コンピュータビジョンと
イメージメディア 104－11
（1997．3．27）

CVCV-WG報告：コンピュータビジョンにおける技術評論と将来展望XIV
-コンピュータビジョンと情報統合-

長屋茂喜
RWCPつくば研究センタ 情報統合研究室

あらまし　実世界を対象とする知能システムの構築を目指す枠組みとして，最近注目を集めている情報統合について解説する．また，これまでに発表された研究成果の幾つかを事例として取り上げ，そこに共通する一般的な情報統合アーキテクチャについて明らかにする．さらに，情報統合の観点から，重要な要素技術であるビジョン技術に対して求められている要請や課題についても取り上げる．なお，本報告はCVCVワーキンググループの活動の一貫として書かれたものである．

CVCV-WG Special Report: Technical Review and View in Computer Vision (XIV)
- Computer Vision and Information Integration -

Shigeki NAGAYA
Information Integration Lab. Tsukuba Research Center.
Real World Computing Partnership

Abstract -This paper discusses information integration, a field that has received considerable attention in recent years, in terms of a framework for constructing an intelligent system that targets the real world. The paper will present several examples of research achievements in this area and will clarify general information-integration architecture common to this research. It will also take up the demands and issues associated with vision technology, which is considered an essential element from the viewpoint of information integration. This report is a part of the activities of the CVCV working group of IPSJ.

1. まえがき

約一年ほど前，人工知能学会誌の「AIマップ」というコラムで，「ビジョン研究から見た統合アーキテクチャ」と題して，ビジョンにおける情報統合について，松山先生，大田先生，金谷先生，上田先生らによる誌上討論が行われた[松山95][AIマップ96]．松山先生が意見を論文として提示し，残る先生方がこれに対する質問・コメントを寄せるという形で議論が交わされたが，そこでの議論は非常に興味深く，多くの示唆を与えるものであった．

ビジョン研究を代表する先生方による誌上討論が行われたという事実が指し示すように，「情報統合」というキーワードが，ビジョン研究者の間でも最近注目を浴びている．情報統合が着目されてきた背景には，より一般的な立場から情報の統合について考え，新しい情報処理のアーキテクチャ(応用システムの設計原理)にまで高めようという気運の高まりがあることや，かつての人工知能研究が果たせなかった，実世界への取り組みを表明し，実世界を扱う研究ビークルに積極的に取り組んでいることが挙げられる．

本稿では，この「情報統合」と呼ばれる枠組みについて述べる．第二章では情報統合とは何か，その目的や位置づけについて，第三章では少しずつ見えつつある一般的な情報統合のアーキテクチャについて述べる．第四章では最新の研究成果の内，幾つかの事例を紹介する．また，ビジョン研究者にとって，有益な情報たるよう，第五章では，情報統合に使われるビジョン技術に対する制約や課題についても触れるつもりである．以上，RWCで実際に情報統合研究に携わるものとしての立場からの見解も交えて説明してゆきたい．

2. 情報統合

2.1 背景

情報統合とは，複数の情報をまとめ，処理することによって，個別の情報からは得られないような情報を得ることである．「個別の情報からは得られない」という言葉には，次の二つの意味がある．

一つは，両眼立体視のように，異なる複数の部分情報から全く新しい情報を取り出すという意味である．一組のステレオ画像を考えたとき，個々の画像には奥行き情報は含まれていないが，画像中の対応点について三角測量の原理を適用することによって奥行き情報を求められる．このように，両眼からの情報を統合することによって，初めて奥行き情報という全く別の情報を得ることができる．

もう一つは，冗長な複数の部分情報をまとめることによって，取り出される情報の精度や信頼性を高めるという意味である．先の両眼立体視を三枚以上の画像に広げた多眼視について考えてみる．立体視では，原理的に，二枚の画像で対応点が求まればその点の空間情報は決定される．したがって，三枚以上の画像を用いることは冗長である．しかし，この冗長性を活用することにより，両眼立体視おける，誤対応，位置決め精度，隠れなどの問題を解決し，奥行き情報の精度・信頼性を高めることができる．

このように，情報を統合することによって，潜在化していた情報を顕在化させたり，情報の冗長性を利用した頑健性を実現できることがわかる．

2.2 枠組みとしての情報統合

前節で述べた二つの例は情報統合の本質をよく表しているが，同時に考え方そのものが全く新しいわけでもないことも示している．実際，これまでにも，画像や音声認識の分野で，○○理解システムなどの形で，情報の統合は既に行われていた．また，人工知能研究では，「記号とパターンの統合」，或いは「Symbol Grounding」などと呼ばれ，重要な研究課題の一つとして，知られていた．

こうした歴史がありながら，第一章で取り上げた誌上討論が指し示すように，情報統合が最近注目されている背景には，より一般的な立場から情報統合について考え，新しい情報処理のアーキテクチャ(応用システムの設計原理)にまで高めようという機運の高まりがある．これまでのように，様々な応用対象ごとにヒューリスティックにシステム設計を行うのではなく，より一般的な観点からシステムを設計する指針を求めようという訳である．また，このアーキテクチャの追求を通じて，それまで同時に取り扱う機会の少なかった，音声や動画像や自然言語理解などの個別技術それ自体も発展するのではないかという期待が大きいためでもある．

2.3 実世界への取り組み

情報統合が注目されているもう一つの理由に，実世界への取り組みを表明している点にある．かつての人工知能研究が，選ばれた小規模のデータに対してのみ有効であり，Toy worldと批判された反省に立って，実世界を対象として再び挑戦しようというわけである．この再挑戦の原動力には，近年のハードウエアのすさまじい進化や，認知科学，生理学等の学

際領域からの寄与，人工知能研究におけるシンボルとパターンの統合に関する多大な成果が得られたことが挙げられる．しかし，これらの他に，この取り組みに自信を与える事実として，パターン認識の分野の一つである，音声認識技術が実用化の段階に入りつつあることが挙げられる．

現在，音声認識技術は，基礎研究のフェーズを脱して，一つの製品[*1]として成り立つようになりつつある[*2]．リアルタイムでの認識はもちろん，約10万語程度の大規模な数の単語を，不特定話者の場合で92%程度，特定話者の場合で97%程度の認識率が得られるようになっている[*3]．また，音声入力の仕方も，単語ごとに区切って入力するだけではなく，ユーザが連続発声した音声も認識できるようになっている．

画像認識と比較的共通する技術を有する音声認識において，こうした技術的な成功が得られた背景には，言語情報[*4]と統合によって非常に良い結果が得られたためである．

1980年代の後半頃まで，主要な研究テーマとなっていたのは，様々な音響特徴量に基づいた音声認識技術の研究であった．このころの標準的な認識率は，約1000程度の単語に対して，せいぜい85%程度であった．音響特徴量から得られる認識精度は既に限界であり，単語数が増えると認識率が低下するというトレードオフの状態にあった．ところが，1990年代の初頭になると，ATIS等の大規模な音声対話データベースから学習・生成した言語情報と音響特徴量を統合するアプローチが注目されるようになった．これにより，言語情報による制約を用いて探索範囲を十分小さく限定したり，文法的に正しくない候補を抑制することが可能になった．例えば，音声認識の過程で，以下のような二つの候補が得られた場合を考えると，

(a).私は家ません。

(b).私は行けません。

(a)と(b)の候補は音響的には非常に近い（違いは子音/k/の有無のみ）．だが，言語情報と統合することで、(a)を文法的にはありえない候補として，除くことが可能になった．こうして、大語彙に対する認識率が向上し，大語彙連続音声認識が可能になったのである[*5]．

このように，音声認識技術は実用化の点で成功をおさめつつある．言語情報との統合で用いられた方法論（大規模な音声対話データベースからの言語情報の生成など）は，情報統合研究においても，大いに参考となる．このため，ビジョンを含めた他の情報統合においても，同様のアプローチを採用することによって，成功をかなり期待できそうな状況である[*6]．こうした理由もあって，現在の情報統合研究のテーマの多くは，ヒューマンインターフェイスや自律ロボットなど，人間や実世界を対象とするビークルが積極的に選ばれている．また，そのための実世界データを大量に蓄積したデータベースの構築も熱心に進められている[*7]．

3. 情報統合アーキテクチャ

本章では，情報統合のアーキテクチャについて述べる．まず，情報統合に多大な影響を与えた[Brooks 86]によるサブサンプションアーキテクチャについて述べる．次に，これまで発表された情報統合の研究成果から，共通して浮かび上がってくるアーキテクチャについて明らかにするとともに，各種の統合方式の特徴について述べる．

3.1 Subsumption Architecture

[Brooks 86]によるサブサンプションアーキテクチャとは，自律走行ロボットに使われたモジュールのアーキテクチャである．このアーキテクチャの特徴は，従来，知能の必需品であると考えられていた知識表現やそれらを用いた推論を否定し，そのかわりに，環境に条件反射的に反応するエージェント・モジュールを用いて，それらの協調・競合によって知的に振る舞う自律走行ロボットシステムを構築した点にある．

扱う対象の抽象度の異なるモジュール群を垂直に結合し，上位のモジュールは下位のモジュールの出力を利用したり，下位のモジュールの動作を抑制することにより，全体として協調動作する．

下層にそれだけで歩き回ることができるモジュールをおき，上位の層には障害物回避を行うモジュールをかぶせ，下位のモジュールを干渉できるようにす

---

*1 読者の方々の中には，TVや新聞などでそれら製品のCMや解説記事を既にご覧になった方もいるかもいるだろう．

*2 事実，米国での音声認識に関する研究の強力なスポンサーであったDARPAは，研究フェーズが製品開発に移ったことを理由に研究予算を大幅に削減した[河合 96]．

*3 米語の場合．日本語で同程度の認識率を得るには対象となる単語数が1万語程度になる．

*4 ここでは，文脈自由文法（CFG）や、n-gram（連続して現れる単語や音韻などの組合せ。bigram,trigram）などをさす．

*5 DARPA主導による半年に一回ごとのベンチマーク競争が行われ，明確な優劣が付けられたことも大きい．

*6 だからといって，情報統合の取り組みが必ず成功するという乱暴なことをいうつもりはないが．

*7 http://www.rwcp.or.jp/people/toyoura/rwcdb/

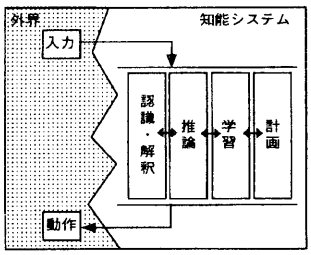

図1 一般的な情報統合アーキテクチャ

る.こうすると,下位の層だけで完全な動きが期待できるし,上位の層では障害物を避けて歩く機能だけを設計することができる.層となるモジュールを積み上げて行くことが可能であるので,より高級な機能を簡単に付与できる.サブサンプション・アーキテクチャは,システム設計の見通しの良さや動物行動学の成果[*8]ともよく一致したこともあって,情報統合のアーキテクチャにも多大な影響を与えた.

3.2 情報統合アーキテクチャの共通項

これまで提案されてきた情報統合アーキテクチャは総じて図1に示されるように,サブサンプションアーキテクチャと[Norman 88]で提唱されている認知モデルの略化したものを組み合わせたような形態となっている.ノーマンのモデルでは,実世界の認識,解釈,推論,学習,計画が縦列に接続し,何れのモジュールも処理をスキップする事はできないが,情報統合アーキテクチャでは,これらのモジュールが入力と行動の間にサブサンプションアーキテクチャのように並列に位置し,いずれかのモジュールで処理が失敗しても良いように成っている.

サブサンプションアーキテクチャとの違いは,実際のほとんどのシステムではモジュールへの入力が下位からのものしかない点である.

3.3 アーキテクチャの特徴-統合方式

アーキテクチャとしての違いが明確に現れてくる部分は入力される情報をどのように統合するかという部分である.ここでは,これら統合方式の内,代表的なものの幾つかを説明する.

(1) 複数のエージェントによる統合

自律的に情報処理する機能を有する複数のエージェントの相互作用から,実世界における集団・社会における協調メカニズムを模した群知能によって統合を行い,知能システムを構築しようとするのが,マルチエージェントによる統合である.このアプローチでは,(a) エージェント間で共有,或いは通信する情報の設計によって,システム全体の協調が設計できる,(b) エージェントごとに単に入力・動作の対を設計することで,エージェントの推論・計画・学習の各機能を構築できる,という利点がある.実際,マルチエージェントによる統合を行っているシステムのほとんどは,エージェントごとに機能を分担させたり,抽象度ごとにエージェント群を階層化するなどのモジュール設計を行って知能システムを構築している.

(2) 充足度を持つネットワークによる統合

[Oka 96]によるマルチモーダル対話システムのための情報統合方式は,連続オートマトンと呼ぶ,充足度を持つ入出力の対のネットワーク表現によって実現されている.この方式の特徴は,1) 音声・ジェスチャ認識系からの認識結果系列から有意な部分を選んでユーザの意図を取り出すことができる点と,2) 充足度による評価を行うことによってパターンによる統合を実現している点,の二つにある.

連続オートマトンの原理を示したのが図2である。アークには、入力(発話あるいはジェスチャーの認識結果)に対応したラベルが付加されている。一方,ノードは一種の変数域となっており,そのノードにいたるまでのアークの履歴とそのノードに到達する充足度の総和が書き込まれる。ネットワークに対して下位に位置する音声認識系・ジェスチャ認識系からの認識結果が,入力信号として与えられる.入力の度に,全アークにおいて,そのラベルと入力との距離によって遷移の充足度を計算され,遷移先のノードが持つ充足度と比較される.ノードが持つ充足度よりも高い場合には,遷移先のノードの充足度およびアーク履歴を上書き更新する。こうして次々と入力ごとに全ノードの充足度を更新してゆく.閾値を越えたノード

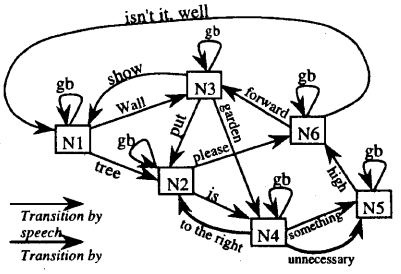

図2 連続オートマトンの原理

*8 一見複雑に見える鳥や魚の行動も,実は単純な反応の連鎖にすぎないとする考え方.Brooksの言葉を借りるならば「非常に単純なレベルの知能を調べてみると世の中の明示的な表象やモデルはまったく邪魔であり,世界をそれ自身のモデルとして用いる方がよい」というわけである.

が現れると，個々のノードごとに，そのノードにいたるまでのアーク履歴をユーザの意図の解釈結果として出力する．閾値は，ユーザ意図の出力を行うトリガーとして用いてられ，解釈結果の充足度の順に各解釈結果の中からもっとも高いものが採用される．

この統合方式の場合，ネットワークの構築が課題となる．入力・動作の組み合わせが小規模であれば，人手による設計も可能であるが，組み合わせ数が膨大になる実世界のデータに対しては，ネットワーク構造を自動的に生成する必要がある．ネットワーク構造の自己組織化手法については，[豊浦 97]による手法が報告されており，良い結果が得られつつある．

(3) 確率モデルによる統合

BayesやDempster-Shapherの確率モデルによる判別関数を用いて，複数の情報を統合する方式を採用している事例が幾つも発表されている[松山 89][Sakamoto 97]．こうした確率モデルを用いた統合方式は，HMMやニューラルネットワークのように，人手では解析・設計が困難であるような事象間の依存関係を与えたサンプルから学習できるため，統合の実装が比較的容易であるという特長を持っている．また，HMMやニューラルネットワークに比べて，サンプルデータ数が少なくてもそれなりに学習できる[*9]ことや，その学習結果が人間にわかりやすい形態であるという利点がある．

この他，[大森 95]に述べられている，結合状態を制御して学習・構築されたニューラルネットワークによる統合方式なども興味深い．

4. 事例紹介

2.1節でも述べたが，情報統合の研究テーマには，ヒューマンインターフェイスや自律ロボットなど，人間や実世界を対象とするビークルが好んで選ばれている．本章では，こうした研究事例の内，ヒューマンインターフェイスと自律動作ロボットに関する研究成果について簡単に紹介する．

4.1 音声・ジェスチャ-MM対話システム

図3は音声とジェスチャの二つのモーダルによって操作可能な対話システムである[Nagaya 96b]．このシステムの特徴は，システムとの対話時に，ユーザが自由に連続発話したりジェスチャを行うことができる点である．つまり，音声・ジェスチャのそれぞれの入力から，有為な部分（単語や身振り）を抽出して統合し，有為な操作かどうかをシステムが判断してくれる．また，同時に複数のユーザに対応でき，知的な電子書記としてユーザの協調作業支援も行ってくれる．

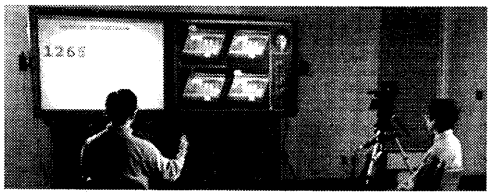

図 3 音声・ジェスチャ - マルチモーダル対話システム

4.2 擬人化エージェントMM対話システム

図4は擬人化エージェントによるマルチモーダル対話システムである[Hayamizu 97]．本システムは，先述のマルチモーダル対話システムと同様に，顔画像認識や音声認識などのさまざまなモーダルを備えた対話システムである．このシステムの特徴は，図4の画面中に示されている，非常に表情豊かで高いユーザビリティを備えた擬人化CGエージェントと，ユーザからの要求に応じた例示型の画像学習機能を備えている点である．対話を通じてエージェントに見せたことのない物体を教示することができる．

図 4 擬人化エージェントマルチモーダル対話システム
([Hayamizu 97]より引用)

4.3 手渡しロボット

図5は人間との協調作業を学習・実行する「手渡しロボット」である[Suehiro 97]．このシステムの特徴は，積み木ブロックおよび人間の手の位置・姿勢をリアルタイムで認識・理解するビジョン部を含め，システム全体が，図6に示されるエージェント・ネットワークで実装されている点である．

このシステムで提唱されている，エージェント・ネットワークは，エージェント同士で共有する情報にパターンを用い，その認識・理解を個別のエージェントに任せことによって，シンボルをエージェント内に隠

*9 事象間の依存関係を表す条件付き確率を表す分布関数が適切に推定できると仮定する

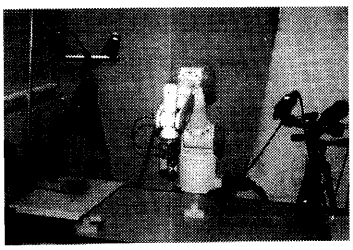

図5 手渡しロボット([Suehiro 97]より引用)

- ブロックをロボットアームで掴み，人間に手渡そうとしている -

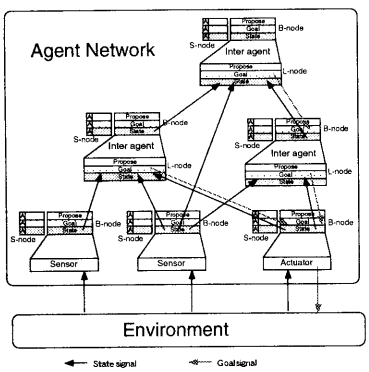

図6 パターンに基づくエージェントネットワーク

蔽する.これによって,共有する情報の相互関係や同一性の判定の問題(symbol grounding)の問題を回避できる.エージェント内部だけでなくエージェントの相互関係についても学習が可能である.

紙面の都合により割愛したが,他にも注目すべき情報統合システムは数多くある.[Matsui 97]の自律移動ロボットや,[Sakamoto97]のマルチモーダル対話システムなど,興味を持たれる読者にはそれらの論文を一読されることを強く勧める.

5. ビジョン技術に求められるもの

5.1 情報統合に適用するための必要条件

(1) ビジョンならではの利点を有する

情報統合に用いられるビジョン系は,超音波センサや赤外線センサなど,ビジョン以外のセンサ・認識系と一緒に用いられることになる.このため,システムに組み込まれるビジョン認識系は,次のような条件が要求されることになる.

1) 計算時間や価格・精度・安定性等の点で他種のセンサより優れる.
2) ビジョン系にしかできないセンシング・認識機能を提供する.例えば,顔認識等のビジョンでしか得られない情報を提供するとか,他種センサが不安定な状況下で安定に動作し,相補的な役割を果たせる等.

(2) 精度・安定性についての条件が明確である

情報を統合し,精度の高い全体情報を得るには,個別の情報の精度が非常に重要な要素となる.高い精度が得られるが,非常に不安定なアルゴリズムよりも,安定して程々の精度が得られるアルゴリズムの方が統合しやすいことがままある.

また,想定されるすべての条件下で精度がなくても,精度の得られる条件がわかっているならば,その条件下でのみ,統合に用いればよい.同様な議論は安定性についても成り立つ.

(3) リアルタイムに結果を返す

[岡96]の指摘にあるように,リアルタイム性は実世界を対象とする上で無視できない,重要な要素である.特に,マルチモーダル対話システムでは,リアルタイム(数十～数百ms)に認識結果が得られるアルゴリズムでなければならない.

注意してほしいのは,この条件が要求していることは,リアルタイムで全計算が終了という意味ではなく,単に結果を返せばよいという意味である.例えば,連続DPのように過去の計算結果を用いて差分のみを計算する構造のアルゴリズムであれば,実際の全計算時間がリアルタイムに終了しなくても,この条件を満たす.また,アルゴリズムの構成を工夫することによって,この条件をクリアすることができる.手渡しロボット[Suehiro 97]はその好例である.積み木ブロックを追跡・姿勢推定する画像センサ系を,画像全体を探索するモジュールと,姿勢推定・追跡するモジュールとにわけ,並列に動作させることにし,姿勢推定を行うモジュールは推定範囲を狭めることによってリアルタイムで動作させ,計算量の多い探索モジュールは,姿勢推定モジュールがブロックを見失ったときに動作させることによって,先に述べたような必要条件を満たしている.

5.2 ビジョン側の課題

(1) 部分情報,多様性,状況依存に関する数理モデルの研究

[AIマップ 96]で指摘されているように,これまでのビジョン研究には,数理解析の見地から頑健性,柔軟性に焦点を当てた判断,識別,認識のための研究がほとんどなかった.様々なビジョン・アルゴリズムについて,こうした理論からの成果が得られていれば,これらアルゴリズムをどのように統合すべきかという見

通しを明確に得ることができる. 今のビジョン応用システムの設計を電気回路の設計[*10]に例えるならば, 使用素子の物理的な特性を全く知らないで設計している状態である. 現状では, アルゴリズムの安定性や精度に対する経験的な知識を用いているにすぎず, こうした状況がビジョン応用システムの設計に職人芸的な難しさを与える一因であると思われる.

(2) 統合を意識したアルゴリズムの再評価・設計

前節でも述べたことだが, ビジョン系を情報統合の構成要素として用いるためには, 精度に対する条件とリアルタイム性の二つが実際にシステムを構成する上で非常に重要である.

統合を前提に, 精度の条件に関する制約を裏返して考えてみると, 条件が明確であるならば, 必ずしも全ての条件下で性能を発揮しなくても良いということである. このことは, 従来の様々なビジョンアルゴリズムについての評価は, そのまま統合要素としての評価にはあてはまらないことを示している. その意味で, 従来のアルゴリズムを評価し直すことが必要であろう.

また, 再評価と同時に, 計算アルゴリズムの構造もリアルタイム向けに再設計することも必要である.

6. むすび

本報告では, 情報統合について紹介するとともに, ビジョン技術に求められている条件や課題を中心に述べた. 情報統合は, 実世界を対象とする知能システムの構築を目指す, 新しい情報処理の枠組みであり, ビジョン技術には音声・言語技術と同様に, 強力な中枢をになうことが期待されている. ビジョン技術にとって, 情報統合は一つの応用の枠組みであるが, ビジョン技術が目指す高度な視覚機能の実現と情報統合が目指す高度な知能の実現の方向は同一であり, 本稿で述べた情報統合からの要請は, 今後のビジョン研究の一つの方向を示す重要な道標になると考えられる.

*10 実際にはパターンの統合であるからこれほど明快ではないであろうが

参考文献

[松山 95] 松山隆司, AIマップ-ビジョン研究から見た統合アーキテクチャ, 人工知能学会誌, Vol. 10, No. 6, pp.888-894, (1995).

[AIマップ 96] 大田友一, 金谷健一, 上田博唯, 松山隆司, 「AIマップ-ビジョン研究から見た統合アーキテクチャ」へのコメントと回答, 人工知能学会誌, Vol. 11, No. 2, pp.216-223, (1996).

[岡 96] 岡隆一, パターン情報から情報統合へむけて-場という概念を中心に-, 人工知能学会誌, Vol. 11, No. 2, pp.177-184, (1996).

[麻生 96] 麻生英樹, 情報統合-課題と展望-, 人工知能学会誌, Vol. 11, No. 2, pp.185-192, (1996).

[伊庭 96] 伊庭幸人, 基礎的問題から見た情報統合, 人工知能学会誌, Vol. 11, No. 2, pp.193-200, (1996).

[中島 96] 中島秀之, 情報統合のための有機的プログラミング, 人工知能学会誌, Vol. 11, No. 2, pp.201-208, (1996).

[Brooks 86] R. Brooks, A Robust Layered Control System for A Mobile Robot, IEEE Journal of Robotics and Automaton, Vol. RA-2, No.2, pp.14-23, Mar. 1986.

[Otsu 94] Nobuyuki Otsu, Toward Flexible Information Processing in Real World, RWC Technical Report, Vol. TR-94001, pp.1-6 (1994).

[大森 95] 大森隆司, 記号とパターンの注意による統合過程のモデルと情報処理アーキテクチャ, RWC情報統合ワークショップ, pp.124-134, (1995)

[國吉 95] 國吉康夫, 本村陽一, 開一夫, 原功, 麻生英樹, 松原仁, 赤穂昭太郎, 末広尚士, マルチロボットシミュレータと情報統合研究, RWC情報統合ワークショップ, pp.304-313 (1995).

[松山 89] 松山隆司, Dempster-Shaferの確率モデルに基づくEvidential Reasoningの論理的意味に関する考察, 人工知能学会誌, Vol. 4, No. 3, pp.340-350, (1989).

[湯浅 95] 湯浅夏樹, 三谷純司, 外川文雄, データベースからの学習に基づく音声と身振りの認識系統合モデル, RWC情報統合ワークショップ, pp.317-326, (1995).

[伊庭 95] 伊庭幸人, 学習と階層 -ベイズ統計の立場から-, RWC情報統合ワークショップ, pp.189-198

(1995).

[Nagaya 96a] S. Nagaya, T. Endo, Y. Itoh, J. Kiyama and R.Oka, A Proposal of Novel Information Integration Architecture - Open Coorperative Work Space-, in Proc. 1996 on International Conference on Multi-sensor Fusion and Integration for Intelligent Systems, Washinton D.C., pp.425-432, Dec. 1996.

[Oka 97] Ryuuichi Oka, Spotting Method Approach Towards Infromation Integration, RWC Sympodium, pp.175-182 (1997).

[豊浦 97] 豊浦潤, 岡隆一, テキストの知識ベース化のための自己組織化ネットワークの提案, 信学技報, NLC-97, pp.24-30 (1997).

[Mukai 97] T. Mukai, T. Nishimura, S. Nagaya, J. Kiyama, H. Kojima, Y. Itoh, S.Seki, K. Takahashi and R.Oka, Multi-modal and Realtime Dialog Through Gesture-Speech Inteface on Personal Computer, RWC Sympodium, pp.1-7 (1997).

[Hayamizu 97] S. Hayamizu, K. Sakaue, O. Hasegawa, K. Itoh, T. Yoshimura, K. Hashida, T. Akiba, H. Asoh, S. Akaho, T. Kurita, K Tanaka and N.Otsu, Multimodal Interation system at the Electrotechnical Laboratory, RWC Sympodium, pp.16-22 (1997).

[長谷川 96] 長谷川修, 坂上勝彦, 伊藤克恒, 栗田多喜夫, 速水悟, 田中和世, 大津展之, 擬人化エージェントによる対話型視覚情報学習システム, 第二回画像センシングシンポジウム, pp.145-150, (1996).

[Sakamoto 97] K. Sakamoto, H. Hinode, S. Seki, K. Watanuki, J. Kiyama and F.Togawa, Multimodal Inteface Prototype Based on Rhythm and Timing of Interaction, RWC Sympodium, pp.31-36 (1997).

[Suehiro 97] T. Suehiro, H. Takahashi and H. Yamakawa, Research on Real World Adaptable Autonomous Systems -Development of a Hand-to-Hand Robot-, RWC Sympodium, pp.398-405 (1997).

[Matsui 97] T. Matsui, H. Asoh, I. Hara and N.Yamasaki, An Office Conversant Mobile Robot That Learns by Navigation and Conversation, RWC Sympodium, pp.59-62 (1997).

[中川 88] 中川聖一, 確率モデルによる音声認識, 電子情報通信学会 (1988).

[小林 94] 小林哲則, 対話音声の認識技術, 日本音響学会誌, Vol. 50, No.7, pp.574-580 (1994).

[速水 94] 速水悟, 管村昇, 音声対話システムの研究と実用化の動向, 日本音響学会誌, Vol. 50, No.7, pp.563-567 (1994).

[河合 96] 河合剛, 多様化する米国音声言語研究 - DARPA時代の終焉と困惑-, 情処研報, SLP-12-7, pp.41-48, Jul. 1996.

[高橋 94] 高橋勝彦, 関進, 小島浩, 岡隆一, ジェスチャ動画像のスポティング認識, 信学論(D-II), J76-D-II, pp.1552-1561. (1994).

[関 95] 関進, 高橋勝彦, 小島浩, 岡隆一, ジェスチャ動画像の実時間認識システム, 1995年信学会総大 (1995).

[Campbell 96] L. Campbell, D. Becker, A. Azarbayejani, A. Bobick, A. Pentlaand, Invariant features for 3-D gesture recognition, in Proc. 1996 on International Conference on Automatic Face and Gesture Recognition, Vermont, pp.157-162, Oct. 1996.

[Nagaya 96b] S. Nagaya, S. Seki and R. Oka, A Theoretical Consideration of Pattern Space Trajectory for Gesture Spotting Recognition, in Proc. 1996 on International Conference on Automatic Face and Gesture Recog-nition, Vermont, pp.175-182, Oct. 1996.

[Yamato 92] J. Yamato, J. Ohya and K.Ishii, Recognizing human action in time-sequential images using hidden markov model, In Proc. 1992 IEEE Conference on Computer Vision and Pattern Recogntion, pp.379-385. IEEE Press, 1992.

[Ekman 69] P.Ekman and W. Friesen, The Repertoire of Non-verbal Behaviar : Categories, Origins, Usage and Coding, Semiotica, Vol. 1, pp.49-98, Mar. 1969.

[中村95] 中村裕一, 西谷正志,大田友一, プレゼンテーション映像における話者の行動理解, 信学技報, PRU95-143, pp.51-56. (1995).

[Norman 88] Norman, D.A, The Psychology of Everyday Things, N.Y., : Basic Books, Inc. (1988).

[Marr82] D. Marr, Vision, W.H. Freeman and Company, New York (1982).

[白井 87] 白井良明, パターン理解, オーム社 (1987).

[Minsky 85] Marvin Minsky, 心の科学, 産業図書株式会社 (1985).