

言語の獲得には、外界を知る必要があります。画像認識が基盤技術として、明確にしておかなばなりません。そこで、自然言語認識の考察の前に、画像認識技術を考察することになりました。

1．はじめに

言語獲得では、画像を見て、瞬時にシカとか人間とか、人工物とかを判断しなくてはなりません。そのための学習はそう多くない画像の提示で行えなくてはならず、ビックデータを必要とするディープ・ラーニングには向かない技術であると思えます。また、画像のここが、参照画像と異なるとか、同じとか語れるようにならねば、会話が続かないでしょう。とするならば、画像を解析し、ある一定のアルゴリズムをもって認識していくという技術が求められます。

2．「理解する」とは

計算機は物事を理解できないと言われます。ならば、「理解する」とはどういうことでしょうか。計算機はレジスタの内容のみを決まった命令で実行するだけだから「理解していない」のでしょうか。とすれば、あるデータの固まりをメモリに持って、複数のコマンドを同時並行処理ができて、データの参照、変更、データを観て、特定の処理を実行できるように成っていれば、「理解した」と言えるということになります。そういう意味では、ニューロコンピュータは入力データセットから並行処理で出力を決定しますから、「理解している」に近いかも知れません。「理解する」とは、データを外界の構造を反映したものとして表現し、そこに同時並行的に多数の処理を柔軟に組み立てて処理できること・・・と言えそうです。

3．ポテンシャル法

画像認識システムを設計しようとするとき、曖昧性が無数にあるので、一度に複数の画素とか線分を何であるかを知りたくなります。固まりとして知りたくなります。つまり、画像を「理解」したコンピューティングをしたくなるのです。人間も、画素毎に逐一解析しているのでなくて、纏まった画素群を一度に処理しているように感じられます。そうしないと、曖昧性を払拭できないのです。というか、ものすごく計算量の多い処理になってしまいそうです。

そこで、ポテンシャル法を応用を考えてみました。ポテンシャルは各画素が距離に反比例する値を他の画素に与え、その合算値によって、所定の解析処理をしようというものです。対向線の存在とか隘路では、空間画素のポテンシャルがそうでないよりも大きいでしょう。また、とがった線分でも通常よりもポテンシャルが高い点が存在します。また、鋭

角側の空間のポテンシャルが鈍角側の空間のポテンシャルよりも大きくなります。あとは、画素群を順次スキャンして、ポテンシャルの極値を探せば、一気に線や領域や対向線の解析が済みます。



あるいは、電気回路の銅線のモデルのように、線上は同一のポテンシャルとなるように画素の重みを計算し（電子の動きを弛緩法でシミュレートする）それで空間ポテンシャルの等高線を得ていく方法もあるでしょう。これは計算量が大きくなるので、画素が固定ポテンシャルを持つとする方法よりも不利です。



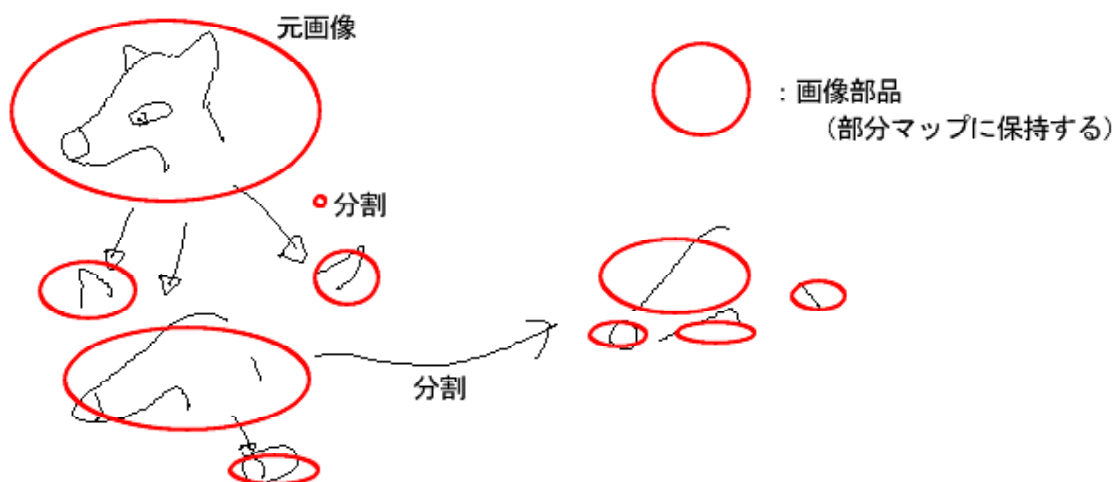
いずれのポテンシャル法を用いても、円の場合は、外郭曲線の内点の GRAD（グラジエント）は0になって、簡単に判別できます。

4. 画像認識

画像は全て、標準の大きさのマップで表現します。大きな画像も標準のマップに投影して、解析を一貫性ある評価基準で行うようにします。大きな線分とか、雑音画素というものを評価するにはそうするよりないです。

画像は、部品と外枠を同定することが認識処理になります。部品は領域と線分と点です。外枠は、それらを包含する閉じ曲線です。解析は先ず部品群を捉えます。そうして、ポテンシャルの最低の閉曲線を外枠候補としていきます。正確な外枠の認定は、知識になる、モデル画像データとの一致を見て判断します。モデルは、解析結果が入っていて、その一致度を評価しながら部品とか外枠を決定します。人間の顔では、髪の毛は外枠にも成りませんが、額の輪郭で、外枠を決めることも必要に応じて行われることです。これは全て、知識とのマッチングによって判断するところです。

認識は全解析結果を一つのマップに納めることもあり得ますが、特徴（部品）毎に部分マップを起こして、それに画像部分をコピーして持つ方が融通が利きます。特徴はそれだけで、画像検索のキーになりますし、特徴は部品として直ぐにマッチングができます。部品の大局的な性質をまとめて表現する事になってきて、「理解する」ことに繋がるからです。抽出した特徴は一つ一つマップになるのです。注視しなかった部品（特徴）はマップにしません。データとして保存しません。



マップの各画素は次のような情報を持ちます。

- (1) この画素が属するのは、線か、領域か、点か、特徴点かのタイプ。
- (2) 幅、長さ、角度、曲率、繰り返し周波数、配置、密度、色の属性
曲率は接する円の半径で表現する。

周波数は、同じパターンの点所在する範囲でフーリエ変換して求める。

- (3) 確信度

これらの配置はマップの中に位置づけることで表現します。

画像認識の大きな領野は次の通り。

(1) 基本画像認識

特徴図形 (部品、外枠) を切り出しながら、基本的な属性を解析していきます。

(2) 図のオーバーラップ解析

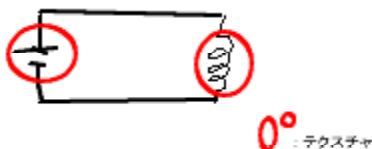
2つの図形の重ね合わせ (曲線による分割) を、重み付き投票で得ていきます。線分群の属性の違いの分布で投票します。

(3) 生物の体勢解析

人が立っている画像から、かがんでいる画像を推論するというようなことです。知識のモデルにテクスチャマッピングし、視点から画像を生成します。

(4) テクスチャ解析

点列とか、線分列とかの境界は明確でないが固まって存在する領域の形状を推定します。ポテンシャル法で行います。ある値以上の合成ポテンシャルの場所を塗りつぶし領域として切り出します。



(5) 繰り返しパターンの周波数と画像範囲の解析

フーリエ解析など、パターンマッチングするなどの重たい処理ですので、専用のCPUが必要と思われます。

5 . おわりに

画像認識は次の3つの手法があります。

(1) 解析

(2) 確率 (重み付き投票法)

(3) 知識の利用

画像はとにかく曖昧性の度合いが大きなものです。これには、解析で基本的なオブジェクト (線とか領域) とその属性を得て、それらの分布から部品と外枠を重み付き投票で決定していき、最後に知識からの援助で画像認識結果を確定していくことになります。

おわり