Vol.2012-CVIM-182 No.3
2012/5/23

# 過去の街並みを可視化するスマートフォンを用いた拡張現実

佐藤 慎也[1]　岡谷 貴之[1]　出口 光一郎[1]

**概要**：本研究では，身近なスマートフォンを用いて，過去の街並みを可視化する拡張現実 (AR) を実現する．ある街の通りでスマートフォンをかざすと，その方向の過去の街並みがスマートフォンの画面上に表示され，あたかもその画面が窓となり，その中に時間をさかのぼった景色が見えるようなものである．画面に表示する画像は，あらかじめ撮影した全方位画像から切り出して生成する．このとき，端末の姿勢を内部センサおよび画像を組み合わせて推定し，リアルタイムで切り出し領域を変えて表示することで，上述の拡張現実感を得られるようにする．端末の姿勢がユーザの見たい方向と一致しない場合でも，端末の表示画面側にあるカメラを用いてユーザの顔を検出し，端末とユーザの顔の相対位置を推定し，この不一致を補正する．ユーザが今いる位置は，端末の GPS データと，端末のカメラがその場で撮影した画像を両方利用して推定する．具体的には，GPS によっておおまかな位置を推定し，それによって全方位画像のデータベースから対象とする画像集合を絞り込む．その後，撮影画像との間で画像特徴を用いた類似度を求め，位置を高精度に推定するとともに，表示すべき画像を決定する．

**キーワード**：拡張現実，スマートフォン，位置推定，姿勢推定，GPS

## 1. はじめに

本研究では，身近なスマートフォンを用いて，過去の街並みを可視化する拡張現実 (AR) を実現する．図 1 に示すように，ある街の通りでスマートフォンをかざすと，その方向の過去の街並みがスマートフォンの画面上に表示され，あたかもその画面が窓となり，その中に時間をさかのぼった景色が見えるようなものである．

画面に表示する画像は過去にその場所を撮影した全方位画像から，実際に今ユーザが向いている方角に対応する部分を切り出して使う．今いる場所に対応する全方位画像を表示する為に，ユーザ (スマートフォン) の自己位置を GPS 及びスマートフォンで撮影したその場所の画像を組み合わせて求める．スマートフォンをかざす方角を変えると，姿勢センサ，GPS 及びカメラから取得した画像によりその動きを検知し，切り出し領域をリアルタイムで変化させることで，ユーザは画面を窓として昔の景色を見ていると感じられるようにする．ユーザは同じ方向を向いているのに端末の向きだけが変わった場合も考慮し，端末の画面側のカメラを使ってユーザが見たい方向を検出し，その方向の景色を表示する．

このシステムを考えた背景には，東日本大震災の津波被害がある．被災した市街地で，時間軸を超えた拡張現実感を実現することで，例えば防災や教育の為，あるいは都市計画の種々の用途に役立てることを目指している．

(a)　(b)

**図 1**　(a) 本研究で提案する拡張現実をユーザが体験する様子 (b) スマートフォンの画面があたかも窓となり，その向こうに昔の景色が見えているように，過去の画像を表示

## 2. 関連研究

現実環境に仮想環境をリアルタイムに重ねて表示する AR 技術は，様々な分野・用途向けに研究開発されている．代表的なものに，トラッキングを行うことで実空間の三次元認識を行い，それを基に精度の高い位置合わせが可能な情報表示を実現する PTAM(Parallel Tracking And Mapping)[1] がある．

本研究に関連の深いものを挙げると，実空間へ仮想情報を表示する研究としては，全方位画像をリファレンスデータとして用い，SURF を用いた画像マッチングを行って，現在カメラに写っている映像が全方位画像のどの方向に対

[1] 東北大学
Tohoku Uniersity


**図 2** シーンの全方位画像から部分領域を切り出してスマートフォンの画面に表示する．

応するかを推定する方法がある [2]．また，失われた文化遺産を CG で合成し，現実世界の映像と重ねて再現研究も様々な研究機関で行われている．代表的なものに，古代の飛鳥京の復元 CG 映像を現在の明日香村の景観に合成する研究がある [3][4]．過去の写真や映像を現在のその場所の映像にシームレスに重ねて表示するために，現在と過去の特徴点の対応を用いて過去に撮影された地点と同じ地点にユーザを誘導する方法がある [5]．森林調査支援を目的に，全方位画像を利用して樹木の CG モデルを現在見えている森林の画像に重ねて表示するという研究もある [6]．

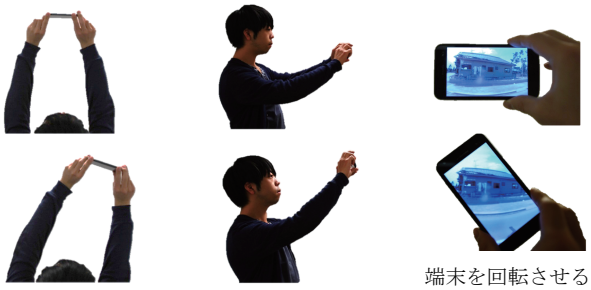
一般的な AR では，現実環境と仮想環境との映像上の幾何学的な位置合わせが必要である．これは，カメラで撮影した実写の映像に，CG などで合成した映像を重ねる（映像上に映像を重ねる）からである，本研究のシステムは，端末の表示画面上の画像と周囲の景色がある程度一致すればよく，通常求められるような高い位置合わせ精度は必要ない．自己位置推定に GPS を一部利用するが，AR の用途で GPS を利用する方法の研究には [7] などがある．

## 3. 端末の姿勢と連動させた画像の表示

### 3.1 概要

ユーザがシーンの見たい方向にスマートフォンをかざすと，その方向の（そのシーンの過去の時刻の）画像が表示されるようにしたい．対象とする（過去の時点の）シーンの，ユーザのいる場所とほぼ同じ場所から撮影した全方位画像が与えられているものとする．この全方位画像から切り出した部分領域を，スマートフォンの画面上に表示する（図 2）．スマートフォンの姿勢を求め，それに合わせて全方位画像からの切り出し領域を変化させ，これを実時間で行うと，「仮想的な窓を通して過去のシーンを見る」という拡張現実を実現できる．

シーンの物体表面がユーザおよびスマートフォンから十分遠いところにあると仮定し，スマートフォンの 3 次元空間における姿勢（回転の 3 自由度）のみから，全方位画像における表示画像の切り出し領域を決める．このとき，



水平方向の端末の動かし方　鉛直方向の端末の動かし方　端末を回転させる際の動かし方

**図 3** ユーザに求めるスマートフォンの動かし方．

ユーザが自由にスマートフォンを把持してよいことにすると，見たい方向とスマートフォンの姿勢が一致するとは限らないので，先述のような拡張現実感が実現できない．そこで，ユーザは図 3 のようにスマートフォンを動かすものと仮定する（あるいはそのようにユーザに要請する）こととする．

ただし，ユーザにこの動作の完全性を期待するのは難しいことがわれわれの予備的実験でわかった．そこで，ある程度の傾きのずれ（すなわち，ユーザの頭部とスマートフォンの画面を結ぶ直線方向とスマートフォンの画面の姿勢）が発生することは前提とし，これを端末のインカメラでユーザの顔の位置を検出することで補正する．具体的には，ユーザの顔位置から，ユーザの前額面に対する画面の傾きを（荒い精度で）算出し，これを元に，ユーザが本来向いている方向に対応する部分領域を全方位画像から切り出す．

### 3.2 端末の姿勢推定

スマートフォンの姿勢を求めるには，内蔵された加速度センサやジャイロセンサの出力を用いるセンサベースの方法と，スマートフォン内蔵のカメラを使うビジョンベースの方法の 2 通りが考えられる．前者は推定精度はあまり高くはないが高速に処理でき，後者は高精度だが計算量が大きい（特に時間遅れが問題となる）．実写映像に合成映像などを重ねる AR では，映像どうしの幾何学的な整合性が特に重要であり，後者のビジョンベースの方法が有利である．しかし本研究は，映像どうしを重ねるわけではないので，それほど高い幾何学的整合性を実現する必要はない．ユーザが向いている方向が推定できればよいので，センサベースの方法を中核的な方法として採用する．具体的には，各センサの特徴を考慮して，速い動きはジャイロセンサで検出し，遅い動き（ドリフトの補正）に加速度センサおよび地磁気センサを利用する．

スマートフォンの方位角，ロール角，ピッチ角をそれぞれ $\theta_x$，$\theta_y$，$\theta_z$ とし，これに対応するジャイロセンサの出力であるスマートフォン筐体各軸回りの角速度を $\omega_x$，$\omega_y$，

$\omega_z$ とすると，（積分定数を除いて）

$$\theta_i = \int \omega_i \cdot dt \qquad (i = x, y, z) \qquad (1)$$

と表せる．

この姿勢に対する全方位画像上の切り出し領域は，図 2 に示す画像座標上を移動する．$u$ 方向，$v$ 方向の移動量 $\Delta u$, $\Delta v$，端末の画面鉛直方向まわりの回転角 $\Delta\phi$ は，全方位画像のサイズを $W_{pano} \times H_{pano}$ とすると，全方位画像（円柱パノラマ画像)）の横方向が $2\pi$，縦方向が $\pi$ の角度範囲に相当するので，

$$\Delta u = \theta_x \cdot \frac{W_{pano}}{2\pi} \qquad (2)$$

$$\Delta v = \theta_y \cdot \frac{H_{pano}}{\pi} \qquad (3)$$

$$\Delta\phi = \theta_z \qquad (4)$$

と書ける．

このとき，スマートフォンのインカメラで検出したユーザの顔位置を元に，全方位画像からの切り出し領域を補正する．検出した顔の位置が端末とユーザの前額面が平行な時と比べてどれだけずれているかを求める．このずれの値を $d$ とし，端末の画面幅を $W_{disp}$，インカメラの画角を $\alpha$ とすると，ユーザの前額面に対する端末の傾き $\psi$ は，

$$\psi = \frac{\alpha \cdot d}{W_{disp}} \qquad (5)$$

となる．この $\psi$ の値を用いてジャイロセンサで求めた端末座標 $x$ 軸周りの回転角を補正し，ユーザが本来向いている方角に対応する領域を全方位画像から切り出す．

上述のように姿勢はジャイロセンサの出力を積分して得るため，ドリフトの影響で時間が経つにつれて誤差が蓄積する．そこでジャイロセンサで求めた姿勢を加速度センサと地磁気センサを用いて補正する．加速度センサと地磁気センサを組み合わせると，ユーザのいる場所を原点とした天頂方向を $z$ 軸，北極点の方向を $y$ 軸とする座標系における方位角，ピッチ角，ロール角が計算される．方位角は端末座標 $z$ 軸負方向が向いている方角，つまりユーザの向いている方角である．これにより端末の姿勢が求まり，切り出し領域の絶対位置を求めることが出来るので，ジャイロセンサで求めた姿勢を補正して，切り出し領域の位置をユーザが向いている方向に補正できる．

具体的な補正の方法は次の通りである．加速度センサ，地磁気センサから，方位角，ピッチ角，ロール角が得られる．それぞれ $\theta'_x(0 \leq \theta'_x < 2\pi)$，$\theta'_z(-\pi < \theta'_y \leq \pi)$，$\theta'_y(-\pi < \theta'_y \leq \pi)$ とすると，これらから求めた切り出し領域の $u$ 方向，$v$ 方向の移動量 $\Delta u'$，$\Delta v'$，端末の画面を中心とした回転角 $\Delta\phi'$ は，

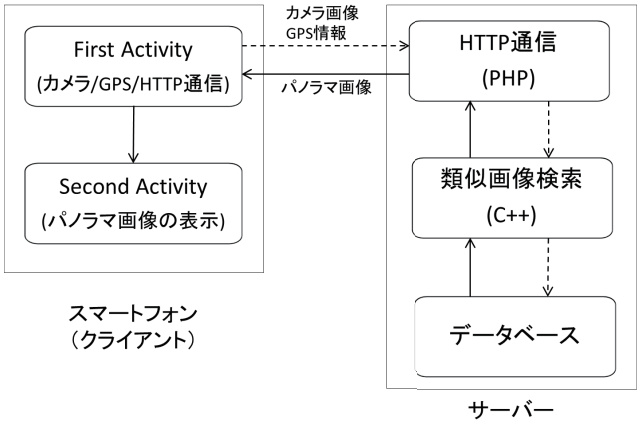



**図 4** システム全体の構成

$$\Delta u' = \theta'_x \cdot \frac{W_{pano}}{2\pi} \qquad (6)$$

$$\Delta v' = \theta'_y \cdot \frac{H_{pano}}{\pi} \qquad (7)$$

$$\Delta\phi' = \theta'_z \qquad (8)$$

と書ける．この値と，ジャイロセンサで求めた切り出し領域の位置がこの値とずれている場合に補正を行う．補正方法としてはジャイロセンサで求めた移動量 $\theta_i$ に補正量 $\delta i$ を加える事により，誤差の蓄積している $\theta_i$ の値を徐々に $\theta'_i$ に近づけている $(i = x, y, z)$．姿勢角の補正は，ユーザに違和感を与えない滑らかなものとするために，補正量を $\theta_i$ と $\theta'_i$ の差に比例した値とし，比例定数を $k_i$ として

$$\delta_i = k_i|\theta_i - \theta'_i| \qquad (i = x, y, z) \qquad (9)$$

とする．

## 4. 自己位置推定

端末の自己位置は，GPS の緯度経度情報を元にした k-近傍探索と，スマートフォンで撮影した画像の画像特徴量に基づく最近傍探索を組み合わせた方法で推定する．まず初めにスマートフォンのセンサから取得した現在地の緯度経度に最も近い緯度経度情報を持つ画像を $k$ 枚に絞り込む．ここでは $k = 20$ を選んだが，これは GPS センサの精度によって変更するのが適当である．次に，撮影した画像と絞り込んだ $k$ 枚の画像それぞれから特徴点を抽出し，特徴点集合の画像類似度から対応を求め，撮影画像に最も近い画像を選びだす．特徴点の検出と特徴量の記述には SURF[8] を用いた．最近傍探索には FLANN(Fast Library for Approximate Nearest Neighbors)[9] を用いた．

## 5. スマートフォンへの実装

スマートフォンには Android 端末を用いて，アプリケーションの開発を行った．使用端末は SAMSUNG 社製の GALAXY SII LTE で，OS は Android 2.3.6 である．

アプリケーションの構成は図 4 のようになっている．ま

**図 5** 被災地でのデモの様子 (左上) 端末に表示された過去の景色 (左下) 実際に今見えている景色 (中央上) 動かす前 (中央下) 鉛直上向きに移動 (右上) 水平方向へ移動 (右下) 回転の動き

ず，Android 上でアプリケーションを起動させるとカメラが起動し，今いる位置の (端末が今向いている方向の) 周辺の景色を撮影する．GPS データをセンサにより取得した後，SD カードに保存されたカメラ画像と GPS 情報が HTTP の POST でサーバに送られる．サーバはカメラ画像と GPS のデータを PHP で受け取り，取得したカメラ画像をファイルに保存し，GPS 情報を C++で実装された類似画像検索プログラムに渡し，それを入力として検索を行う．類似画像検索プログラムでデータベースからユーザが今いる位置の全方位画像を決めると，この画像はサーバ上のファイルに出力され，PHP で Base64 で文字列にエンコードされ，JSON 形式で Android に送信する．Android では HTTP 通信のレスポンスとしてこれを受け取り，Base64 でデコードした画像を SD カードに保存して，これを読み込んで画面に表示する．3 節の方法でユーザが見ている方角に合わせて全方位画像の一部が切り取られて表示される．

実際に東日本大震災の被災地でこのアプリケーションを実行した様子を図 5 に示す．このように端末画面には過去に撮影した全方位画像の一部が窓から景色を見るように映っているのがわかる．現在は無くなってしまった建造物も端末画面を通して窓越しにそこに在るように見えている．

## 6. まとめ

過去の街並みを可視化する拡張現実技術として，過去に撮影した全方位画像を切り出して画面に表示し過去の街並みを可視化する拡張現実アプリケーションを作成した．表示する全方位画像は，GPS と，スマートフォンのカメラで撮影した今いる場所の画像とデータベースの全方位画像の画像類似度を組み合わせて検索する．全方位画像の一部を切り出してユーザの見ている方向をセンサで検知して，見ている方向に対応する領域を表示し，顔検出によりユーザが本来向いている方向と端末の向きのずれを検出し，そのずれを補正する．

### 参考文献

[1] G. Klein and D. Murray: *Parallel tracking and mapping for small ar workspaces*, Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 1-10 (2007).

[2] N. Yazawa, H. Uchiyama, H. Saito, M. Servieres, G. Moreau and E. IRSTV: *Image based view localization system retrieving from a panorama database by surf*, IAPR Conference on Machine Vision Application, Yokohama, Japan (2009).

[3] 角田哲也, 大石岳史, 池内克史, バーチャル飛鳥京：複合現実感による遺跡の復元, 日本情報考古学会第 23 回大会, pp. 79-86 (March. 2007).

[4] 角田哲也, 大石岳史, 池内克史, 複合現実感における建物の陰影表現, 画像の認識・理解シンポジウム (MIRU 2005) (July 2005).

[5] 谷川陽彦, 久保守, 村上健一郎, 拡張現実感を利用した森林調査支援システムにおける魚眼画像と樹木モデルの位置合わせ, 映像情報メディア学会技術報告, Vol. 34 ,No. 22, pp. 25-28 (June. 2010).

[6] 笠田和宏, 鳴海拓志, 谷川智洋, 廣瀬通孝, 撮影位置への誘導による過去映像と現在風景のシームレスな接続, 映像メディア学会技術報告, Vol. 34, No. 25, pp. 117-122 (June. 2010).

[7] 横地祐次, 池田聖, 佐藤智和, 横矢直和, 特徴点追跡と GPS 測位に基づくカメラ外部パラメータの推定, 情報処理学会論文誌, Vol. 47, No. SIG5(CVIM13), pp.69-79 (March 2006).

[8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, *SURF: Speeded-Up Robust Features*, Computer Vision and Image Understanding, Vol.110, pp. 346-359 (2008).

[9] Marius Muja and David G. Lowe, *Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration*, in International Conference on Computer Vision Theory and Applications (VISAPP'09) (2009).