

NO.

自動樹木検索システム開発業務

報告書

平成3年2月

国際協力事業団

林 開 発

91-03

J R

JICA LIBRARY



1089066131

2215

NO.

自動樹木検索システム開発業務 報 告 書

平成 3 年 2 月

国際協力事業団

林 開 発

91-03

J R

国際協力事業団

22115

目次

要約	1
1. システム開発の背景	2
2. システムの目標	3
2.1 システムに対する要求項目	3
2.2 システムの設定目標	3
2.2.1 システムに対する考慮点	3
2.2.2 システムの基本構造	4
3. システムの構造と構成	6
3.1 システムの構造および流れ図	6
3.2 システムの構成	7
3.2.1 システムの構成図	7
3.2.2 システムを構成する装置	8
4. 理論的背景	10
4.1 特徴抽出部	11
4.1.1 各パラメータの定義	11
4.1.2 各パラメータの算出方法	12
4.2 データベースの表現と構造	18
4.2.1 データベースの表現方法	18
4.3 判定係数の算出と算出方法	21
4.3.1 判定の流れ	21
4.3.2 マハラノビスの汎距離による判定係数の算出	22
4.3.3 補正を加えた判定係数の算出	26
(a) 特異値をもつ樹種の削除	26
(b) 樹種の分布の割合による重み付け	29
4.3.4 認識率の定義	31
5. 装置性能確認試験の方法	33
5.1 検索対象樹種	33
5.2 シミュレーションによる性能の確認方法	33
5.2.1 被検索樹葉	33
5.2.2 性能の確認方法	38
6. 装置性能確認試験の結果	39
6.1 判定係数の算出法とその有効性	39
6.1.1 本システムによる判定結果	39
6.1.2 判定係数の算出法と認識率の関係	40
6.2 特異値をもつ樹種の削除	42
6.3 樹種間の分布の度合いによる重みづけ	43
7. システムとしての性能の検証	45
8. 結 論	48
APPENDIX	49
記号表	50
用語定義表	51
画像処理用語表	52

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是有用木材、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性がある。

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、すでにデータベースに格納されている特徴の分布を用いて判定係数を算出し、さらに補正などを加えて、その大小で最も可能性の高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。

1. システム開発の背景

森林、中でも熱帯林は、熱帯地方の人々の食糧や木材や薪炭材等の燃料の供給源であるだけでなく、世界全体にとっても、野生鳥獣のうち約半分が生息していると推定される「生物種の宝庫」として、又、地球規模での気候の緩和など、環境調節機能を有するものとして、更には、温暖化の原因物質である二酸化炭素の吸収源として、その重要性の認識が年々高まっている。しかし、その一方で、急激に減少しつつある事が、近年の調査で明かとなり、大きな社会問題となっている。森林の減少の要因は様々であるが、自然回復力を無視した焼畑式移動耕作、及び、有用樹林の強度伐採がその主な原因とされており、それに対する森林の再生等の処置は未だに効果を挙げていないのが現状である。一方、最近における世界経済の拡大傾向は木材需要の増加を促し、林業をその主な産業としている開発途上国では、人口増加や貧窮等の経済的理由を背景として、ますます熱帯降雨林の無理な伐採を強いられている。その為、先進国初め開発途上国においても最近では森林再生時の植林に有用な木材、造林適性樹種の研究開発が促進されているが、その基本となる樹木の検索にあたっては現在特別な専門学者あるいは数少ない木材利用専門学者に頼らざるを得なかった。そこで、最新のコンピュータ技術を利用した樹種の検索システムを開発することで、今後の有用樹種の拡大、選定ならびに適性造林樹種の選定の手段となる事が期待できる。

本報告書は、国際協力事業団の業務依託に基づき開発された「自動樹種検索システム」に関する報告書である。本業務依託により開発したシステムは、葉面の画像を中心とした特徴を画像処理装置により求め、これらを用いて自動的に樹種を検索するものである。本報告書は、システムの目標、システムの構成、理論、装置の性能の検証試験・結果、及び、システムとしての性能の検証で構成されている。

2. 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りである。

- (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%以上の正当率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%程度の正当率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の設計を行う必要がある。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 システムを作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、対象とする樹種数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インドネシアで採取された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、平面で扱い得る樹葉のみとする。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。

3. 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、特徴抽出部、データベース部、検索部の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。

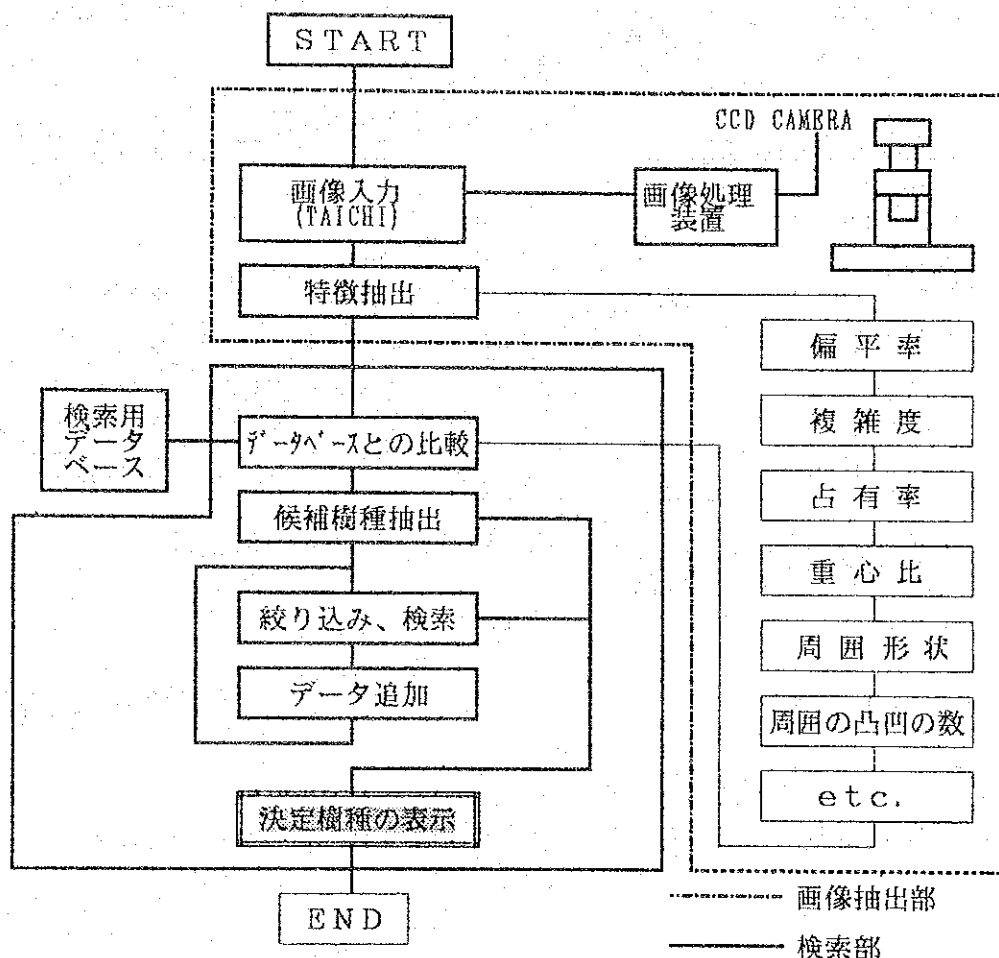


図3.1 自動検索システムの基本的な流れ

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。

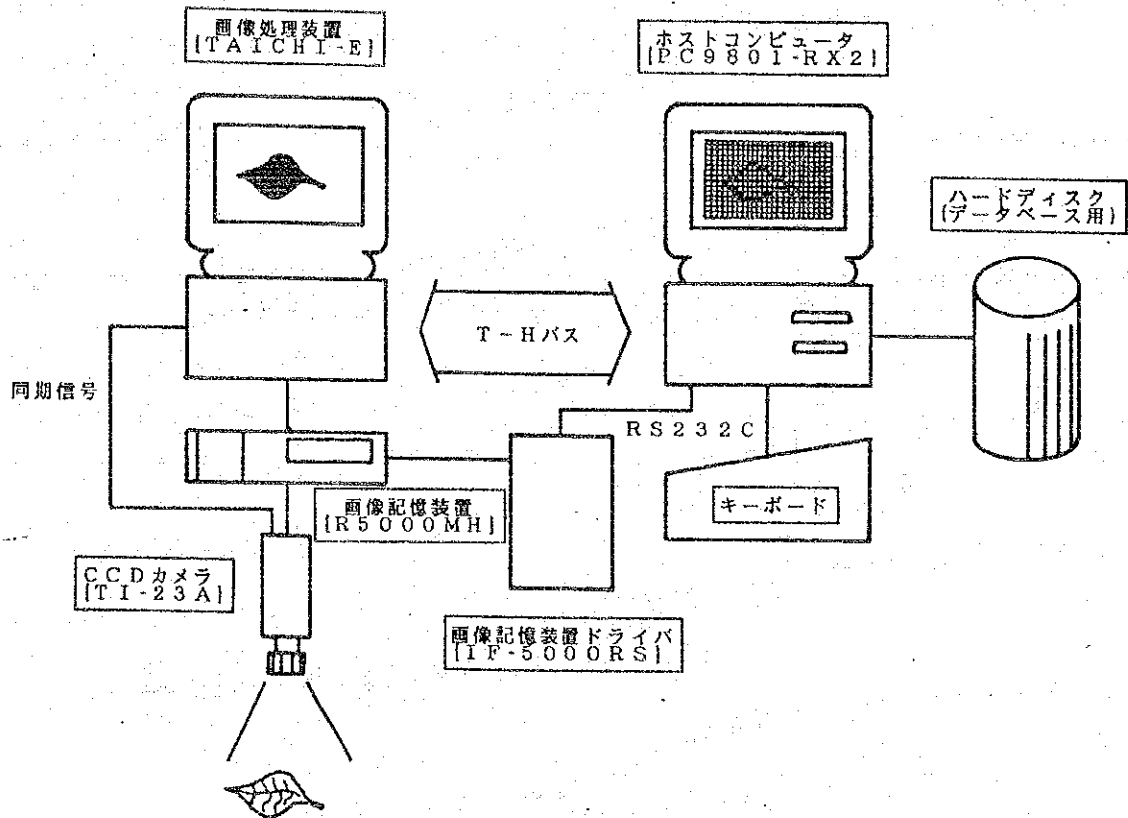


図 3.2 実験装置の構成図

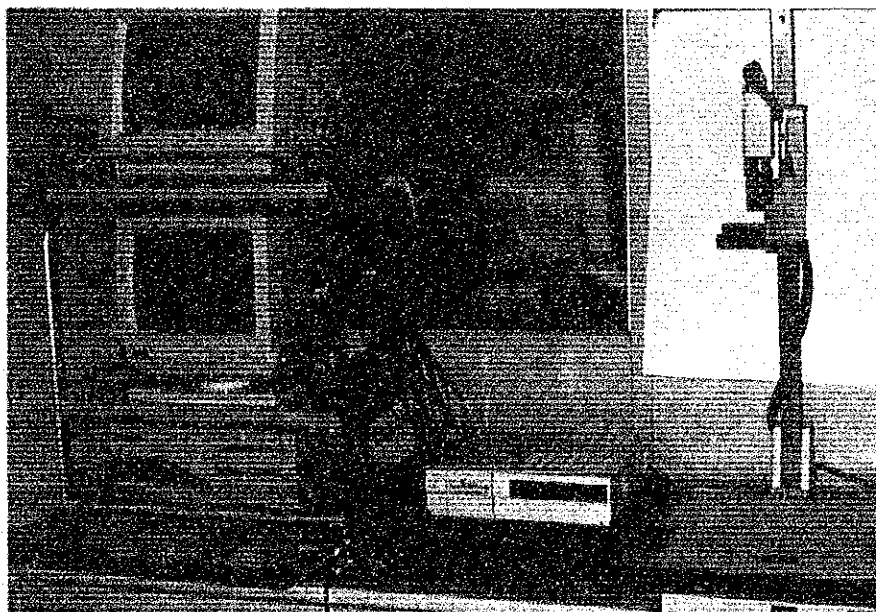
3.2.2 システムを構成する装置

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。

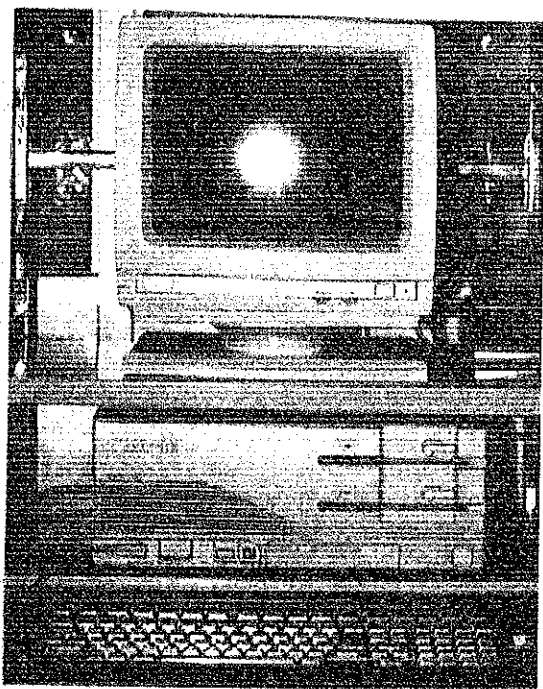
表 3.1 実験装置の概要表

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画素 (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-B)	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 (1F-5000RS)	外形寸法	62W×99H×200D (mm)

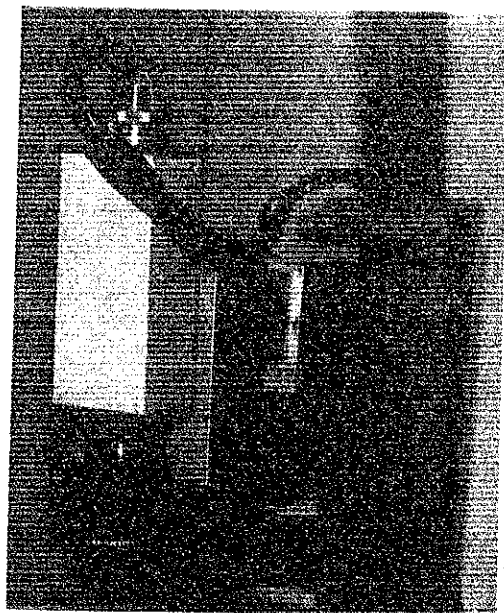
(1) 樹木自動検索装置



(2) 樹木自動検索装置構成機器

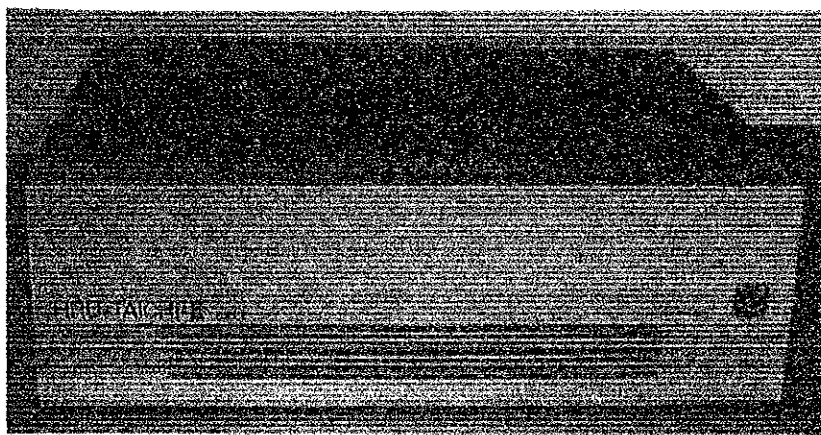


コンピュータ (PC-9801-RX2)

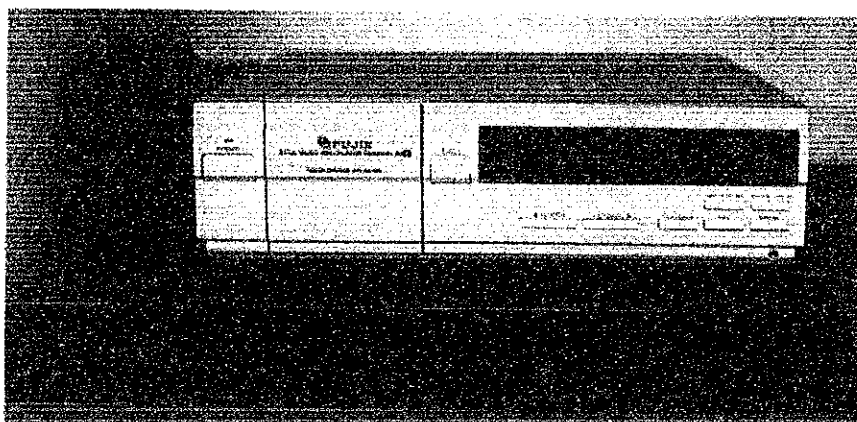


CCDカメラ

画像処理装置
(TAICHI-E)



画像記録装置
(R-5000-MH)
(IF-5000RS)



4. 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、本システムは、(1)特徴パラメータを算出する特徴抽出部、(2)データを保存するデータベース部、(3)特徴パラメータから樹種名を検索する検索部、の三つから構成される。

- (1) 特徴抽出部： 各標本から画像処理により23種のパラメータ値を抽出する。抽出する過程は、巻末のAPPENDIXに記してある。この時、各標本から得られる特徴は、樹葉の形状的な性質に限られる。これは、樹葉の色、葉脈、表面あらさなどの特徴も、抽出可能であるが、採取してからの経年変化やその他の自然的な要因に作用されやすい為で、本システムでは用いていない。
- (2) データベース部： 各標本から得られた特徴は、予めその各特徴を保存したデータベースと比較、判別される。母集団から得られた各パラメータの値は、ばらつきを持つ分布を形成し、その分布を正規分布に近似することで、データベースを構成している。
- (3) 検索部： データベースには、各樹種の特徴が収められている。その各樹種の特徴と標本との可能性を算出し、求められた判定係数をもとに判別分析を行なうところが判断部である。判別方法として、多変量及び一変量のマハラノビスの汎距離が採用され、その値の大小で候補樹種の選定が行なわれる。

4. 1 特徴抽出部

葉の特徴を抽出する場合に外形的な形状、葉脈等の内部の特徴、色や硬さなどの質的な特徴などを抽出することができる。葉脈は、樹種によっては画像処理による抽出が可能であるが、全ての場合に適用できる訳ではない。又、色は、樹葉を抽出してからの時間や、季節、及び、その樹種の植生していた場所等にも左右される。これらの事を考え、本研究では、外形的特徴からの情報のみを扱うこととした。以下に、その特徴抽出部の詳細を示す。

4. 1. 1 各パラメータの定義

本研究では、葉の各特徴をそれぞれ、分布のもつ特徴パラメータとして扱い、そのパラメータ群で各樹種を特徴付けできるものとする。パラメータは以下の23種である。

表4. 1 各パラメータの定義と意味

番 号	パラメータ	記 号	パラメータの持つ意味
1	長さ	$L \text{ (mm)}$	樹葉の長さ
2	幅	$W \text{ (mm)}$	樹葉の幅
3	面積	$S \text{ (mm}^2\text{)}$	樹葉の面積
4	頂点数	Ch	樹葉の頂点数
5	包絡点数	Cv	樹葉の包絡点数
6	重心率	Lg	樹葉の重心率
7	偏心率	L/W	樹葉の偏心率
8	重複率	L^2/S	樹葉の重複率
9	占有率	$S/(L*W)$	樹葉の占有率
10	重心比率	Lg/L	樹葉の重心比率
11	頂点比率	Ch/I	樹葉の頂点比率
12	包絡面積比	SV/S	樹葉の包絡面積比
13	包絡形状		樹葉の包絡形状
14-23	周囲形状		樹葉の周囲形状

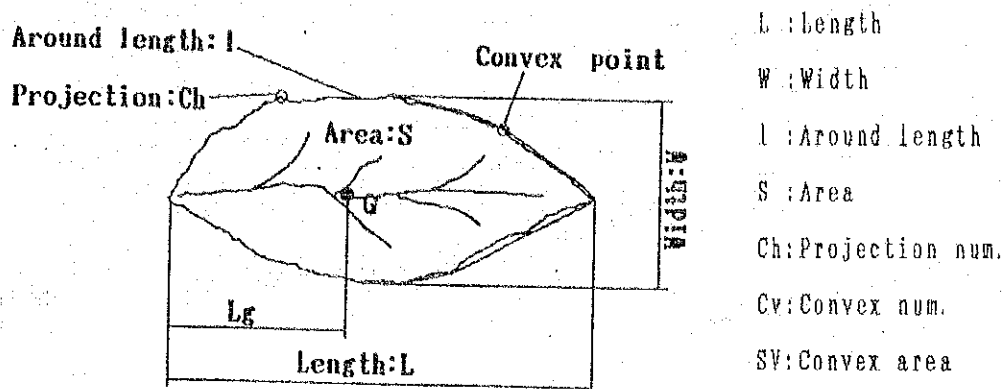


図 4. 1 葉の各部の定義

4. 1. 2 各パラメータの算出方法

特徴量のなかには、長さ、幅、周囲長などの画像処理装置の機能により算出される次元をもつ 7 種のパラメータの他に、コンピュータで演算を行ない算出される無次元量パラメータも存在する。それらの算出過程に於ける処理方法について以下に記す。

(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最大長さである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、

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

$$\text{最大幅} : B = (1^2 + 5^2)^{1/2} / bl = 26^{1/2} / bl \quad \dots\dots (4-2)$$

となる。ここで、 bl は単位長さ当りの画素数である。

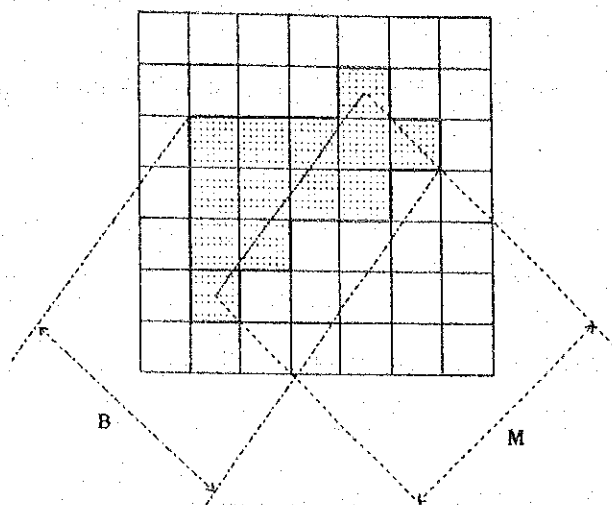


図 4. 2 図形の最大長および最大幅の例

(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。 g 個の画素から構成される閉図形の面積 S は、

$$S = g / bl^2 \quad \dots\dots (4-3)$$

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。

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

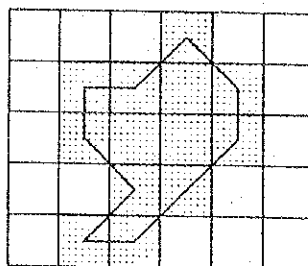


図 4. 3 面積の例

(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 4.2 の場合、周囲長 l は、

$$\text{周囲長} : l = (4/b1 + 6 \cdot 2^{1/2}/b1) \dots\dots\dots (4-5)$$

で表わされる。

(4) 頂点の算出法

頂点は円状探索法、及び、偏角差分関数による頂点決定法と同様の方法を用いて抽出を行なった。

円状探索法とは、画像処理装置により行なわれる8点近傍の境界画素抽出により、得られたデータをもとにおこなう細線化処理の事で、図 4.4 のように、基準点周りを前探索点から、反時計回りに画素を探索して、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言う。

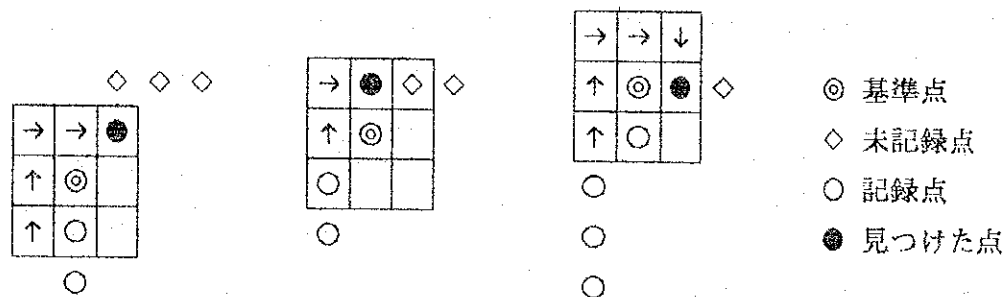


図 4. 4 円状探索法

偏角差分関数とは、図 4.5 に示すように、ある点と前後 g 番目（現状では、カメラの倍率の違いによる誤差をなくす為、倍率により変化させている）の点を結ぶ線のなす角度を記録されたすべての点に関して計算する。そして、その角度がある範囲内（現状 20度以上 160度以内）にあり、かつ前後の点で計算した角度よりも 80度に近いとき、その点を頂点と判断する。例えば、下記の図では、 g を4としていて、中央の画素のなす角度が前後の画素のそれよりも鋭角なので、この画素は頂点として認識される

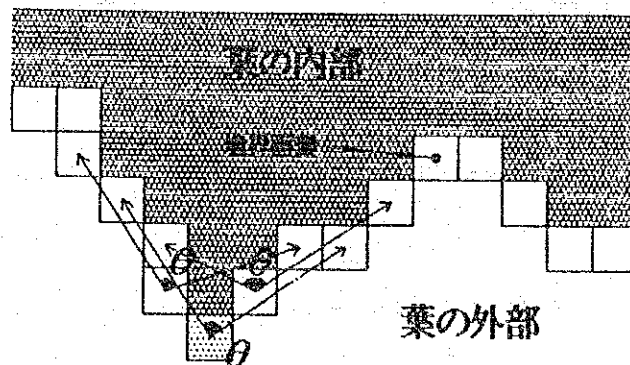


図 4. 5 偏角差分関数

(5) 周囲形状量の算出

周囲形状パラメータは図 4.6 の様に重心から境界画素までの距離の最小値から最大値までの区間を 10 等分した時、それぞれの各区間に含まれる画素数を総画素数で除した値で構成され、これらの特徴量から周囲の細かい複雑さを見ることができる。

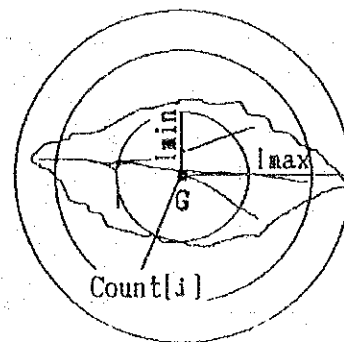


図 4. 6 周囲形状パラメータ

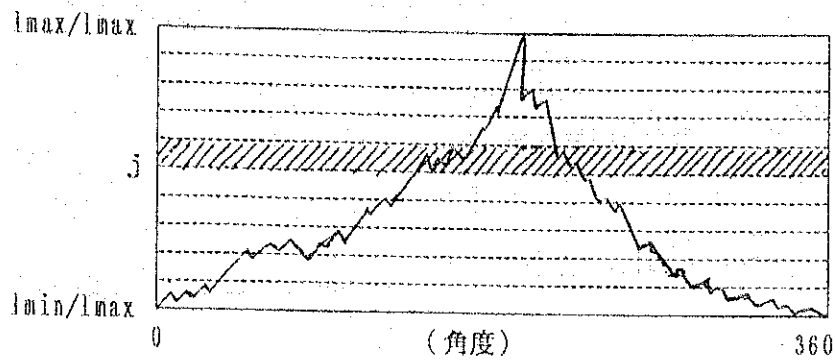


図 4. 7 周囲形状パラメータの算出過程

図 4.7 に示すように、区間 j にはいつてきた画素の数を数え、以下の式で周囲形状 $[j]$ が求まる。

$$\text{周囲形状 } [j] = \frac{\text{10等分した内の } j \text{ 番目にはいつてきた画素数}}{\text{総画素数}} \quad \dots\dots\dots (4-6)$$

(6) 包絡面積比

包絡面積比とは、包絡点の囲む閉包の面積と図形の面積との比の事で、このパラメータにより、葉の周囲への張り出し具合をみることができる。

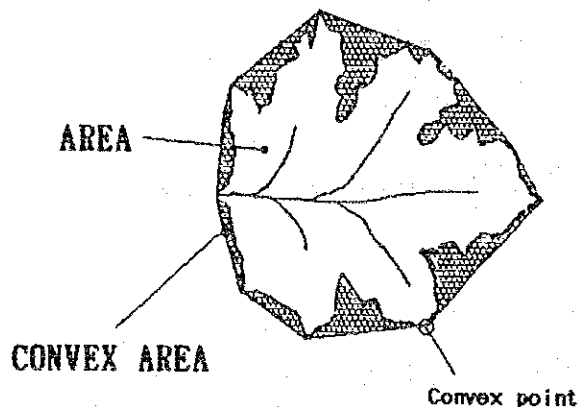


図 4. 8 包絡面積比の定義

このパラメータは、既に挙げられた周囲形状パラメータと類似の特徴を有するが、周囲形状パラメータが周囲の細かい部分に着目しているのに対し、このパラメータは、図形中の大きな凹部の割合の検出を行なうことを目的としている。

4.2 データベースの表現と構造

本診断システムは分布を持つ対象物を診断するためにデータベースとの比較を行ない判定するものである。よって、データベース部はその対象物を表現するために多くの情報を持たなくてはならない。しかし、あるパラメータの分布を表現するのにパターンマッチング的固定を行なうとすれば、分布の形までをデータとしてもたなければならない。その為、数値的处理が簡易で、統計的处理が可能である正規分布で表すものとし、データベースには平均値と標準偏差、分散・共分散行列が書込まれる。したがって、データベース部は、以下のようなになる。但し、ここでは、パラメータを $P_1 \sim P_P$ 、樹種を $G_1 \sim G_n$ としている。

4.2.1 データベースの表現方法

(a) 正規分布

正規分布は、ガウス分布とも言われ、その確率変数 x は、その確率密度関数が下記で表される時、正規分布 $N(\mu, \sigma)$ を持つという。

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (x - \mu)^2 \right] \quad \dots\dots (4-7)$$

但し、 $-\infty < x < \infty$ とする。

また、正規分布は、確率密度関数にあたえられる μ 、 σ をそれぞれ平均、標準偏差としてもつ。

(b) 平均値と標準偏差

ある樹種のあるパラメータについてそれぞれ平均と標準偏差が存在する。
したがって、その構造は以下に示す通りである。

表 4. 2 データベースのデータ構造

	P ₁	P ₂	...	P _j	...	P _p
G ₁	μ_{11}, σ_{11}	μ_{12}, σ_{12}	...	μ_{1j}, σ_{1j}	...	μ_{1p}, σ_{1p}
G ₂	μ_{21}, σ_{21}	...	...	...	...	...
	...	...	...	...	...	...
G _i	μ_{i1}, σ_{i1}	...	...	μ_{ij}, σ_{ij}	...	μ_{ip}, σ_{ip}
	...	...	...	...	...	...
G _n	μ_{n1}, σ_{n1}	...	...	...	...	μ_{np}, σ_{np}

また、平均と標準偏差は以下の式で与えられる。

$$m_{ij} = \frac{1}{N} \sum_{m=1}^N \mu_{j,m}^{(i)} \quad \dots\dots\dots (4-8)$$

$$\sigma_{ij} = \left[\frac{1}{N} \sum_{m=1}^N (\mu_{j,m}^{(i)} - m_{ij}) \times (\mu_{j,m}^{(i)} - m_{ij}) \right]^{1/2} \quad \dots\dots\dots (4-9)$$

(c) 分散・共分散行列

分散・共分散行列は、樹種毎に存在する。分散とは、あるパラメータにおいて、自分自身の平均からの隔たりを表し、共分散とは、あるパラメータ同士の散らばり具合を表す。ここでは、樹種*i*についての分散・共分散行列について述べる。

表4.3 分散・共分散因子

	P ₁	P ₂	...	P _j		P _p
P ₁	S ₁₁	S ₁₂	...	...	...	S _{1p}
P ₂	S ₁₂	...	...	...	...	...
	...	...	...	...	...	...
P _j	...	...	...	S _{ij}	...	...
	...	...	...	...	...	...
P _p	S _{1p}	...	...	...	...	S _{pp}

それぞれの因子は以下の式で与えられる。

$$S_{jk}^{(i)} = \frac{1}{N} \sum_{m=1}^N (\mu_{j,m}^{(i)} - m_{ij}) \times (\mu_{k,m}^{(i)} - m_{ik}) \quad \dots\dots\dots (4-10)$$

4.3 判定係数の算出と算出方法

ここでは、判定に用いられる判定係数の算出とその方法について述べる。

4.3.1 判定の流れ

判定の流れは以下の図 4.9 に示す。

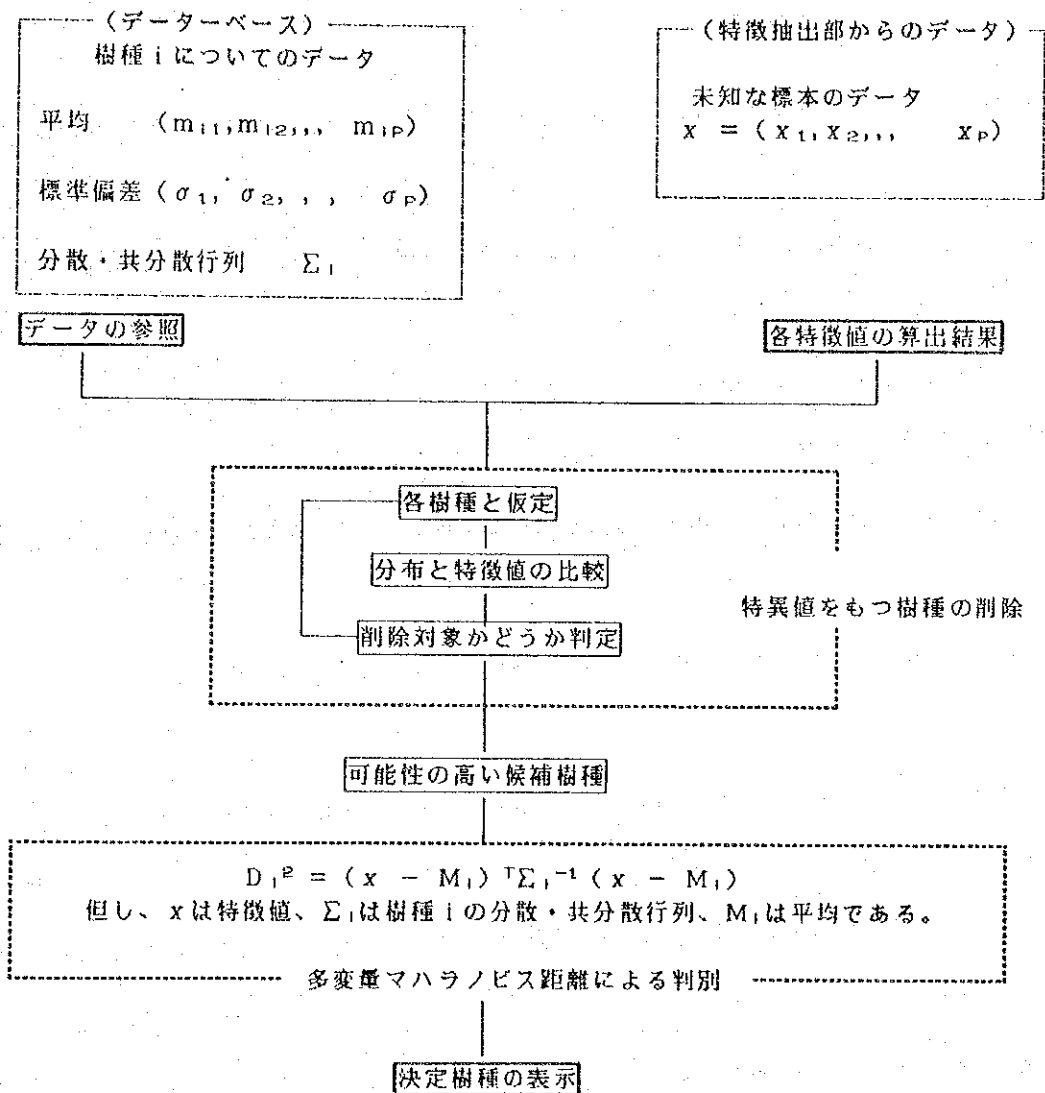


図 4.9 判断部の流れ図

4.3.2 マハラノビスの汎距離による判定係数の算出

多群の正規分布母集団から対象を判別する為に、本研究では、判定分析法におけるマハラノビスの汎距離を用いた判別式を採用した。マハラノビスの汎距離は、インドのマハラノビスによって考案された幾何的距離の事で、正規分布からの離れ具合を示す指標である。ここでは、理解を容易にする為、一変量においてのマハラノビスの汎距離から説明し、本研究で用いた多変量に拡張するものとする。

(a) マハラノビスの汎距離の定義

右図 4.10 のように二群の分布が与えられた時、ある未知な標本 x がどちらの分布に近いかを考えた時、その分布の勾配 (標準偏差) をも考慮に入れた指標が必要となる。それがマハラノビスの汎距離と言われるもので、平均からの差を標準偏差で除した値で、この例では、マハラノビスの汎距離は以下の様に与えられる。

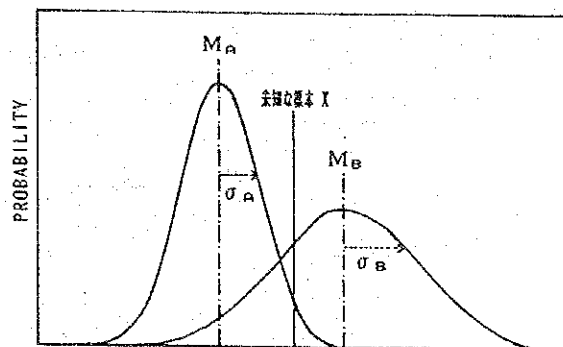


図 4.10 マハラノビスの汎距離

$$\text{分布 A} \quad dk_A = (x - M_A) / \sigma_A \quad \text{分布 B} \quad dk_B = (x - M_B) / \sigma_B$$

$dk_A < dk_B$ 未知 x は分布 A に属する可能性が高い

$dk_A > dk_B$ 未知 x は分布 B に属する可能性が高い

(b) 一変量のマハラノビスの距離

判断対象物の各特徴値から各パラメータの分布へのマハラノビス距離を次式で定義する。例えば、樹種 i のパラメータ j へのマハラノビス距離は以下の式で与えられる。ここでは、パラメータ $P_1 \sim P_P$ 、樹種 $G_1 \sim G_n$ があるとした場合 1 枚の葉から測定される各パラメータの値を $x_1 \sim x_P$ とする。

$$dk_{ij} = \frac{|x_j - m_{ij}|}{\sigma_{ij}} \quad \dots\dots\dots (4-11)$$

表 4. 4 一変量マハラノビスの汎距離のデータ構造

	P_1	P_2	...	P_j	...	P_P
G_1	dk_{11}	dk_{12}	...	dk_{1j}	...	dk_{1P}
G_2	dk_{21}	dk_{22}	...	dk_{2j}	...	dk_{2P}
	...	...	...	...	...	...
G_i	dk_{i1}	dk_{i2}	...	dk_{ij}	...	dk_{iP}
	...	...	...	...	...	...
G_n	dk_{n1}	dk_{n2}	...	dk_{nj}	...	dk_{nP}

ここで各パラメータを独立に考え、判断対象物から樹種 G_i へのマハラノビス距離の総和 T_i を次式より求める。

$$T_i = \sum_{j=1}^P \frac{|x_j - m_{ij}|}{\sigma_{ij}} \quad \dots\dots\dots (4-12)$$

$i = 1$ から $i = n$ までの T_i を求め、最も T_i の小さい場合の G_i をその樹種と判断する。

(c) マハラノビスの汎距離の一変量と多変量の比較

マハラノビスの距離の一変量の定義は、式 4.11 で与えられた。これを多変量に拡張する為、以下の対応関係を考える。

表 4. 5 マハラノビス距離の一変量と多変量の比較

	一変量マハラノビス 汎距離	多変量マハラノビス 汎距離
変数	X	$X_1, X_2, \dots, X_p$
平均	m	$m_1, m_2, \dots, m_p$
分散	σ^2	分散・共分散行列 Σ
定義	$dk = \frac{ X - m }{\sigma}$	$D = \frac{(X - M)^T \Sigma^{-1} (X - M)}{p}$

但し、 Σ 、 M 、 X は以下の通りとする。

$$\Sigma = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{12} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1p} & S_{2p} & \dots & S_{pp} \end{bmatrix} \quad M = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{bmatrix} \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

即ち、各パラメータ変数が、変数ベクトルに、分散が分散・共分散行列に対応する。つまり、一変量マハラノビスの汎距離が各パラメータ毎の演算であったのに対し、多変量マハラノビスの汎距離はすべてのパラメータを一度に考えたベクトル演算であることになる。

(d) 多変量のマハラノビスの汎距離

上記の様な一変量マハラノビスの汎距離との対応関係により、判断対象物の各パラメータ値から各樹種の各パラメータへの多変量マハラノビス距離を次式で定義する。

$$D_i^2 = (X - M_i)' \Sigma_i^{-1} (X - M_i) \quad \dots \dots \dots (4-13)$$

但し、 Σ_i 、 M_i 、 X は以下のように定義した。

$$\Sigma_i = \begin{bmatrix} \begin{matrix} (i) & (i) \\ S_{11} & S_{12} \end{matrix} & \begin{matrix} (i) \\ S_{1P} \end{matrix} \\ \begin{matrix} (i) & (i) \\ S_{12} & S_{22} \end{matrix} & \begin{matrix} (i) \\ S_{2P} \end{matrix} \\ \begin{matrix} (i) & (i) \\ S_{1P} & S_{PP} \end{matrix} & \begin{matrix} (i) \\ S_{PP} \end{matrix} \end{bmatrix} \quad M_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{iP} \end{bmatrix} \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_P \end{bmatrix}$$

但し、 X は特徴値群、 Σ_i は樹種 i の分散・共分散行列、 P_i は平均である。パラメータ $P_1 \sim P_P$ 、ここで樹種 $G_1 \sim G_n$ があるとした場合1枚の葉から測定される各パラメータの値を $x_1 \sim x_P$ とする。

$i=1$ から $i=n$ までの D_i^2 を求め、最も D_i^2 の小さい場合の G_i その樹種と判断する。

4.3.3 補正を加えた判定係数の算出

判定係数を求める過程で、補正を行なう。その補正を加えるときの判断部の構造について、以下に述べる。

(a) 特異値をもつ樹種の削除

本システムにより認識を行なう場合、評価用データベースとの比較を行なうことにより樹種の相違を認識するのであるからデータベースの形態によって認識率、処理時間等が大きく変わってくると思われる。

昨年までの結果を調べたところ、誤認をしたケースについてその原因を見てみると、マハラノビスの汎距離の極端に悪いパラメータがある反面、偶然に良い結果を生むパラメータも含まれるため、総合的にみると判定係数が平均化されてしまい誤認する場合が多かった。そこで、判断部の形態として以下の様なモデルを提案する。

(1) 判断構造

ある未知な標本が入ってきた時、誤認識される対象は限られた一部の標本であり、その場合どれかのパラメータにおいて限りなく他の樹種に近い値をとるので誤認される。しかし、その他のあるパラメータにおいて、誤認される樹種と離れている値をとれば、そのパラメータにて分離が可能である。

例えば、図 4.11 のように、全対象群 A は、ある 1 というパラメータで、 B_1 、 B_2 の 2 群に分けられある樹種が対象から削除され、候補が絞られる。そこで、判断構造の高効率化を行ない、誤認識を避ける為、候補樹種の選出を行なう。これは、多変量マハラノビスの汎距離において判定分析を行なう為の前処理としておこなわれる。

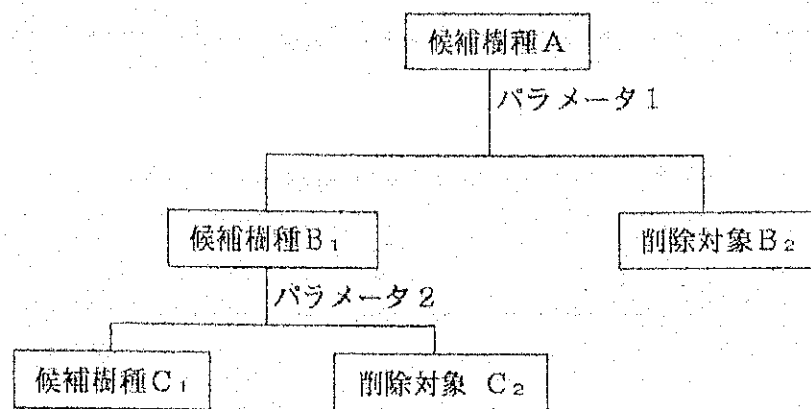


図 4. 1 1 判断構造のツリーの図

(2) 判断方法

未知な標本が、一枚入力された時、もし真の樹種であれば、母集団の分布の $\pm 3\sigma$ の間に存在する確率は 99.97% であるというのが、正規分布の性質から導かれる。したがって、ある樹種と仮定した時、そのなかのあるパラメータにおいて、少なくとも $\pm 3\sigma$ 越えた値を示すパラメータがあれば、あの樹種は対象にはならないとして候補樹種から削除をおこなう。以下にその流れ図を示す。

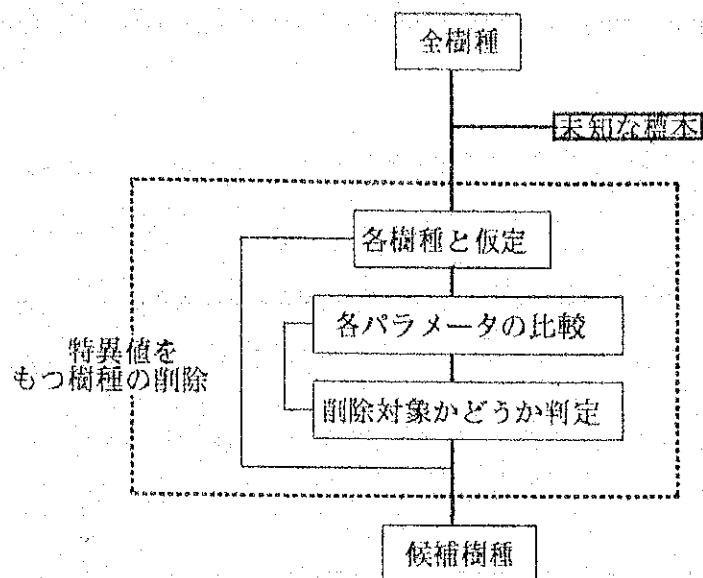


図4. 12 候補樹種の選出の流れ図

この処理では、判定係数に多変量マハラノビスの汎距離を用いた時の計算量の増加に対応し、構造化による処理時間の短縮を図る事と、データベースに格納されていない未知の樹葉が標本として入力された場合に、システムに”その標本については認識不可能”と判断させる事を目的としている。

(b) 樹種の分布の割合による重み付け

ある樹種とある樹種の分離を考えた時、23種のパラメータのなかには分布がかなり接近したものや、離れたものも存在する筈である。分布の接近したものは、マハラノビスの汎距離ではその差をみることができない。しかし、分布の離れたものはその差からどちらに属するか決める事ができる。この様に、マハラノビスの汎距離を用いるとしても、ある2種の判別をするのにその分布の違いでどのパラメータをみるべきかが変わるのも当然である。したがって、パラメータ毎に強調をいかにするかがここでは問題である。

(1) 重みの定義

δ を重みの因子とし、 $\delta(k)_{AB}$ はパラメータ k における樹種 A と樹種 B の分布を重なり具合を考慮した時の重みを表す。定義から、以下のことは自明のこととする。

$$\delta(k)_{AB} = \delta(k)_{BA} \quad \dots\dots\dots (4-14)$$

今、パラメータ k における樹種 A と樹種 B の分布の重なった部分の面積を以下の式で与える。但し、樹種 A と樹種 B は値 z において分布が重なっていて、樹種 A の平均が樹種 B の平均よりも小さいと仮定し、式 (4-2) の確率密度関数 $f(x)$ を用いて $\phi(z)$ を表すものとする。

$$\phi(z) = \int_z^{\infty} f_A(x) dx + \int_{-\infty}^z f_B(x) dx \quad \dots\dots\dots (4-15)$$

以下に図 4.13 でその関係を示す。

確率密度 $\phi(z)$ は、ある正規分布において、変数 z の位置が示す確率密度を与える事が分かる。よって $\phi(z)$ は、その値を標準正規化した時、その定義を $0 \leq \phi(z) \leq 1$ であたえる。

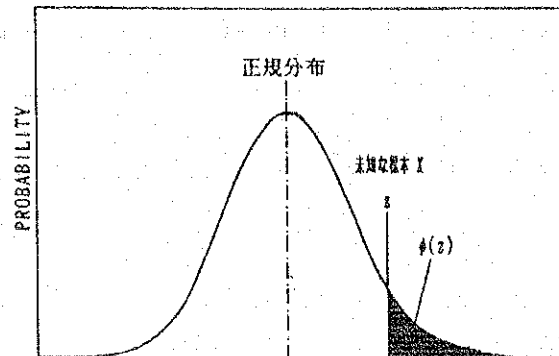


図 4. 1 3 $\phi(z)$ の定義

その確率密度 $\phi(z)$ を用いて、重みの因子を以下の様に定義した。

$$\delta = 2 \exp [\phi(z) / (\sigma_A + \sigma_B)] \quad \dots\dots\dots (4-16)$$

これにより、重み因子は、その定義を $1 \leq \phi(z) \leq 2$ とする。分布の中心に近いほど大きくなり、離れるにしたがって、1 に近くなる。この重み因子から以下の重み係数を定義した。パラメータ k の樹種 i の重み係数を $\varepsilon(k)_i$ とした。

$$\varepsilon(k)_i = \frac{1}{n} \sqrt{\sum_{j=1}^n \delta(k)_{nj}} \quad \text{但し、} j \neq n \quad \dots\dots\dots (4-17)$$

$$= \sqrt{(\delta(k)_{n1} + \delta(k)_{n(n-1)} + \delta(k)_{n(n+1)} + \delta(k)_{np})} \quad \dots\dots\dots (4-18)$$

4.3.4 認識率の定義

本研究では判断の能力の基準として、認識率を定義する。認識率とは、未知な標本をもってきたとき、その標本を正しく認識した割合の事で以下の式で定義する。

$$R = \frac{C}{E} \quad \dots\dots\dots (4-19)$$

R : 認識率

E : 総試行回数

C : 正しく判断した回数

認識率の収束条件

本研究では、認識率が一定値に収束したと判断する基準として、認識率の収束条件式を以下の様に定めた。今、総試行回数をE回とした時に、得られた認識率を 認識率[E] という記号で表す事にする。収束条件式は、過去3回の認識率の推移で表される。

$$\text{収束1} = (\text{認識率}[E] - \text{認識率}[E-1])/100$$

$$\text{収束2} = (\text{認識率}[E] - \text{認識率}[E-2])/100$$

$$\text{収束3} = (\text{認識率}[E-1] - \text{認識率}[E-2])/100$$

この収束条件式3つがそれぞれ1%を越えない時、認識率が収束したとしてその時の値を認識率とする。以下にその流れ図を示す。

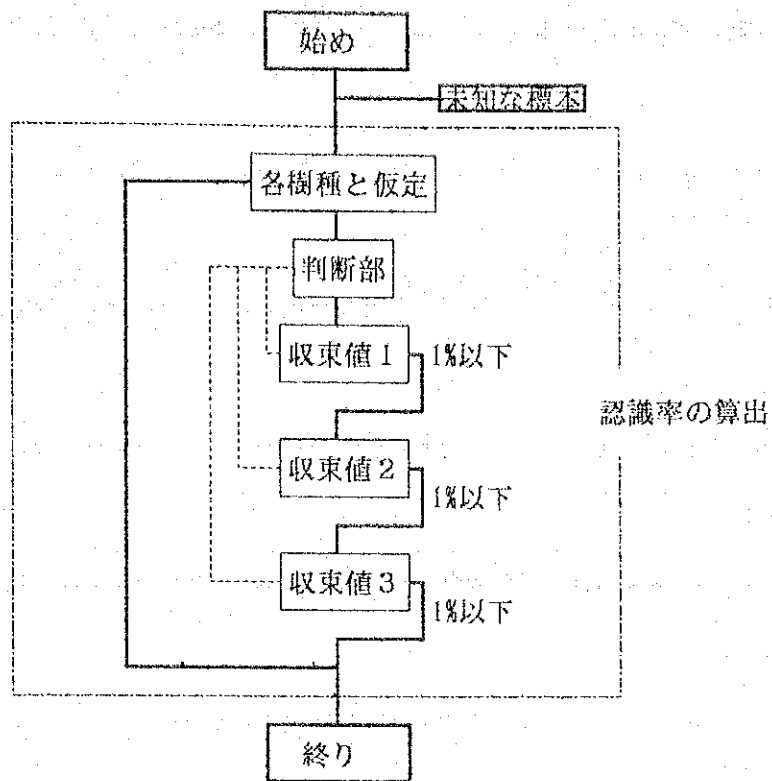


図4. 1 4 認識率の収束条件の流れ図

従って、装置の能力を求める為に、認識率を求める過程で、本研究では、収束条件式3つが1%以下に収束するまで無作為に標本の抽出をし、試行を繰り返すものとする。

5. 装置性能確認試験の方法

5. 1 検索対象樹種

本研究で用いられた樹種の葉は、インドネシア・東カリマンタンに植生しているフタバガキ科及びそれに近い樹種をサンプルとして用いている。

尚、これらの樹種は、専門家でも樹葉のみで判定することは困難であるといわれており、樹種名の鑑定については、インドネシアのボゴール国立植物園に依頼した。

樹種数は 50 種とする。次頁にその試料の各データを記す。

5. 2 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、インドネシアで採取された 50 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。

5. 2. 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。

No.	Species Name	Family Name	Diameter (cm)	Height (m)	Q'ty (pcs.)	Date (1990)	Place *	Remarks
1	Doryobalanops Beccarii Dyer	Dipterocarpaceae	15	17	200	Aug 20	L K	
2	Shorea Ovalis (Korth.) Bl.	↑	25	29	250	↑	↑	
3	Shorea Sp.	↑	14	16	250	↑	↑	
4-1	Shorea Polyandra Ashton	↑	10.5	15	150	Aug 21	↑	
4-2	↑	↑	9.8	14	130	↑	↑	
5-1	Shorea Leprosura Miq.	↑	11	18	100	↑	↑	
5-2	↑	↑	13	19	175	Aug 22	↑	
7	Shorea sp.	↑	41	37	300	Aug 23	↑	
8	Beilschmiedia pulverulenta Kostum	Lauraceae	53	28	275	↑	↑	
9-1	Shorea Pauciflora King	Dipterocarpaceae	13	20	125	Aug 24	↑	
9-2	↑	↑	21	24	150	↑	↑	
13	Shorea Laevis Ridl.	↑	11	9	335	↑	B S	
14	Shorea Seminis (de Vriese) Sloat.	Dipterocarpaceae	21.5	14	250	Aug 28	↑	
15	Dryobalanops Beccarii Dyer	↑	17.2	13	250	↑	↑	

* L K : Lempake Forest, Samarinda, East Kalimantan
B S : Bukit Soeharto Forest, Samarinda, East Kalimantan

No.	Species Name	Family Name	Diameter (cm)	Height (m)	Q'ty (pcs.)	Date (1990)	Place ‡	Remarks
16	Cotylelobium Lanceolatum Craib	Dipterocarpaceae	9.5	9	250	†	B S	
17	Dipterocarpus Caudiferus Merr.	†	9.2	7	250	Aug 29	†	
18	Shorea Gibbosa Brandis	†	11	10	250	Aug 30	†	
19	shorea sp.	†	10.5	9	250	†	†	
21	Shorea Smithiana Sym.	†	25	12	250	†	†	
23	Payena Acuminata (Bl.) Pierre	Sapotaceae	6.2	9.5	250	†	†	
24	Sindora Leiocarpa Backer ex de Wit	Fabaceae	-	-	250	†	†	
25	Baccaurea Kunstleri King er Gage	Euphorbiaceae	12	19.5	250	†	†	
26	Cryptocarya sp.	Lauraceae	14.5	10	250	†	†	
27	Alstonia Pneumatophora den Berger	Apocynaceae	16	12	250	†	†	
28	Castanopsis Motleyana King	Fagaceae	7.8	12	250	Sep 1	†	
29	Sandoricum Borneense Miq.	Meliaceae	16.6	13	250	†	†	
30	Palaequium Rostratum (Miq.) Burck	Sapotaceae	9.6	13	250	†	†	
31	Durio of. Macrophyllus King Ridey	Bombacaceae	36	20	250	†	†	

* B S : Bukit Soeharto Forest, Samarinda, East Kalimantan

No.	Species Name	Family Name	Diameter (cm)	Height (m)	Q'ty (pcs.)	Date (1990)	Place *	Remarks
32	Dyera Costulata (Miq.) Hook.f.	Apocynaceae	43.2	22	250	↑	B S	
33	Archidendron of Bubalinum Nielsen	Fabaceae	26.2	26	250	↑	↑	
34	Schima Wallichii (DC.) Korth	Theaceae	21	15	250	Sep 1	↑	
36	Litsea Firma (Bl.) Hook.f.	Lauraceae	8.9	12	250	↑	↑	
37	Podocarpus Bulmei Enidl	Podocarpaceae	22.3	26	250	↑	↑	
38	Anisoptera Grossivenia Sloot.	Dipterocarpaceae	23	26.7	200	Sep 3	↑	
39	Diospyros Borneensis Hirn	Ebenaceae	12.8	15	250	↑	↑	
40	Dipterocarpus gracilis Bl.	Dipterocarpaceae	18.6	18	250	Sep 5	↑	
41	Cratxylum Arborescens (Vahl) Bl.	Hypericaceae	6.8	12	150	↑	↑	
42	Artocarpus Dadah Miq.	Moraceae	14.5	16.8	250	↑	↑	
43	Adidandra Borneensis Kobuski	Theaceae	12	11	200	Sep 6	↑	
45	Tristania Whitiana Griff	Myrtaceae					↑	
46	Aquilaria Malaccensis Lamk	Thymelacaceae					↑	
47	Dacryodes rostrata (Bl.) H.J.L	Burseraceae					↑	

* B S : Bukit Soeharto Forest, Samarinda, East Kalimantan

No.	Species Name	Family Name	Diameter (cm)	Height (m)	Q'ty (pcs.)	Date (1990)	Place *	Remarks
48	Carallia Brachiata (Lour.) Merr	Rhizophoraceae					B S	
49	Rhodamnia Cinerea Jack	Myrtaceae					↑	
50	Artocarpus cf Nitidus Trec	Moraceae					↑	
51	Drypetes sp.	Euphorbiaceae					↑	
52	Gluta Renghas Linn	Anacardiaceae					↑	
53	Syzygium sp.	Myrtaceae					↑	
54	Payena Acuminata (Bl.) Pierre	Sapotaceae					↑	
55	Castanopsis Acuminatissima (Bl.) A. DC	Fagaceae					↑	
56	Drimycarpus Luridus (Hook.f) Ding Hou	Anacardiaceae					↑	
57	Meiogyne Virgata (Bl.) Miq.	Annonaceae					↑	
58	Koompasia Malaccensis Maing ex Banth	Fabaceae					↑	

* B S : Bukit Soeharto Forest, Samarinda, East Kalimantan

5.2.2 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。

(a) 判定係数の算出にマハラノビスの汎距離を用いた時

(1) 一変量マハラノビスの汎距離を用いた判別

(2) 多変量マハラノビスの汎距離を用いた判別

(b) 判定係数の算出に補正を加えた時

(1) 特徴値が特異値となる樹種を削除をする補正を付加した時の判別

(2) 樹種間の分布に応じた重みづけを行った時の判別

6. 装置性能確認試験の結果

6. 1 判定係数の算出法とその有効性

判定係数を求める際の各判別式の違いを考察する。

6. 1. 1 本システムによる判定結果

本研究で用いられた判別式は、過去の研究における一変量マハラノビス距離を拡張した多変量マハラノビスの汎距離である。図 6.1 は、多変量マハラノビスの汎距離を用いたときの判別結果である。

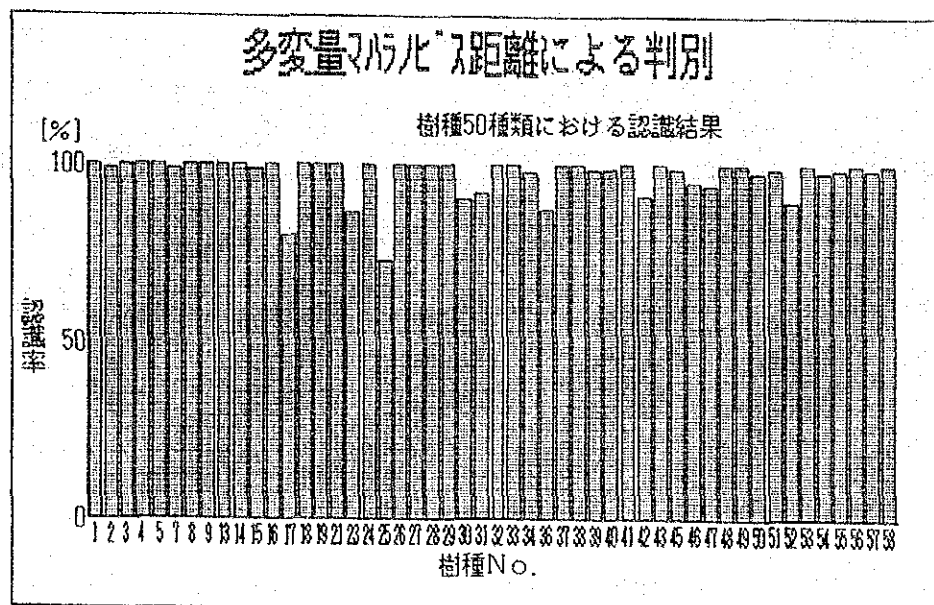


図 6. 1 認識率結果

多変量マハラノビスの汎距離による判別式の有効性を評価する為、以下のように考察を行なった。

6. 1. 2 判定係数の算出法と認識率の関係

昨年の結果を評価するため、判定係数の算出を昨年同様、一変量マハラノビスの汎距離を用いた場合と、多変量マハラノビスの汎距離を用いた場合の違いについて検証を行なった。

(a) 10種の母集団による認識率

樹種10種の認識結果を以下に示す。それぞれ多変量マハラノビスの汎距離と一変量マハラノビスの汎距離を用いた時の認識率結果を示す。但し、ここでは未知な標本として10枚の樹葉の平均を入力している。

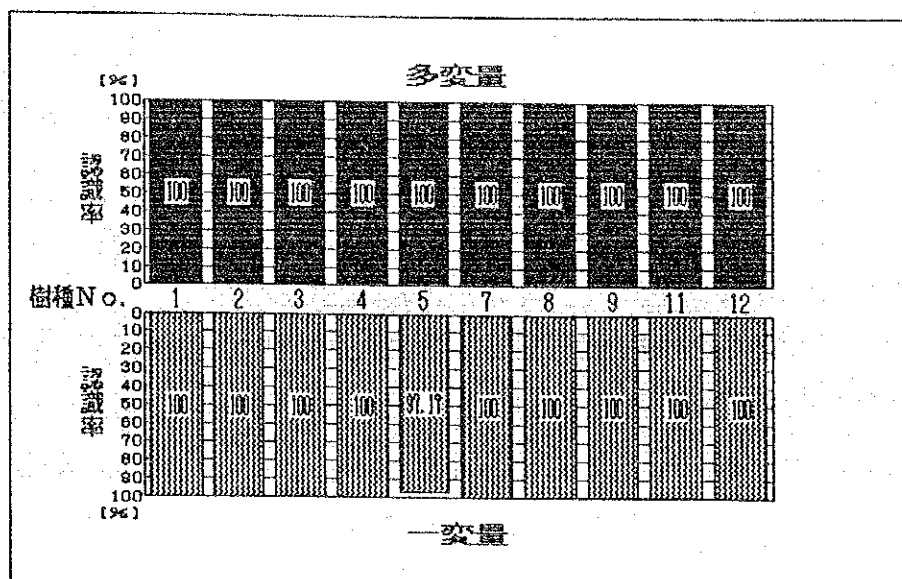


図6.2 認識率結果(10種中)

一変量マハラノビスの汎距離と多変量のそれとの認識率の差を対象物10種にて評価するには、どちらも認識能力が対象10種に限った場合では十分であるので、判別式の有効性の評価をすることはできなかった。

(b) 50種の母集団を用いた場合

樹種数を最大の50種にしたときの一変量、及び多変量マハラノビスの汎距離を用いた認識結果について以下に示す。

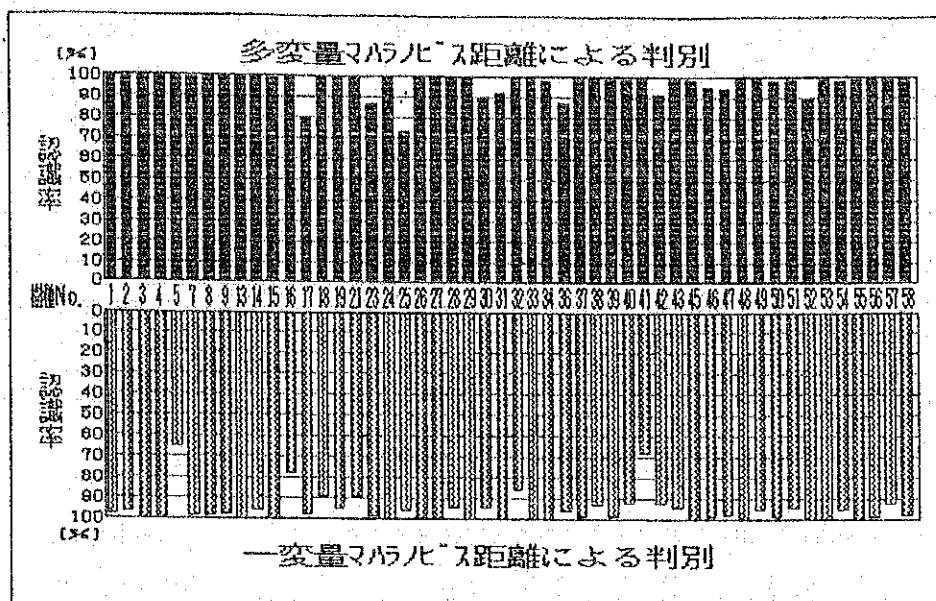


図6.3 認識率結果

多変量マハラノビスの汎距離による判別結果は、一変量のマハラノビスの汎距離による判別結果よりも、全般的に向上していることが分かった。但し、多変量における結果においても、樹種25のように極端に認識率の劣る樹種も存在する。そこで、平均認識率を求めたのが以下の表6.1である。それによれば、多変量による判別結果のほうが、良好であることが確認された。

これにより、対象となる樹種が増えた時は、判定係数に多変量マハラノビスの汎距離を用いた方が、良い結果が得られる事が分った。

表6.1 平均認識率

多変量	一変量
97.3%	94.0%

6.2 特異値をもつ樹種の削除

多変量マハラノビスの汎距離を用いて判別を行なわせる前に、特異値をもつ樹種の削除をすることで、対象となる樹種とそうでないものとに分離した上で、判別を行なわせる。ここでは、その処理をするかしないかによって、どれだけ認識率が異なるかを実験した時の結果をのべる。標本とした樹種は無作為に選んだ5種とする。

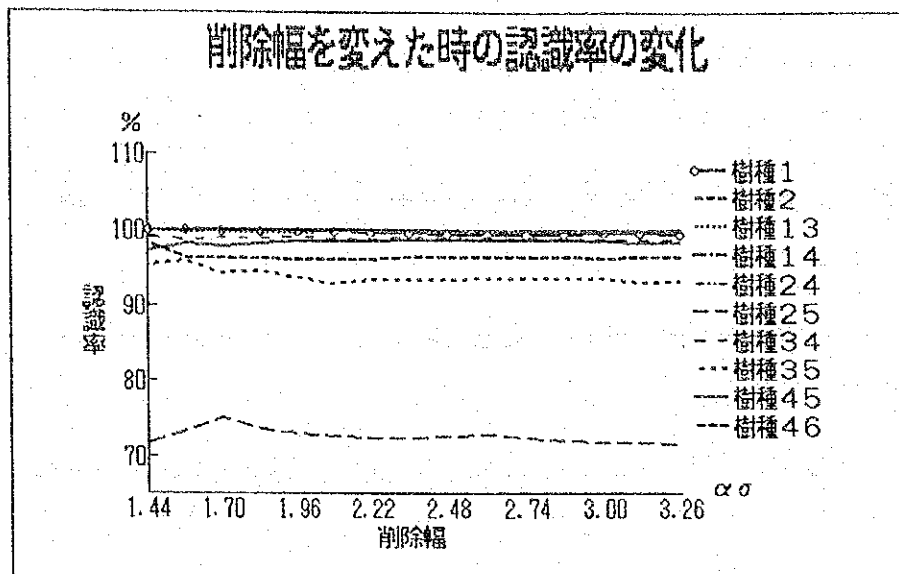


図 6.4 削除幅を変えた時の認識率変化

特異値をもつ樹種の削除というのは、全対象のなかから今問題にしている対象を判別するために、各パラメータにおいて、予めある幅をしきい値として、それを越えるものに関しては対象から削除する過程をいう。図 6.4 は、そのしきい値を変化させた時の認識率の推移を表わし、しきい値を大きくするに従い、削除される対象は減少し、この処理は効力を失う。上記の実験結果によれば、予め対象を絞り込むことで候補を選出し判別を行なう時と、すべての対象から分離を行なう時の認識率にそれほどの差をみることができなかった。これより、判断部を構造化したときに、削除される樹种群には対象とする樹種が含まれることが無く、全対象を対象としたときに得られる結果と一致することが分かる。

6.3 樹種間の分布の度合いによる重みづけ

表 6.3 は判断構造にこの重み付けの処理を取り入れるか否かによって、どれだけ認識率が異なるかを実験した時の結果をのべる。データとして用いた樹種は無作為に選んだ8種とする。上記の4種は、重み付けをすることで認識率が向上した場合で、下記の4種は、重み付けをすることで、認識率が悪くなった場合である。

表6.3 重み付けによる認識率結果

樹種 NO	認 識 率	
	重み付けあり	重み付けなし
25	93.40 %	72.90 %
31	100.0 %	92.45 %
47	98.11 %	94.34 %
52	92.45 %	89.62 %
17	52.83 %	80.19 %
23	80.19 %	86.92 %
27	63.21 %	100.0 %
40	82.08 %	99.06 %

重み付けにより、認識率の向上が認められたケースもあるが、逆に認識率の悪くなる方へ重み付けをされたものもあった。以下にその原因について考察する。

(1) 重み付けの大きさはその定義から 1 と 2 の間の値をとる。一変量では、各パラメータごとのマハラノビスの汎距離の大きさは、値として 3 前後の大きさをもつ。多変量の場合であっても、その行列因子にかけられた重みの係数は、マハラノビスの汎距離が示す値と大きさをほぼ同じとしなければいけない。例えば、もし、その大きさがあまりにも大きければ、マハラノビスの汎距離で得られた値の差にはほとんど関係することなく、その重みの因子の大きさで候補樹種の選出が行なわれることになる。

(2) 重みの係数の与え方は、例えば、樹種 A のあるパラメータにおける重みであれば、その樹種 A と他の樹種との分布の重なり具合が与える。つまり、分布が重なっているパラメータに対しては、分布が重なっていないパラメータとその評価の大きさを変えることを目的としている。

この場合に、以下の例を考える。樹種 A、B という 2 群の分離を行なう時、パラメータ 1 で分布が離れていて、パラメータ 2 で分布が接近しているとす。その時、パラメータ 1 において樹種 A の分布に近い値が与えられた時は、その重みの定義から、判別式においてその大きさを小さくする方向に重みが働く。また、パラメータ 2 において、樹種 A からやや遠い値が与えられた時、その値を今度は小さくする方向に重みが働く。つまり、この例において結果的には、判別能力を弱くしているといえるだろう。単に、分布の重なりから重みを計算すると、その結果、特徴量と分布との位置関係によっては判別能力を落とす結果になると推測する。

以上に述べたように、重み付けについては、性能を向上させる可能性は見出したものの、現状では、最適な条件の定義にまでは至っていない。

7. システムとしての性能の検証

判定係数算出の過程で用いられた多変量マハラノビス距離の判別式による判別結果は、ほぼ全般的に認識率の向上を図れたものの一部の樹種について過去の研究による一変量マハラノビス距離による判別式のほうが良好であるという結果を得た。この因果関係については、検証を行っていないが、性能試験として確認された認識率結果に基づきどの樹種がどちらの判別式による分離が良好であるか明らかにされている。

また、前項までの結果から、ある想定した樹種を未知な標本と仮定した時の認識率は算出され、装置としての性能は確認された。しかし、全く未知な標本を判定するシステムとしての検証はまだである。ここでは、システムとしての信頼性を定義し、その判断構造を決定する。

1. 認識率

認識率とは、ある未知な標本（被判定物）を入力し、その標本をある樹種と判定した時のその樹種への可能性の割合を表わすものである。即ち、誤認識の割合を表わす。ここで、誤認識とは次の2種が存在する。

ある樹種Aからみて誤認識とは

- ①樹種Aにおいて真の標本を入力した時、違（Aでない）と認識された時
- ②違（Aでない）の樹種において、違の標本を入力した時、Aと認識する時の2種存在する。

認識率とは、上記の①における誤認識の度合いを示すものであり、装置の判別能力を示す指標である。しかし、これはこのシステムにおいて、被判定物を判定した時のその真の可能性を示すものではない。それは、②の誤認識を考慮にいれていないからである。

2. 判定率

ここで言う判定率は、被判定物の判定結果におけるその樹種への可能性を表わす。つまり、2種類の誤認識の原因を考慮したものであり、ある未知な標本がある樹種と認識された時のその確からしさの可能性を表わす。

3. 判断構造

こうして得られた判定率の高い判別式の方を優先させることにより、より良い判断構造を以下の様に提案する。

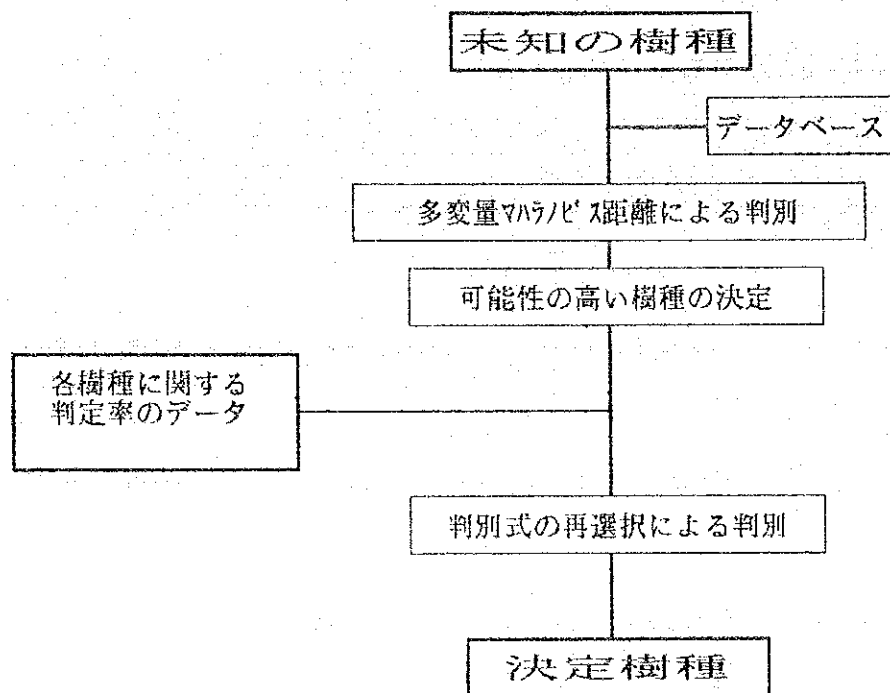


図 7. 1 樹種の選出の流れ図

このようにして、巧に多変量と一変量マハラノビス距離による判別式をそれぞれ分離能力の良い樹種に用いることにより識別能力の向上を行うことができると想像する。本研究では、すべての樹種において、事前に判別能力として信頼度を算出することで、以後の未知な樹種の標本に対する判別能力を格段と向上させる事が可能である。これにより、各樹種ごとに良好な判別式の方を用いることで、以下の様な検証結果を得た。

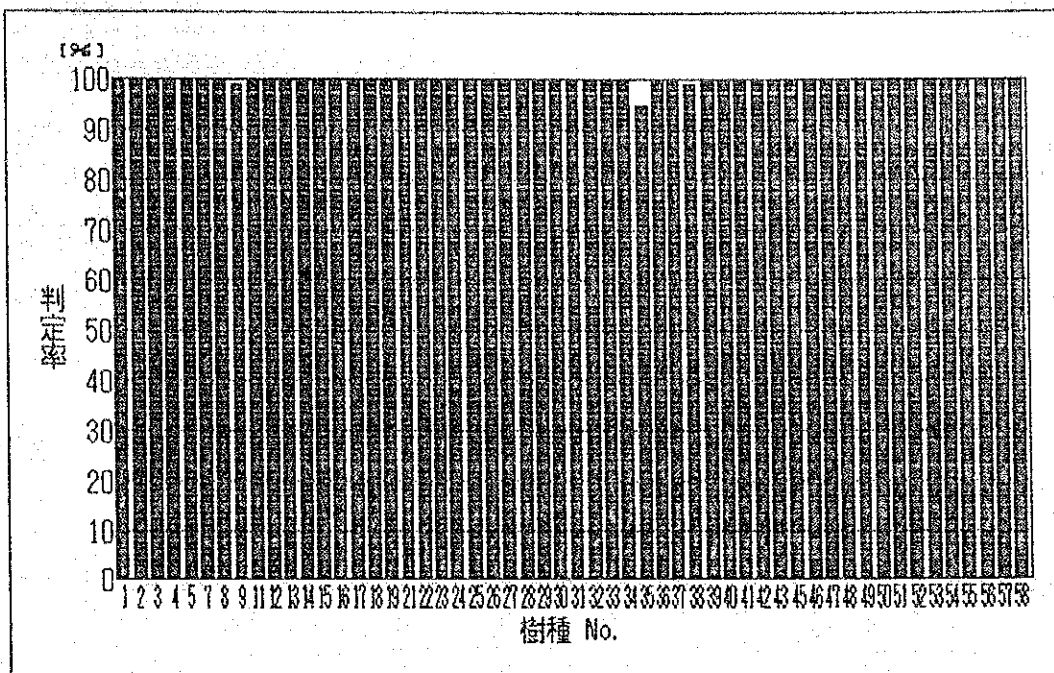


図7.2 システムとしての判別能力

8. 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、専門家によりインドネシアで採取されたおよそ50種の樹種を用いて、本システムの性能の確認を行なった結果、最も判定率の悪い樹種においても約95%で判定を行うことができた。したがって、本システムは、樹葉の外形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、実用可能である事が確認できた。

これらの試験を通して、当初の目標である性能を満たすシステムを構築することができた。

APPENDIX

1. 記号表、用語表

2. パラメータ23種の各分布表

記号表

L	:	葉の長さ	(mm)
W	:	葉の幅	(mm)
S	:	葉の面積	(mm ²)
l	:	周囲長	(mm)
S_v	:	包絡面積	(mm ²)
Ch	:	頂点数	(個)
L_g	:	葉柄からの重心長	(mm)
C_v	:	包絡点数	(個)
l_{max}	:	葉の重心からの最大長さ	(mm)
l_{min}	:	葉の重心からの最小長さ	(mm)
D_i	:	樹種 i のマハラノビスの汎距離 (多変量)	
dk_i	:	樹種 i のマハラノビスの汎距離 (一変量)	
M_i	:	樹種 i のパラメータの平均値群	
m_{ij}	:	樹種 i のパラメータ j の平均値	
X	:	未知な葉の特徴値群	
x_i	:	未知な葉のパラメータ i の特徴値	
σ_{ij}	:	樹種 i のパラメータ j の標準偏差	
Σ_i	:	樹種 i の分散・共分散行列	
$S_{kj}^{(i)}$	:	樹種 i のパラメータ k とパラメータ j の分散・共分散因子	
$\mu_{j,m}^{(i)}$	:	樹種 i のパラメータ j における標本 m 番目の値	
$\delta(k)_{ij}$	:	樹種 i のパラメータ k における重みを考えた時のパラメータ j との重みの因子	
$\varepsilon(k)_i$	:	樹種 i のパラメータ k における重み係数	
$f(x)$	:	確率密度関数	
$\phi(z)$	:	確率密度のこと、点 z における存在確率を表し、 $f(x)$ の積分で与えられる	
R	:	認識率	
E	:	総試行回数	
C	:	正しく判断された回数	
p	:	総パラメータ数	
n	:	総樹種数	
N	:	標本数	
i	:	樹種に用いる添え字	
j, k	:	パラメータに用いる添え字	
m	:	標本に用いる添え字	
g	:	画素に用いる添え字	

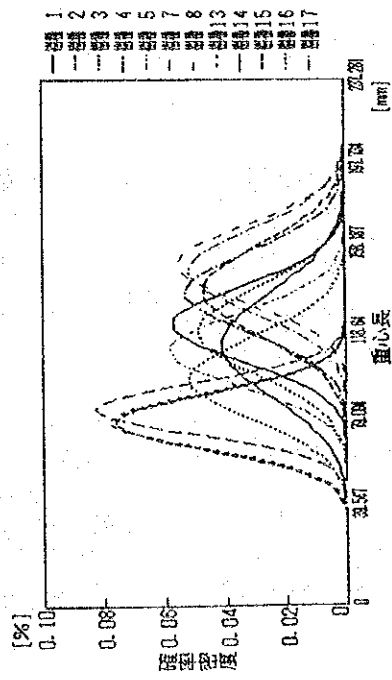
用語定義表

特徴分布	標本から各特徴値を算出した時に得られる分布のこと。特徴分布は、 x 軸に各特徴値の大きさを、 y 軸には、その値の頻度をとるものとする。
母集団	ある樹種の樹葉全体のこと。但し、過去に於ける結果報告から、約 200枚にて、近似的に母集団を表現する事ができると報告されていることから、本研究では、採取してきた約200枚の樹葉をもちいて母集団と定義した。
パラメータ	各樹種を他と分離する為に定義した特徴。
判定係数	標本の特徴値がどの位の確率で母集団に属するかを示す指標となる値。
削除幅	ある標本を判定する際、判定係数を計算する前に、特徴値について定めた限界の値の範囲。
特異値	標本から得られる各特徴値のうち、比較する特徴分布の削除幅を越えた値。
標本	システムが対象とする各サンプル。

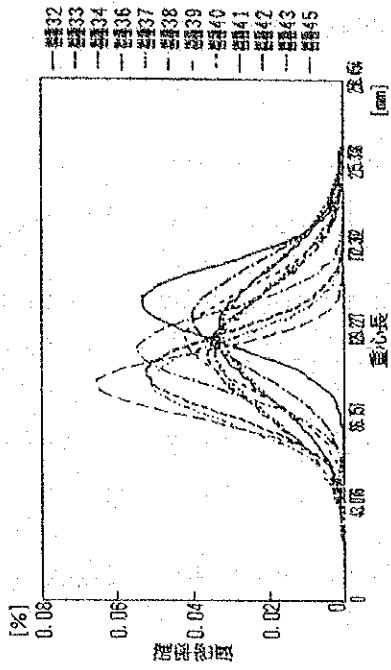
画像処理用語表

画素	カメラに内蔵されている、縦・横 512個の方眼状の目。その一つを画素といい、1 画像は合計で約26万画素に256階調の輝度で収納される。
二値化	対象とする図形の各画素をある閾値を境に指定濃度に変換する操作。様々な画像処理を行なう為の前処理や、情報量圧縮などの目的にも用いられる。
背景画素	対象とする図形以外に、その背景を構成している画素。
境界画素	背景画素から8近傍以内にある背景画素以外の画素の事で対象となる図形の外郭をなす画素。
8近傍	ある画素を処理の対象とした時、その画素を囲む周囲の8つの画素。
細線化	図形の外形情報のみを必要としたり、画素のつながりを必要とする時行なう処理で、8近傍内の画素のつながりを一画素とする処理。
包絡点	図形を囲む最小凸多角形を考えた時、その最小凸多角形を構成する各点。

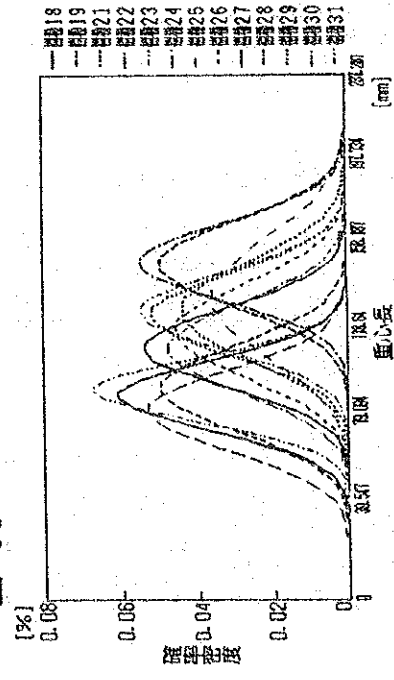
重心長パラメータにおける確率密度の分布



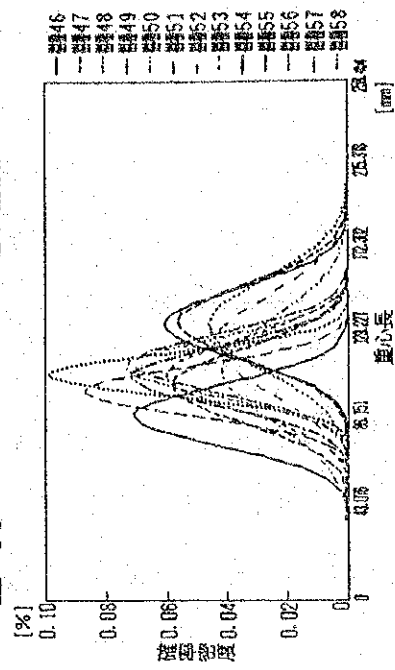
重心長パラメータにおける確率密度の分布



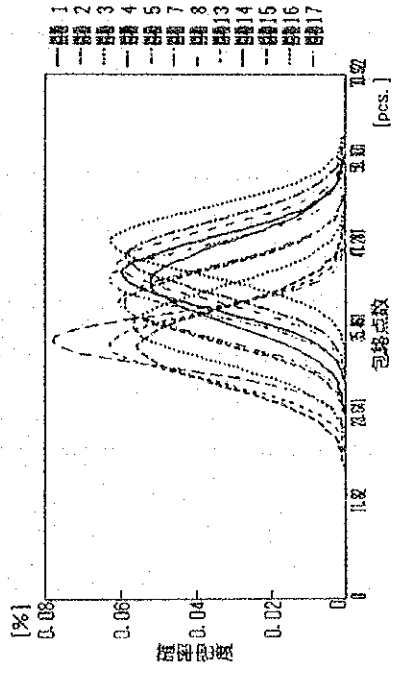
重心長パラメータにおける確率密度の分布



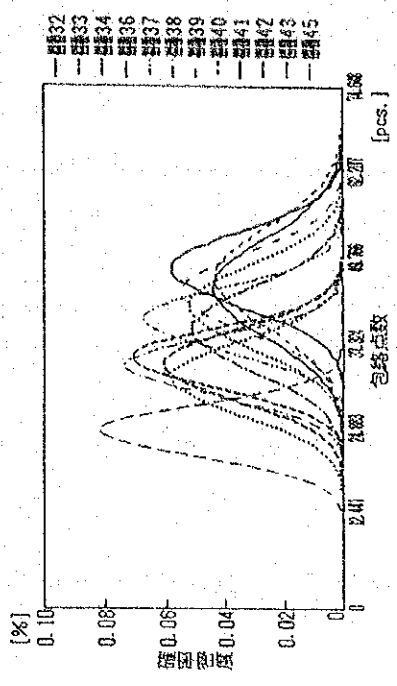
重心長パラメータにおける確率密度の分布



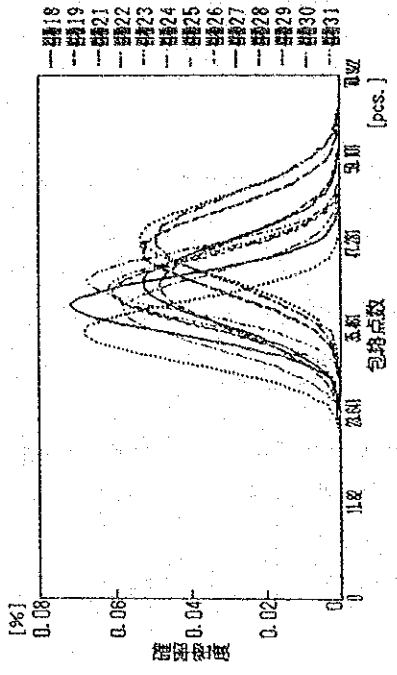
包絡点数パラメータにおける確率密度



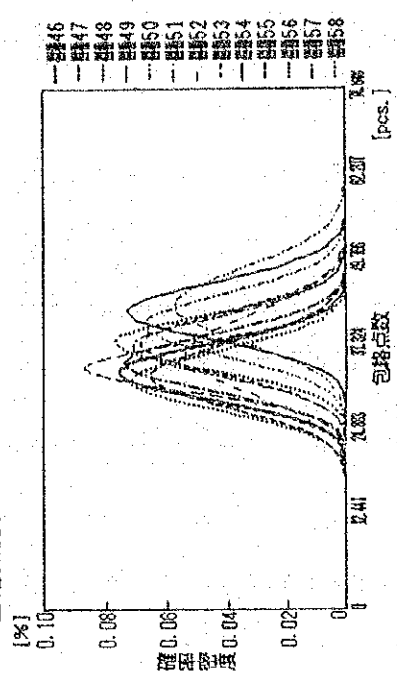
包絡点数パラメータにおける確率密度の分布



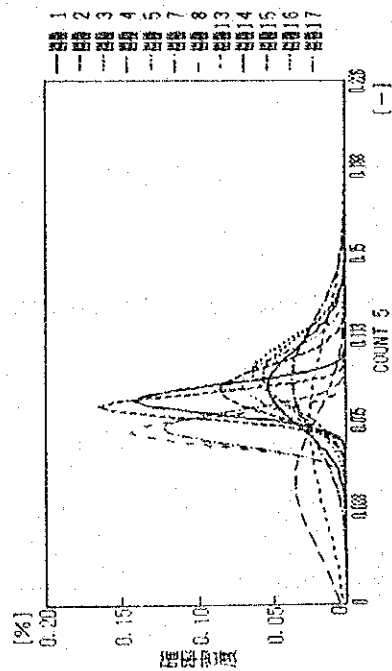
包絡点数パラメータにおける確率密度



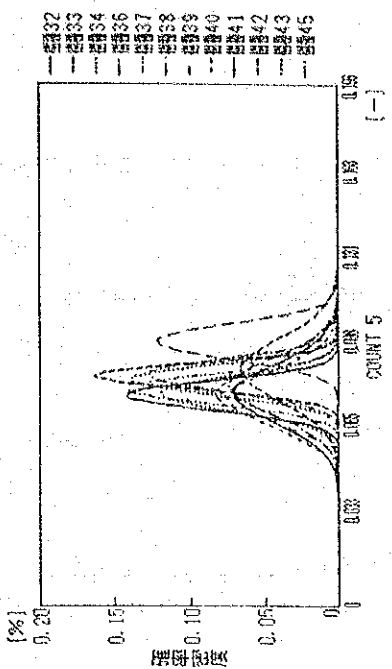
包絡点数パラメータにおける確率密度の分布



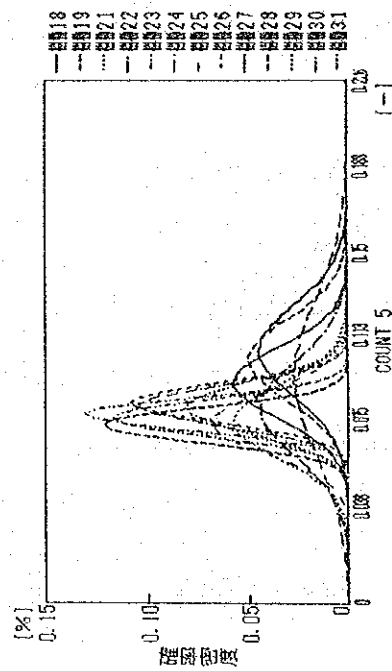
COUNT 5パラメータにおける確率密度の分布



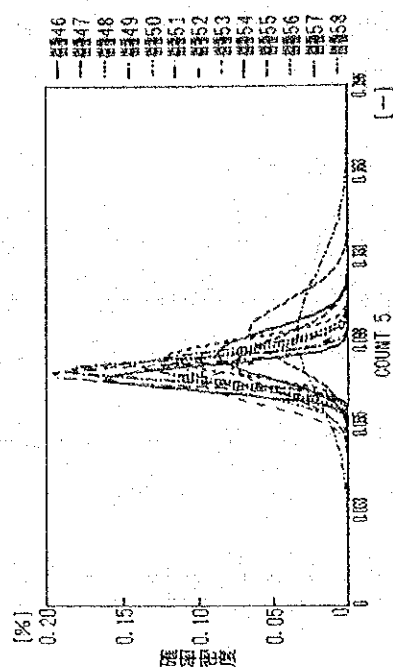
COUNT 5パラメータにおける確率密度の分布



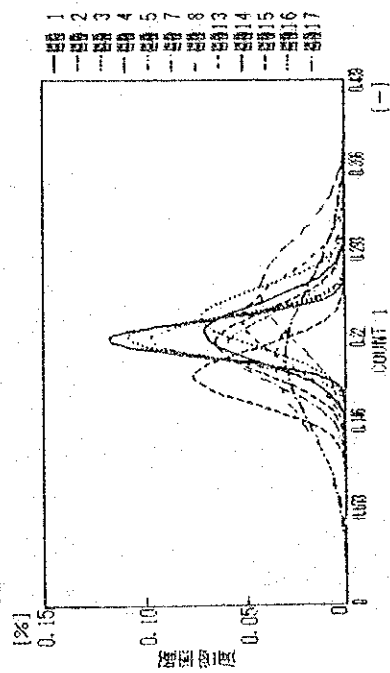
COUNT 5パラメータにおける確率密度の分布



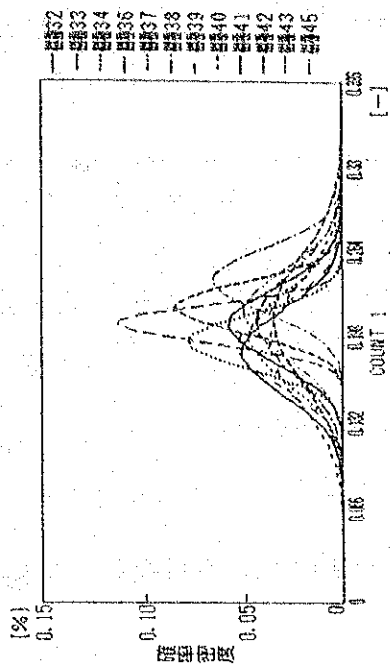
COUNT 5パラメータにおける確率密度の分布



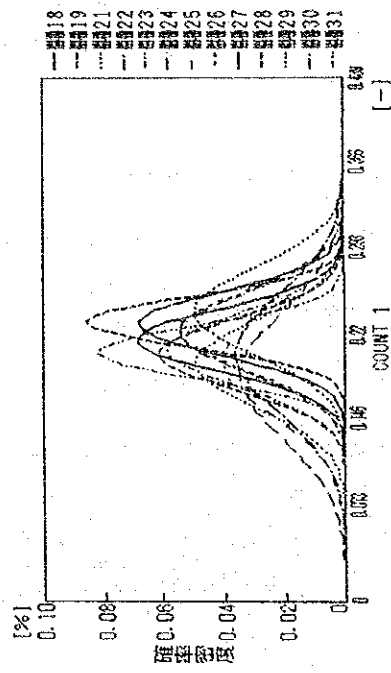
COUNT 1パラメータにおける確率密度の分布



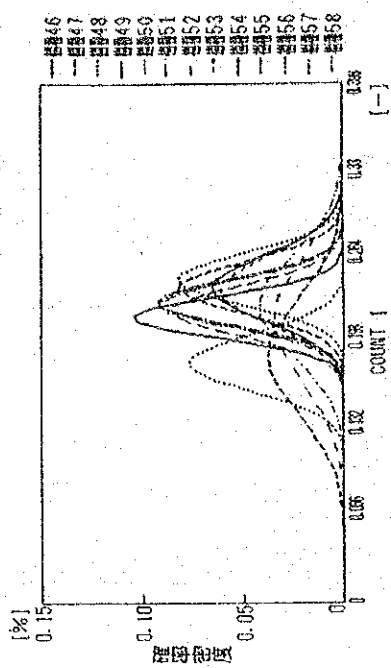
COUNT 1パラメータにおける確率密度の分布



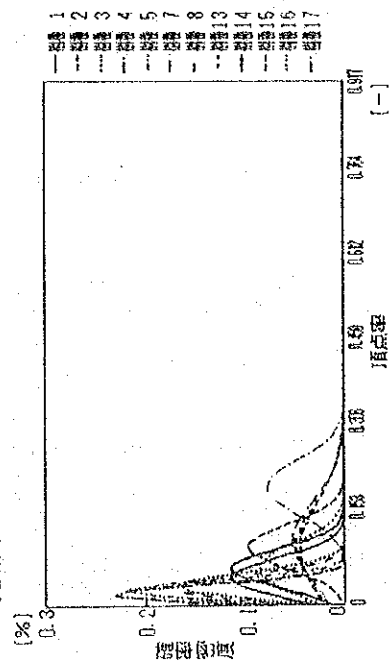
COUNT 1パラメータにおける確率密度の分布



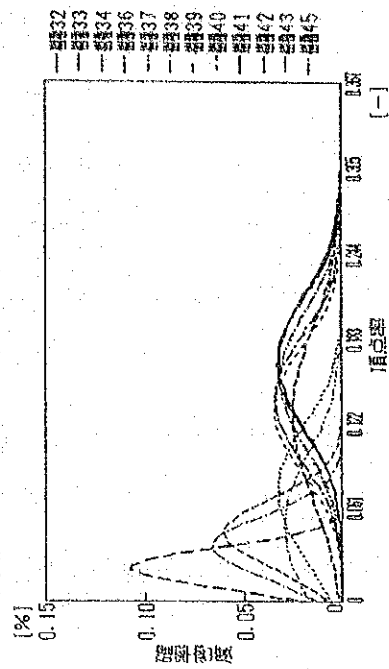
COUNT 1パラメータにおける確率密度の分布



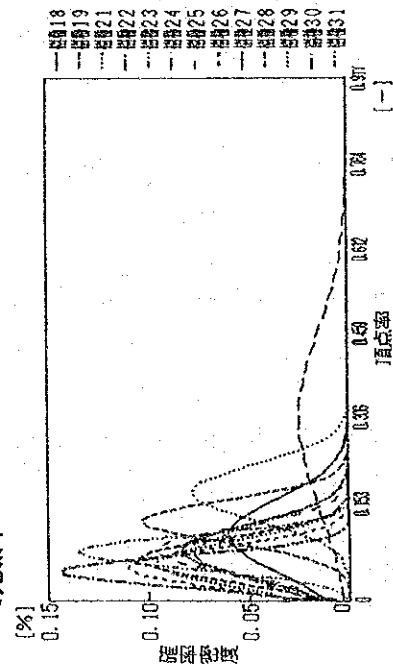
頂点率パラメータにおける確率密度の分布



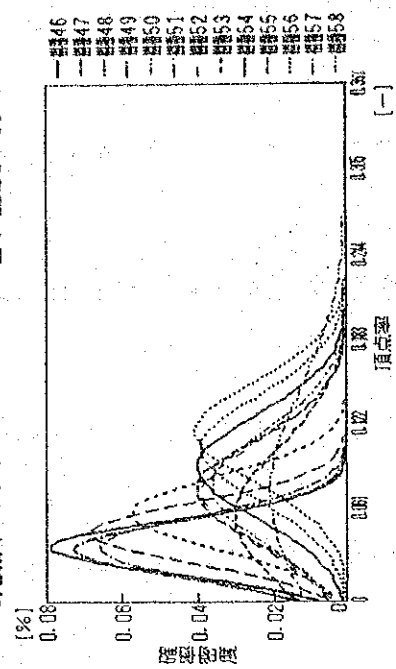
頂点率パラメータにおける確率密度の分布



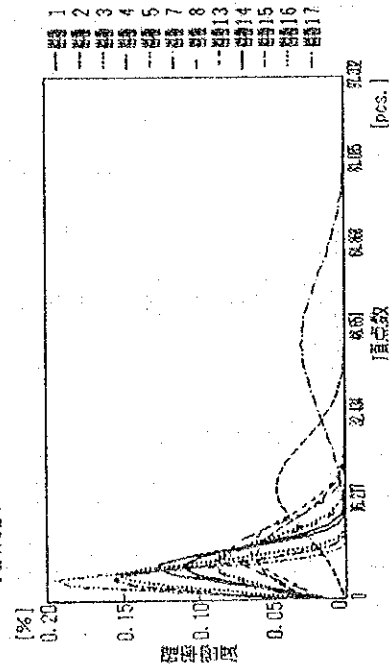
頂点率パラメータにおける確率密度の分布



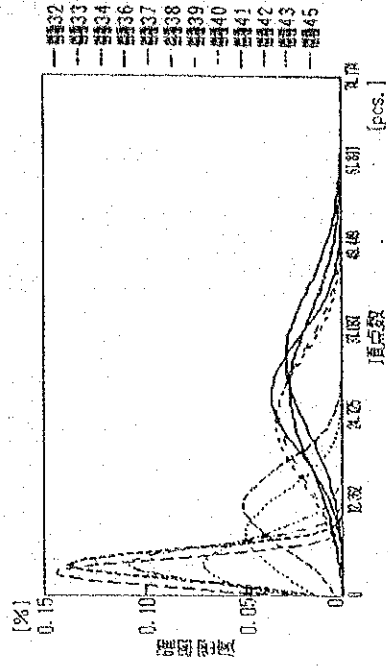
頂点率パラメータにおける確率密度の分布



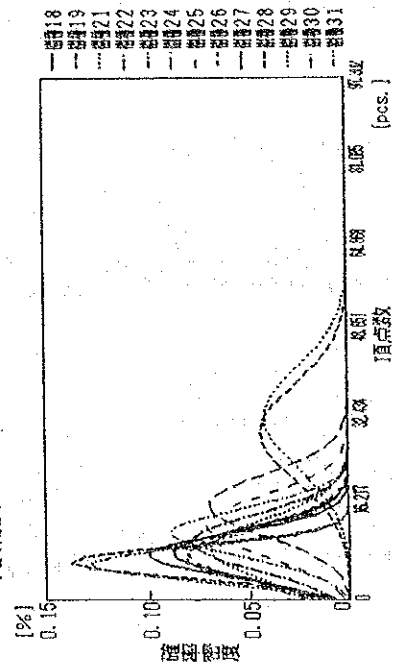
頂点数パラメータにおける各樹種の分布



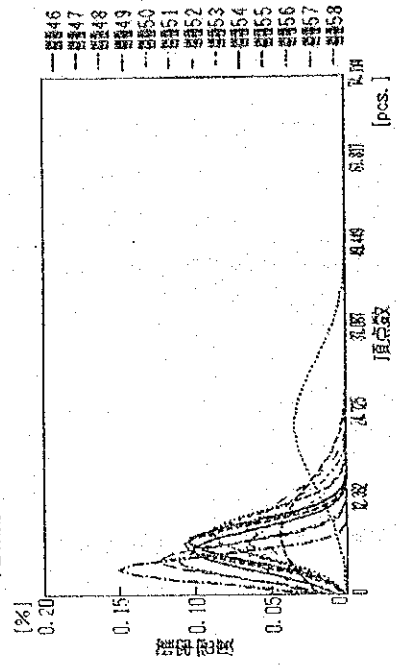
頂点数パラメータにおける各樹種の分布



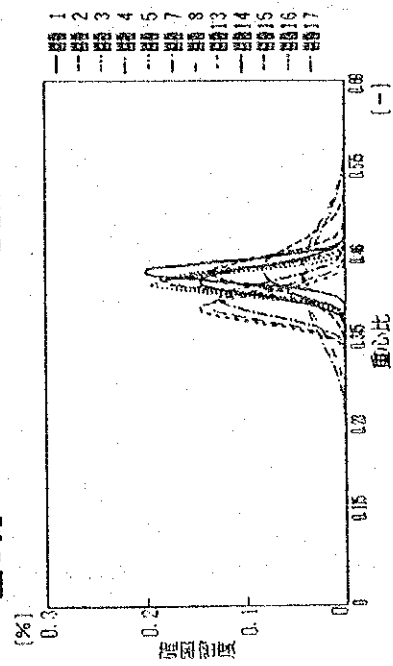
頂点数パラメータにおける各樹種の分布



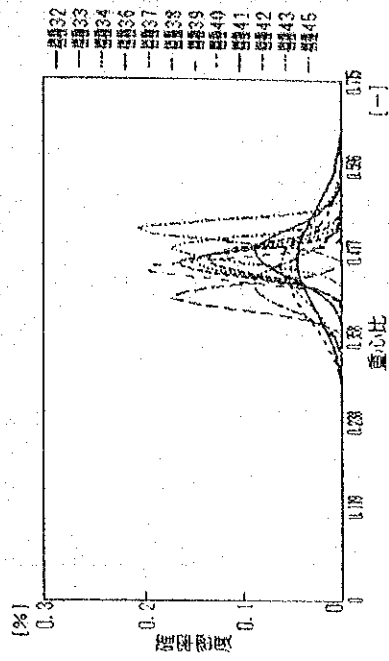
頂点数パラメータにおける各樹種の分布



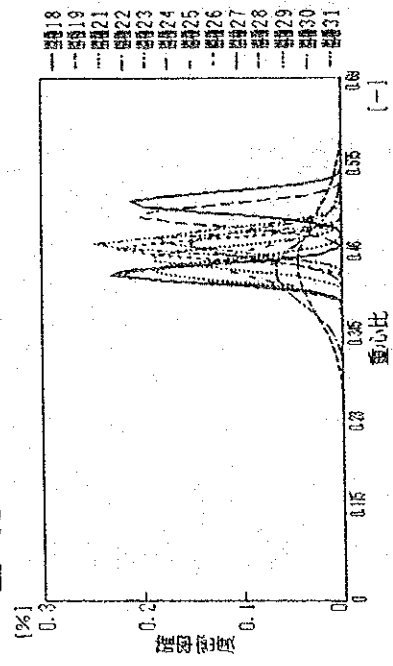
重心比パラメータにおける各樹種の分布



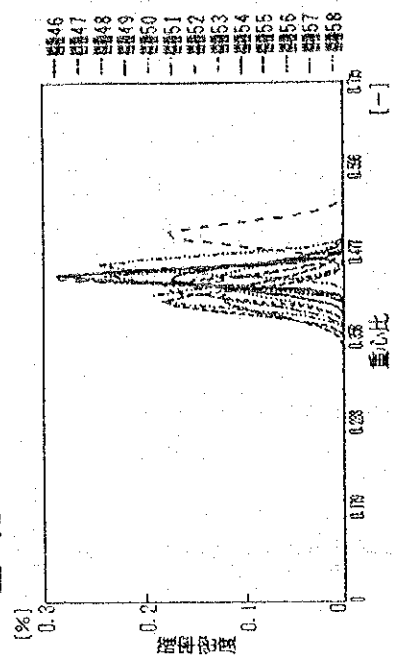
重心比パラメータにおける各樹種の分布



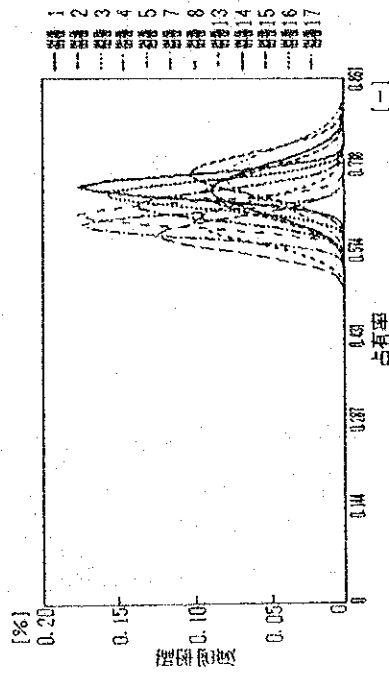
重心比パラメータにおける各樹種の分布



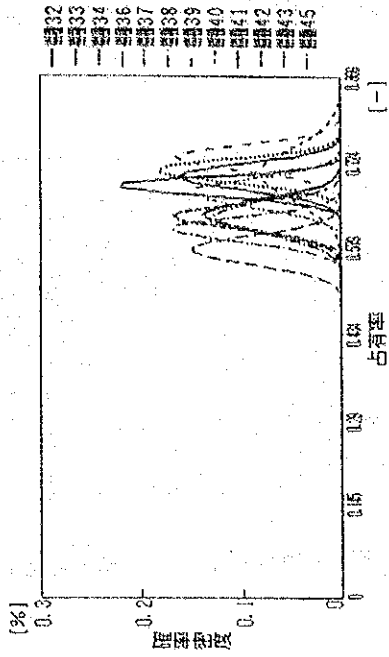
重心比パラメータにおける各樹種の分布



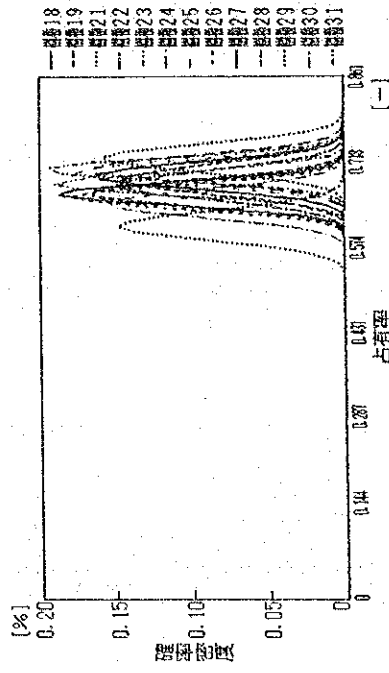
占有率パラメータにおける各樹種の分布



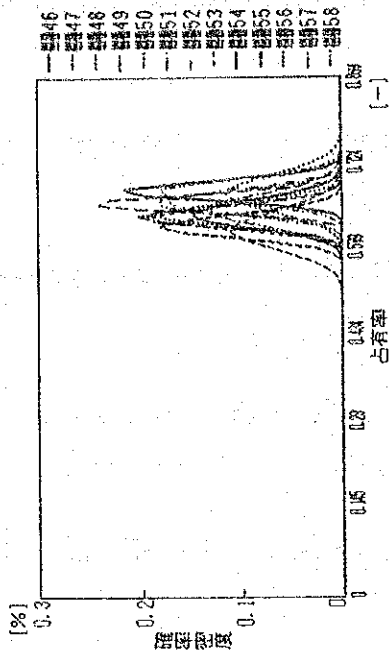
占有率パラメータにおける各樹種の分布



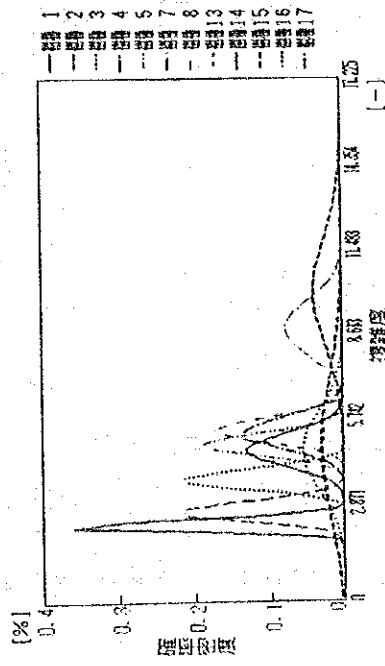
占有率パラメータにおける各樹種の分布



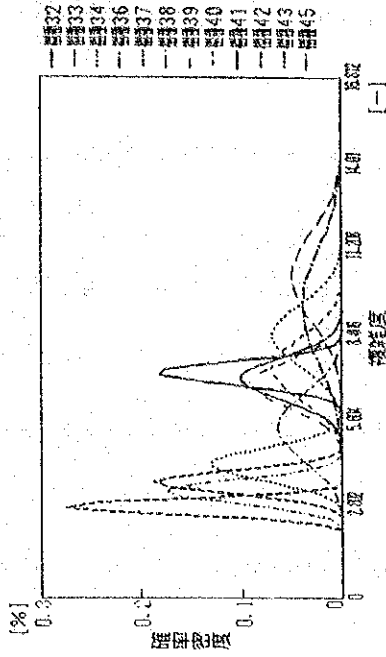
占有率パラメータにおける各樹種の分布



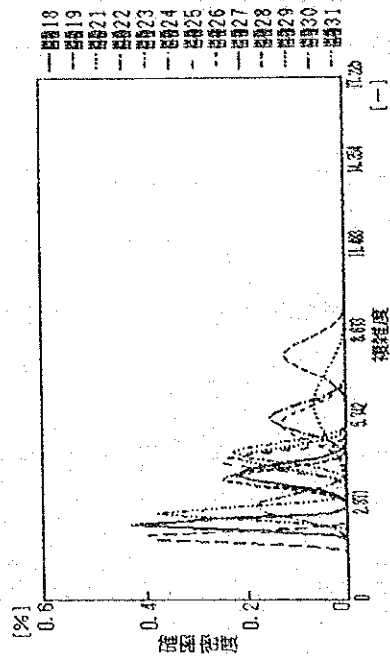
複雑度パラメータにおける各樹種の分布



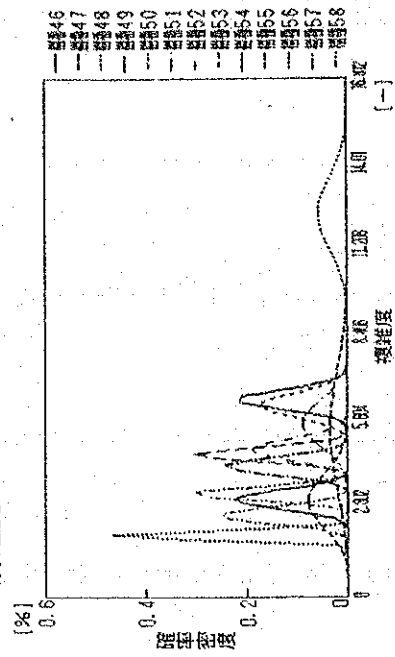
複雑度パラメータにおける各樹種の分布



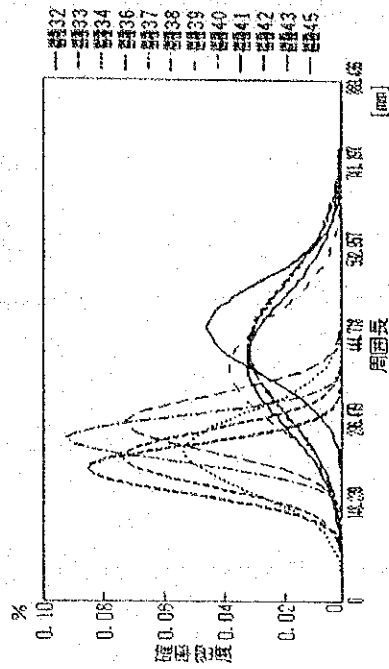
複雑度パラメータにおける各樹種の分布



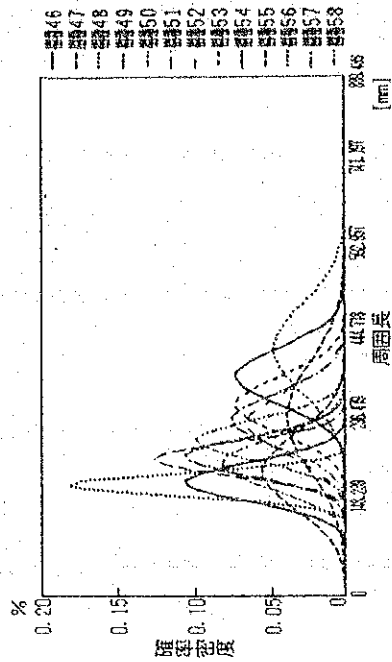
複雑度パラメータにおける各樹種の分布



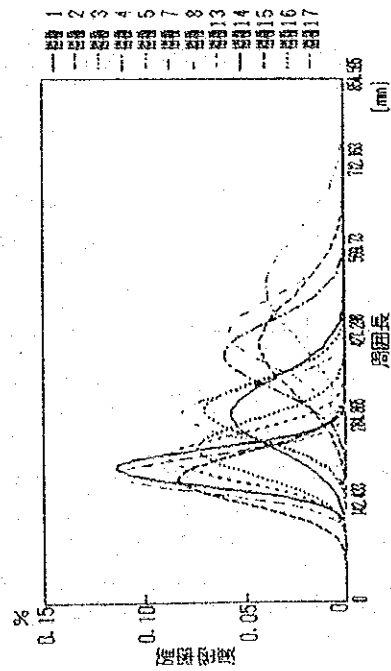
周囲長パラメータにおける各樹種の分布



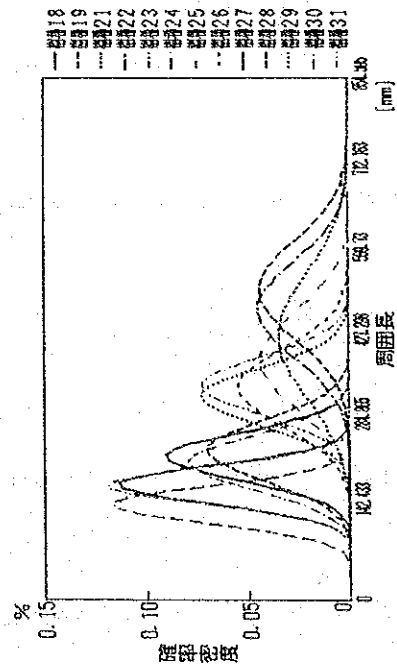
周囲長パラメータにおける各樹種の分布



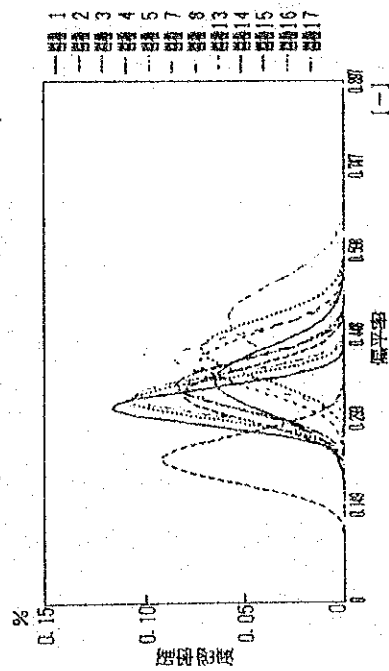
周囲長パラメータにおける各樹種の分布



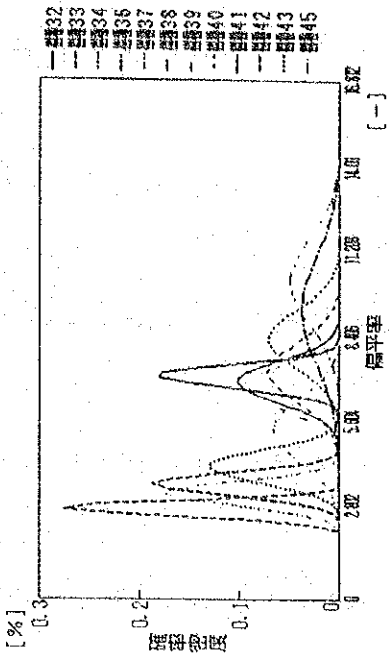
周囲長パラメータにおける各樹種の分布



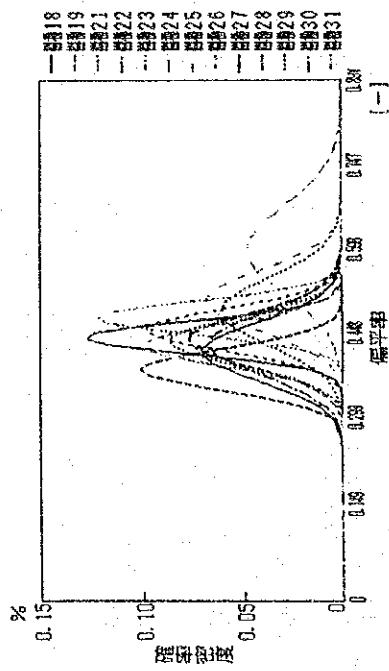
偏平率パラメータにおける各樹種の分布



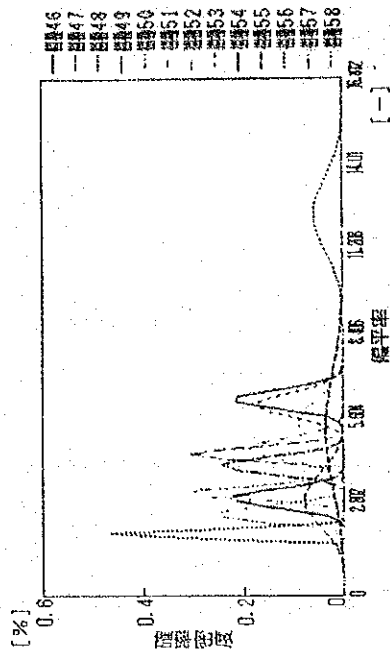
偏平率パラメータにおける各樹種の分布



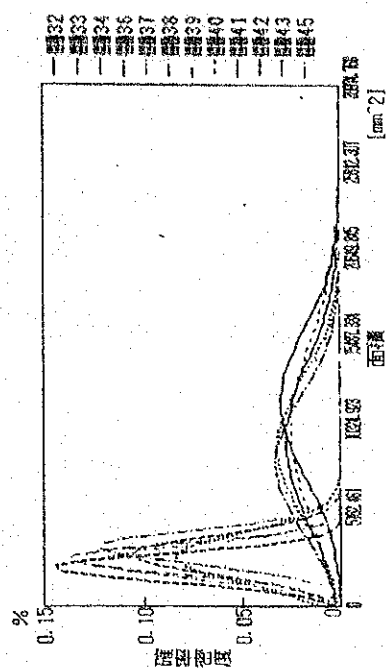
偏平率パラメータにおける各樹種の分布



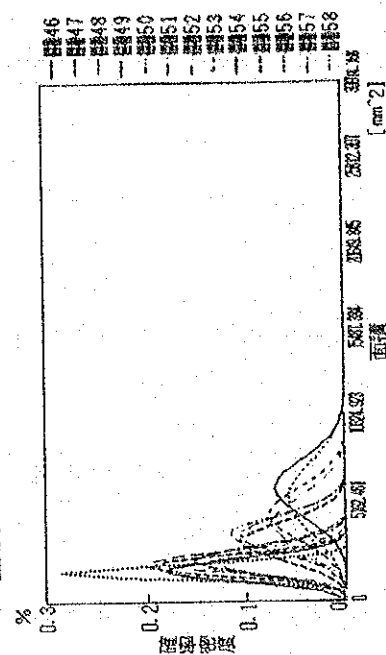
偏平率パラメータにおける各樹種の分布



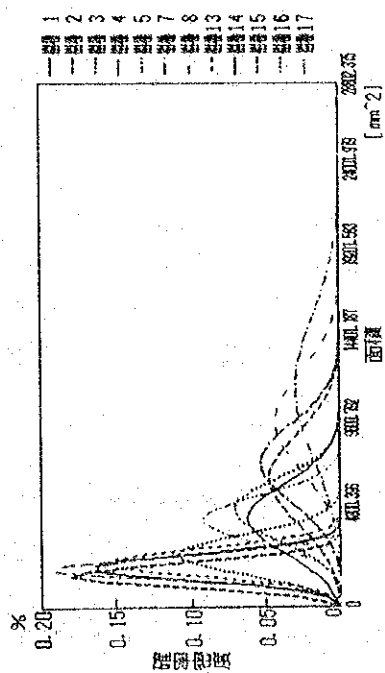
面積パラメータにおける各樹種の分布



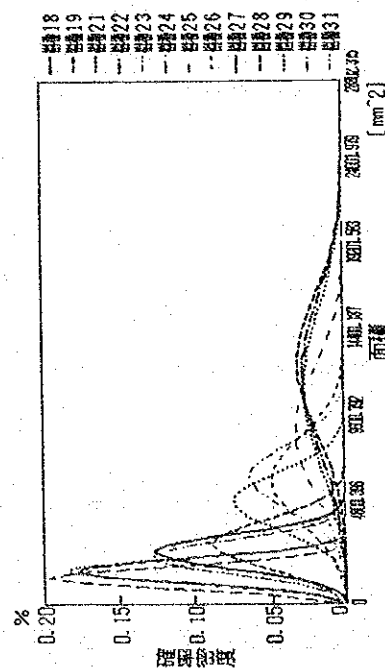
面積パラメータにおける各樹種の分布



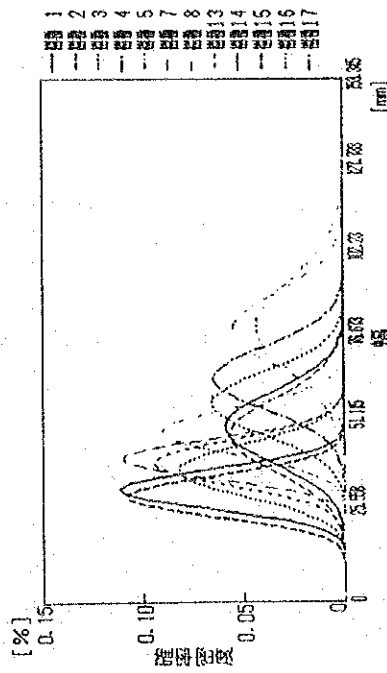
面積パラメータにおける各樹種の分布



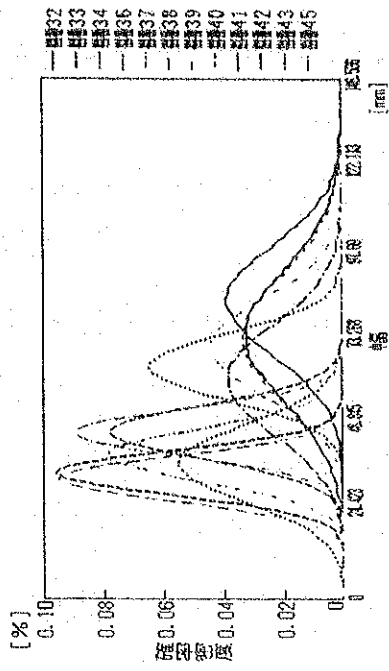
面積パラメータにおける各樹種の分布



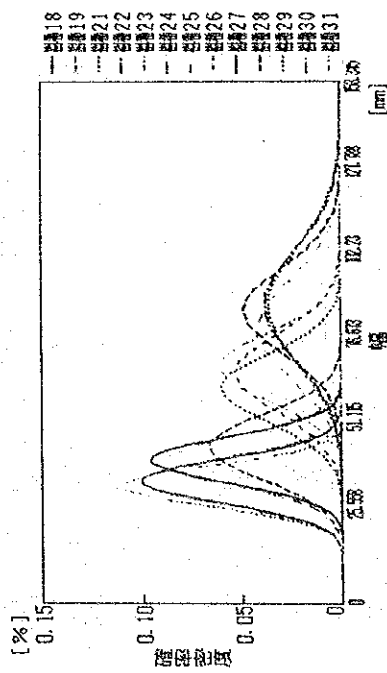
幅パラメータによる各樹種の分布



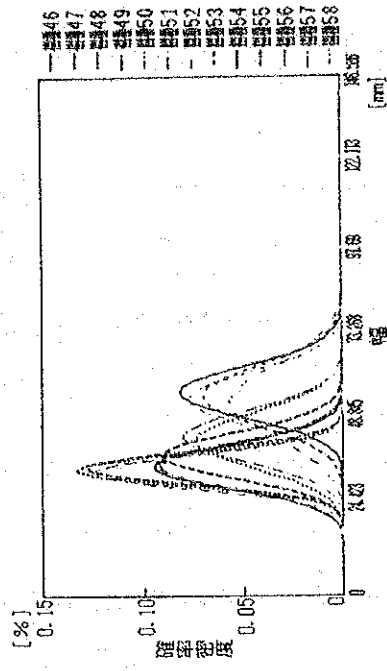
幅パラメータによる各樹種の分布



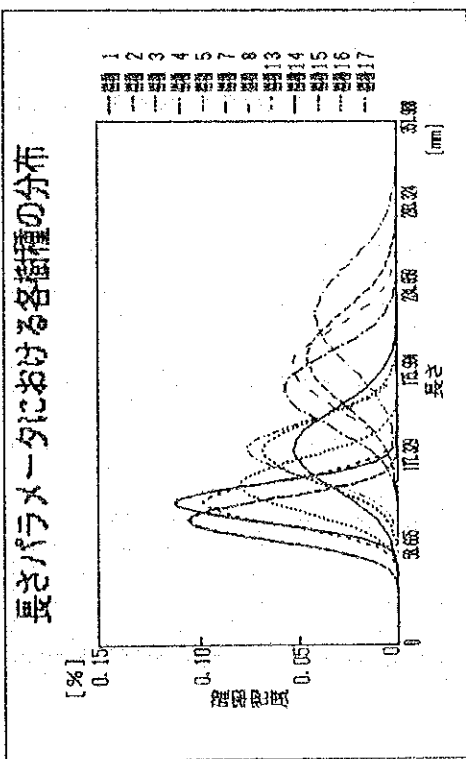
幅パラメータによる各樹種の分布



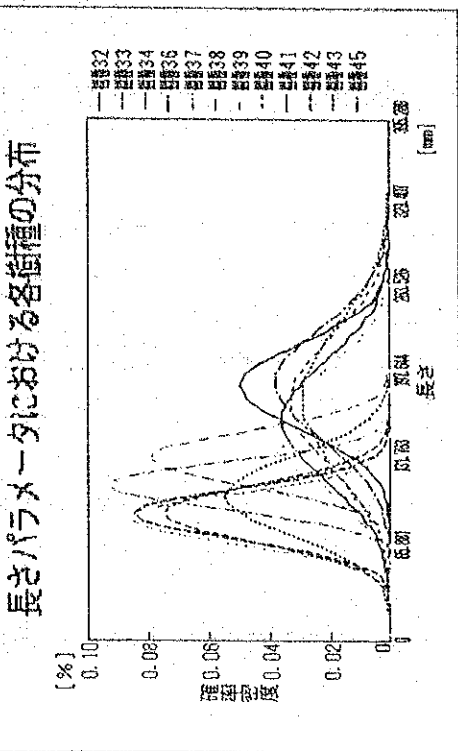
幅パラメータによる各樹種の分布



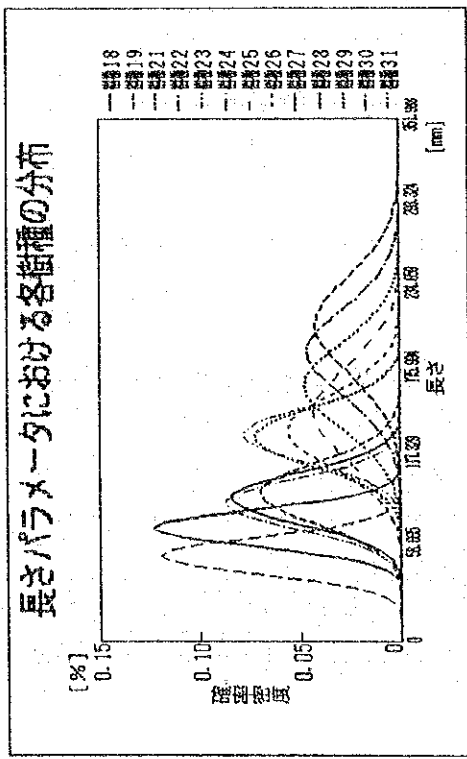
長さパラメータにおける各樹種の分布



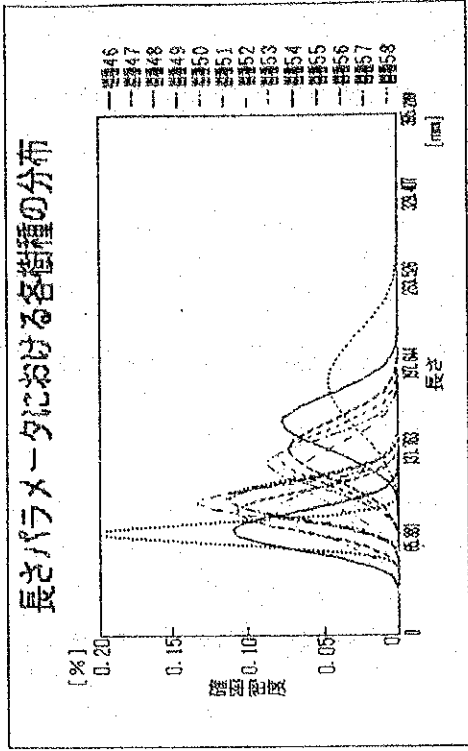
長さパラメータにおける各樹種の分布



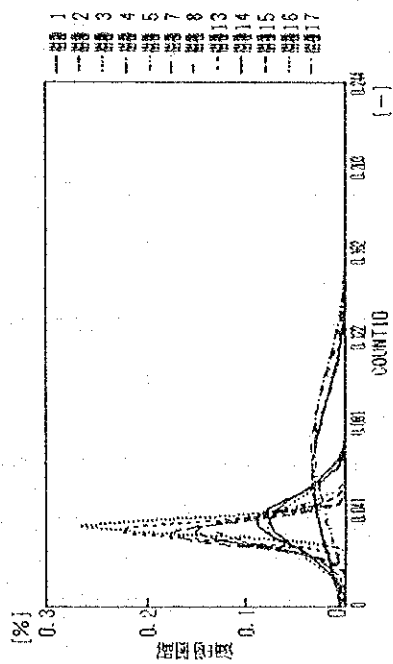
長さパラメータにおける各樹種の分布



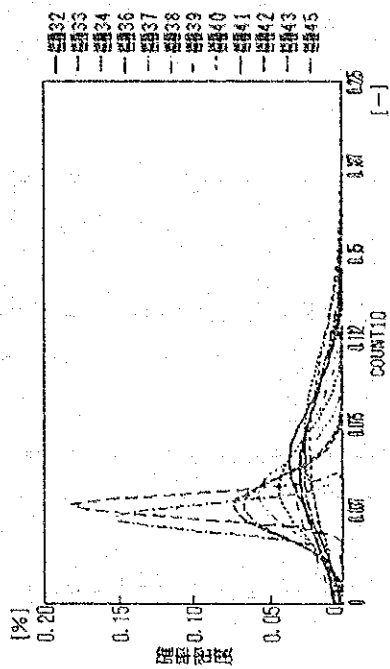
長さパラメータにおける各樹種の分布



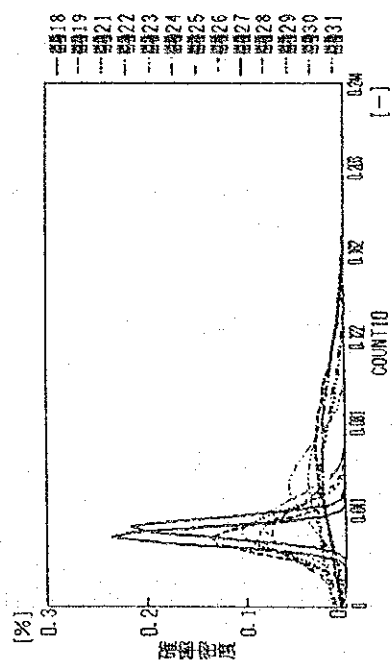
COUNT10パラメータにおける確率密度の分布



COUNT10パラメータにおける確率密度の分布



COUNT10パラメータにおける確率密度の分布



COUNT10パラメータにおける確率密度の分布

