

カメラ画像からテキスト情報を取得できる文字検出・認識技術

カメラで撮影した文字列を認識しユーザーに適切な情報を提示

現代社会では、看板や、標識、案内板、メニューなど、街のいたるところに文字情報が存在しており、人々はそれらの文字情報を頼りに日々の生活を送っています。従来の文字認識技術が対象とするスキャナで取得した文書画像に比べて、街中を撮影したカメラ画像は文字の背景や照明の条件など様々な変動要因が加わるため、文字列を確実に検出し認識することは容易ではありません。

そこで東芝は、カメラ画像からテキスト情報を取得するための画像認識技術として、カメラ画像文字認識技術を開発しました。これまで、文字認識、顔認識、及び人物検出で培った当社固有の画像認識技術を活用することで、様々な変動に対してロバストになり、国際学会が提供する評価セットに対してトップクラスの文字行検出性能を実現しました。

カメラ画像文字認識と技術開発の背景

今日、屋外には看板や、標識、案内板など、屋内にはレストランのメニューや商品のタグなど、いたるところに文字情報があふれています。例えば、旅行先で看板や、案内板、メニューなどに書かれている外国語の文字がわからないため、道に迷ったり料理の注文ができなかったりといった経験をする場合があります。人々は街中にあふれる文字情報を頼りに生活しています。

このような背景の下、東芝は、カメラ画像から文字列を検出し認識するための技術として、カメラ画像文字認識技術の研究開発を進めてきました。この技術は情景文字認識技術とも呼ばれ、

日常生活の中で目にする風景を撮影したカメラ画像から、テキスト情報を取得するものです。例えば、スマートフォンやタブレットの内蔵カメラを用いて、外国語の案内板を母国語で画面に表示したり、買い物中に商品のタグを読み取り、ユーザーが知りたい情報をクラウドシステム上の関連情報から提示したりするなど、様々なシーンで利用できます(図1)。

近い将来、眼鏡のように掛けるだけで、外国にいてもあたかも母国にいるかのように、目の前の外国語を読み取って母国語に翻訳して表示するウェアラブルデバイスが実現できると考えられます。

カメラ画像文字認識技術の構成

カメラ画像文字認識技術は、三つの処理で構成されます(図2)。1番目の文字行検出処理は、撮影されたカメラ画像から文字列を含む領域を検出する処理です。2番目の文字行認識処理は、検出された文字列領域に対して文字認識を行い、文字認識結果をテキストとして出力する処理です。3番目の言語解析処理は、文字認識結果に対して言語解析を行い、意味を理解して適切な情報をユーザーに提示する処理です。

文字行検出技術

当社は、これまでOCR(光学式文字認識)技術の研究開発に取り組み、郵便物の宛名自動読取区分機や、帳票読

案内板の文字列を検出、翻訳して、行きたい場所へ案内



(a) 旅行先の案内板を母国語で読みたい

商品のタグを読み取り、人気ランキングなどユーザーが欲しい情報を提示



(b) 商品の情報を知りたい

図1. カメラ画像文字認識技術の利用シーン—旅行中にスマートフォンなどで撮影した案内板を翻訳したり、店内の商品タグから商品情報を読み取りユーザーの求めている情報を提示したりするなど、様々なシーンで利用できます。



図2. カメラ画像文字認識技術の構成—カメラ画像から文字列を検出する“文字行検出”、検出結果を認識したテキストとして出力する“文字行認識”、及び認識結果に対して言語解析を行う“言語解析”の三つの技術で構成されます。



(a) 背景が写真の文字列



(b) 影が掛かった文字列



(c) ゆがんだ文字列

図3. 文字行検出結果の例—カメラ画像文字認識技術を用いることで、様々な撮影環境の変動要因に影響されることなく、文字列を検出できます。

取装置、ドキュメントリーダなど様々な製品を実用化してきました。これまでの文字認識技術は、主に紙の文書を読み取り得られる文書画像を対象としていました。一方、カメラ画像は、(1)様々な種類のフォントや背景、(2)照明条件に伴う輝度変化や影の発生、(3)撮影方向による変形など、様々な変動要因が加わります。カメラ画像文字認識技術の中でも、特に最初の処理でキーとなる文字行検出処理は、これらの変動に対してロバストである必要があります。

そこで当社は、顔検出や人物検出などの画像認識の研究で培った特徴抽出技術や検出技術を活用して、新しい文字行検出技術を開発しました。文字領域の検出には顔検出や人物検出で用い

られる特徴量を使用し、文字行生成には画像中の直線検出に用いられてきたHough変換を拡張して使用しました。これらの処理により、カメラ画像の様々な変動要因に対してロバストな文字行検出技術を実現し、これまで検出が難しかった背景が写真の文字列や、影が掛かった文字列、斜めから撮影してゆがんだ文字列なども検出できるようになりました(図3)。更に、文字認識分野で世界最高レベルの国際会議であるICDAR(International Conference on Document Analysis and Recognition)が2013 Robust Reading Competitionのために提供した評価セットに対して、学会トップクラスの文字行検出性能(F値76.2%)を達成しました。

今後の展望

カメラ画像文字認識技術は、実世界のいたるところに存在している文字列というアナログ情報を、文字コード列というデジタル情報に変換するための技術です。当社は、この技術をアナログ情報で構成される実世界とデジタル情報で構成される仮想世界を橋渡しする重要技術と位置づけており、今後も様々なシーンへの応用を視野に入れた研究開発を進めていきます。

登内 洋次郎

研究開発統括部
研究開発センター
インタラクティブメディアラボラトリー主任研究員