

本の携帯電話事業者は近い将来、Android 端末を発売する計画があることを明らかにしています。

## ■ 端末が先か、キラー・アプリが先か

HTC の T-Mobile G1 は非常に開発しやすい端末です。アプリケーションは Java で開発できますし、GPS と 6 軸の電子コンパスを内蔵しています。Android は電子コンパスのインタフェースを基本仕様に入れている点も魅力です。そして、無線 LAN や  HSDPA 経由の高速通信でサーバーとインタラクティブにやり取りしながらさまざまな機能を実現できます。

Android のようなオープンな OS を使わないと、端末のデバイスをフル活用できません。従来のように携帯電話事業者の縛りが多い仕様では、世界展開が難しいという問題もあります。オープン化は今後の携帯電話 OS の大きな流れで、そこに AR アプリケーションも載せていけると思います。

問題は、T-Mobile G1 のように AR アプリケーションを開発しやすい端末の数です。AR 向きの端末が増えないと、特定の用途にしか広がりません。よく携帯電話端末メーカーと話をしますが、「ニワトリが先か卵が先か」という話になります。つまり端末の普及が先か、キラー・アプリケーションの出現が先かということです。この答えとしてメーカーはキラー・アプリケーションが先、アプリケーション開発会社は、端末が普及していないと作らないといっています。つまり、まだ混沌とした状況にあるのも事実です。

 HSDPA = High Speed Downlink Packet Access。第3世代移動通信システム [W-CDMA] の高速版。伝送速度は下りが最大 14.4M ビット/秒。

# シーン認識エンジン「SREngine」を使ったモバイルARアプリ

author : 金村 星日

株式会社光星 チーフマネージャー



私が個人で開発しているAR技術を使ったプロジェクト「SREngine」について紹介します。なお、このプロジェクトはまだ開発の初期段階のため、今後大きく変更する可能性があります。

SREngineの「SR」はscene recognitionの略で、SREngineを直訳すると、シーン認識エンジンとなります。つまりSREngineは、風景や景色といったシーンを特定したり、認識したり、学習したりすることができます。シーン認識では、画像処理技術を利用してシーンの④特徴成分を抽出しています。この特徴成分を④パターン・マッチングすることで、シーンを学習したり、認識したりするのです。次ページの写真4.1は街の風景に、SREngineが黒い点で特徴成分を表示したものです。

このようにSREngineは特徴成分でシーンを認識するため、外部のセンサーや④GPS、④Wi-Fiなどを使った位置測定を一切使わなくても、シーンを特定できるという利点があります。

④特徴成分=SREngineの2009年3月時点の開発版では、ComerまたはFeature Detectionなどのアルゴリズムを利用し、特徴成分を抽出している。

④パターン・マッチング=特徴的な部分を抽出して、照合(マッチング)する手法のこと。

④GPS=Global Positioning System。全地球測位システム。

④Wi-Fi=無線LAN規格であるIEEE802.11シリーズの相互接続を確認して認定するWi-Fiアライアンスが、認定した製品に与えるブランド名称。一般に無線LANのことを指すことも多い。

## ■写真 4.1 SREngine の特徴点

風景の特徴点を認識する。



☞ アノテーション=あるデータに対して付与するタグやメタデータのこと。

☞ Web API = Web サイトやアプリケーションの開発を効率的に進めるための技術。アプリケーションの開発者は、決められたAPIを呼び出すようにプログラミングすれば、サービスを利用できる。

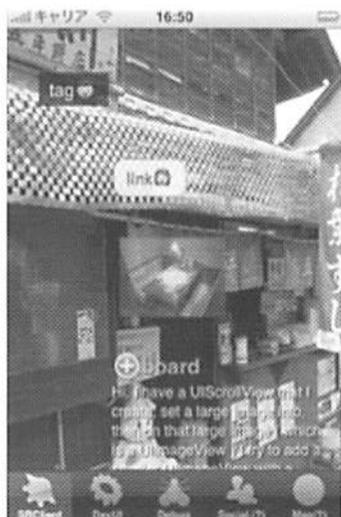
SREngine の利用例はいくつかあります。詳しくは後述しますが、例えば取り込んだ映像の周辺飲食店の詳細情報をほぼリアルタイムで表示するといった機能があります。USB カメラを通して取り込んだ画像を認識し、リアルタイムにシーンの特徴成分を抽出してパターン・マッチングします。飲食店と認識すると、あらかじめ取り込んでおいたさまざまな電子情報の☞アノテーションを風景に重ね合わせます(写真 4.2)。

飲食店の情報は、「Hot pepper」や「食べログ」といった情報サイトの情報を☞Web API を使って取得することもできます(写真 4.3)。これらの情報は無料でダウンロードして利用できます。

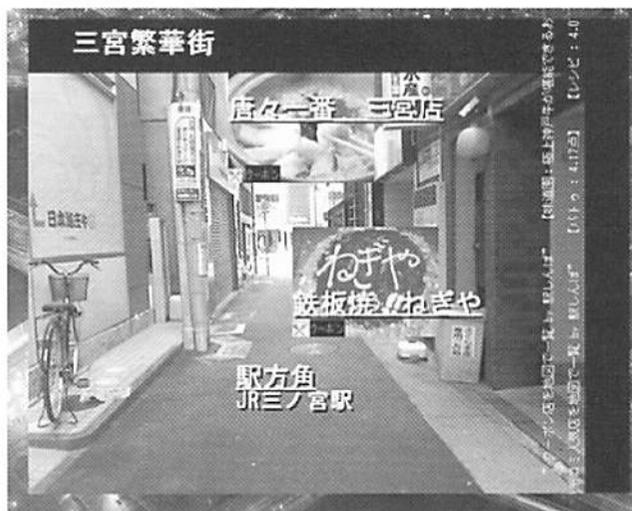
## ■ iPhone に SREngine を実装

SREngine のシステムについては後述しますが、まずは iPhone 上に実装したときの利用例を紹介し

## 写真 4.2 飲食店の前で SREngine を利用した様子



## 写真 4.3 Web API で取得した情報を表示



#### ■写真 4.4 カフェの前で SREngine を利用した様子



ましよう。

カフェの前で iPhone をかざすと、電子情報のアノテーションが表示されます(写真4.4)。友人が「先に店内に入っていますよ」といったメッセージを残して、そのメッセージを見ることができます。iPhone を振るとアノテーションがリセットされます。

外に出て、ビルに対して iPhone をかざすと、どのフロアにどんなテナントが入っているのかという情報を見ることができます(写真4.5)。カレンダーをクリックすると、カフェの営業時間を見られます。また、本日のコーヒーは何かといった情報も見られます。

ほかにも、駅構内などのポスターの詳細情報を見るといった使い方もできます。例えば、マンションのポスターをカメラライブビューでとらえると、完成したエントランスの画像や残り戸数を表示すると

## 写真 4.5 ビルの前で SREngine を利用した様子

テナント情報を表示。



いった用途に利用できます(次ページの写真 4.6)。ポスターを貼った時点でエントランスが完成してなくても、販売個数が変わっても、SREngine で最新の情報をほぼリアルタイムで表示できます。広告の情報をアップデートしたり、新しい付加情報を付けたりするという使い方が、将来的にはできるでしょう。

### ■ 情報へのアクセス方法が変わる

SREngine のこうした使い方は、検索という行動に大きなインパクトを与えます。

パソコンやインターネットの普及によって、サイバーワールドが登場しました。私たちは現実世界で生きていますが、サイバーワールドの中でメールやファイル転送、Web ブラウザを使ったさまざまな

## ■写真 4.6 ポスターの前で SREngine を利用した様子

ポスターに記載していない最新情報も表示。



インターネット・アプリケーションを便利に使っています。

こういったサービスを利用するときに、ユーザーはキーワードや URL を使って情報にアクセスしています。これまでにパソコンやモバイル端末など、サービスを利用する端末は増えたり変わったりしましたが、情報にアクセスする手法は大きく変化していません。

ところが、SREngine は紹介した通り、ユーザーがキーワードや URL といったものを意識せずに、情報にアクセスできます。仮想的に入力なしでサイバーワールドの情報にアクセスする手法を、SREngine で提供できるのではないかと考えています。

## ■ 広告の形を変える

SREngine は広告の形も変える可能性がありま  
す。ユーザーが空間に SREngine の端末をかざし  
て、飲食店やポスターなどを認識すると、広告情報  
を表示するといった使い方ができます。

ユーザーは目的に合った広告を好意的にとらえる  
傾向にあるということが、いくつかの調査でわかっ  
ています。例えば検索連動型広告では、約 6 割が広  
告は自分に役立つと考えているそうです。逆にまっ  
たく邪魔だというユーザーは、1 割にも満たないそ  
うです。

SREngine の広告は、ユーザーの関心とかなり一  
致するのではないかと考えることができます。なぜ  
なら、ユーザーは注意を向けた局所的な空間に  
SREngine の端末をかざしています。表示される広  
告はその空間に関連しているため、ユーザーの関心  
と広告とが合致している傾向が高いでしょう。前述  
したマンションのポスター広告のように、既存の広  
告の情報をアップデートするなど新しい付加価値を  
付けることもできます。

## ■ スタンドアローン型は ノート PC ベース

SREngine のシステムを紹介します。ローカルで  
すべて完結するタイプの「Stand-alone Edition」と、  
ネットワークを介して利用する「Client/Server  
Edition」の 2 種類があります。

Stand-alone Edition は、Intel Core 2 Duo の CPU  
が搭載されているノートパソコンにウェブカメラを  
つなぐことで動作します。Stand-alone Edition は、

④マッシュアップ= Web上の複数のサービスや技術を組み合わせて新しいサービスなどを作ること。

④画角=カメラで撮影するときの被写体の写る範囲を、カメラレンズを中心にした角度で示したもの。

私が開発した「Fallen」という AR アプリケーションに実験的に実装しました。

対応する OS は、Windows XP または Windows Vista。CPU は 2.4GHz Intel Core 2 Duo を使っています。ウェブカメラには、ロジケールの QCam Pro for Notebooks を利用しています。オプションとして、ネットワーク接続環境があれば④マッシュアップを利用したり、URL をクリックしたりすることができます。

ローカルタイプの Stand-alone Edition をテストする過程で、さまざまな課題が浮き彫りになりました。まず、ウェブカメラを付けたノートパソコンを広げて街を歩いていると、「変な人がいる」という目で見られます。また、ノートパソコンは小さく軽くなったとはいえ、まだ重いという課題があります。ノートパソコンをベースにした SREngine を街で利用するには抵抗があります。

次に、USB カメラへの対応をどうするかという課題があります。USB カメラは数多くの製品が市場にあり、レンズのスペック、④画角、オートフォーカス機能があるかどうかなどさまざまです。SREngine で画像処理をする際に、こうしたスペックの違いが問題になります。仮に一つの USB カメラに限定しても、設定などを変えられるので画像処理をする際に問題となります。

さらに、街のシーンには膨大な情報量があるため、ローカルで全データを保存するのは実用的ではありません。ネットワークへの常時接続が必要になります。SREngine は、GPS や Wi-Fi の位置測定技術を利用しなくてもシーンの特徴点を用いて動作しますが、位置情報を利用すれば検索対象を絞り込めます。

シーン認識の精度を向上できます。

またローカルでは利用できませんが、ユーザー間でシーンの情報を共有できれば、新しいソーシャルメディアに発展する可能性があります。

## ■ iPhone とサーバーで ネット対応版を開発

これらの課題をクリアできれば、SREngine は周辺情報を検索する新しい手法になると考えました。そこで、クライアント・サーバー型の開発とともに、さまざまなモバイル端末を比較検討しました。開発環境がオープンかつ無償で手に入ることと、カメラを搭載していることから、アップルの iPhone が候補に挙がりました。

ただし、iPhone は CPU が非力だとよくいわれます。画像処理の性能は高いけれど、CPU の演算スピードはそれほど速くないため、iPhone 上で SREngine を動作させるのは難しいのではないかと考えました。しかし、さまざまな工夫を施して、SREngine の iPhone 版「SREngine for iPhone」を作成しました。

iPhone で利用している機能としては、初めから組み込まれているカメラ、携帯電話の 3G (第 3 世代) ネットワークへの接続機能と Wi-Fi 無線 LAN への接続機能を使っています。無線 LAN と 3G とを比較すると、3G の方が体感で分かる程度遅いです。ただし、3G でも十分使えます。

iPhone の OS はバージョン 2.2 を使っています。開発環境は、アップルの iPhone SDK です。特別なプライベート・ライブラリなどは一切必要ありません。

ん。iPhone 以外に新しいハードウェアも、特別なソフトウェアも追加する必要がありません。

## ■ サーバー側はオープンソースで構成

SREngine のサーバーの構成は、Web サーバーとデータベース・サーバーです。一般的に多くの Web サービスが「LAMP」という構成を使っています。LAMP の L は Linux、A は Apache、M は MySQL などのデータベース・サーバー、P は PHP、Python、Perl といったスクリプト言語です。それらを合わせて LAMP と呼び、この LAMP の構成を SREngine のサーバーに採用しています。

Web サーバーとしては Apache、Django というフレームワークを使っています。データベースには MySQL あるいは SQLite3、開発環境としては Python を利用しています。

このようにサーバー側は、すべてオープンソースのソフトウェアで構成しています。開発・運用コストが比較的安く抑えられています。

SREngine の独自アルゴリズムは、シーン認識処理を分離できるように設計されています。このため、モバイル環境にも柔軟に対応できます。

役割を大きく分けると、クライアント側 (iPhone) では特徴点を抽出し、サーバー側ではパターン・マッチングを行っています。この方法の利点は、クライアント側で特徴点情報を最適化してから送信するため、余分なネットワーク・トラフィックが発生しないことです。サーバー側ではデータベース検索を最小に抑えるために、特徴点に対して十分な前処理を施します。

このような分散処理により、シーン認識の精度を

高めることができ、また、ボトルネックとなるディスクアクセスを最小にしています。

## ■ 他のシステムと連携して発展

SREngine のプロジェクトは 2009 年 4 月時点で、1 年にも満たないかなり未熟なものです。数多くの課題が山積していますが、エンジンそのものの機能強化を進めていきたいと考えています。

シーンを認識する精度を高くするほど、認識しにくくなるというジレンマがあります。実用的で使いやすいようになるように、開発を進めていきます。

シーン認識の精度をさらに高めるために、GPS や Wi-Fi などの位置推定技術を導入する予定です。Wi-Fi を使った位置推定技術であるクウジットの PlaceEngine の試用版を実装してテストをしたところ、良好な結果を得ました。

SREngine ではセンサーに依存しないことと、コストをかけないことを特徴としています。このため、ボトムアップなアプローチを考えています。ユーザーにさまざまな用途を考えてもらって、集合知でこれまでにないモバイルアプリケーションへと発展できればよいと思っています。

また、SREngine のサーバー側は主に Web 技術を使っています。このため、Web のマッシュアップ用の Web API を比較的容易に使えます。Web API を通して、グルメや美容、駅情報といった情報を今すぐにも活用できます。

プラットフォームについては、iPhone 以外を利用する可能性もあります。iPhone 以外で理想的なプラットフォームが今後登場すれば、移行するつもりです。

クウジット=2007年7月設立。位置情報を獲得するPlaceEngineなどの事業を展開している。

PlaceEngine =クウジットが提供する無線LANを利用した位置情報サービス。接続している無線LANアクセス・ポイントから現在位置を特定する。

☞ ARToolKit = AR アプリケーションを開発するためのオープンソースのC言語ライブラリ。奈良先端科学技術大学院大学の加藤博一教授が開発した。

一番重要なポイントは、オープンなプラットフォーム環境ということです。現時点では、シーンにテキスト情報と画像をアノテーションとして合成させています。ほかのAR技術と融合すると、もっと面白いことができるのではないかと思います。例えば「☞ ARToolKit」を使った技術を融合させることで、よりリッチなオブジェクトをシーンごとに出すことも考えられます。

SREngineは単体で成立するものではありません。ほかのテクノロジーと組み合わせることで、やっとユーザーがアプリケーションとして使えるものになります。

## ■ 今後必要になる SREngine 以外のエンジン

まだ構想段階ですが、SREngine以外に少なくとも二つのソフトウェア・エンジンを提供する必要があります。現時点のSREngineは、まずシーンを認識し、学習済みのシーンか、あるいは誰かが既に登録したシーンであれば、情報が出てきます。そうでなければ情報が表示されないため、サービスの開始当初はほとんど使い物にならないでしょう。

登録情報がない場合でも、場所に応じた情報を出す必要があるでしょう。ここでは、場所に応じた情報を持つ「ロケーションエンジン」が大変重要になってきます。例えば天気予報などの情報は、GPSの位置情報の精度で十分表示できます。街を歩いていると、「3時間後に降水確率が非常に高まる」という情報を表示するサービスなどはかなり有効でしょう。

こういったアプリケーションで重要なのが、どん

な情報を提示するかということです。モバイル端末は画面サイズに制約があるので、ユーザーが有益な情報だと思えるものを選択的に、しかも自動で提示する必要があります。どういった情報が有用なのかは、ユーザーごとに異なります。ユーザー・プロフィールなどと連動するエージェント的で賢いエンジンが求められます。そういったエンジンを、「インテリジェンスエンジン」と呼ぶことにします。

ロケーションエンジンとインテリジェンスエンジンは、SREngineがある程度実用的なレベルに到達した段階で開発に着手しようと考えています。これらが実現すれば、既存のさまざまなサービスと連携することで、新しいモバイル・プラットフォームが生まれると考えています。

# 「世界をクリックカブルに」 AR は Web に相当する プラットフォームになる

iPhone で動く AR アプリケーションとして話題の「セカイカメラ」クローズドβ版が 2009 年 2 月に公開された。開発を進める頓智・の井口尊仁社長は、「ワールドワイドで 2 億、3 億ユーザー規模のプラットフォームにしたい」、「クリックカブルな世界を作りたい」と意気込む(写真 A.1)。

——なぜサービス名は「セカイカメラ」で、社名は「頓智・」なのですか。

セカイカメラは「世界を丸ごと扱えるカメラ」という意味です(次ページの写真 A.2)。僕は大学の哲学科の学生だった時代に隣の理工系の学生に教えてもらってプログラミングにはまったのですが、そのとき「世界の人間の思念というのは演算可能なのではないか」と直感的に思いました。グーグル創立者のセルゲイ・ブリン氏が言う“世界の情報をすべて整理して届ける”という思いとすごく近いと考えているのですが、コンピューティングによってすべての人間の思考や概念が扱えるよう

■写真 A.1 頓智・の井口尊仁社長



撮影：的野弘路

■写真 A.2 2009年2月にファッション展示会「rooms」会場で公開されたセカイカメラ  
クローズドβ版

クウジツと協力し、位置情報の取得には「PlaceEngine」を利用していた。



になり、しかもそれをアウトプットして共有できるのではないかと直感的に感じました。セカイカメラの原点には、そういった自分のインスピレーションがあります。

一方、哲学科の学生というのは“人間にとっての世界のとらえ方”みたいなことを一生懸命考えて、ハイデッガーはこう言っているとか、いやいやニーチェはこうだ、といったことをいつも思索しています。世界観は非常に多様で、右や左あるいは上流、下流というものではなく、人それぞれが独自の見方を持っている、という考え方がありました。ですので、セカイカメラには“世界は多様である”という意味も込めています。

社名は良い名前が思い浮かばなくて苦労しました。グーグルはものすごく大きな数字の単位をもとにした名前です。つまりそこに“世界中の情報をす

べて束ねる”という哲学が入っていると思います。

セカイカメラの哲学を社名にどう込めるのかというときには、とても困りました。僕らはグーグルのようにロケット・サイエンス級の最難関技術を持つ会社ではないし、博士号を持つエンジニアがごまんといるとか、何億円という予算があるとか、そういうわけでもありません。

となったときに、“頓智”ではないかと思いました。発想の逆転、追い詰められたときにサバイバルのためにひらめく逆転のアイデアこそ、僕らの真骨頂であると。頓智の後ろについているドットは、ライブニッツの単子論(モノドロロジー)の考え方からとりました。人間が持っているオルタナティブな思念、その一つひとつが世界の中心という考えから付けています。

## ——頓智の発想はセカイカメラにどう生かされているのでしょうか。

セカイカメラの技術は“頓智の嵐”です。よくギークな方からは中身がないとか、テクノロジー的に見るべきところがないといったことを言われますが、「いや、いいんです」と。もちろん、基礎研究にお金をかけて画像認識をやったりしてもよいのですが、とにかく大変です。なかなか動かないかもしれないし、ユーザーにとってはかえって不便かもしれない。

セカイカメラは位置と方向ですべてを解決しようという考え方で作られています。だから、対象物が移動するとトラッキングできませんし、厳密な位置を指定した表示もできないという欠点があります。セカイカメラの「エアタグ」は、なんとなく近

 エアタグは実映像にオーバーレイされるアイコン。リンク情報などをもち、クリックすると情報の閲覧などができる。

④QRコード=デンソー（現デンソーウェーブ）が開発した2次元コード。バーコードと異なり2方向に情報を持つので、表示面積が小さくても情報量を多くできる。

④RFID=Radio Frequency Identification。ID番号を持つ小さなタグを、無線で認識できるようにする技術。

④電子コンパス=地磁気から方位を検出するデバイス。地磁気センサーとも呼ばれる。

④加速度センサー=端末の傾きや振動を検出できるセンサー。携帯電話機のカーソルの移動やゲームのコントロールに利用できる。

くに浮いている感じです。アバウトにできているのです。

エアタグを厳密に表示させようとする、動きがぎくしゃくしたり、ハードウェアへの負担が重くなったりします。④QRコードや④RFIDを使えば厳密な表示ができると思いますが、現在のコンシューマ向けの機器でやろうとすると破綻します。ですから僕は、その部分はとりあえず置いておこうと。アバウトに位置と方向だけで作って、トラッキングはユーザーに指でスクロールして手動でやっていただければいいのではないかと。セカイカメラのエアタグがぶかーんと浮いているレベルなら、現在のデバイスの能力とソフトウェアの実装で実現できます。

でも、このエアタグを使って“エアキャラクタ”や“エア看板”という仕組みを作ってみたところ、企業やコンシューマの方が面白いと言ってくれています。日本の通信事業者やメーカーの方々は、サービスをがっちり組み上げるという考え方をしますが、僕たちはユーザーに楽しんでもらえるしかけを優先しました。

もちろん、iPhoneには④電子コンパスがないので、向いている方向は取得できません。でもそこでダメだと言ってしまうと、まったく先に進めないの、逆に今のiPhoneでどこまでできるのかを追求したいのです。ニュースや交通の混雑情報、あるいはテレビ番組表など、方位が重要ではない情報はたくさんあるわけですから。

もっとも、ほぼ同時にAndroid版のセカイカメラも開発していて、iPhone版よりも進んでいます。こちらでは6軸センサー（電子コンパス+④加速