

Web 情報処理を用いた拡張現実感システムの設計

衛門 一樹[†]松浦 寛^{††}西山 裕之[†][†] 東京理科大学理工学部^{††} 東京理科大学理工学研究科

1 はじめに

近年、翻訳システムは Web 情報処理を用いた翻訳が可能である。Web 情報処理とは、ユーザが入力した情報に対して Web 上から情報の収集及び統合を言う。Web 上を膨大なデータベースとして捉えることで様々なサービスを取り入れたシステムの構築が可能となる。また Web 上で翻訳が可能となったことで翻訳を行いたい時、Web に繋がっていることで動的に翻訳が可能である。また文字認識を翻訳に用いることでユーザが解読できない文章に対しても翻訳が可能である。その例として視覚障害者のために画像内の文字列を音声化する研究 [1] がある。しかし文字認識は静止画に対して行うため、動的な翻訳システムにすることが難しい。

そこで動的な翻訳システムを行うために拡張現実感を用いる。拡張現実感とは現実の環境から知覚に与えられる情報に、コンピュータが作り出した情報を重ね合わせ、補足的な情報を与える技術である。近年よく用いられる手法としてカメラの映像上に補足情報を与えている。またその手法はリアルタイム文書画像検索の研究 [2] にも応用されている。よって拡張現実感を利用することでリアルタイムに文字認識を利用した翻訳が可能と考えられる。

そこで本研究では上記の拡張現実感と Web 情報処理を組み合わせ、文字認識で入力を行い、リアルタイムに翻訳が可能なシステムの設計を行う。また本研究では設計を基に試作機の実装を行う。

2 本システムの設計

従来の画像を使った翻訳システムは、カメラで静止画を生成してからその静止画に対して文字認識処理を行い、翻訳処理の結果を表示させることが一般的である。用途としてはユーザが解読できない文章に対して、この翻訳システムを使用する。

しかし既存のシステムでは静止画を収集した後処理を行うため、動的な利用が難しい。したがって本システムにおいては拡張現実感を用いてリアルタイム性を追求し、Web の情報と組み合わせることで屋外でも使用が可能となるシステムの設計を行う。本システムのユーザのシステム利用イメージを図 1 に示す。図 1 のように、ユーザが「解読できない文章の意味を即座に知りたい」という要望を持っていた時、本システムを使用することで理解することが可能である。本システムの特徴は以下の通りである。

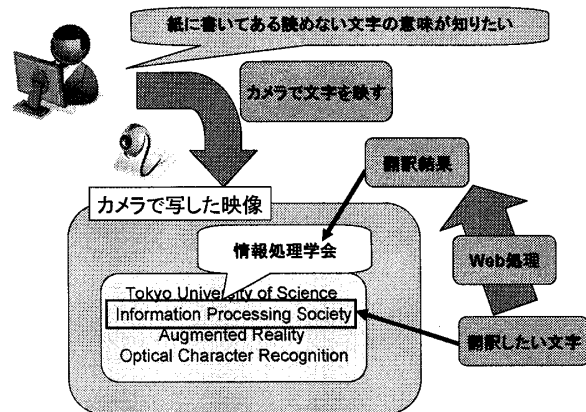


図 1: システム利用イメージ図

- 画像からの文字抽出と Web 情報処理の統合
- 抽出した文字の位置に翻訳結果の補足表示

図 1 においてユーザは文字範囲の指定を行い、指定された文章を英語から日本語に翻訳した結果を吹き出しとして表示を行う。

3 実装システム

図 2 に本システムの流れを示す。本システムは画像

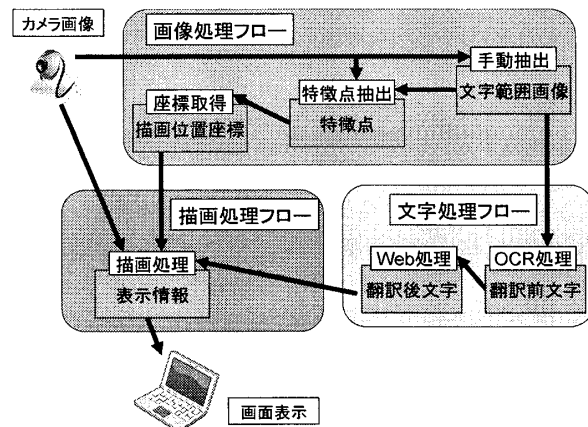


図 2: 本システムにおける情報の流れ

処理と文字処理、描画処理に分けて様々な機能を実現している。画像処理においては文字抽出する範囲の指定、指定した範囲の座標、文字範囲、カメラ画像全体のそれぞれにおいて特徴点抽出を行う。文字処理では

Web design information processing systems using augmented reality.

Kazuki Emon[†], Hiroshi Matsuura^{††}, Hiroyuki Nishiyama[†]

[†]Faculty of Sci. and Tech, Tokyo University of Science

^{††}Graduate School of Sci. and Tech, Tokyo University of Science

文字範囲に対して文字認識, Web 上で処理するための送受信を行う。描画処理では指定範囲の座標に対する表示を行う。また実装システムは Web カメラを用いて PC 上で試作機を開発する。

3.1 画像処理

本システムでの画像処理は OpenCV を利用して文字範囲の手動抽出, 文字範囲とカメラ画像の特徴点抽出, 文字範囲の座標取得を行う。文字範囲の手動抽出では PC を使用するため, マウスまたはそれに順次たもので抽出を行う。抽出する方法はマウスによるポインタのドラッグアンドドロップで線を描き, その線を対角線として四角形の四頂点を割り出し, 四角形に囲まれた部分をユーザが指定した文字範囲としている。そのイメージを図 3 に示す。



図 3: 文字手動抽出のイメージ

次に特徴点抽出を行う。カメラ画像全体に対して毎フレーム行い, 文字範囲に対しては文字範囲を手動抽出した際に行う。次にカメラ画像全体と文字範囲の特徴点を利用して描画位置の座標を取得する。

3.2 文字処理

本システムでの文字処理は OCR ソフトを利用した文字認識処理と WebAPI を利用した翻訳処理を行う。文字認識処理は指定した文字範囲をカメラ画像から切り取り, OCR ソフトを用いて PC 上で処理が可能となる文字データに変換する。次にその文字データをクエリとして WebAPI を用いて翻訳処理を行う。OCR ソフトと WebAPI の機能を変えることで様々な言語に対応することが可能である。今回実装したシステムでは英数字の読み取りが可能な Tesseract OCR[3] を使用し, WebAPI は Google AJAX Language API[4] を使用して英語から日本語へ翻訳処理を行う。

3.3 描画処理

本システムでの描画処理は画像処理で得た文字範囲の座標と文字処理で得た翻訳結果を合成し, 吹き出しに翻訳結果を表示する処理を行う。描画には OpenGL を使用する。またリアルタイムに動作しているため, 描画位置は毎フレーム変更される。最終的な実行画面を図 4 で示す。

4 評価

本システムではリアルタイムに動作を行っているため, フレーム毎の処理時間が遅ければリアルタイム性が損なわれる。そのため動作時の処理時間について評価を行った。図 5 はフレームごとの処理時間の推移である。図 5 から 3 回程度, 1 秒以上処理時間が長いフレームが存在した。これは翻訳時に Web から翻訳結果を受信する際の待ち時間が原因で処理時間が長くなっている。また翻訳結果受信後の処理時間は各フレームごとに 0.4 秒程度であるため, 処理時間は十分な速度でリアルタイム性も損なわれてはいないと言える。



図 4: 実行画面

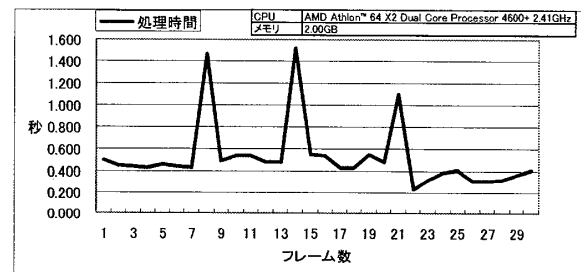


図 5: フレームごとの処理時間の推移

5 おわりに

本研究では拡張現実感と Web 情報処理を組み合わせ, 文字認識で入力を行い, リアルタイムに翻訳が可能なシステム的设计と試作機の実装を行った。本システムを利用することでユーザが読解できない文章の翻訳を行い補足情報として提供することで, 単語及び文章の意味を即座に理解することが可能となった。そして評価からリアルタイム性が損なわれていないことを示した。

今後は英数字の OCR だけでなくほかの言語にも対応し, かつ携帯端末上で動作するシステムの開発を行う。また文字認識だけでなくランドマーク認識といったシステムとの組み合わせを行う。

参考文献

- [1] 清田 公保, 島川 学, 本田 幸代, 江崎 修央, 砂崎 由樹, “視覚障害者のためのカメラ映像内の文字情報読み上げシステム”, 情報処理学会研究報告, UBI, 2006(14), pp.139-143, 2006
- [2] 中居 友弘, 黄瀬 浩一, 岩村 雅一, “Web カメラを用いたリアルタイム文書画像検索とその拡張現実への応用”, 情報処理学会研究報告, CVIM, 2006(93) pp.41-48, 2006
- [3] “Tesseract OCR”, <http://code.google.com/p/tesseract-ocr/>
- [4] “Google AJAX Language API”, <http://code.google.com/intl/ja/apis/ajaxlanguage/>