) and Vegetation(gray)

4. The following is a comparison between the performances of k-means algorithm and the mixture model approach. The data set is a random sample from the three component mixture normal population with the following parameters :

Mean vectors : $\begin{array}{l} (\ 0 \quad 0\ ) \\ (\ 0 \quad 50\ ) \\ (\ 60 \quad 0\ ) \end{array}$

Dispersion matrices: 1) $\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$ 2) $\begin{pmatrix} 225 & 0 \\ 0 & 225 \end{pmatrix}$ 3) $\begin{pmatrix} 400 & 0 \\ 0 & 400 \end{pmatrix}$.

Mixing proportion : (0.34 0.33 0.33)

Results :

a) k-means algorithm

```
Seeds are as follows:
-2.693713    2.059846
 4.393713   48.604930
70.246826    2.413622
After classification the following statistics are obtained :
Number assigned to each group :
      60  39  51
Mean vectors for each group :
    -1.893713 1.859846
     3.323714 47.60347
    68.687468 1.943136
```

Dispersion matrices for each group :

1) 104.344966 -6.568898

-6.568898 105.774576

2) 255.285726 -22.754473

-22.754473 170.728603

3) 270.457421 19.288843

19.288843 422.055709

Performance:

Original group	# observations classified into group		
	1	2	3
1	45	2	6
2	10	30	5
3	5	7	40

b) Mixture model approach :

Estimate of mixing proportion for each group

0.399 0.257 0.344

Number assigned to each group

52 47 51

Estimates of correct allocation rates for each group

0.983 0.975 0.988

Estimate of overall correct allocation rate 0.983

Estimate of Mean vector for each group :

-0.184661	0.184974
0.224324	49.563152
61.804466	0.664675

Estimate of Covariance matrices for each group

1) 98.376648

7.882300 110.117767

2) 264.230896

1.127595 179.682510

3) 285.739532

11.067012 428.002472

Performance:

Original group	# observations classified into group		
	1	2	3
1	51	2	0
2	0	43	0
3	1	2	51

References

[1] Everitt, B.S. and Hand, D.J. (1981), *Finite Mixture Distributions*, Chapman and Hall.

[2] Hand, D.J. (1981), *Discrimination and Classification*, Wiley.

[3] Tou, J.T. and Gonzalez, R.C. (1974), *Pattern Recognition principles*, Addison-Wesley.

[4] Duda, R.O. and Hart, P.E. (1973), *Pattern Classification and Scene Analysis*, Wiley.

[5] Basford, G.J. and McLachlan, K.E. (1984), *Mixture Models*, Marcel Dekker.

[6] Richards, J.A. (1986), *Remote Sensing Digital Image Analysis*, Springer Verlag.

[7] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum likelihood from incomplete data via EM algorithm (with discussion), *J.R.Stat.Soc.* B 39, 1-38.