

ペン型スキャナを用いたマルチモーダル統合認識

4 R - 4

富田民則 坂尾秀樹 難波康晴 青島弘和 柳 邦宏

(株)日立製作所 システム開発研究所

1. はじめに

紙上に印刷された文字をはじめ、実世界の対象を手軽にコンピュータに入力することが可能なペン型スキャナを試作した。このペン型スキャナにより書類上の所望の部分の文字を部分的にピックアップできる。本稿では、モダリティの持つ情報を互いに補足するマルチモーダルインタフェースの特性[1]を利用した、マルチモーダル統合認識の一手法と、その手法に基づき、ペン型スキャナを用いて試作した統合認識システムについて述べる。

2. ペン型スキャナ

試作したペン型スキャナの外觀を図1に示す。ペン軸に小型 CCD カメラが取り付けられており、ペンを書類に向けるとペン先周辺の紙面の映像が CCD カメラによって映される。その映像はパーソナルコンピュータ上にモニタ表示することが可能である。ユーザはその内容を確認しながらペン側面に用意されたボタンを押すことで、静止画像としてキャプチャすることができる。

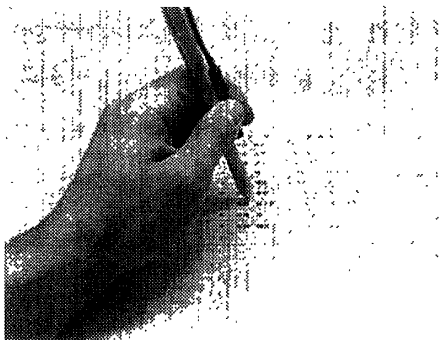


図1 ペン型スキャナ外觀

ペン型スキャナは、ポインティングデバイスであるペンと画像入力装置を一体化したものであるため、書類の指定した一部分をピックアップして

Multi-Modal Analyzer using Pen-Scanner.
Taminori TOMITA, Hideki SAKAO, Yasuharu NAMBA,
Hirokazu AOSHIMA and Kunihiro YANAGI
Systems Development Laboratory, Hitachi, Ltd.

入力するだけでなく、位置の指示と、対象の入力を同時に行うことが可能である。したがって、従来のマルチモーダルユーザインタフェースにおいて操作対象の指示がモニタ画面等に限定されがちであった点を、実世界に手軽に広げる有効な入力装置であると考えている。

3. 統合認識

マルチモーダル統合認識の一例として、音声認識結果をキーとして印刷文字の誤認識の訂正を行うケースを考えた。図2に示すように音声入力「電話番号」とともに印刷文字の入力があった場合、その印刷文字は電話番号(数字)である可能性が高いものとして、認識を行うことができる。その結果、印刷文字認識単独ではできなかった誤認識の訂正が可能となる。

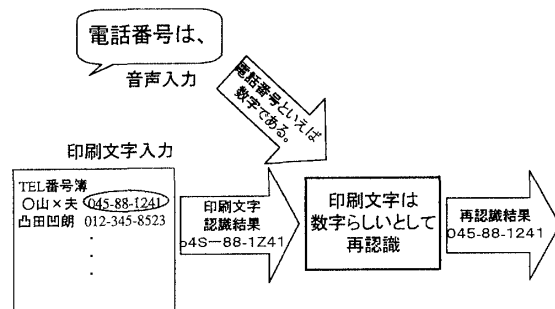


図2 統合認識

4. システム構成

3.で述べた例を実験するシステムを試作した。システム構成を図3に示す。

システムはペン型スキャナ、スキャナ画像用のキャプチャボードを内蔵したパーソナルコンピュータ、印刷文字認識、統合認識を行うWSと音声認識用WSから構成される。

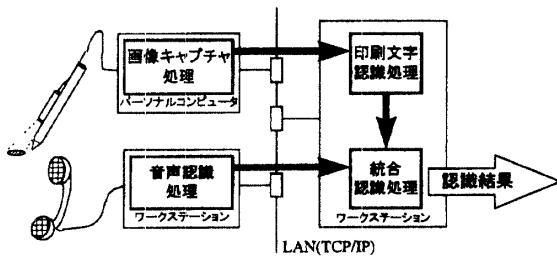


図4 システム構成

4.1 画像キャプチャ処理

ペン型スキャナの画像をユーザの操作によりキャプチャする。キャプチャした画像は入力時刻情報を付加した後、印刷文字認識を行うためWSに送信される。

4.2 印刷文字認識処理

キャプチャ画像に対して印刷文字認識を行う。認識結果と入力時刻情報を統合認識処理に送信する。

4.3 音声認識処理

入力音声の認識を行い、音声の入力時刻情報と認識結果を統合認識処理に送信する。

4.4 統合認識処理

音声入力時刻近くに入力された印刷文字が存在する場合、音声認識結果と整合がとれるように必要なら印刷文字認識結果の修正を行う。図5にその処理手順を示す。

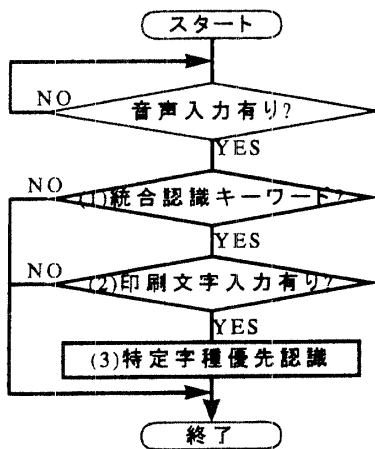


図5 統合認識処理手順

ある特定のキーワードと、それに対応する文字種を知識として設定しておき、音声認識結果として、

そのキーワードが存在し(1)、かつ、その音声入力時刻付近に印刷文字入力があった場合(2)、その印刷文字に対する認識を、キーワードに対応する文字種を優先して行う(3)。

5.動作実験

図6に実際の動作例を示す。実験はキーワードとして「電話番号」、対応する文字種として「数字」という知識を登録して行った。

時刻 T1 に印刷文字画像が入力され、その認識結果は「た3ハ5」となった。その後、時刻T2に音声「内線番号」が入力されると、電話番号は数字であるという知識に基づき数字を優先して再認識が実行され、その認識結果は「3256」に訂正された。

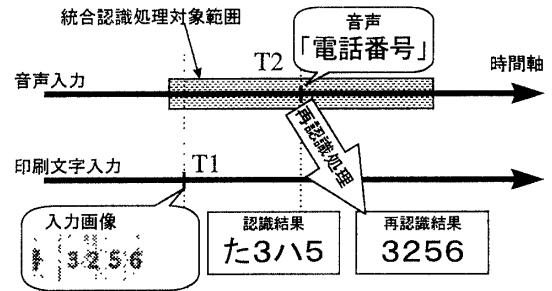


図6 動作例

6.おわりに

マルチモーダルユーザインタフェースの有効な入力装置と考えられるペン型スキャナを用い、印刷文字認識に音声認識結果を作用させることによって誤認識の訂正を行うマルチモーダル統合認識の一手法を提案し、試作システムを開発、誤認識訂正に有効であることを確認した。

他のモダリティを前提とした場合の統合認識の有効性の検証と、モダリティを互いに作用させる双方向統合認識が今後の課題である。

参考文献:

[1]長尾:マルチモーダルインタフェースとエージェント,人工知能学会 Vol.11No.1pp.32-40(1996)