

TR-A-0167

自由手書き文字認識の研究

長石 道博

1993. 3.25

ATR 視聴覚機構研究所

〒619-02 京都府相楽郡精華町光台 2-2 ☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411

Facsimile: +81-7749-5-1408

自由手書き文字認識の研究

視覚研究室 長石 道博

研究の目的： 一般人が意味のある日本語文章を罫線上に書いた自由手書き文字を実際に収集、分析してその認識方法を検討し、コンピューター上で認識方法の検証を行なう。

研究経過と成果の概要：

1、研究の動機

文字認識技術は、マンマシンインターフェイスの向上を図る有力な手段として古くから研究・開発が行なわれている。活字体（アルファベット等）の認識やオンライン手書き文字認識の基礎的な技術開発が行なわれており、その成果としてこれらの技術を応用した製品（AIノート、TRAN、ワードバンクペン、スキャニオン等）が市場に出ている。一方、これからのマルチメディア時代に向けて、ペン入力コンピューター等、誰でもが簡単に入力できる手段が一層要求されるようになってくる。この要求を満たすには、手書き文字、特に枠などの制限がなく自由に書いた文字を識別できる技術が不可欠である。この研究は将来非常に有用性がある技術なので、自由手書き文字の認識を研究テーマとして選んだ。

2、手書き文字認識の問題点

手書き文字認識研究の事例は非常に多く、認識の応用商品も見られるが、近年認識性能の向上が限界に達しており、手書き文字を十分認識することが依然実現できていない。これは次のような問題点があるからと考えられる。まず、認識率が使用するデータに大きく依存する点である。例えば、学習データをほぼ完璧に認識できる認識方法でも、データが異なると急速に認識率が低下する場合がある。そのため文字データ自体の分析・評価が必要である。次に、現在の認識方法自体の手書き文字の変動（大きさや形など）の吸収が十分でない点である。枠や書き順（オンラインの場合）等の制限がある場合、比較的認識率のよい認識方法でも、枠が無くなったりすると、認識することが非常に困難になる場合がある。従って、より手書き文字の変動に対して柔軟性のある認識方法の開発が必要である。以上の問題点を踏まえて、1)手書き文字のデータ自身がどのような性質をもっているのか、2)自由手書き文字を認識する為の新しい文字認識方法を提案できないか、という観点から自由手書き文字を認識の研究を始めた。

3、自由手書き文字の収集とその分析

自由手書き文字の筆記条件とその収集方法（パソコンに表示した文字列を書き取る）についての検討・提案を行なった上で、実際に自由手書き文字を収集、分析を行なった。その結果、自由手書き文字が制限付き手書き文字に比べ複雑で、大きな変動をもっていることを定量的に示した。更に提案した自由手書き文字の妥当性を大学の試験答案からの文字と比較・検証し大規模な自由手書き文字のデータベースを作成することの出来る可能性を示した。

また、自由手書き文字の文字変動が文書内の筆記位置に依存することが解かった。これはデ

データベース作成におけるデータの均一性に問題が生ずる可能性があることを示す。また、逆にこのような筆記位置に関する依存性を利用した文字認識方法を提案した。

4、新しい手書き文字認識方法の提案

今までの変動の分析結果から、自由手書き文字の文字変動が非常に大きく複雑なことが判明した。現在まで行なわれてきた認識方法の改良の積み重ねだけでは、このように変動の大きい自由手書き文字に対し満足な認識結果を得ることが難しいと考えられる。今までとは全く異なった新しい観点から文字認識を考えてみる必要があると考えられる。

そこで、心理学や生理学知見である「視覚の誘導場理論」に基づく認識方法、文字切り出し方法を提案し、実際に認識や切り出しが可能であるという結果を得た。

誘導場を用いて文字を認識する際には、文字の種類や変形の様子を誘導場の分布の違いを定量化することで可能になる。そして、本認識方法が活字体をはじめ手書き数字でも、従来法に比べて十分な認識性能をもっていることが示された。

誘導場を用いる文字の切り出しでは、従来の斜影による1次元的な方法では切り出せなかった文字でも、2次元的に切り出すことが出来る。特に手書き文字のように隣接する文字の間隔が狭い場合、斜影による方法に比べて有利であることが、大学の試験答案からの実画像に対して示された。

更に、錯視図形に於ける視覚の誘導場の分布から、場を用いて人間の視覚現象も説明できることが示唆された。このことは、「場」を用いることで、人間の文字認識の仕方に近い認識装置を実現できる可能性を示唆するものと考えられる。

自由手書き文字認識の研究

Recognition of Unconstrained Handwritten Characters

長石 道博

Michihiro NAGAISHI

ATR視聴覚機構研究所

Visual Perception Department

ATR Auditory and Visual Perception Research Laboratories

2-2 Hikaridai, Seika-cho, Soraku-gun,

Kyoto 619-02, Japan

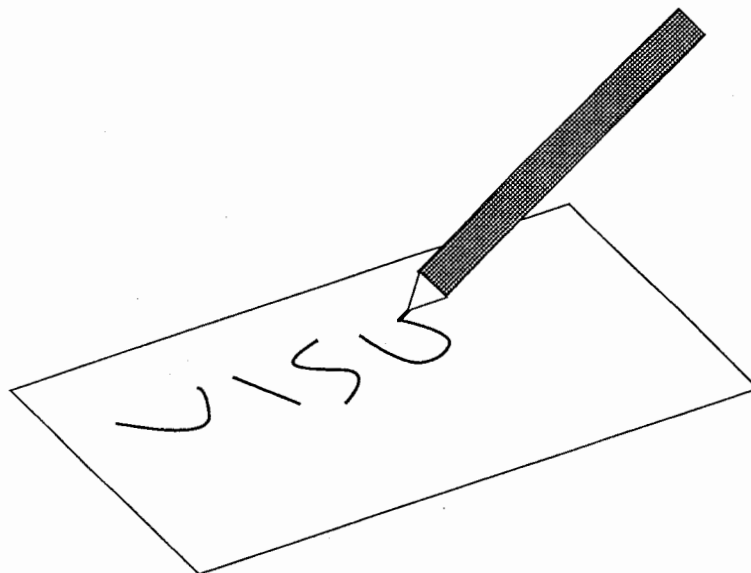
ATR

Advanced Telecommunications Research Institute International

-
- 1、自由手書き文字認識の問題点
 - 2、文字収集
 - 3、手書き文字の変動分析
 - 4、視覚の誘導場理論
 - 5、誘導場による認識
 - 6、誘導場による切り出し
 - 7、錯視図形がつくる誘導場

手書き入力・・・より人間に近づいた
ヒューマンインターフェースの実現を目指して

Handwriting is a natural and desirable method of inputting data into a computer.



「我々がごく普通にものを考えながら紙に書く動作」



自由手書き文字認識の必要性

It is need to recognize unconstrained handwritten characters.

ATR

Advanced Telecommunications Research Institute International

問題点
Problem

認識率がデータに大きく依存する



文字データ自体の分析・評価が必要

現状の認識方法では手書き文字の変動吸収が不十分



より文字変動に対して柔軟性のある新しい認識方法の開発

目的
Purpose

自由手書き文字認識の必要性
It is need to recognize unconstrained handwritten characters.

特徴分析
Feature analysis

自由手書き文字データベース
Database of unconstrained handwritten characters

自由手書き文字の定義
自由手書き文字とは何か？
What are unconstrained handwritten characters ?

収集方法の検討
いかにして自由手書き文字を収集すべきか？
How should we collect unconstrained handwritten characters ?

自由手書き文字の変動分析
自由手書き文字特有の特徴とは何か

ATR

Advanced Telecommunications Research Institute International

手書き文字の収集法

Collection method of handwritten characters

実際に書かれた文書そのまま集める (自由手書き文字)

大学の試験答案などを収集
筆記者への制約は一切ない
文字データの作成効率が非常に悪い

見本を見ながら枠内に記入 (制限付き手書き文字)

多くの文字認識用データベースが採用
枠や字形などの筆記制限あり
文字データの作成効率が非常に高い

表示されていく文字列を書き取る

講義中のノートをとる場面を模擬した方法
筆記制限はない
自由手書き文字を効率的に収集



Advanced Telecommunications Research Institute International

手書き文字文書の一例

A handwritten document sample

* 大学の試験答案から

こり(酸塩)のたち2い

とた2い3が実状2い

小も9と3=4が"の2い

ATR

Advanced Telecommunications Research Institute International

自由手書き文字の筆記条件

Writing Conditions of Unconstrained Handwritten Characters

自由手書き文字とは何か？

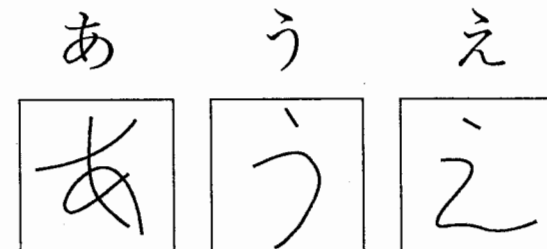
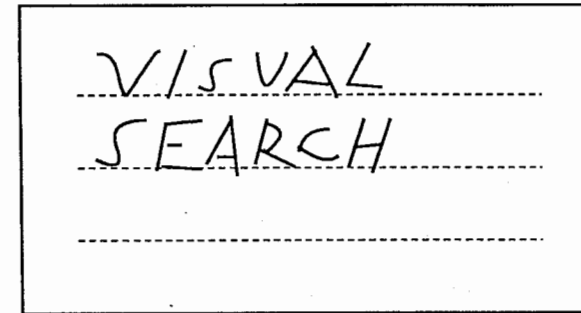
What are unconstrained handwritten characters ?

筆記内容が無意味文字の羅列ではなく、単語や行、段落等の単位で意味をもった文字列を含んでいる。

The contents must not be a meaningless group of characters, but must include characters which have meanings as parts of words, lines or paragraphs.

罫目等の文字枠や文字の書き方（楷書体、書き順）等の制限はしない。

There are no constraints such as a specified frame in which the characters must be written. Writing on ruled lines is allowed, because memos are often written on ruled paper.



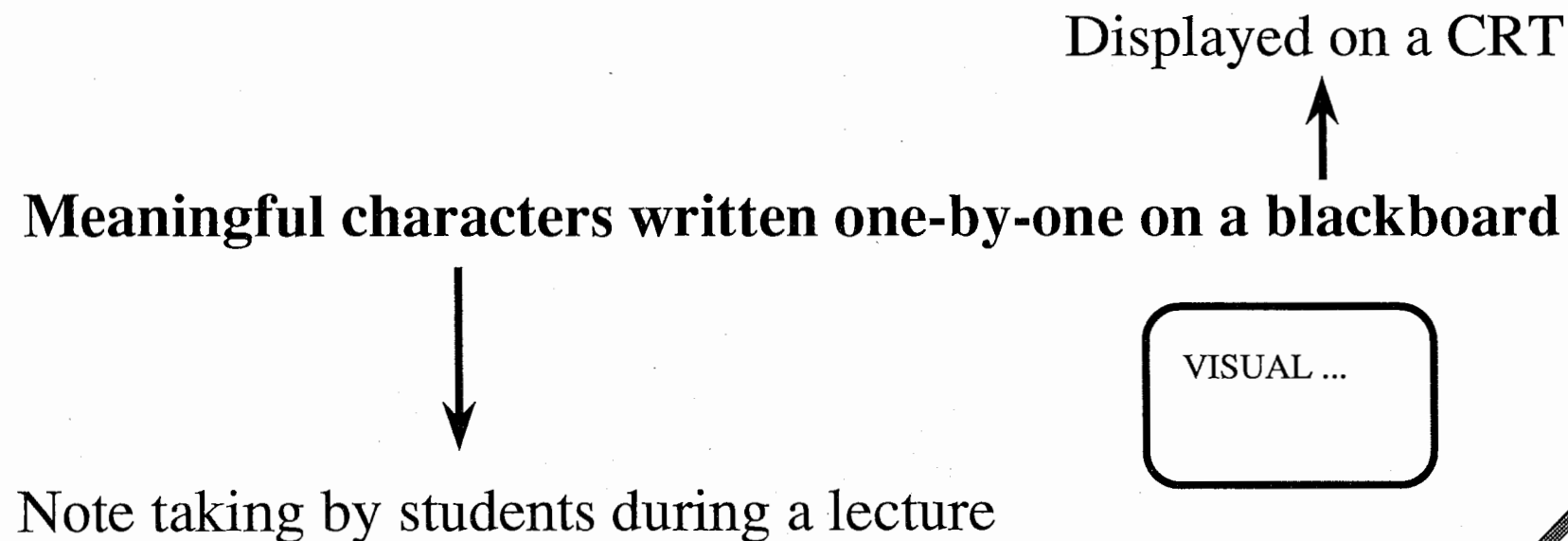
circumscribed rectangle frame

ATR

Advanced Telecommunications Research Institute International

筆記の疑似表示を用いた収集方法

Collection Method Using Pseudo-Representation of Writing

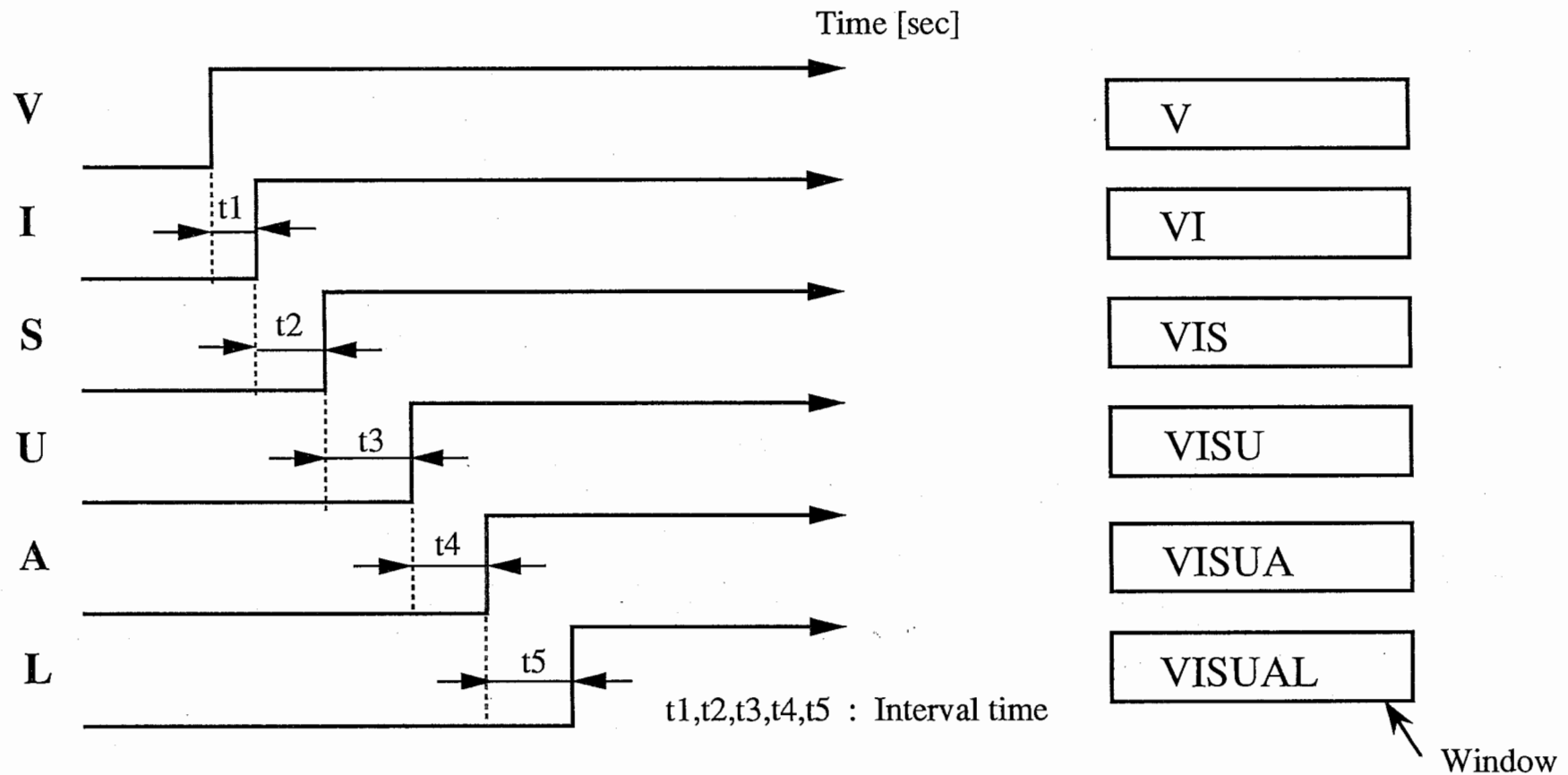


* When the screen becomes full, it is cleared. Therefore, it is impossible to wait until the end of the text before starting to write.

* This method forces the writer to wait until the character is displayed, so the writer must write down the characters as they appear one after the other.

ATR

Advanced Telecommunications Research Institute International



文字列表示

Sequence of presenting characters

The character display rate is set to match the time needed to actually write.

For example, the interval "t1" between displaying "V" and "I" is shorter than the interval "t4" between "U" and next character "A", because "A" has more strokes so it takes longer to write.

変動エントロピ

Variation Entropy

$$h = -\sum_i^M \frac{n(i)}{N} \log \frac{n(i)}{m}$$

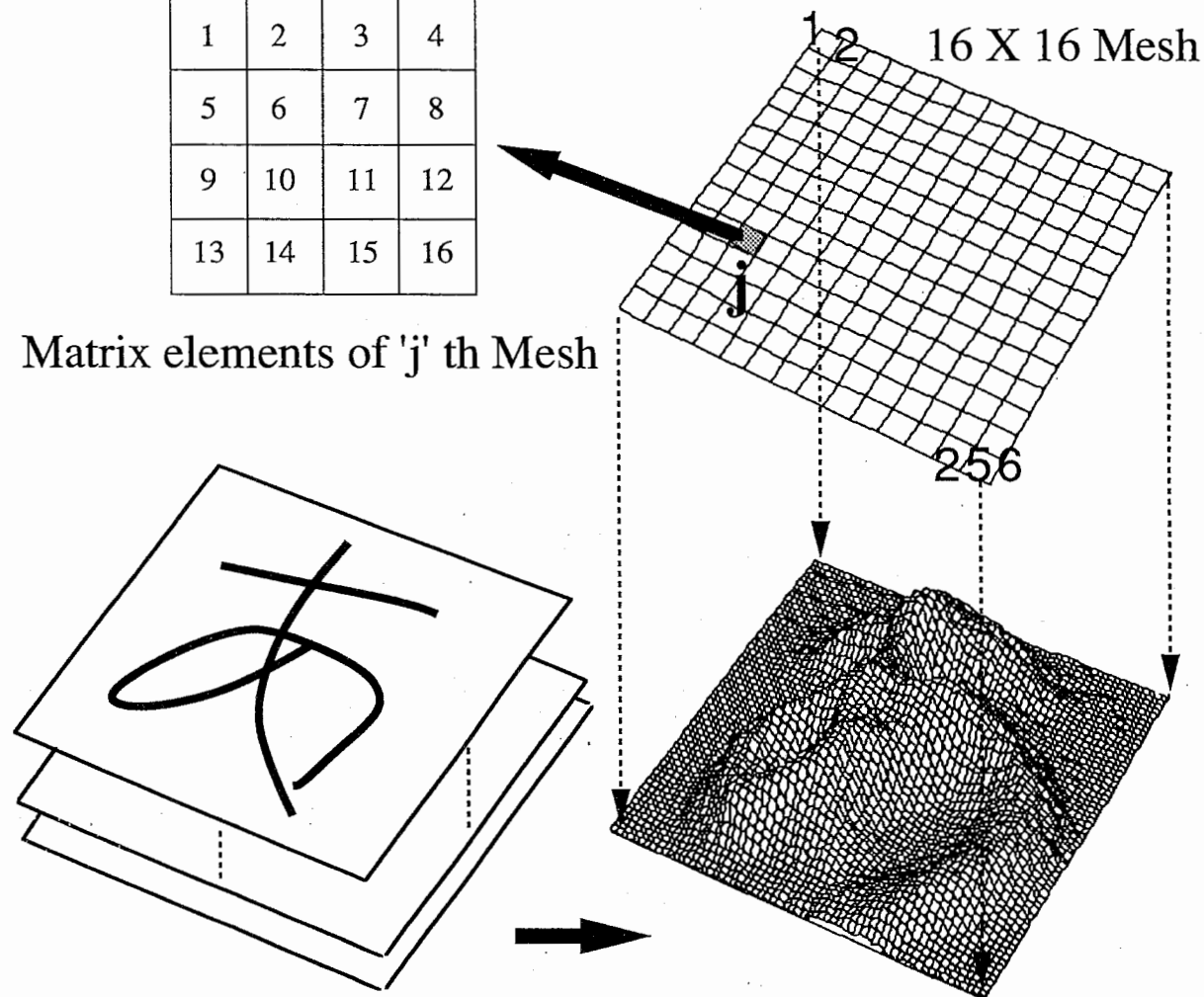
$$(N = \sum_i^M n(i))$$

$$0 \leq n(i) \leq m$$

$$(i = 0, 1, \dots, M = 64 \times 64)$$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Matrix elements of 'j' th Mesh



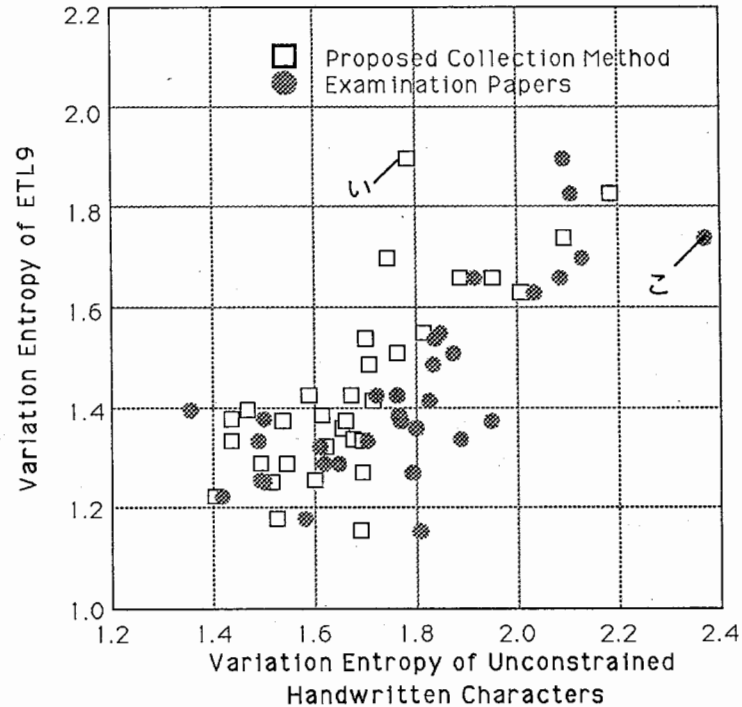
Superimposing

Three-dimensional figure

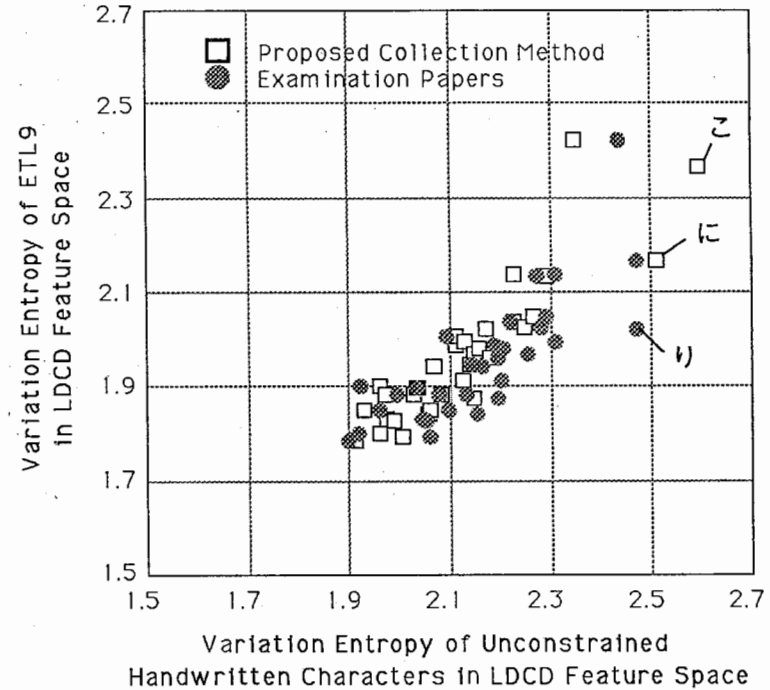
ATR

自由手書き文字と制限付き手書き文字の変動の比較

- * 制限付き手書き文字として ETL9 を使用
- * 平仮名のみ比較

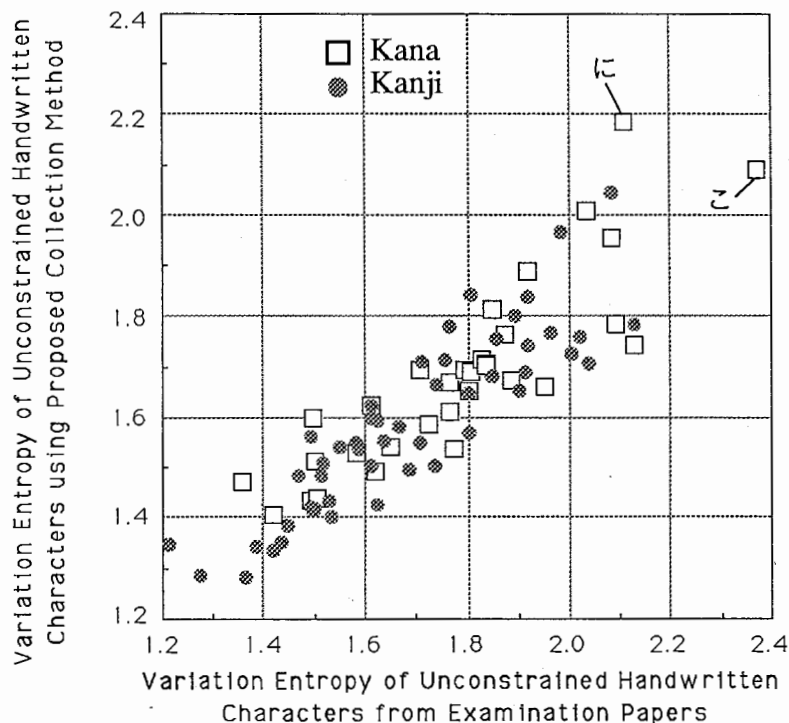


変動エントロピ

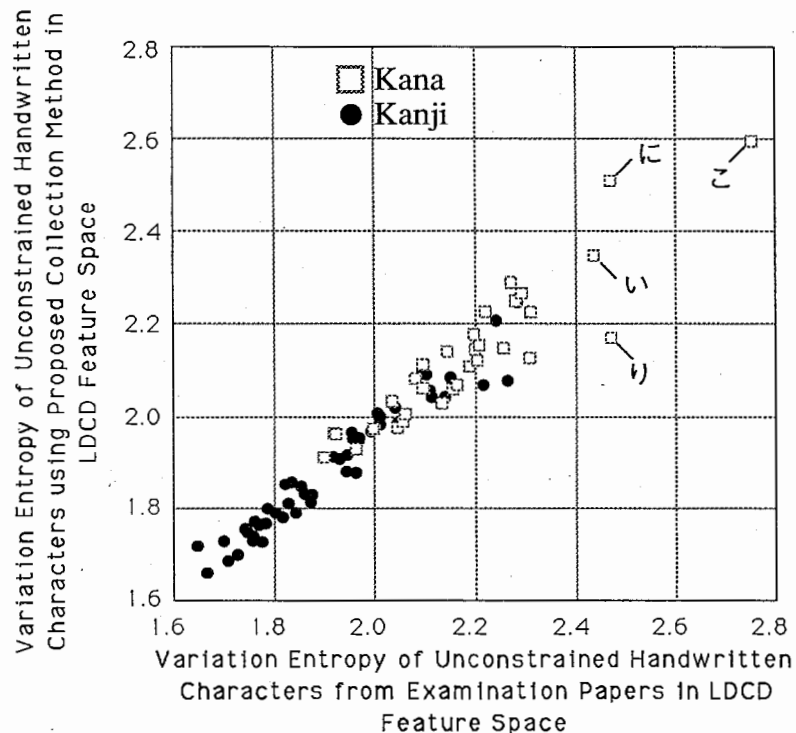


LDCD 特徴のエントロピ

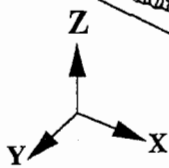
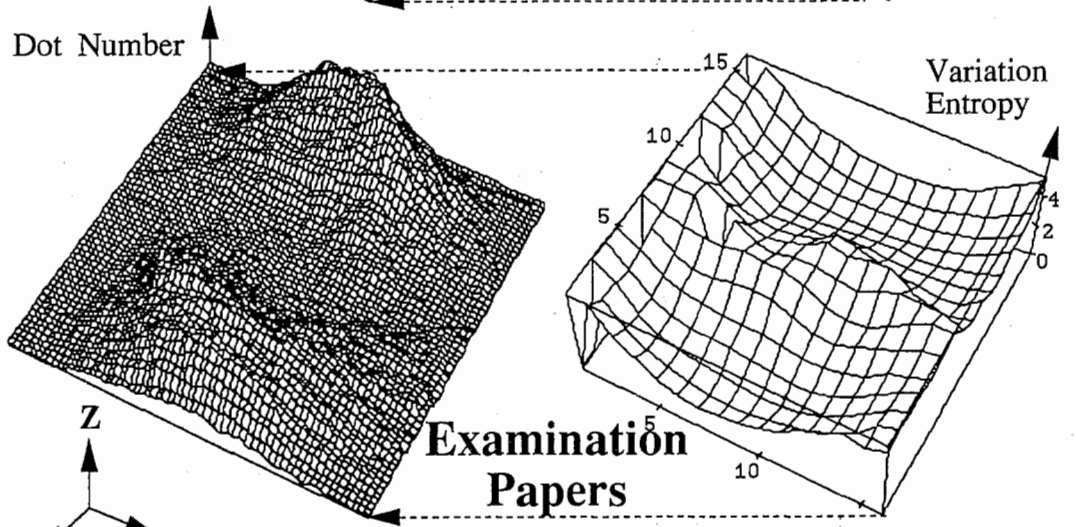
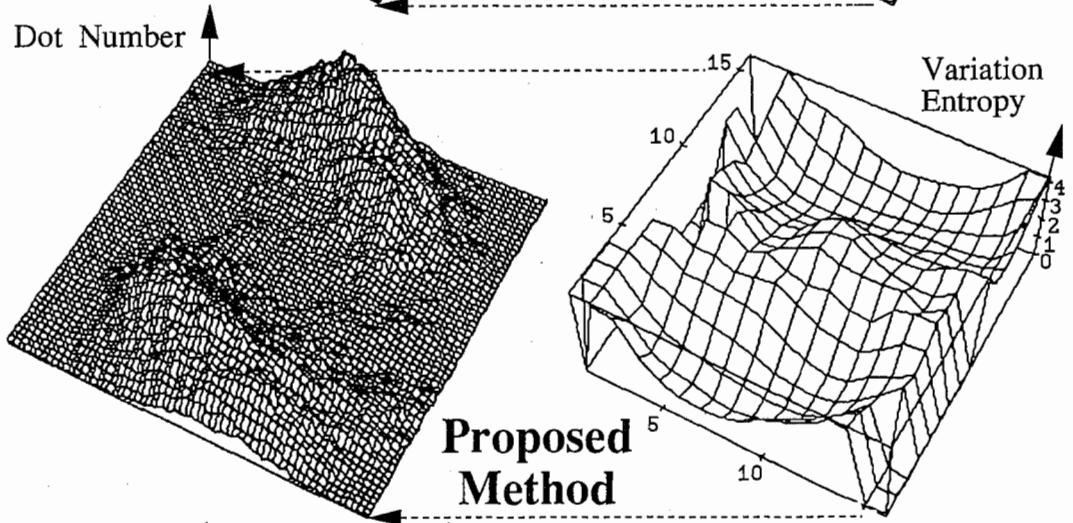
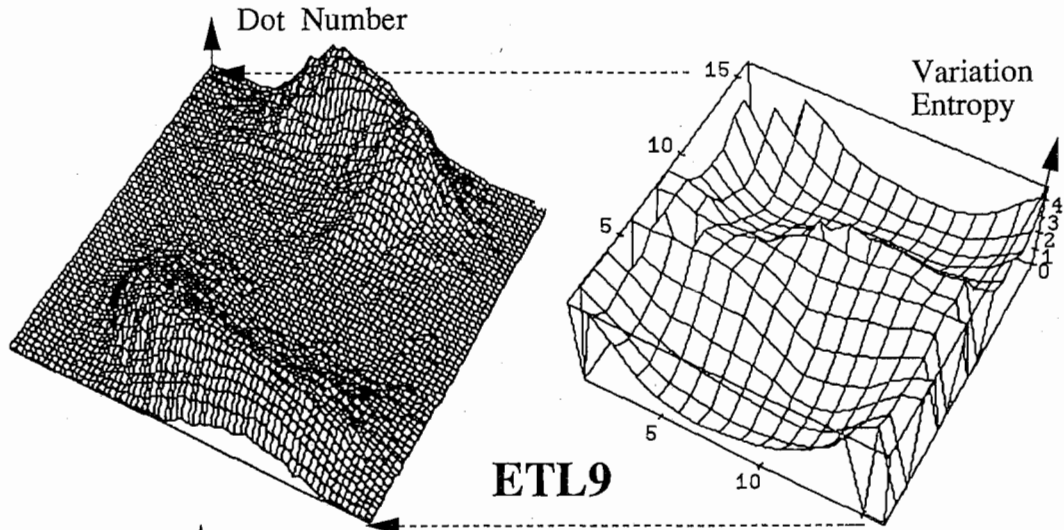
自由手書き文字 (実際の文書と疑似表示による文書)の 変動の比較



変動エントロピ



LDCD 特徴のエントロピ

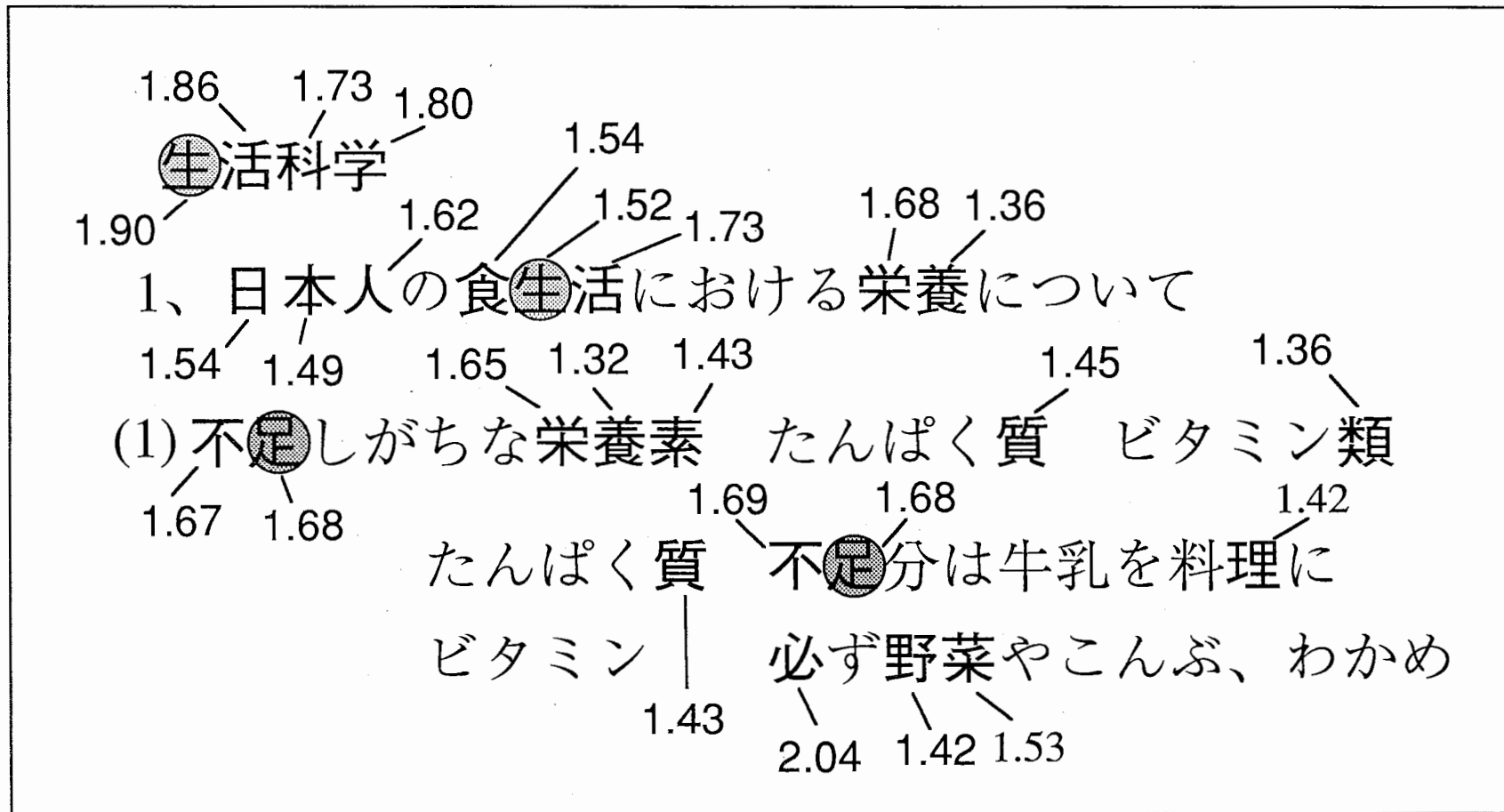


Three-dimensional Figure

Variation Entropy

文書内における各文字の変動エントロピー

Variation entropy of individual characters in the sample document



まとめ
Conclusion

自由手書き文字認識の必要性
It is need to recognize unconstrained handwritten characters.

特徴分析

自由手書き文字データベース
Database of unconstrained handwritten characters

自由手書き文字の定義

収集方法の検討

自由手書き文字の筆記条件
Writing Conditions of
Unconstrained Handwritten Characters

自由手書き文字の変動分析
自由手書き文字特有の特徴とは何か

疑似表示を用いた収集方法
Collection Method
Using Pseudo-Representation of Writing

制限付き手書き文字に比べて
大きな文字変動

文書内位置に依存する文字変動
Character vary depending
on their positions in a document

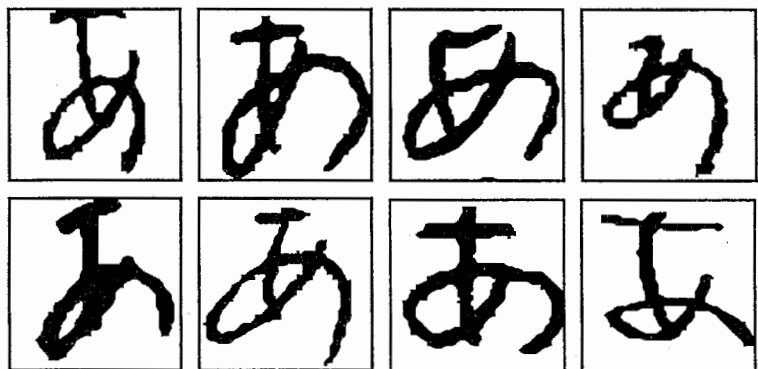
新しい観点にもとづく文字認識方法の検討

ATR

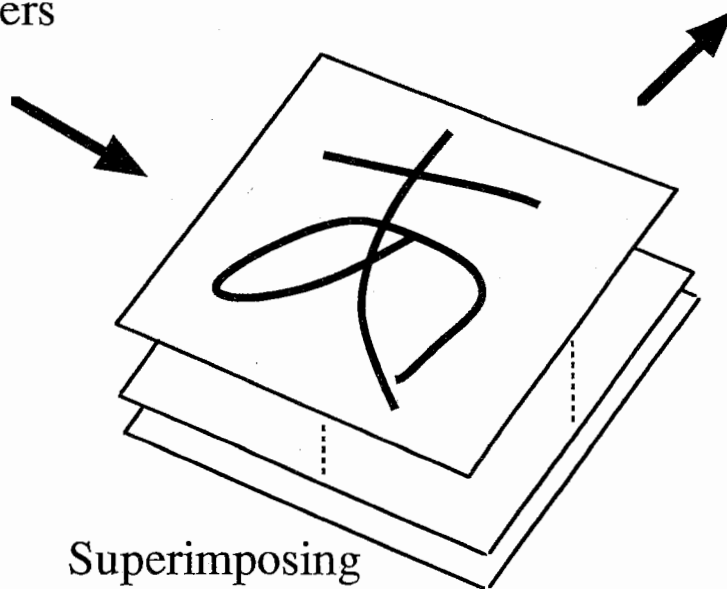
-
- 1、はじめに
 - 2、視覚の誘導場理論
 - 3、誘導場理論の一般化とその推定
 - 4、誘導場を用いた文字の識別
 - 5、実験
 - 6、今後の課題

手書き文字の重ね合わせ

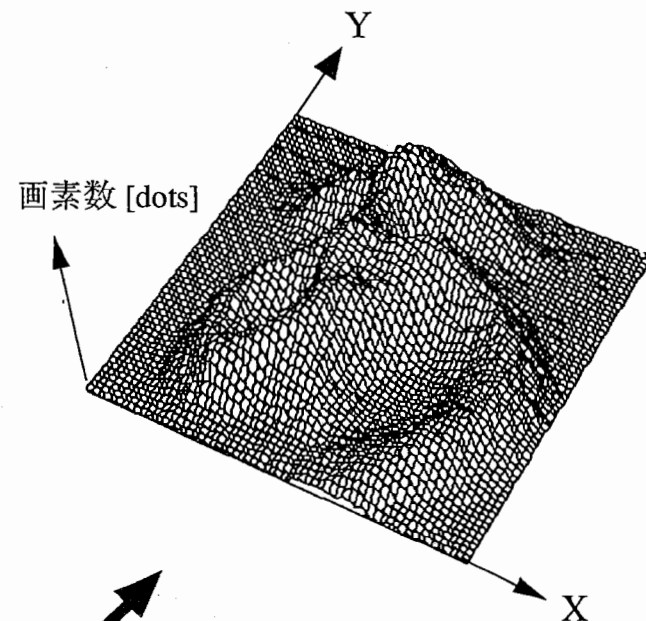
Superimposing of handwritten characters



Handwritten characters



Superimposing



3 - dimensional figure

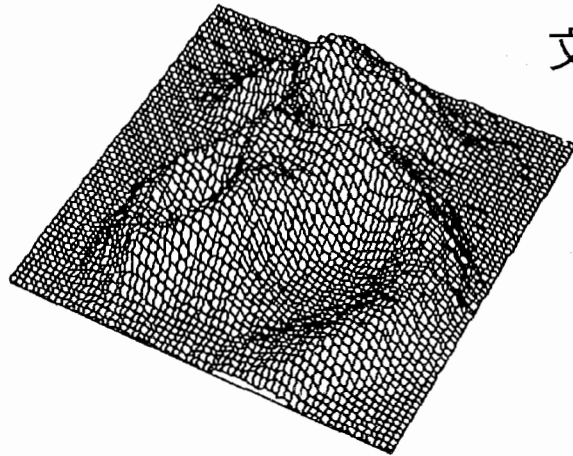
文字が書かれる領域を規制するもの

書道における空間配置の
重要性

重ね合わせた文字画像の
広がり

自然現象のパターン形成での
物理法則

文字の書かれる領域を規制する物理法則



「場」
field

視覚の誘導場理論

The theory of the field of induction on the retina

ATR

Advanced Telecommunications Research Institute International

視覚の誘導場に関する研究

生理学

網膜に於ける誘導 (本川) 鯉の網膜上で誘導現象を発見
視覚モデル (藤井) 生理学的な知見より構成



錯視図形の説明

心理学

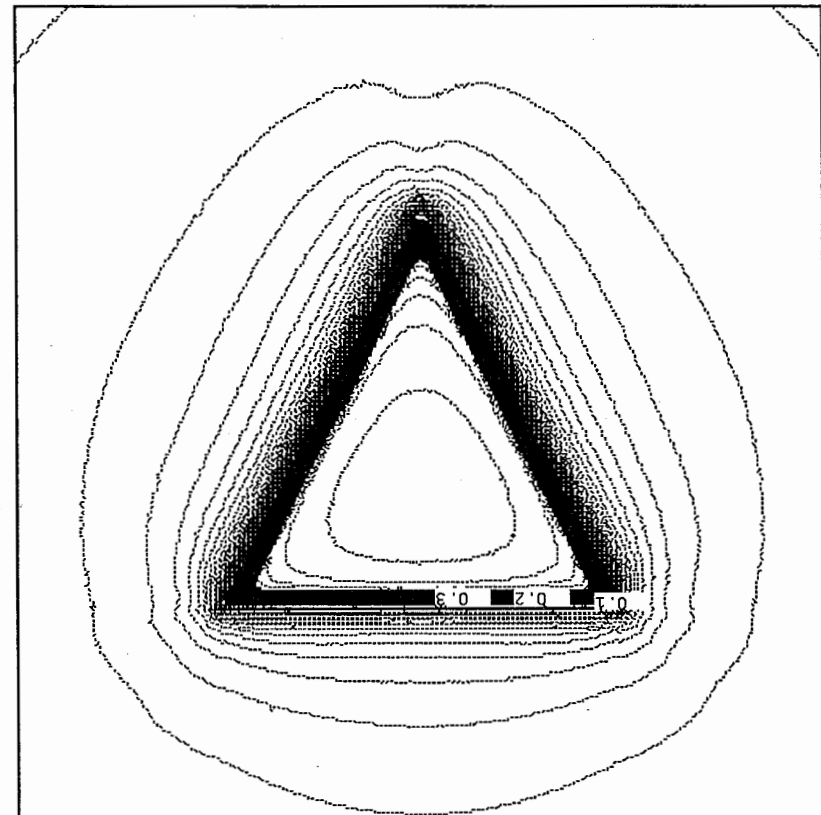
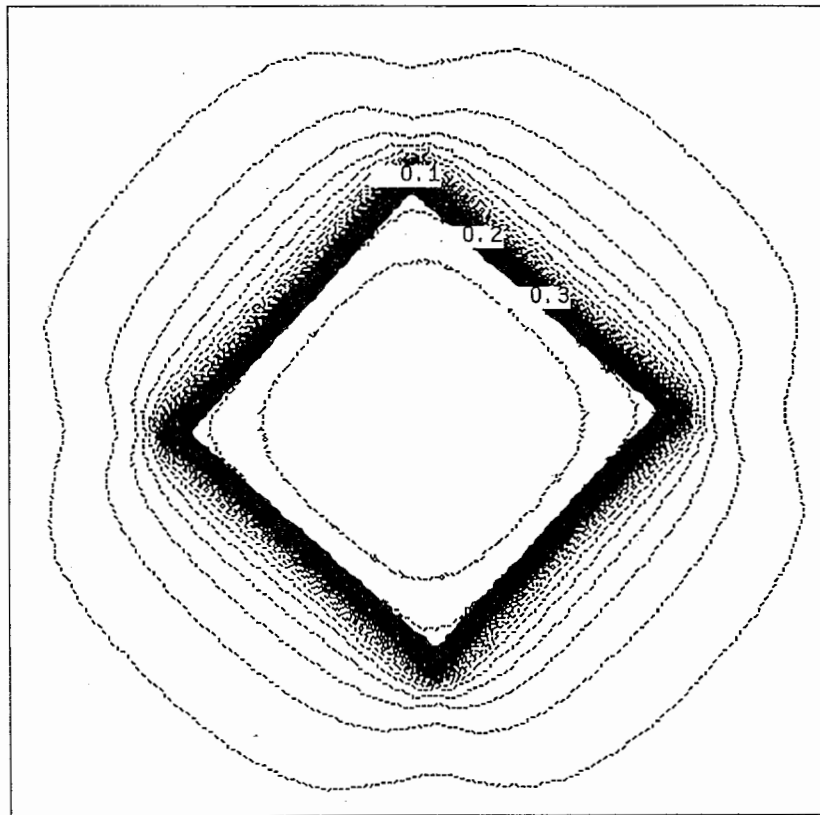
感応理論 (小保内) 神経の興奮性と抑制性活動を考慮
光刺激閾法 (横瀬) 図形、文字の場の理論的予測
場による文字認識の可能性を示唆



横瀬の心理的ポテンシャル場理論

図形の「場」

Induction fields of figures



* この図形の「場」は、本報告で提案したデジタル画像における誘導場の計算方法に基づいて計算したものです。

太線：図形
細線：等ポテンシャル線

ATR

Advanced Telecommunications Research Institute International

心理的ポテンシャル場理論

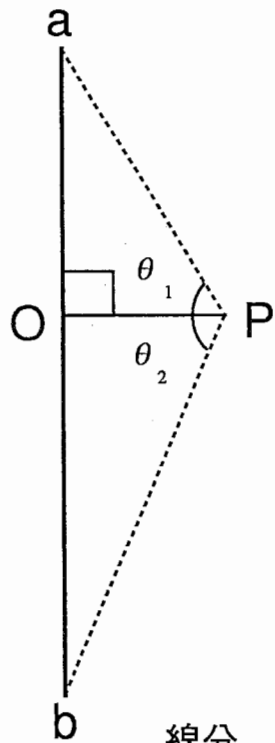
The theory of psychological potential field

点Pにおける「場」の強さ (心理実験から)

$$M = f\left(E \frac{H^b}{D^d}\right)$$

D 図形からの距離
 H 図形と背景の明度対比
 E 図形の形の因子

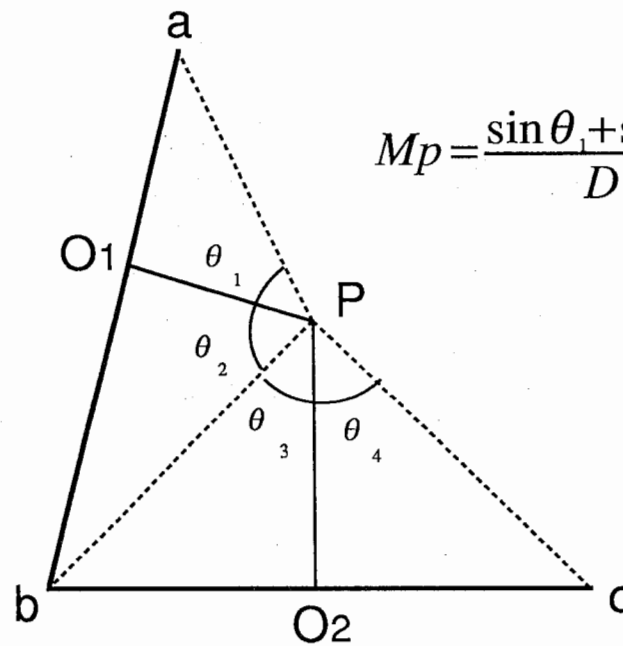
↓
理論式の提案



線分
A line.

$$M_p = \frac{1}{D} \int_{-\theta_2}^{\theta_1} \cos \theta \, d\theta$$

$$= \frac{\sin \theta_1 + \sin \theta_2}{D}$$



重ね合わせ
Superposition.

$$M_p = \frac{\sin \theta_1 + \sin \theta_2}{D_1} + \frac{\sin \theta_3 + \sin \theta_4}{D_2}$$

$$PO_1 = D_1$$

$$PO_2 = D_2$$

ATR

心理的ポテンシャル場の一般化 (1)

Generalization of psychological potential field

電磁気学的解釈が可能

理論予測可能

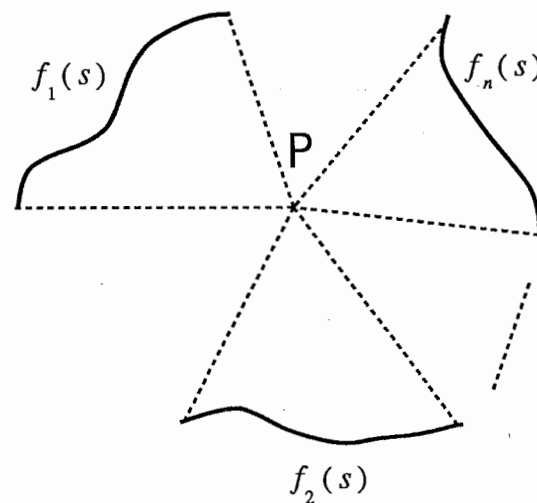
生理学的知見と一致 (Müller-Lyerの錯視図形)



視覚の誘導場理論の基礎理論として
吟味する余地あり

クーロンポテンシャルの重ね合わせ

$$M_p = \int \frac{ds}{R(f_1(s))} + \int \frac{ds}{R(f_2(s))} + \dots + \int \frac{ds}{R(f_n(s))}$$
$$= \sum_i^n \int \frac{ds}{R(f_i(s))} \quad R(f_i(s)) : \text{点Pと各曲線 } i \text{ との距離}$$



複数の線分
Lines.

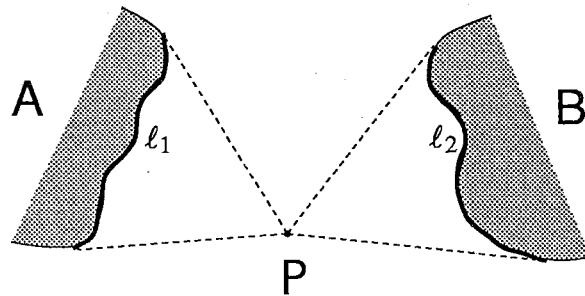
ATR

心理的ポテンシャル場の一般化 (2)

Generalization of psychological potential field

面図形の場合、電荷は輪郭線のみ分布

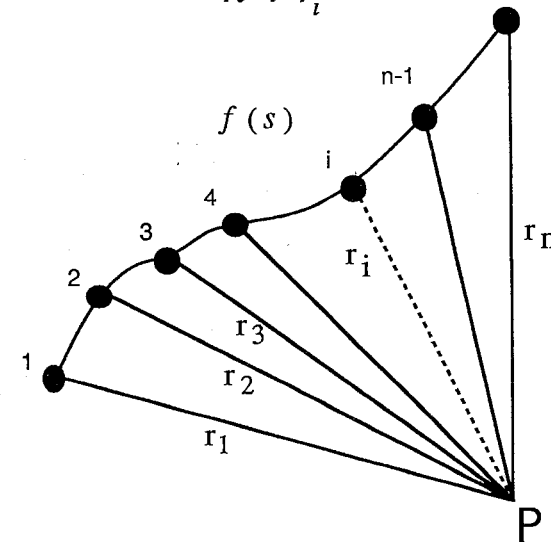
$$M_p = \int_{l_1} \frac{ds}{R_1(s)} + \int_{l_2} \frac{ds}{R_2(s)}$$



面図形のポテンシャル場
The potential field of a two-dimensional figure.

デジタル画像の場合

$$M_p = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i}$$



点列のポテンシャル場
The potential field of an array of dots.

心理的ポテンシャル場の性質

The nature of psychological potential field

心理的ポテンシャル場の分布と人間の文字パターンの読み方

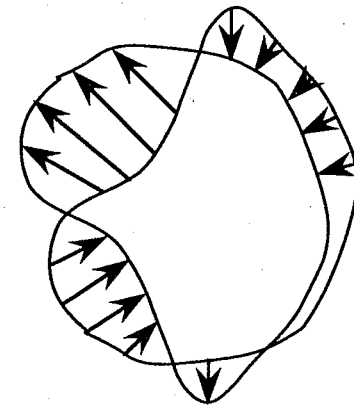
The distribution of psychological potential field and reading of characters



同一カテゴリー 「場」の分布が似ている

文字のトポロジー性を失うことなく
文字変形を行ない、一致できる

別カテゴリー 「場」の分布が異なっている



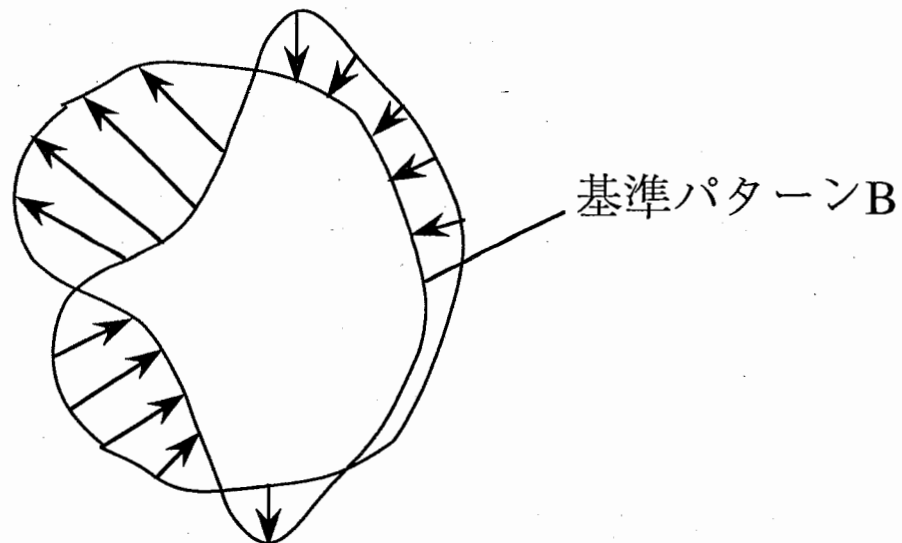
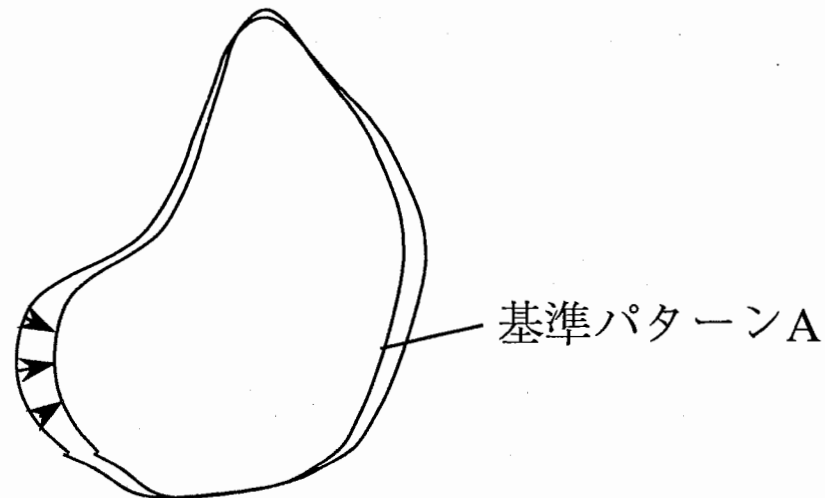
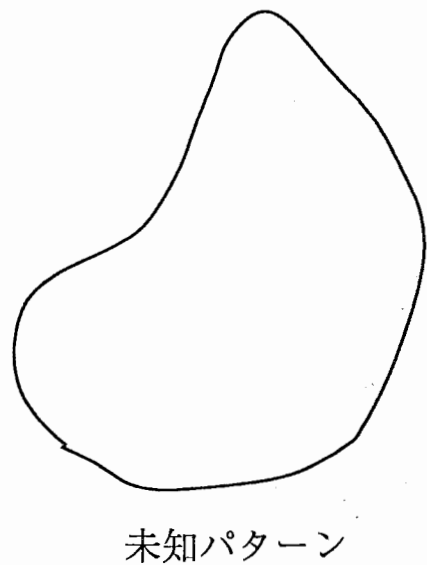
ATR

Advanced Telecommunications Research Institute International

ポテンシャル場を用いた識別

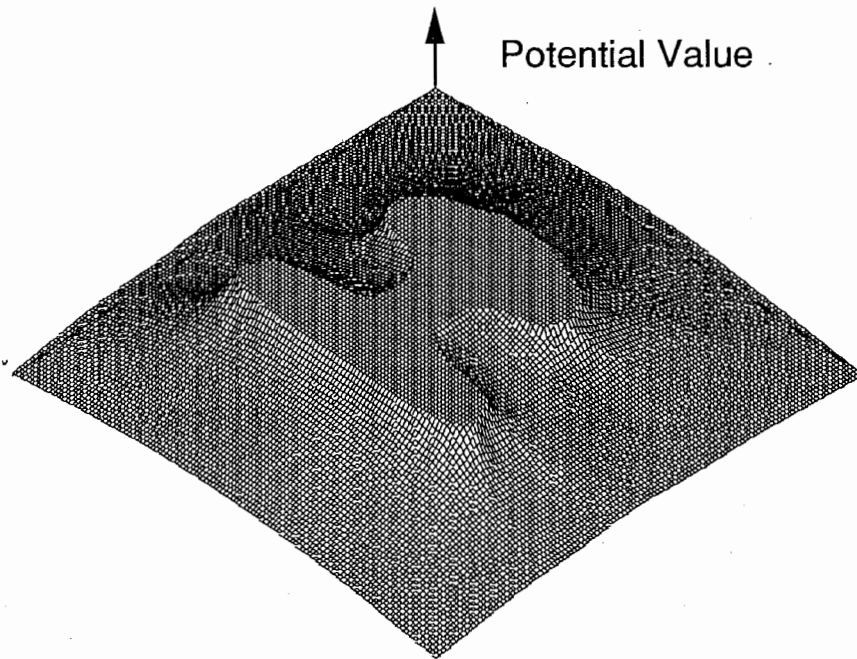
Recognition based on the potential field

3次元の弾性体と見なす

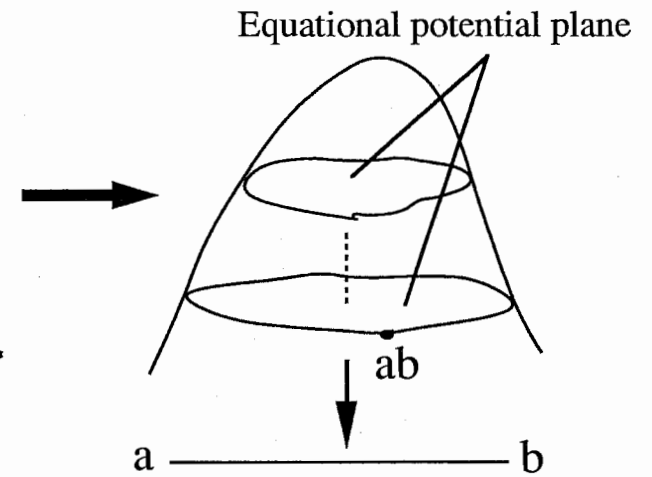
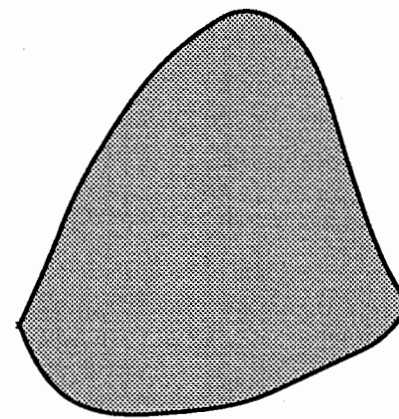


3次元図形を1次元で扱う

Consideration of a 3-dimensional figure as a 1-dimensional array



3-dimensional representation
of the field '土'.



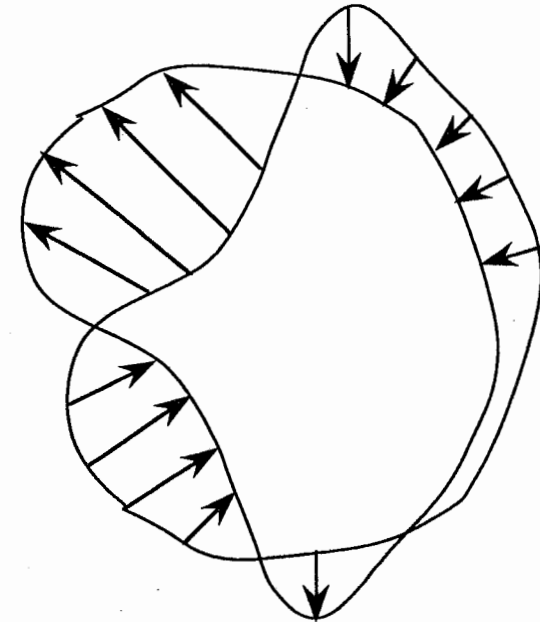
弾性エネルギー

Elastically energy

相異なる2つの図形がつくる等ポテンシャル面で、一方の輪郭線が他方の輪郭線に一致しようと変形した場合のエネルギー



各輪郭線に張られた膜が、変形によって生ずる弾性エネルギー $U = U_1 + U_2$



膜の変形に要する仕事

$$U_1 = \alpha \iint (u_x^2 + u_y^2) dx dy \quad (\alpha = \frac{\tau}{2})$$

輪郭線が変形に要する仕事

$$U_2 = \frac{\beta}{2} \int u_{ss}^2 ds \quad (\beta = \frac{\tau^2}{\rho^2} C)$$

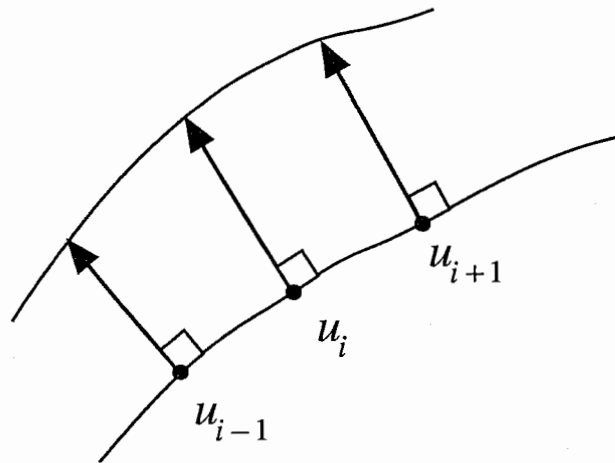
u_i : 変位ベクトル

ρ : 膜の密度

τ : 膜の張力

等ポテンシャル面の変位の計算

Calculating border line displacement on an equational potential plane



$$u_i = u(x(i), y(i))$$

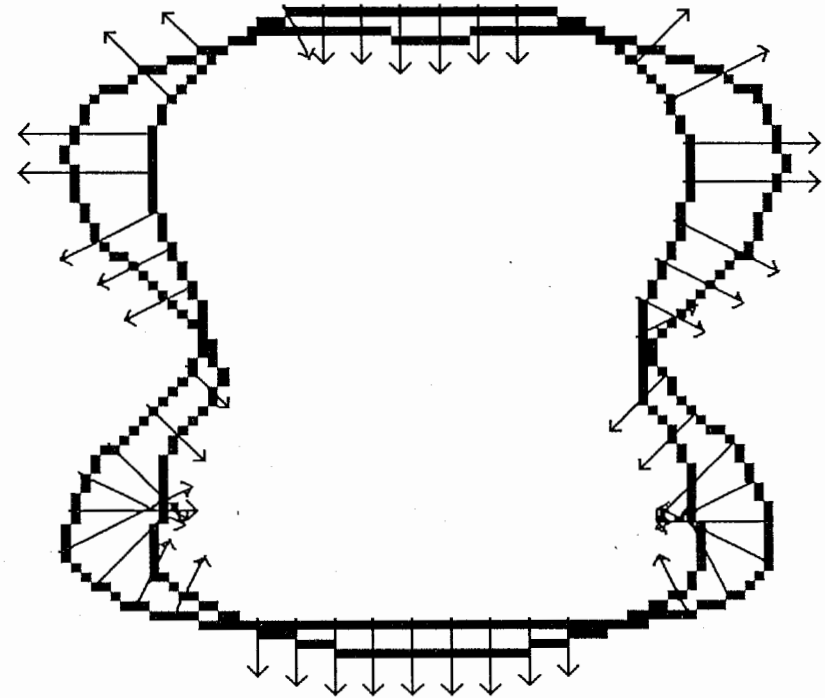
$$u_x = \partial u_i / \partial x = x(i) - x(i-1)$$

$$u_y = \partial u_i / \partial y = y(i) - y(i-1)$$

$$u_{xx} = \partial u_x / \partial x = x(i+1) - 2x(i) + x(i-1)$$

$$u_{yy} = \partial u_y / \partial y = y(i+1) - 2y(i) + y(i-1)$$

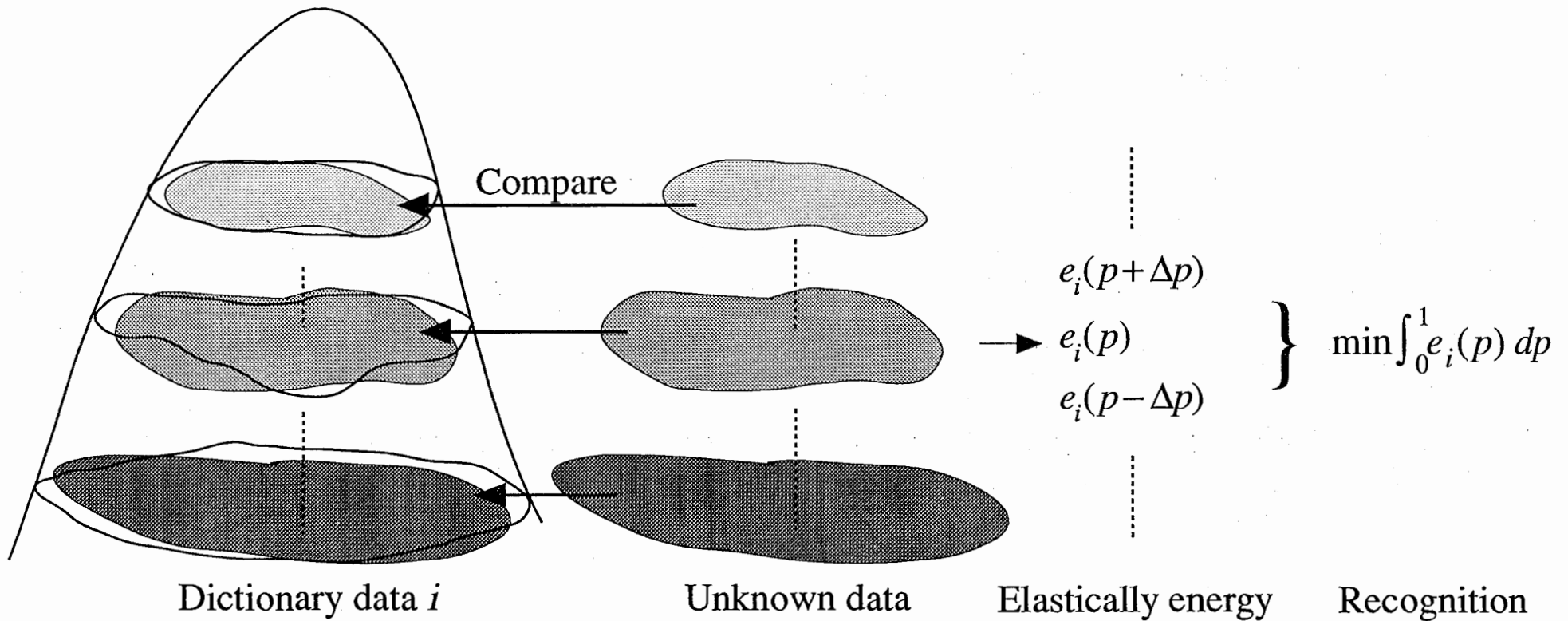
$$u_{ii}^2 = u_{xx}^2 + u_{yy}^2$$



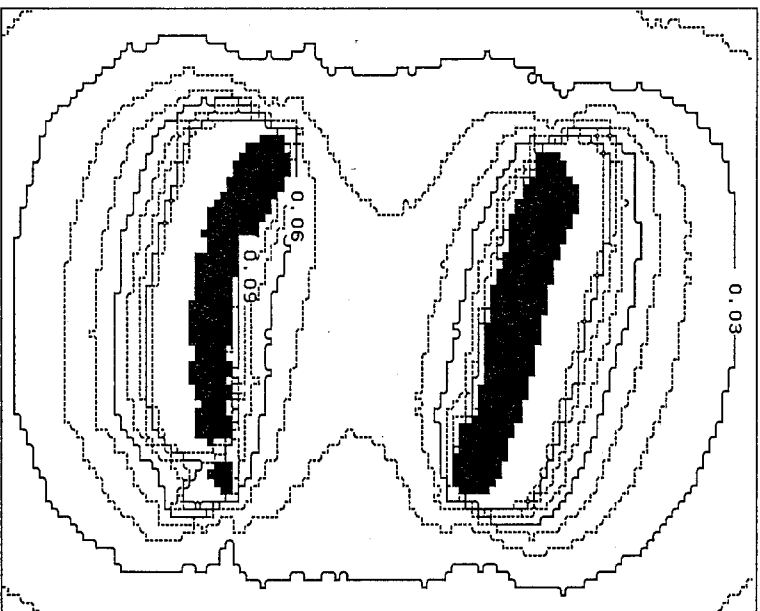
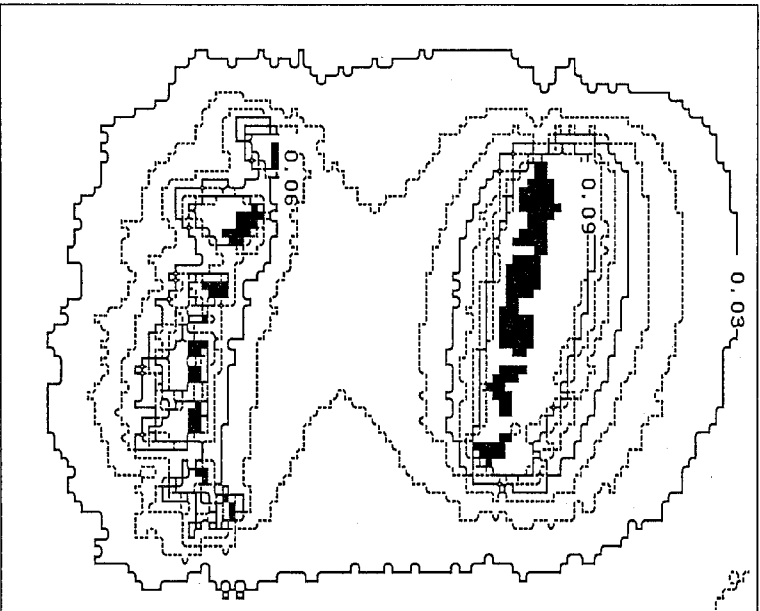
弾性エネルギーを指標にした識別

Recognition by evaluating elastically energy

等ポテンシャル面毎に変形による弾性エネルギーを求める

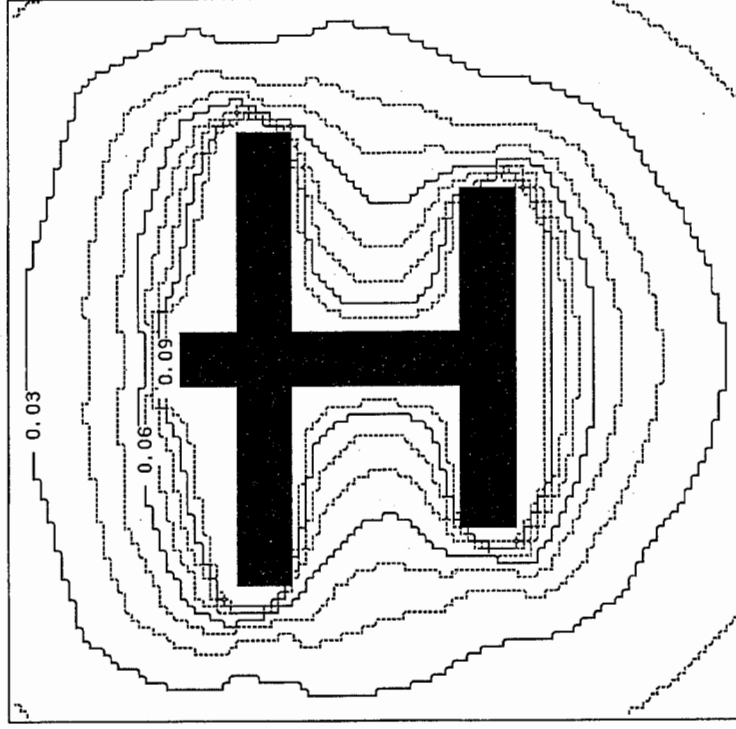
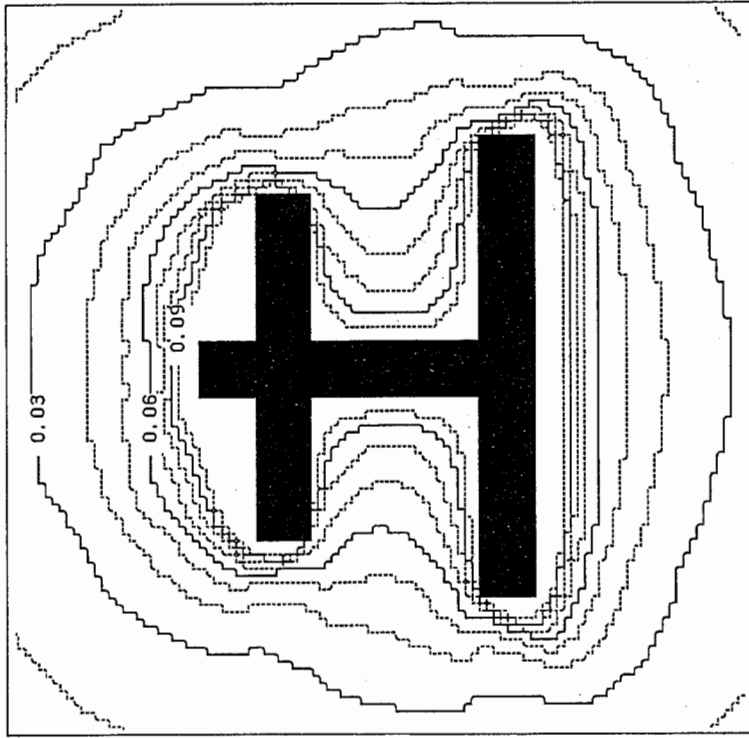


手書き文字「こ」のポテンシャル場 Potential field of the handwritten character 'こ'



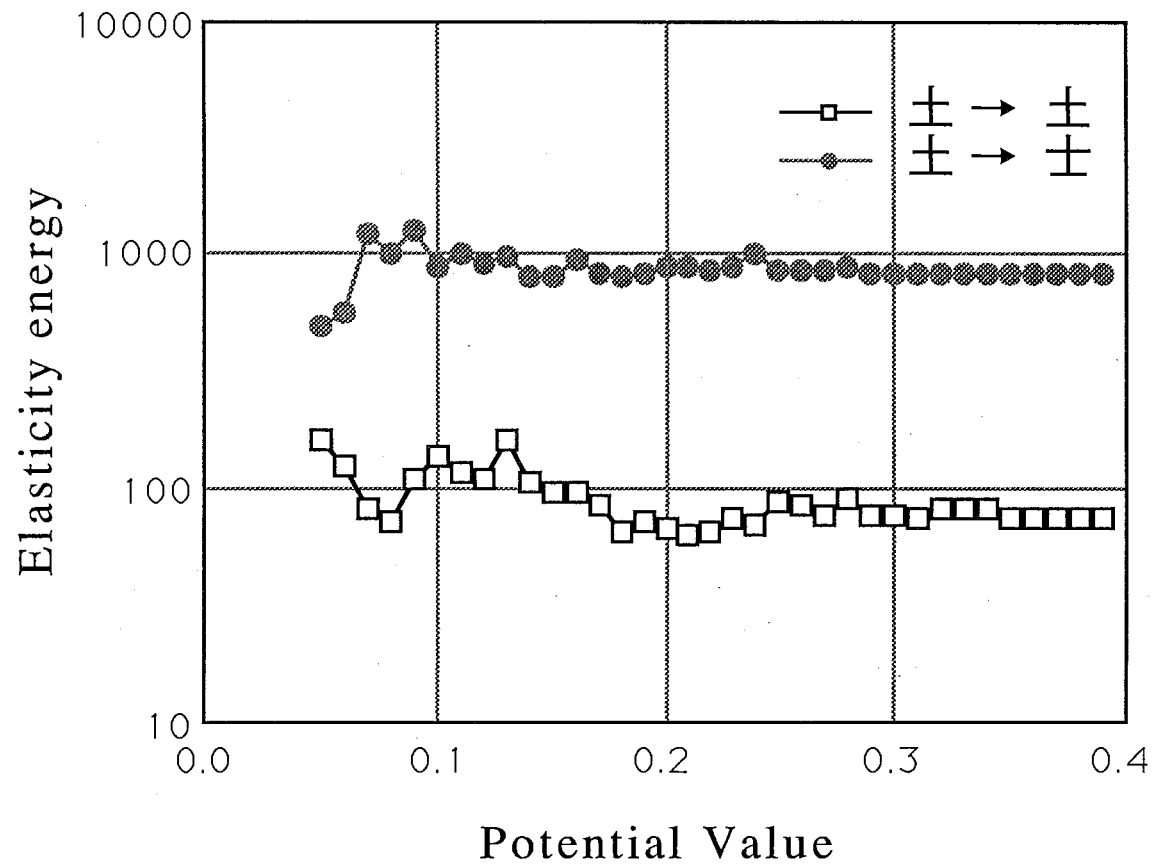
「士」と「士」のポテンシャル場

Potential fields of '士' and '士'



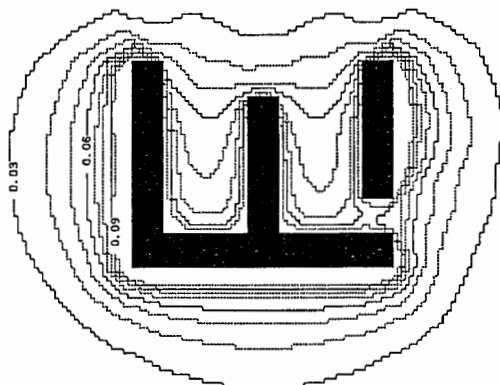
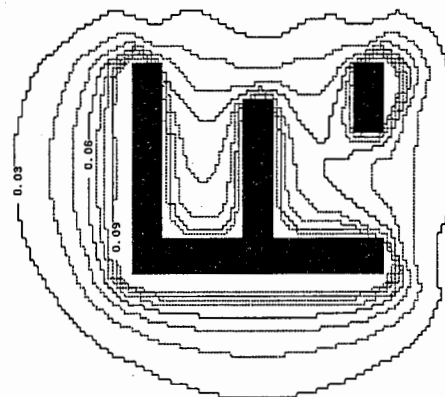
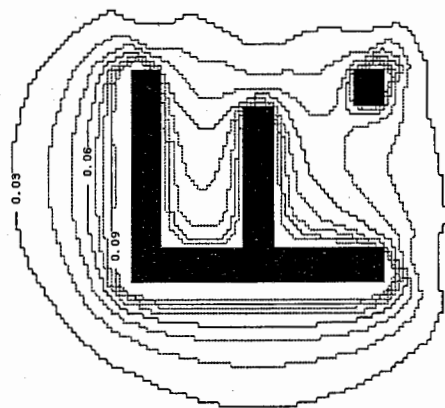
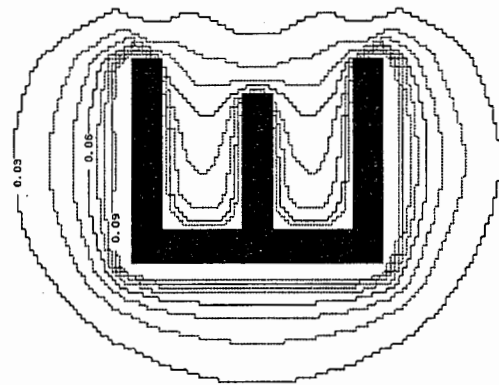
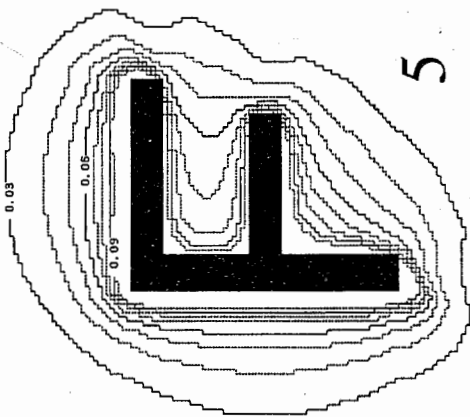
変形に要する弾性エネルギー

Elasticity energy from deformation of the potential field



各パターンのポテンシャル場

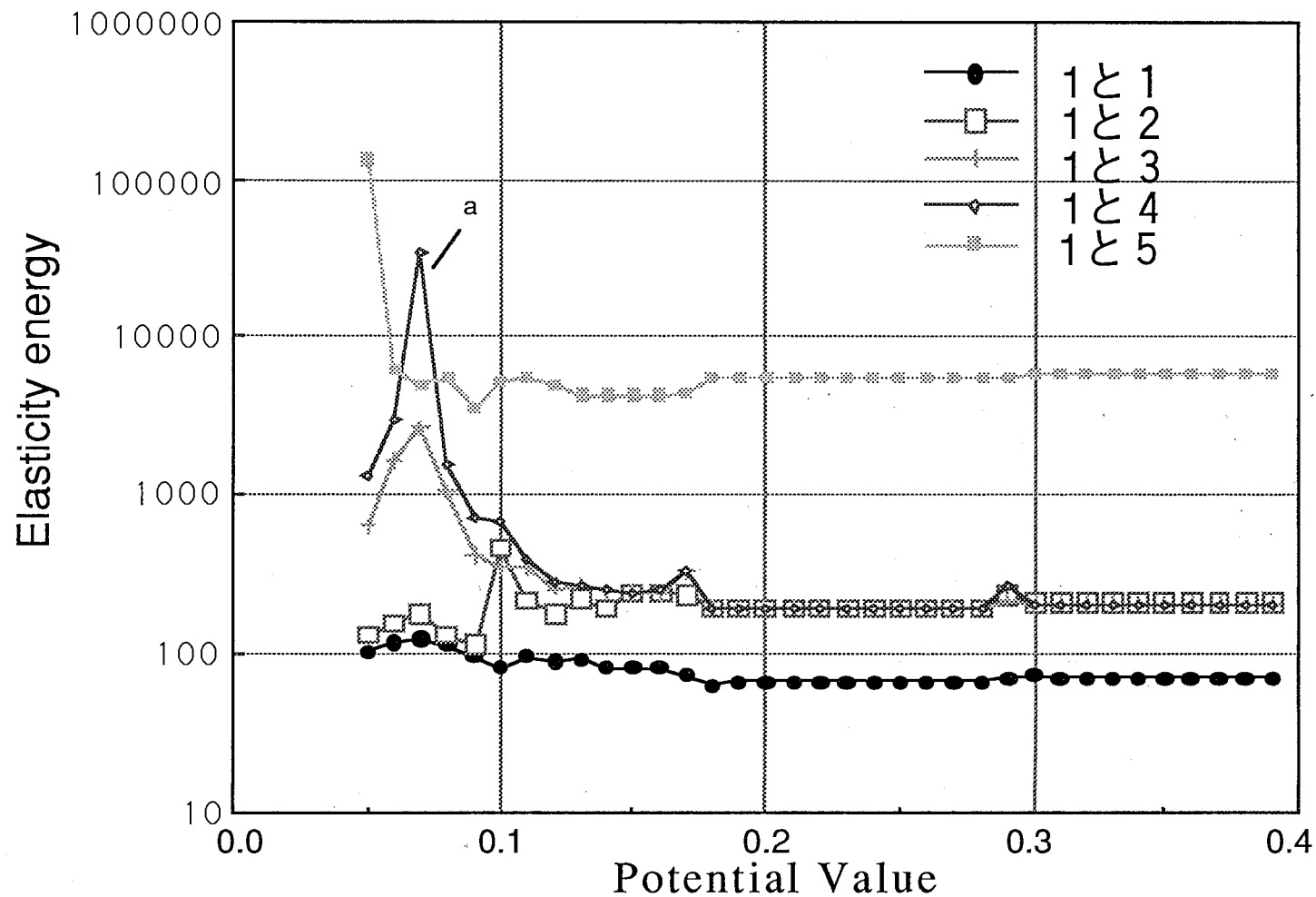
Potential fields of various patterns



4 ATR

変形に要する弾性エネルギー

Elasticity energy from deformation of the potential field

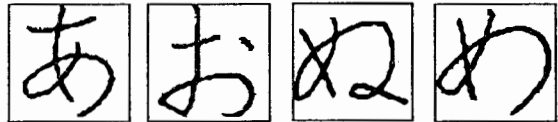


辞書と未知文字間の弾性エネルギー

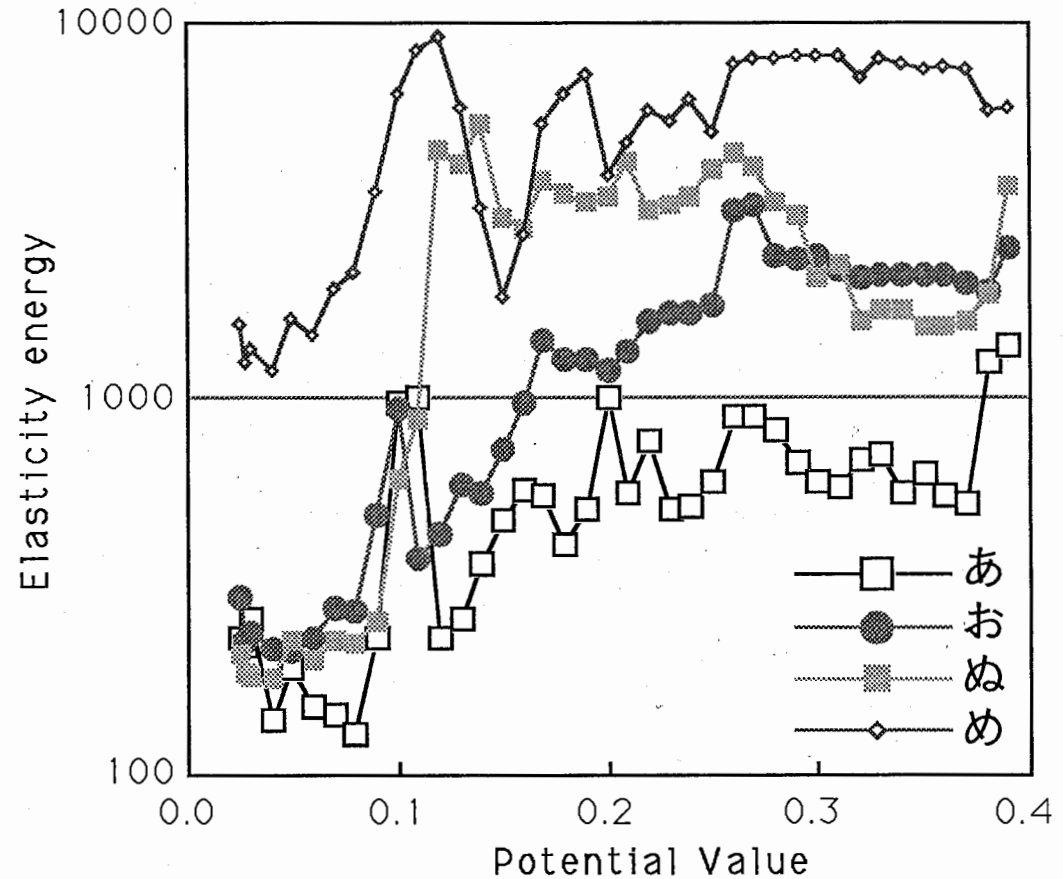
Elasticity energy between dictionary characters and an unknown character



Unknown character *



Dictionary character **



* ETL9を使用

** 「現代書道字典」(木耳社刊)

まとめ

視覚の誘導場理論による文字認識の提案

生理学、心理学的知見

ポテンシャル場の推定

場の分布と人間による文字の読まれ方の類似性に着目

場の違いを、変形による弾性エネルギーで評価

誘導場による文字認識の可能性

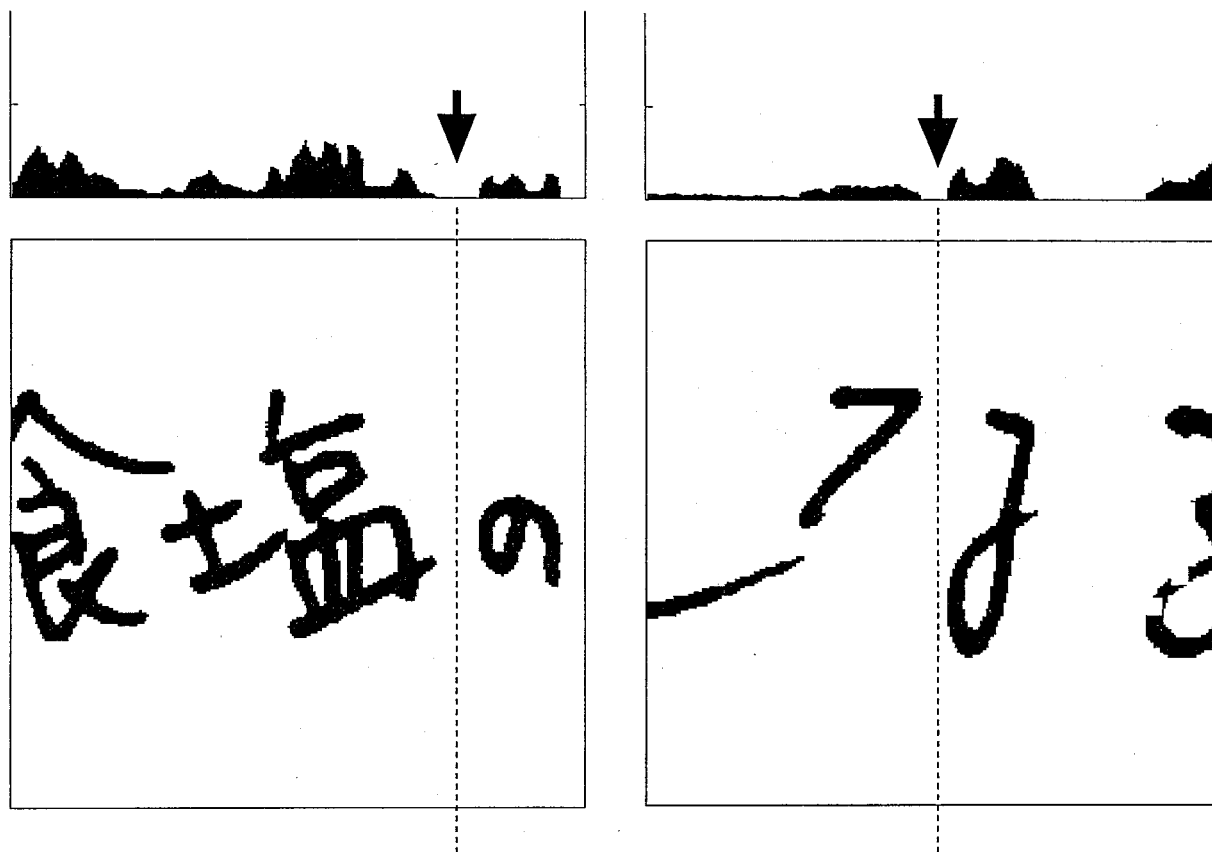
弾性エネルギーによる文字の類似性、変形の定量表現

文字変形等の視覚的把握が容易

-
- 1、射影による文字列の切り出し
 - 2、文字間隔と視覚の誘導場
 - 3、手書き文字の誘導場の分析
 - 4、文字切り出し方法の提案
 - 5、実験
 - 6、今後の課題

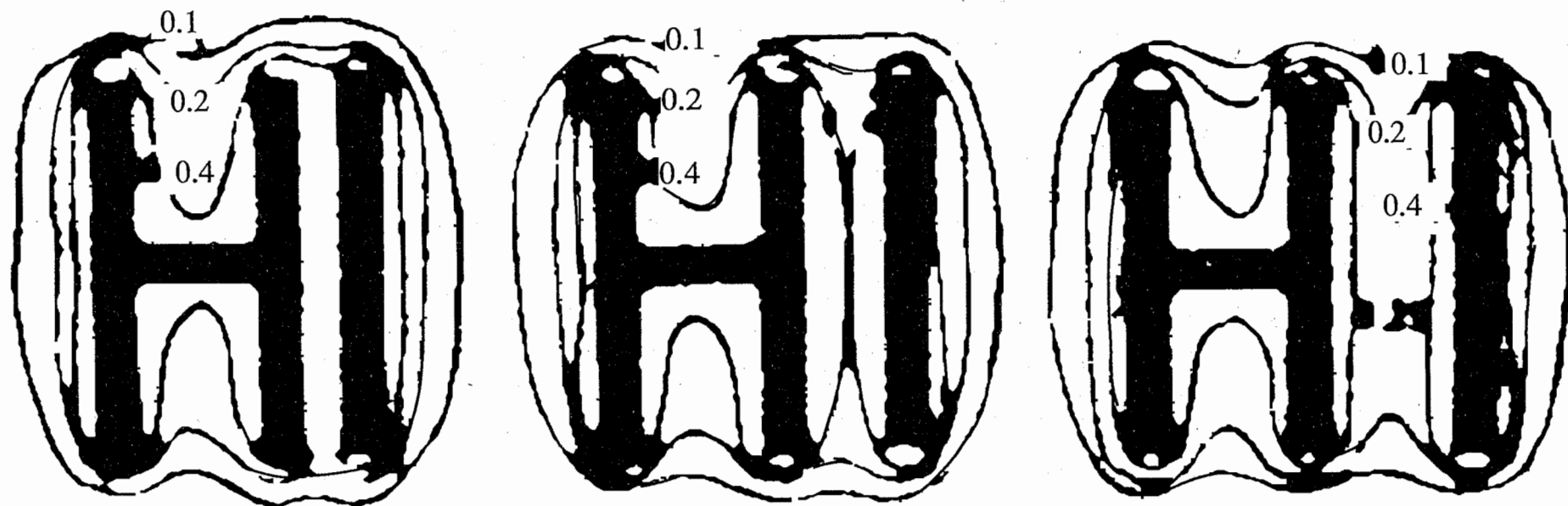
手書き文字の射影

Projection of handwritten characters



文字列の視覚の誘導場

The field of induction on the retina of characters. (YOKOSE, 1970)



2次元的な領域による文字の切り出し

A method of segmenting that evaluates the area of a character

射影による1次元的な
切り出しの限界

新しい方法

2次元的な切り出し

手書き文字の視覚の誘導場の分析

文字領域の規定の仕方の基礎的な検討

視覚の誘導場による
文字認識

複数の文字による
視覚の誘導場

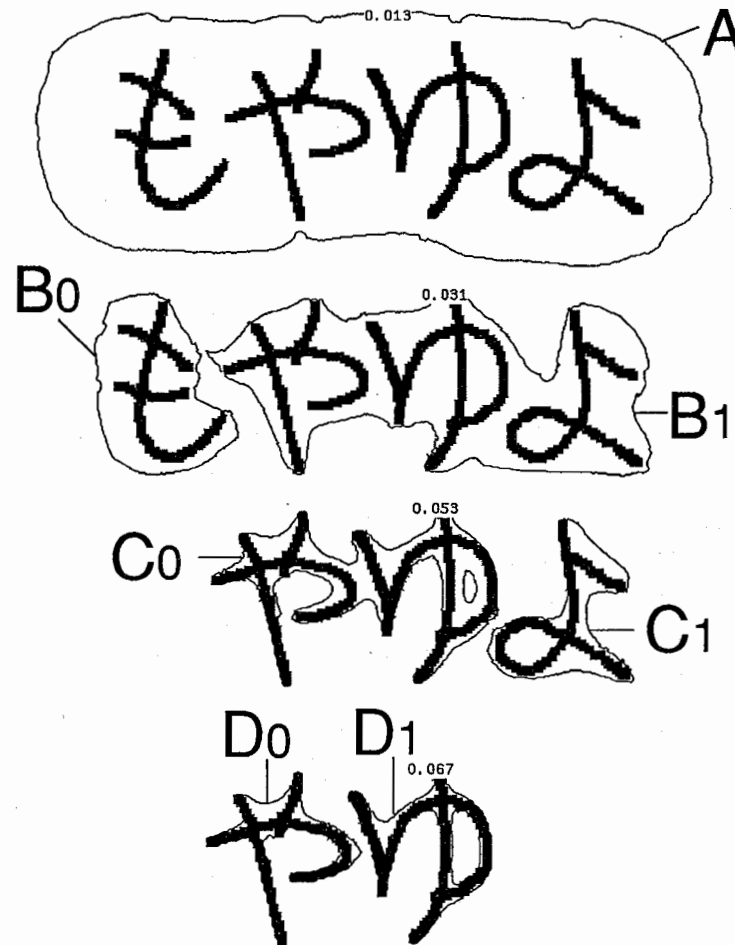
可能性

ATR

Advanced Telecommunications Research Institute International

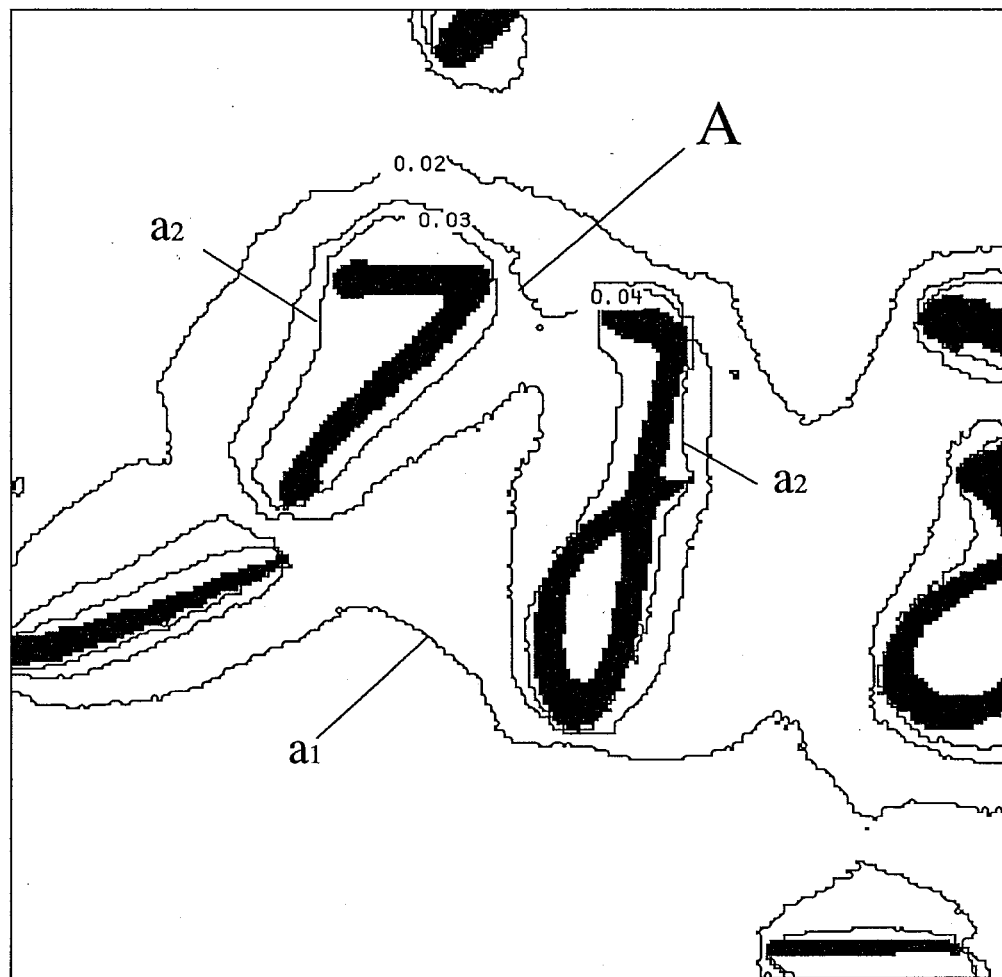
複数の文字による視覚の誘導場

The field of induction created by characters



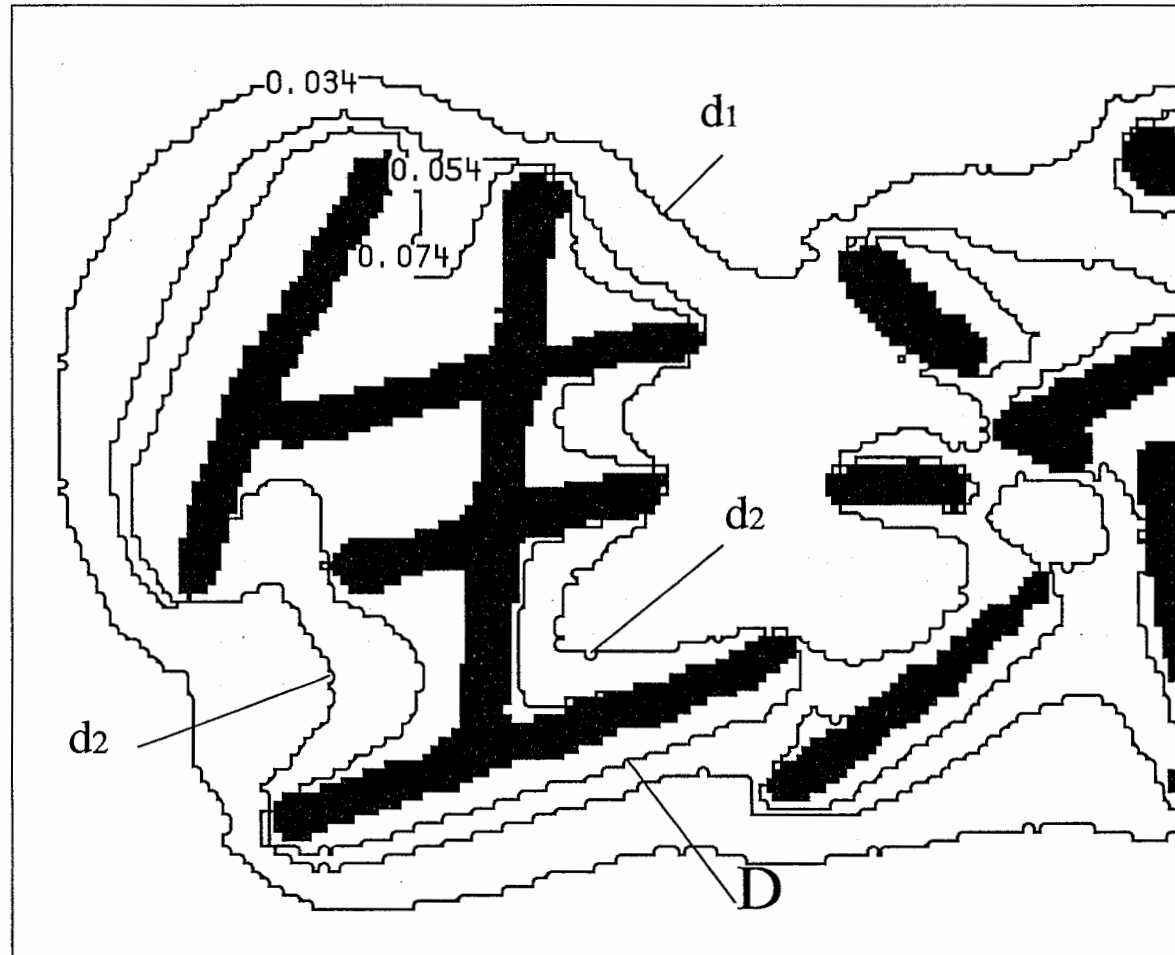
文字「な」の周辺の視覚の誘導場の分布

Distribution of the field of induction around 'な'.



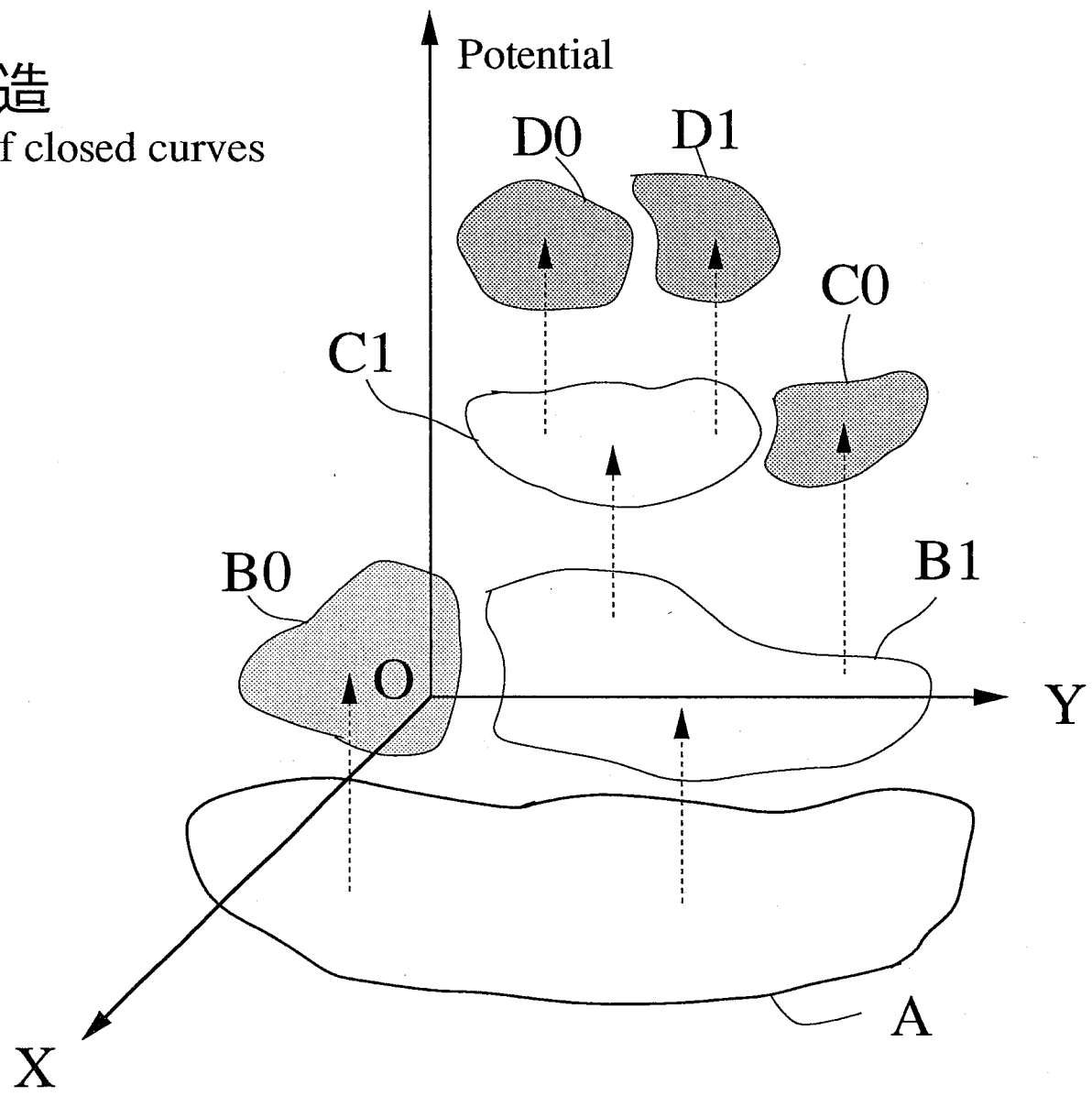
文字「生」の周辺の視覚の誘導場の分布

Distribution of the field of induction around '生'.



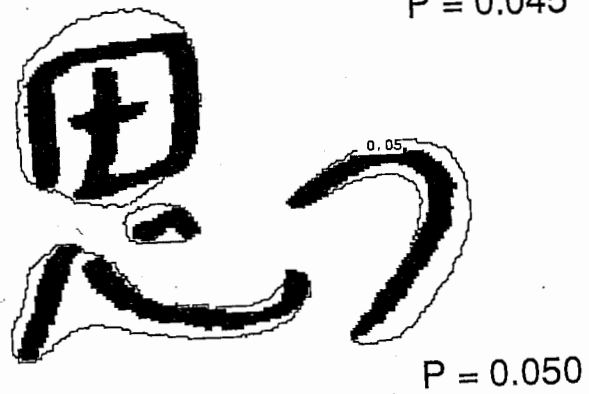
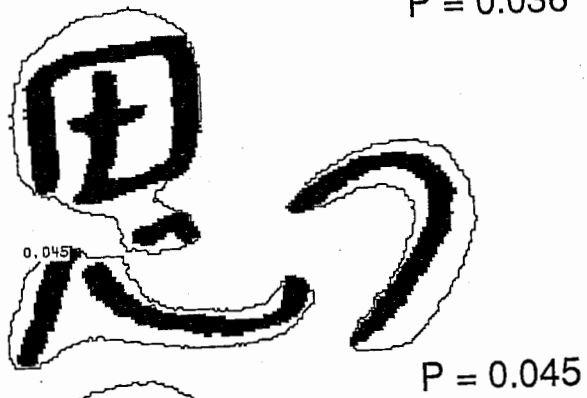
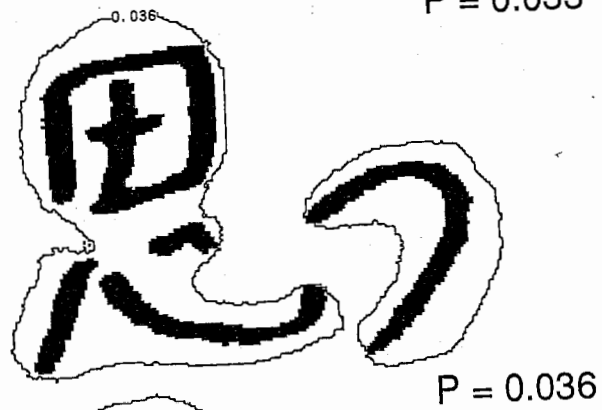
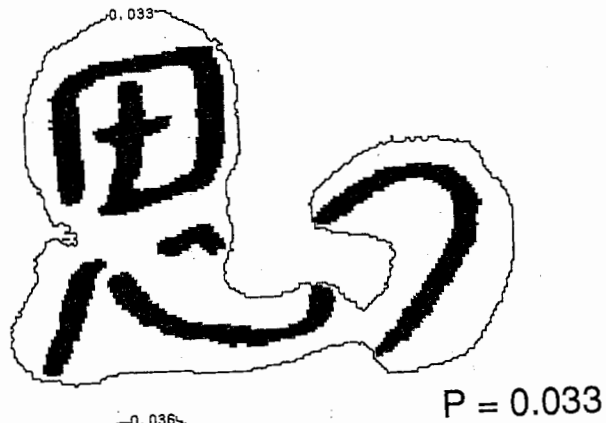
閉曲線の階層構造

Hierarchical structure of closed curves



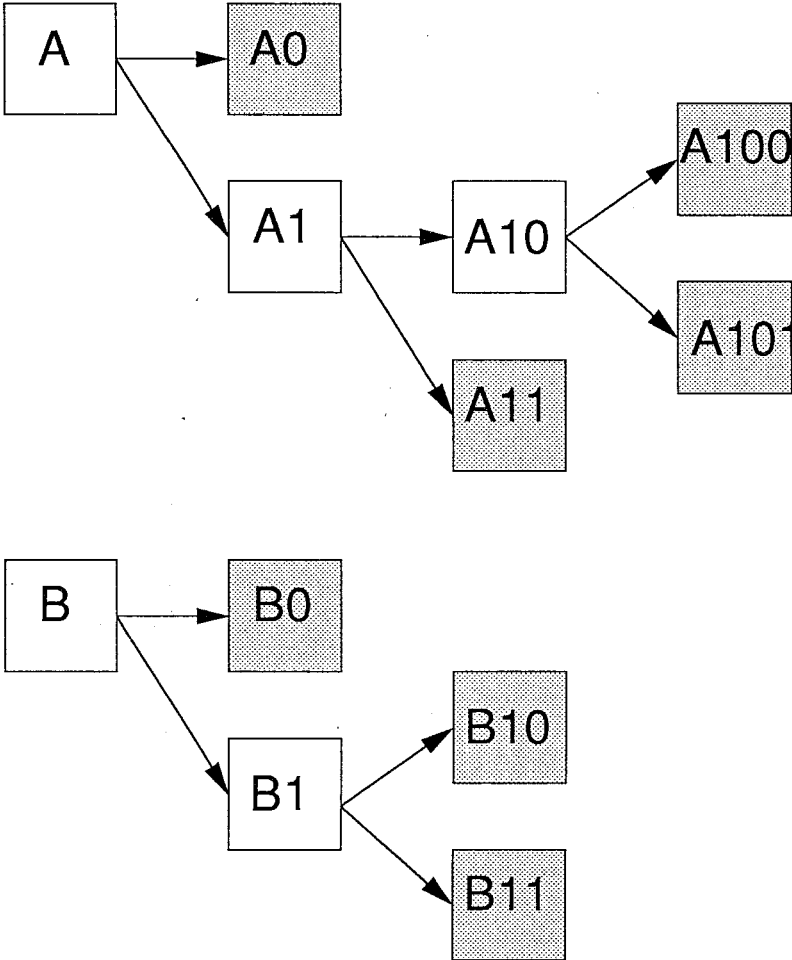
文字要素の分離

Separation of character elements



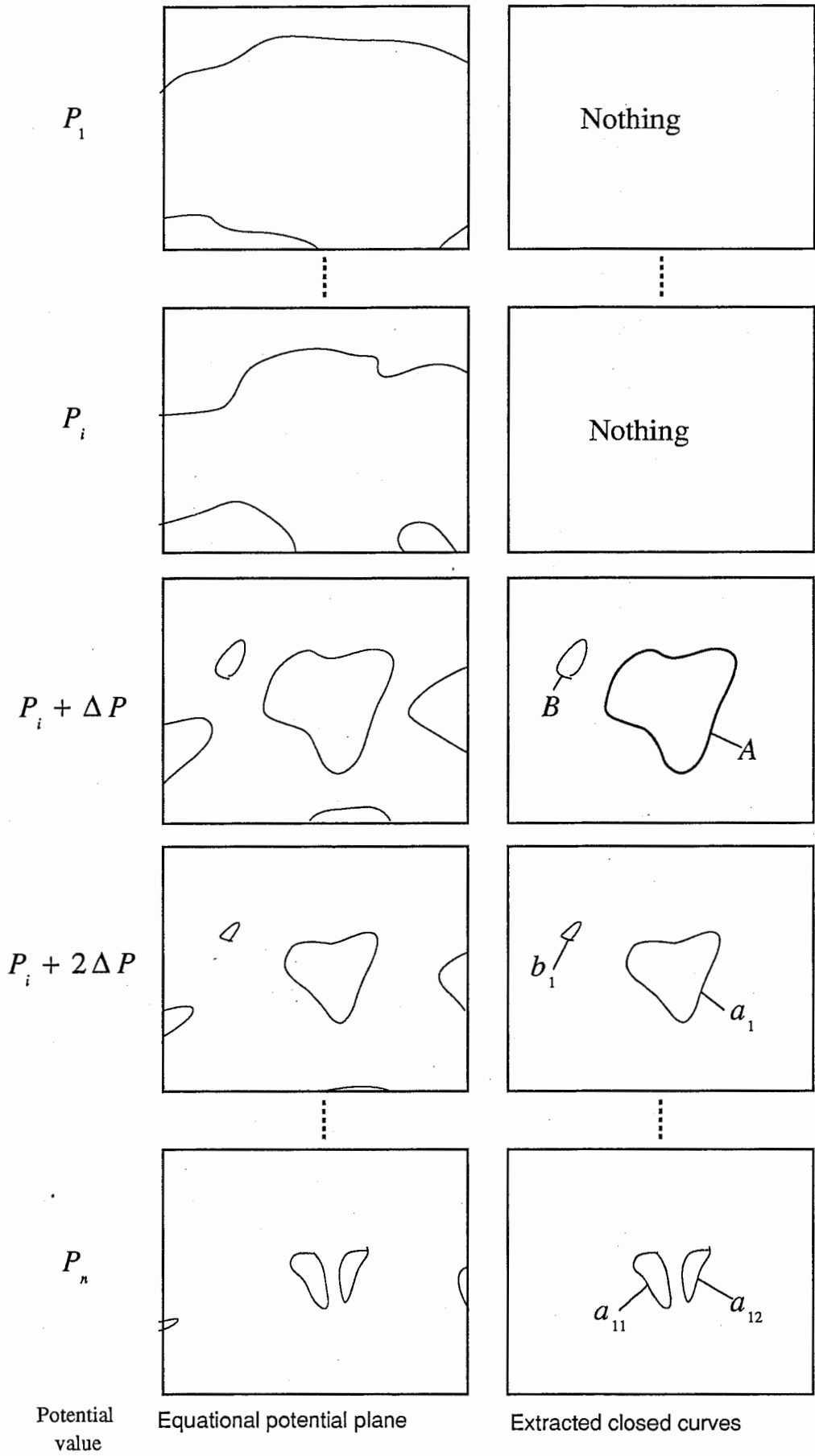
閉曲線の包含関係

Inclusion of closed curves



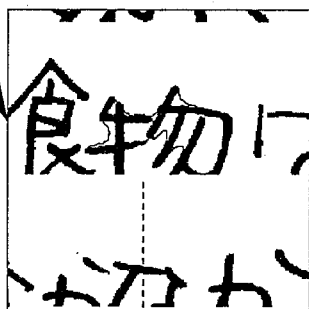
等ポテンシャル面からの閉曲線の抽出

Extraction of closed curves on equational potential planes.



活が豊かにな
含む食物は
にながなか

Document Image



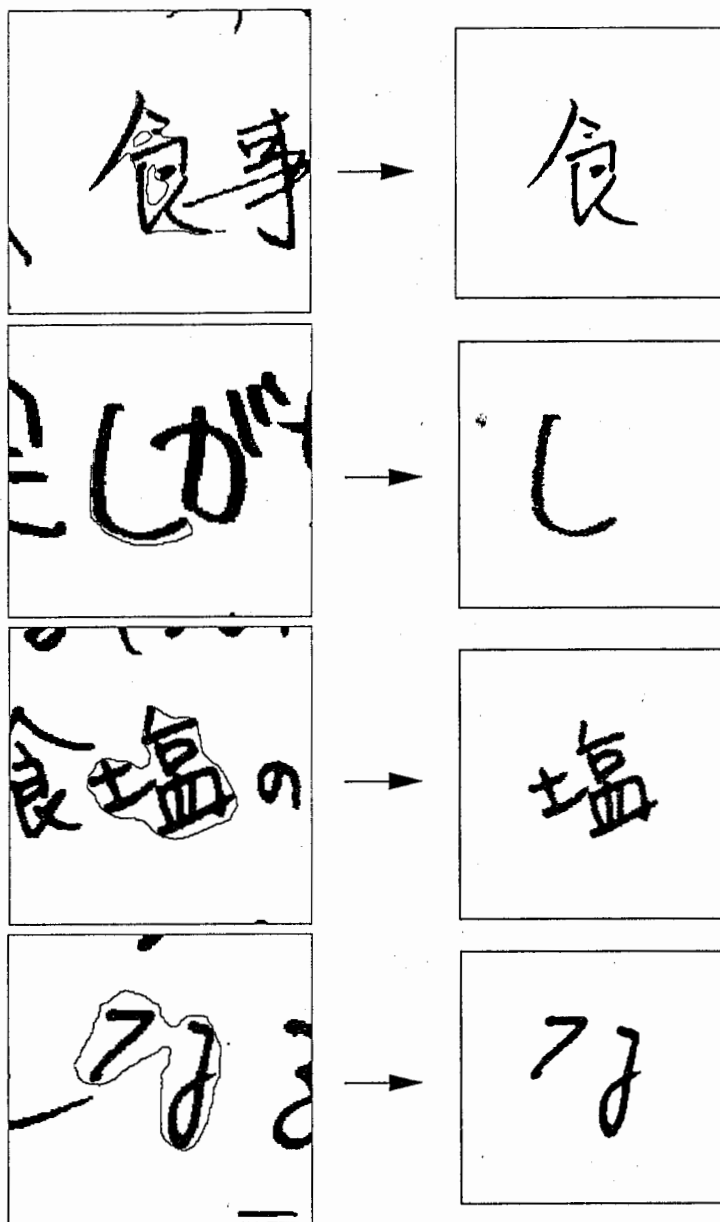
Extracted Closed Curve

物

Extracted Character

文字切り出し結果 (1)

Experimental results of segmenting characters

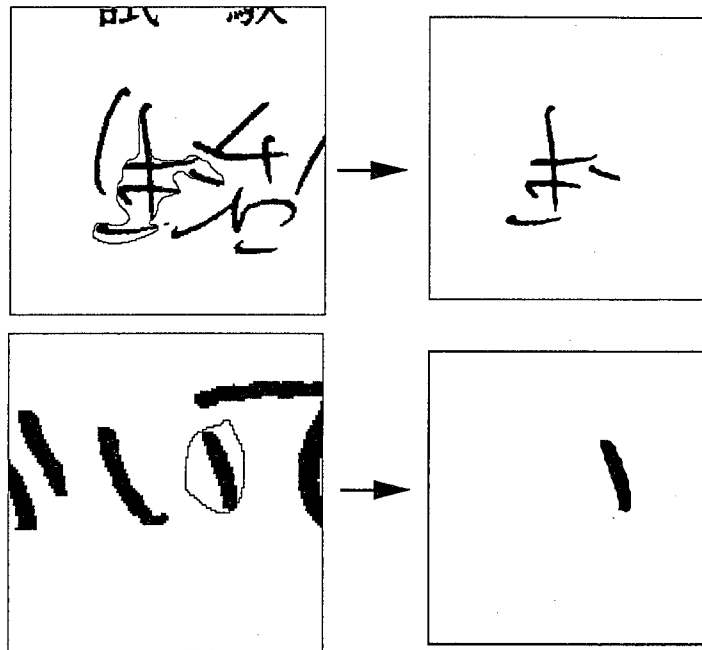


Extracted Closed Curve

Extracted Character

文字切り出し結果

Experimental results of segmenting characters.



文字切り出し結果 (失敗例)

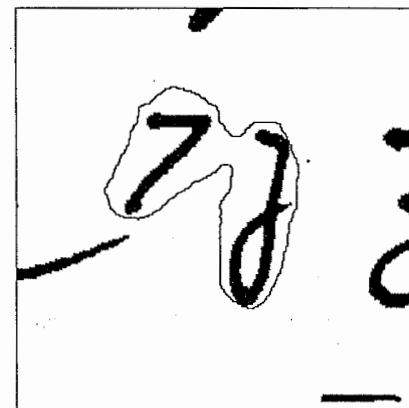
Experimental results of segmenting characters.

考察

複数の文字の視覚の誘導場

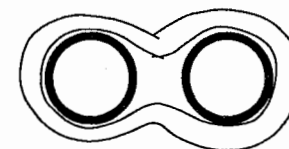
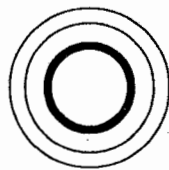
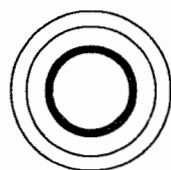
閉曲線群、包含関係が文字領域を規定

2次元的な文字の切り出しの優位性



なぜ閉曲線領域に文字や文字成分があるのか

誘導場の遠隔作用、遮蔽の性質



ATR

Advanced Telecommunications Research Institute International

まとめ

射影による1次元的な
切り出しの限界

複数の文字による
視覚の誘導場

2次元的な切り出し

↓

手書き文字の視覚の誘導場の分析

↓

視覚の誘導場理論による文字の切り出し
閉曲線群、包含関係が文字領域を規定する上で重要

↓

2次元的な文字の切り出しの優位性

今後の課題

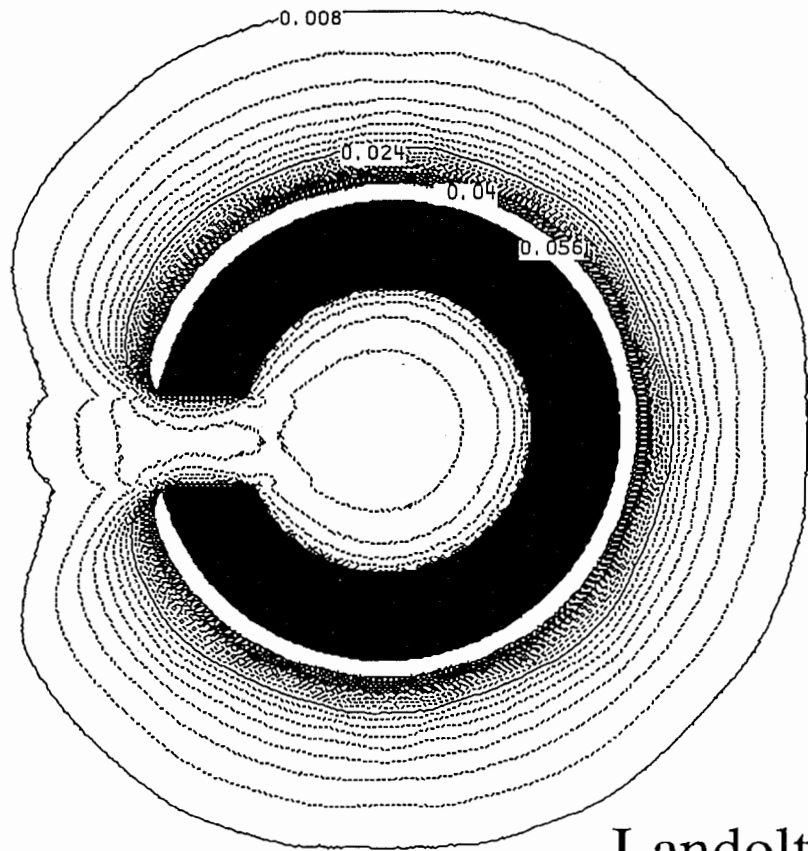
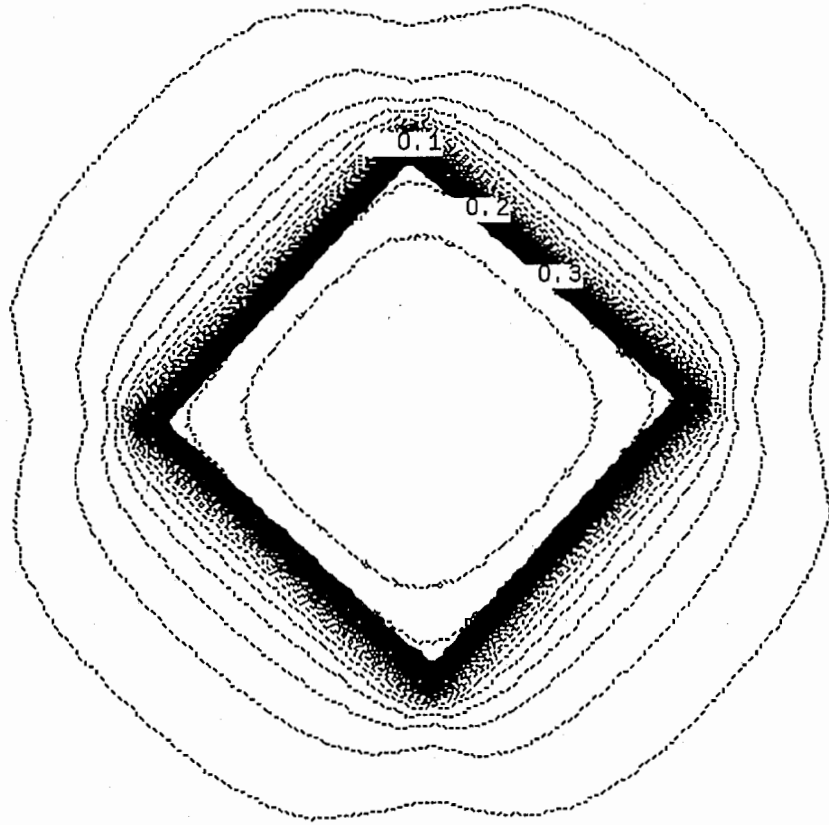
現状：視覚の誘導場の初歩的な応用に過ぎない



視覚の誘導場の理論面の整備

認識
切り出し

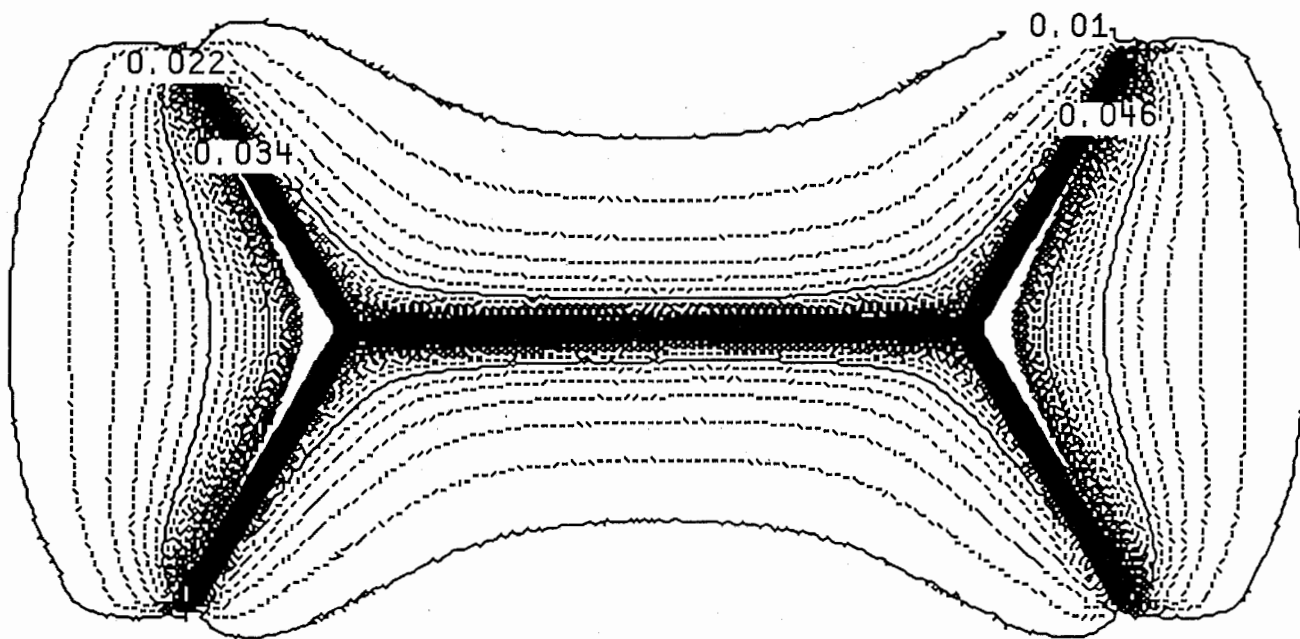
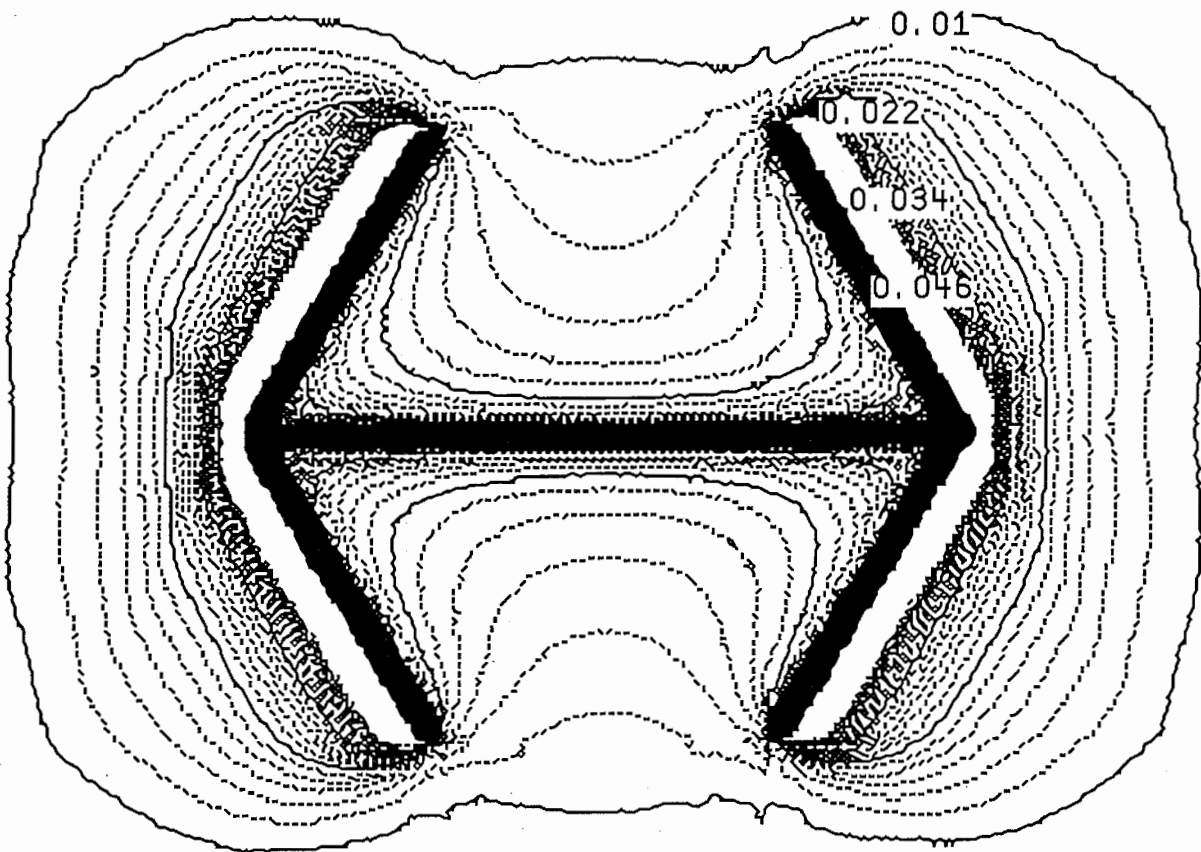
生理学的知見との対応付け



Landolt 環

ATR

Advanced Telecommunications Research Institute International

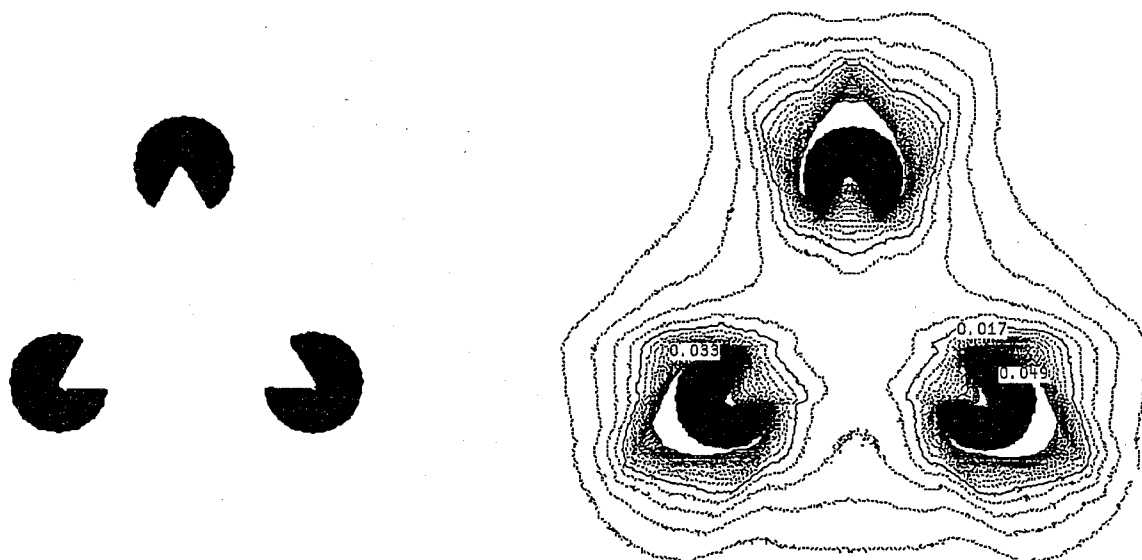


Müller - Lyer の図形のつくる誘導場

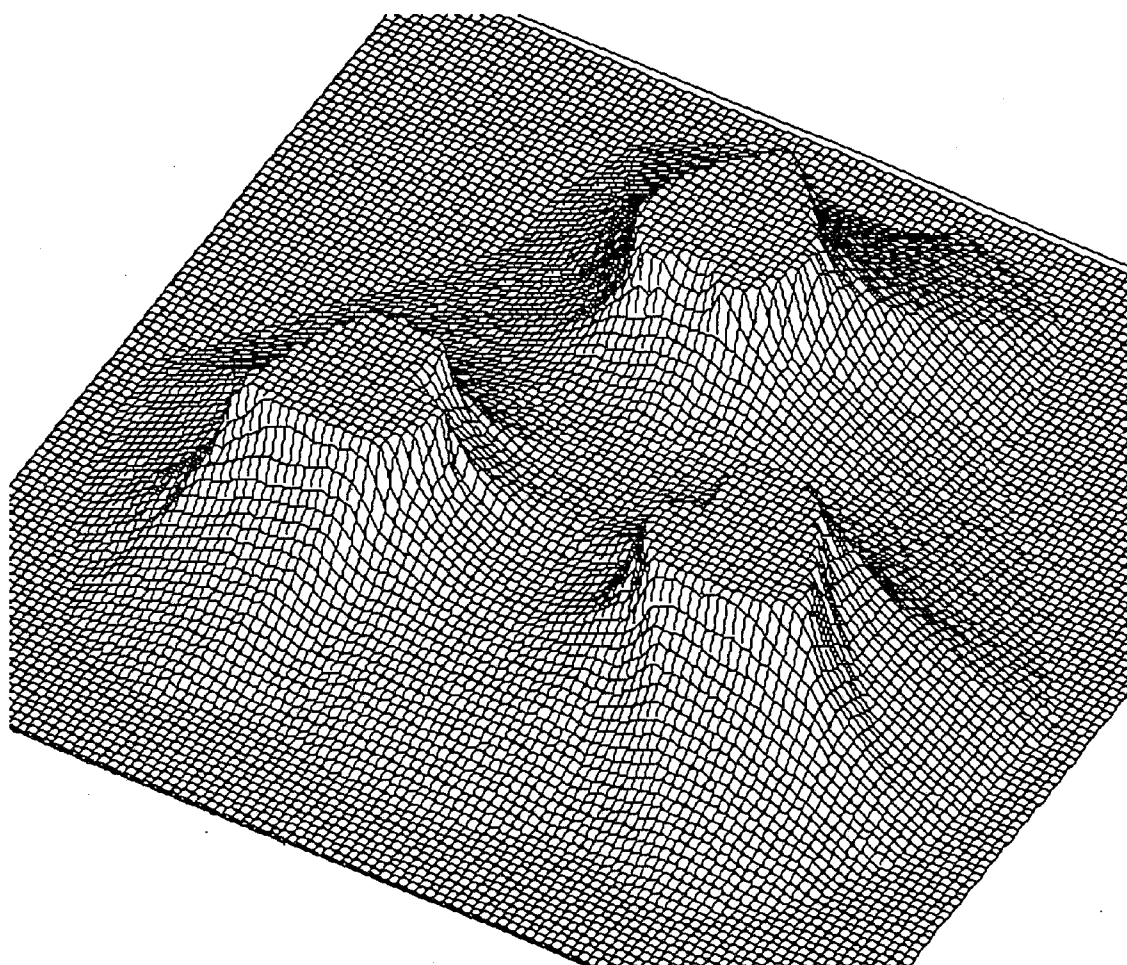
ATR

Advanced Telecommunications Research Institute International

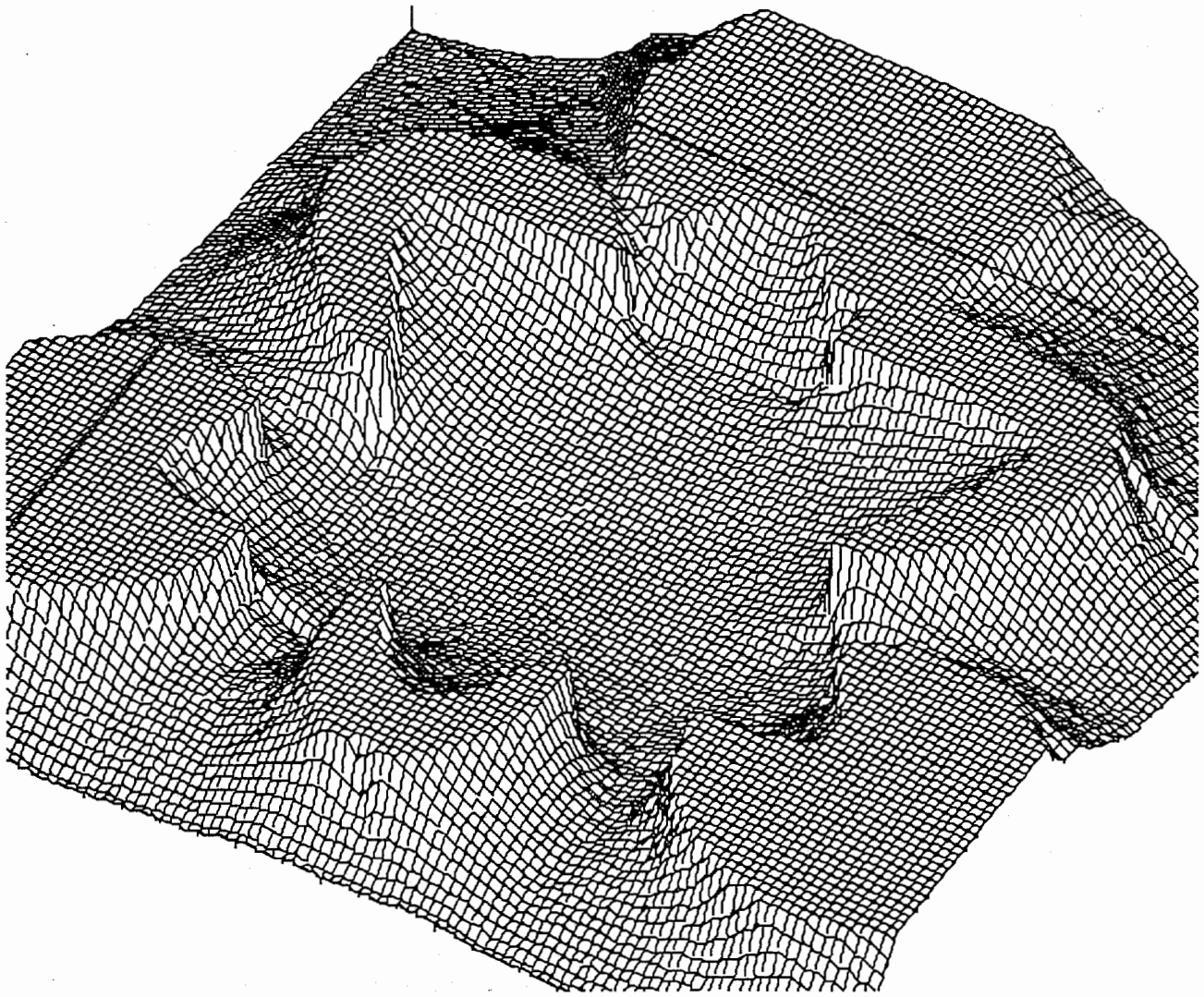
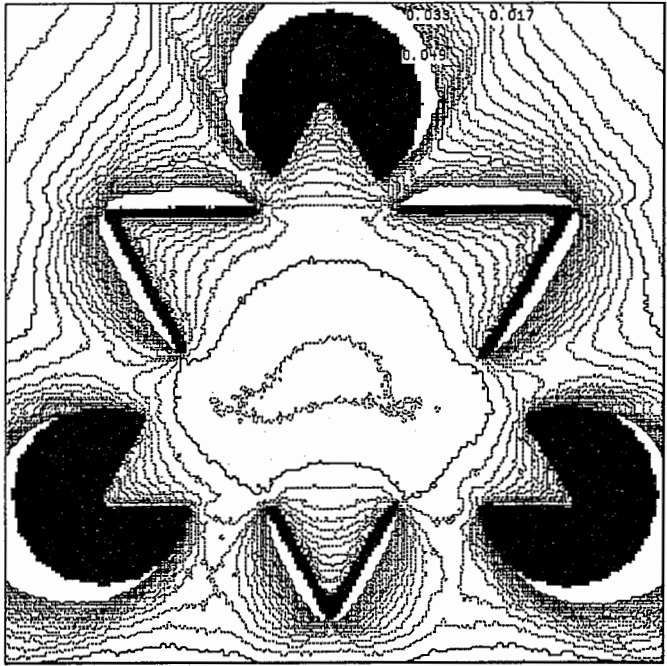
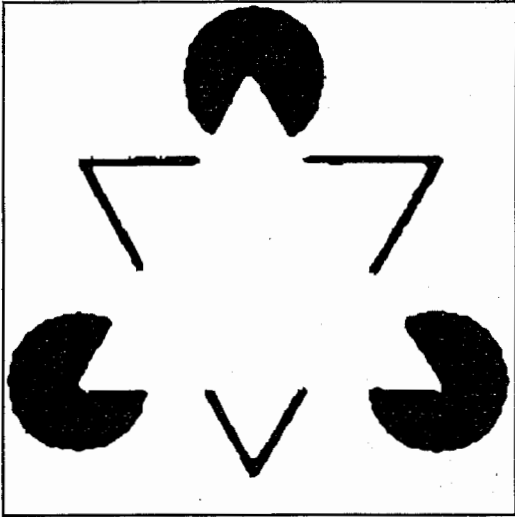
カニツツアの図形のつくる誘導場



Kanisza triangle image and its induction field

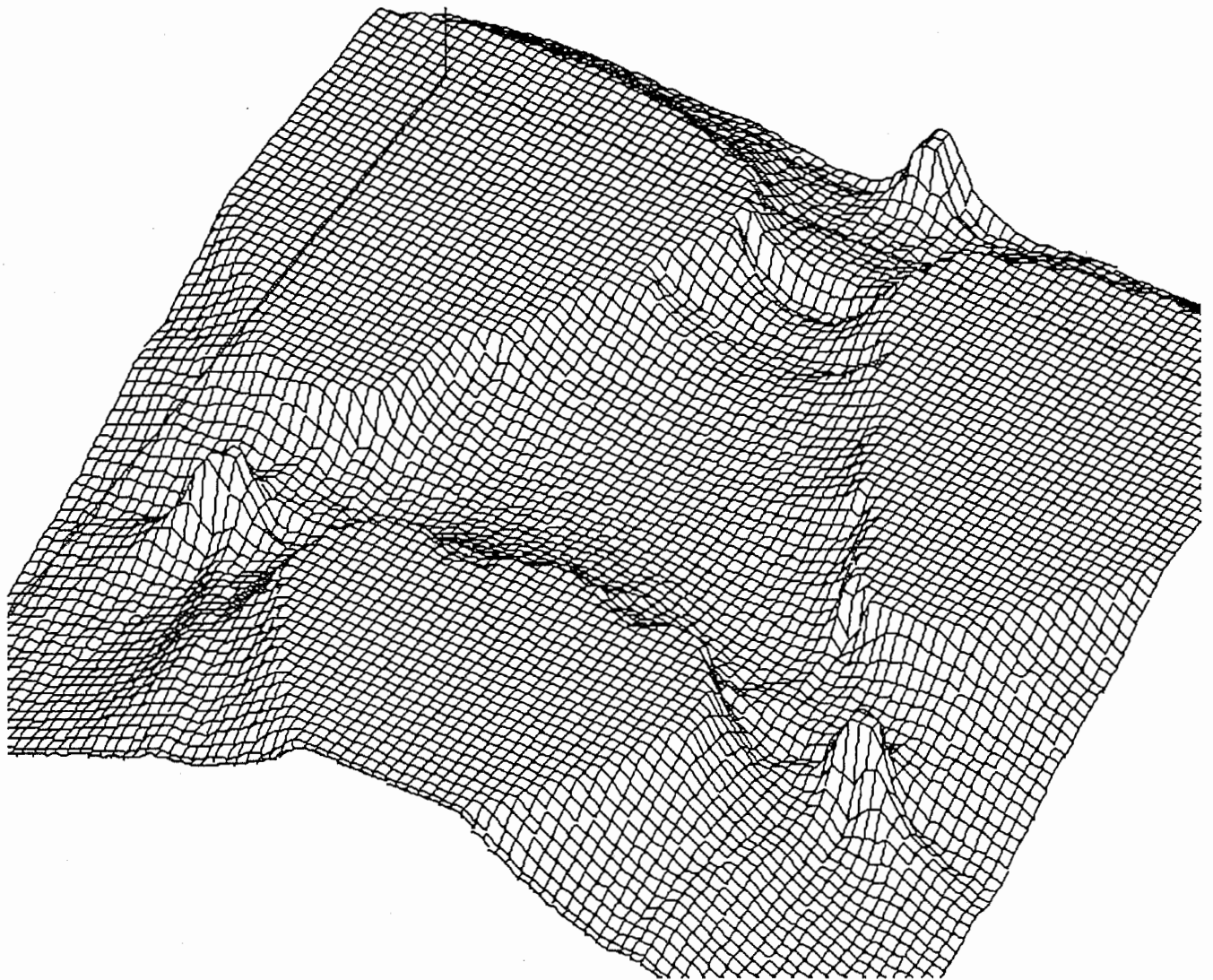
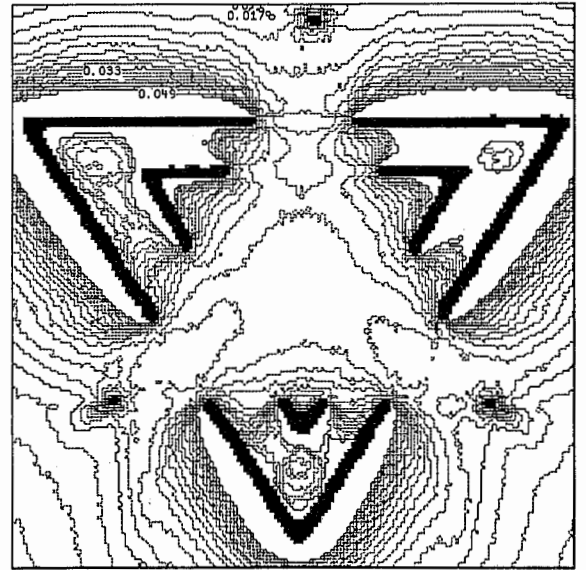
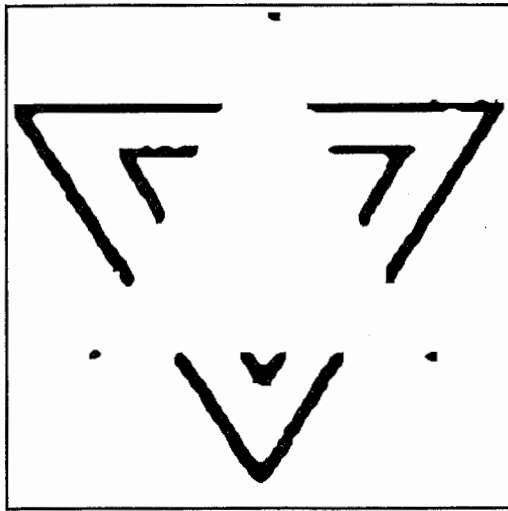


Three-deimensional displacement of the field of induction of Kanisza triangle image



ATR

Advanced Telecommunications Research Institute International



ATR

Advanced Telecommunications Research Institute International