

No.

# 自動樹木検索システム開発業務 報告書

平成2年3月

国際協力事業団

林 開 発
8 9 - 3 8
J R



# 自動樹木検索システム開発業務

## 報告書

JICA LIBRARY



1085165(7)

21565

平成2年3月

国際協力事業団

国際協力事業団

21565

目 次

要約

第1章 システム開発の背景	1
第2章 システムの目標	2
2.1 システムに対する要求項目	2
2.2 システムの設定目標	2
2.2.1 システムに対する考慮点	2
2.2.2 システムの基本構造	3
第3章 システムの構造と構成	5
3.1 システムの構造および流れ図	5
3.2 システムの構成	6
3.2.1 システムの構成図	6
3.2.2 システムを構成する装置	7
第4章 理論的背景	9
4.1 抽出する特徴パラメータの概要	9
4.2 特徴パラメータの定義および抽出方法	10
4.2.1 特徴パラメータの定義	10
4.2.2 特徴パラメータの抽出方法	11
4.2.3 特徴パラメータの抽出手順	15
4.3 特徴パラメータの分布の表現方法	16
4.3.1 分布の形状およびその表現	16
4.3.2 データベースの内容	18

4.4 特徴パラメータの分布の評価	18
4.4.1 評価関数	20
4.4.2 評価基準	20
4.4.3 データベースの評価の流れ	20
4.5 分布をもつ特徴からの検索方法	21
4.5.1 マハラノビス距離	21
4.5.2 マハラノビス距離による判定	22
4.5.3 認識率と誤認識率	24

## 第5章 性能確認試験の方法 25

5.1 検索対象樹種	25
5.2 撮影倍率の影響	27
5.2.1 基準長さ	27
5.2.2 頂点数に関する補正	27
5.3 シミュレーションによる性能の確認方法	29
5.3.1 被検索樹葉	30
5.3.2 性能の確認方法	30

## 第6章 性能確認試験の結果 31

6.1 被検索樹葉の枚数の影響	31
6.1.1 被検索樹葉が1枚の場合の認識率	31
6.1.2 被検索樹葉の枚数と認識率の関係	32
6.1.3 標本数と認識率の関係	33
6.1.4 マハラノビス距離	34

6.2 特徴パラメータの影響	37
6.2.1 単一の特徴パラメータによる認識率	37
6.2.2 特徴パラメータの有効性	53
第7章 結 論	54
第8章 問題点および今後の課題	55
8.1 対象となる樹葉の収集に関する提言	55
8.2 データベースの構造およびデータの扱いに関する提言	57
8.3 検索の結果の表示に関する提言	58
APPENDIX 1 標本の樹種と母集団の組合せによる 全マハラノビス距離の分布の違い	59
APPENDIX 2 特徴パラメータ別の確率密度分布および近似した正規分布	63
APPENDIX 3 特徴パラメータ抽出用プログラムのフローチャート	83





## 要 約

熱帯地域に植生する樹木の種類は数千、同じ科・属の樹種が数百以上にもまたがる。しかし、これらの内で有用な樹種は数十種類に過ぎないと言われている。この有用樹種の分布状況の把握や樹林保護の観点から造林に適性な樹種の認定などが重要な課題であり、先進国始め開発途上国においても最近では有用木材、造林適性樹種の研究開発が促進されている。しかしながら、その基本となる樹木の検索・同定にあたっては専門学者あるいは数少ない木材利用専門家の経験と知識に頼らざるを得ないのが現状である。そこで少数の専門家に代わる樹木の自動検索システムを開発する必要がある。

本システムは国際協力事業団からの業務委託により開発を行った自動樹木検索システムである。基本的には無作為に採取された樹木の葉に画像処理を行い抽出された特徴パラメータと、データベースとしてコンピュータ内に保存された各樹種に関する樹木の葉の特徴パラメータとを比較して検索し樹種を特定するというものである。本システムに要求される機能としては、限定された50種以上の樹木の分類検索を現段階では少なくとも80%以上の確度で行えること、また対象の樹木に関する特定の知識を有しない人でも簡単に操作できること、他にシステム自体の問題としては検索結果の信頼性を評価できること、対象物の増加や樹種の変化に伴い柔軟に対応できる構造にすることなどをあげ検討を行った。開発されたシステムは大別して特徴抽出、データベース、検索の3つの部分から構成される。まず、特徴抽出部にはCCDカメラを通して得られた樹葉の外部形状を画像解析することにより幅、面積など数十種の特徴パラメータを抽出する。データベース部ではそれぞれの特徴パラメータが分布を持つため、標準偏差、平均値を使い各分布の特徴の表現を行っている。検索部では特徴抽出部で得られた特徴パラメータとすでに持っている特徴の分布を用いてマハラノビス距離を計算し、その大小で最も確からしい樹種を決定している。なお、これらの制御を行うコンピュータには、安価で機能性の高いパーソナルコンピュータを用いた。



# 第1章 システム開発の背景

熱帯地域に植生する樹木の種類は数千、同じ科・属の樹種が数百以上にもまたがる。しかし、これらの内で有用な樹種は数十種類に過ぎないと言われている。この有用樹種の分布状況の把握や樹林保護の観点から造林に適性な樹種の認定などが重要な課題であり、先進国始め開発途上国においても最近では有用木材、造林適性樹種の研究開発が促進されている。しかしながら、その基本となる樹木の検索・同定にあたっては植物分類学者あるいは数少ない木材利用専門家の経験と知識に頼らざるを得ないのが現状である。そこで少数の専門家に代わる樹木の自動検索システムを開発する必要がある。

本報告書は、国際協力事業団からの業務依頼に基づき開発された「自動樹木検索システム」に関する報告である。本業務依頼により開発した自動樹木検索システムは、樹葉の形状を画像処理技術により求め、これらを使って自動的に樹種を検索するものであり、本報告書は、(1)システムの目標、(2)システムの構成、(3)理論、(4)適応例等から成る。

## 第2章 システムの目標

### 2.1 システムに対する要求項目

本業務依頼により開発する樹木自動検索システムが有すべき機能上の仕様は以下の通である。

- (1) 検索機能 …… 入力された50種類以上の樹木の分類検索が可能なこと
- (2) 使用者 …… 対象の樹木に関する特別の知識を有しない人
- (3) 検索能力\* …… 上記50種類以上の樹木を少なくとも80%以上の確度で分類検索が可能なこと
- (4) その他 …… 操作が簡単で取扱が容易なこと

\* 具体的な数値は明確にされていないが、植物分類学者でもおよそ80%程度の確度であると言われている。

### 2.2 システムの設定目標

#### 2.2.1 システムに対する考慮点

具体的な利用状況を想定し、前項の要求目標以外に考慮すべき点について検討を行った。

- (1) 簡易性 …… 本システムの利用者は、樹木やコンピュータなどに特別な知識を持たない人が利用することを前提にシステム的设计を行う必要がある。従ってなるべく扱いやすく、誤操作を招かないようにする。特に特徴パラメータのデータベース作成時のデータ入力にはなるべく拘束を少なくし、簡単に入力できる方法を採用した。従って、本システムでは、

- ① システムを起動する場合に特別の操作を含まないこと
- ② 対象となる樹葉に特別の処理を施す必要のないこと
- ③ 検索過程に特別の知識や判断を要求しないこと

等を満足する必要がある。

- (2) 操作性 …… 操作は単純にし、各ステップを指示に従い進めることにより、データの蓄積や検索ができるようにする。
- (3) 信頼性 …… 実際に検索システムを実用化する上で、検索システムでは正しい答を判定する割合(認識率)が最も重要であり、その割合は、限りなく100%に近いことが望ましい。検索の認識率が低ければ、その検索結果の信頼性は低下し、余計な労力を増やす結果となることは明らかである。一方、判定結果の確度が目標値を越えることは当然であるが、不確実さをどの様に示すかが重要である。判定結果の確実さを明確にし、他の樹種である確率も併せて示す。
- (4) 柔軟性 …… 実用化の段階では対象物の種類の数が多くなり、検索効率が問題となる。また、対象とする樹種が変化した場合等に、容易にシステム内容の変更が可能でなければならない。本システムは、将来対象物が増加したり、樹種が変化した場合にも対応できる構造とし、更に認識率の低下を防ぎながら検索効率を上げていくことが可能な様に柔軟性のある構造とする。

## 2.2.2 システムの基本構造

前項の検討結果に基づき、システムの基本構造を以下の様に定めた。

- (1) 検索方法 …… 樹葉の形状に関する特徴を画像解析により求め、これらの特徴パラメータをデータベースと比較し検索を行なう方式とする。
- (2) 対象樹種 …… 本システムは、データベースの構築が可能であれば対象樹種に制限はない。しかしながら、本年度はJICA短期派遣専門家によりパプアニューギニアで採取された50種以上の樹種を対象とする。
- (3) 特徴抽出法 …… 特徴の抽出には、対話式を始め数々の方法が考えられる。本システムでは、操作する人の意志が入らぬように画像解析によるデジタル値を用いる。また、特徴抽出の対象は、平面で扱える樹葉のみとする。即ち、CCDカメラを通して樹葉を撮影し、その画

像から種々の特徴データを引出す。

- (4)特徴パラメータ …… 樹葉の特徴を表すパラメータには、形状（長さや幅など）に関するパラメータ、質（色彩や硬さなど）に関するパラメータ、微視的（葉脈や細胞組織等）なパラメータ等が考えられる。画像解析では、いずれも取扱可能であるが、樹葉に特別な処理を施さないことや樹葉の保存状態や季節により特徴値が大きく異なることによる誤判断を避けるために、本システムでは形状に関する特徴のみを扱うこととする。

# 第3章 システムの構造と構成

## 3.1 システムの構造および流れ図

本システムの基本的な考え方は、画像処理により樹木の葉より抽出された特徴パラメータを、データベースとしてコンピュータ内に保存された各樹種に関する特徴パラメータを用いて検索し樹種を特定するというものである。従って本システムは大別して、(1)特徴抽出、(2)データベース、(3)検索 の3つの部分から成る。基本的な構成および流れを図 3.1 に示す。

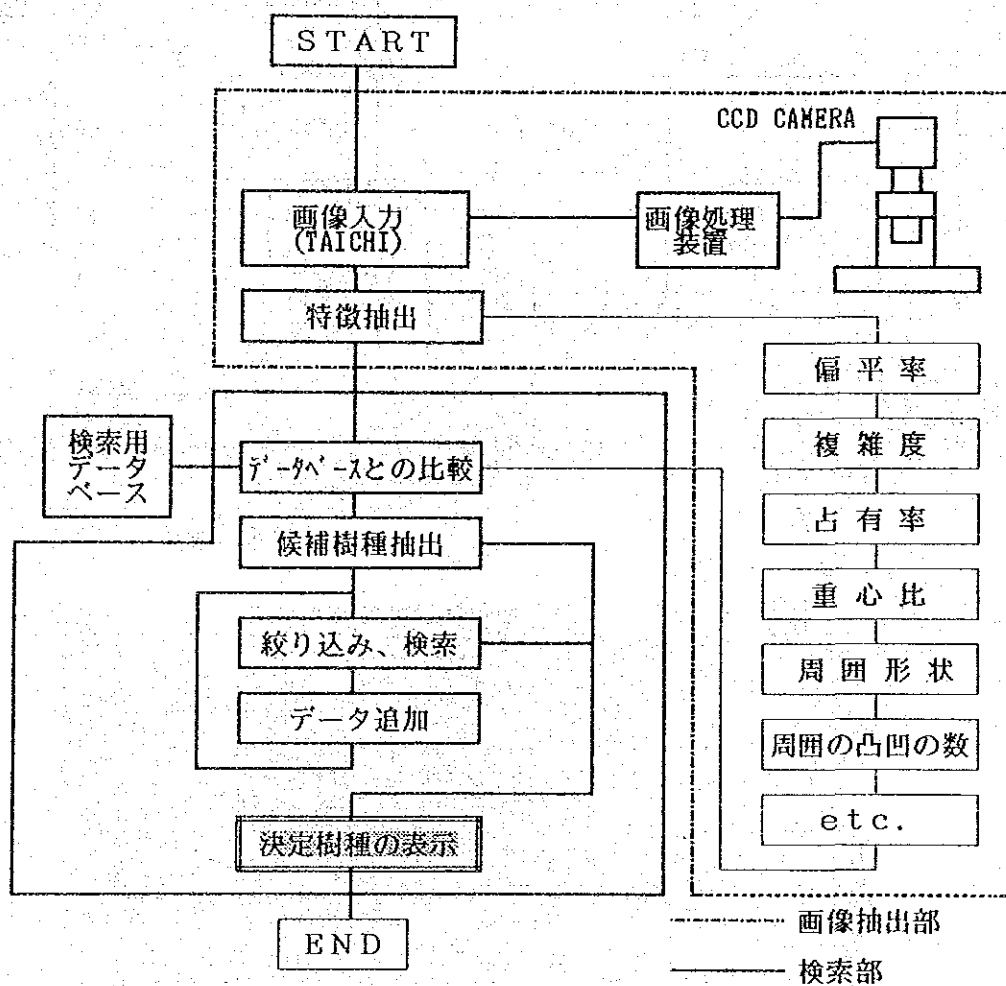


図3.1 自動検索システムの基本的な流れ

### 3.2 システムの構成

#### 3.2.1 システムの構成図

本システムの構成を図 3.2 に示す。

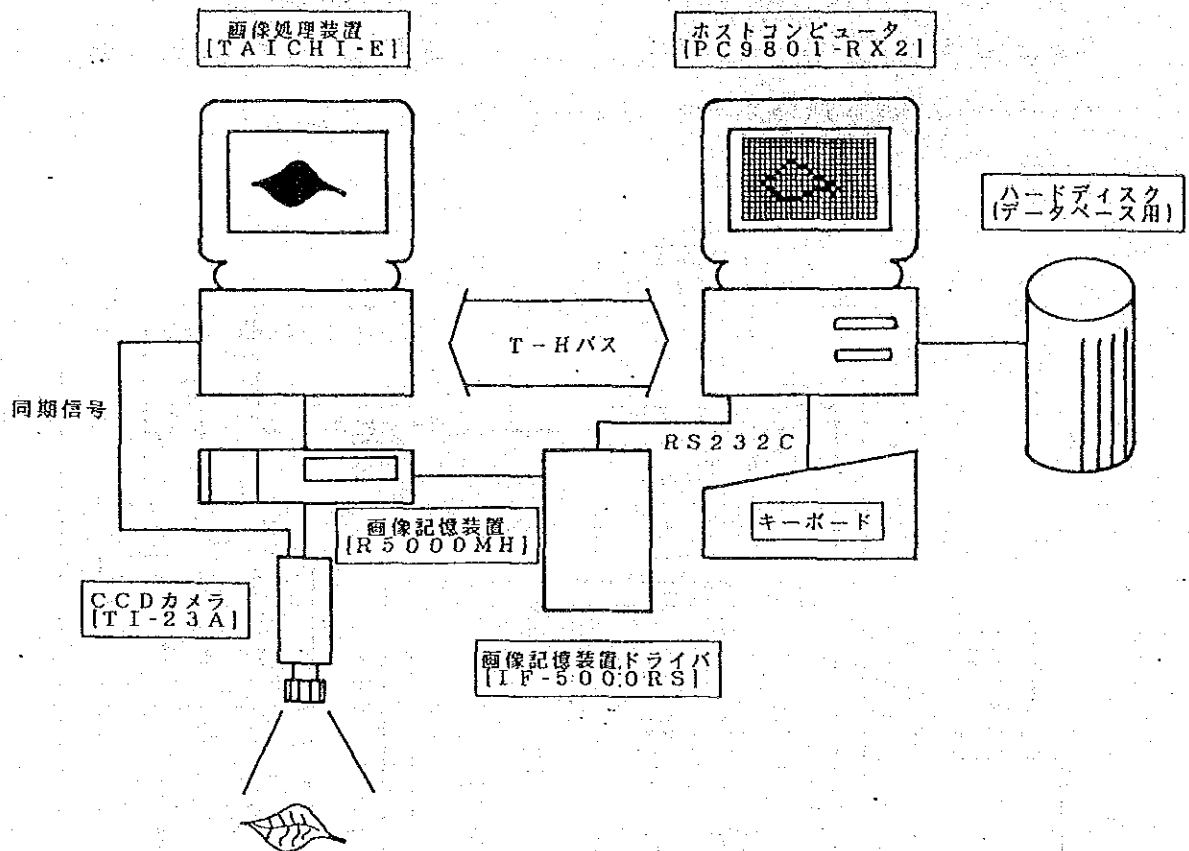


図3.2 実験装置の構成図



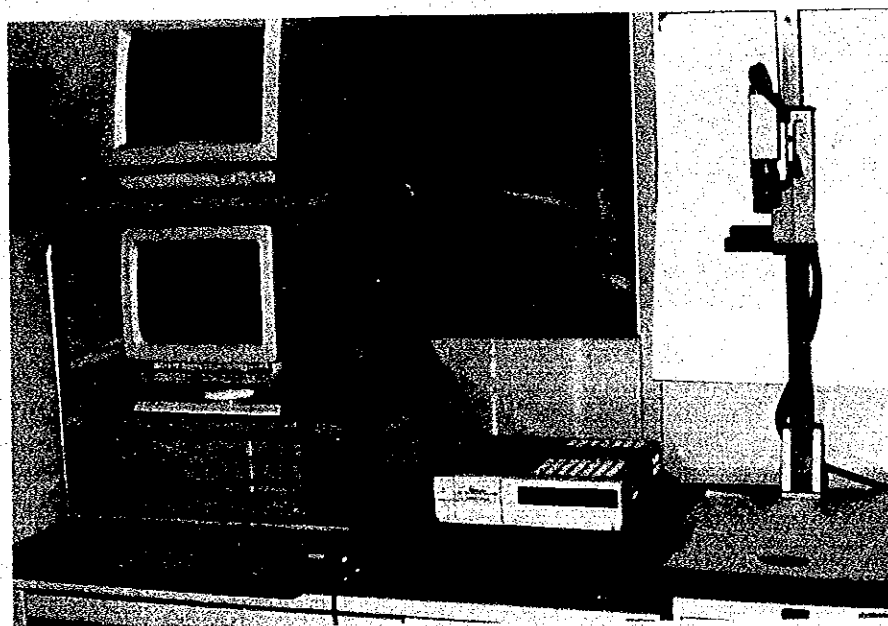
### 3.2.2 システムを構成する装置

本システムを構成する装置の概要を表 3.1 に、装置全体および各装置の写真を図 3.3 に示す。

表 3.1 実験装置の概要表

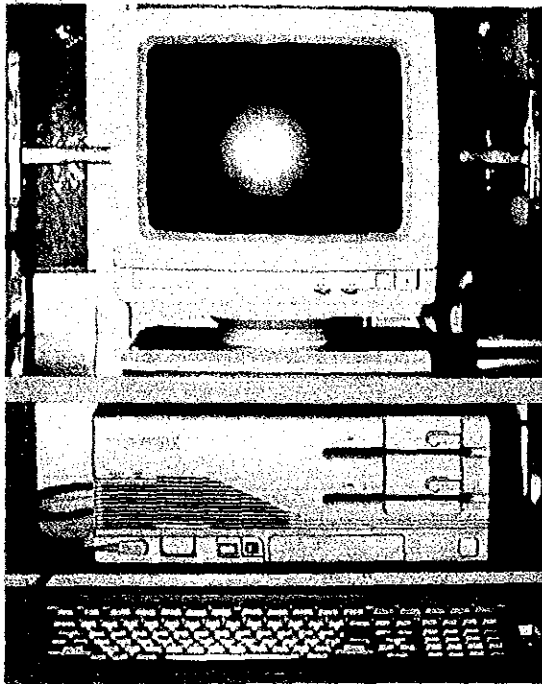
装置名	仕様・特徴	機能・大きさ
CCDカメラ (TI-23A)	撮像素子  画面サイズ 解像度 外形寸法	インターライン方式、 CCD固体撮像素子画素数 266664素子 (有効画素数 492V×512H)  1/2インチ水平 380本、垂直 350本 44W×45H×117.5D (mm)
レンズ	絞り 焦点距離	1:2.8 35mm
画像処理装置 (TAICHI-E)	機能 処理速度	画像処理機能 1画素 80nsec
コンピュータ	PC-9801 RX2	40MB ハードディスク内蔵
画像記憶装置 (R5000MH)	機能 信号方式 外形寸法	録画再生機能 NTSC標準方式準拠 330W×99H×365D (mm)
画像記憶装置 ドライバ (IF-5000RS)	外形寸法	62W×99H×200D (mm)

#### (1) 樹木自動検索装置

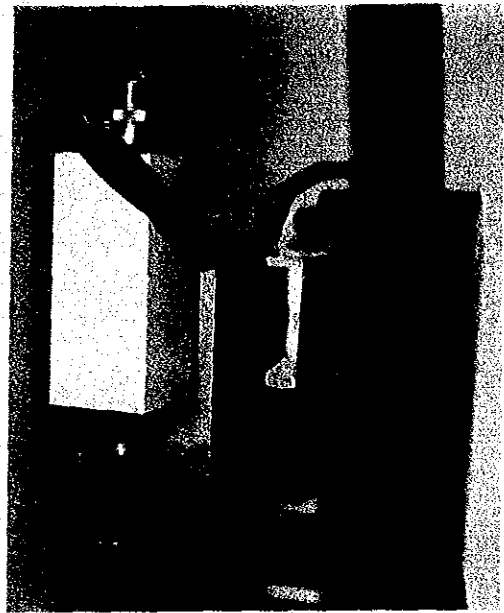




(2) 樹木自動検索装置構成機器

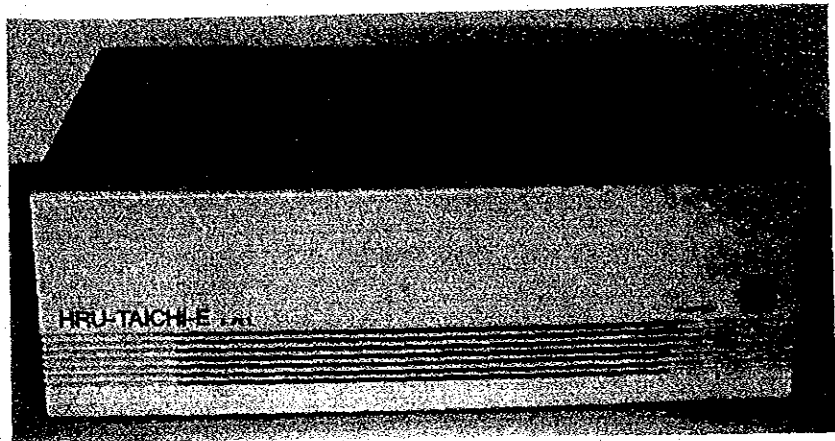


コンピュータ (PC-9801 RX2)

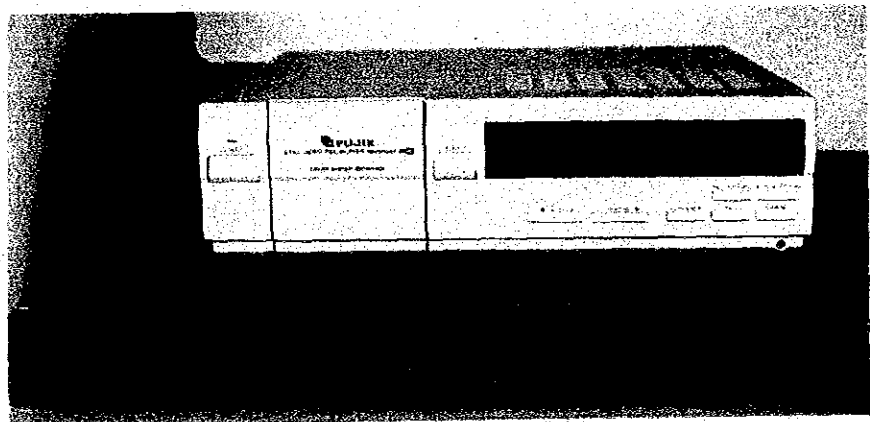


CCDカメラ

画像処理装置  
(TAICHI-E)



画像記録装置  
(R-5000-MH)  
(IF-5000RS)





## 第4章 理論的背景

本システムは、各種の樹葉がもつ特徴のばらつきを分布で表現し、これをデータベースに保存して、実際に検索したい樹葉からデータベースに保存されている特徴と同様の特徴を画像解析により求めデータベースと比較しながら検索を行うものである。従って、(1)抽出する特徴パラメータ、(2)データベースにおける特徴の分布の表現方法、(3)分布をもつ特徴からの検索方法 について明確にする必要がある。

### 4.1 抽出する特徴パラメータの概要

本システムでは、使いやすさ、特徴データの抽出の確実さなどから、CCDカメラを通して得られた樹葉の外形状を画像解析することにより、各種の特徴パラメータを抽出することとした。一般に画像処理を用いて対象物の特徴パラメータを抽出する場合、外形的な形状の特徴、表面の模様等の特徴、色や濃淡などの質的な特徴から特徴パラメータを抽出することができる。本システムで、形状から得られる特徴パラメータに限定した。理由は以下の通りである。

- (1) 表面の模様である葉脈の特徴は今後パラメータとして抽出できる可能性があるが、現段階では樹種による鮮明度の違いから一様に葉脈を取り出すことは難しい。
- (2) 色や濃淡は樹葉を採取してからの時間や季節などに左右され、また、照明や環境の違いにより変化するため、特徴が定まらない危険がある。本システムでは汎用性を考慮しているため、特殊な状況でのみ用いることができる色や濃淡などの質的な特徴は採用しないこととした。
- (3) 細胞や染色体の構造や細かな葉脈などは重要な特徴データであるが、これらを得るためには特殊な処理を施さなければ成らないため、本システムでは採用しない。

## 4.2 特徴パラメータの定義および抽出方法

### 4.2.1 特徴パラメータの定義

本システムで採用した特徴パラメータは、(1)基本パラメータ および (2)誘導パラメータから成る。基本パラメータは、樹葉から直接得られるパラメータであり、誘導パラメータは基本パラメータあるいは画像処理装置により得られるデータから演算により得られるパラメータである。

図 4.1 に、基本パラメータの定義を、また表 4.1 に全特徴パラメータの定義及び意味を示す。なお、表中の次元数は、1つの標本からその特徴パラメータが抽出できる数に対応している。

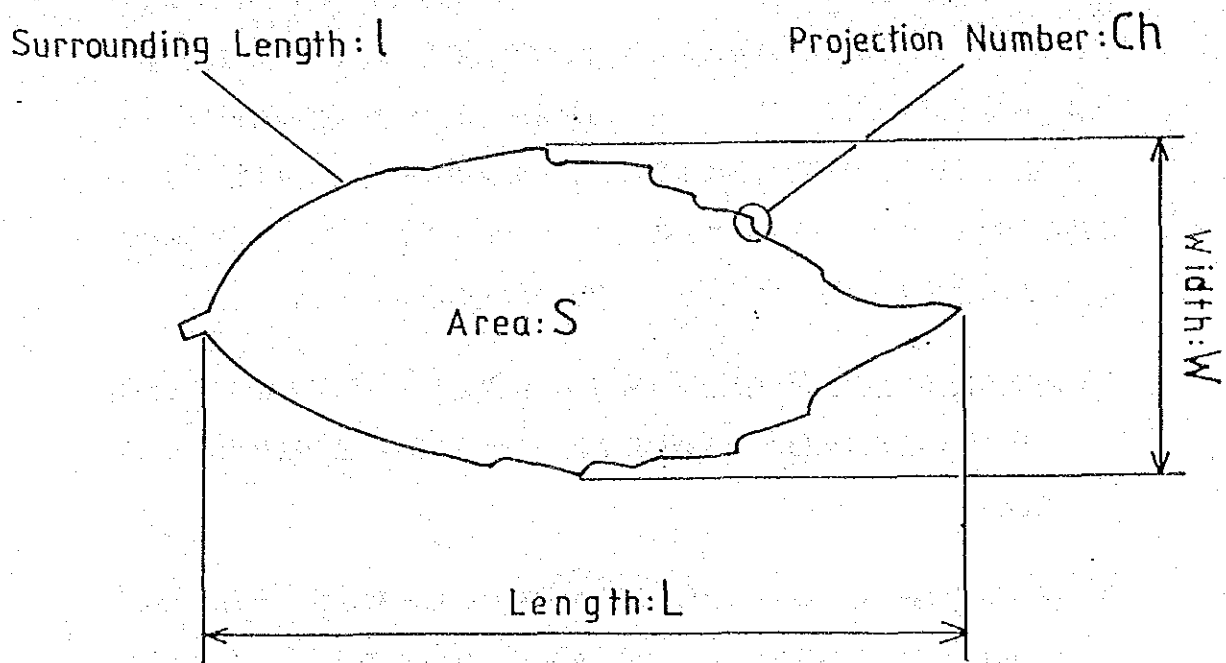


図 4.1 樹葉から直接得られる基本パラメータの定義

表4.1 各パラメータのもつ意味

パラメータ	記号	単位	パラメータのもつ意味	次元
長さ	L	mm	実際の樹葉の最大長さ	1
幅	W	mm	実際の樹葉の最大幅	1
周囲長	l	mm	実際の樹葉の寸法	1
面積	S	mm <sup>2</sup>	実際の樹葉の寸法	1
頂点数	Ch		樹葉周囲の突起の数	1
偏平率	W/L	…	葉の縦長、及び横長を見分ける	1
複雑度	l <sup>2</sup> /S	… …	葉の周囲の複雑さを見る	1
占有率	S/(L*W)	…	葉の四隅への張り出し具合を見る	1
重心比	Lg/L	…	先細と先太を見分ける	1
頂点率	Ch/P	…	単位長さあたりの突起の数を見る	1
周囲形状 (C <sub>1</sub> -C <sub>10</sub> )	…	…	周囲の細かい部分の複雑さを見る	10

#### 4.2.2 特徴パラメータの抽出方法

本システムでは、対象物をCCDカメラで撮影し、得られた図形を2値化（画像を白黒2階調に変換）した後、構成する画素（画像を格子状に分割した時の各点、または要素）を基に各値を算出した。前項の表4.1で示した各特徴パラメータの抽出方法は以下の通りである。

- (1) 長さおよび幅の抽出方法 …… 対象となる閉図形の中で、最大長および最大長に垂直方向の長さ、最大幅をそれぞれ長さL、幅Wと定義した。

CCDカメラおよび画像処理装置により得られる2値化された図形は、512×512の画素により記録される。単位長さ当りの画素数は撮影倍率により異なる。単位長さ当りの画素数は、撮影倍率を変えた場合には基準長さ(5.2.1, P.27参照)を撮影することにより求めることができる。

図4.2の場合、

$$\text{最大長} : M = (4^2 + 5^2)^{1/2} / bl = 41^{1/2} / bl \dots \dots (4.1)$$

$$\text{最大幅} : B = (1^2 * 5^2)^{1/2} / bl = 26^{1/2} / bl \dots \dots (4.2)$$

となる。ここで、blは単位長さ当りの画素数である。

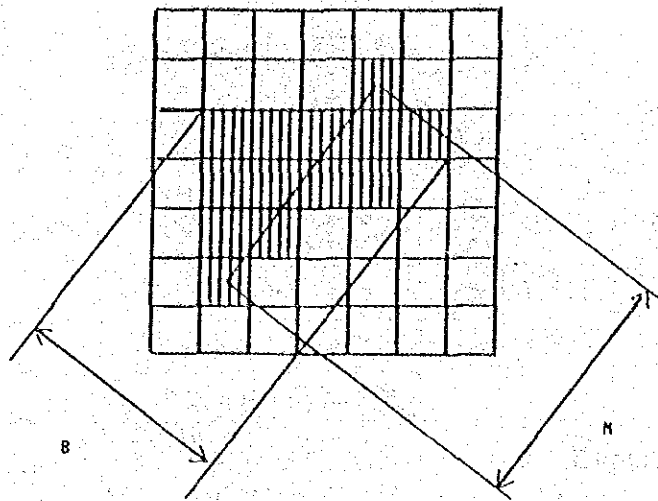


図 4. 2 図形の最大長および最大幅の例

(2) 面積の特徴抽出方法 …… デジタル図形では図形の面積は図形に含まれる画素数から求めることができる。n 個の画素から構成される閉図形の面積 S は、

$$S = n/b^2 \quad \dots\dots\dots(4.3)$$

として求めることができる。

図 4.3 の場合、面積 S は式(4.4)で求められる。

$$S = (4 \times 5 - 7)/b^2 = 13/b^2 \quad \dots\dots\dots(4.4)$$

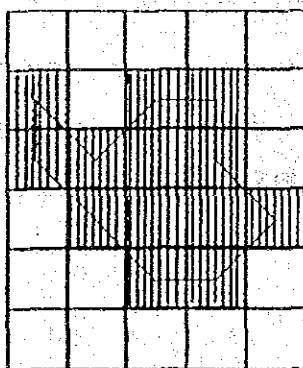


図 4. 3 面積の例



(3) 周囲長の特徴抽出方法 …… 周囲長は、境界画素が構成する長さを意味する。境界画素とは、2値化された閉図形に含まれる画素でその1部が背景に接している画素をいう。図 4.3 の場合、周囲長  $l$  は、

$$\text{周囲長} : l = (4/b1 + 6*2^{1/2}/b1) \dots\dots(4.14)$$

で表わされる。

(4) 頂点数の抽出方法 …… 本システムでは頂点は円状探索および偏角差分関数による頂点決定法と同様の方法を用いて頂点を決定しその数を数えた。撮影倍率の違いにより、頂点と判定するためのしきい値は異なるため、事前に検定が必要である。円状探索および偏角差分関数を用いた頂点の探索法を以下に示す。

(a) 円状探索追跡法 …… 偏角差分関数により頂点を決定するのに必要な境界画素のつながり方をデータを得るために円状探索法を用いる。図 4.4 のように基準点周りを探索して、点のつながり方を調べながら、未記録点を追跡していく方法を円状探索法と言う。

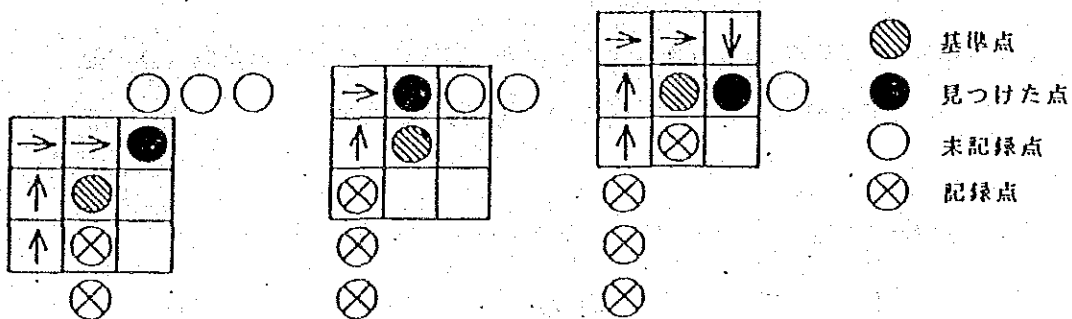


図 4. 4 円状探索

(b) 偏角差分関数による頂点決定法 …… 図 4.5 に示すように、ある画素と前後  $n$  番目の画素を結ぶ線のなす角度  $\theta$  を、記録されたすべての点に関して計算しその角度がある範囲内でかつ前後の点で計算した角度よりも  $90$  度に近いとき、その点を頂点とした。撮影倍率(カメラの位

置やレンズの焦点距離など)の違いにより抽出される頂点数が変化しないように $n$ と $\theta$ を設定する必要がある。

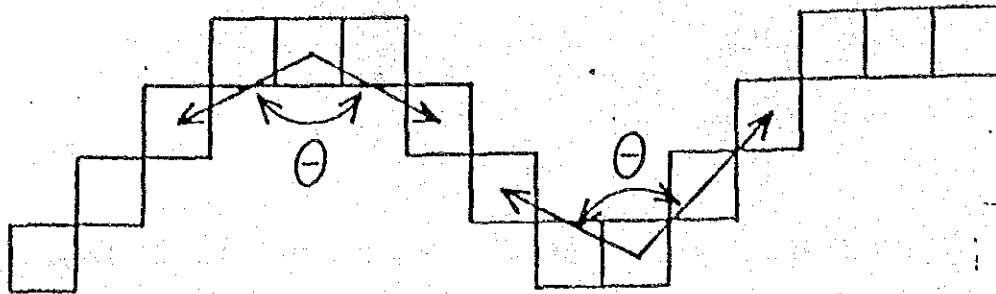


図 4. 5 偏角差分関数

- (5) 周囲形状 …… 周囲形状は、樹葉の周囲の凹凸の様子を表わす特徴パラメータである。境界画素の序数と重心から境界画素までの距離  $l_g$  をその最大値  $l_{g,max}$  で除した無次元距離  $l_g/l_{g,max}$  の関係を図 4.6 に示す。  
 $l_g/l_{g,max}$  の軸を  $n$  区間に分割した線をそれぞれ  $1, 2, 3, \dots, m, n$  とし、線  $l$  の下に入る画素数を全が素数で除した値を  $C_1, C_2, C_3, \dots, C_m, C_n$  とすると  $n$  個の特徴量を定義できる。

本システムでは、 $l_g/l_{g,max}$  を 10 等分し、10 種類の特徴量を定義した。それぞれ、COUNT1 ~ COUNT10 で表されている。

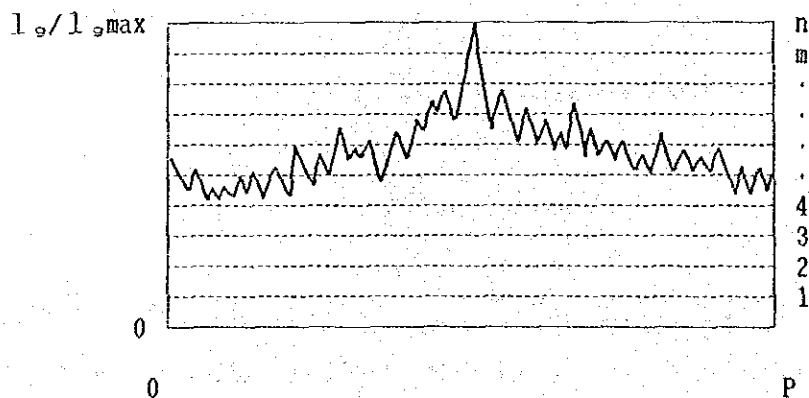


図 4. 6 重心から境界画素までの距離の関係

#### 4. 2. 3 特徴パラメータの抽出手順

本システムでは、図 3.2 に示した自動検索システムを用いて特徴パラメータの抽出を行なった。

樹葉の実画像から特徴パラメータ抽出し、その特徴パラメータをデータベースへ保存する手順を図 4.7 に示す。図中、画像記憶装置は、システムの構築、データベース作成および装置の信頼性を検討する上で必要な対象となる樹葉の実画像データを保存するもので、本システムの開発途中で必要な装置である。したがって、完成された自動検索システムにおいては不要となる。また、図中の「データベースの構築」の部分は、データベースを作成する際に用いられ、検索作業では不要である。

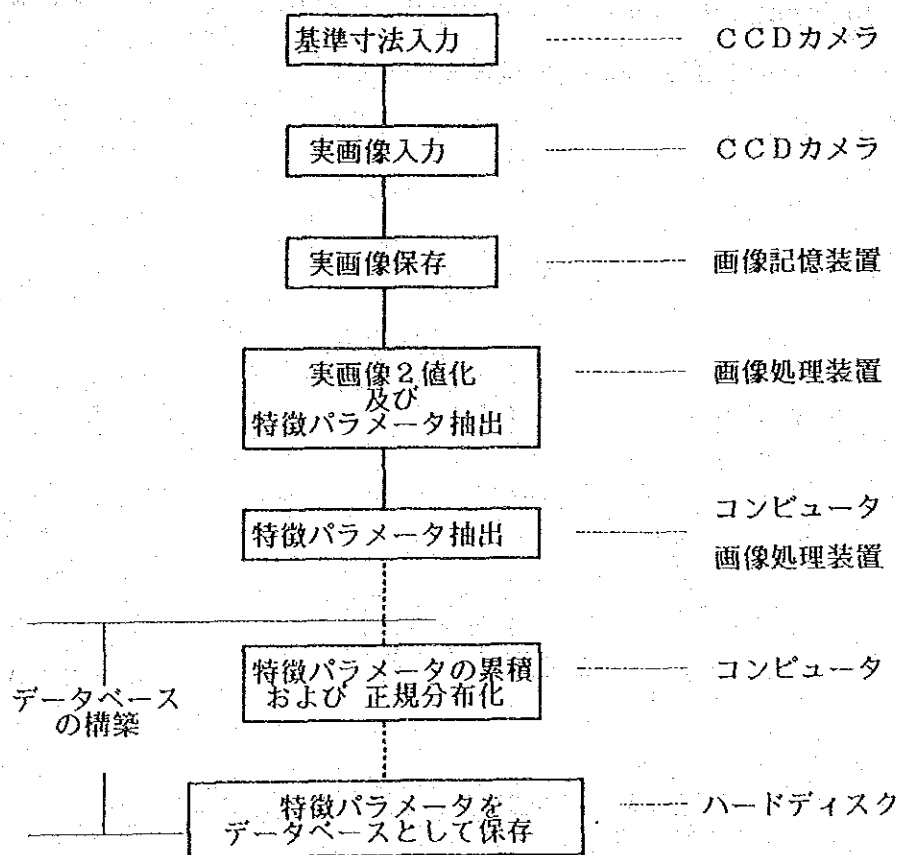


図 4. 7 特徴パラメータ抽出の手順

### 4.3 特徴パラメータの分布の表現方法

本システムはある任意の樹葉の特徴パラメータとコンピュータに保存されたデータベースを比較し樹種を特定するものである。したがってデータベースとして多くの樹種の複数の特徴パラメータを保存する必要がある。どの特徴パラメータも、樹種が同一であっても1枚1枚の葉によりその値が異なるため、各特徴パラメータの母集団は広い分布を示すことになる。したがって、データベースを構成する各特徴パラメータの分布は母集団を正確に表現できるものでなくてはならない。

#### 4.3.1 分布の形状およびその表現

- (1) 分布の形状 …… 前項で示した各特徴パラメータを約200の標本について求めた。その一例を図4.8に示す。いずれの特徴パラメータも平均値や頻度の値は異なるが形状はほぼ同様となる。

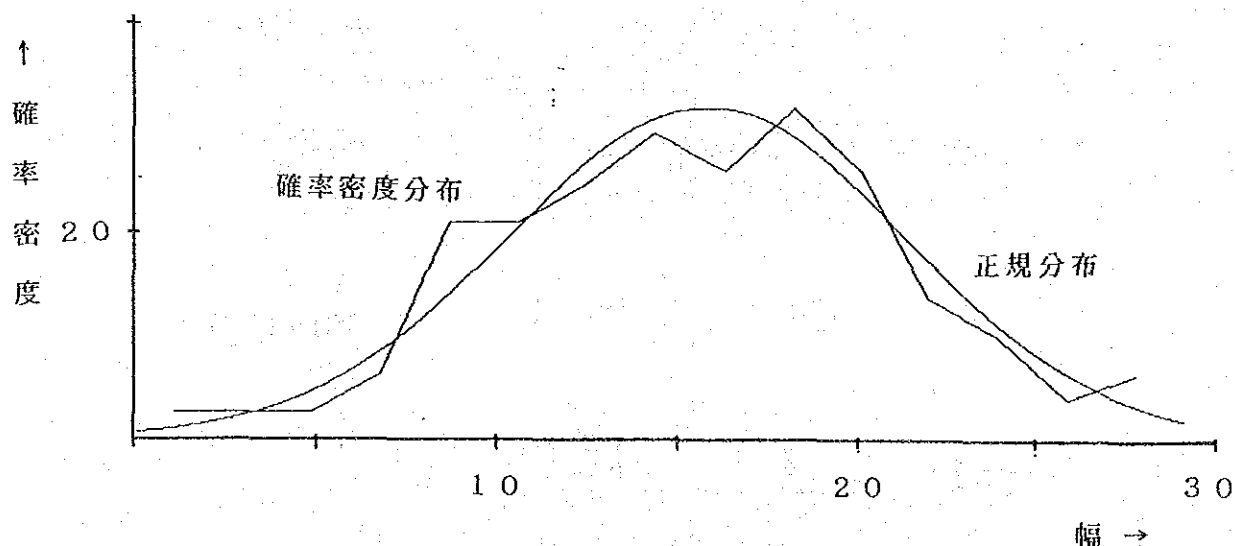


図4.8 分布の形 (樹種Gの幅の例)

- (2) 分布の表現 …… 前項で示した通り、各特徴パラメータは正規分布に非常に近い形状を示す。したがって、本システムでは各特徴パラメータの分布を正規分布に近似し母集団を表現し、データベースにはそれぞれの平均

値と標準偏差を記録した。データベースに記録される平均値 $\bar{X}$ と標準偏差 $\sigma$ は、式(4.6)および式(4.7)で示される。

$$\text{平均値 } \bar{X} = \left( \sum_{k=1}^N \frac{X_k}{N} \right) \dots\dots\dots (4.6)$$

$$\text{標準偏差 } \sigma = \left( \sum_{k=1}^N \frac{(X_k - \bar{m})^2}{N} \right)^{1/2} \dots\dots\dots (4.7)$$

本システムでは、全ての特徴パラメータの分布を正規分布で近似した。頻度分布曲線と近似した正規分布の一例を図 4.9 に示す。(その他の特徴パラメータについては、APPEBDIX 1 を参照)

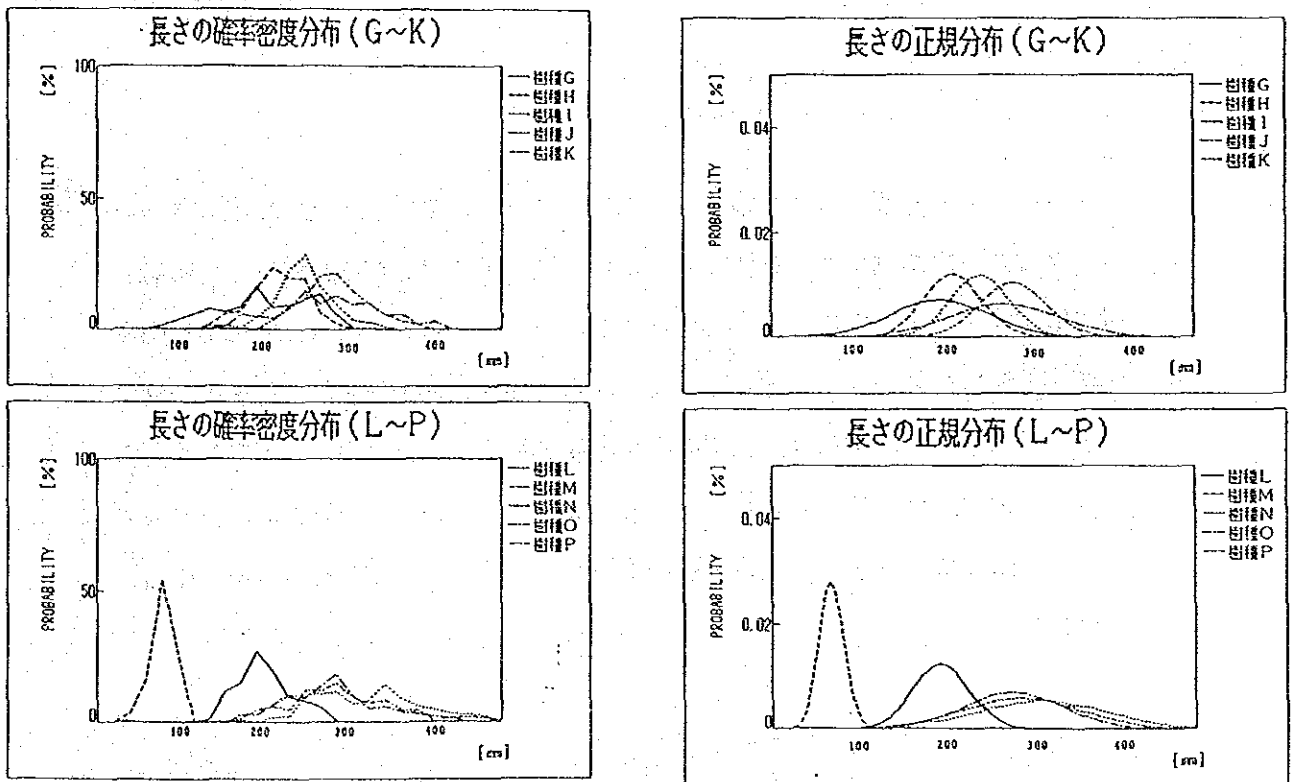


図 4.9 頻度分布曲線と近似した正規分布の一例

#### 4.3.2 データベースの内容

母集団の各特徴パラメータの分布が正規分布で近似できることから、データベースには平均値と標準偏差を記録すれば良い。表 4.2 にデータベースに記録される各特徴パラメータに関するデータの構造を示す。

ここで  $G_i$  ( $=G_1 \sim G_m$ ) は樹種を、 $P_j$  ( $=P_1 \sim P_n$ ) は特徴パラメータを示す。

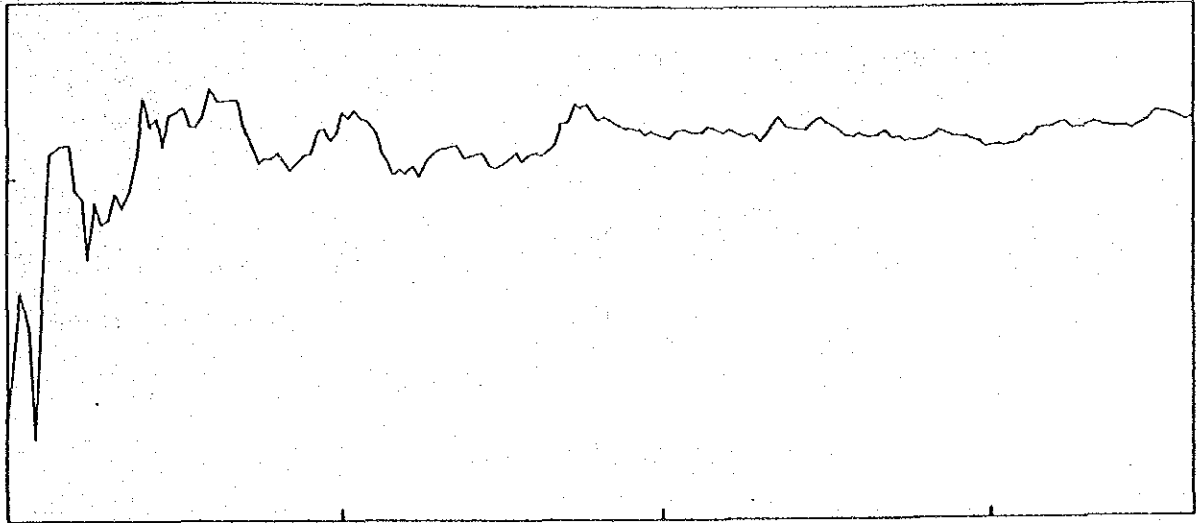
表 4.2 比較される分布のデータ構造

	$P_1$	$P_2$	...	$P_j$	...	$P_n$
$G_1$	$\bar{X}_{11}, \sigma_{11}$	$\bar{X}_{12}, \sigma_{12}$	...	...	...	...
$G_2$	...	...	...	...	...	...
	...	...	...	...	...	...
$G_i$	...	...	...	$\bar{X}_{ij}, \sigma_{ij}$	...	...
	...	...	...	...	...	...
$G_m$	...	...	...	...	...	$\bar{X}_{mn}, \sigma_{mn}$

#### 4.4 特徴パラメータの分布の評価

本システムの主要部分であるデータベースには、母集団を正確に表現できている平均値  $\bar{X}$  と標準偏差  $\sigma$  が必要である。母集団の平均値と標準偏差は、その母集団を構成する標本の数により変化することは明らかである。図 4.10 に母集団を構成する標本の数の変化に伴う平均値  $\bar{X}$  および標準偏差  $\sigma$  の変化の様子の一例を示す。図から明らかな通り、標本の数が増すと平均値と標準偏差は振動しながらある一定値に近づいていく。したがって、幾つの標本から平均値および標準偏差を求めれば母集団が十分にその樹種を表現していると言えるかを明らかにする必要がある。

## 平均値



## 標準偏差

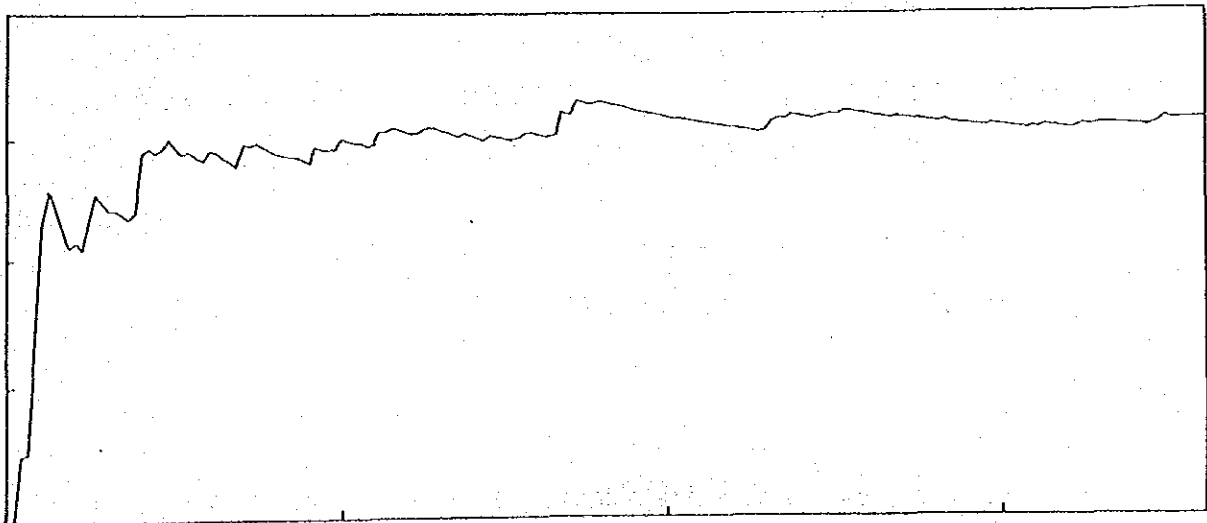


図4.10 標本の数と平均値および標準偏差の変化の関係

#### 4.4.1 評価関数

母集団がある樹種のある特徴を十分に表現できているかを評価する基準として、本システムでは、式(4.8)(4.9)で示される2つの評価関数を用いた。

##### (1) 固有区間当たりの平均変化率

$$A_i = \frac{\left| \sum_{n=i}^{i+9} X_n - \sum_{n=i+1}^{i+10} X_n \right|}{\sum_{n=i+1}^{i+10} X_n} \dots\dots\dots (4.8)$$

##### (2) 固有区間当たりの無次元化した標準偏差

$$B_i = \frac{\left( \sum_{n=i}^{i+9} \left( X_n - \frac{\sum_{k=i}^{i+9} X_k}{10} \right)^2 \right)^{1/2}}{\left( \sum_{n=i}^{i+9} \frac{X_n}{10} \right)} \dots\dots\dots (4.9)$$

#### 4.4.2 評価基準

前項で定義された式(4.8)(4.9)の2つの評価関数を用い、 $A_i$  および  $B_i$  の値が次の2つの条件を同時に満たした場合に収束したものと定義した。

- (a) 第1条件： $A_i$ が、連続して10回以上 1%以下になる
- (b) 第2条件： $B_i$ が、連続して10回以上 2%以下になる

#### 4.4.3 データベースの評価の流れ

実際にデータベースを構築する場合には、前述の評価関数を用い収束条件を満足するまで標本の数を増やしていく。これらの作業の流れを図 4.11 に示す。



《平均値》

《標準偏差》

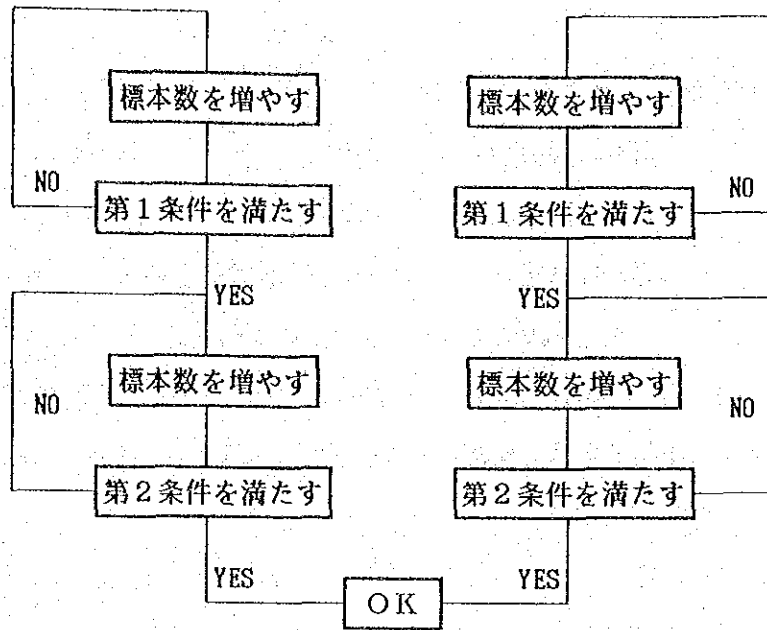


図4.11 特徴パラメータの分布の収束判断の流れ図

4.5 分布をもつ特徴からの検索方法

本システムにおける検索方法は、標本となる樹葉から得られる各特徴パラメータの値が、データベースに記録された各樹種とどの程度近いかを定量的に算出し、一番近いと算出された樹種をその標本の樹種であると判定するものである。ここで、データベースにある樹種との近似度の定量的計算には、マハラノビス距離を用いる。

4.5.1 マハラノビス距離

マハラノビス距離とは、標本から得られたある特徴パラメータの値がその特徴に対応する母集団の分布どの位置に相対的に相当するかを示す無次元の値である。すなわち、標本のある特徴パラメータの値のマハラノビス距離は、式(4.10)で与えられる。

$$dk = \frac{|X - \bar{X}|}{\sigma} \dots\dots\dots (4.10)$$

ここで、dk : マハラノビス距離

X : 標本の特徴パラメータの値

$\bar{X}$  : 母集団の特徴パラメータの分布の平均値

$\sigma$  : 母集団の特徴パラメータの分布の標準偏差

式(4.10)は、ある1種類の樹種(被検索物)の1つの特徴パラメータについてのマハラノビス距離である。実際のシステムでは、m種の樹種にそれぞれn個の特徴パラメータが存在する。

ここで、未知の標本q(樹種qの樹葉)から計測される特徴パラメータj(j=1~n)の値を  $X_{qj}$  とする。また、樹種i(i=1~m)の特徴パラメータjの母集団の分布を  $P_{ij}$  とすると、全てのマハラノビス距離すなわちマハラノビス距離の一般式は、

$$dk_{ij} = \frac{|X_{qj} - \bar{X}_{ij}|}{\sigma_{ij}} \dots\dots\dots (4.11)$$

ここで、 $dk_{ij}$  : 樹種iの特徴パラメータjから得られるマハラノビス距離

$X_{qj}$  : 未知の標本qから得られる特徴パラメータjの値

$\bar{X}_{ij}$  : 母集団の特徴パラメータの分布の平均値

$\sigma_{ij}$  : 母集団の特徴パラメータの分布の標準偏差

で表される。

#### 4.5.2 マハラノビス距離による判定

被検索物である未知の標本qから各特徴パラメータの値  $X_{q1} \sim X_{qn}$  を求め、データベースにある各樹種の  $X_1 \sim X_n$  に対応する特徴パラメータの分布  $P_{ij}$  を用

いてマハラノビス距離  $dk_{ij}$  を算出し、それぞれの樹種  $i$  毎にその総和  $T_i$  を算出する。ここで、 $T_i$  は、全マハラノビス距離と呼ばれ、式(4.12)で定義される。

$$T_i = \sum_{j=1}^m dk_{ij} = \sum_{j=1}^m \frac{|X_{ij} - \bar{X}_{ij}|}{\sigma_{ij}} \dots \dots \dots (4.12)$$

全マハラノビス距離  $T_i$  は、標本と同じ樹種の母集団から得られた特徴パラメータを用いて算出された値が最も小さくなる確率が高い。したがって、本システムでは、樹種  $G_i$  ( $i=1$  から  $n$ ) に対するそれぞれの  $T_i$  を求め、その値が最も小さい樹種を標本の真の樹種と判定するようにした。

理解を容易にするために、2種類の樹種を1つの特徴パラメータで判定する場合の概念図を図 4.12 に示す。

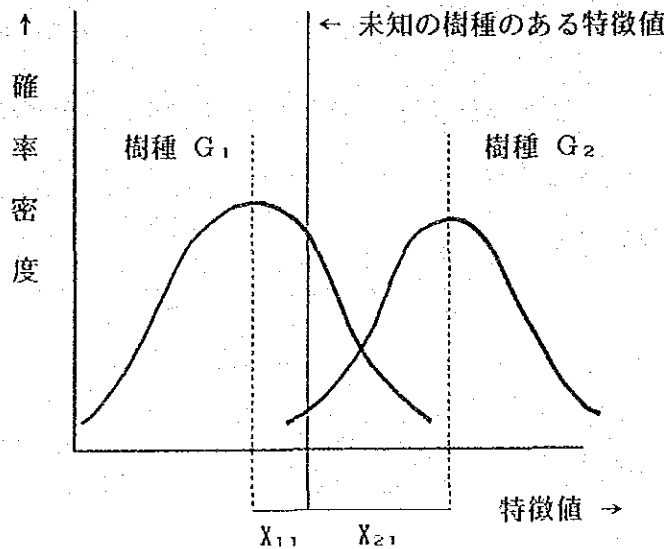


図 4. 1 2 2 種類の樹種を 1 つの特徴パラメータで判定する例

図 4.12 において、樹種  $G_1$  と樹種  $G_2$  の標準偏差  $\sigma_i$  が等しければ、特徴パラメータの値  $X_{ij}$  が  $X_{21} > X_{11}$  であるから、樹種  $G_2$  に比べ樹種  $G_1$  の全マハラノビス距離  $T_i$  が小さくなる。したがって、この未知の樹種は  $G_1$  であると判断される。樹種および特徴パラメータの数が増加しても同様な手法を用いるこ

とにより未知の樹種を特定することが可能である。

真の樹種の全マハラノビス距離  $T_i$  は、特徴パラメータの数が増加にしたがい他の樹種のそれと比べて小さくなる確率が增大する。したがって、特徴パラメータの数を増加するにしたがい正しく認識する確率も上昇する。

#### 4.5.3 認識率と誤認識率

システムの検索能力の評価は、次式で示される認識率及び誤認識率を用いた。

$$R_c = 1 - R_w = \frac{C}{E} \dots\dots\dots (4.13)$$

$R_c$  : 認識率

$R_w$  : 誤認識率

$C$  : 正しく認識した回数

$E$  : 検索を試行した回数

なお、誤認識率  $R_w$  の中で、 $E$  回試行を行い  $C$  回誤って樹種  $G_i$  と認識した場合の  $G_i$  に対する誤認識率を  $R_w(G_i)$  と表す。

## 第5章 性能確認試験の方法

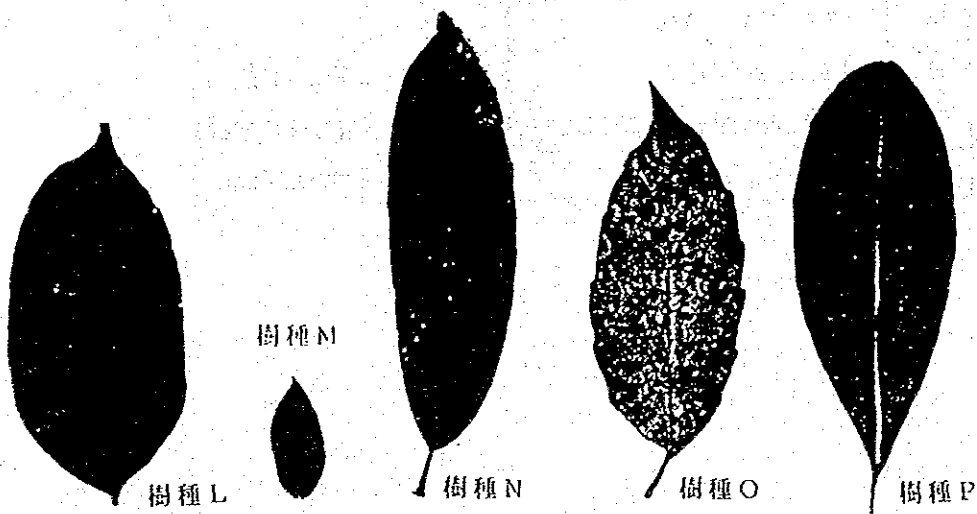
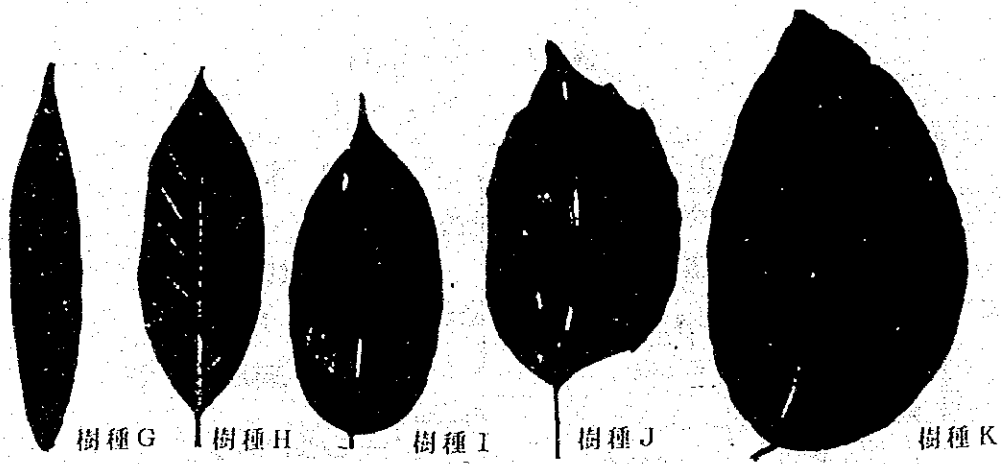
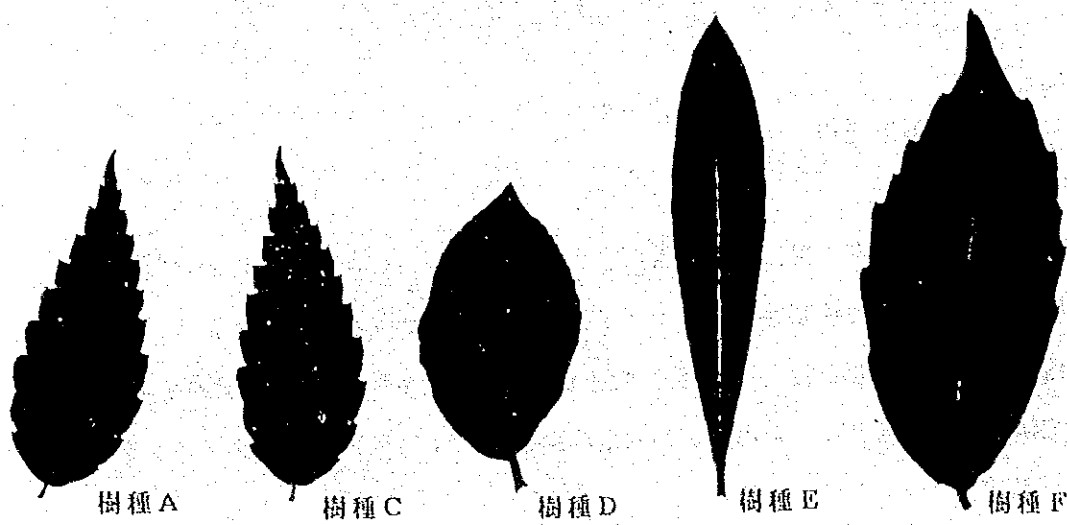
### 5.1 検索対象樹種

パプア・ニューギニアにある熱帯雨林の伐採現場から採取された約50種類の樹葉の中から10種類の樹葉を選び本システムの性能確認試験を行った。表 5.1 に採用した樹種を示す。また、図 5.1 にそれぞれの代表的と思われる樹葉を示す。

本報告書の範囲では、各樹種を表および図中に示すG～Pのアルファベットで示した。

表5.1 樹種A～Pの樹種名

呼名	樹種	科名
G	MANILTOA PLURIJUGA	RUBIACEAE
H	TIMONIUS TIMON	
I	PTEROCARPUS INDICUS	
J	CORDIA DICHOTOMA	BORAGINACEAE
K	INTSIA BIJUGA	LEGUMINOSAE
L	MANILTOA PSILOGYNE	
M	ALBIZIA PROCERA	
N	MANGIFERA MINOR	ANACARDIACEAE
O	PIMELODENDRON AMBOINICUM	EUPHORBIACEAE
P	ALSTONIA SCHOLARIS	APOCYNACEAE



0      100 mm

図5. 1 検索の対象とした樹葉

## 5.2 撮影倍率の影響

標本の実画像を撮影しそれから各特徴パラメータの値を算出する場合、次元をもつパラメータ（長さL，幅W，周囲長l，面積S・・・）や頂点数のようにしきい値を設定して求めるパラメータでは、撮影倍率が異なることにより得られる値が変化してしまう。したがって、撮影倍率の違いによる測定値の変化を自動的に修正する必要がある。

### 5.2.1 基準長さ

特徴パラメータのうち次元をもつ長さL，幅W，周囲長l，面積Sについては、CCDカメラの位置やレンズ系を変化した場合の実長と画素数の関係を補正するために、必要に応じ撮影前に基準長さを入力し、画素数と実長の関係を定めるようにした。

### 5.2.2 頂点数に関する補正

(1) 補正の手順 …… 頂点数は、報告者らにより開発された偏角差分関数により頂点を抽出法(4.2.2.(3)参照)を用いた。この方法は、画素の配列状況から角度を求め一定の角度以下を頂点と認識する様にしてある。したがって、1画素の長さが撮影倍率により異なると画面上で同一の角度でも実際は異なった角度、あるいは異なった大きさの突起となってしまう。

角度を求める際に幾つの画素間で角度を求めるかをしきい値と呼び、撮影倍率としきい値の関係を予め調べる必要がある。

本システムでは、図5.2に示す手順により、撮影倍率（例えば、CCDカメラの位置）を変化した場合に、実際の目測により測定した頂点数と本システムで計測した頂点数が一致するしきい値を求めた。ただし、角度は20度に固定した。

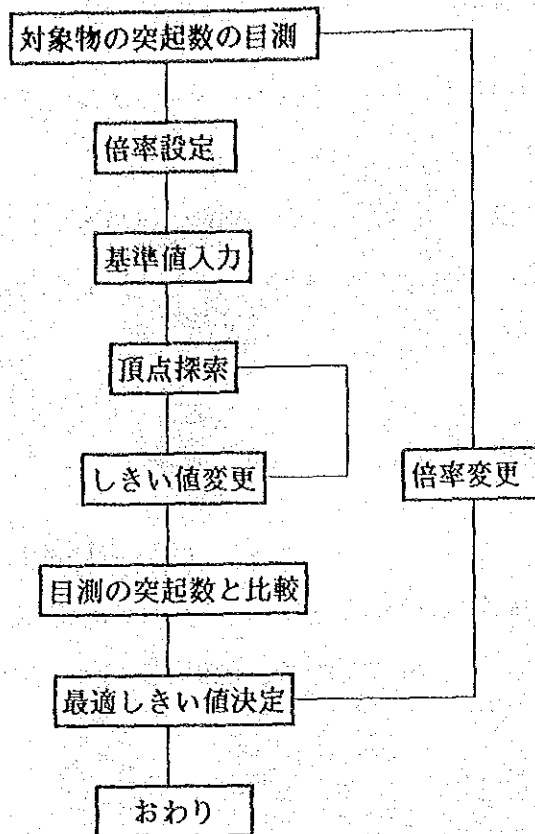


図 5. 2 撮影倍率の変化にともなうしきい値の決定手順

## (2) 頂点数計測の際の撮影倍率としきい値の関係

偏角差分関数により頂点を抽出する際のしきい値（角度、および画素間隔）を変化させ、実際の目測により測定した頂点数と一致する点のしきい値を求めた。すなわち、撮影倍率を変化することは、真の単位長さ当りの画素数（1画素が実寸法で何画素に相当するか）が変化することを意味する。したがって、目測により得られる頂点数と計測により得られる頂点数が一致するためのしきい値を撮影倍率を変化して求めた。但し、角度は20度に固定で行なった。

図 5.3 に検定結果である撮影倍率（単位長さ当りの画素数）としきい値の関係を示す。



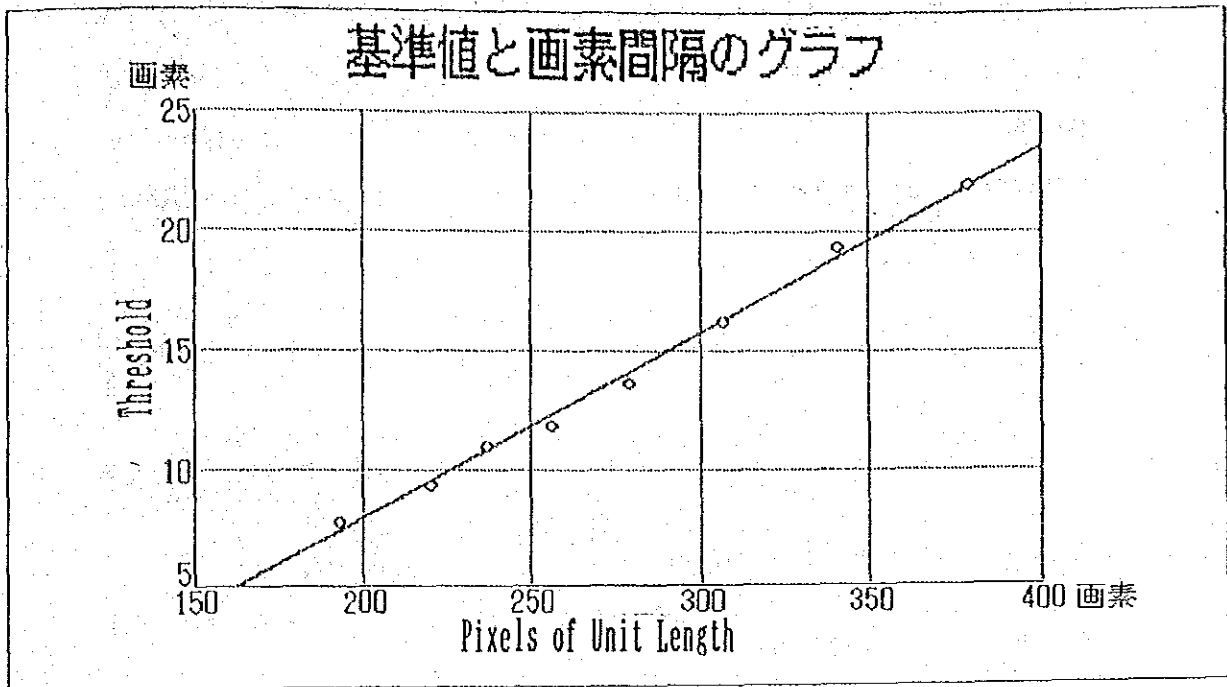


図 5. 3 基準値としきい値の関係

図からも明らかなように、撮影倍率としきい値は直線関係にある。したがって、本システムでは、カメラの位置やレンズ系を変化した際の撮影倍率を前項で示した基準寸法を用いて算出し、さらに図 5.3 の関係からしきい値の画素数を決定して頂点数を算出するロジックを加えてある。

### 5. 3 シミュレーションによる性能の確認方法

本システムの性能を確認するために、バプア・ニューギニアで採取された 10 種類の樹葉を用いて、繰り返し検索テストを行ない認識率を求めた。

テストに使用した樹種は、表 5.1 で示した樹種 G ~ 樹種 P である。

### 5.3.1 被検索樹葉

本システムの性能を確認するためには、データベースを構築した樹葉群とは別に同程度の枚数の樹葉群を用いて実際に検索試験を行なうことが望ましい。しかし、ここではデータベースを構成する樹葉の中から樹葉を数枚分のデータを無作為に抽出し、これらを未知の樹葉として検索を行なった。この作業を認識率が一定値に収束するまで繰り返し行なった。本試験でこの方法を採用した理由は以下のとおりである。

- (1) 対象とした樹葉は海外で採取されたものであり、十分な数を確保できない。
- (2) データベースの各特徴パラメータの分布を作成する際に、十分な枚数の樹葉を用いており、数枚抽出しても分布は変化しない。

### 5.3.2 性能の確認方法

本システムの性能を確認するために行なったシミュレーション・テストの方法は、データベースのとして記録されているそれぞれの樹種について、その中から1枚から10枚のデータをランダムに抽出し、これを被検索樹葉として検索を行なう。この作業を平均認識率が一定値に収束するまで繰り返し行なった。

具体的には、次の2点のテストを行なった。

- (1) 認識率の確認 …… データベースに記録されている全ての特徴パラメータを用いて、認識率を求めた。また同時に、判定を行なった際のマハラノビス距離を記録し、その分布図を求めた。
- (2) 特徴パラメータの有効性の確認 …… 各特徴パラメータそれぞれ単独で用いて検索を行ない認識率を求めた。

## 第6章 性能確認試験の結果

### 6.1 被検索樹葉の枚数の影響

#### 6.1.1 被検索樹葉が1枚の場合の認識率

標本として樹葉を1枚だけ用いて検索を行なった場合の認識率を図6.1に示す。

図から明らかなように、樹種により認識率は異なりその値は55~100%の間でばらつきがある。現状で記録されている樹種およびこれらの特徴パラメータでは、樹種Mが最も判別しやすく樹種Iが最もしにくい。これは、樹種Iの分布が他の何れかの樹種の分布と比較的にしており、誤認識するためである。すなわち、本システムの検索方法では、パラメータの分布の位置関係が関与しており、さらに認識率を向上させるためには、誤って認識しやすい樹種を予め調べ、これらを分離することが可能な特徴を中心に再検索するなどのチェック機能などを加える必要がある。

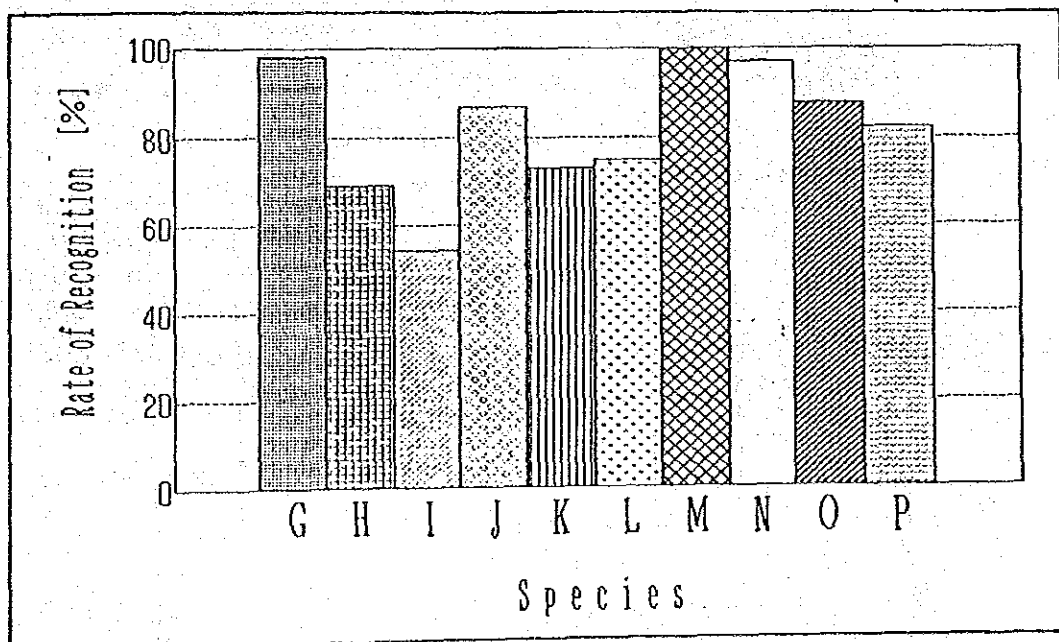


図6.1 標本樹葉1枚の場合の認識率(全特徴パラメータ)

### 6. 1. 2 被検索樹葉の枚数と認識率の関係

被検索樹葉の枚数を1～10枚に変化させた場合の認識率を図6.2に示す。

この際、被検索樹葉の樹種は未知であるがそれらは同一の樹種である。

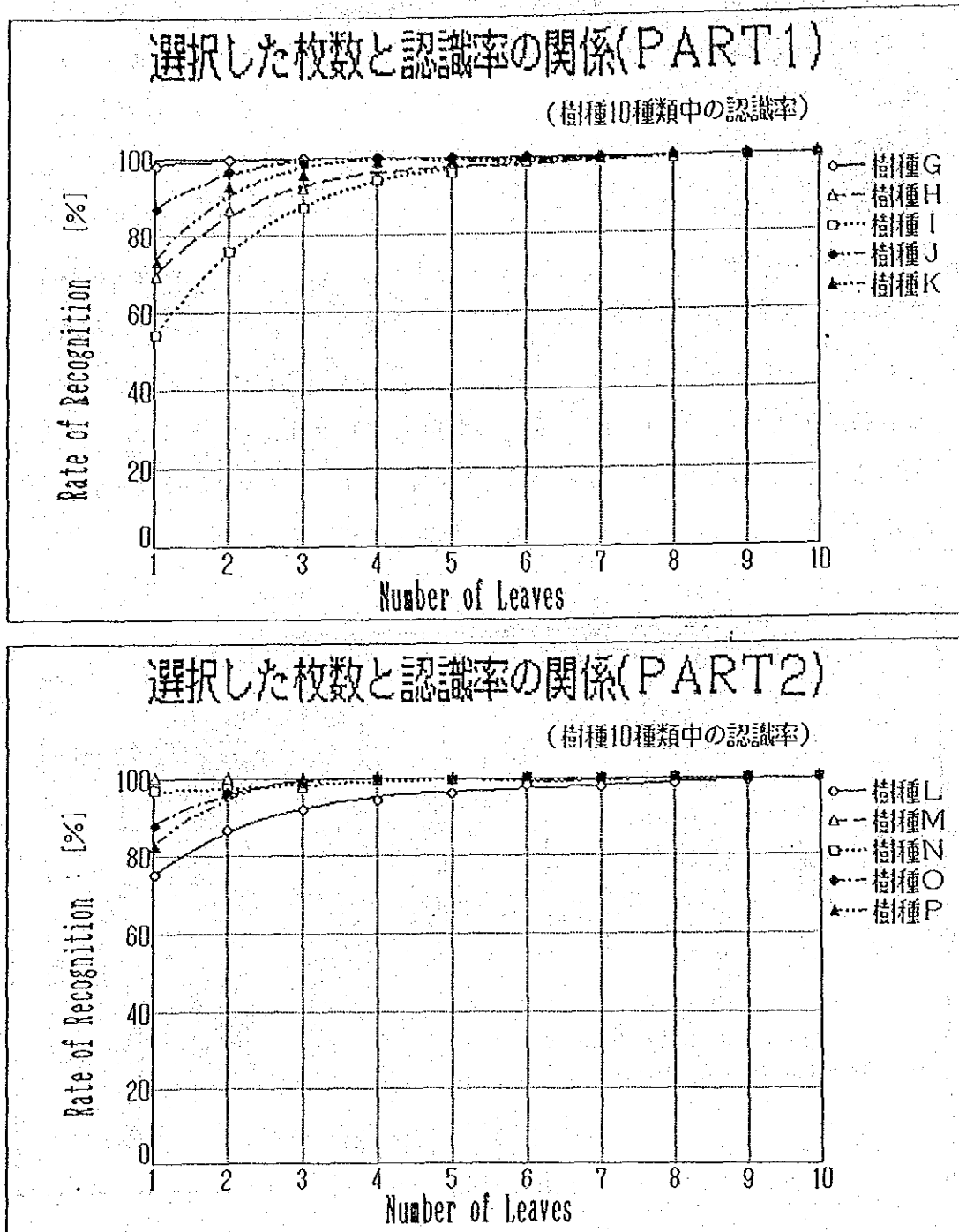


図6.2 標本樹葉枚数と認識率の関係(全特徴パラメータ)

いずれの樹種でも標本の枚数を増やすと認識率は増加する。その増加率は樹種により異なり、1枚で最も認識率の低い樹種Iでは9枚でほぼ100%に達するが、樹種Iに比べ1枚ではかなり認識率の高い樹種Lは、10枚の標本を使わないと100%に達しない。しかし何れの場合でも7枚程度以上でほぼ100%に近い認識率を得ることができる。

### 6.1.3 標本数と認識率の関係

前項で示した通り、標本数を増やすと認識率は増加する。この理由は、以下の通りである。

ここで、任意の樹種の任意の特徴パラメータを考える。母集団の特徴パラメータを $N(\mu, \sigma^2)$ の正規分布とすると、その確率密度関数は式(6.1)で表わされる。

$$f(X) = \frac{1}{(2\pi)^{0.5}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \dots\dots\dots(6.1)$$

ここで、母集団からn個の標本  $x_1, x_2, \dots, x_i, \dots, x_n$  をランダムに抽出し、それぞれのある特徴パラメータの値を  $X_1, X_2, \dots, X_i, \dots, X_n$  とすると、その特徴パラメータの平均値 $\bar{X}$ は式(6.2)となる。

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} \dots\dots\dots(6.2)$$

ここで、複数枚の樹葉を繰り返し選択し、ある特徴パラメータを抽出すると、選択された樹葉の特徴パラメータの平均値 $\bar{X}$ の分布は新たな確率変数となり、その確率密度関数は、式(6.3)で示される。

$$f(\bar{X}) = \frac{n^{0.5}}{(2\pi)^{0.5}\sigma} \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right) \dots\dots\dots(6.3)$$

すなわち、特徴パラメータの平均値 $\bar{X}$ の分布は、 $N(\mu, \sigma^2/n)$ で表現される正規分布となり、標本数nが増加するにしたがい広がりの小さな分布、すなわち

平均値に近い値となる確率が高くなる。したがって、これらの値から求められるマハラノビス距離は、小さくなり認識率が増加する。

図 6.3 に、標本数  $n$  を変化した場合の正規分布の形の変化を示す。また、図 6.4 に本システムの場合の一例として標本の枚数を変化した場合の頂点率の平均値の分布の違いを示す。

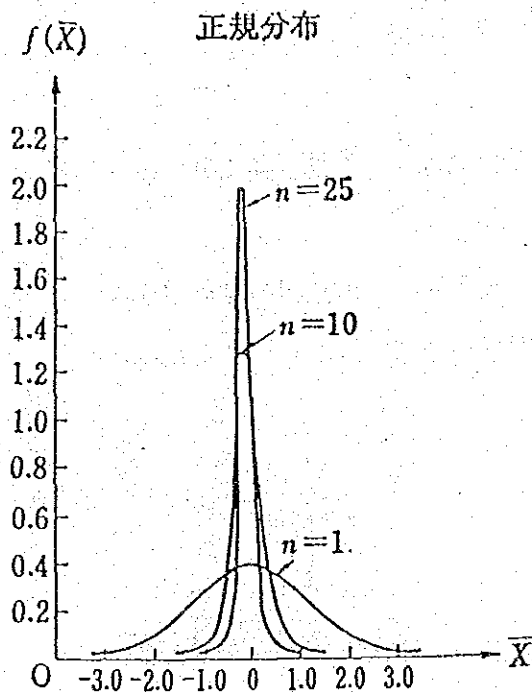


図 6.3 正規分布の先鋭化

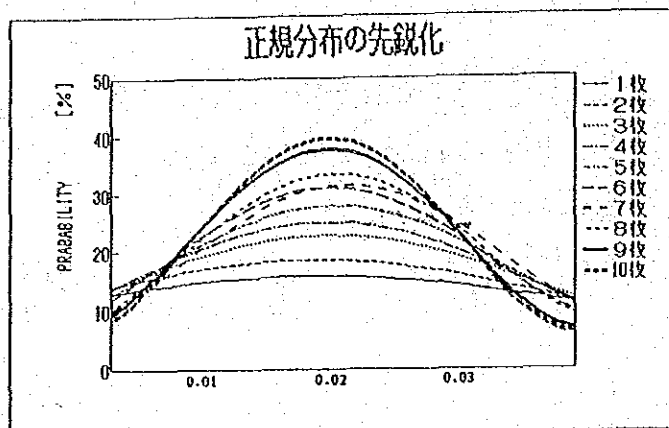


図 6.4 枚数による分布の変化の一例

#### 6.1.4 マハラノビス距離

図 6.5 および図 6.6 に樹種 L を検索した場合の全マハラノビス距離の分布の一例を示す(その他の樹種の内、樹種 I に関する全マハラノビス距離の分布の変化を APPENDIX 1 に示す)。各図中の 4 枚は、標本数を 1, 3, 5, 10 と変化した場合である。図 6.5 は、樹種 L (未知の樹葉であるが実際は樹種 L) を検索した場合に、樹種 L の母集団のデータを用いて算出されたマハラノビス距離の分布を、また図 6.6 は、樹種 I の母集団のデータを用いて算出されたマハラノビス距離の分布を示す。

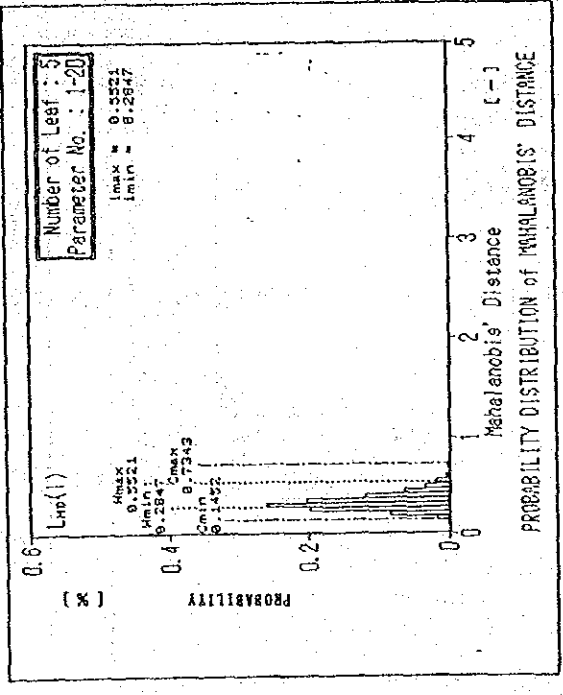
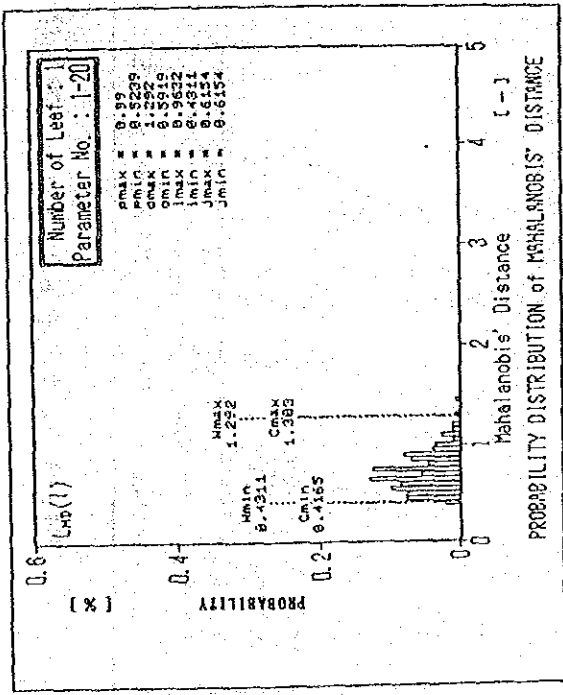
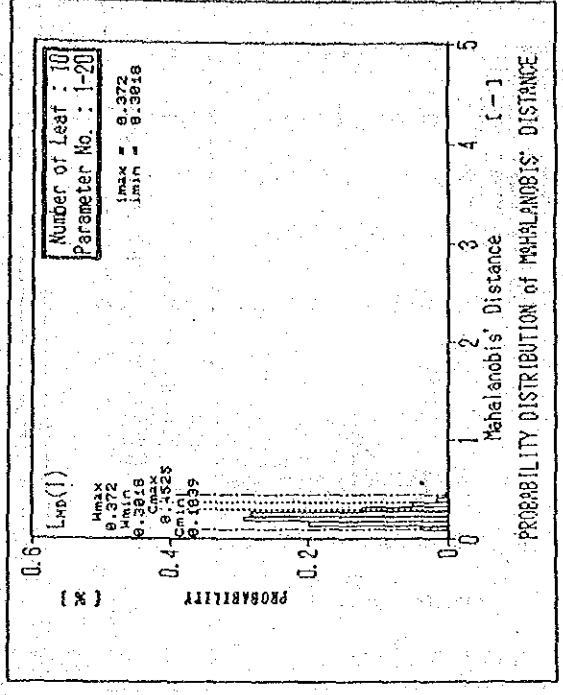
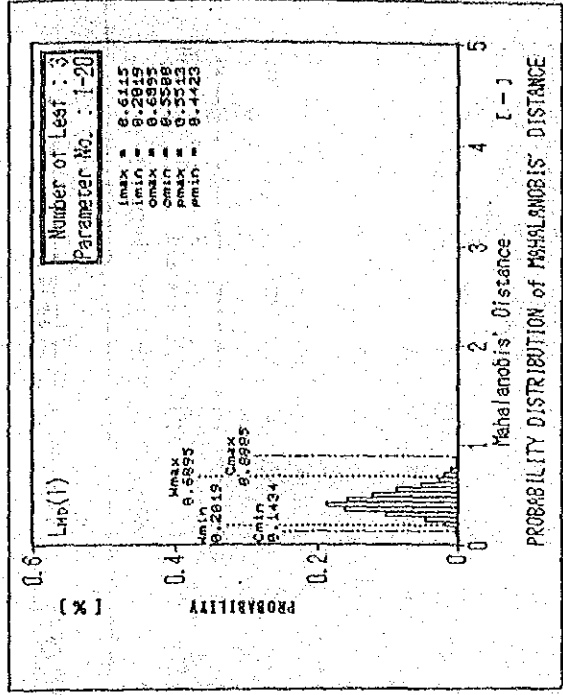


図6.5 標本樹葉の樹種Lの時樹種Lと認識したマハラノビス距離の分布

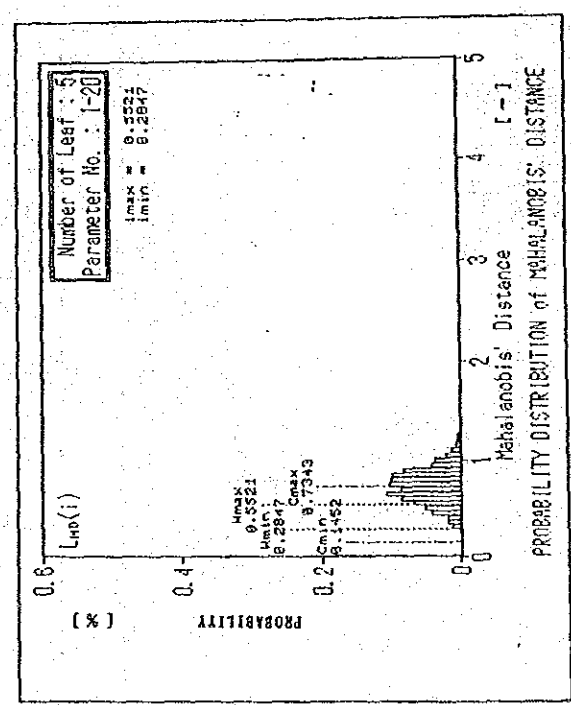
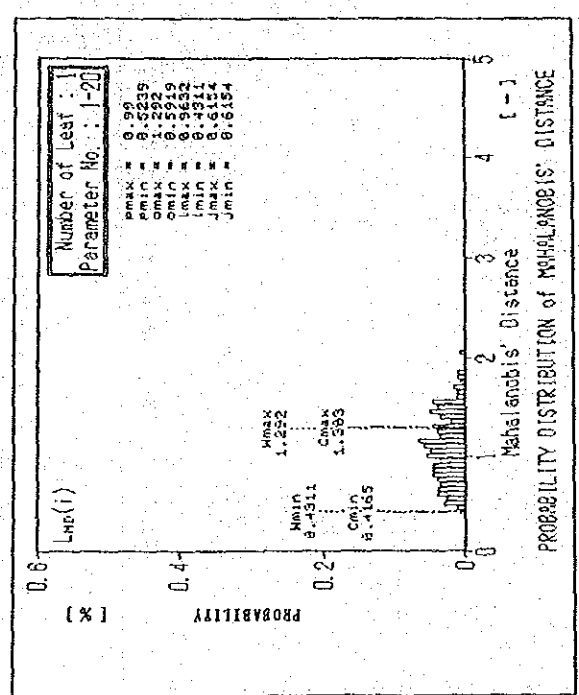
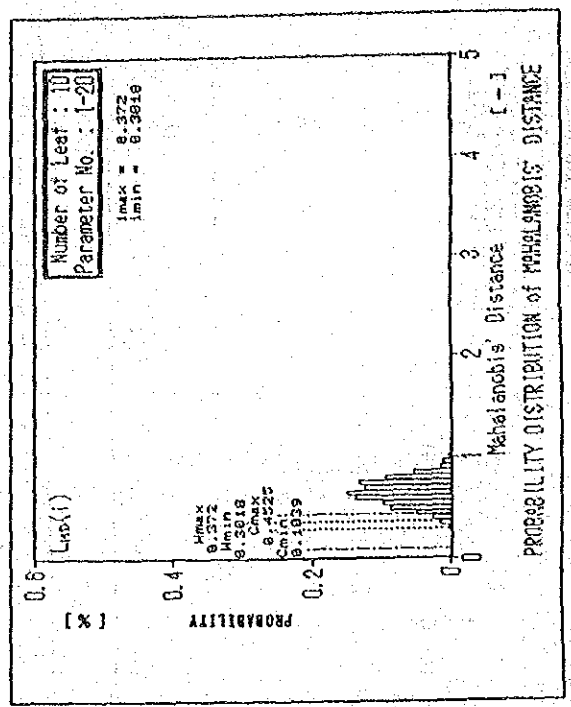
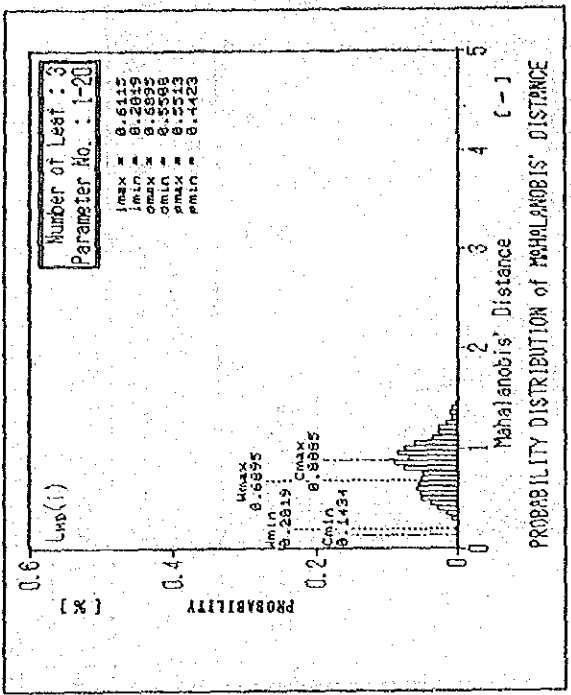


図 6. 6 標本樹葉の樹種 I の時樹種 I と誤認識したマハラノビス距離の分布



これらの図から明らかなように、真の樹種の母集団のデータから算出されたマハラノビス距離の分布(図 6.5)は、標本数が増加するに従い0に近づきながら先鋭化して行くのに対し、他の母集団のデータから算出されたマハラノビス距離(図 6.6)は、0でないある値に近づきながら先鋭化して行く。標本数が少ない場合には、この2つの分布の一部が重複するために樹種Lを樹種Iと判定する場合がある。しかし、標本数が増加してこの2つの分布が完全に離れた場合には、LをIと判定することはなくなる。樹種I以外の母集団のデータを用いて算出されたマハラノビス距離も同様の挙動を示すため、標本数が増加するのにもとまない真の樹種のマハラノビス距離の分布が他の樹種のそれから次第に離れ、最終的には完全に分離して認識率が100%に到達する。

標本数の増加に伴う認識率の上昇の割合が樹種によって異なるのは、各樹種間の各特徴パラメータの位置関係が関与しているためである(APPENDIX 2参照)。各特徴パラメータは非常に重なり合っており、その重なり具合が樹種により異なっていることや特徴パラメータによりその認識能力が異なっているためである。

以上の結果から標本数を増やすことにより認識率を上げ信頼性を増やすことができる。

## 6.2 特徴パラメータの影響

特徴パラメータの有効性を評価するために、1種類の特徴パラメータを用いて認識率を求めた。

### 6.2.1 単一の特徴パラメータによる認識率

図 6.7 ~ 図 6.16 に、各樹種を単一の特徴パラメータで検索した場合の認識率を示す。殆どの樹種および特徴パラメータで、標本数の増加にもとまない認識率が増加しており、単一パラメータの場合でも標本数の増加は認識率に良好な影響を与えている。これらは、6.1.3項で述べた通り、標本から抽出される特徴パラメータの平均値が母集団の平均値に近づき、平均値の分布が先鋭化されるためである。

しかしながら、樹種Oを特徴パラメータ COUNT2 で検索した場合(図 6.15 参

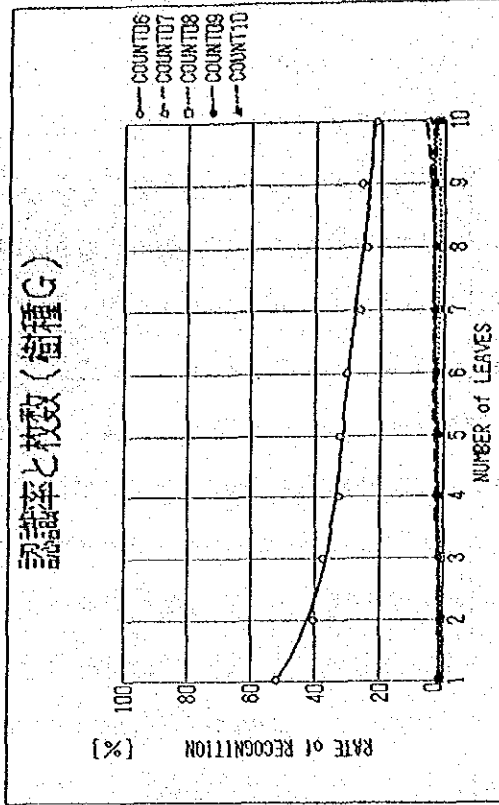
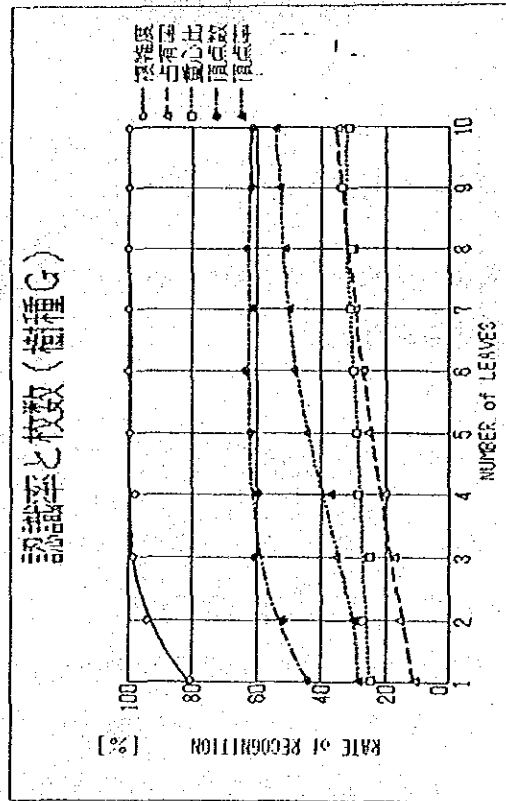
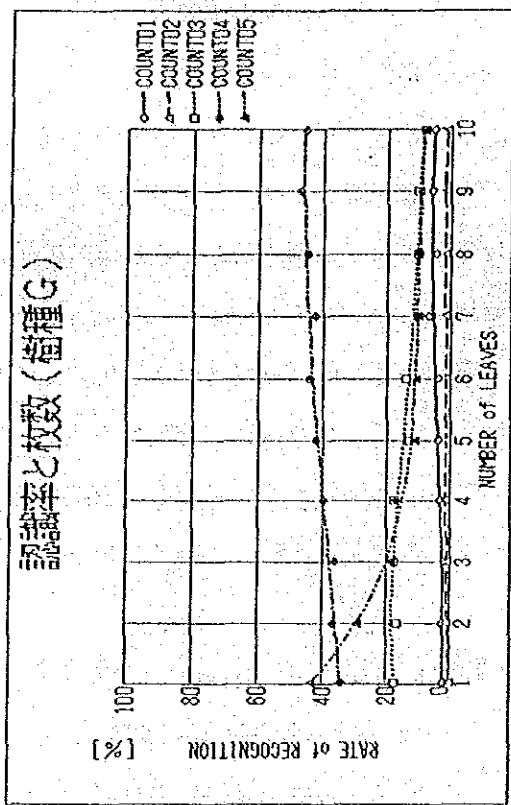
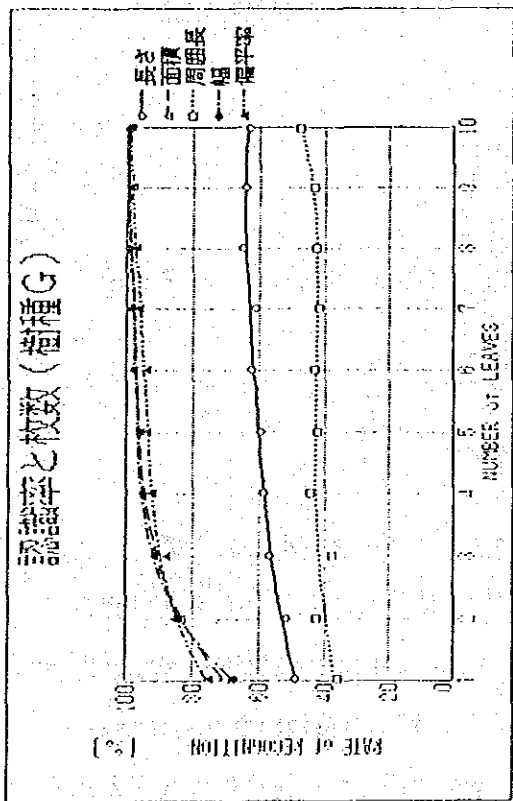


図6.7 各特徴パラメータの枝数と認識率の関係 [標本樹葉の樹種 - G]

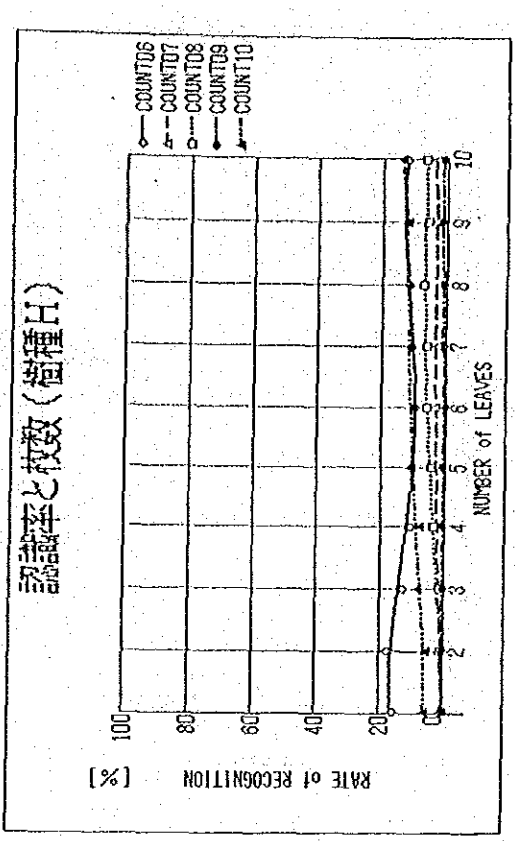
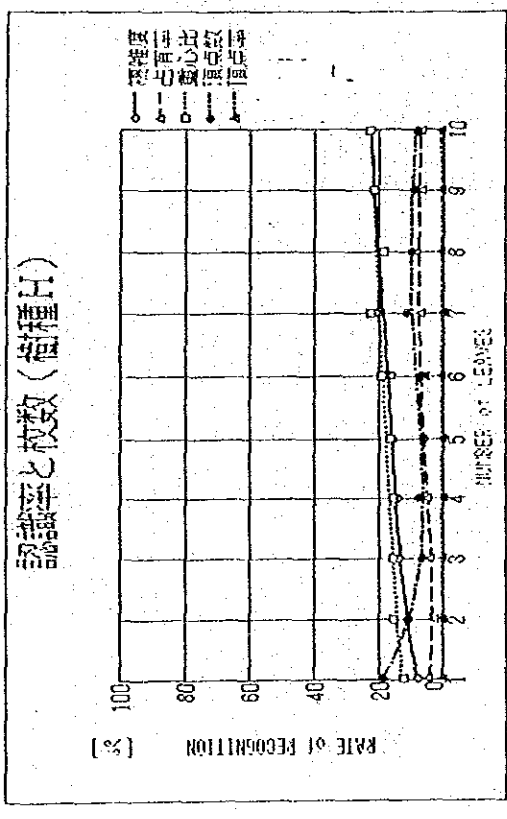
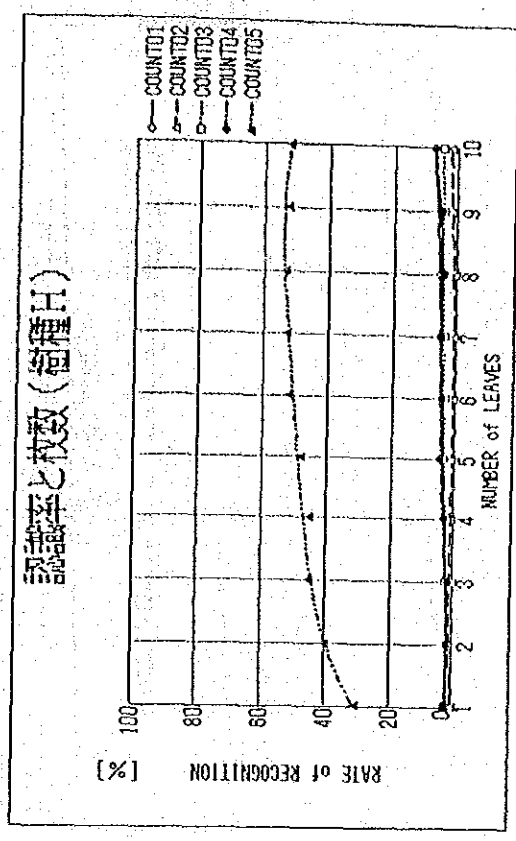
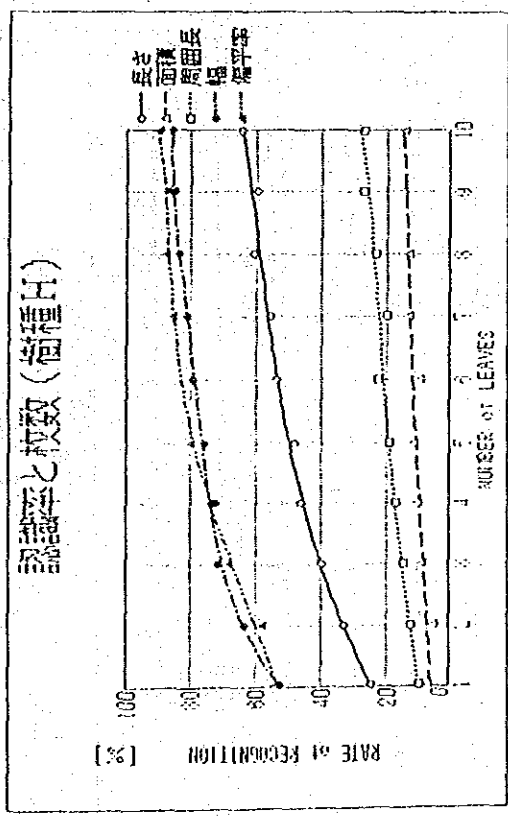


図 6. 8 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種 - H]

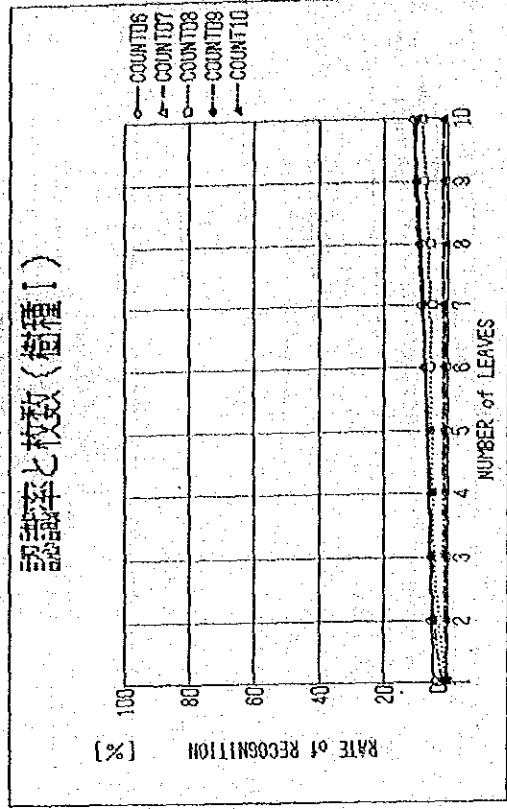
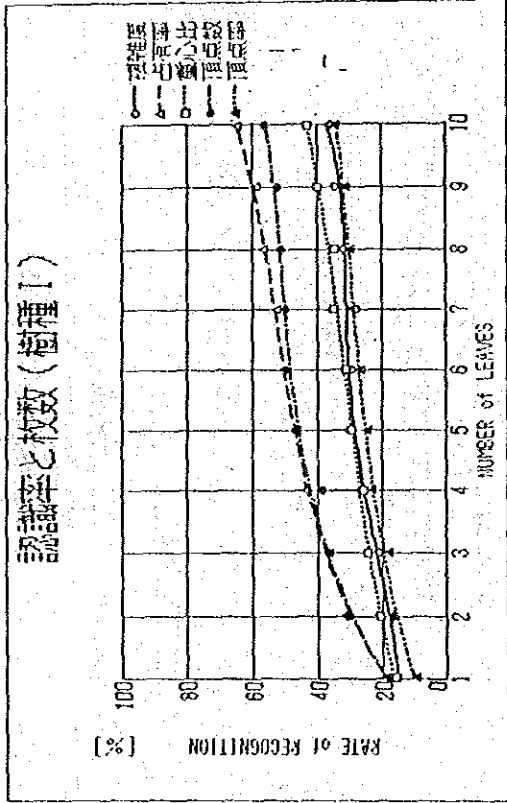
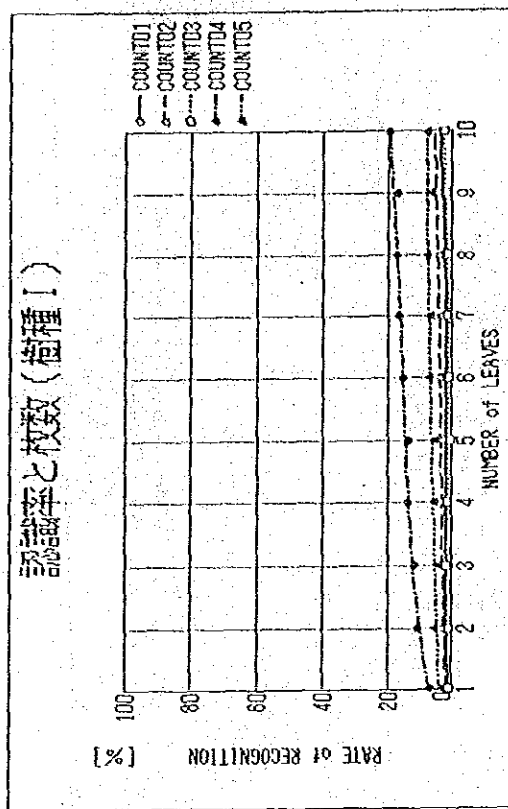
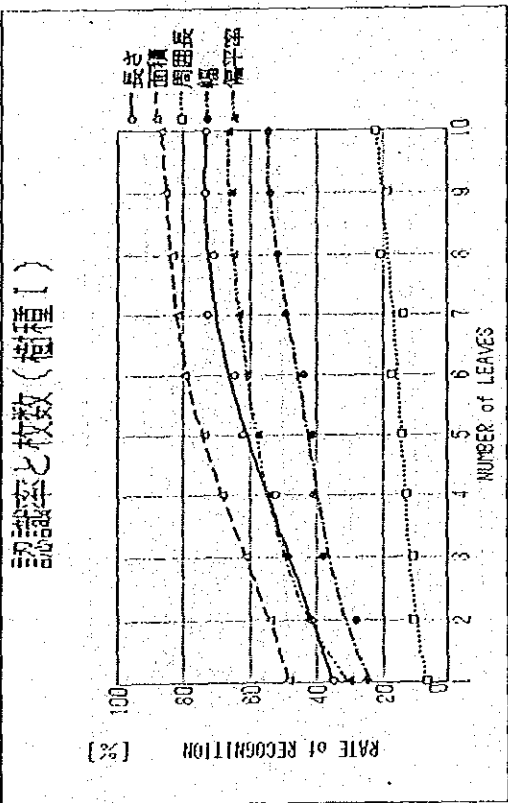


図 6.9 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種 - I]

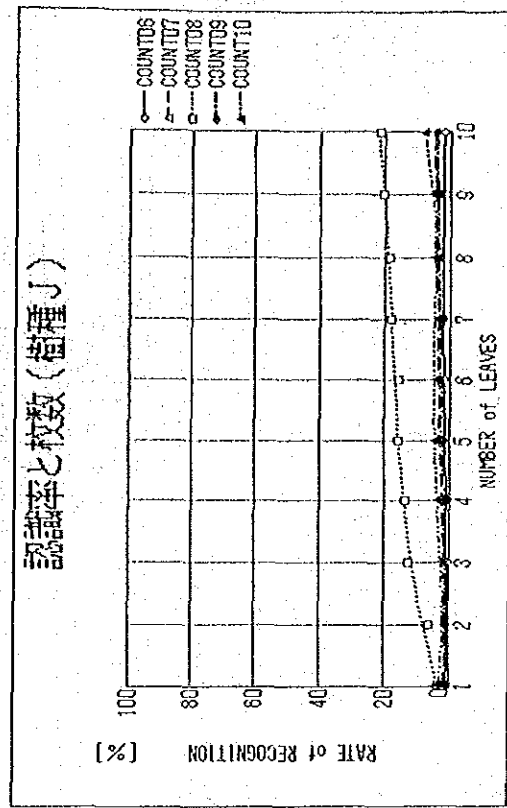
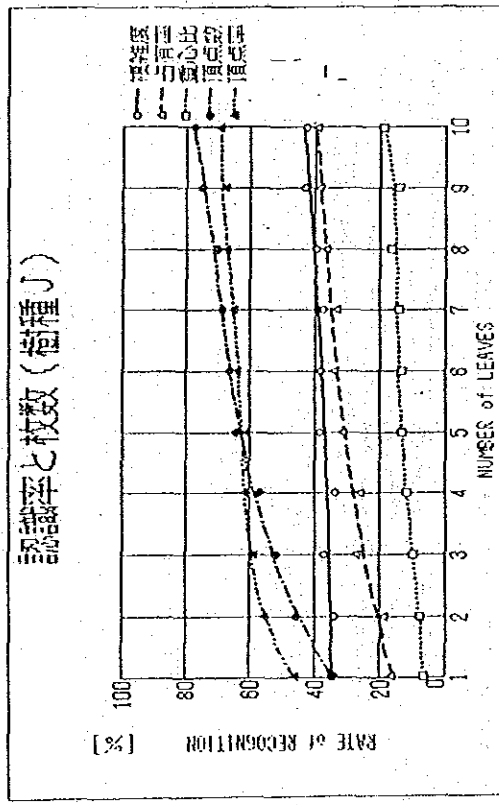
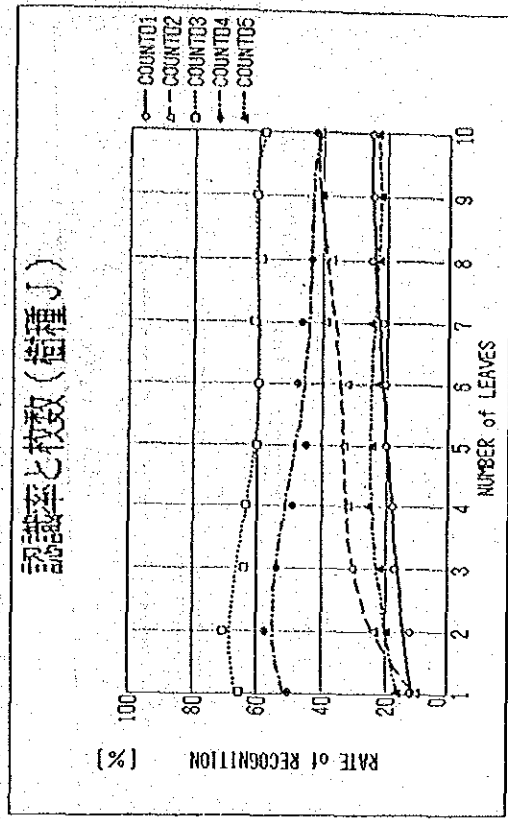
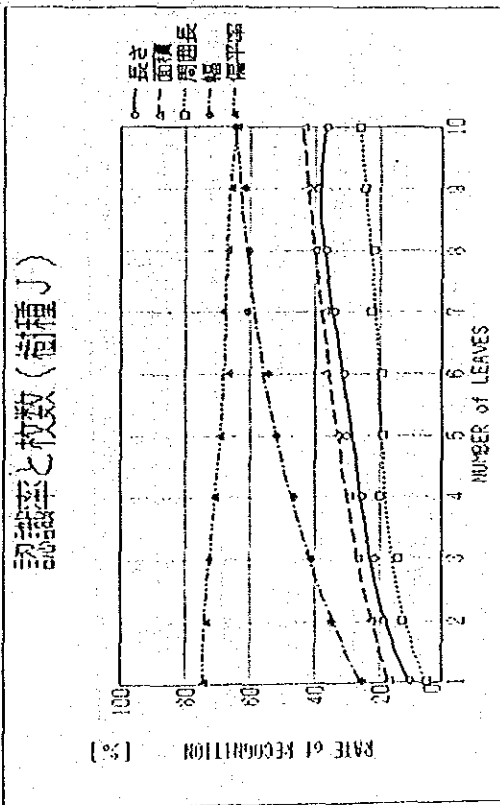


図 6.10 各特徴パラメータの枝数と認識率の関係 [標本樹葉の樹種 - J]

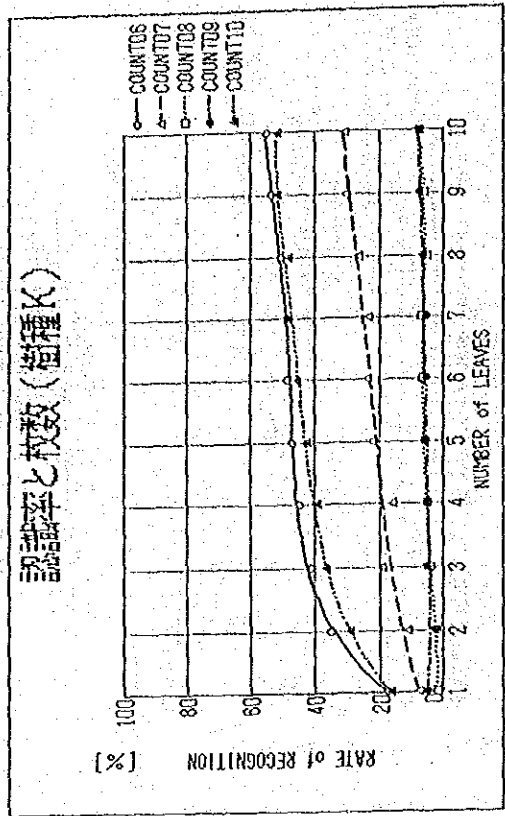
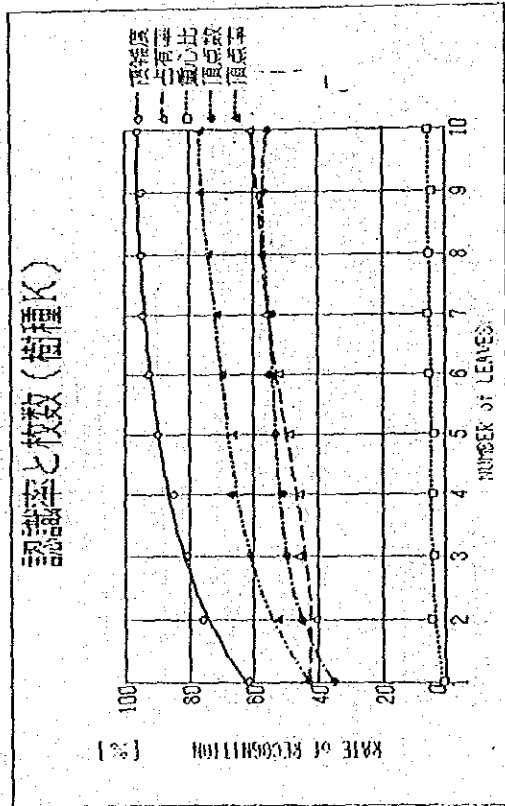
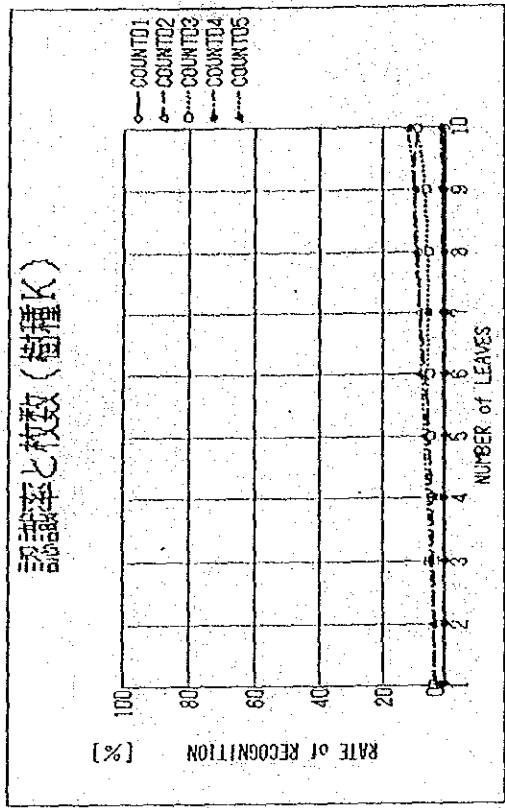
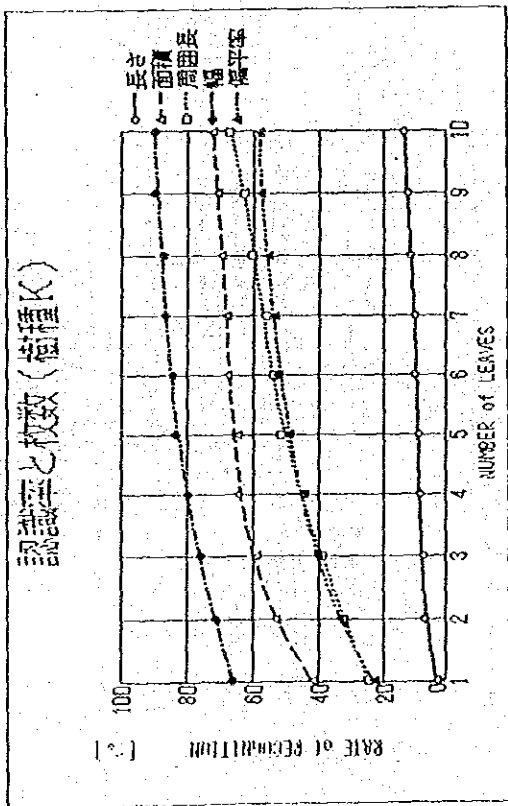


図6.11 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種-K]

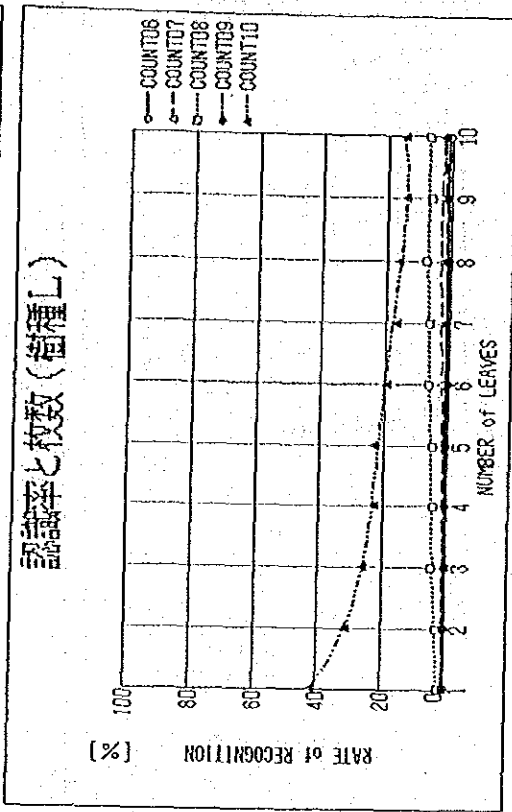
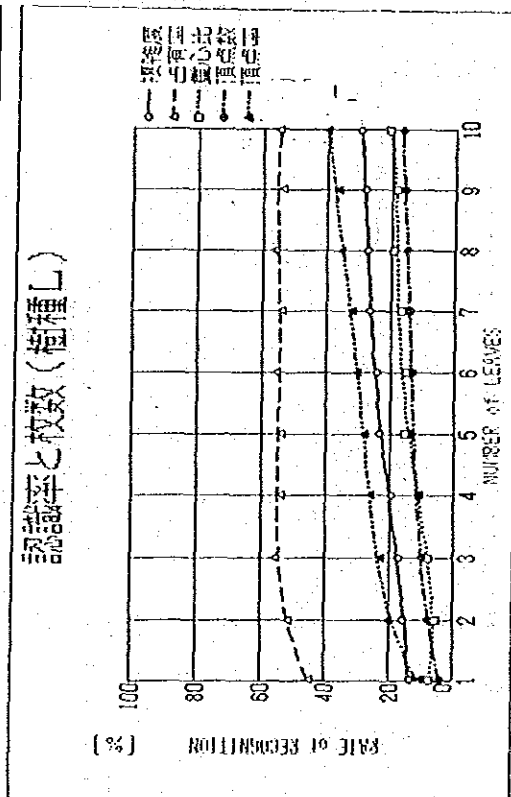
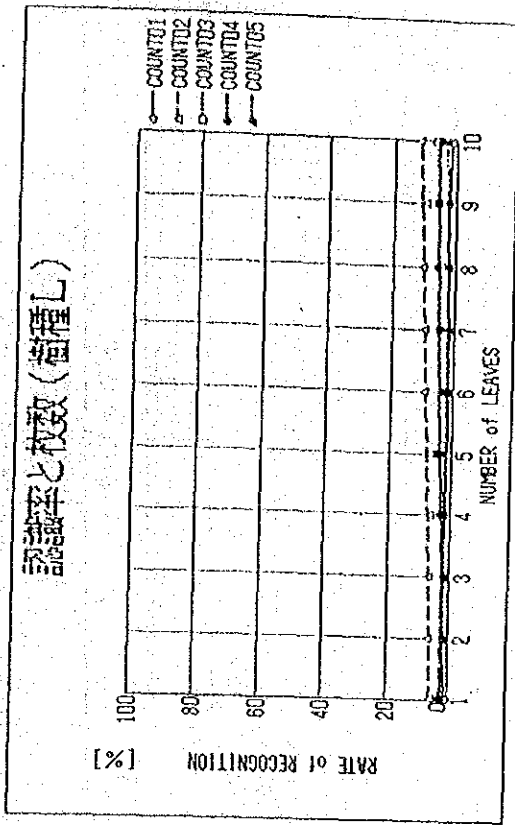
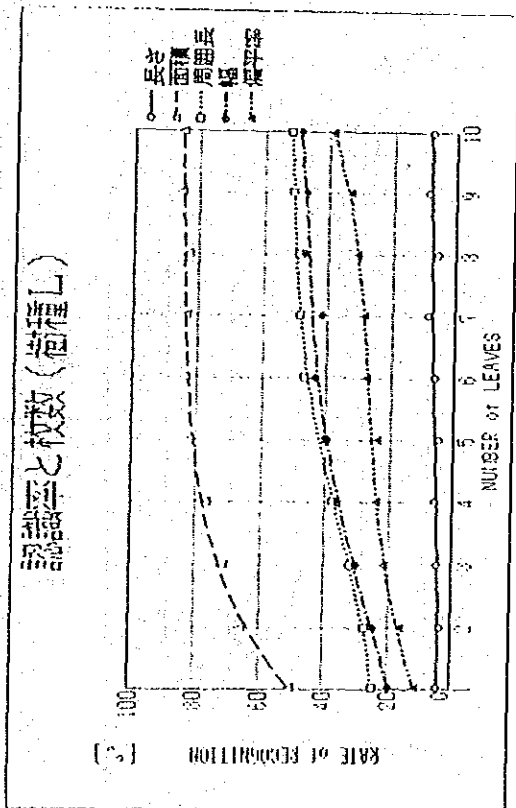


図6.12 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種-L]

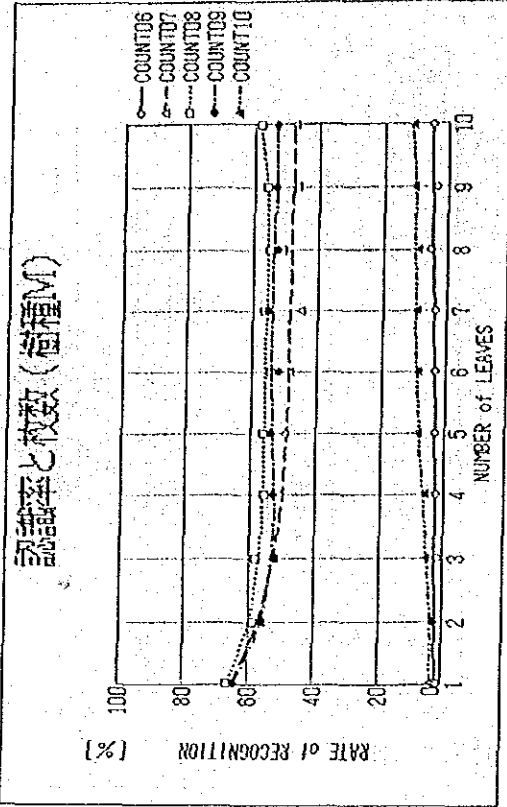
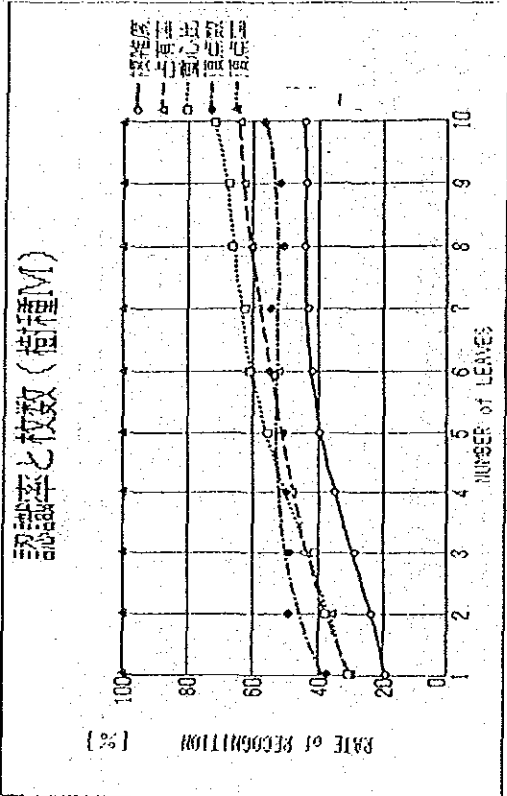
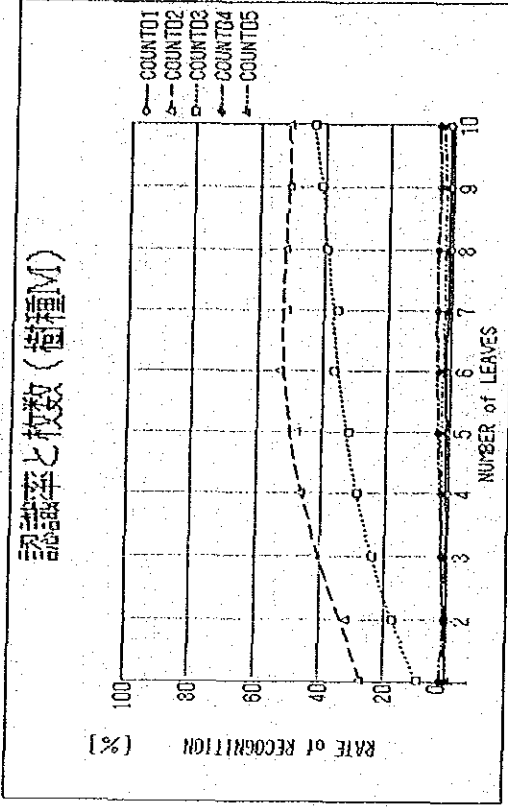
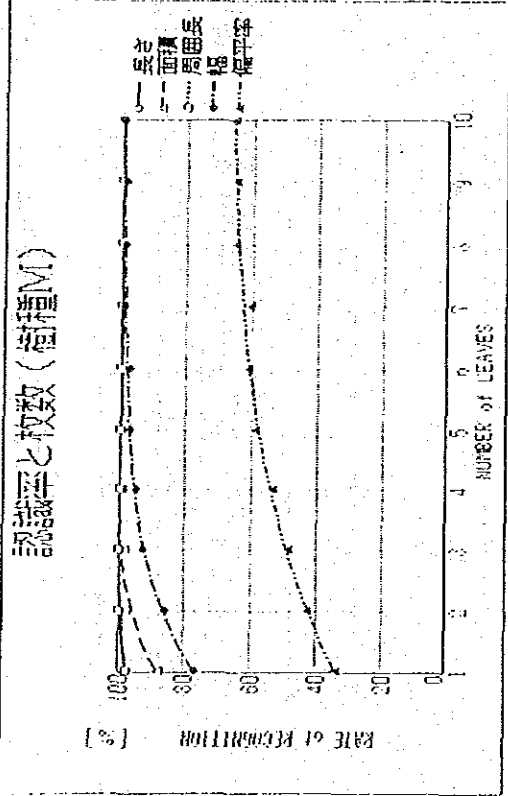


図6.13 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種 - M]



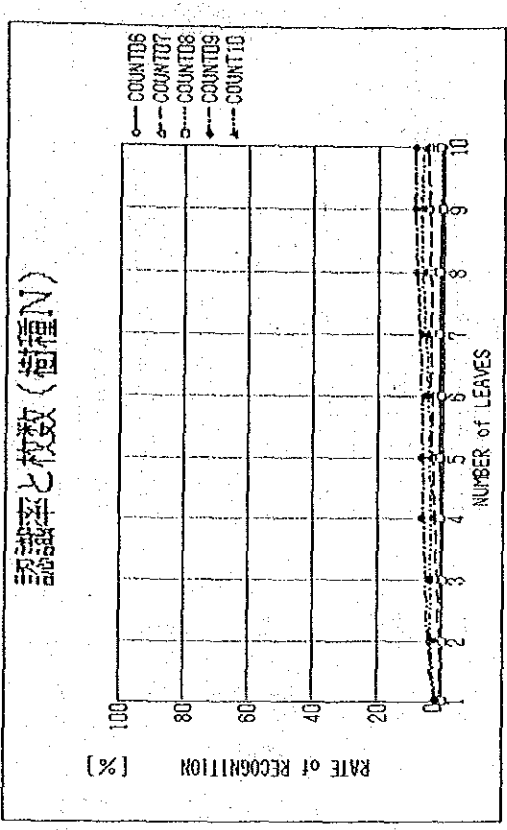
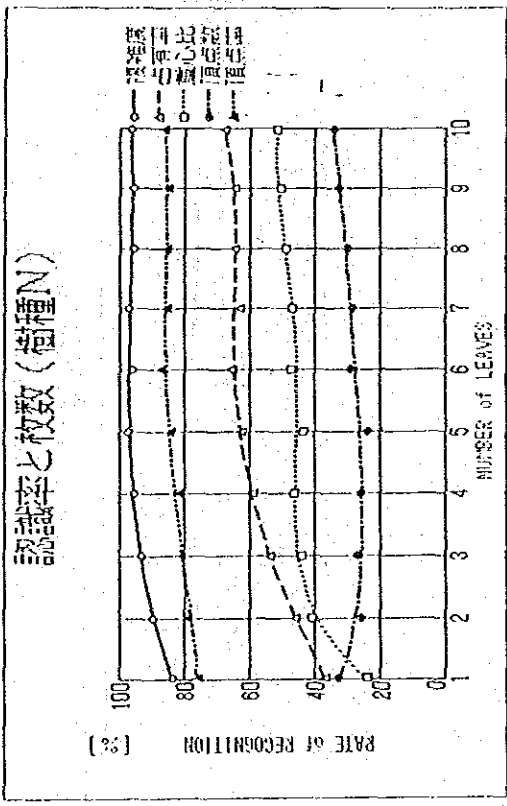
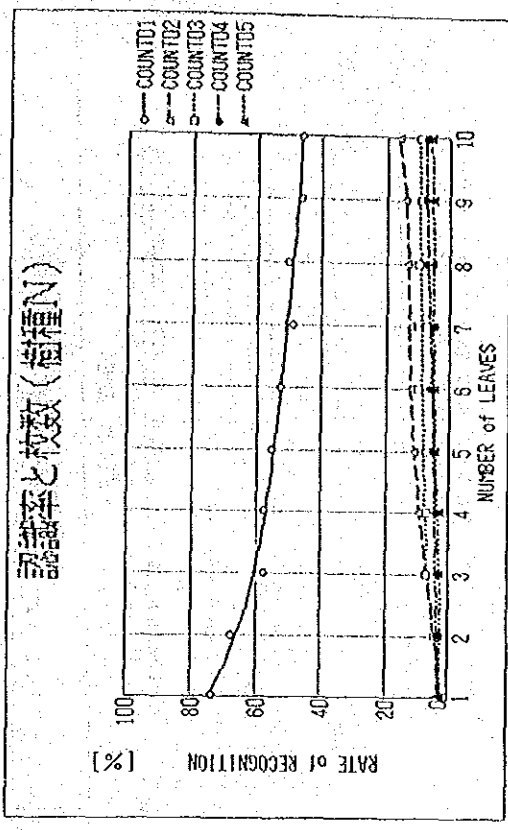
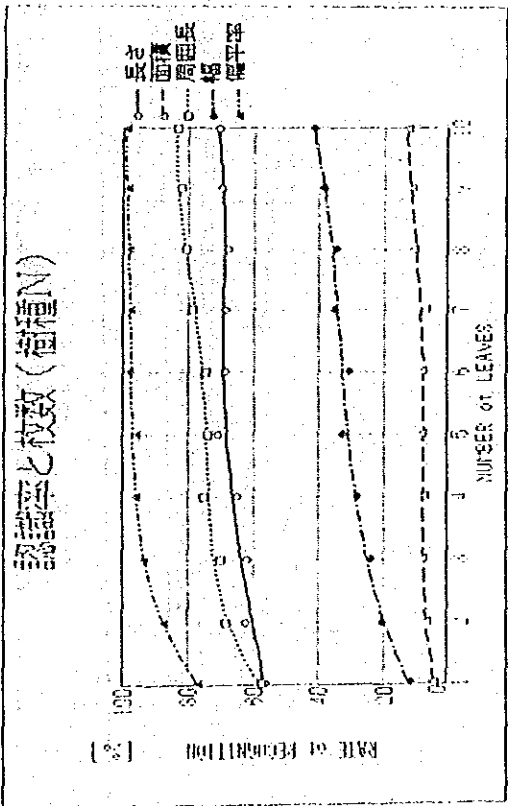


図 6. 1.4 各特徴パラメータの枝数と認識率の関係 [標本樹葉の樹種 - N]

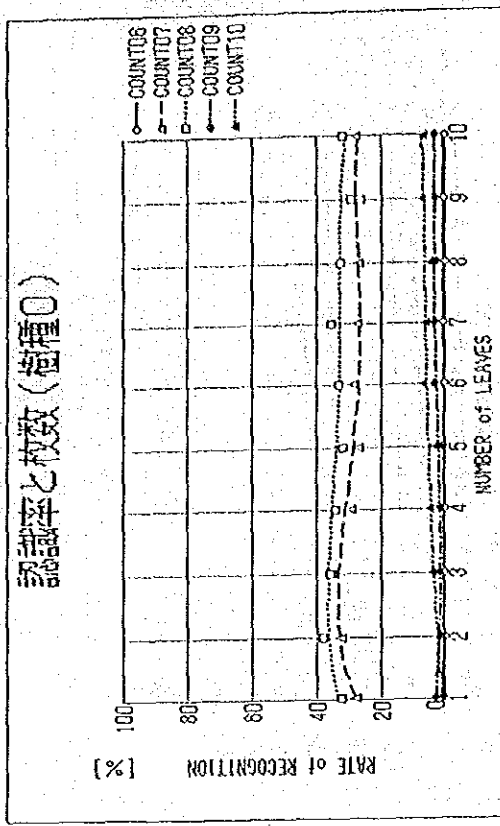
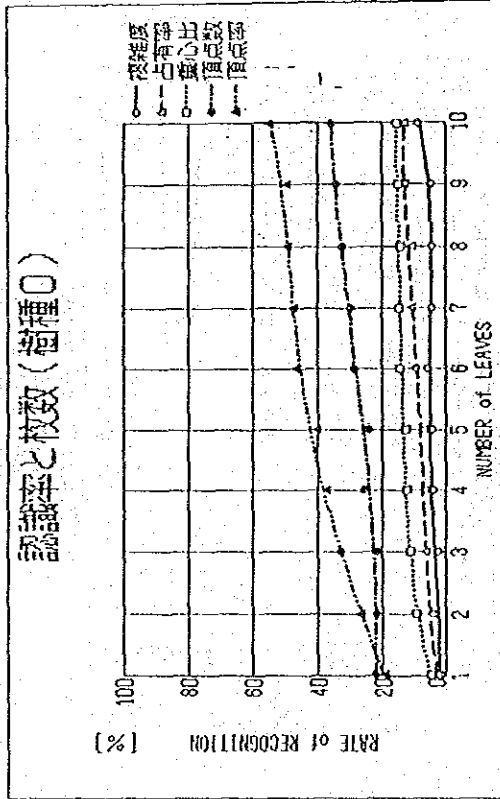
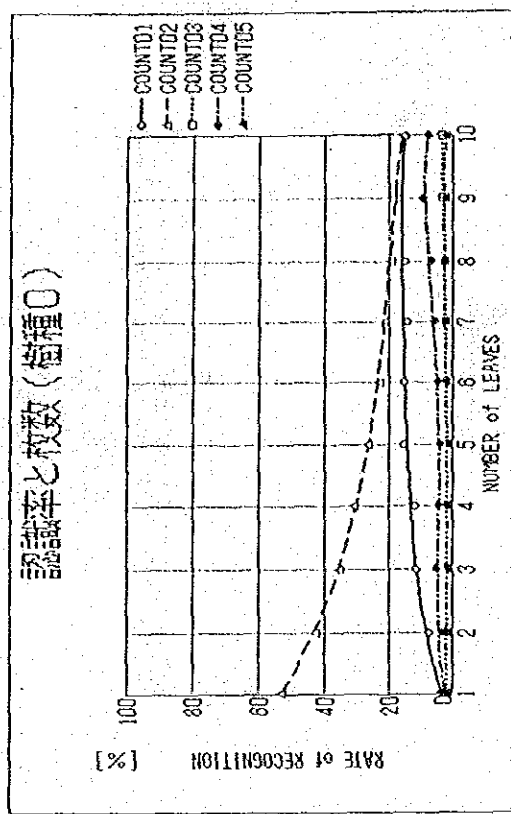
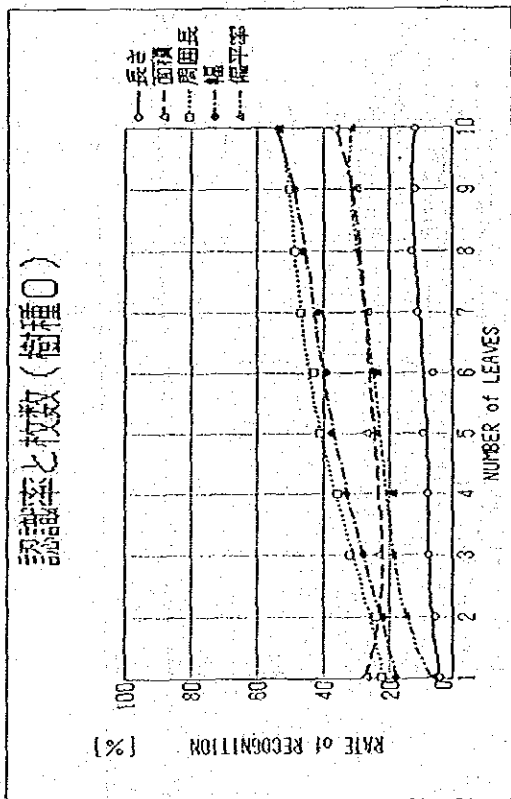


図6.15 各特徴パラメータの枝数と認識率の関係 [標本樹葉の樹種-0]

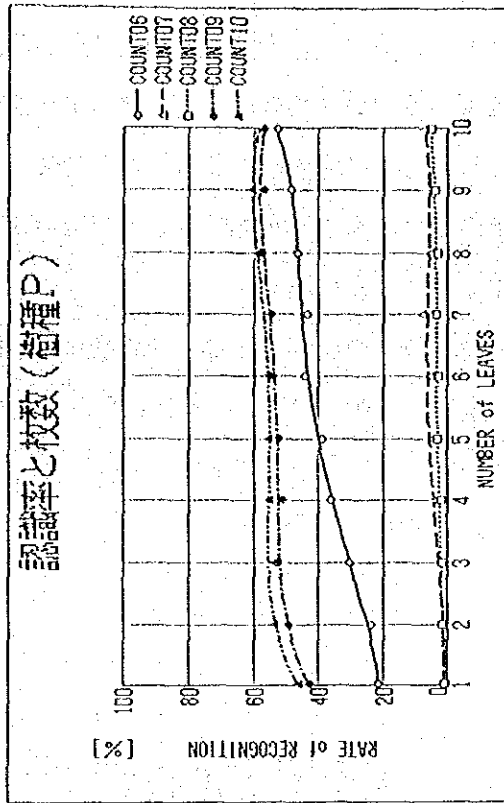
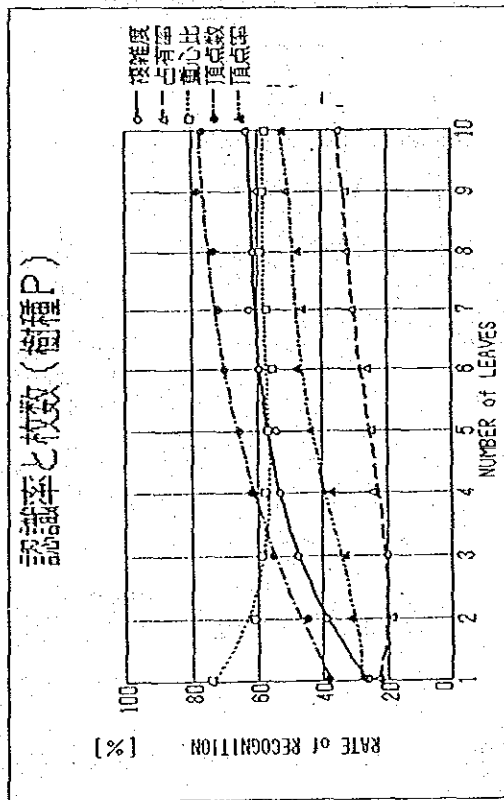
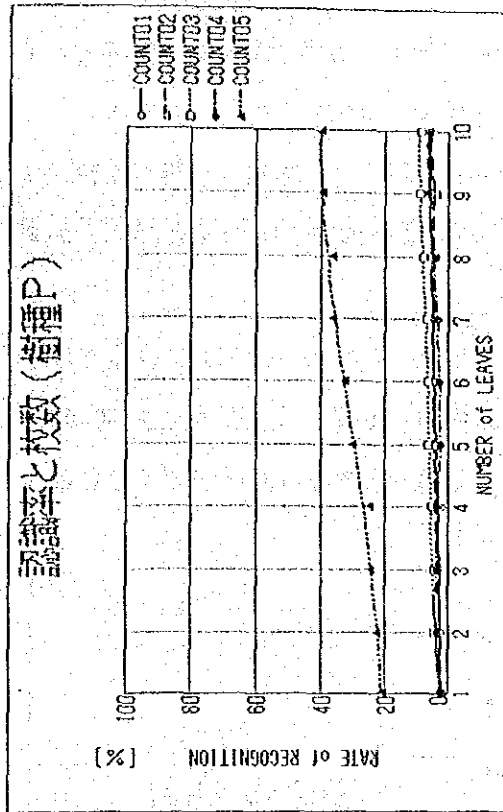
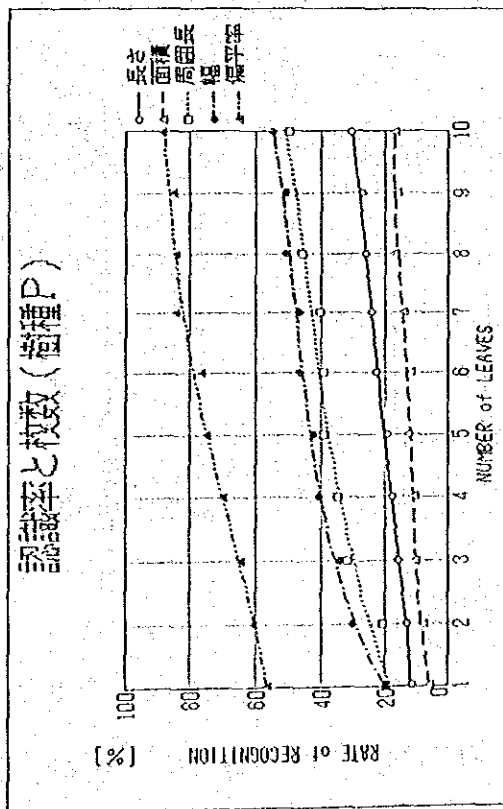


図 6. 16 各特徴パラメータの枚数と認識率の関係 [標本樹葉の樹種 - P]

照)のように、標本数の増加に伴ない認識率が低下する場合がある。標本として樹種Oを用いた場合に、主に誤って認識する樹種は樹種Mと樹種Jである。

標本として樹種Oを用いてその枚数を増加した場合に、樹種Oと正しく判定する確率(認識率)と樹種Mおよび樹種Jと誤って認識する割合(誤認識率)の変化を図6.17に示す。

標本数を増加すると樹種Mの誤認識率は多少上下しながら全体的には徐々に下がっているが、樹種Jの誤認識率は全体的に増加しておる。樹種OをCOUNT2で検索した場合の認識率の低下は主に樹種Jとの分布の位置関係が原因であると考えられる。

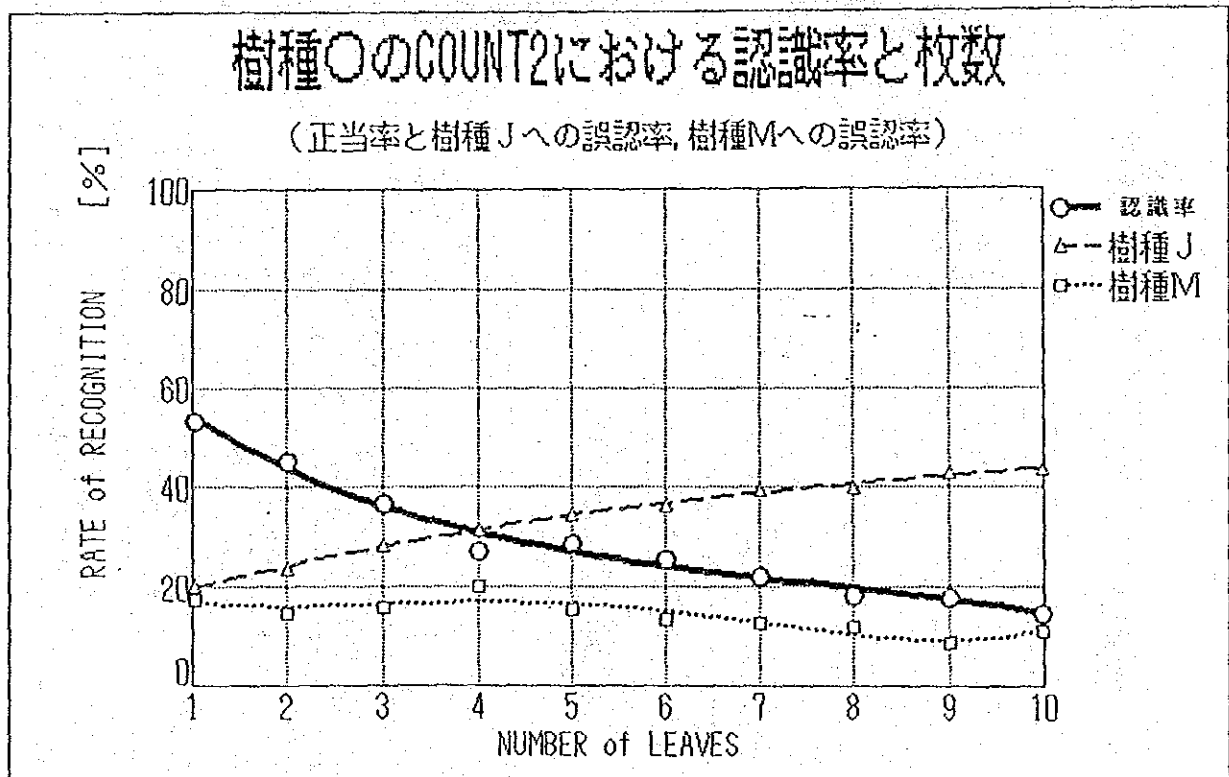


図6.17 標本樹種Oの時の樹種M及び樹種Jの誤認識率の枚数による変化

さらに、図6.18に標本数の違いに対する標本平均の確率密度分布近似曲線の変化を示す。

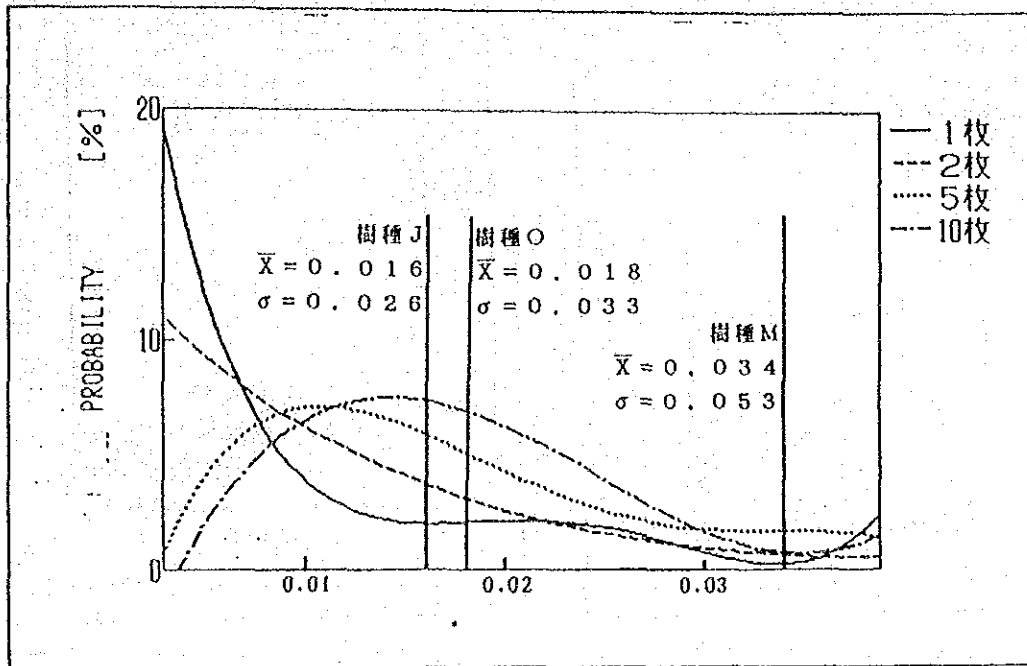


図 6. 18 標本樹葉枚数に対する標本平均の確率密度分布近似曲線の変化

図 6.18 を見ると、標本数が増加するに従い標本平均の分布のピークが徐々に右にずれてきている。そのため、樹種 J の平均値付近の確率密度が高くなり、樹種 J と誤って判断する割合が高くなってきているものと思われる。

さらに、図 6.19 ~ 図 6.21 に示すマハラノビス距離の分布を見ても、真の樹種 O と誤認した樹種 J はほぼ似たような分布になっており、同様の事が言えるであろう。

標本数をさらに増加させると標本平均のピークが樹種 J の平均値を越えて真の樹種 O の平均値に近づき、その結果認識率が増加することが予想できる。

この事は、マハラノビス距離の分布図より、誤認したときのマハラノビス距離幅 ( $W_{max}$  と  $W_{min}$  の間隔) が狭くなってきていることから予想される。

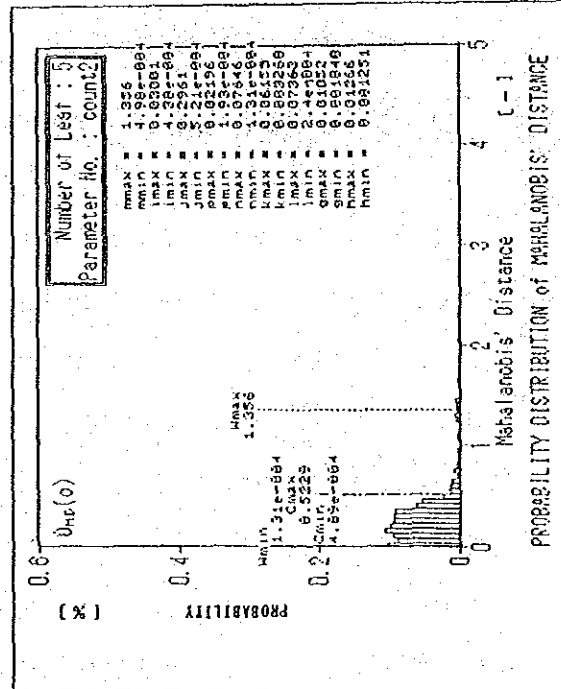
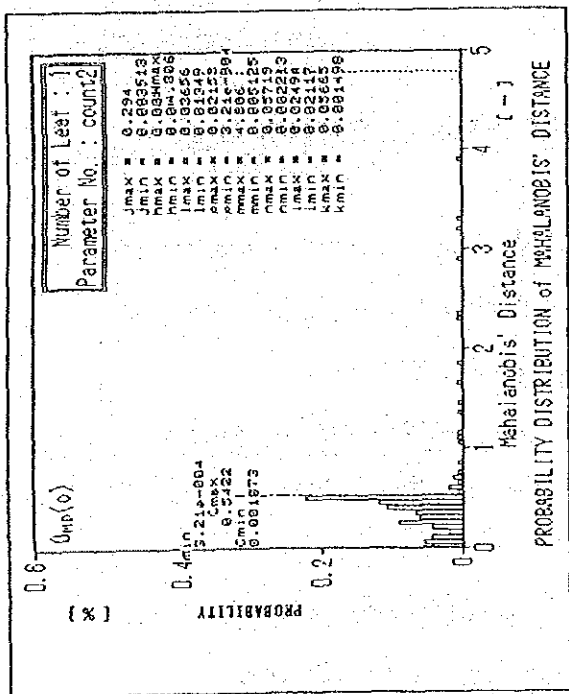
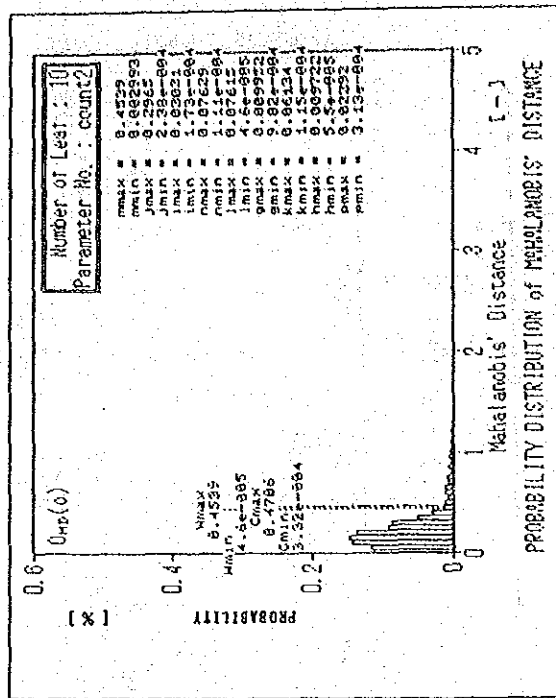
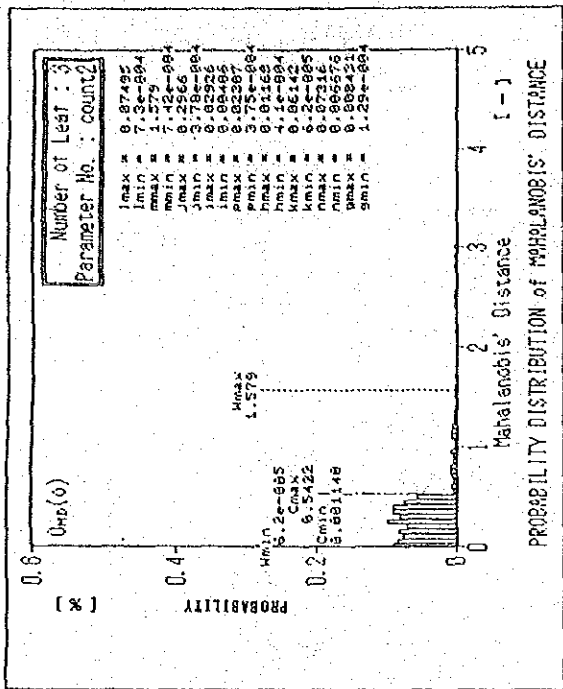


図 6.19 標本樹葉の樹種のマハラノビス距離の分布

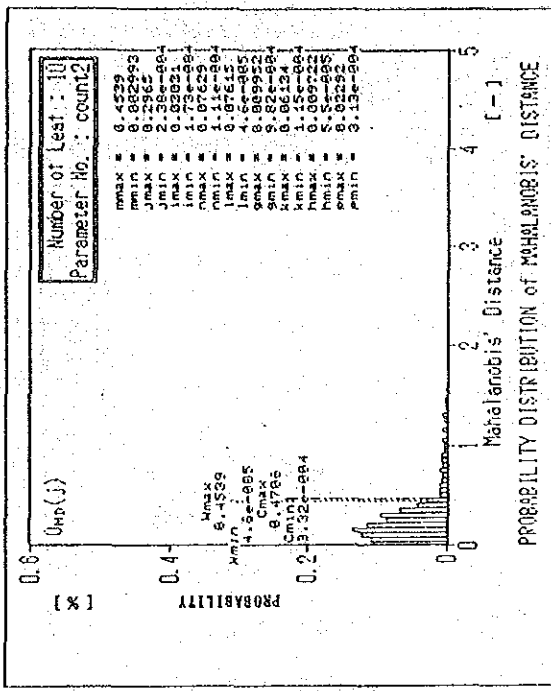
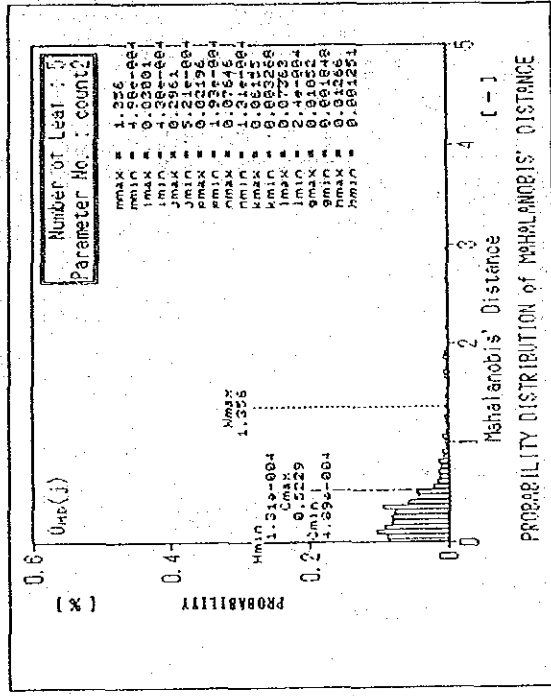
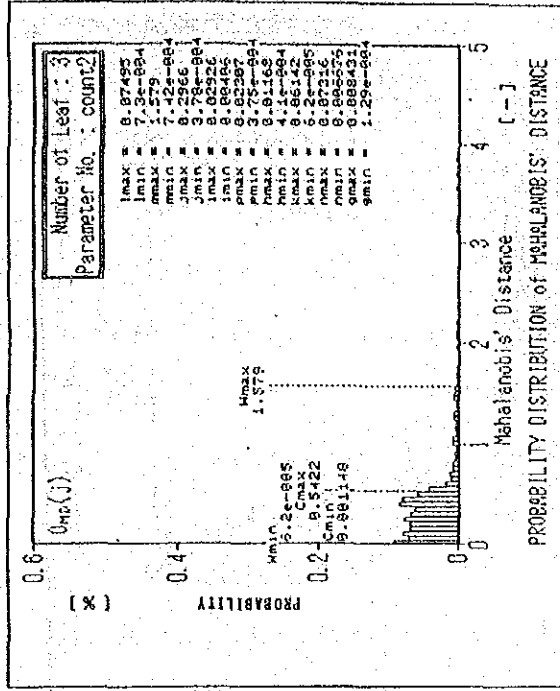
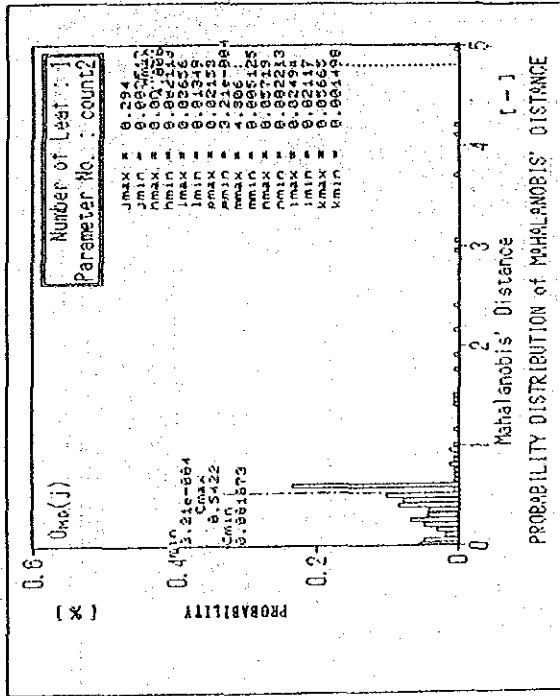


図 6.20 標本樹葉の樹種 0 の時樹種 J と誤認識したマハラノビス距離の分布

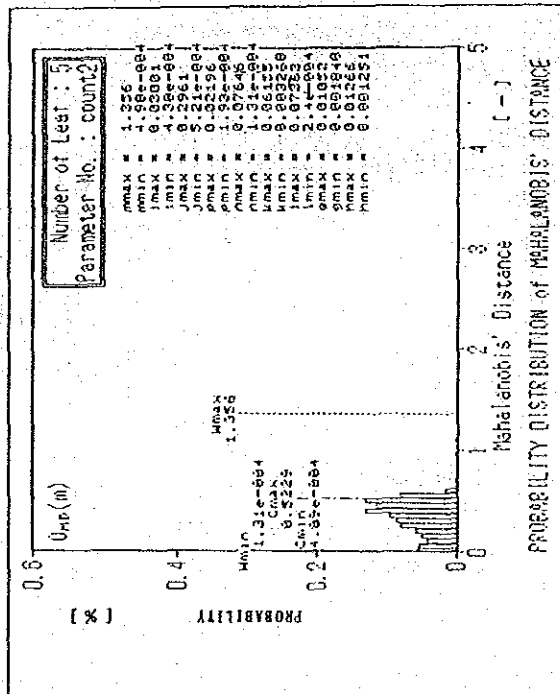
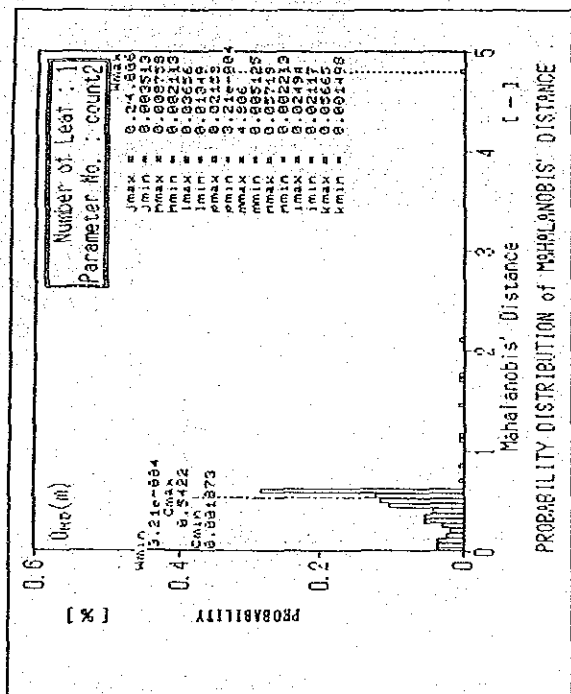
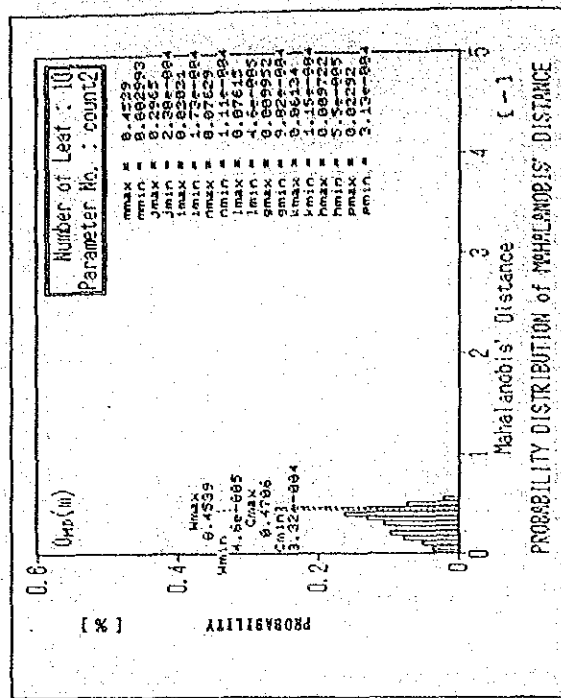
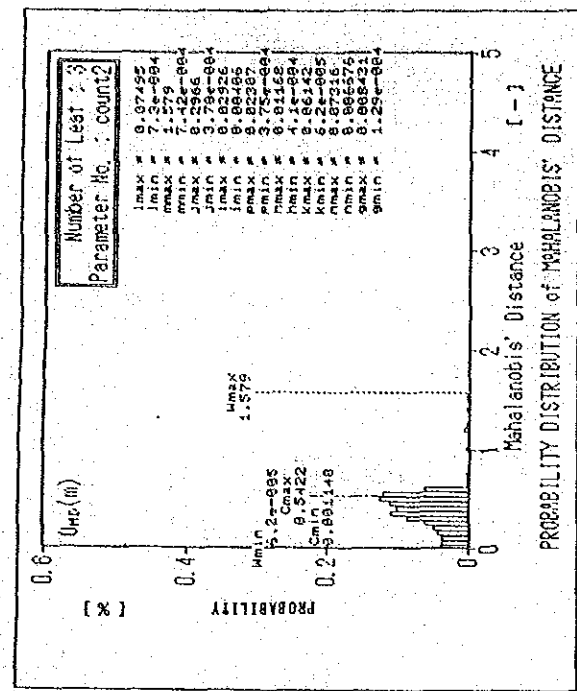


図 6. 2. 1 標本樹葉の樹種 O の時樹種 M と誤認識したマハラノビス距離の分布



## 6. 2. 2 特徴パラメータの有効性

樹種Hに関し、単一の特徴パラメータで最も認識率の高い2パラメータ(TOP2:面積と偏平率)と周囲形状度を除いて最も認識率の低い2パラメータ(WORST2:頂点数と占有率)およびこれらの4つを全て用いて検索した場合の認識率と標本数の関係を図 6.22 に示した。

図 6.2 の樹種Hと図 6.22 を比べると、全パラメータを用いた場合、TOP2を用いた場合および MIX(=TOP2+WORST2) を用いた場合のそれぞれによる認識率には余り大きな差はない。この事は、全体の特徴パラメータを用いた場合でも単独で有効な特徴パラメータの影響が大きくきいてくることが分る。

また、図 6.8 の頂点数および占有率だけを用いた場合の認識率と図 6.22 の WORST 2 の認識率を比べると、2つの特徴パラメータを同時に用いた WORST 2 による認識率のほうが、単一の特徴パラメータで得た認識率の和よりも高くなる。

この様に、樹種により単一で有効な特徴パラメータは異なる点や、単一では余り有効でない特徴パラメータも組合せにより有効となる点を考え合わせると、特徴パラメータの数が多いほど認識率を高くする可能性がある。

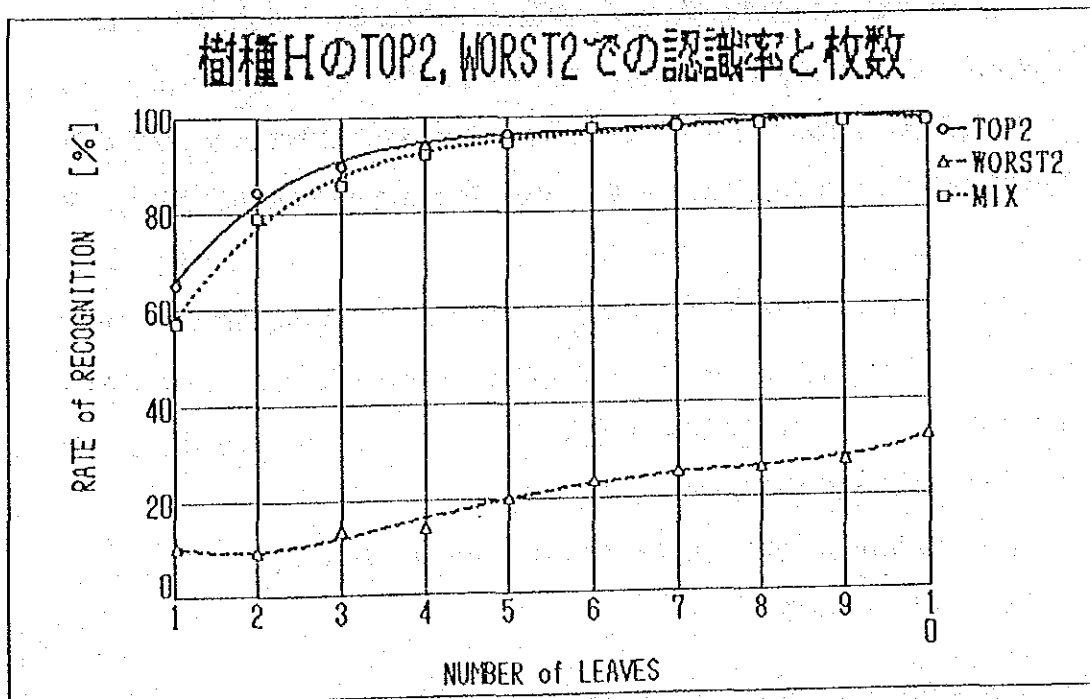


図 6. 22 樹種HのTOP2とWORST2のパラメータによる枚数と認識率の関係

## 第7章 結 論

国際協力事業団からの業務依頼により『自動樹木検索システム』を開発し、その性能確認試験を行なった。

本業務依頼により開発されたシステムは、樹葉の外形形状から得られる特徴を画像処理により抽出し、これらを用いて樹種を自動的に検索する方式のシステムである。今回、システムの開発と同時に、専門家によりバブア・ニューギニアで採取されたおよそ50種の樹種から10種類を選択し、本システムの性能の確認を行なった。これらの試験を通して、以下の結論を得た。

(1) 本試験の範囲では、標本樹葉枚数が1枚で判定を行なった場合の認識率は約50%～約90%である。しかし、標本樹葉を複数枚にして判定を行なった結果、5枚程度の標本で約100%の認識率を得た。

この結果から、本システムの検索方法の信頼性を確認することができた。

(2) 標本樹葉枚数を増やすに従い標本平均の分布が先鋭化し、それに伴い認識率を向上することができる。

この結果から、標本平均を新たな確率変数として用いることのできる検索対象物に関しては、本システムにおける検索方法を有効に利用することができる。

(3) 単独では認識能力の低い特徴パラメータと認識能力の高い特徴パラメータを組合せて使用した場合、全体の認識率は認識能力の高い特徴パラメータに大きく依存する。また、単独では認識能力の低いパラメータでも、これらを組み合わせることにより認識能力を高めることができる。

これらの結果から、単独では認識能力の低いと考えられるパラメータも有効に利用することが可能である。

以上の結果から、本システムは、樹葉の外径形状を用いて樹種を自動的に判断する『自動樹木検索システム』として十分に有効な装置であり、実用可能である。

## 第8章 問題点および今後の課題

### 8.1 対象となる樹葉の収集に関する提言

(1) 樹葉の採取方法 …… 本システムの性能および有効性を試験するとともにデータベースを構築する目的から、本システムの設置予定地であるパプアニューギニアに生息する樹木の樹葉を収集した。収集に当っては、専門家が現地に入り、およそ50種類の樹葉を集め、乾燥処理および整理をした上で日本に空輸した。しかし、処理時間の不足や移送途中に時間が掛りかなりの樹葉が腐ったり霉が生えたりして、データベースを製作するに十分な枚数を確保できなかった。採取は、50種類ほど行なわれたが、利用できた樹種はわずか10種類であった。

確実なデータベースを構築するために、今後は現地で直接画像入力することが望ましい。

(2) 樹葉の管理方法 …… 樹葉の特徴パラメータはその樹葉の状態で異なる。図 8.1 は同一の樹葉から乾燥前(A)と乾燥後(C)に抽出した複雑度の例である。また、図 8.2 は、樹種Aと樹種Cに関する標本数の認識率の関係を示す。この場合は、特徴パラメータとして複雑度と偏平率のみを用いた。

この結果を見ると、認識率に関してはパラメータ数が2と少ないにもかかわらず、標本樹葉枚数10枚で約90%の認識率が得られた。これは、複雑度の分布の違いに依存している。また、これらの図から明らかなように、人間の目では似ているものでも特徴パラメータに違いが生じる。したがって、実際に検索を行なう場合には、一定の管理状態で構築されたデータベースを用いて、同様の状態の標本を使用する必要がある。

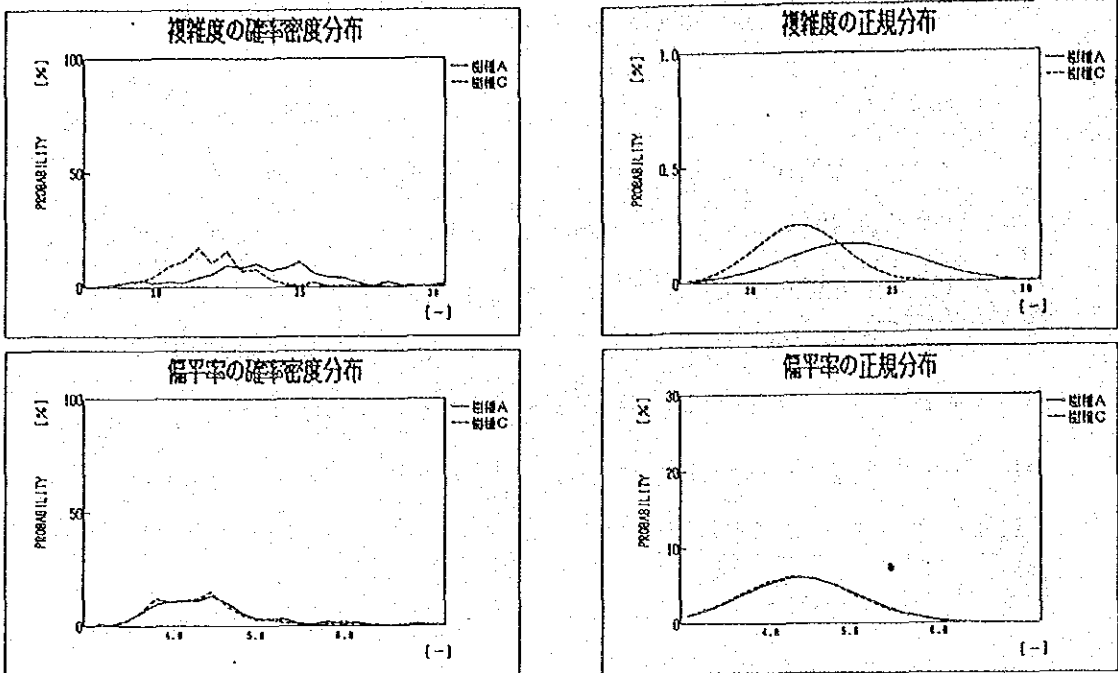


図8.1 樹種A及び樹種Cの特徴パラメータの分布

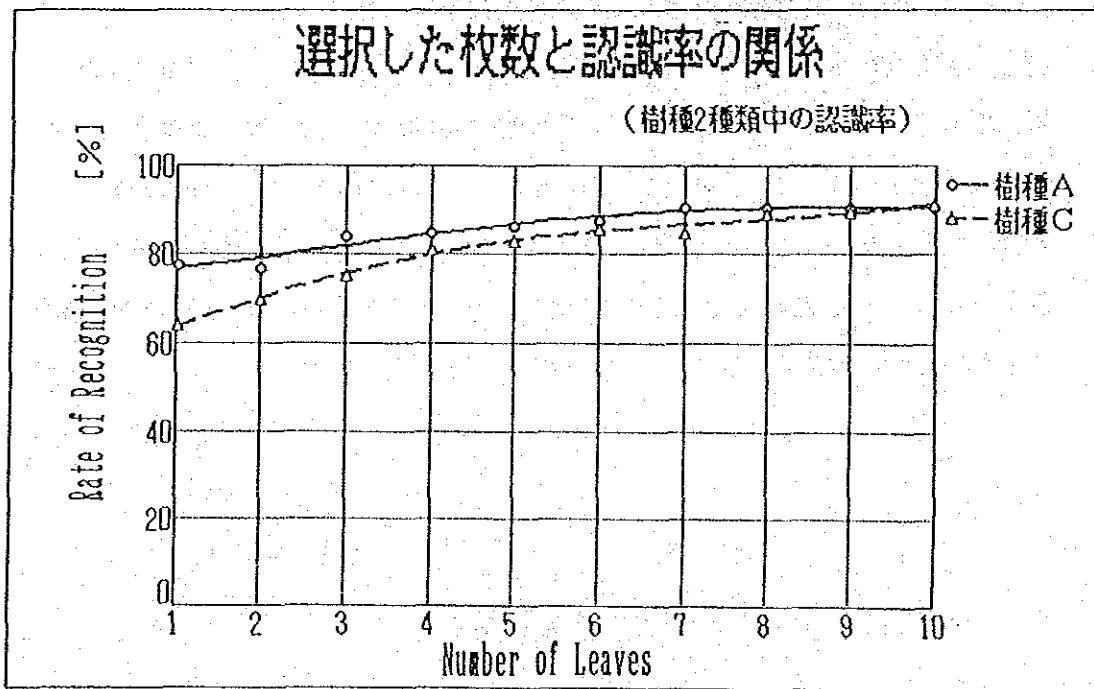


図8.2 標本樹葉の枚数と認識率の関係

## 8. 2 データベースの構造およびデータの扱いに関する提言

(1) 対象の樹種数とデータベース …… 本報告書では、10種類の樹葉を対象に取上げ、システムの性能試験を試みた結果である。対象となる樹種数が増加すれば、特徴パラメータの分布の位置関係は一層複雑になり認識率は本報告書の結果より低下することが予想される。しかしながら、本システムの検索方法では、前項で示した通り、人間の目では判別できないようなわずかな違いでも特徴パラメータに差が生じることから、樹種数が増加しても充分検索可能であると考えられる。しかし、少しでも認識率を高くする必要から、

- (a) データベースの構造化およびプロダクション・ルールの採用
- (b) 誤認識しやすい樹種間の差別法の確立
- (c) 特徴パラメータの見直しと追加

などを検討し、認識率の低下を防ぎながら検索効率を上げる努力が重要である。

(2) 確率密度分布の近似方法 …… 本報告書の範囲では殆どの特徴パラメータの確率密度分布は正規分布に近似することができる。しかしながら、幾つかは正規分布では近似できない特徴パラメータが存在する(APPENDIX 2 参照)。本報告書の範囲で正規分布で近似することが必ずしも適当ない特徴パラメータは、周囲形状を示す COUNT である。特に COUNT の内、COUNT1 や COUNT2 は、殆どの樹葉が0となりその他がわずかな値をもつ。したがって、頻度は0が最も多くなり COUNT数が増加するにしたがい頻度が下がる傾向にあり、偏った頻度分布を示す。この偏りのある頻度分布を対象の正規分布で近似することは難しい。また確率密度のピークが2つないし3つある分布等も存在する。これらの分布については、特徴パラメータの決め方を検討したり、近似する方法を検討する必要がある。

現状のシステムの周囲形状は、各境界画素と重心の距離をその最大値で無次元化し、その最大値と0の間を10区間に区切りそれぞれの区間毎に樹葉内部の画素数を数え、これを周囲形状のパラメータとした。しかし、COUNT の小さい区間では抽出結果の値が0となるものが多く含まれているため得られた分布は偏ってします。今後は、各境界画素と重心の距離をその最大値で無次元化した値の最大値と最小値の間を10区間に区切りそれぞれの区間毎に樹葉内部の画素を計算するこ

とにより偏りのない特徴パラメータを定義できると考える。

(3) 頂点数に関する考察 …… 現状のシステムの頂点数は、人間の目で突起と認めるものを基本に考えてある。しかし、これは個人差があり極めて主観的になってしまう。今後は、しきい値を順次変化して各大きさの突起を求め、これら全てを特徴パラメータとして扱うことが考えられる。この方法により、楓の様に大きな突起の中に小さな突起を持つような樹葉も扱うことが可能になると考えられる。

### 8.3 検索の結果の表示に関する提言

現状のシステムでは、データベースにある特徴パラメータの分布を用いて全マハラノビス距離を求め、最も小さくなる樹種を真の樹種と判定する。今後は、真の樹種の最大のマハラノビス距離を考慮に入れ、たとえ、最小値であってもある値以上の場合、「樹種不明」あるいは「樹種Yの可能性はX%である」といった形式の解答を得られるようなシステムにかえる必要がある。

## APPENDIX 1

標本の樹種と母集団の樹種の組合せによる  
全マハラノビス距離の分布の違い

標本を樹種Iとした場合の標本数の  
違いによる全マハラノビス距離の分布





# 標本の樹種と母集団の樹種の組合せによる全マハラノビス距離の分布の違い

(標本を樹種 I とした場合の標本数の違いによる全マハラノビス距離の分布)

本頁より4頁にわたり、標本として樹種 I を用いた場合の全マハラノビス距離の分布を示す。各図中、左上に示めす大文字のアルファベットは標本の樹種（ここでは、樹種 I）を示し、括弧内の小文字のアルファベットはマハラノビス距離を算出する際の母集団の樹種を示す。

例： I<sub>MO(h)</sub>……標本が樹種 I で、樹種 H の特徴パラメータの分布を用いて求めた全マハラノビス距離

樹種 I を標本として、樹種 I の特徴パラメータの分布を用いて求めた全マハラノビス距離は、標本数が増加するにしたがって分布の幅が狭くなりその平均値は 0 に近づく。これに対し、樹種 I 以外の特徴パラメータの分布から算出される全マハラノビス距離は、平均値が 0 以外に漸近する。前者の分布と後者の分布は、標本数がある枚数以上では完全に独立し判定が可能となる。

