

Fig. 1

1. Symbols Related to Matrix

$\mathbf{A}$	Matrix
$\mathbf{a}$	Vector
a_{ij}	Element in i-th row and j-th column of matrix $\mathbf{A}$
$\hat{a}_t$	Element in t-th row of vector $\mathbf{a}$
$\text{tr}\mathbf{A}$	Trace of matrix $\mathbf{A}$
$\det\mathbf{A}$	Determinant of matrix $\mathbf{A}$
$\ln\mathbf{A}$	Logarithm of matrix $\mathbf{A}$
$\text{diag}\mathbf{A}$	Diagonal matrix of matrix $\mathbf{A}$
$\text{vec}\mathbf{A}$	Function to vectorize matrix $\mathbf{A}$
$\widehat{\text{vec}}\mathbf{a}$	Function to matrix vector $\mathbf{a}$

Fig. 2

2. Symbols Related to Probability Theory

Ω	Sample space
$\mathcal{F}$	Complete additive class
P	Probability measure in measurable space $(\Omega, \mathcal{F})$
$(\Omega, \mathcal{F}, P)$	Probability space
$\mathbb{R}^d$	d-dimensional real coordinate space
$\mathcal{B}(\mathbb{R}^d)$	Borel set class of $\mathbb{R}^d$
$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$	Measurable space related to $\mathbb{R}^d$
X	d-dimensional random vector, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ from $(\Omega, \mathcal{F}, P)$ -measurable function
$\mathbf{x}$	d-dimensional variable vector in $\mathbb{R}^d$
$P_X(\mathbf{x})$	Distribution of X in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
$F(\mathbf{x})$	Cumulative distribution function of X in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
$p_X(\mathbf{x})$	Probability density function of X in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
D	Statistical distance
$\mathbf{D}$	Statistical distance matrix of D
$\mathbb{E}[X]$	Expected value of X
$\text{cov}[X]$	Covariance matrix of X
$\text{corr}[X]$	Correlation matrix of X
$\text{cov}[X_1, X_2]$	Cross-covariance matrix of X_1 and X_2
$\text{corr}[X_1, X_2]$	Cross-correlation matrix of X_1 and X_2
$\boldsymbol{\mu}$	d-dimensional mean vector
$\boldsymbol{\Sigma}$	d-dimensional covariance matrix
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	d-dimensional normal distribution defined by means vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

3. Symbols Related to Hierarchical Clustering

S	Cluster label set
$ S $	Cardinality of set S
n_i	Cluster label
$\delta(a, b)$	Distance function, map of which is $\delta: S \times S \rightarrow \mathbb{R}$

Fig. 3

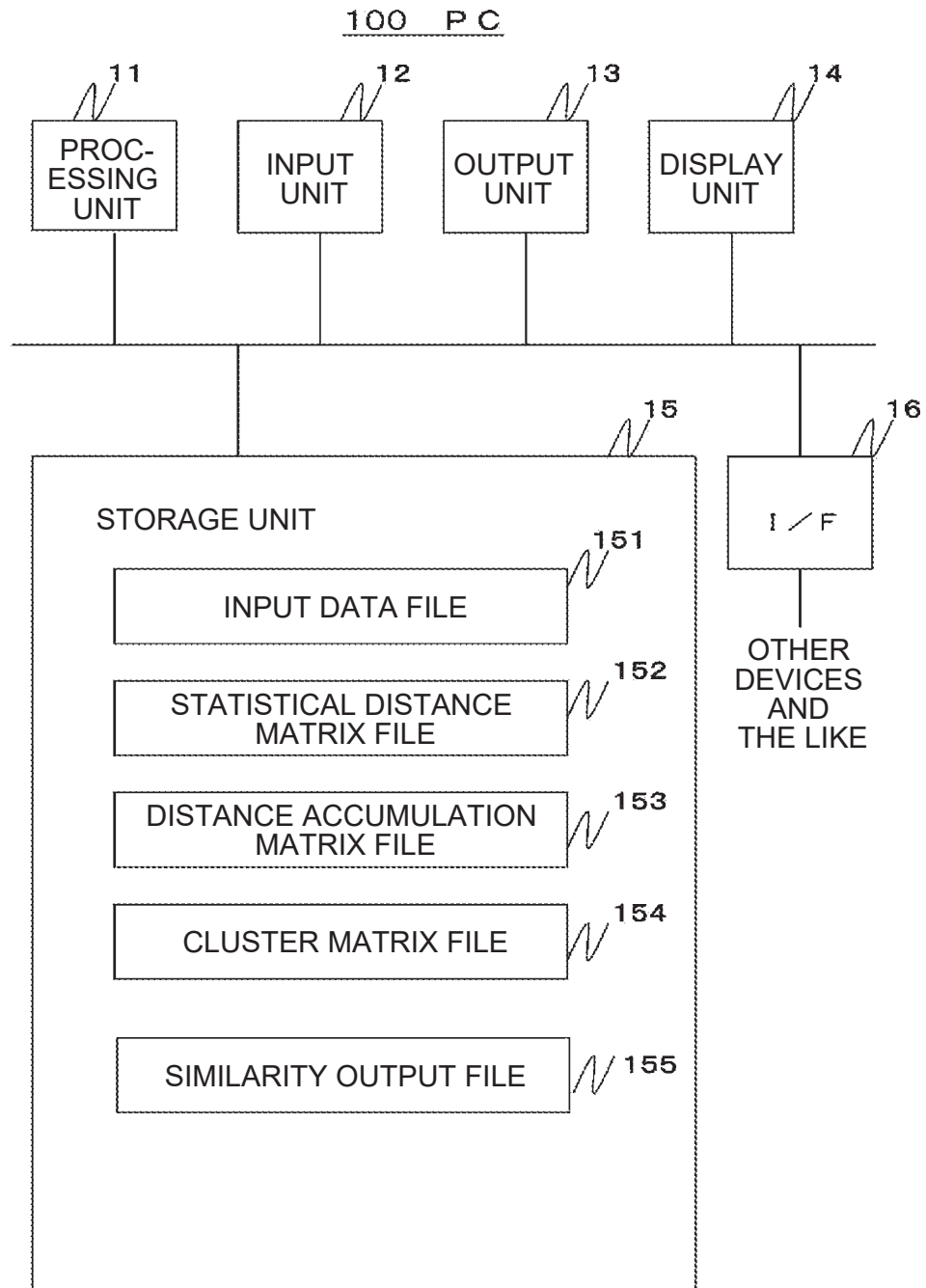


Fig. 4

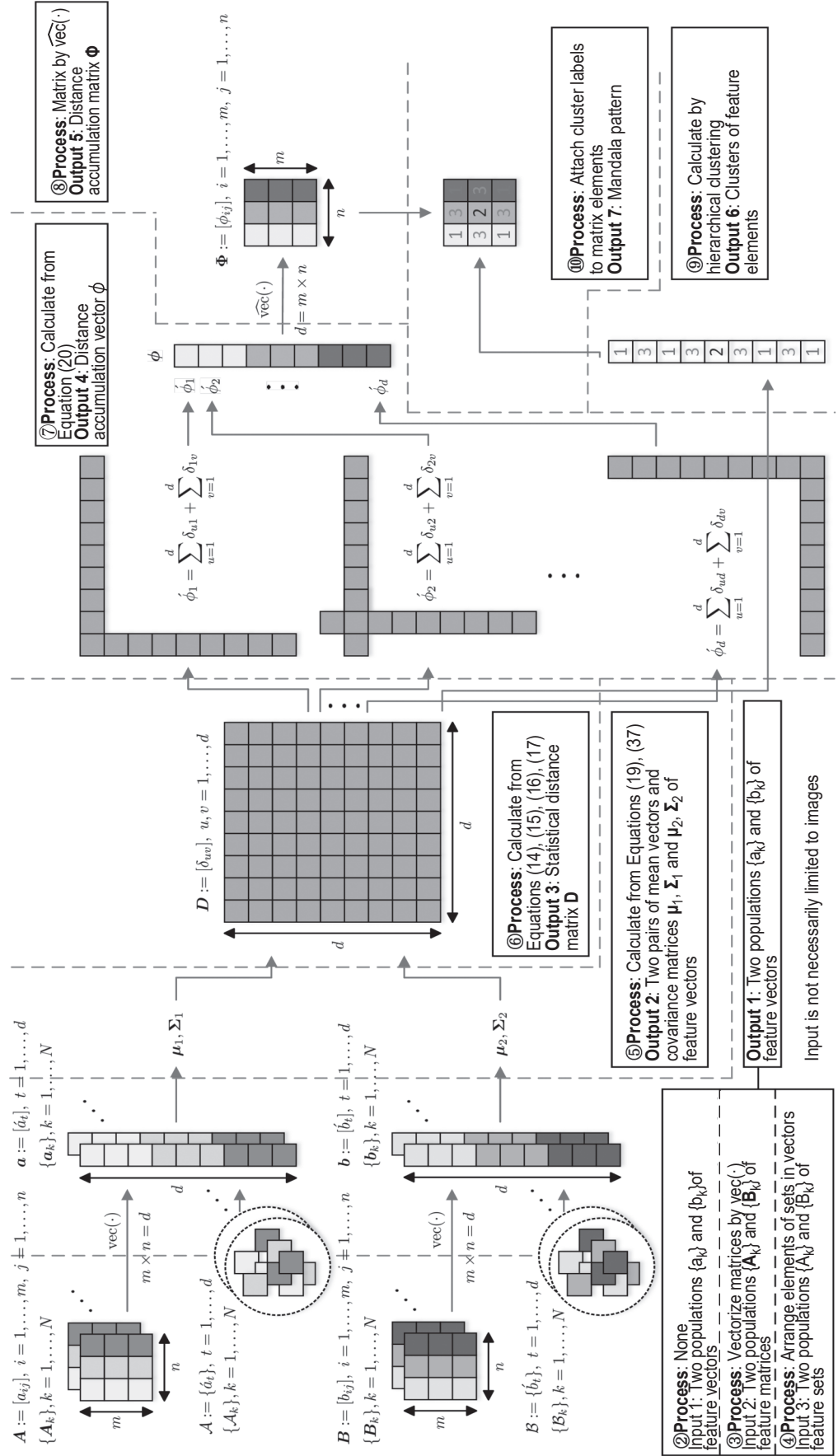


Fig. 5

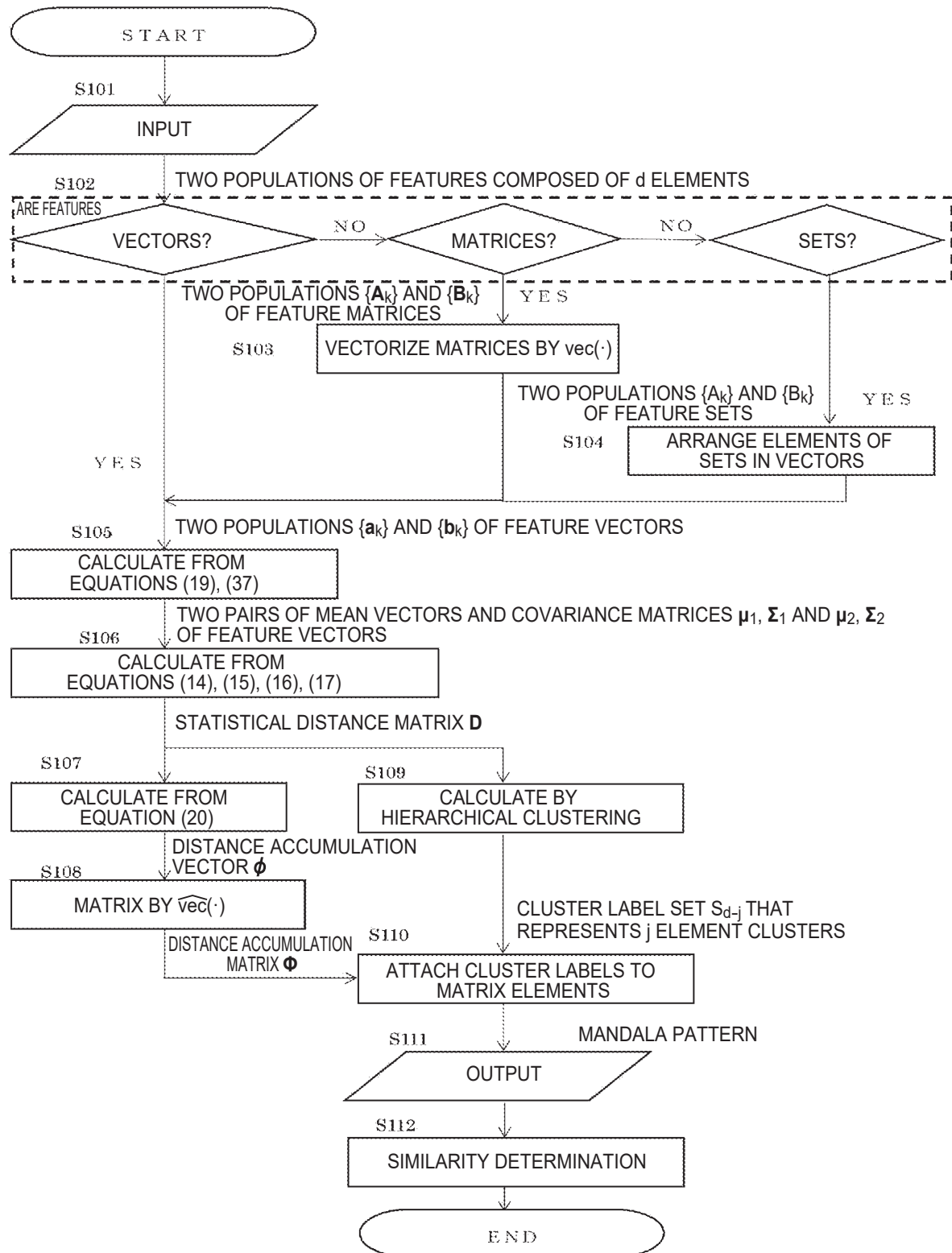


Fig. 6



([6] A. Krizhevsky, V. Nair, and G. E Hinton.
 CIFAR-10 (Canadian Institute for Advanced Research).
 Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.
 'Learning multiple layers of features from tiny images',
 Alex Krizhevsky, 2009. "The CIFAR-10 dataset", April 8, 2009
<https://www.cs.toronto.edu/~kriz/cifar.html>)

Fig. 7

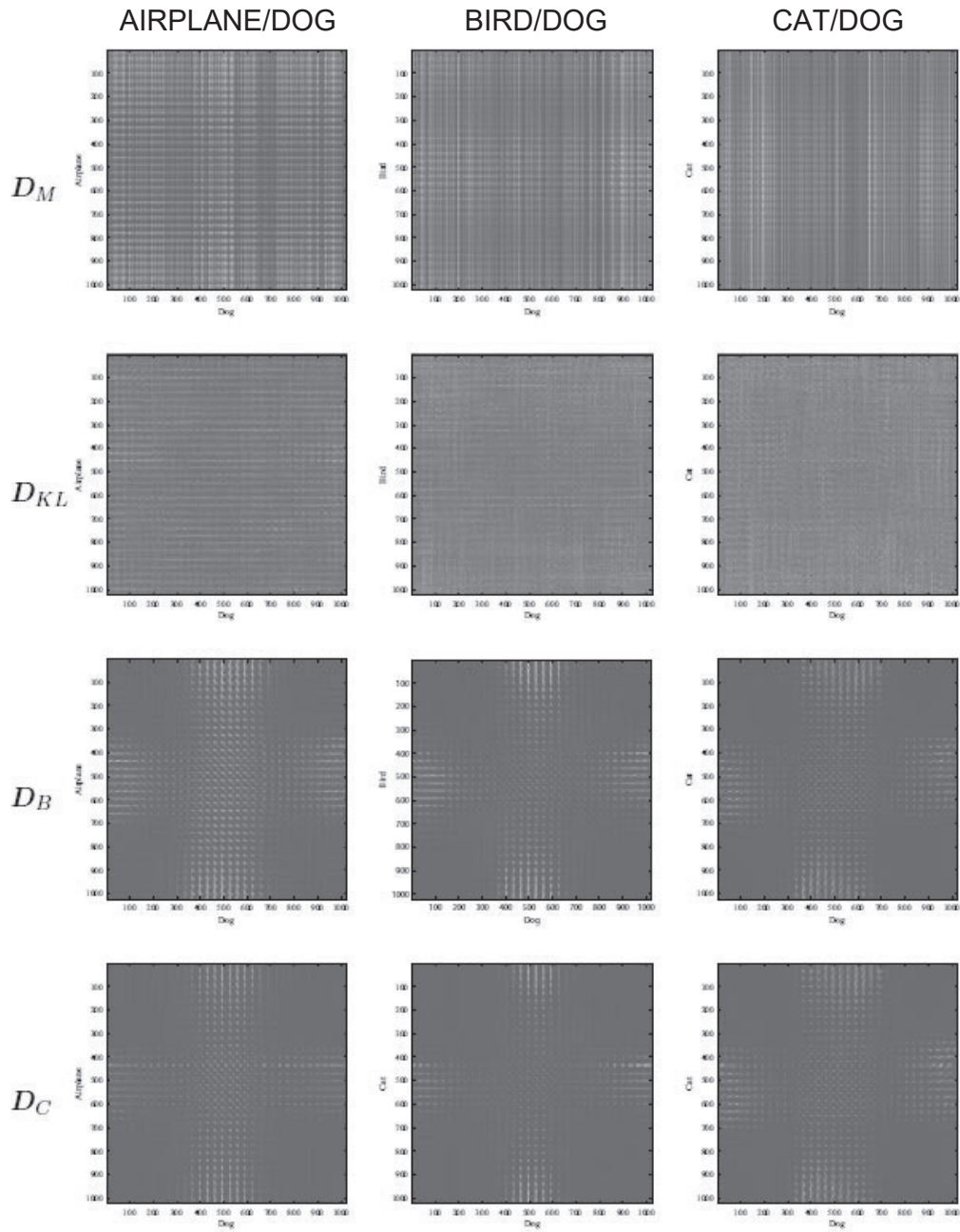


Fig. 8

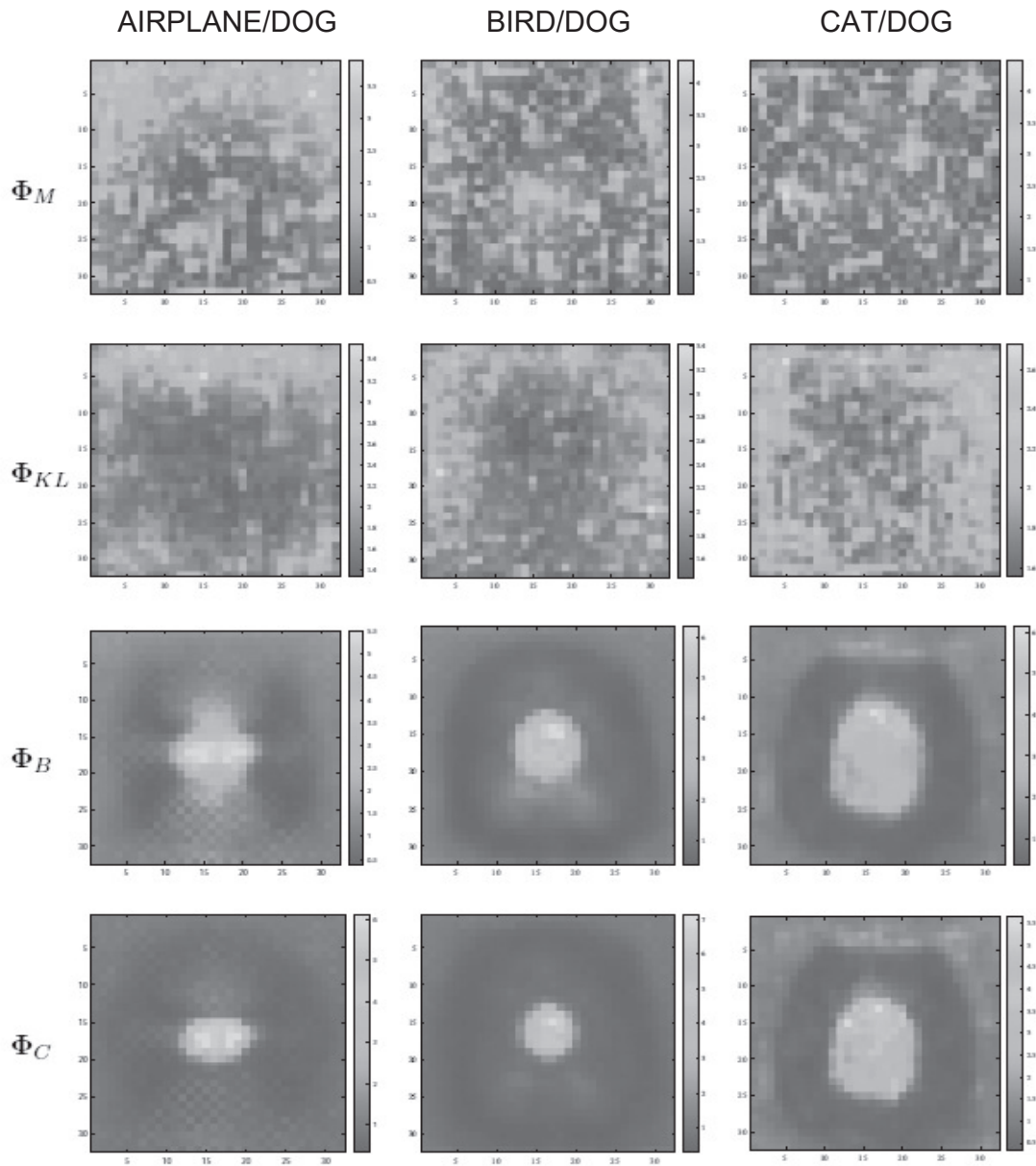


Fig. 10

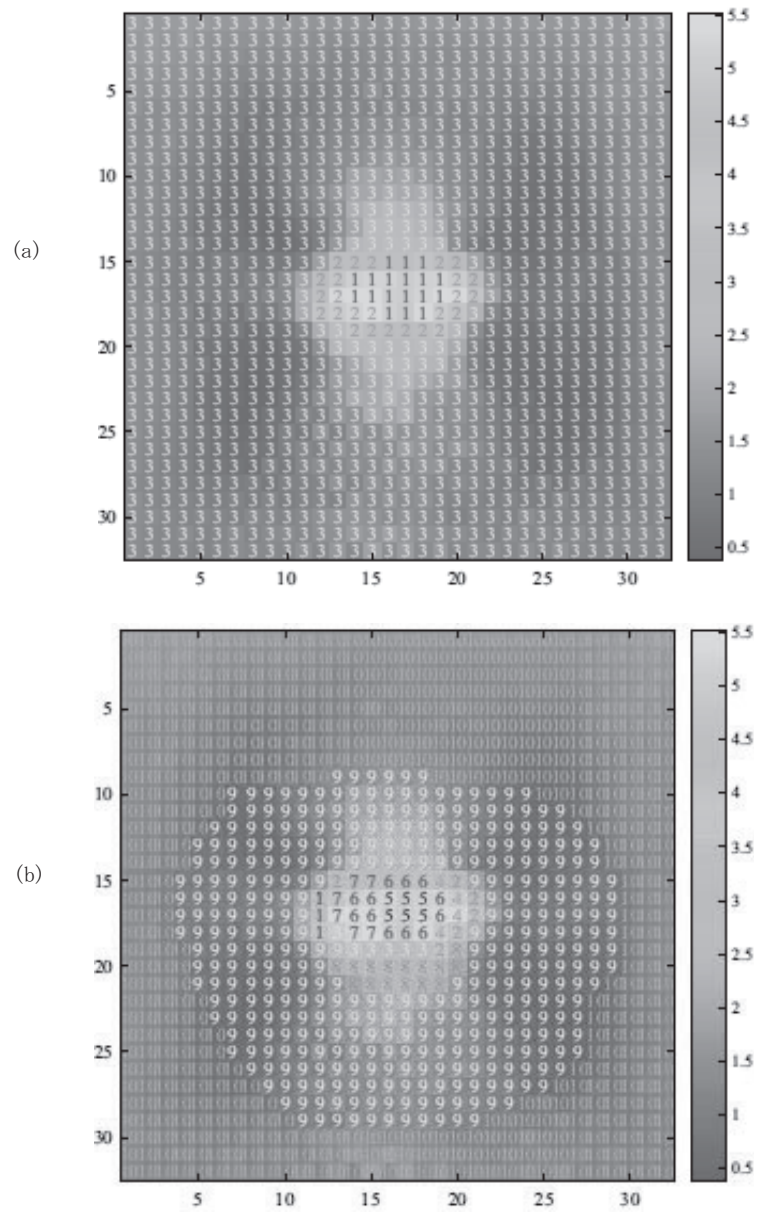


Fig. 11

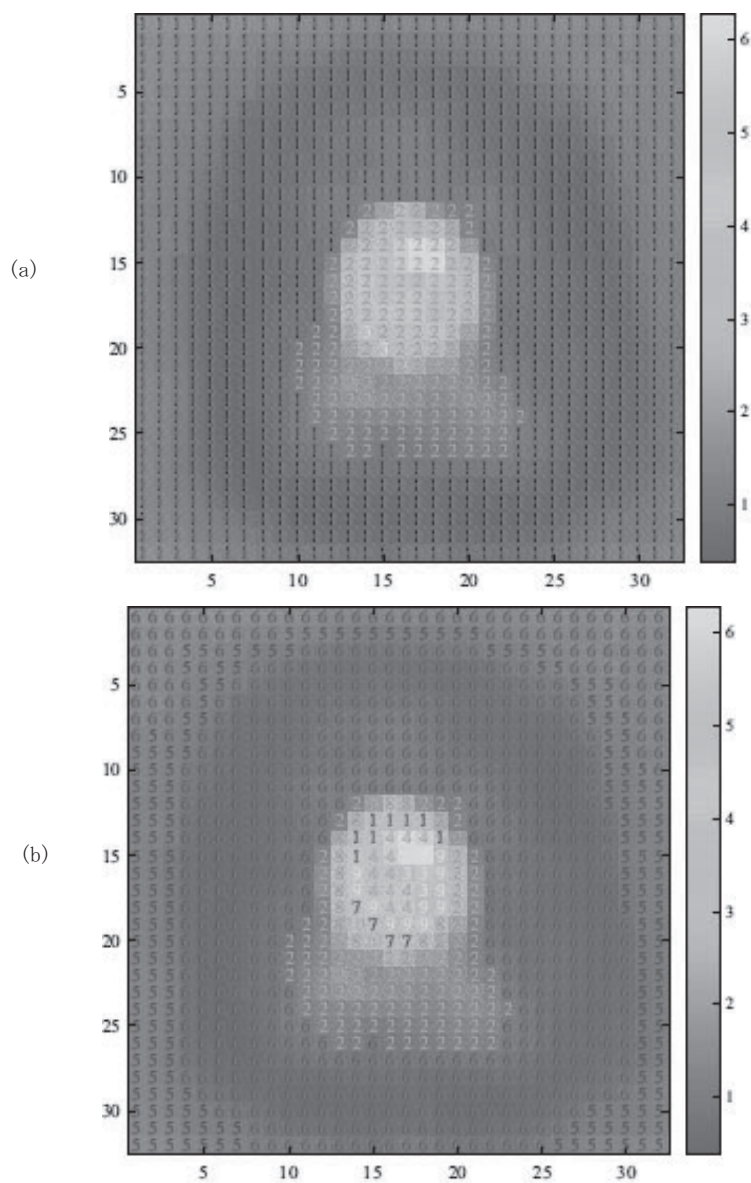


Fig. 12

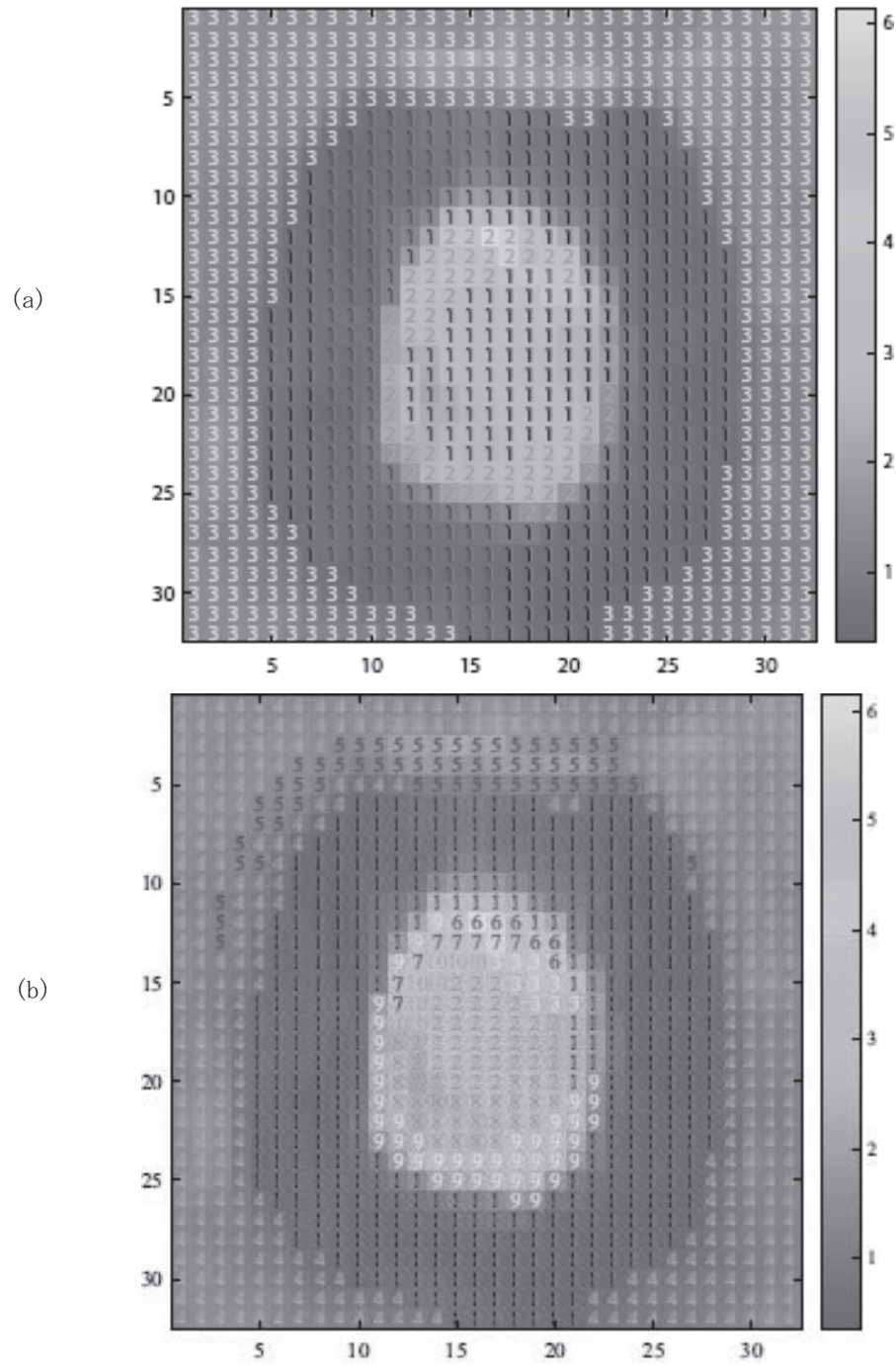


Fig. 13

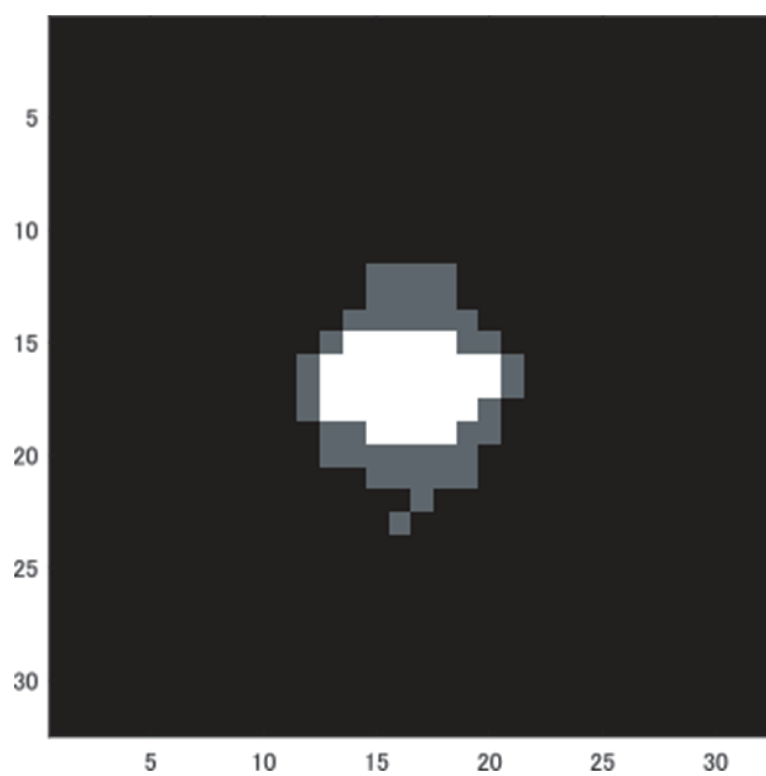


Fig. 14

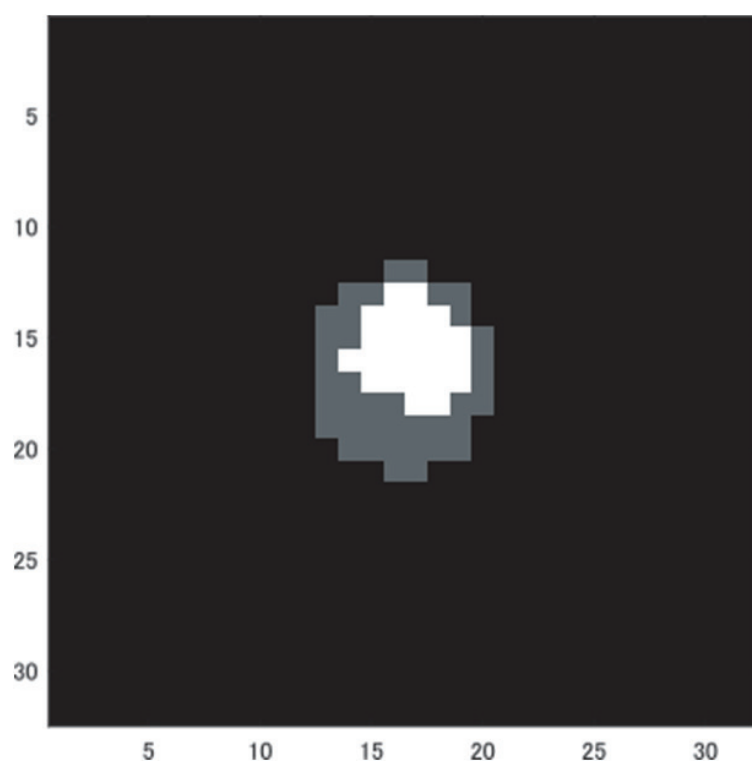


Fig. 15

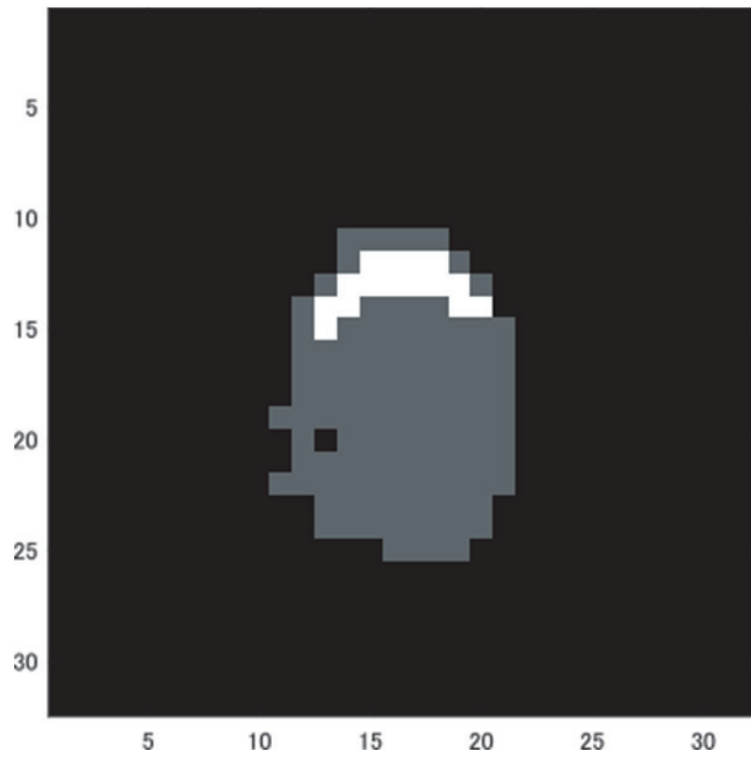


Fig. 16

	AIRPLANE AND DOG	BIRD AND DOG	CAT AND DOG
N_{low}	951	962	895
N_{medium}	41	38	114
N_{high}	32	24	15
$N_{\text{medium}} + N_{\text{high}}$	73	62	129
ρ	0.4381	0.3871	0.2481

Fig. 17

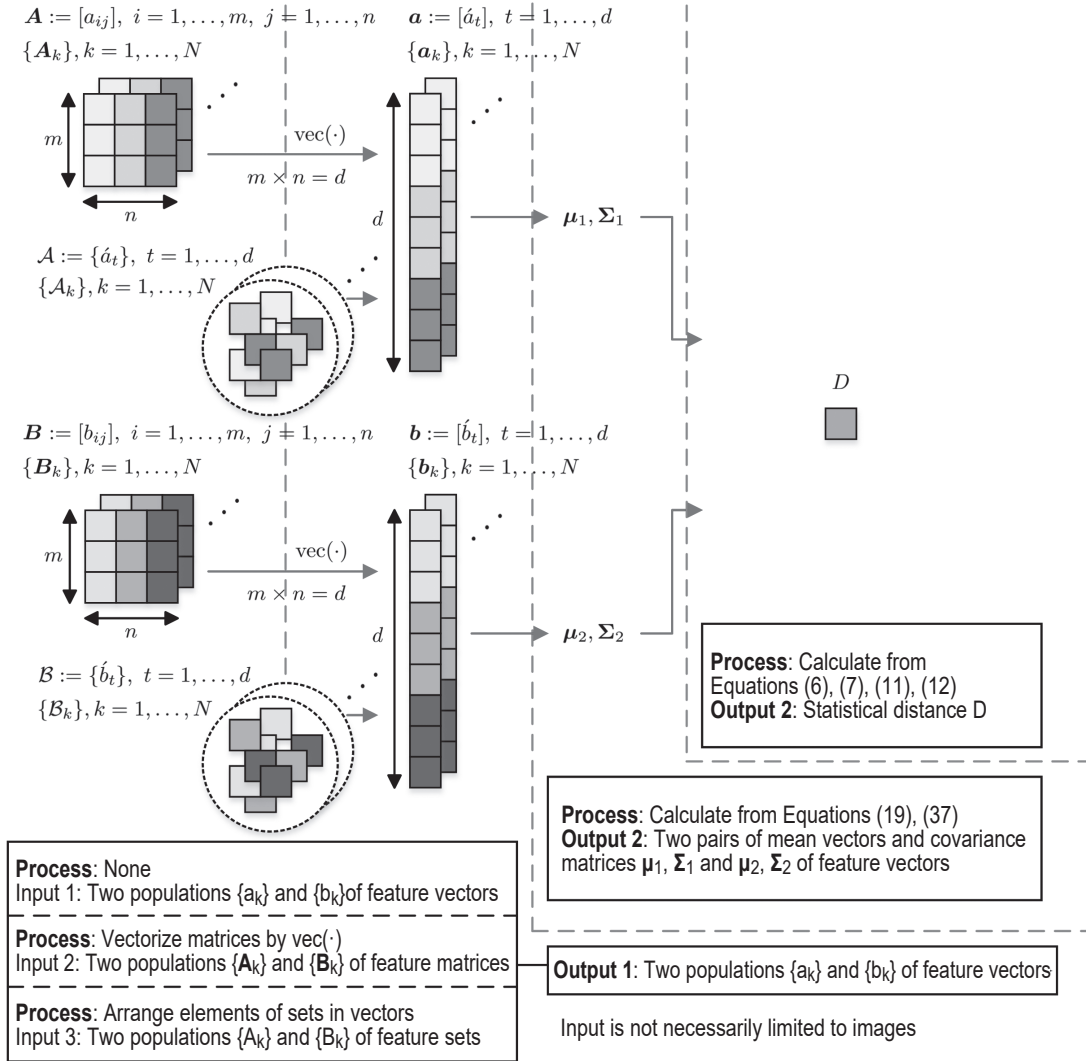


Fig. 18

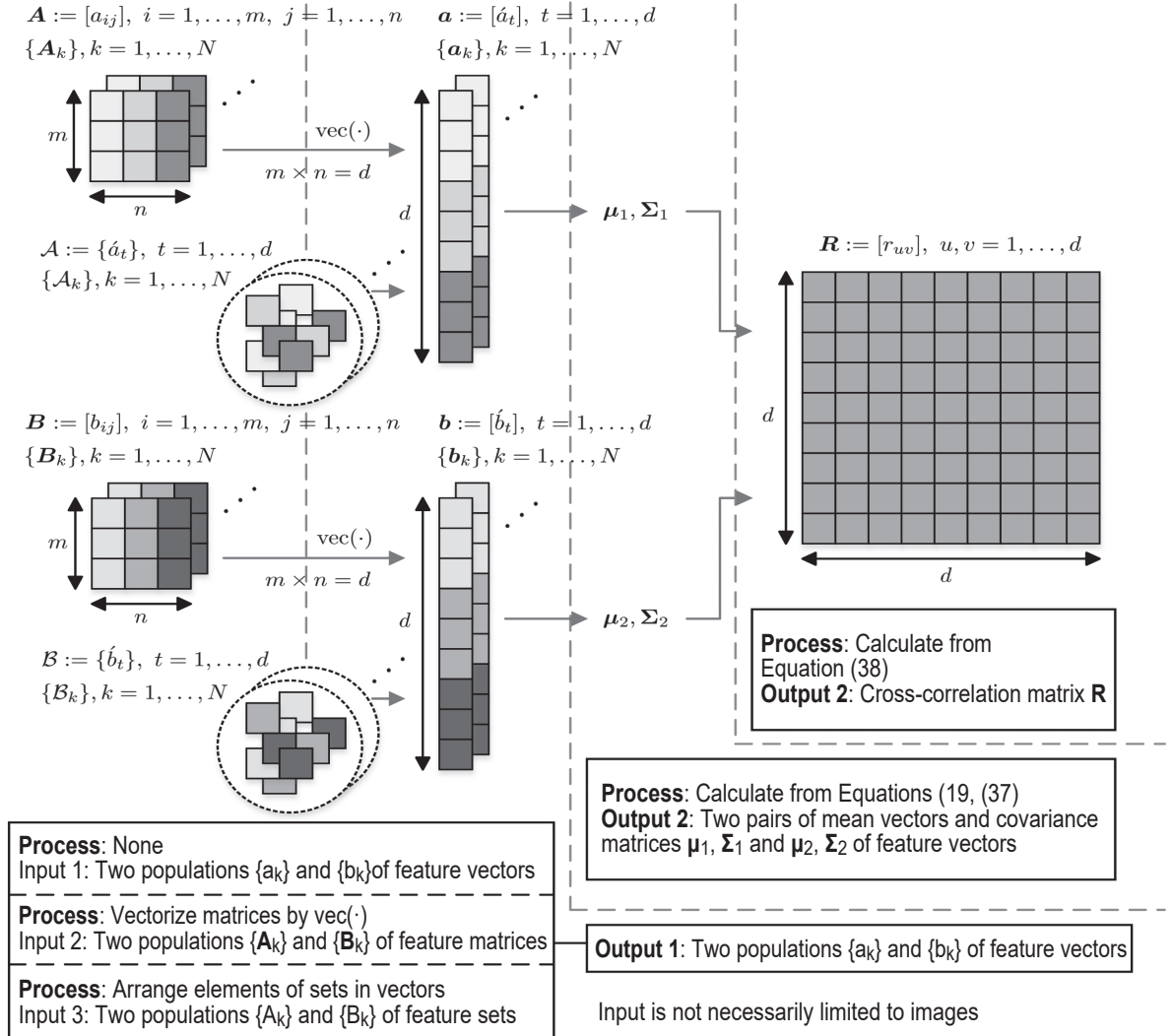


Fig. 19

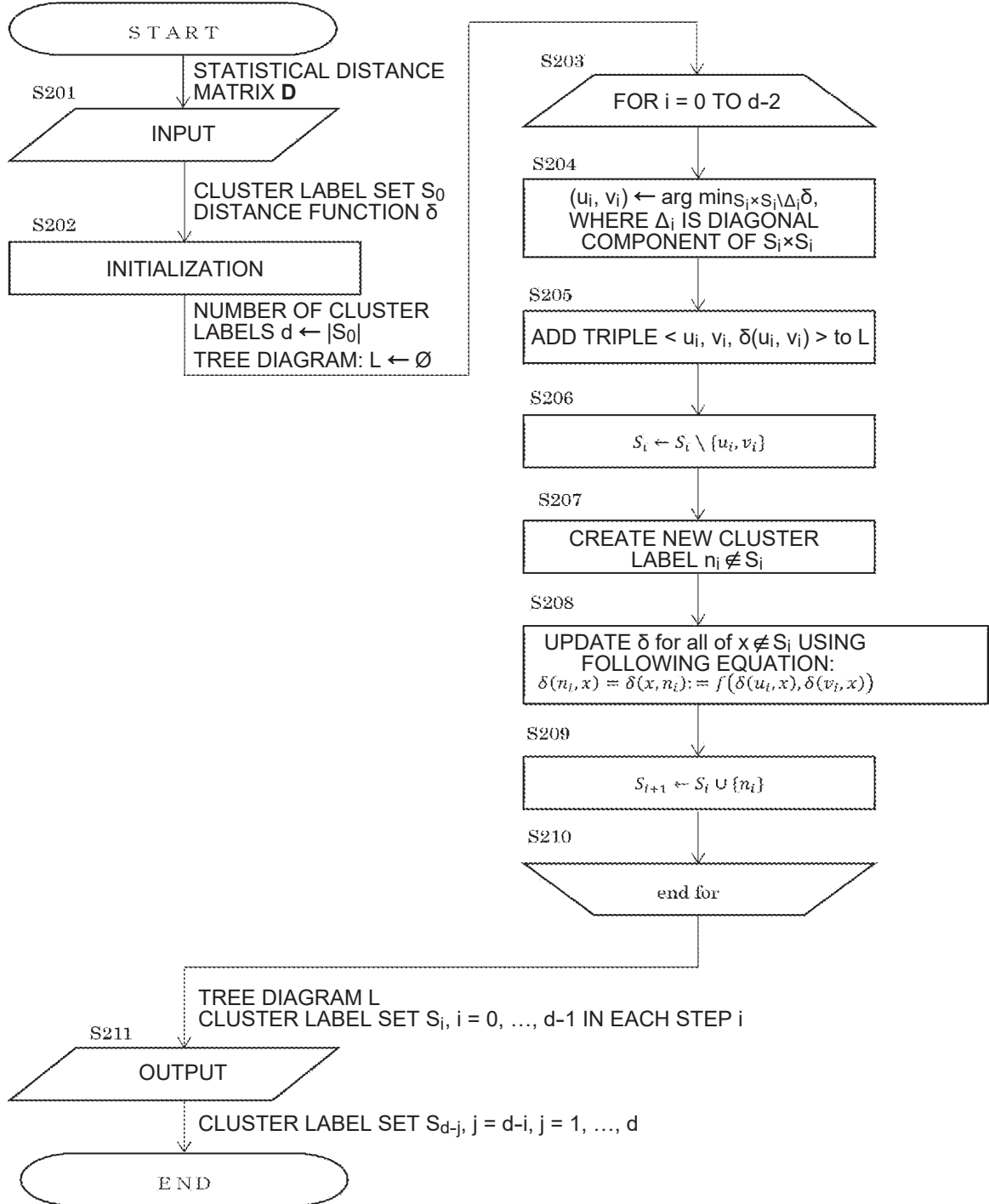


Fig. 20

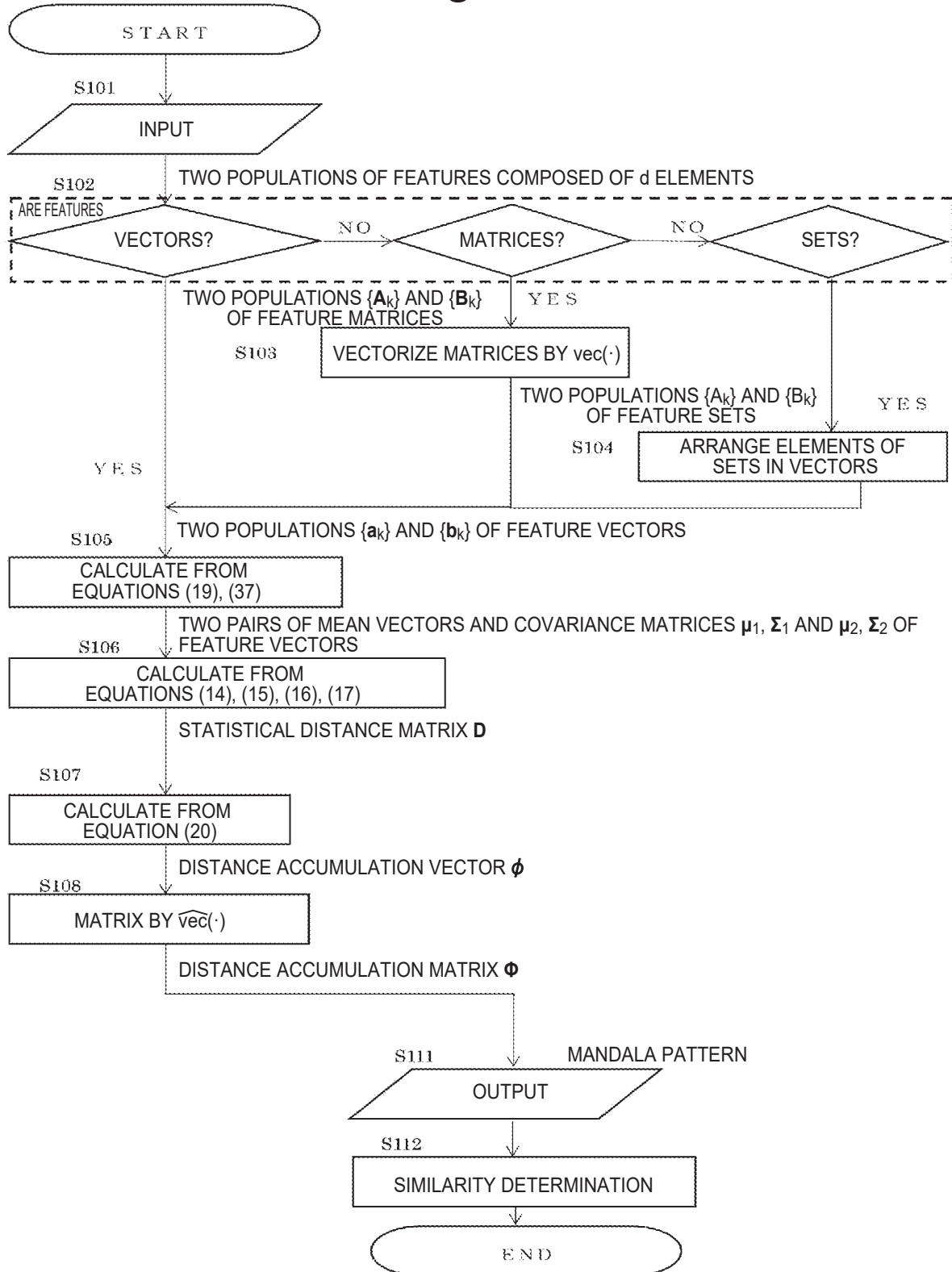


Fig. 21

