様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２８　年　　５　月　３０　日現在

機関番号：　１４４０１
研究種目：　基盤研究(A)（一般）
研究期間：　2012～2015
課題番号：　２４２４００３１

研究課題名（和文）センシングで得られるプライバシー情報の開示に調和したユーザ利得の創出

研究課題名（英文）Creation of profit harmonized with disclosure of privacy information

研究代表者
馬場口　登（BABAGUCHI, Noboru）

大阪大学・工学（系）研究科（研究院）・教授

研究者番号：３０１５６５４１

交付決定額（研究期間全体）：（直接経費）　　35,200,000円

研究成果の概要（和文）：本課題では、ユーザがプライバシー情報(属性、顔、表情、動作、移動履歴、位置軌跡、興味など)を開示することに応じて高い利得（有益な情報提供など）を得ることの可能な情報基盤HIFI（Harmonized Information Field）の構成に関する研究を実施した。HIFIは実世界における空間的に限定された場を対象に設計され、HIFI内部の処理メカニズムを、プライバシー情報の収集、保護、組織化、活用の各プロセスに分け、各々を具体化した。

研究成果の概要（英文）：In this research, we aimed to construct HIFI (Harmonized Information Field) where a visitor at HIFI can get some profits, e.g. useful recommendations, in accordance with disclosure of his/her privacy information such as attribute, face, expression, action, position, and interest. HIFI is designed for a distinct place in the real world, consisting of processing modules of sensing, protection, organization and utilization of privacy information.

研究分野：　情報学　知覚情報処理

キーワード：　プライバシー情報処理　センシング　ユーザ利得

１．研究開始当初の背景

研究代表者は、本研究課題に着手するまでに PriSurv (Privacy Protected Video Surveillance System)・センシング Web・MPP (Mobile Privacy Protection)プロジェクトなど、視覚情報のプライバシー保護を中心に研究開発を進めてきた。そこでの背景概念は、プライバシー情報を自ら制御できる範囲で開示あるいは保護することであった。プライバシー情報の利活用をさらに進めるという観点から、ユーザがプライバシー情報を開示する度合いに応じた利得をそのユーザが得るという考え方の重要性を認識するようになった。

そこで、プライバシー情報の開示と利得が調和し、実空間とサイバー空間がリンクした情報空間（本研究では Harmonized Information Field:HIFI)の構築を最終目標とし、センシングを通して開示されたプライバシー情報が安心して集積され、有効活用されることにより、ユーザに上質な情報をフィードバックするメカニズムの構成するプロセスを明らかにするために本研究課題を開始した。

２．研究の目的

本研究課題では、ユーザ（HIFI への来場者）がプライバシー情報(ユーザ ID と結びつけられた人間に関する種々の情報、すなわち顔、容姿、動作、移動履歴、位置軌跡、嗜好、興味など)を開示することに応じて高い利得を得ること、すなわち有益性、上質性などを有する情報提供が実現されるメカニズムを明らかすることが目的である。HIFI は実世界における空間的に限定された場を対象に設計される。HIFI 内部の処理メカニズムを、プライバシー情報の収集、保護、活用の各プロセスに分け、各々を実現する手法を開発する。

３．研究の方法

HIFI の構成要素技術[論文 4、発表 10]を以下の(1)(2)(3)に分け、研究目的の達成を図る。

(1)プライバシー情報の収集

①HIFI 出入口での収集・管理

来場者の顔、服装、性別、年齢といった静的なプライバシー情報を HIFI の出入口において収集する情報エントリーシステム IES (Information Entry System)を構築する。その際、来場者自身が情報開示度を決定できることを重視する。来場者が意図せずプライバシー情報が流出することを避けるため、何を開示／秘匿するかは来場者自身が選択する方式とする。加えて、その選択を円滑に行えるようにするためのインタラクション機構を開発する。なお、上述のインタラクションやプライバシー情報の収集には、来場者に大きな負荷がかからないようにする。

②HIFI 内部での収集

HIFI 内部における来場者の位置情報など、動的なプライバシー情報を収集するための技術を開発する。動的なプライバシー情報の具体例として、HIFI 内部における各来場者の位置履歴および来場者間のグループ関係（家族、友人、カップルなど）に焦点を当てる。

(2)プライバシー情報の保護

①属性の匿名化

ユーザ（来場者）側、サービス提供者側の 2 つの観点から、属性の保護手法を検討する。

ユーザ側の観点からのアプローチでは、ユーザ毎に開示したい情報や求めるサービスの質・量には違いがあると考え、異なる情報を開示する 5 種類の ID 種別（実名、3 種類の仮名、匿名）を定義し、ユーザの主観に応じて ID 種別を選択した後、ID 種別によって異なる匿名化を行う。3 種類の仮名は、個人情報、当日のみの行動情報、当日までの行動情報の開示/非開示により分類する。また、実名を除く ID 種別は実世界との可到達性を遮断するため、同一のデータを持つ個人が k 人以上になるよう、属性の値を汎化する k-匿名化処理を ID 種別毎に施す。

サービス提供者側の観点からのアプローチでは、可到達性を遮断するために用いた k-匿名化処理に焦点を当てる。従来の k-匿名化手法では、匿名化後のデータの情報損失が少なくなるよう汎化処理するため、サービス提供者が活用したい属性の値が詳細に記述されているとは限らない。そこで、具体的なサービスとしてユーザへの情報推薦を考えた際、推薦に必要な情報の保存という観点から評価するために TF-IDF 法を匿名化処理に導入する。TF-IDF 法による評価値は、ある推薦情報が匿名化後のデータ内でどれだけ固有でかつ重要な情報であるかを示すため、TF-IDF 法による値がより高くなるよう汎化することにより、情報推薦に有用な匿名化データを導ける。

さらに、汎化に利用される階層も、推薦に適するように自動生成する。汎化する属性には、数値属性とカテゴリカル属性の 2 種類が存在する。数値属性に対しては、CF-Tree によるクラスタリングを応用し、非階層的クラスタリングによるクラスタの値域をノードとし、階層的クラスタリングの結果に従い併合することにより階層を生成する。一方、カテゴリカル属性に対しては、基本となる階層構造を用意した後、単語間の距離を測定する Normalized Web Distance を用い、最も類似している属性値の下位ノードとして追加することにより生成する。

②プライバシー保護画像の定量評価

視覚情報の画像には撮影者、観察者、被写体といったステークホルダーが存在し、「誰が」「誰を」見るかによってプライバシーの侵害と感じる度合いが異なると予想される。これは、プライバシーの問題の主観性に加えて、観察者が被写体に関してどれだけの知識を持っているかにも左右される。一方で、視覚情報に対するプライバシー保護では、主に

ぼかしや塗りつぶしなどの画一的な画像処理が用いられており、上記のような状況においてどの程度プライバシーが保護されるかは明らかではなかった。

そこで本研究では、プライバシー問題の主観性に対して、被写体の同定可能性（被写体が同定される割合）によりプライバシー侵害の程度を定量化するとともに、観察者が被写体をどれだけ知っているか（親密さ）、及び被写体の外見が目立つか（顕著さ）の 2 点に着目し、画像処理によるプライバシー保護の度合いを明らかにするために被験者 100 名以上の大規模な調査実験を行った。

③プライバシー保護映像の自動生成

モバイルデバイスにより撮影された映像には、その映像中で主要な人物（重要人物）と、偶然映り込んだ人物（非重要人物）が含まれる。本研究では、重要人物については映像撮影などの許可が得られるものと仮定し、全ての人物を重要人物・非重要人物に分類した上で、非重要人物に対して自動で選択的にプライバシー保護処理を適用する手法を提案する。ここでは人物を完全に除去する保護処理を考える。

プライバシー保護においては、人物の検出漏れなどによる保護漏れが大きな問題となることから、本手法では非重要人物の除去ではなく、フレーム全体の背景を推定し、重要人物のみを抽出して背景上に重畳するアプローチを採用する。さらに、重要人物・非重要人物の識別性能がプライバシー保護性能に影響することから、条件付き確率場を利用して、重要人物の空間的特性（重要人物は並んで出現しやすいなど）を援用する。

(3)プライバシー情報の活用

プライバシー情報を活用したサービス提供を実現するための基盤技術を開発する。具体例として、来場者がこの後訪問する可能性の高いエリアの情報（注目情報や評判など）を提供するサービスを考え、これを来場者側がプッシュ型で享受できるようにするための訪問エリア予測手法を開発する。また、同様のエリア情報をエリア外観の撮影等により来場者がプル型で取得できるサービスを考え、これを継続的に利用した際に生じるプライバシー問題を緩和する技術を開発する。

４．研究成果

(1)プライバシー情報の収集

①HIFI 出入口での収集・管理

本研究で開発した IES の概要を図 1 に示す。IES は、来場者の顔や全身を観測するためのセンサ群、および来場者とのインタラクションを実現するためのタッチモニタから構成される。タッチモニタには HIFI 内部を模した 3 次元仮想環境が表示され、これを介したインタラクションにより来場者はプライバシー情報の開示度を適切に選択するための知識を得る。仮想環境の表示は、タッチモニタの中央上部に配置されている「来場者種別」のボタンに触れることにより、その種別に応じた内容へとリアルタイムに変化する。来場者種別には、全てのプライバシー情報が収集対象となる「実名」、一部の情報のみが収集される「仮名」、いずれの情報も収集されない「匿名」などがあり、自らの情報開示を制御できる[発表 6]。

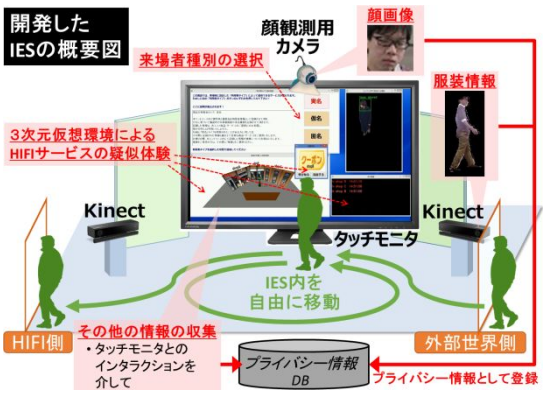

図 1 IES の概要図

プライバシー情報の収集は来場者が来場者種別を選択するためにモニタ中央上部を注視するタイミングで行う。具体的には、来場者の正面顔をタッチモニタ上部に設置したカメラで観測し、その画像から顔、性別、年齢などの情報を収集する。同時に、タッチモニタ脇のセンサにより服装情報も収集する。これらの情報は選択された来場者種別に応じて適応的に登録・破棄される。

来場者への負荷を評価する目的で、10 名の被験者に IES を利用してもらい、その際に感じた負荷の大きさを 5 段階（1 が最も負荷が小さく、5 が最も大きい）で評価してもらった。この結果、10 名中 9 名の被験者が 2 以下の評価値を付けたことから、IES が来場者にとって比較的負荷の小さいシステムであることが確かめられた[発表 9]。

②HIFI 内部での収集

まず、来場者（ユーザ）の移動軌跡の抽出手法は、全てのユーザの移動軌跡の推定や更新をモバイル端末や固定カメラから得られたセンサ情報を反映したコスト関数の最適化（軌跡群最適化）により行う[発表 8]。

軌跡群最適化に用いられるコスト関数は、ステップベクトル（一歩あたりの歩幅と方向）と移動速度の関係に関する項、他の端末が発する Bluetooth の電波強度とユーザ間距離の関係に関する項、固定カメラから得られる位置と推定された位置の関係に関する項の 3 項からなる。

ステップベクトルと電波強度はユーザの所持するモバイル端末で取得する。モバイル端末、固定カメラ、移動軌跡を推定するためのサーバはネットワークを介して接続されており、モバイル端末や固定カメラから新たなセンサ情報が得られるたびに（これをイベントと呼ぶ）、サーバは軌跡群最適化を行うことによって、全てのユーザの移動軌跡を一括して推定する。

この最適化問題は大規模なものであるが、比較的発生頻度の低い、Bluetooth の電波強度取得イベントや固定カメラ映像取得イベントが発生した際の情報のみを用いて最適化を行い、その後、発生頻度の高いステップベクトル取得イベントの影響を考慮するという、段階的な最適化手法を開発した。後者は最適化のための繰り返し計算が不要なため、計算負荷を著しく低減できる。

また､ユーザが HIFI 内部に入場した際に、各ユーザの歩幅や各モバイル端末の方位センサの偏角を学習するための領域を設け、ユーザやモバイルセンサの特徴を考慮した推定を行う手法を開発した[発表 1]。

開発した手法の推定精度と計算時間について、実験を通して評価した。T 字型の領域（横線 7m、縦線 10m、固定カメラを T の交差点付近に設置）において、3 名のユーザが動きながら 60 秒間滞在する場合、推定誤差の平均は､時間の経過とともに大きくなり 60 秒後には約 3m となった。精度が良くない原因の一つとして、固定カメラ映像取得イベントがうまく検出されなかったことが挙げられる。ユーザの歩幅やモバイル端末の方位センサの偏角特性を学習するために有効な動きについても検討を行い、様々な方向に渡って計 10m 程度移動するのが適切であることがわかった。学習の組合せに応じた推定結果の一例を図 2 に示す。

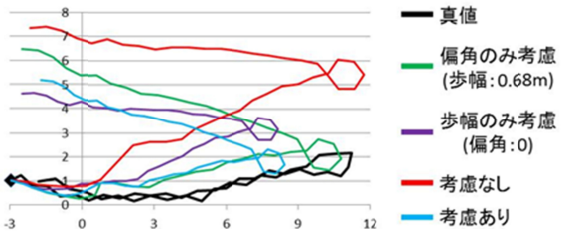

図 2 学習の組み合わせに応じた推定結果

次に、来場者の位置履歴情報、すなわち軌跡から来場者間のグループ関係を推定する手法を考案した[発表 5]。基本的には、互いの軌跡が時空間的に類似している来場者同士を一つのグループとして統合することによりグループ関係を推定することが可能であるが、例えば一時的に別行動をとりその後合流する場合など、状況によっては同一グループに属する来場者の間で軌跡が類似しないこともある。この点を考慮し、本手法では、相異なる軌跡同士が同一のグループに属しているか否かを判定するための識別モデルを状況ごとに複数種類構築する一方、判定対象の利用者群に対してはそれらのモデルを確率的に適用することにより、高い精度で来場者間のグループ関係を推定することを実現した。

表 1 は、本手法の精度を評価するため、3 種類の軌跡データセットを対象に実験を行った際の推定精度をまとめたものである。表中の「従来手法」とは、状況に関係なく常に同じモデルを用いる手法のことを指す。特に推定精度が低く、より複雑な環境であると言えるデータセット 3 において従来手法からの改善幅が最大となったことから、本手法は複雑な環境にも対応し得ることが示唆される。

表 1 本手法と従来手法の精度比較

データセット ID	1	2	3
従来手法の精度	77.2%	71.4%	41.9%
本手法の精度	78.9%	73.5%	44.8%

(2)プライバシー情報の保護

①属性の匿名化

ユーザの観点からは、実名、仮名、匿名といった 3 種類だけでなく、仮名を詳細に分割することにより、よりユーザの主観に適した保護処理を実現する枠組みを構築した。また、性別、生年月日、職業、身長の 4 つの属性を仮想的に与えシミュレーションした結果、75,000 人以上であれば、全ての属性の値が抑圧状態にならず利活用できる状態になるため、既存の k-匿名化処理が適用可能であることを示した[発表 11]。

TF-IDF 法を用いた匿名化処理では、上記の 4 属性の情報を持つ 100 名の仮想データに対して実験した結果、従来の情報量を重視した匿名化では、表 2 の通り生年月日のデータが抑圧されるものの、提案手法では生年月日を重要視した結果、表 3 の通り生年月日のデータを残した匿名化処理が実現されたことが分かる。身長を重要視した際は、10cm 毎の匿名化データが多く生成されたこと、汎化の階層を自動生成した際も同様の結果が得られたことから、情報推薦に適した匿名化データの生成が可能となった[発表 4,7]。

表 2 従来の情報量に基づく匿名化の結果例

性別	生年月日	職業	身長	生成数
男女	*	大分類	*	43 人
男女	*	*	10cm 毎	40 人

表 3 生年月日を重視した匿名化の結果例

性別	生年月日	職業	身長	生成数
男女	10 年毎	*	*	43 人
男女	10 年毎	*	10cm 毎	32 人

②プライバシー保護画像の定量評価

図 3（a）に、評価対象とした画像処理の一例（ぼかし）を、また図 3（b）に評価結果を示す（左が親密さ、右が顕著さ）。この結果から、よく見知った人物や特徴的な人物については、評価対象とした最大のぼかし度合いにおいても 6 割程度のサンプルで人物を同定可能であることが明らかになった[論文 2]。これは、ぼかしによるプライバシー保護処理の効果が限定的であることを示す。本研究では、この他に、拡大縮小（保護処理を適用しない場合に相当、画像中でのサイズによる同定可能性を示す）や、塗りつぶしなどを評価対象とした。この成果は、今後の画像・映像

に対するプライバシー保護処理に関する研究のみならず、実用的なシステムにおいてもその設計指針となるものであり、学術的・社会的な価値が高いと考える。

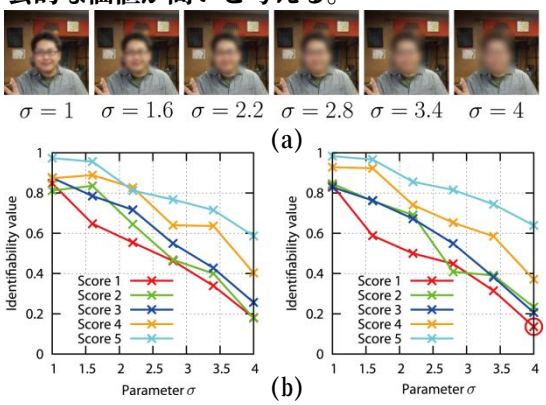

図 3 評価対象の画像処理例と評価結果

③プライバシー保護映像の自動生成

図 4（左）から、本研究で提案した条件付き確率場による識別性能は、人間には及ばないものの、ベースラインとなる SVM を用いた識別や既存手法に比べて高いことが分かる[論文 1]。また、図 4（右）のフレーム例から、背景推定によるプライバシー保護処理の効果を確認した。

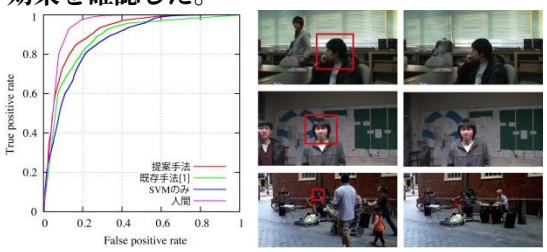

図 4 ROC 曲線による識別性能の評価結果（左）、フレーム例（中央）とプライバシー保護処理結果（右）、中央の赤枠は重要人物

(3)プライバシー情報の活用

①来場者の訪問エリア予測

来場者の訪問エリアは、その来場者の性別・年齢やグループ関係などと強い相関があると考えられるが、そのようなプライバシー情報と現実のエリア訪問行動の関係を数学的なモデルとして記述することは困難である。そこで、サービス対象の来場者と同じプライバシー情報を設定した仮想来場者がエリア訪問行動を行うシミュレーションシステムを開発し、その結果を統計的に解析することにより当該の来場者の訪問エリアを予測する手法を考案した[発表 2]。開発したシミュレーションシステムの外観を図 5 に示す。

現実の複合商業施設を訪れた被験者の訪問エリア履歴を正解データとして精度評価実験を行った結果、次時刻の訪問エリアの予測に関してはプライバシー情報を考慮しない場合よりも高い精度が得られることが確認できた。

図 5 シミュレーションシステムの外観

②来場者の位置情報による情報提供

来場者が能動的に対象エリアの外観を撮影してサーバに送信し、サーバ側で画像からエリア ID を認識して当該エリアの情報を返送するというサービス形態では、来場者が位置履歴情報の開示を選択していない場合でもエリア ID の系列という形で位置履歴情報が流出し得る。この問題を回避するため、来場者からサーバに画像が送信される際、一部の情報を改変することによりサーバ側でのエリア ID の認識を困難にする一方、サーバからはエリア ID の候補とその代表画像の組をエリア情報と共に複数返送するという枠組みを考案した[発表 3]。来場者側では、改変前の画像をサーバから返送されてきた画像と照合することにより、エリア ID を一意に認識することが可能となる。

10 か所のエリアを対象に画像からエリア ID を認識する実験を行った結果、改変前の画像では 99.5%であったサーバ側の認識率（プライバシー流出度に相当）が 44.5%まで抑えられた一方、来場者側の認識率としては 76.5%を確保できた。

５．主な発表論文等

〔雑誌論文〕（計１８件）

[1]Y. Nakashima, N. Babaguchi, J.-P. Fan: "Privacy Protection for Social Video via Background Estimation and CRF-based Videographer's Intention Modeling", IEICE Transactions on Information and Systems, 査読有, Vol.E99-D,No.4, pp.1221-1233, April 2016. DOI: 10.1587/transinf.2015EDP7378

[2]Y. Nakashima, T. Ikeno, N. Babaguchi: "Evaluating Protection Capability for Visual Privacy Information", IEEE Security & Privacy, 査読有, Vol. 14, No. 1, pp. 55-61, February 2016. DOI: 10.1109/MSP.2016.3.

[3]N. Nitta, N. Babaguchi: "Digital Diorama: Privacy-Preserving and Intelligible Sensing-Based Real-World Content", ITE Transactions on Media Technology and Applications, 査読有, vol.3, no.3, pp.184-193, July 2015. DOI: 10.3169/mta.3.184

[4]N. Babaguchi, Y. Nakashima:

"Protection and Utilization of Privacy Information via Sensing", IEICE Transactions on Information and Systems, 査読有, vol. E98-D, no. 1, pp. 2-9, January 2015. DOI: 10.1587/transinf.2014MUI0001 ＜招待論文＞

〔学会発表〕（計 39 件）

[1]今西健児，伊藤義道，馬場口登: "個人及びデバイス特性を考慮した軌跡群最適化による複数ユーザの位置推定"，電子情報通信学会技術研究報告，PRMU2015-168，pp. 19-24，March 2016.

[2]宮崎永爾，中村和晃，馬場口登: "複合商業施設における回遊行動シミュレーションに基づく来場者グループの訪問店舗予測"，電子情報通信学会 2016 年総合大会，D-8-1，p. 96，九州大学，March 2016.

[3]藤井宏次朗，中村和晃，馬場口登: "画像認識に基づく情報提供サービスのための利用者位置履歴の保護"，電子情報通信学会 2016 年総合大会，D-21-5，p. 218，九州大学，March 2016.

[4]新井健介，河野和宏，馬場口登: "推薦対象の属性から構築した階層構造を用いた TF-IDF 法による匿名化処理"，電子情報通信学会技術研究報告，vol. 115，no. 479，EMM2015-81，pp. 31-36，屋久島，March 2016.

[5]小野士，中村和晃，馬場口登: "グループの行動状態を考慮した群集中のグループ検出"，電子情報通信学会技術研究報告，vol. 115，no. 388，PRMU2015-106，pp. 81-86，信州大学，December 2015.

[6]大西祐貴，小野士，中村和晃，馬場口登: "プライバシー情報利活用空間への入場時における利用者ガイダンスシステム"，第 14 回情報科学技術フォーラム(FIT2015)，K-002，pp. 437-438，愛媛大学，September 2015.＜第 14 回情報科学技術フォーラム FIT 奨励賞受賞＞

[7]新井健介，河野和宏，馬場口登: "TF-IDF 法によるユーザへの情報推薦のための匿名化処理"，電子情報通信学会技術研究報告，vol. 115，no. 38，IT2015-10，EMM2015-10，pp. 51-56，京都，May 2015.

[8]遠藤健，伊藤義道，馬場口登: "移動軌跡群最適化に基づく屋内環境における複数ユーザの位置推定"，電子情報通信学会技術研究報告，PRMU2014-161，pp. 11-16，March 2015.

[9]小野士，中村和晃，馬場口登: "実空間における適応型サービスのための情報エントリーシステム"，2014 年映像情報メディア学会年次大会，7-4，大阪大学，September 2014.

[10]N. Babaguchi: "Protection and Utilization of Privacy Information "，First International Workshop on Information Hiding and its Criteria for Evaluation (IWIHC2014)，(in conjunction with ASIACCS 2014)，1 page，Kyoto，JAPAN，June 2014. ＜招待講演＞

[11]新井健介，河野和宏，馬場口登: "ユーザの主観に適応した ID 種別毎のプライバシー保護"，電子情報通信学会 2014 年総合大会，D-21-3，p. 194，新潟大学，March 2014.

〔図書〕（計 1 件）

[1]N. Nitta，R. Akai，N. Babaguchi: "People Counting Across Non-overlapping Camera Views by Flow Estimation Among Foreground Regions", Human Behavior Understanding in Networked Sensing - Theory and Applications of Networks of Sensors，eds. P. Spagnolo，P. L. Mazzeo，C. Distante，Chapter 11，Springer Verlag，pp.239-259，2014.

〔その他〕
ホームページ「センシングで得られるプライバシー情報の開示に調和したユーザ利得の創出」
http://www2c.comm.eng.osaka-u.ac.jp/proj/hi-fi/index.html

６．研究組織
(1)研究代表者
馬場口　登（BABAGUCHI，Noboru）
大阪大学・大学院工学研究科・教授
研究者番号：30156541

(2)研究分担者
なし

(3)連携研究者
新田　直子（NITTA，Naoko）
大阪大学・大学院工学研究科・准教授
研究者番号：00379132

伊藤　義道（ITO，Yoshimichi）
大阪電気通信大学・工学部・准教授
研究者番号：10263203

河野　和宏（KONO，Kazuhiro）
関西大学・社会安全学部・准教授
研究者番号：60581238

中村　和晃（NAKAMURA，Kazuaki）
大阪大学・大学院工学研究科・助教
研究者番号：10584047