

チュートリアル
画像解析入門
～簡単なこと、困難なこと～

国立情報学研究所

北本 朝展

<http://agora.ex.nii.ac.jp/~kitamoto/>

自己紹介

- 東京大学生産技術研究所(工学系研究科博士課程) 1992-1997
- 文部省学術情報センター 1997-2000
- 国立情報学研究所 2000-2006
- 大学院ではリモートセンシング画像処理および画像データベースの枠組みと設計に関する研究をおこなった。
- その後はメテオインフォマティクスやデジタルアーカイブの研究へ。

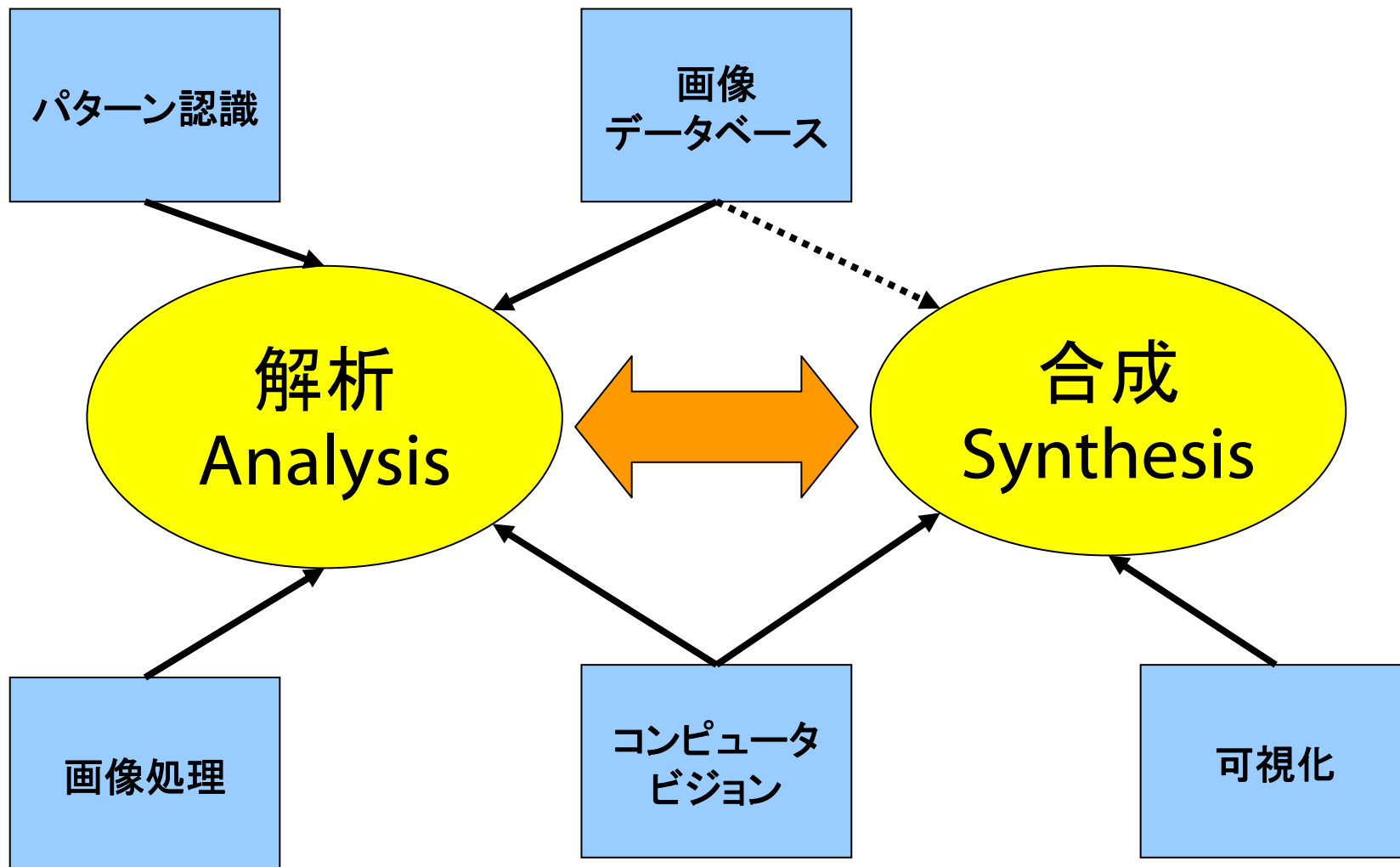
融合研究プロジェクト

- 情報研傘テーマ
 - 地球環境ポータル、極限環境生物データベース
- 遺伝研傘テーマ
 - 3D画像による体脂肪計測法
- 統数研傘テーマ
 - オーロラ画像処理
- 極地研傘テーマ
 - ??
- 育成融合テーマ
 - シミュレーションデータの可視化

画像解析から見た融合研究テーマ

- 情報研傘テーマ
 - 画像データベース、コンピュータビジョン
- 遺伝研傘テーマ
 - パターン認識
- 統数研傘テーマ
 - 画像処理
- 育成融合テーマ
 - 可視化(画像合成、コンピュータグラフィクス)

画像解析から見た融合研究テーマ



画像データベース

- 画像情報処理の中では、それなりに活発で他分野との接点も多い分野。
- 画像に付与された「メタデータ」を使って、データの選択や検索をすることが多い。
- 画像の内容を対象として検索することは、現在でもかなり難しい。
- 次々に登場するメディアとニーズに伴って、新しい研究者・企業の参入も多い。

コンピュータビジョン

- 画像情報処理の中で、最も数理的で正統的なアプローチを好む分野。
- 物体に電磁波が当たり、それを反射・散乱・吸収してカメラに到達し、それを(デジタル的に)記録して画像が得られる、という各過程を、数理的に厳密に記述したいのが研究の動機。
- 物体の3次元構造のモデル化・復元や、物体表面における電磁波の振る舞い、カメラキャリブレーションなどの問題に関心が強い。

パターン認識

- 画像情報処理の中で、工学的にも哲学的(?)にもテーマを広げられる分野。
- 広義には信号一般、狭義には画像に表現されているパターンを「認識」するための数理的な方法を追究したい。
- 実世界のノイズに満ちたパターンを扱うためには、モデルの適応や学習が必要。
- そもそも「パターン」とは何か？

画像処理

- 画像情報処理の中で、伝統的で中心的な位置を占めるものの、やや泥臭い分野。
- 画像信号の中から、コンピュータが扱いやすい特徴を抽出し、信号品質の改善や特徴づけをおこなう。
- 画像内容には立ち入らない場合が多い。
- ある処理に有効な特徴は、必ずしも人間が認識しやすいものとは限らない。

可視化

- 狭い意味での画像情報処理には含まれず、コンピュータグラフィクス(CG)に近い分野。
- ただし実データを扱うのがCGとは異なる。
- 画像情報処理は解析(Analysis)あるいは帰納が中心であるのに対し、可視化は合成(Synthesis)あるいは演繹が中心。
- より大きな視点では、解析と合成が相補的に組み合わさって「画像情報学」を構成する。

実際の画像の例



東京より電車でわずか40分。
思えば、ぶらり熱海へ小旅行。

<http://www.kkratami.com/>

さて画像解析は困難な問題なのか？

- デジタル画像は高次元データであり、的確に解析するには情報が不足している。
- ところが、なぜか、人間はこの困難な問題を何なく簡単に解いている。
- どうやら動物もそれなりに問題をうまく解いているようだ(現にちゃんと生きている)。
- こんな簡単なことが、なぜできないの？
- 正直、途方にくれてしまう。。。

実世界の縮退

- 画像は3次元の世界を2次元の平面に記録したものであり、記録の過程で多くの情報が失われてしまっている。
- 画像は電磁波の強度の分布を記録したものであるが、同様の強度分布を生成する実世界の状況が複数存在する。画像が暗いのは、物体が黒いのか、照明が暗いのか？
- したがって、一意に正しい画像の理解を得ることは困難である

高次元データ

- 16画素 × 16画素のアイコン画像、各画素は白黒の2値しかとらないとする。
- こんなに小さな画像でも、生成可能な画像の数は $2^{(16 \times 16)} = 2^{256} = 10^{77}$ 通り。
- その中で意味のあるパターンはいくつあるのか、列挙することは不可能。
- より少数のルールで意味のあるパターン（とその変異）を表現しなければいけない。

高次元データ

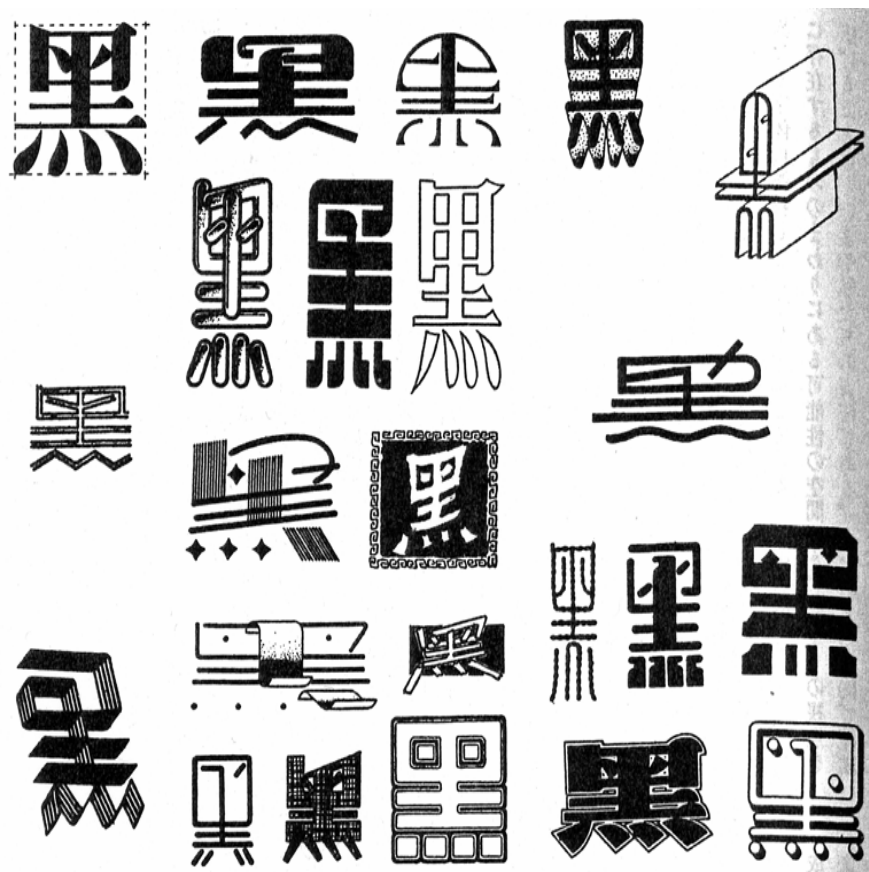


文字「A」の認識

- これは全部文字Aである。
- このすべてのパターンを「文字A」と認識できる機械をどのように作ればよいか。
- 同じ字体の文字「B」は生成できるか？

メタマジック・ゲーム、D.R.ホフスタッター、白揚社、1990

文字「黒」の認識



- これは全部文字「黒」である(中国風)。
- 日本人なら見てわかるが、西洋人には同じとは思えないらしい。
- 文字「A」の認識と本質は同じ。

メタマジック・ゲーム、D.R.ホフスタッター、白揚社、1990

文字認識

- あるパターンを入力とし、これを有限個のカテゴリに対応づけて出力する(分類)。
- 理論的にはすべての可能なパターンを列挙し、その一つ一つがどのカテゴリに属するかを決定すればよい。
- しかし境界付近には、必ず紛らわしいパターンが出現する。
- どのカテゴリにも属さないパターンを棄却することはさらに困難な課題。

二つのアプローチ

- 構文的パターン認識(Syntactic Pattern Recognition)
 - 文字のパターンを要素に分解し、その接続関係をグラフ構造などで表現する。
 - カテゴリごとに決定した規則を用いて推論し、最も規則と一致するカテゴリに分類する。
- 統計的パターン認識(Statistical Pattern Recognition)
 - 要素には分解せず、画像特徴量を抽出して高次元特徴空間に写像する。
 - 多数のデータから決定したカテゴリごとの領域との距離を測定し、最も近いカテゴリに分類する。

構文的パターン認識

- 文字「A」とは、上に尖った山が一つあり、横棒が一つあるパターンである。
- 一見、こちらの方が高度な処理ができそうに思えるのだが、過去の研究ではこちらのアプローチはあまり成功していない。
- その根本的な原因は、可能な空間の広大さに関する過小評価と、変動への脆弱性にあると考えられる。

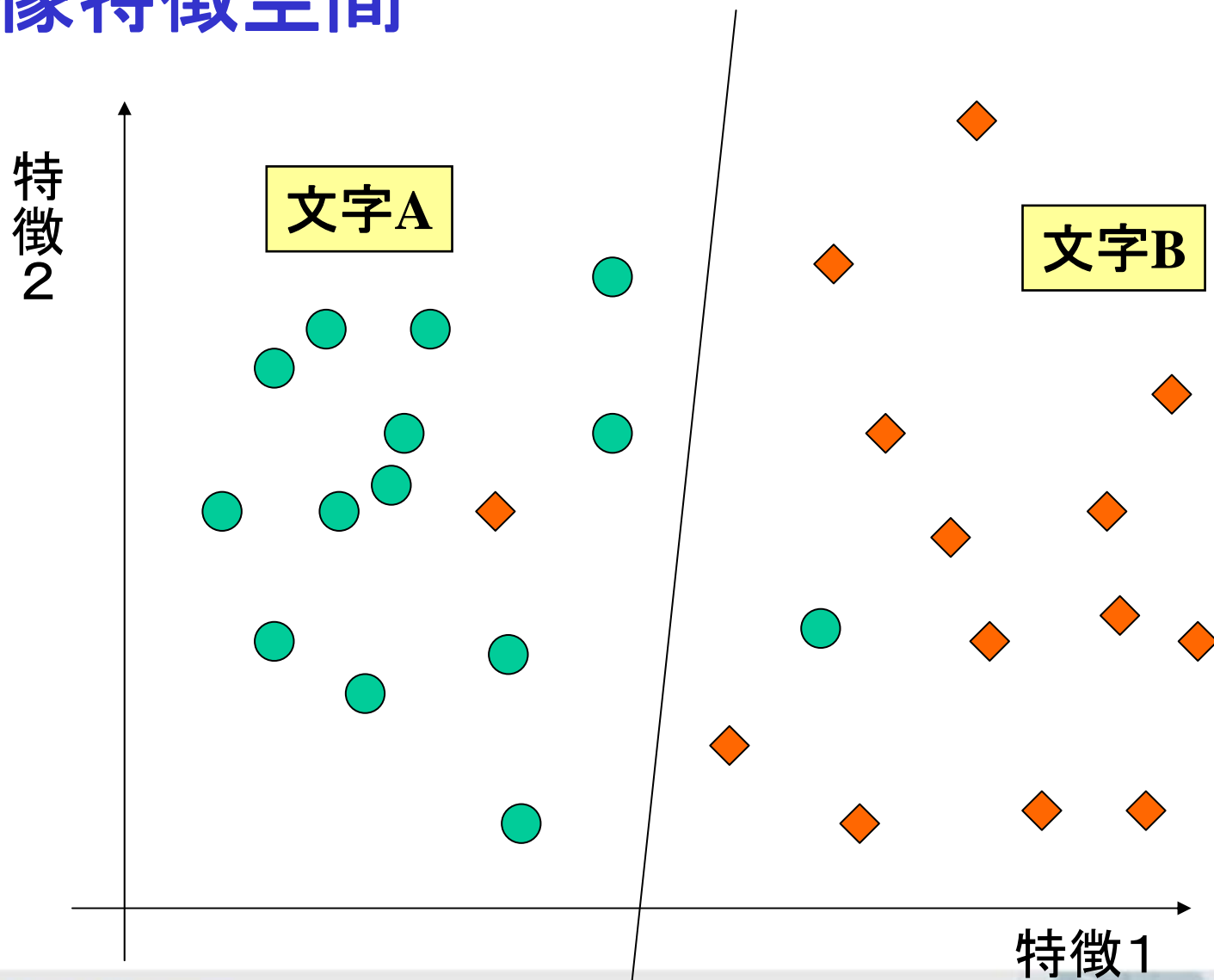
典型的失敗例

1. 文字「A」は、上に尖った山が一つあり、横棒が一つあるパターンで問題ないよね。
2. あれ、この文字だとうまくいかないな。横棒がちょっと斜めになっても大丈夫なようにパラメータを調整しよう。
3. あれ、尖った山ではなくて、山が平らな場合もあるんだな。だったら、場合分けして平らな場合もOKとなるようにしよう。
4. あれ、まだダメだな。そうかこんな可能性があったか。ならこのパラメータをちょっといじって、、
5. このような試行錯誤を繰り返すうちに、可能性の組み合わせ爆発に負けて泥沼にはまってしまう。

統計的パターン認識

- パターンの「文字性」に着目するのではなく、有効な画像特徴量の抽出に着目する。
- 多くの文字Aパターン(コーパス)について画像特徴量を抽出すると、特徴空間中で似たようなところに分布することがわかる。
- 多数の例から学習することで、文字Aに特有のパターンを間接的に「認識」できる。

画像特徴空間



画像特徴空間

- 画像情報処理の立場としては、文字Aの特徴量が他の文字の特徴量と重ならず、かつ小さな空間(次元)に集中するように、画像特徴量を選び出す技術が決め手となる。
- 画像特徴空間中で識別面を学習する問題は、確率モデルの学習として別に議論する。
- 明日のチュートリアル「カーネル法によるデータ解析入門」では、おそらくこのあたりの説明がある(はず)。

文字認識の精度向上

- 実際には、現在の商用製品では、代表的なフォントで明瞭に印刷されていれば、99%程度の認識精度に達している。
- 画像処理における特徴量の選択においては継続的な技術の向上がある。
- しかし画像以外の情報、例えば辞書との併用がかなり精度向上に寄与している。
- 画像情報と知識との統合が重要という結論。

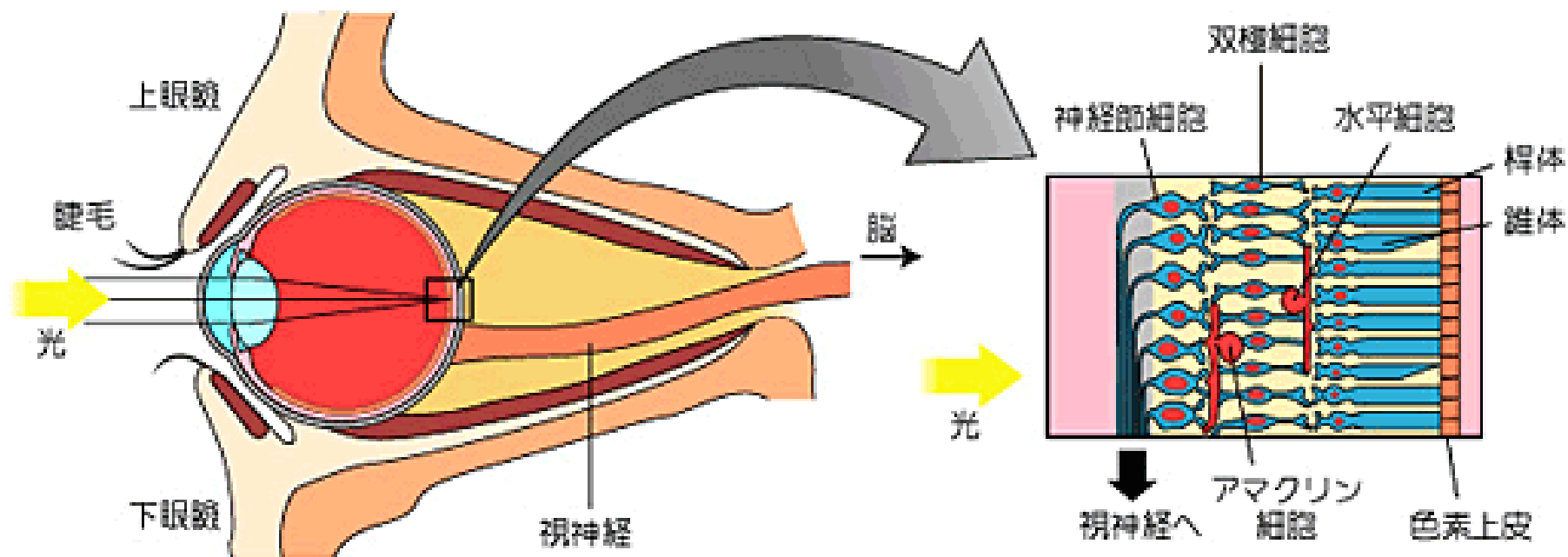
人間の認知

- 人間の意識的な活動は、構文的パターン認識のアプローチに近いかもしれない。
- が、通常の読書では、文字Aには横棒が1本、などといちいち認識しているとは思えない。
- となると、通常は統計的パターン認識のように、無意識的に画像特徴量を拾って、既知のカテゴリに当てはめているのかもしれない。

視覚情報処理

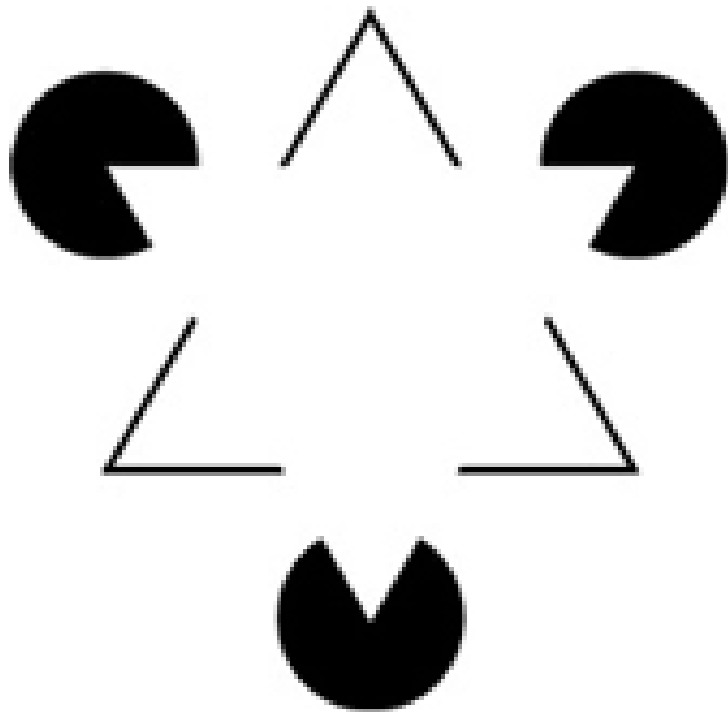
- われわれ人間は視覚情報をうまく処理しているのだから、まず人間の仕組みに学ぶべきではないか？
- 眼の構造などに関する研究は、色彩情報処理などの分野で多くの成果を生んだ。
- 錯視などの認知に関する研究は、人間の視覚情報処理が非常に巧妙な仕組みをもっており、しかも実世界を「ありのまま」に見ているのではないことを明らかにした。

人間の眼の構造



<http://eye.pfeizer.co.jp/>

主観的輪郭

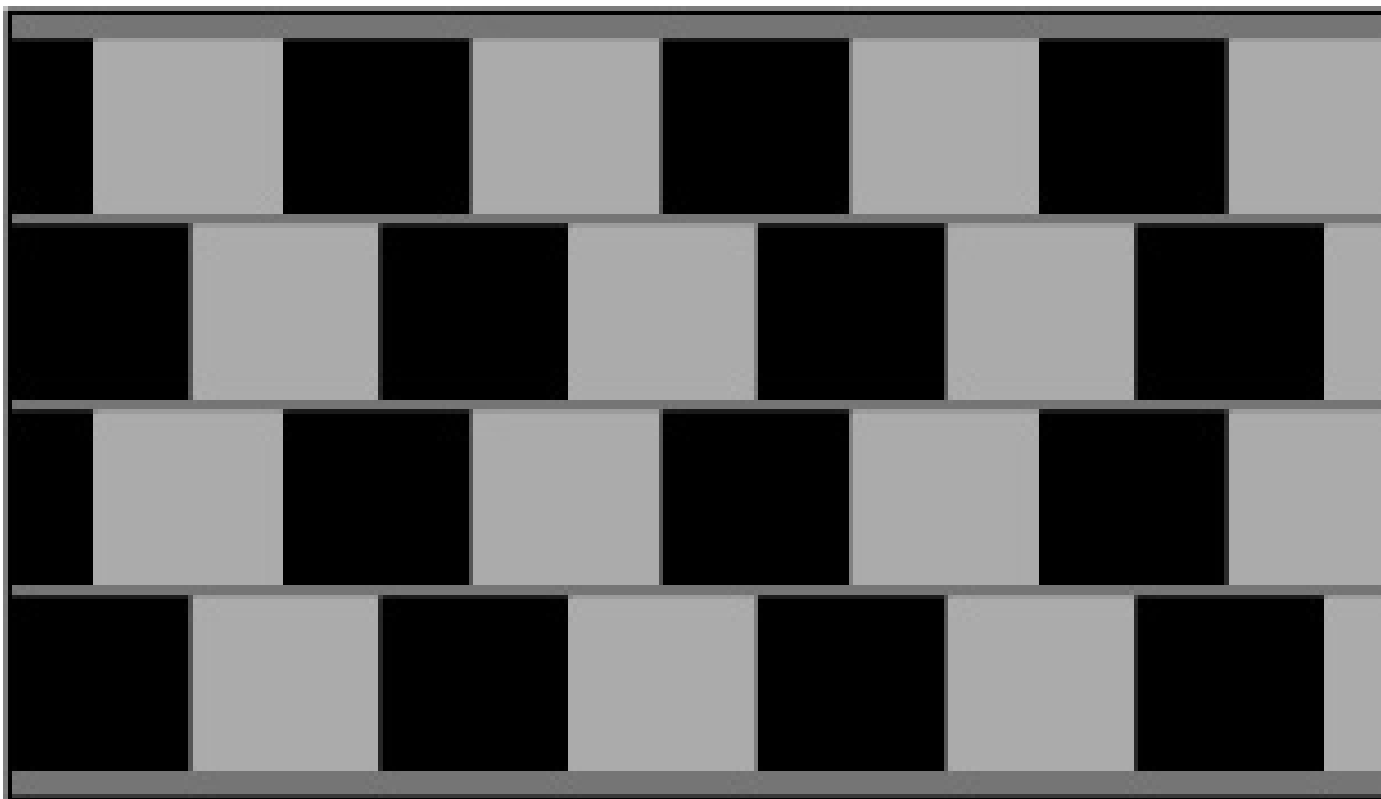


カニッツアの三角形

- 画像の濃淡というレベルでは存在しないのに、視覚レベルではあたかも存在するように知覚してしまう。
- 脳の判断により、輪郭が「あるはず」の部分を勝手に補間していると考えられる。

<http://eye.pfeizer.co.jp/>

錯視



<http://eye.pfeizer.co.jp/>

人間とコンピュータの違い

- 人間の視覚システムが進化の過程で獲得した実世界解釈機能は、錯視も生じるけれども生存競争には有利だったのだろう。
- コンピュータ視覚にはこうした仕組みがビルトインされていない(今のところ)。人間が画像中に見ていると思っているものが、実際の画像信号には出現していない場合もある。
- 特定の目的に関しては、脳のメカニズムがベストではなく、必ずしも真似る必要はない。

セグメンテーション問題

- 画像の領域分割(セグメンテーション)は画像解析に立ちはだかる大きな壁である。
- 画像を意味のある部分領域、あるいは構成要素に分割する(分節)することにより、「図(前景)」と「地(背景)」の分離が可能となる。
- 分割するためには、領域が何であるかを知っているとやりやすい。
- 領域が何であるかは、領域が分割されているとやりやすい。

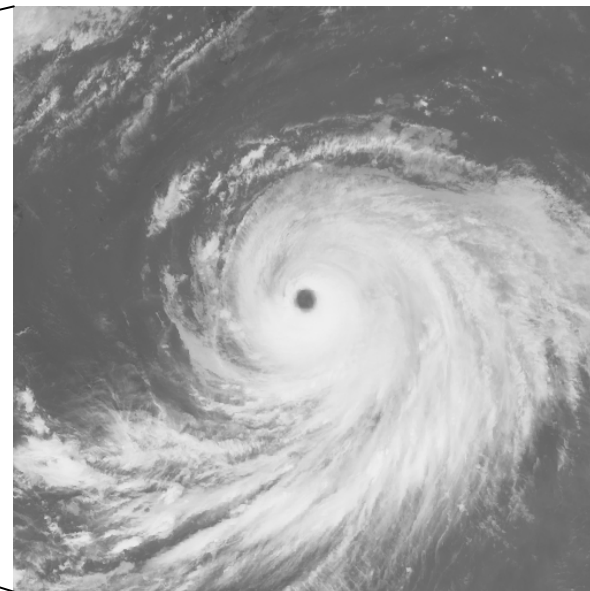
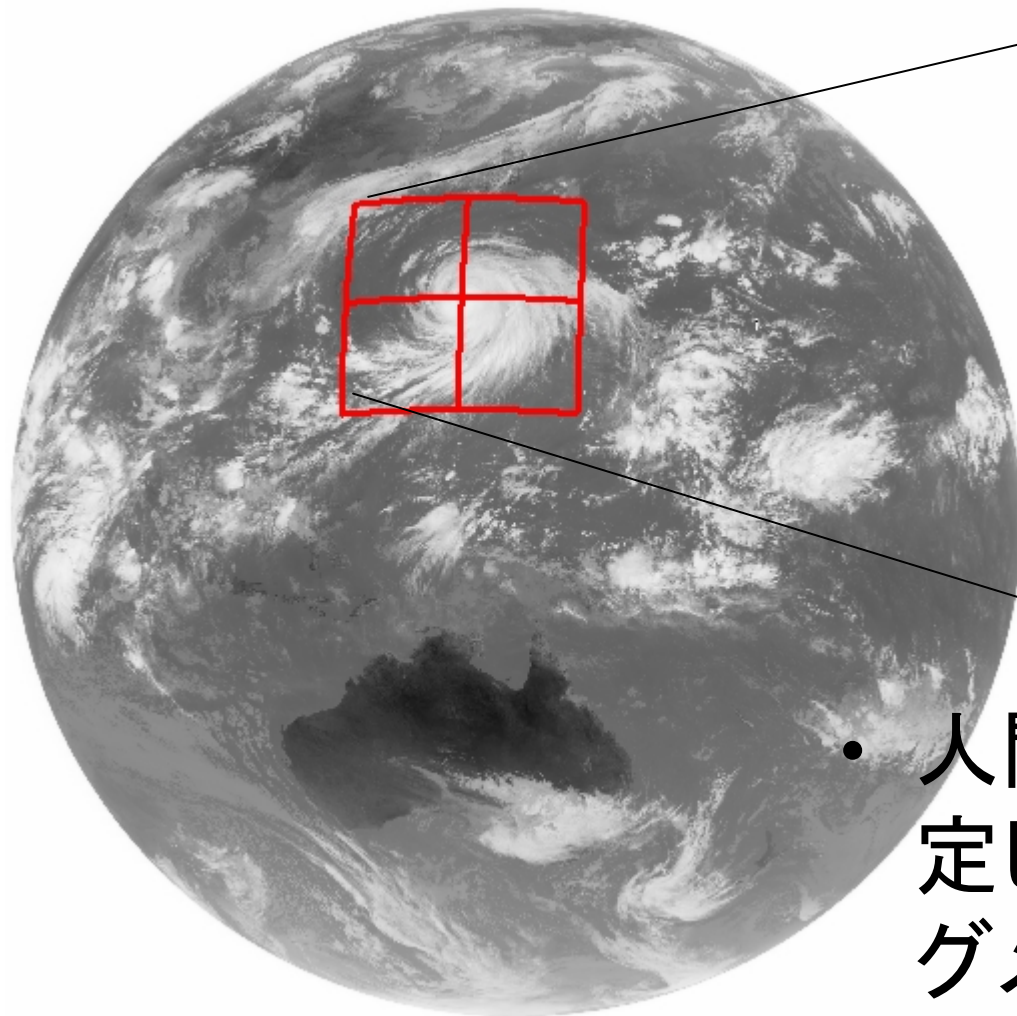
セグメンテーション問題

- 頑健なセグメンテーションは大変に難しい。このことが高レベルの画像解析（画像理解）を困難にしている。
 - 一般的な問題を解くのが難しいとしても、問題を限定すれば解けるかもしれない。
 - コンピュータがセグメンテーションしやすいようにお膳立てしてやれば、人間以上の能力を発揮することもある。
 - そもそもセグメンテーションが必要なのか？

問題をコントロールする

- 工業への応用などでは、問題を限定し、お膳立ても整えることで、画像処理が有効に活用されている。
- 例えば背景(ノイズ)を除去する、照明条件を一定にする、撮像系の歪みを抑えるなど。
- 自然界の観測では制御可能な要因が限定されるが、人間が少しだけ介入することにより、問題が簡単になる例もある。

人間の介在



- 人間が中心位置を指定してやることで、セグメンテーションが容易になる。

セグメンテーション不要問題

- Shape and motion from image (画像列からの形状や動きの復元)など。
- この種の問題はかなり成功している。
- なぜなら、物理世界の制約を入れることで問題が解けるし、それが実際に正しい解を与えるためである。
- 人間にとっての意味はあまり関係なので、物理世界の性質のみに集中できる。

物理レベルと意味レベル

- 単なる信号変換は物理的なレベルで扱うことができる。
- 画像圧縮(JPEG等)になると画質の評価という問題で人間という要素を避けることができないが、それでもなんとかなる。
- 意味のある領域と無視してもよい領域を分割するという意味レベルに達した段階で、問題が混迷を深めるように見える。

画像の階層構造

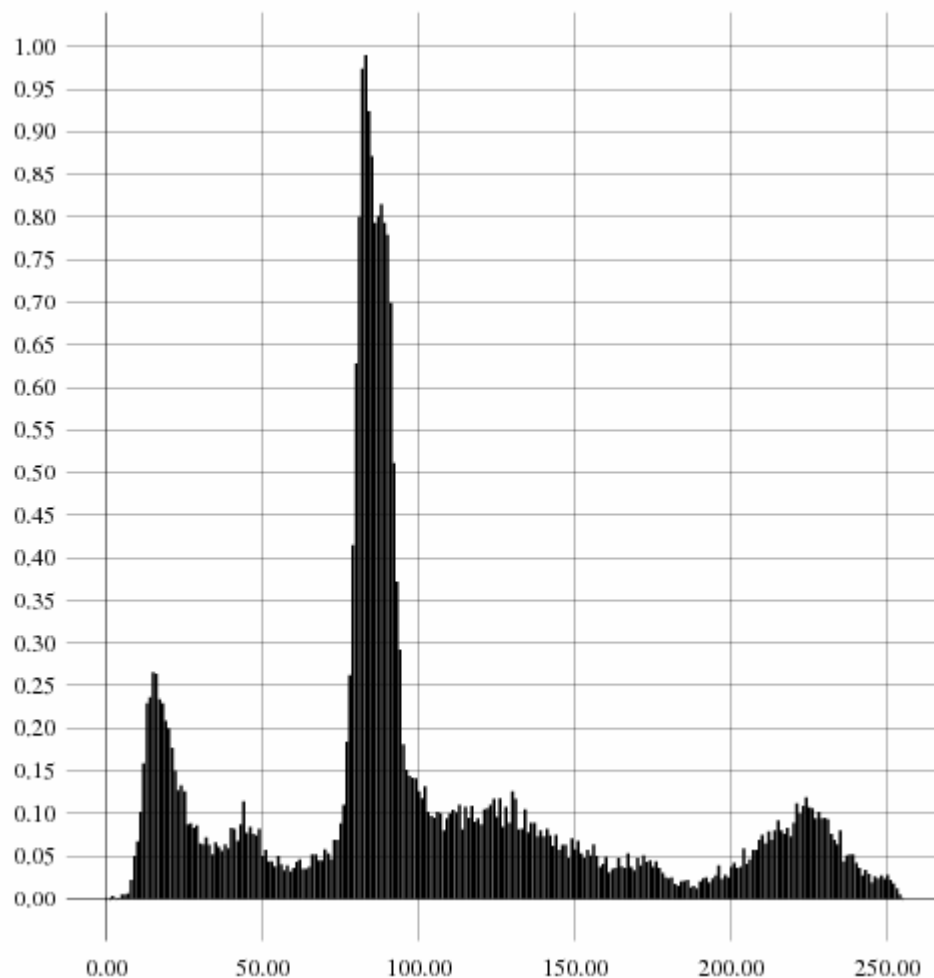
意味・感性レベル	画像の意味、あるいはその画像かうける印象に着目する
シーンレベル	画像全体の領域の空間的構成やその関係に着目する
領域レベル	何らかの観点で一様な画素の連続領域に着目する
トークンレベル	画素が何らかの観点で連続した最小単位のまとまりに着目する
画素レベル	1画素ごとの特徴に着目する

原画像



- カラー画像
- JPEG圧縮
- ○年○月○日撮影
- 被写体は××

画素レベル



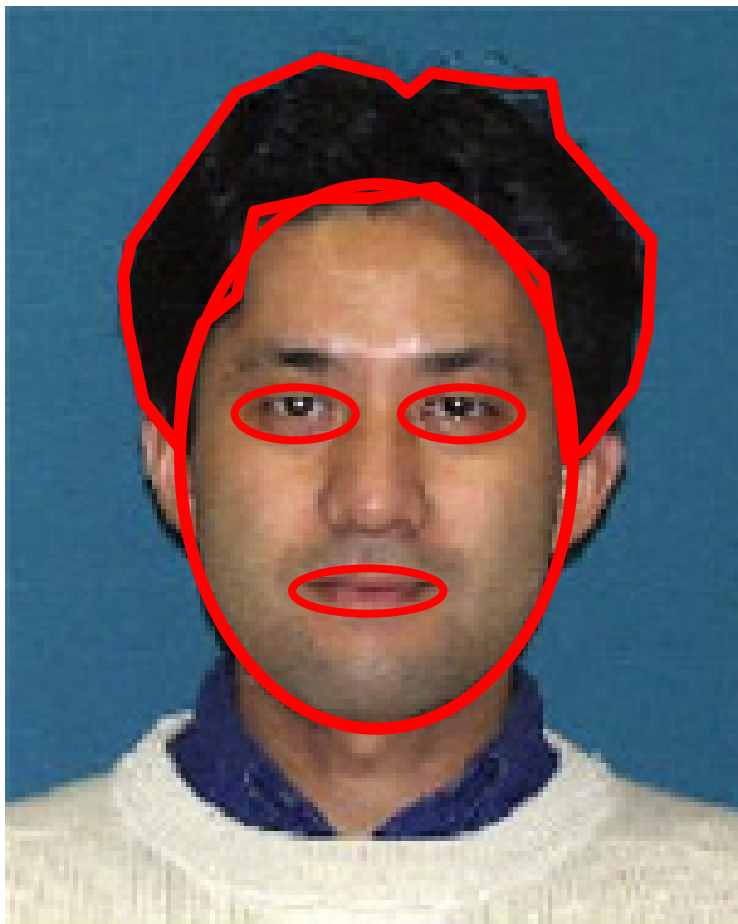
- 個々の画素値（濃淡、カラー）について何が言えるか。
- 例えば画像ヒストグラムを作ることにより、相対的に明るい画素と暗い画素との分布を調べることが可能。

トークンレベル



- 濃淡値が大きく変化する画素を検出（エッジ検出）することにより、境界らしき画素を検出する。
- 局所的なレベルでの処理なので、大局的に意味があるかは不明。

領域レベル



- 何らかの観点から一様な領域を抽出し、基本的な構成要素の集合として画像を表現する。
- また構成要素間の空間的位置関係も抽出できる。

シーンレベル



- 「人間の顔」を含む画像であるとの理解に達する。
- この写真に写っている人は〇〇さんである。
- これは男性、××才ぐらいである。

意味・感性レベル



- この人の表情は怒っている？それとも笑っている？
- この人に似たような顔を持つ人は誰かいる？

最近の研究の流れ

- コンピュータが高速化し、メモリが大容量化したという進歩を最大限に活用することで、画像解析の本質的な困難さを減らそう。
 - 画像中の物体の3次元構造がわかるようにカメラを移動させよう。センサも安いからたくさん使ってしまうおう。
 - いろいろな方向から見たときのアピアランスも全部記憶してしまえ。
- できれば厄介な画像認識は避けて、むしろメディアとしての面白い活用を考えたい。

画像解析の秘訣

- 画像の性質を決め付けるな。画像の変異の可能性は人間が想像する以上に大きい。それをすべて受け入れる心の準備をしよう。
- 高品質画像を手に入れよう。低品質画像に対して必死に頑張るより、高品質画像をシンプルに処理する方がはるかに楽。
- 試行錯誤は重要だが、注意しないと組み合わせ爆発に敗れがちなことを忘れずに。
- 本当に画像解析が必要な問題かをよーく考えよう。無理に使う必要はないかも。

問題をうまく定義する

- 文字Aの認識、椅子の認識、これらはすべてとんでもなく多様なインスタンス（アピアランス）を認識対象としている。
- 人間が「一まとまりのカテゴリ」と思い込んでいるものは、実は膨大な変異のインスタンス（アピアランス）を含んでいることがある。
- 問題をうまく定義しよう。分類したいカテゴリは、本当に一まとまりなのか？

まとめ

- 画像解析が難しいのは、高次元データであること、的確な解釈をするには情報が不足しているからである。
- パターンとは何かを明示的に書き下すより、全体的な特徴を捉える方法が現在は優勢。
- 人間とコンピュータでは見え方が違う(人間の見え方を模擬することはおそらく可能)。
- セグメンテーションは未だに難問。
- 画像は階層構造で表現すると整理できる。

参考文献

- パターン認識・理解の新たな展開、小川英光編著、電子情報通信学会、1994。
 - 約10年前に日本のトップ研究者たちが問題意識を整理した本。
- 新編 画像解析ハンドブック、高木幹雄、下田陽久監修、東京大学出版会、2004。
 - 画像処理手法をかなり網羅している。高価ではあるが研究所（あるいは研究グループ）に1冊は常備しておきたい。
- わかりやすいパターン認識、石井他、オーム社、1998。
 - パターン認識については最も手頃な入門書のひとつ。
- ビジョン—視覚の計算理論と脳内表現、デビッド・マー、産業図書、1987。
 - 視覚情報処理に関する大きな枠組みを提案した名著。一昔前の本ではあるが、その意義は決して色あせていない。

ソフトウェアツール

- 画像表示、変換
 - ImageMagick (OSS), Adobe Photoshop
- 画像認識
 - OpenCV (OSS)
- 総合ツール
 - Matlab
- その他分野に特化したツールが多数ある。
ただし使えるツールを探すのは難しい。

コンタクト

- 共同研究のコンタクトをお待ちしています。
- 融合研究プロジェクトに興味のある方(少なくとも地球環境ポータルは人手不足か)。
- 大学院生で画像情報処理に興味がある or 必要に迫られている方。
- 必ずしも融合研究でなくても、共同研究の種がございましたら、ご遠慮なくどうぞ。

コンタクト

- 個人ウェブページ
 - <http://agora.ex.nii.ac.jp/~kitamoto/>
- 情報研傘テーマウェブサイト
 - <http://earth.nii.ac.jp/>（試験公開中）
- 私が関わっている他プロジェクトのウェブサイト
 - <http://www.digital-typhoon.org/>
 - <http://eye.tc/>
 - <http://dsr.nii.ac.jp/>
 - <http://www.biportal.jp/>