

1. はじめに

視覚が外界認識に大きな力となるのは言を待たないでしょう。さらに、思考で生成される論理的な事象についてもコンセプトは視覚系のオブジェクトと同じように扱っていくこととなりますから、視覚系は更に大きな位置を汎用人工知能で持つこととなります。

視覚系はオブジェクトの形を主に認識する外側膝状体系と距離と動きを中心に高速に認識する上丘系とがあります。本論では、主に外側膝状体系を考えます。

2. 視覚領野のサブ領野群とそれらの働き

視覚領野は幾つかのプロセスの違いからできるサブ領野からなります。まず、静止画像認識（学習とパターンマッチング）があり、静止画像から立体視認識を行いオブジェクトとその空間配置を認識します。また時間的に静止画像は変化していきます。その動画像認識が静止画像認識の次に行われます。3次元世界のオブジェクトですから一つの視点から見ますと、オブジェクトが重ね合わされ一方が他方を遮蔽することがあります。部分から遮蔽された部分を立体視を中心としたプロセスにより静止画像から完全なオブジェクトの姿を推測することをします。その上に、特に人間の顔認識機構（相貌認識）があり、笑い顔とか人物の特定とかを行う領野があります。最後に、身体性座標空間にオブジェクトを配置し、認識の基盤としていきます。

（1） 静止画認識

2次元の視覚画像を解析し、画像の重ね合わせとそのクラスタリングで、オブジェクトの同定と属性の分類を獲得し、パターンマッチングで、既存の画像と対応させて、オブジェクトの関連情報を再生（発火）します。

（2） 立体視認識

3次元オブジェクト空間を画像認識結果から獲得していきます。また、パターンマッチングによりオブジェクトの関連情報を再生（発火）します。

（3） 動画認識

時間的なオブジェクトの形状や空間配置、属性の変化を記録し、傾向を学習し、ある状況からの時間的展開を推測（シミュレーション）します。

(4) オブジェクトの重ね合わせ状態の認識

1つのオブジェクトを別のオブジェクトが遮って隠したときに、画像からもとのオブジェクトの完全な状況を推測していきます。

(5) 画像（オブジェクト）の属性群認識

画像の感覚を得ることやオブジェクトの属性を認識します。

(6) 相貌認識

人物の認識を行う。特に顔の認識に力をいれて実現するサブ領域です。

(7) 身体性座標空間での認識の完成

視覚領域で得られるデータは身体性座標空間として保存します。また、必要に応じて、オブジェクトに関する部分的なデータからシーンなどの全体状況を再生します。

2. 1 静止画認識

(1) フォーカス処理

視線を中心にした視野の変遷を記録していき、外界の大きな枠組から特定のオブジェクトの認識を行います。フォーカスにより解析対象の選別を行うこともします。論理 SLAM により視野空間を合成して身体性座標空間の部分空間にオブジェクト配置を構築します。

論理 SLAM としたのは、座標変換をするだけで無く、座標変換を越えた、論理的な配置コマンドによる座標空間の他の座標空間への変換です。例えば、「私の右側に花瓶がある」ですと、これは身体性座標空間から、世界座標空間で捉え直す必要があります。世界座標では、基準が南向きだったとしたら、身体が東むきならば、花瓶は南側にあるだろうという推論を行います。

(2) 2値化処理

画像は色を持っていますが、その色の選別をしながら2値画像を多重につくり画像解析していきます。

意志からのコマンド入力により2値化の制御を行います。

(3) 線分分布解析

画像解析の基盤は、線画を得るところから始まります。閉じ曲線はオブジェクトの区切りであり、オブジェクトの枠組です。その中に、部品として各種の更に小さな閉じ曲線群が配置されます。そのため、画像を画素でなく線分分布として解析の基盤を造っていきます。

線分を認識プロセスの基盤とするメリットは、画像が他の画像によって遮蔽されていても、途中の画像の状況からパターンマッチングを機械的に実施することができることにあります。

連続する2つの線分の相対位置関係か、2つの線分の成す角度をもって、(key,value)として、keyを列にして量子化したvalue値の行にこの画像の名前を設定して、配列を作ります。配列は、どの線分を基点にしても良いように連続する線分数だけにの配列を作り、パターンマッチング(検索)のために用意します。

パターンマッチングは、2つの画像の配列のvalueが一致する線分群部分の数を評価して行います。

(4) 多段階解像度管理機構の構築

画像の解像度を下げて、大局的な図形の傾向を見ることができます。長い線分か、大きな曲がり角か、分岐があるかなどです。細かな解像度では、図形がギザギザを持っているとか、穴がいっぱいあいた帯であるとかが見てきます。高速に線分分布の属性を得るのに多段階の解像度をもって解析することは効率の点で重要です。

(5) 正規化と画像の重ね合わせ

枠の形状の特徴を画像の大きさと傾きを一定にして重ね合わせ、そのクラスタリングをすることにより得ることができます。そのクラスタリングの準備とクラスタリングによってオブジェクトとクラスタを対応させるという管理も行っていきます。

線分による画像表現する場合は、線分分布解析で説明した配列の重ね合わせを行います。生画像を基盤に重ね合わせをする場合には、縦長方向と横長方向の2つの量子化した向きを決めて重ね合わせます。

(6) 正規化画像のクラスタリングと特徴図形の獲得

クラスタリング解析を行い図形の特徴を分離・獲得していきます。その分離の方法は、交差法によります。特徴図形は交差法の中で、最長一致のものを特徴図形とすることにより行います。交差法により、特徴図形の連鎖のグラフ（ステートマシン）が形造られます。

(7) 特徴図形による画像の近似

クラスタリングによって得た図形の組み合わせとして、オブジェクト画像を近似し、パターンマッチングの準備をします。

線分分布解析の線分の代わりに、特徴図形によって図形を展開した結果をもとにその特徴図形群の相対配置関係を配列にすることで、実現します。

特徴図形は相似であれば同一のものとして処理しなくてはなりません。画像の中にある状況に応じた大きさの特徴図形と相似のものを検出できなくてはなりません。そのため、各解像度の段階に応じて、特徴図形の線分分布配列を持ちます。全ての候補特徴図形の全ての解像度の配列をもって認識しようとしている画像とのマッチングを行います。

(8) パターンマッチング用配列の獲得と利用

パターンマッチングは、線分分布の完全一致と特徴図形の分布解析によって実現します。その解析は、特徴群の (key,value) としての配列にオブジェクトの (key,value) 展開で対応する配列のエントリにオブジェクトの識別子を埋め込んで、近傍検索していくことで行います。

(9) コンセプトの構築

外界のセンサーデータを交差法で特定のオブジェクトに対応した部分 (key,value) の集まりに篩い分けしていくことで、コンセプトが形成されます。コンセプトには、識別子が一意に汎用人工知能により与えられます。

(10) 各種データの保存

RDB にコンセプトを永続化します。分散環境でのデータ保存になります。

2. 2 立体視認識

(1) 単眼立体視

一つのレンズ系で獲得する画像と画像内のオブジェクトの3次元配置を得る能力です。

(2) 複眼立体視

二つのレンズ系をもって高度の3次元オブジェクトの配置関係を得る能力です。

(3) 座標空間の構成

生活の基準になる世界座標空間（基本的に北極から南極をみた世界になります）と身体正面を基準にした身体性座標空間があります。世界のオブジェクトとオブジェクトの属性の配置を管理していきます。

座標系の解像度は配列の解像度です。2つのオブジェクトが同じ配列のエントリに入りそうになったら配列は更に解像度を増したものを用意し、その衝突配列のエントリからこの解像度を増したものを指して、そこにオブジェクトを別々のエントリに入れるようにします。

2. 3 動画認識

(1) 時系列の構成とチャンク化

静止画像の時間的変化を静止画像データのシーケンスとして保存し、さらに立体視の時間変化をシーケンスとして持ちます。その中で、連続しない部分（断片化属性を持つ部分）をもってシーケンスはチャンクとして切り出されます。

(2) コマンドの生成

意志のもと、視覚データは他のセンサーデータとともに記号化されます。もっともおおざっぱなチャンクはある時刻での視覚認識データとセンサーデータの集まりです。そのチャンクはコンセプトとしてコマンド表記されます。そのコマンドは、(動詞[格,オブジェクト・属性]・・・) というフレームが与えられていますから、交差法によって多数のデータを与えられて、交差法により精密なコマンドとして獲得されていきます。

2. 4 他のオブジェクトによる遮蔽状態の認識

(1) 座標空間の断片の組み合わせ処理

パターンマッチングによるオブジェクトの断片の認識を行う。

大枠の空間知識と断片同士の補完組み合わせを行う。

2. 5 画像（オブジェクト）の属性認識

線の持つ曲率とか、長いとか、重そうとか、尖っていたそうとかの属性を画像と線分布配列から解析して得ていきます。画像の持つ感覚を得ていく認識部です。この部分はいかなる解析プロセスを技術者が用意するかという問題です。

ですが、画像データの重ね合わせとクラスタリングをしていく汎用人工知能の持つ過程で、特徴として、尖っている、曲率が大きいとかの属性は抽出されていくでしょう。それは、他の肌センサーと結びつく正に感覚的な属性となっていきます。

2. 6 相貌認識

静止画像認識を顔の属性に特化して精密なコンセプトのもと構築したサブ領域です。顔の構造と筋肉の動きを精密に分類し（交差法で）、コンセプト化したものがパターンマッチングのデータになります。

2. 7 身体座標空間への反映

(1) SLAM の構成

視界は局所時間での作業座標空間で、ここに視覚による3次元立体空間が作られます。それを身体座標空間へ反映させる必要があります。

論理による座標変換と数学的な座標変換によるものがあります。基本的に論理によるSLAMを実行します。

(2) 世界座標空間から身体座標空間への認識状況の反映

オブジェクトの配置は世界座標空間で管理しますが、姿勢の状況に応じて身体を中心とした座標系に必要なデータを持ってきて、外界を認識する必要があります。

3. 視覚領野の総合的働き

凧揚げの例をあげてみましょう。「風の強い日、凧をあげました。」という文で、身体性座標空間の凧上げのシーンが発火されます。木が生えている丘の上に凧をあげている人が想定されるシーンです。ここで、「枝に糸が絡まってしまった。」という文が続くと、木と枝と凧の糸を連想した状況が発火して、そういうことは可能性があるという評価がくだします。そのあとの難儀な作業も連想として、シーンのイメージで発火されます。このように、推論とか状況の評価とか思考の実現は、視覚系のイメージを基盤に行われます。

連想はイメージの特徴的な状況をエキスとして、そのエキスがパターンマッチングするとして実現されます。パターンマッチングが基本的なプロセスと言えます。

思考には常に評価が行われる過程です。思考の起点となる事象も評価され、思考の起点からの直観プロセスの結果も最適なものが評価されます。また思考が首尾一貫していることも評価によって保障されます。

4. おわりに

思考は視覚系のイメージを基盤として実現されます。問題発見も、問題設定も、問題解決も評価をしながらの思考が行われます。視覚センサーからのデータをイメージとして（デジタルツインとして外界を表現します）獲得していく過程は本論で示す通り、大がかりな解析と学習と認識機構によります。

この他にもいろいろ、実際に汎用人工知能を造るにあたって実現するプロセスがあるでしょう。例えば、ディープラーニングです。そのようなプロセスでも最終的には思考過程で使うイメージ生成をもって、結果を出すことが目的のものになります。

画像の属性認識で、解析プロセスは、普通、1ピクセル毎に処理していきます。それはコンピュータとしてのもので、人間の脳は画像を一気に解析するように思われます。そのことを実現することも考えて見たい物です。ピクセルごとにプロセッサを配置する（FPGAなら可能でしょう）と良いように思われます。

おわり