

スマートフォンカメラビューと 歩行軌跡のマッチングによる周辺人物の認識

岩橋 宏樹^{*1} 樋口 雄大^{*1} 山口 弘純^{*1,*2} 東野 輝夫^{*1,*2}

^{*1} 大阪大学 大学院情報科学研究科

^{*2} 独立行政法人 科学技術振興機構, CREST

{ h-iwahashi, t-higuti, h-yamagu, higashino }@ist.osaka-u.ac.jp

1 はじめに

自身の近辺にいる友人など、ソーシャルな関係を持つ他のユーザとのつながりをモバイル端末を通じて支援するサービスが、近年、高い注目を集めている。イベント会場など多くの人々が集まる環境において友人のいる場所までユーザを導くモバイルソーシャルナビゲーション [1] がその一例である。こうしたサービスにおいて、ユーザが周囲の人々との位置関係を把握するためには、自身および近隣のモバイル端末の位置を正確に推定する技術が欠かせない。一方、モバイル端末向けの測位システムである GPS や WiFi 測位は、一般に数 m~数十 m の誤差が生じるため、周辺の人物と位置推定結果とを対応付けることが難しいという課題がある。

そこで本稿では、スマートフォンのカメラ画像から抽出した周辺人物の動きの特徴量と、測位システムから得られる近隣スマートフォンユーザ群の動きの特徴量をマッチングすることで、カメラビュー内の人物を認識する手法を提案する。周辺人物との位置関係を直接比較する代わりに、移動軌跡の形状をマッチングの対象とすることで、位置推定誤差が認識精度に与える影響の軽減を図る。認識結果をカメラ画像に重畳することで、周辺人物との位置関係を直感的にユーザへ提示する AR ナビゲーションの実現を目指している。シミュレーション実験により、提案手法が周辺の人物を高い精度で認識できることを示す。

2 提案手法

2.1 想定環境

環境内のすべての人物がスマートフォン (クライアント) を保持していると仮定し、測位に必要な測定情報 (e.g., WiFi の受信電波強度) がクライアントから測位サーバへ定期的に収集されるものとする。測位サーバはこれらの情報をもとにクライアント間の相対位置を推定し、その結果をクライアントへフィードバックする。提案手法は測位方式とは独立であり、数 m 程度の測位精度を実現できれば、任意の方式を適用することができる。なお、3章の性能評価では、自律航法と Bluetooth の受信電波強度を用いた相対位置推定手法 [1] を想定して実験を行う。

各クライアントは、ユーザの操作に応じて、スマートフォンの内蔵カメラから自身の周辺の人物の画像を定期的に取得できるものとする。これらの画像に対して顔検出アルゴリズム [2] を適用することによって、図 1 (a) のように、画像内の人物の顔の位置と向きを推定結果が得られる。顔のサイズは人によらずほぼ一定であることから、検出された顔領域のサイズとカメラからの距離の関係をあらかじめモデル化しておくことで、カメラビュー内の人物との距離を推定することができる。また、画像

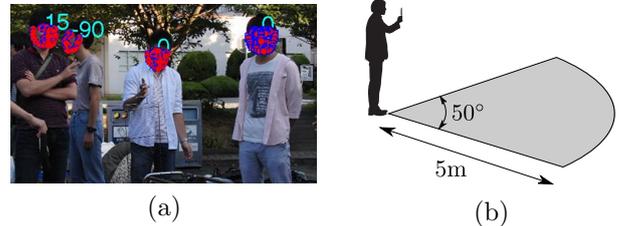


図 1: カメラ画像を用いた周辺人物との相対位置の認識

中の顔の位置とカメラの画角との関係から、カメラの撮影方向に対する周辺人物の相対的な方位を求めることが可能である。さらに、スマートフォンの加速度センサおよび電子コンパスの測定値からカメラの姿勢を推定すれば、カメラビュー内のクライアントとの相対位置を推定することができる。文献 [2] の顔検出アルゴリズムでは、カメラ画像内の人物を約 90% の精度で検出可能である。また、予備実験により、人物とカメラとの距離が離れるにつれ、顔の検出精度が徐々に低下することを確認している。本稿では、同図 (b) のように、ユーザの現在位置および撮影方向を基準として半径 5m, 方位角 $\pm 25^\circ$ の扇形の領域内にいる人物との相対位置を正確に推定できると仮定する。

2.2 マッチングアルゴリズム

すべてのクライアントからなる集合を S とする。各クライアント $A_i \in S$ は、近隣クライアント $A_j \in S$ の相対位置の推定値を τ 秒毎に測位サーバから受信し、過去 $N\tau$ 秒間の相対位置の履歴 $P_j^{est} = \langle \hat{p}_{j,t-(N-1)\tau}, \dots, \hat{p}_{j,t-\tau}, \hat{p}_{j,t} \rangle$ を保持する。また、 A_i は端末の内蔵カメラから τ 秒毎に周辺人物の画像を取得し、それらの画像に対して顔検出アルゴリズムを適用することで、カメラビュー内の人物との相対位置を推定する。ここで、過去 $N\tau$ 秒間に A_i のカメラビュー内に入ったすべての人物の集合を $M_i \subset S$ 、カメラ画像から推定された人物 $B_k \in M_i$ の相対位置の履歴を $P_k^{cam} = \langle p_{k,t-(N-1)\tau}, \dots, p_{k,t-\tau}, p_{k,t} \rangle$ とする。なお、ある時刻 t' において人物 B_k が一時的にカメラ画像内から検出されなかった場合には、 $p_{k,t'} = null$ とする。

P^{est} や P^{cam} に含まれる相対位置の推定誤差が認識精度に与える影響を軽減するため、提案手法では、クライアントの移動軌跡の形状の類似性をもとに、カメラビュー内の人物 $B_k \in M_i$ に対応するクライアント $A_{k'} \in S$ を推定する。まず、 P_k^{cam} をもとに、カメラビュー内の人物 $B_k \in M_i$ の相対位置の τ 秒ごとの変化量 $u_{k,t} = p_{k,t} - p_{k,t-\tau}$ (移動ベクトル) を算出し、 P_k^{cam} から移動ベクトルの系列 $U_k^{cam} = \langle u_{k,t-(N-2)\tau}, \dots, u_{k,t-\tau}, u_{k,t} \rangle$ を求める。なお、 $p_{k,t}$ と $p_{k,t-\tau}$ のいずれかが $null$ の場合には $u_{k,t} = null$ とする。同様に、測位サーバから受信した $A_j \in S$ の推定位置の履歴 P_j^{est} も、移動ベクトルの系列 $U_j^{est} = \langle \tilde{u}_{j,t-(N-2)\tau}, \dots, \tilde{u}_{j,t-\tau}, \tilde{u}_{j,t} \rangle$ に変換する。

2つのベクトル系列間の類似度の指標として、提案手法では、式 (1) で定義される編集距離を用いる。ここでは、ベクトル系列の先頭要素間のユークリッド距離に応じた

Trajectory-based People Recognition for AR Social Navigation with Smartphones

Hiroki Iwahashi^{*1} Takamasa Higuchi^{*1} Hirozumi Yamaguchi^{*1,*2} Teruo Higashino^{*1,*2}

^{*1} Graduate School of Information Science and Technology, Osaka Univ., Osaka Japan

^{*2} Japan Science and Technology Agency, CREST

ペナルティを課しながら、先頭要素を逐次的に取り除いていくことで、系列間の乖離度を算出している。このとき、ペナルティに上限を設けることで、各要素が持つノイズが類似度の評価値に与える影響の軽減を図っている。

$$ED(U, V) = \begin{cases} n \cdot 2\epsilon & \text{if } m = 0 \\ m \cdot 2\epsilon & \text{if } n = 0 \\ \min\{ED(\text{Rest}(U), \text{Rest}(V)) + c, \\ ED(\text{Rest}(U), V) + 2\epsilon, \\ ED(U, \text{Rest}(V)) + 2\epsilon\} & \text{otherwise} \end{cases} \quad (1)$$

ここで、 U, V は比較対象のベクトル系列、 n, m はそれぞれ U, V の要素数、 $\text{Rest}(U)$ は U から先頭要素を取り除いた系列を表す。 c は U, V の先頭要素 u, v が $\|u - v\| < \epsilon$ を満たすとき $c = \|u - v\|$ 、それ以外の時 $c = 2\epsilon$ となるペナルティ関数である。 ϵ はペナルティの上限を表す定数であり、本稿では経験的に $\epsilon = 0.5$ とする。

マッチングを行うためには、 $B_k \in \mathcal{M}_i$ と $A_j \in \mathcal{S}$ のすべての組み合わせについて、式 (1) で定義される軌跡間の編集距離 $ED(U_k^{cam}, U_j^{est})$ を算出する。なお、 U_k^{cam} の中に $null$ 要素が含まれる場合には、それらの要素をあらかじめ系列から削除するとともに、 U_j^{est} から対応する時刻の移動ベクトルを取り除いた上で編集距離の計算を行う。

最後に、カメラビュー内のそれぞれの人物 $B_k \in \mathcal{M}_i$ について、 $ED(U_k^{cam}, U_j^{est})$ が最小となるようなクライアント $A_j \in \mathcal{S}$ (同じ編集距離を持つ複数の候補が存在する場合には、それらの集合) を認識結果として出力する。

3 性能評価

3.1 シミュレーション設定

提案手法によるマッチングの精度を検証するため、シミュレーションによる性能評価を行った。30m×30mの懇親会会場を想定し、クライアント数は20とする。一般に、懇親会会場では、複数の参加者からなるいくつかのグループが構成される。そこで、クライアントは4人1組のグループ単位で移動すると仮定し、各グループの平均的な振る舞いを表す5つの参照点を作成する。それぞれの参照点は、Random Waypoint モデルに従ってフィールド内を独立に移動する。移動先の waypoint はフィールド内からランダムに選択され、参照点は目的地に向かって一定の速度で移動する。目的地の waypoint に到着すると、確率 p で次の目的地へと移動を開始し、確率 $(1 - p)$ で10秒間その場に静止する。本稿では、 $p = 0.5$ とする。グループの参照点が次の目的地を選択すると、そのグループに属するクライアントは、参照点を選択した waypoint を中心とする半径1.5mの領域内から自身の行き先 waypoint をランダムに選択する。以上により、各クライアントは、グループの参照点を中心として、集団的に振る舞う。

上記のモビリティモデルに対して文献 [1] の相対位置推定アルゴリズムを適用することで、測位サーバから受信される相対位置 P^{est} を求める。測定値の誤差は、[1] で定義された自律航法の誤差モデルおよび Bluetooth の受信電波強度のモデルに基づき決定する。このアルゴリズムにより、10m以内の距離にある近隣クライアントとの相対位置が平均3~4mの精度で推定される。

また、各時刻におけるクライアント間の正しい相対位置情報から図1(b)で定義されるカメラビュー領域に相当する範囲を切り出すことで、カメラ画像に基づく相対位置の推定結果 P^{cam} を擬似的に生成する。なお、カメラの撮影方向は、ユーザの移動方向に一致すると仮定する。

画像の撮影間隔は $\tau = 2$ 秒とし、マッチングのウィンドウサイズ N は1~15の範囲で変化させた。以上の条件の下で200秒間のシミュレーション実験を行い、 τ 秒毎

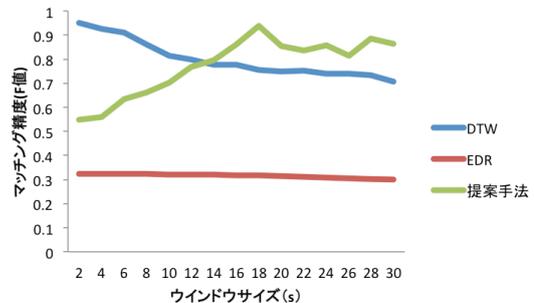


図2: マッチング精度

に、過去 $N\tau$ 秒間の相対位置の履歴を用いてカメラビュー内の人物とグループとのマッチングを行った。提案手法では、カメラビュー内のある人物に対して複数の候補がマッチする可能性があることから、マッチングの適合率と再現率をそれぞれ求め、F-measure によって精度を評価した。また、比較対象として、移動軌跡間の類似性の指標として一般に広く用いられている Dynamic Time Warping (DTW) および Edit Distance on Real Sequences (EDR) を適用した場合についても評価を行った。

3.2 シミュレーション結果

シミュレーションによる平均マッチング精度の評価結果を図2に示す。DTWでは、移動ベクトル間のユークリッド距離をペナルティとして P^{est} と P^{cam} の乖離度を評価する。このため、 P^{est} または P^{cam} のいずれかに誤差が極端に大きい要素が含まれていると、他の要素の類似性とは無関係に、2つのベクトル系列が大きく乖離していると判定される。ウィンドウサイズが大きくなるにつれてマッチング精度が徐々に低下しているのはこのためである。EDRでは、類似度算出のためのペナルティとして常に一定の値を用いることで、こうしたノイズの影響の軽減を図っている。しかし、複数のベクトル系列の組み合わせに対して同じ類似度が出力される場合が多く、マッチングの候補を一意に絞り込むことが難しい(適合率が低い)。一方、提案手法では、ユークリッド距離を直接ペナルティとして用いることで類似度の評価結果の多様性を保ちつつ、ペナルティに上限を設けることで、DTWで問題となるノイズの影響を軽減している。ウィンドウサイズ N が十分に大きい場合には、90%前後の精度が実現されており、マッチング結果を少ない候補数に絞り込みながら、高い適合率が得られることが確認できた。

4 まとめと今後の課題

本稿では、モバイル端末間の相対位置情報とカメラビューから推定される周辺人物のトポロジ情報をマッチングすることで、ユーザの近隣の人物を認識する手法を提案した。シミュレーション実験により、カメラビュー内の人物を約90%の精度で識別できることを示した。カメラ画像に基づく相対位置推定誤差のモデル化や、Android 端末上への実装と評価を行うことが今後の課題である。

参考文献

- [1] Higuchi, T., Yamaguchi, H. and Higashino, T.: Clearing a crowd: context-supported neighbor positioning for people-centric navigation, *Proc. of Pervasive '12*, pp. 325–342 (2012).
- [2] Zhu, X. and Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild, *Proc. of CVPR '12*, pp. 2879–2886 (2012).