

対話支援型問題解決システムにおける ジェスチャ解析の一手法

賀川経夫 西山泰幸 遠藤勉

(大分大学工学部知能情報システム工学科)

1 はじめに

本研究では、小学校1年の算数のドリルテキストを対象とし、教師である人間との対話を用いて問題の解決や知識の獲得を行なうシステムの構築に取り組んでいる。実際のドリルテキストを用いた対話においては、言語だけでなく図を指示したり、線や丸を描く等の動作が伴うのが普通である。そこで、本システムにおいても、より円滑で自然な対話を実現するためには、システムが教師の各種動作を認識する必要がある。さらに、このようなジェスチャ解析を行なう場合、対話文理解処理との協調も考慮しなければならない。

本システムでは、教師のジェスチャはCCDカメラを用いて連続画像として入力される。本稿では、対話と共に行なわれるジェスチャ解析のための連続画像の処理手法に関する検討を行なう。

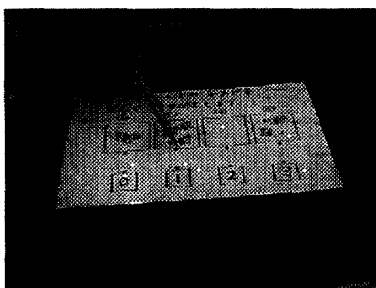


図 1: 連続画像 (1 フレーム分)

2 ジェスチャ解析の概要

本研究での対話に伴うジェスチャの解析においては、指示物体 (ペン等) の先の動きに着目する。すなわち、連続画像の各フレームからペン先に相当する点を指示点として捉え、あらかじめシステムが持っているテキスト画像との対応づけを行ない、その点の軌跡からジェスチャの特徴抽出を行なう。なお、連続画像はビデオレ-

ト (30 フレーム / 秒) で取り込んでいる。入力画像の一例を図 1 に示す。

2.1 指示点の抽出

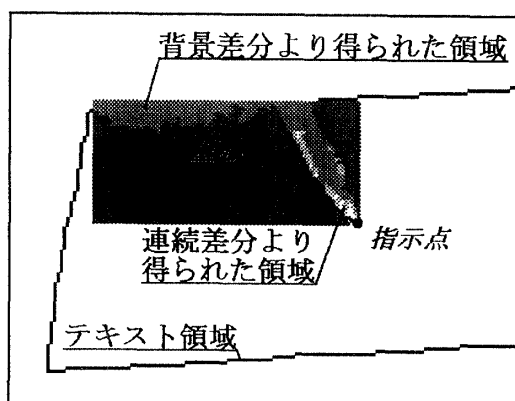


図 2: 指示点の検出

指示点の抽出においては、あらかじめ取り込まれた背景画像 (手やペン等の指示物体が存在していない時のドリルテキストの画像) との差分画像である背景差分と各フレーム間の連続差分を用いる。

まず背景差分を用いて指示物体領域とその重心を求め、次に連続差分画像において指示物体領域と重なっている部分を求める。そして、その重なった領域内の点のうち、重心点から最も距離の大きな点を指示点の候補とする。このとき前フレームでの指示点との距離が閾値より大きな場合は、指示点の検出を行なわない。また、重なった領域が得られなかった場合は、前フレームでの指示点をそのまま現フレームでの指示点とする。指示点抽出の概要を図 2 に示す。

2.2 知識との対応づけ

指示点の抽出を行なった後、システムが知識として持っているテキスト画像との対応づけを行なう。テキスト画像に関しては、あらかじめ図形や文字等の認識処理が行なわれており、図形構造表現という形式で構造化された知識としてシステム内に保持されている。従って、この画像中のどこに指示点が存在しているかを検出することにより、テキスト中のどの対象物を指示しているのかということが明確になる。対応づけの方法はまず前述の背景差分を用いてあらかじめ処理画像内でのテキスト

A Method of Gesture Analysis for Dialogue-Based
Problem Solving System

Tsuneo KAGAWA, Yasuyuki NISHIYAMA,
Tsutomu ENDO

OITA University, Dannoharu 700, Oita, Japan.

の領域を抽出しておき、その領域よりテキストを構成する4本の直線とそれらの交点を求める。次に得られた直線を等分に分割し、テキスト領域にグリッドを作成し、そのグリッドを用いて知識として保持している画像の座標を計算し対応づけを行なっていく。(図3)

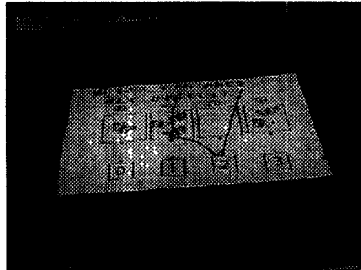


図3: 指示点の追跡

上記の処理によって得られた点は、 $point(i, x, y)$ という形式により点列として記述される。ここで、 i は連続画像のフレーム番号、 x, y はテキスト中の座標である。

2.3 動作の特徴抽出

教師のジェスチャ動作に対応した指示点系列の特徴抽出を行なうために $point(i, x, y)$ の集合の分割を行なう。分割によって得られた指示点の集合を操作セグメントと呼ぶことにする。操作セグメントを求める手続きは以下の通りである。

- (1) $d_0(=distance(point(i, x, y), point(i-1, x, y)))$, $d_1(=distance(point(i, x, y), point(i+1, x, y)))$ を求め、その比 $\frac{d_0}{d_1}$ が $\frac{d_0}{d_1} < TH_0, TH_1 < \frac{d_0}{d_1}$ を満たす時その点を分割点とする。すなわち指示点の前後の距離に大きな変化があった場合に、その点を特徴点とし、その点を基準として分割を行なう。
- (2) (1)によって得られた各セグメントの全長と始点-終点間の距離を用いて *linear* (直線), *circular* (円), *dot-shaped* (停留, 点) の3種類のタイプに分類する。
- (3) 各セグメントをそのタイプによって方法を変えて再分割処理を施す。特に、前後の指示点の移動方向が大きく変わっているところは特徴点とされその点で分割される。

最終的に得られた操作セグメントのうち、そのセグメントを構成する点の個数が閾値以上であるものに対して、特徴の抽出を行なう。特徴の抽出は、システムがあらかじめ持っている知識(図形構造表現)と対応づけることにより行なう。得られる特徴としては、以下に挙げるようなものである。

- *linear*: 線の始点と終点のそれぞれ再近傍にある対象物のインスタンスは異なるがクラスが等しい場合 *connect*(図形要素1, 図形要素2)とする。対話解釈においては、「せんでむすぶ」等の語と対応づけさ

れる。

- *circular*: 円で囲まれている対象物をそのクラス毎に分類し *multi-point*(対象物1, ..., 対象物n)とする。「まるでかこむ」、「これらのもの」等の語と対応。
- *dot-shaped*: 停留点の再近傍にある対象物を用いて *single-point* (対象物)とする。「これ」等の語に対応。

3 実験

本手法を用いて実験を行なった。教師の指示点の軌跡を図4に、これから抽出された操作セグメントとその特徴を図5に示す。

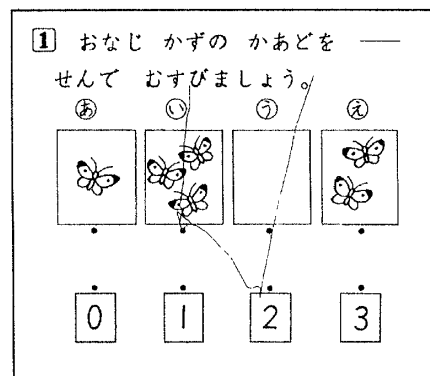


図4: 実験結果

```
seg([13,dot-shaped,[[single-point,[p_form(点2)]]]]).
seg([15,circular,[[single-point,[p_form(点2)]]]]).
seg([17,linear,[[connect,[p_form(点2),p_form(点7)]]]]).
seg([24,dot-shaped,[[single-point,[p_form(点7)]]]]).
```

図5: 操作セグメント

4 おわりに

連続画像として入力されたジェスチャを解釈するための手法に関して述べた。本手法においては、知識として持っている画像との対応づけを行なう際に、指示点の奥行き情報(Z座標)情報は考慮していない。そのため、対応づけの際に誤差が大きくなることがあった。今後の課題としては、(1)対話解析処理との統合、(2)「ぬりつぶす」「せんをかく」等の他の動作との対応づけ、(3)3次元的な動作の処理手法といったものが挙げられる。

参考文献

- [1] T.Endo,T.kagawa:Cooperative Understanding of Utterances and Gestures in Dialogue-Based Problem Solving System,Proc. PACLING'97(1997)
- [2] 岩井, 八木, 谷内田: 単眼動画からの手の3次元運動と位置の推定, 信学論, Vol.J80-D-II, No.1, pp.44-55(1997)
- [3] 賀川, 渡辺, 遠藤: 言語情報を用いた図形の解釈に関する一手法, 信学技報, TL96-5, pp.11-20(1996)