

自然言語処理との共働を考慮した

図形認識について

高木 朗 六沢一昭 伊東幸宏 小原啓義

(早稲田大学 理工学部)

1 はじめに

視覚情報と言語情報は、例えば人間が目で見たことと言葉で表現することや、逆に言葉で言われたことを目で確かめることが出来ることから分る通り、互いに密接なつながりをもっている。従って、両者の間の情報のやりとりを可能にするという問題は興味深い問題であり、工学的にみても重要であると考えられる。しかしながら、視覚情報と言語情報はその形態が大きく異なっており、両者の間の連絡を可能にするためには、自然言語処理、視覚情報処理の両方の側からの検討が必要である。従来、視覚情報処理に関してはパターン認識、computer visionの分野は活発な研究が行われてきているが、言語情報との接続という問題に関しては余り検討が行われていない。現在、我々は簡単な二次元図形から成る動画を例にとり、言語情報との対応を考慮した特徴解析システムを試作しており、本稿ではこれについて述べる。

2 システムの構成

図1にシステムの概要を示す。システムに画面が与えられると、その中に含まれる図形の特徴、動き等を検出しその結果を言語情報に変換し易い形にまとめ出力する。入力画面は130×200のドットマトリックスを1画面とする連続した画面である。各画面に描かれる図形は、線分及び曲率一定の弧から成る二次元図形である。各ドットには色を表す数値が付随している。この図形は複数の部分から成っているものもよく、又、図形又は部分図形の内部に図形が描かれているものもよい。図形又は部分図形は影を縁取りされている。本システムは取り扱う動画の内容は、位置、大きさ等の変化に関連した自動詞的現象(「転がる」、「歩く」等)及び位置の変化に関連した他動詞的現象(「ける」、「投げる」等)である。

部に図形が描かれているものもよい。図形又は部分図形は影を縁取りされている。本システムは取り扱う動画の内容は、位置、大きさ等の変化に関連した自動詞的現象(「転がる」、「歩く」等)及び位置の変化に関連した他動詞的現象(「ける」、「投げる」等)である。

3 検出する特徴

図形処理の結果は、別稿^[1]に述べる様に意味表現に変換され、そこから自然語文が生成される。この為、図形処理における特徴検出は意味表現の構造をふまえて行われる。

動画の場合、そこから引き出される自動詞、他動詞を主動詞にした文の意味表現は図2の構造をもつ。この構造は、アスペクト、格、名詞句及び動詞

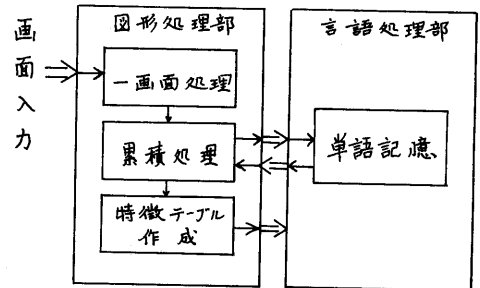


図1 システムの構成

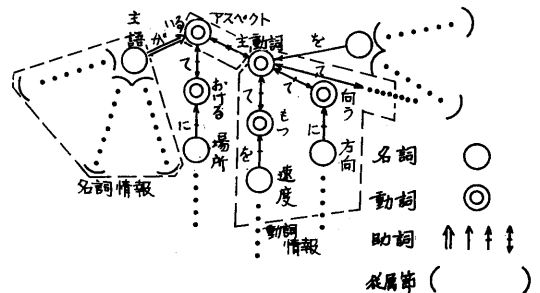


図2 文の意味表現

句の4つの部分に大分割される。この為、本システムは検出される特徴もこの4つの部分から成っている。

尚、静止画の場合、述語情報は全々名詞的特徴から導き出し、動詞情報として陽に画面中には存在しない。この為静止画における述語情報は動詞として検出されない。

(1) 名詞句に関する特徴: 図形又はその部分図形(「机」「足」等)を表現する意味表現は図3の構造をもつ。この為画面中の図形、部分図形毎にそれらを構成する線分、弧を検出し、そこから色、形、大きさ、内包する部分図形、内部図形を検出する。尚、名詞に係る連体修飾の為の情報は全々これらの特徴の中に含まれている。

(2) 動詞句に関する特徴: 属性値変化を表す自動詞(例えば「動く」「ふくらむ」等)の意味表現は図4の構造をもつ。同一図形の複数の属性が変化したり、同一図形の複数の部分が属性値変化を起し、それらが複合して一つの動詞の意味となる場合(「こぼれる」「歩く」等)、更に一つの図形(又はその部分)の動きが他の図形の動きを引き起し、それらの動きが全体として一つの動詞の意味となる場合(「ける」「投げる」等)には、図4の形の意味表現が組み合わされて一つの意味表現が構成される。この為図形処理においては、属性値変化が起っている図形又は部分図形毎に、変化している属性、変化の割合、方向(位置変化の

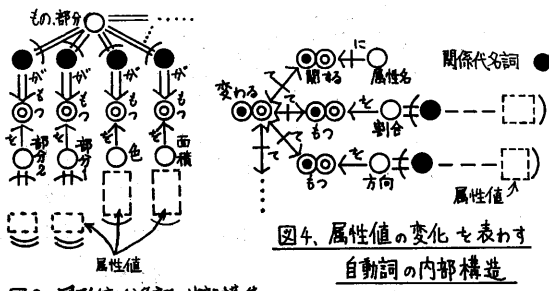


図3. 図形(部分)名詞の内部構造

図4. 属性値の変化を表す自動詞の内部構造

場合には進行方向、それ以外の場合には属性値変化の微係数の符号)を検出する。尚、動詞に係る運用修飾情報は全々これらの特徴の中に含まれている。

(3) 格に関する特徴: 本システムで扱う自動詞がとる格は主格、場所格、方向格(「～から～へ」)である。又位置変化に関する他動詞には目的格がつき作用対象を表すが、これは画面上では一つの図形(又はその部分)の他の図形への接触、離反という形をとられる。この為、図形処理においては、

・どの図形(のどの部分)のどの属性

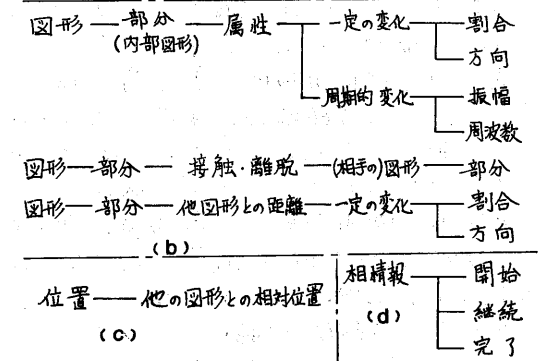
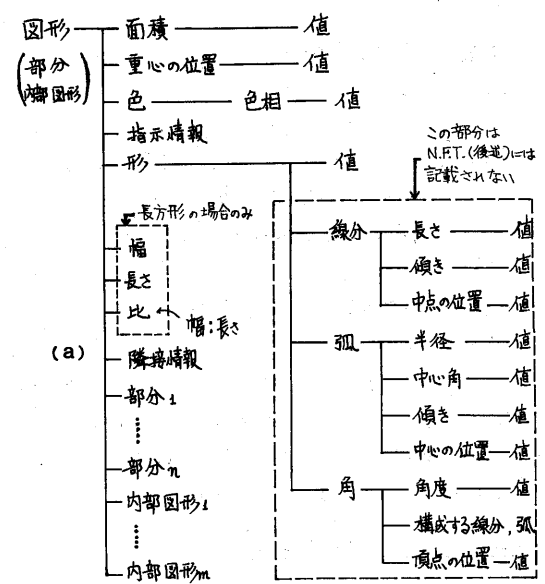


表1 検出する特徴

が変化しているか (主格)

- ・ 属性値変化を起している図形が最近傍の他の図形に対してどの様な相対位置にあるか (場所格)
- ・ 図形に位置変化が起っている場合、その進行方向又は後方にどの様な図形があるか (方向格)
- ・ 1つの図形 (又はその部分) が他の図形に接触した場合、あるいは他図形から離れた場合、その相手の図形は何か (接触の対象又は運動の出発点)

を検出する。

(4) アスペクトに関する特徴: 属性値変化が開始したか、継続しているか、完了したかの3つの場合を検出する。

表1に検出する特徴の一覧を示す。

4. 特徴解析の手順

特徴解析は、1コマ毎の画面に対して処理を行ない、各図形毎に表1(a)の特徴を検出する処理 (以下、1画面処理と呼ぶ) と、これまでの画面を通算して画面情報を累積し、最終的に属性値変化について表1(b)(c)(d)の特徴を検出する処理 (以下、累積処理と呼ぶ) に大別される。

4.1 1画面処理

図5に1画面処理の手順を示す。

(1) 輪郭及びそれを構成する線分、弧、角の検出

これは、主として白井^[2]が提案した方法に依っている。画面中の各図形に対して、まず輪郭を抽出した後、輪郭線の曲率から角点を求める。次いで、角点と角点の間の線分、弧を切り出す。同時にこれらの線分、弧の間の角度、線分の傾き、中点の位置、弧の曲率半径、中心角等が算出される。

(2) 部分分割

輪郭線を下より外折れ角に出会った場合、これまでにたどってきた線分、弧を延長してみ、これに重なる線分、

弧が見つかり、延長した線分、弧を境界線として部分分割する (図6)。部分分割が行われした後、それ以上分割出来ない最小の閉領域を以後図形単位と呼ぶ (図6ではAJKL, BCDE, FGHI, TMNORS, OPQR, U)。図形単位毎に面積 (ドットの数)、重心の位置が算出される。尚、処理例に示す様に、ロボットの手足等は線だけで描かれている。この場合には部分分割は行われず、従って1画面処理の段階では部分図形とは判別されない。

(3) 隣接関係の検出

隣接する図形単位の有無を調べる。互いに隣接する図形単位がある時、その図形単位に隣接する旨と、その図形に対する相対位置 (上下左右等8方向) が記録され、更に境界となる線分、弧が検出される。但し、(2)で部分分割された図形単位は、“隣接”の代わりに同一図形を構成する部分である旨が記録される。この隣接関係は、後述する補完処理及びどの図形単位が集まって一つの図形を構成しているかの判断等に用いられる。

(4) 補完情報検出

図7の様に図形が互いに接して見える場合、四角形が円によって隠された

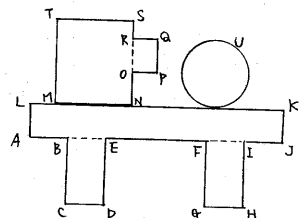
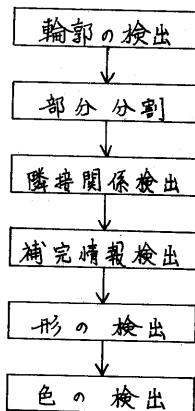


図6 部分分割

図5 1画面解析の処理手順

いと見ることも出来るし、又(a)の場合 ABCDEFGH という輪郭をもつ図形が円と接しているとも見ることも出来る。どちらが正しいかは他の画面を見なければ分からない。そこでこの為の準備として、図7の様に他の図形との境界線に接続する2辺を延長し、これらが重なるか(a)、又は交わる(b)場合、延長して得られる形が本来の形にある可能性があり、この際補うべき線分、弧(図7では EF と、OQ、PQ)が記録される。

(5) 形の検出

輪郭を構成している線分、弧、角から、その図形単位の形が決定される。現在の所、本システムでは、以下の形の検出を行っている。

円、三角形、正方形、長方形、菱形、平行四辺形、台形、四角形

(6) 色の検出

図形単位毎に、その内部のドットに付随している数値を読み取る。

4.2 累積処理

図8に処理手順を示す。最初の画面が与えられると、1画面解析の結果得られた表1(a)に關するデータが図形単位毎に累積処理用のバッファに記憶される。次いで、2枚目の画面が与えられるとこの1画面処理の結果がバッファの内容と比較される。この際、位置や大きさが変わってしまったり、他の図形単位の後に隠れて形が変わってしまった図形単位の存在を考慮しつつ、2枚目の画面の各図形単位とバッファ中の各図形単位の対応付けが行われ、これに基づいて、バッファ中の該当する図形単位の属性に新しい属性値が書き加えられる。これは画面の経過と共に、各図形単位の属性値などの様に変化していったかを知るために行われる処理である。

対応付けが完了すると、隣接している図形単位がある場合、これらが同一

図形を構成する部分なのか、下ま下ま隣接しているだけでなく、実際には別の図形なのかを判断するために、不動点検出と概念記憶参照が行われる。

同様に、3枚目以降の画面に対して同じ処理が繰り返され、最終的に、画面中に存在した図形及びその属性値変化に關するデータがバッファ中に蓄積される。

以下に、各部の処理について述べる。

(1) 対応付け処理

対応付け処理は以下の手順で行なう。

1) その時点のバッファ中の各図形単位について、それまでに記録されている属性値から、新しく入力された画面において各図形単位がどのような属性値を持つかを属性毎に予測する(形は除く)。

2) バッファ中の図形単位から1つの図形単位を選び(これをSとす)、以下の手順で新画面中の図形単位と対応付ける。

① 新画面中の図形単位について、Sの予想位置に最も近いものから順に優先順位を付ける。

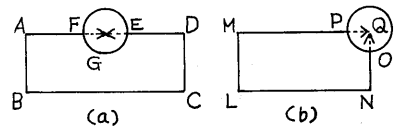


図7 補充情報検出

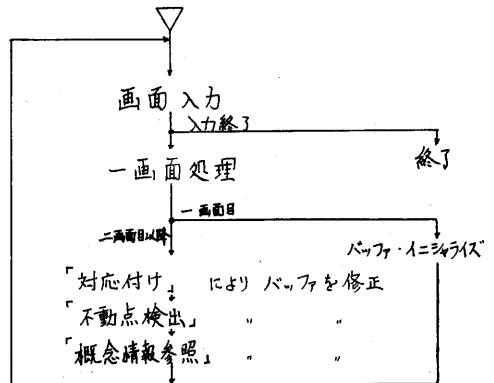


図8 累積処理

② その順位の高いものから順に、その各属性値の予割値と一定の閾値の範囲内で一致する図形単位があるかを調べる。但し、形を比較する際には、単に見えてくる形だけでなく、他の考えられる可能性も考慮する(後述)。一致する図形単位が見つかる両者を対応付ける。見つからなかった場合(例えば対応すべき図形単位が他の図形単位によって完全に隠された場合等)にはその旨を記録する。その上で新しいXについて処理を繰り返す。

3) バッファ中の全ての図形単位が新画面中の図形単位と対応が付くと、バッファ中の各図形単位毎に以下の処理を行なう。

① 後述する形の修正が生じ、形、大きさの真の値がわかった場合、バッファ中の該当データが修正される。

② これまでの画面で互いに接していた図形単位が新画面で離れてしまった場合、隣接関係が修正される。

③ 新画面中の図形単位の新しい属性値もバッファ中の該当図形単位の対応する属性の欄に書き加える。これにより、図形単位毎に属性値の変化系列が作られ、そこから属性値の変化の割合、周期性の有無、変化の方向等が検出される。

④ 新画面でこれまでに離れていた図形が接触した場合、その旨が記録される。又、逆にこれまでに隣接していた図形が離れた場合も、その旨が記録され、更に隣接関係のデータが修正される。

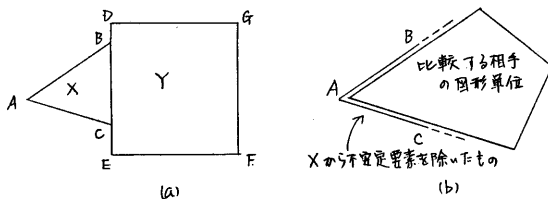


図9 不安定な辺と角

尚、前述の2)②における形の比較において、見えてくる形の直接の比較では一致する図形単位が見つからなかった場合、次の処理を行なう。

補充: バッファもしくは新画面中の着目している図形単位に補充情報が検出されてくる時には、これにより補充を行なうか否かを形を比較する。

不安定要素の除去: 図9の様、ある図形単位X, Yが隣接し、Xの1つの辺全体が境界線になっていて、Yの後にXの一部が隠れている可能性がある。このためこの場合の境界線は必ずしも図形単位を構成するとは限らず、これをそのまま用いると対応付けに失敗する場合が起こり得る。更に、この時、その辺に連続する辺(図示はAB, AC)の長さ及び角(図示はLABC, LACB)も不安定な情報となる。このためバッファもしくは新画面中の着目している図形単位が、こうした不安定な情報をもっている時は、これを除去し、安定な辺のみを取り出して、一方の図形単位の形が、これを包含する形(図9(b))になっていれば、形が一致したと判断する。

変形: 動画の場合、例えばロボット足の様に、折れ曲って変形してしまう図形単位が存在する。こうした場合に対処するために、新画面とバッファ中の図形単位の形が一致しない場合、長さも含めて一方の形が他方を伸ばした形になってくるか否かを調べ、形の比較を行なう。

これらの処理は、画面の性質によって順序を変えることが出来るが現在は①補充②変形③変形+不安定要素の除去④不安定要素の除去、としている。

以上の対応付け処理が終わると、バッファ中に隣接関係が検出されてくる図形単位について、これがそれぞれ接している別の図形があるのか、同一図

形を構成する部分図形であるのかを判断するために、以下の処理を行なう。

(2) 不動点検出

例えば図10の様に、2つの図形単位が重なって互いに接している、位置変化がある場合、画面の経過を追って行くと、両者の間には常に相対位置が不変な点がある、両者が同一図形の構成部分であることがわかる場合がある。このため、着目して互い図形単位の輪郭の角点、角点と角点との間の中点及び重心の、隣接図形単位の重心に対する相対位置を調べ、時間的に不変な点があるか否かを見る。あれば、両者を同一図形の部分図形とみなす。

(3) 概念記憶参照

ここまでの処理を終えた時点でも、例えば図11の机とコップは、依然として互いに隣接する長方形の集まりとしか判断できない。このため、バッファ中の互いに隣接する図形単位を1まとまりとし、後述するN.F.T.の形に変換し言語処理⁽¹⁾へ送る。そこでN.F.T.は意味表現に変換され、概念(単語)記憶中の意味表現と比較される。概念記憶中の意味表現の内N.F.T.から変換された意味表現に含まれる最大のものが選ばれ、これがN.F.T.に逆変換される図形処理へ返される。これを基に同一図形を構成する図形単位がまとめられ、同一図形の部分である旨が記録される。これを隣接図形単位のまとまりから差し引く、残りの部分について同様の処理を繰り返す。

尚、ある時点でも、全ての隣接図形単

位について、どことどことがまとまって同一図形を構成するかが明らかになると、新しい図形が画面に登場しない限り、それ以降の累積処理においては、概念記憶参照は行われたい。

以上の処理によりバッファには

- (i) 属性値変化のあった図形単位について属性値変化の開始・終了時刻、種類(一定あるいは周期的)、変化の様子(一定ならば割合、周期的ならば振幅、周期)、属性が位置で一定の変化の場合進行方向及び逆方向にある図形
 - (ii) 位置が一定に変化した図形単位について進行方向及び逆方向にある図形との距離の変化の様子
 - (iii) 各図形単位について隣接した図形単位とその開始・終了時刻、隣接の方向
- が記録される。

4.3. 検出した情報のまとめ

検出された特徴は表1の木構造に合わせマリスト形式にまとめられる。図12に例を示す。

表1(a)に対応して名詞情報を表わす部分を名詞特徴テーブル(Noun Feature Table 以下N.F.T.と略す)、(b)に対応して動詞情報の様子を表わす部分を動詞特徴テーブル(Verb Feature Table 以下V.F.T.と略す)、(c)に対応して(b)を表わす現象の起る場所を表わす部分をLOCATION、(d)に対応して相情報を表わす部分をASPECTと呼ぶ。これらを全てまとめた命題特徴テーブル(Proposition Feature Table 以下

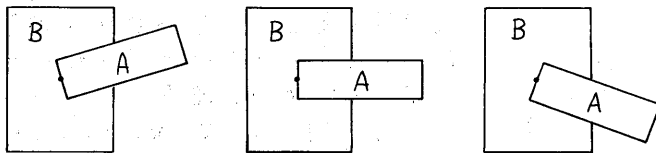


図10 不動点検出 (・不動点)

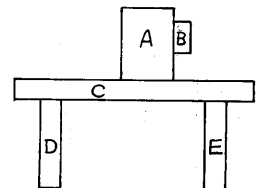


図11 机とコップ

P.F.T.と略す)と併ぶ。

'OBJECT' は図形, 'SUBPART' は部分, 'INOBJ' は内部図形を表す。

NFT では, 図形が複数の部分から成っている場合, それらのうちで最も多くの他の部分図形に接続しているもの, もしくは面積が最も大きいものを SUBPART 1 とし, 残りを SUBPART 2 ~ とし OBJECT の下に書かれる。SUBPART 2 以降の部分には SUBPART 1 に対する相対位置が書かれる。OBJECT のすぐ下の SIZE, COLOR はそれぞれ図形全体の面積, 色を表す(部分により色が異なる場合, 色は不定となる)。

図形が1つの部分から成っている場合, OBJECT のすぐ下に形の記述が書かれ, SUBPART は付かない。

VFT には, 同一図形について同時に起こった属性値変化が1つにまとめられ書かれる。

同一図形の複数の部分について属性値変化が検出された場合, OBJECT の下に属性値変化の検出された部分

```

(OBJECT 1 (SUBPART 1 (COLOR (HUE 15))
  (SHAPE (TRIANGLE))
  (SIZE 200))
  (SUBPART 2 (LOC (SUBPART 1)
    (YAXIS (+)))
    (COLOR (HUE 30))
    ...
(OBJECT 2 (COLOR (HUE 60))
  (SHAPE (CIRCLE))
  (SIZE 300)
  ...
(VFT (OBJECT 2 (LOC (CHANGE (CONSTANTLY)
  (RATE 10)
  (XAXIS (-))
  (TO (OBJECT 1))))))
(LOCATION ((OBJECT 3) (YAXIS (+))))
(ASPECT (PROGRESSIVE))
  
```

図12 P.F.T. の例

が並列に書かれる。また同一図形あるいは部分について複数の属性値変化が検出された場合, OBJECT あるいは SUBPART の下に, 変化のあった属性名が並列に書かれる。

LOCATION は, 基準となる位置を与える他図形がない場合, 空欄となる。

5 処理例

図13(a), 14(a)に入力図形の例を示す。又, その処理結果を図13(b)(c), 14(b)を示す。

```

(OBJECT 1 (COLOR (HUE 20)) ←色相値が20の色
  (SHAPE (CIRCLE)) ←形が円
  (SIZE 314) ←面積314
(OBJECT 2 (COLOR (HUE 10))
  (SHAPE (CIRCLE))
  (SIZE 201))
(OBJECT 3 (SUBPART 1 (COLOR (HUE 30)) ←形が長方形
  (SHAPE (RECTANGLE)) ←長39 幅14
  (LENGTH 39)
  (WIDTH 14)
  (RATIO 35) ←比(幅/長さ×100)35
  (SIZE 409))
  (SUBPART 2 (LOC (SUBPART 1) } 胸の後に接続
    (ZAXIS (-)))
  (COLOR (HUE 30))
  (SHAPE (RECTANGLE)) ←面積不定
  (SIZE))
  (SUBPART 3 (LOC (SUBPART 1) } 胸の下.及び
    (YAXIS (-)) } 右脚の前に接続
    (SUBPART 4)
    (ZAXIS (+)))
  (COLOR (HUE 30))
  (SHAPE (RECTANGLE))
  (LENGTH 35)
  (WIDTH 7)
  (RATIO 19)
  (SIZE 195))
  (SUBPART 4 (LOC (SUBPART 1)
    (YAXIS (-))
    (SUBPART 3)
    (ZAXIS (-)))
  (COLOR (HUE 30))
  (SHAPE (TRAPEZOID)) ←形が台形
  (SIZE))
  (SUBPART 5 (LOC (SUBPART 1)
    (YAXIS (+))
    (COLOR (HUE 30))
    (SHAPE (CIRCLE))
    (SIZE 201))
  (SUBPART 6 (LOC (SUBPART 1)
    (ZAXIS (+))
    (COLOR (HUE 30))
    (SHAPE (RECTANGLE))
    (LENGTH 28)
    (WIDTH 7)
    (RATIO 25)
    (SIZE 223))
  (SIZE)) ←ロボット全体の面積は不定
  (COLOR (HUE 30))) ←ロボット全体の色は色相値30
  
```

図13(b) 検出された名詞情報

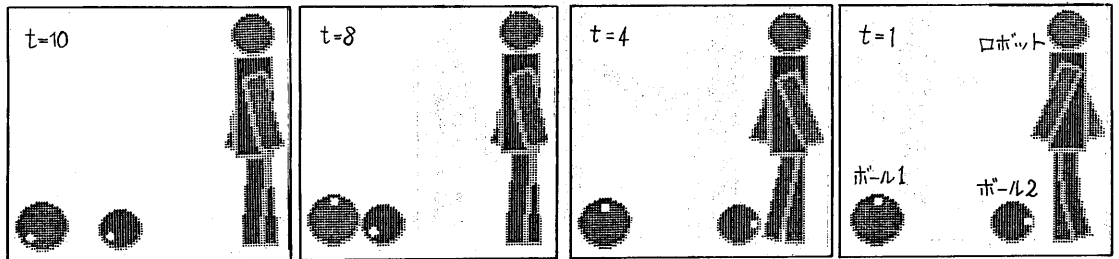


図13(a) 入力図形(一部のみ)

```

( (VFT (OBJECT 3 (SUBPART 3 (LOC (CHANGE (CONSTANTLY) (RATE 2) (XAXIS (-)))))) } (ロボットの左脚の)位置が、一定に、左へ(X方向マイナス)、2の割合で"変化
  (LOCATION) (PROGRESSIVE))) ← 継続
  (ASPECT (PROGRESSIVE))) ← 継続
  (VFT (OBJECT 3 (SUBPART 3 (TOUCH (OBJECT 2)))) (ロボットの左脚が) ボール2に 接触
  (LOCATION) (PERFECT))) ← 完了
  (ASPECT (PERFECT))) ← 完了
  (VFT (OBJECT 2 (SEPARATE (OBJECT 3 (SUBPART 3)))) (ボール2が) ロボットの左脚 から 離れる
  (LOCATION) (PERFECT))) ← ボール2
  (ASPECT (PERFECT))) ← ボール2
  (VFT (OBJECT 2 (LOC (CHANGE (CONSTANTLY) (RATE 8) (XAXIS (-)) (TO (OBJECT 1)) (FROM (OBJECT 3)))))) } (ボール2の)位置が、一定に、左へ、8の割合で、
    (ROTATE (CONSTANTLY) (ANGV 60) (SIGN (+))) } (ボール2の)角度が、一定に、60の割合で 増加 (時計と逆方向に回転)
    (DISTANCE (OBJECT 1) (CHANGE (RATE 7) (SIGN (-))))
    (DISTANCE (OBJECT 3) (CHANGE (RATE 8) (SIGN (+))))
  (LOCATION) (ASPECT (BEGIN))) ← 開始
  (VFT (OBJECT 2 (TOUCH (OBJECT 1))))
  (LOCATION) (ASPECT (PERFECT)))
  (VFT (OBJECT 1 (SEPARATE (OBJECT 2))))
  (LOCATION) (ASPECT (PERFECT))) ← ボール1
  (VFT (OBJECT 1 (LOC (CHANGE (CONSTANTLY) (RATE 6) (XAXIS (-)) (FROM (OBJECT 2))))))
    (ROTATE (CONSTANTLY) (ANGV 30) (SIGN (+)))
    (DISTANCE (OBJECT 2) (CHANGE (RATE 5) (SIGN (+))))
  (LOCATION) (ASPECT (BEGIN)))

```

6 まとめ

以上、簡単な二次元図形から成る動画を対象とする、言語情報との対応を考慮した特徴解析システムについて述べた。現在、一部の他動詞情報(「ける」「投げる」等)の検出が可能であるが、まだ簡単なものであり、これを拡張していくことが今後の課題である。

図 13(c) 検出された動詞情報(一部のみ)

(参考文献)

- 1) 高木, 伊東 他 (1983) 「視覚情報からの自然語文の生成について」 情処学会 知識工学と人工知能研究会 (1983 2月)
- 2) Shirai, Y. (1975) Analyzing intensity arrays using knowledge about scenes..In Winston (ed.) The Psychology of Computer Vision.
- 3) D.L.Waltz (1981) Toward a detailed model of processing for language describing the physical world. IJCAI 1981
- 4) 曾我 他 (1980) 「言語・画像相互参照によるプロットの把握」 信学技報 AL 80-85

```

( ( (VFT (OBJECT 1 (TOUCH (OBJECT 2 (SUBPART 6)))) (ボール (ロボットの左脚)
  (LOCATION) (PROGRESSIVE)))
  (VFT (OBJECT 2 (SUBPART 6 (TOUCH (OBJECT 1)))) (PROGRESSIVE)))
  (LOCATION) (ASPECT (PERFECT)))
  (VFT (OBJECT 2 (SUBPART 6 (ROTATE (CONSTANTLY) (ANGV 13) (SIGN (-))))))
  (LOCATION) (ASPECT (PROGRESSIVE)))
  (VFT (OBJECT 2 (SUBPART 6 (SEPARATE (OBJECT 1)))) (PERFECT)))
  (LOCATION) (ASPECT (PERFECT)))
  (VFT (OBJECT 1 (LOC (CHANGE (CONSTANTLY) (RATE 10) (XAXIS (-)) (FROM (OBJECT 2))))))
    (DISTANCE (OBJECT 2) (CHANGE (RATE 7) (SIGN (+))))
  (LOCATION) (ASPECT (BEGIN)))

```

図 14(b) 検出された動詞情報(一部のみ)

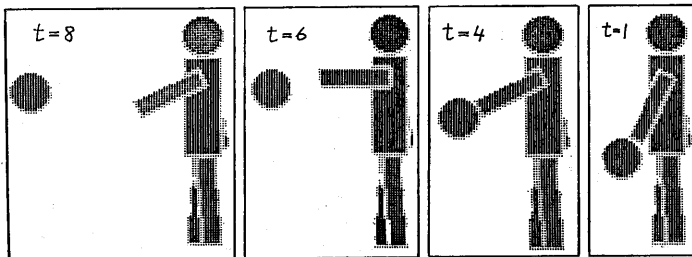


図 14(a) 入力図形(一部のみ)