

高精度な認識を実現する文字認識技術の開発

スマートフォン、クラウドサービスの普及により写真、画像から文字を抽出するサービスが広がりを見せている。しかし、風景写真からの文字抽出は困難で、とりわけ日本語における認識では改善の余地があった。そこで、この風景写真から高精度な文字認識を実現する技術を開発し、本技術を用いたサービスを提供するとともに、プラットフォームとして基盤化しAPI公開を行った。これによりオープンイノベーションによるサービス発展をめざす。

サービス&ソリューション
開発部

やまぞえ たかふみ
山添 隆文
いわさき あつし
岩崎 淳

すみや てつお
住谷 哲夫

1. まえがき

近年、スマートフォン、クラウドサービスの普及により高度な処理を前提としたサービス提供が可能となった。その中で文字を含んだ写真や画像からその文字を認識してさまざまな利便性を提供するサービスが登場している。しかしながら、その画像に文字以外の建物や風景が写り、影やモアレ^{*1}、テカリなどの悪条件が重なると正確に文字を抽出することが難しいという問題があった。

そこで、言語処理を取り入れ悪条件下でも高精度な抽出を可能とする文字認識技術を開発するとともに、本技術と大規模言語データを組み合わせてプラットフォーム^{*2}として基盤化した。

本稿では、開発した文字認識技術を解説するとともに、本技術を活用したAndroidTM^{*3}用アプリ、さらに

は一般開発者やコンテンツプロバイダが基盤をアプリやWebサービス開発に活用できるよう、API (Application Program Interface)^{*4}として公開する取組みにも触れる。

2. 開発した文字認識技術

開発した文字認識技術を図1に示す。文字認識の処理は、画像中から文字らしい領域を見つける文字領域検出処理、見つかった領域にどのような文字が書かれているか調べる文字認識処理、文字認識結果から誤った結果の修正・削除を行う言語処理という流れとなっている。

2.1 文字領域検出処理

本技術で対象とする画像は、一般的なOCRが対象としている文書用スキャナで取り込んだような画像だけでなく、文字の配置や大きさが一

様ではない風景の画像も対象としている。そのため、まず画像のどこにどのように文字が現れているかを、画像に現れる形状パターンとその形状がもつさまざまな特徴から判定する。形状パターンの検出時点では、間違っただけのパターンを正解としてとらえてしまう誤検出が増えるとしても、正解パターンの検出失敗が減ることを優先する。なぜなら後々の処理で誤った結果の除外を行うからである。

風景の画像では、窓枠や木の枝葉といった形状が文字と似通ったパターンも存在するため、文字として判定された形状パターンの配置や傾きなどにより、文字列らしい範囲を絞り込む。

2.2 文字認識処理

次に、絞り込まれた文字らしい並びから歪みを補正して一行の文字列として切り出し、文字列を文字単位

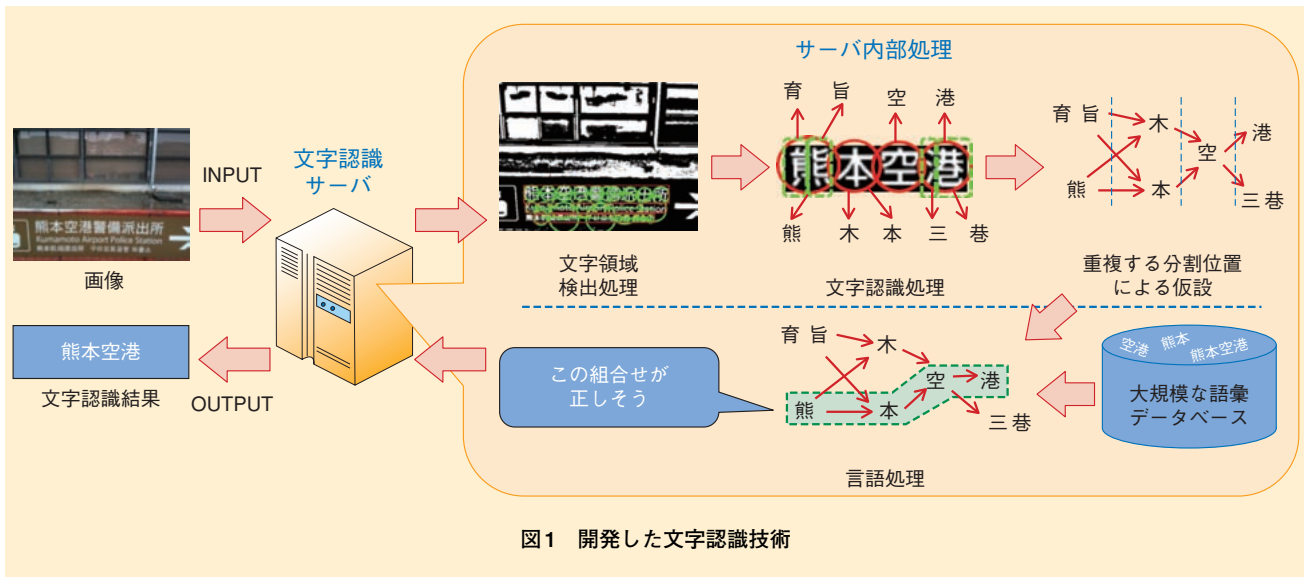


図1 開発した文字認識技術

に分割し、一文字単位でどのような文字が書かれているのか調べる。一文字単位の分割では、例えば「m」という文字なのか、「r」「n」が隙間なく並んでいるのかといった判定は難しいため、文字単位で分割する位置については重複する分割位置による仮説として認識処理を行う。

2.3 言語処理

こうして得られた文字認識の結果は、複数の文字の分割位置や、一文字につき大量の結果の候補から構成される複雑な仮説群となっている。また、文字と似たパターンで構成される背景から誤検出された領域に基づく結果も含まれる。そのような誤りを含む複雑な構造の文字認識の結果に対して言語処理を行うことで、検出語の補正と言葉として意味をなさないものの除外を行う。たとえば、図1の場合、「熊本空港」の「熊」において左右で「育」「旨」に分解

された誤認識結果を含み、「本」においても「本」という認識結果より「木」が正解に近いと判定されたとしても、それぞれのパターンの組合せの中から最も言葉らしい「熊本空港」を言語処理により正解として取り出すことができる。

言語処理としては、独自に収集を進める大規模な語彙データベースを用い、最新の単語を含む100万語を超える語彙に対応する。

3. 文字認識APIの提供

前述の文字認識技術および大規模言語データをプラットフォームとして基盤化し、そのAPIを法人や一般の開発者向けに公開した[1]。本APIは、行画像認識APIと情景画像認識APIの2つのAPIを含む。

・行画像認識API

行画像認識APIは切り出された1行の文字の画像から単語を

抽出し、認識した単語を取得できるAPIである。行画像の文字認識処理は高速に実行でき、3秒程度で認識結果を取得することができる。

・情景画像認識API

情景画像認識APIは文字を含んだ情景画像から文字を抽出し、認識した単語とその位置などを取得できるAPIである。情景画像の認識処理には複雑な計算を必要とするため、認識結果を得るまでにおよそ2分程度の時間がかかる。

文字認識APIのシステム構成概要を図2に示す。各サーバはロードバランサ^{*5}によって負荷分散されている。行画像文字認識については処理が軽いため大量に認識依頼がきても高速に結果を返すことが可能であるが、情景文字認識は処理に長い時間がかかる。そのため、情景文字認識

*3 **Android™**：スマートフォンやタブレット向けのオペレーティングシステム、ミドルウェア、主要なアプリケーションからなるソフトウェアプラットフォーム。米国Google, Inc.の商標または登録商標。

*4 **API**：アプリケーションプログラムに対し

て機能を提供するインターフェース。開発者は制御ロジックとAPIコールを組み合わせ、プログラムを開発する。

*5 **ロードバランサ**：外部からの要求を一元的に管理して同等の機能を有するサーバに要求を転送する装置。負荷分散のために用いる。

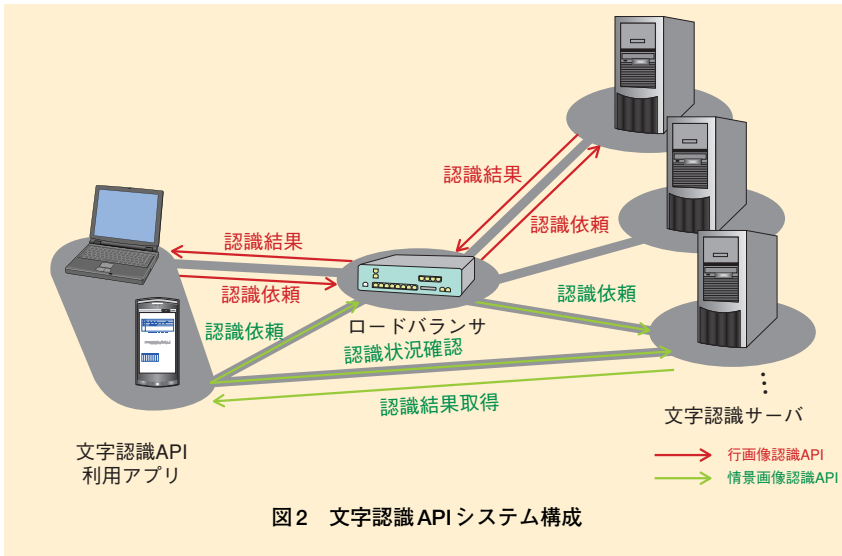
APIでは、処理が終わるまでセッション^{*6}を張り続けるのではなく、認識依頼後に一度セッションを切り、その後文字認識API利用アプリが処理中のサーバに認識状況を直接確認していくことで非同期な構成とし、高いスケーラビリティを実現している。本APIを利用することで、開発者

はさまざまなマッシュアップ^{*7}サービス（ナビゲーション、辞書、翻訳、旅行支援、物品管理など）を構築することができる。本APIの提供により多くのサービスが生み出され文字認識サービスの広がりをみせること、多くの画像データを収集し文字認識の精度を向上させることをね

らいとしている。

4. APIを用いたサービス事例

APIを利用したサービス事例として、「撮って文字入力」というAndroid向けの文字入力アプリ（IME：Input Method Editor）を開発した。従来のIMEでは、テンキーやQWERTYキーをキーパネルとして用いるが、本アプリでは、カメラ映像や画像ギャラリーをキーパネルとして用い、一般的に予測変換といった文字入力候補を表示する欄に文字認識結果を表示し、単語単位で文字認識結果を直接文字入力に利用できるというものである。画像からの文字入力は、利用者が画像から一行の範囲を選択することでを行い、行画像認識APIにより実現する（図3）。また、画像ギャラリーにあらかじめ画像を登録しておくことで、画像全体



Google™：Googleおよびそのアイコンは米国Google Inc.の商標または登録商標。

*6 セッション：サーバとクライアントとの間の通信の意味のあるまとまり。

*7 マッシュアップ：複数の異なるサービスやコンテンツを組み合わせ、1つのサービスを作成、提供すること。

からの単語検出処理が行われ、ギャラリーとして機能する「撮って文字入力」のキーパネルから文字入力候補補として利用することもできる（図4）。この機能は情景画像認識APIにより実現している。

5. 端末アプリへの応用事例

開発の応用例として、文字認識APIの言語処理を利用した料理メニュー翻訳アプリを開発した。本アプリの概要を図5に示す。言語処理は料理名の検出や補正に利用し、機能を端末上に実装することで、通信を必要としないリアルタイムでの動作を実現している。本アプリは外国語の料理メニューを文字認識により読み取り瞬時に日本語に翻訳する。料理メニュー翻訳アプリを利用すると、海外旅行中に読めない料理メニュー（韓国語、中国語、英語）がある場合でも、カメラ越しに外国語メニューの料理名を表示させるだけで日本語訳を同画面に素早く表示させることができる。翻訳言語を選択し、画面中央の枠内に外国語メニューを入れるとすぐに翻訳結果が表示される。また、翻訳前や翻訳後の言葉も、共有ボタンを押すとそのままインターネットで検索することもできる。今後は外国の看板やショッピング中の商品情報など、携帯電話で文字の入力が難しい外国語をカメラ越しで表示させるだけで、その意味を日本語で教えてくれる携帯サービスの実現をめざす。



図4 撮って文字入力（ギャラリー利用）



図5 料理メニュー翻訳アプリ概要

6. あとがき

今回開発した技術、基盤を用いることで画像や写真から簡単に文字を抽出することが可能となる。この技術・基盤を携帯電話で撮影した写真に適用することでモバイルにおけるサービスのさらなる進化を促進する。今後は大規模な画像データの収集

による性能向上や手書き文字、他言語への対応などの機能追加、さらには今回開発した基盤のさまざまなサービスへの応用を検討していく。

文献

- [1] NTTドコモ：“文字認識API.”
<http://recognize.jp/>