

概念を獲得するには、画像と言葉に橋を渡さねばなりません。画像認識して、画像の中のフォーカスが当たっているものを意味ネットワークで捉えて、それを同時に発せられた言葉の意味ネットワーク（コマンド表現でもよい）に対応づけること・・・これが概念を獲得することです。

画像は視覚センサーの値のセットですから、そのままでは、概念構築の対象としては広過ぎます。ピンポイントで特別な事象を選択しないとイケない。しかも、そのピンポイントの事象は他者と簡単に・・・一意に共有されるものです。ピンポイントの事象はフォーカスが当たったオブジェクト群と格と行動と属性（特に能力）ですから、そのフォーカス機構はもう既に「概念」を強力に決めるものであり、他者と共有されるものですから、強制的なものです。ある「概念」がある画像空間では、もうその「概念」にフォーカスが当たるということであるわけです。私たちは自由ではないのです。環境のなかの存在であるということです。

熱中するものにフォーカスが当たる。それは他人に直ぐにわかるもので、他人も・・・一瞬かもしれませんが、その対象にフォーカスしてしまう。あくびをすれば、他者にもあくびが伝搬する。生き生きとした職場では情熱を皆もって、一つの価値を共有していく。アイデアが良く生まれる。優秀な人は、周りの人のレベルをも上げる。そうしたことは、個人の自由にはならないもので、人間が他者とフォーカスを共有できるように仕組みられた業なのだということが分かってきます。ロボットもこのフォーカス能力を持たねば、人間と共通の概念を構築できないということでもあります。

で、ロボットが人間がなににフォーカスしているかは、たとえば、目線のあたるところで分かります。さらに、サッカーや野球ですと焦点のボールとその近傍の人間。新規なもの、危険なもの、手で持っているもの、指差しているもの、植物よりも動物、動物よりも人間。さらに、会話ならば文脈で、フォーカスは慣性を持っているということ。・・・とうとう、フォーカスには文法がありそうです。

で、画像認識ですが、図面でなく、一般画像から意味ネットワークを抽出するプログラムを作っていかなければなりません。ロボコンとかでみますと、結構この辺の技術は進歩しているようで、「コンピュータ・ビジョン」の関連資料で勉強して行く必要があります。まずは、オブジェクト切り出しと、それがなんであるかを解析することをコーディングしていくことになるでしょう。