イメージの合成について　　2016/01/13（水）

自然言語認識と動画像認識にでは、知識ベースの基盤であるイメージを生成していくことが必要です。

そのイメージ生成ですが、自然言語認識では、文を幾つか得て段落となし、段落がいくつか纏まって文章を成すというふうに、言われていることを一気に認識するのではなく、部分的なフォーカスである文を収集して行きながら統一した文章のイメージを構築して行きます。

一方、動画像認識でも、外界を一気に認識するのではなく、フォーカスによって小さな視野の解析と認識を行い、フォーカスを移動しながら全体を認識していくものです。

このように、認識処理が部分的な認識であるイメージ断片を、次々に合成していき、最後に全体のイメージに至るものです。

フォーカスされて得るイメージは断片ですが、オントロジーで接地された纏まったものです。一方で、目標の全体イメージを得て行くには、断片の形状とか過去の体験から得られた知識を持って、ジグゾーパズルを解くような作業になります。高速にその作業を実施するにはあらかじめ全体のイメージの原型というものがあって欲しいものです。自然言語認識では、詳細な記述の前になにか概略を与えます。その概略も既知のひな形であります。動画像認識ではフォーカスの周りには大きな漠然とした全体構造が与えられているのが普通です。このように、先ずは知識があって、その知識によって最終概念構造のイメージが創られて、各フォーカスされたデータから得られるイメージによって全体概念のイメージが明確になっていくというのが手順なのです。

イメージの概略はオブジェクトの共起関係・・・オブジェクトがありそうというデータがあるだけ・・・ですからグラフで表現されるでしょう。それとデパートのように何階かの部屋の連なりがあるというような少しだけ配置を示すデータがあるでしょう。それはマップで表現されます。イメージが明確に成っていくにつれ、マップは大きくなったり、小さくなったりします。デーパートの屋上というように配置とかオブジェクトの属性を指定するデータも必要です。それにはコマンドが役に立ちます。オントロジーによりコマンドを表現するのです。更に、このイメージにこのオブジェクトが明確に存在するというのはコマンドで表現します。

例文を挙げましょう。

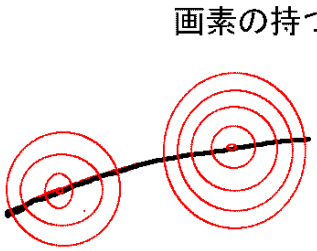
「長野県の観光に来た。先ずは新幹線で東京駅から長野駅に行った。善光寺でそばを食べて、更に松本に行った。松本城を見て、書店で長野県の観光図書を買った。それから特急

あずさで帰った。」

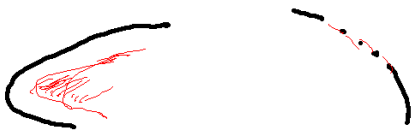
なんとなくイメージが分かると思います。「長野県の観光に来た」では全体を示すイメージが得られるでしょう。長野県の地図を思い浮かべて、どこを回るのかなという身構えができます。「新幹線で東京駅から長野駅に行った」では、この人が東京都に先ずいたのだな・・・長野県の観光が終われば東京都に戻るのだなというイメージが知識から得られます。「行った」とか「着いた」、「移動した」とか、「乗った」、「降りた」といった動詞はシーンの区切りを表わす特徴的なものだということも学習によって得られるでしょう。これで、シーン切り替えの候補となるフラッグを発火させることに成るでしょう。実際にそのシーンが予想のものかは、重み付投票でシーンの支持を重畳させていくことで確信に変えて行きます。例えば、長野ですと、善光寺があるし、そばを食べることもあるでしょう。そんな長野市の枠組みを構成するオブジェクトとして枠組みイメージは持っているのです。

動画像認識においては、枠組みとなる線と枠組みのなかに配置する部品となる図形要素群が重要となります。部品画像が自然言語のオブジェクトに相当するのです。

画像解析の手法として強力と思われるのに、ポテンシャル法があります。その技術について少し触れたいと思います。

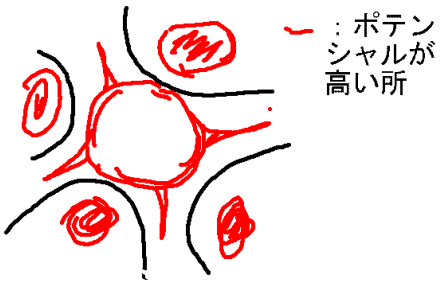


画素毎にポテンシャルをもっていて、ポテンシャルは遠くに離れるほど値が小さくなっていくものとします。そうすると、図形の空間はポテンシャルが重畳していって、ポテンシャル値の濃淡ができます。

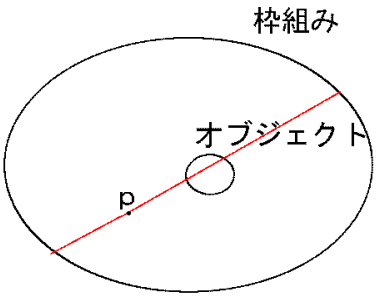


：ポテンシャルが高いところ

濃淡は、尖ったところの周辺が高くなります。また、断線しているところも埋め合わせるポテンシャルの濃い所ができて、断線が繋がります。ただ、この断線処理には線の勢いというか、端点周辺の特別なポテンシャル処理が必要にはなります。



複雑な形状もポテンシャル法によって、総合的に解析できるようになります。



画像解析には枠組みの同定と同僚となる部品の同定が重要です。それは、任意の点 P から線を引いて、その線と図形が反対方向で交われば枠組みで、そうでなければ部品だと判断することで行われます。

おわり