

Hadoop を用いた赤外線測域センサデータ処理

佐藤 充[†] 石松 俊文[†] 川田 翔士^{††} 岩楯 翔仁^{††} 白井 暁彦^{††}

大塚 真吾^{††}

[†] 神奈川工科大学情報学部情報工学科 〒243-0292 神奈川県厚木市下荻野 1030

^{††} 神奈川工科大学大学院工学研究科 〒243-0292 神奈川県厚木市下荻野 1030

E-mail: [†]{s1021079,s0921009,s1285014}@cce.kanagawa-it.ac.jp, ^{††}ubicode@shirai.la,
^{†††}{shirai,otsuka}@ic.kanagawa-it.ac.jp

あらまし 近年創出されているコミュニケーションロボット、デジタルサイネージといったデジタルコンテンツを体験しているユーザに対して、アンケートなどの主観的な印象評価ではなく、物理的な評価を行うことで、システムの質的向上や工学的なコンテンツ開発ノウハウを蓄積できる。さらに、その応用として広告メディアの価格決定などを行うことが可能となる。我々はこの課題に対し、非接触でユーザ属性を測ることが可能な赤外線レーザーによる Time of Flight を用いた円形スキャンが可能な赤外線測域センサを使用し、ユーザ非装着型エンタテインメント空間評価手法 ResBe(Remote entertainment space Behavior evaluation) の研究を進めてきた。ResBe の問題点としては、ユーザの特徴的な行動抽出に時間がかかる点があげられる。そこで、本稿では Hadoop を用いることで、行動抽出の高速化の検討を行う。

キーワード Hadoop, 赤外線センサ, 測域センサ

1. はじめに

近年、センシング技術の発展に伴い多様なセンサより取得した膨大なデータを元に、ユーザ行動評価を行う研究が、戦略的創造研究推進事業における東倉らによる「共生社会に向けた人間調和型情報技術の構築」[1] などで進められてきている。空間内におけるユーザの自然な行動を、アンケートや主観的評価ではなく物理的に測定する事は、インタラクティブなコンテンツなどの質的向上やデジタルサイネージなどの広告メディアを設置する際に設定する料金の参考データとなるなどその意義は大きい[2]。

例えば、駅構内での広告掲載の料金設定は、広告が見られる機会は駅の利用者に比例して増加するという考えから、1日あたりの乗降客数や路線数などのデータをもとに算出されていると考えられる[3]。しかし、そのような手法で定めた広告料設定が、果たして設置された場所に対する対価として本当に適切であるかは疑問である。

我々はこれらの問題に対して、ユーザの自然な行動を物理的に測ることが可能な赤外線レーザーによる Time of Flight を用いた赤外線センサを使用し、ユーザ非装着型エンタテインメント空間評価手法 ResBe(Remote entertainment space Behavior evaluation) の研究[4][5]を進めており、ResBe から得られたデータを HeatMap 手法によって可視化を行うことで、ユーザの意味のある行動の抽出の可能性について検討を行なっている。

この手法の問題点としては、ユーザの特徴的な行動抽出に時間がかかる点があげられる。そこで、本稿では Hadoop、お

び、Hive を用いることで、行動抽出の高速化の検討を行う。

2. 関連研究

本研究の関連研究としては、里らによる遠隔臨場感システムにおける HMD 装着による注視点測定[6]や、上岡らによる血流・血圧や呼気、GPS 装着による測位など装着物を使った測定法[7]、江田らによる個体識別を目的としないセンサネットワークから、人の動的な行動把握を行う事で省エネを目指す手法[8]、無線ネットワークから得られる赤外線センサデータを用いて人の移動や在席状況を同時に推定することで消費エネルギーのムダを省く手法[9]。そして、中根ら[10]による、各階に設置されたセンサから得られた消費電力および人の行動といったセンシングデータを、全体から俯瞰できるような可視化システム、森による、床や椅子、ベッドや家電など日常使用するものをセンサとして用いる事で、室内に人の存在を検出するようなセンシングルームを構築し、人の行動パターンを予想する研究[11]、体験者へのアンケートやゲームシステムにおけるコイン投入数などのマーケティング手法、船津らによる CG アニメーション生成における確率場によるシミュレーション[12]などが存在する。

事例としては、「情報爆発時代に向けた新しい IT 基盤技術の研究」として京都大学キャンパス内に構築されている多視点の映像・音声記録システム、モーションキャプチャシステム、アイマレコーダー、生体データ計測機器が利用できる実験工房「IMADE-PROJECT」[13]、吉田らのショッピングカートや携帯電話などの端末に取り付けられた RFID や IC などのタグ

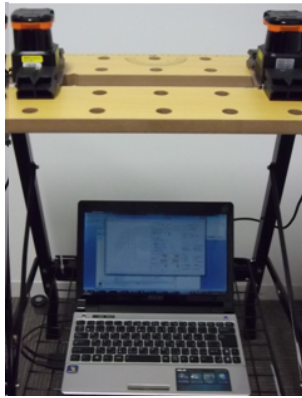


図 1 ResBe-T に使用した測域センサとネットブック

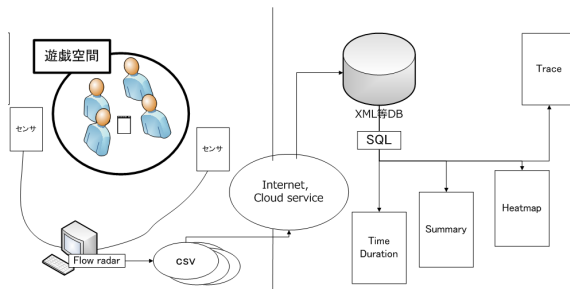


図 2 ResBe-T システム構成

を店内の各種センサで読み取ることによるスーパー内での顧客、従業員の行動把握や作業の効率化、陳列・ポップなどの配置検討に利用するもの、駅構内や空港などに設置されているデジタルサイネージコンテンツの視認率といった広告効果・広告料評価システム[14]や、椿らのデジタルサイネージに搭載された認識カメラで撮影した映像から背景差分を用いて空港内保安検査場の混雑情報を別の場所に提示するデジタルサイネージシステム[15]などが存在する。

3. ResBe-T

まず、我々が以前提案したユーザ非装着型エンタテインメント空間評価手法 (ResBe) の可視化手法である ResBe-T についての説明を行う。我々は、赤外線センサを使用する事で非装着かつ被験者に計測を意識させず評価を行い、人が訪れた事や軌跡を計測するだけではなく、人が空間内のどの位置に多く・長く滞在したかを可視化する事で、より体験空間内での人の流動や空間使用率など被験者状態の識別が可能となる。

3.1 システム構成

赤外線センサ（北陽電機株式会社製 UTM-30LX（以下測域センサ））を USB2.0 でネットブック（図 1;ASUS 社 UL20A/Windows7）に 2 台接続する。ネットブック内の北陽電機社製流動計測ソフトウェア「Flow Radar」を用いてデータを取得し csv ファイル形式として出力、Dropbox を使った遠隔ファイル同期および TeamViewer によるモニタリングを利用し遠隔監視システムとして構築した（図 2）。

3.2 可視化手法

測域センサより取得した csv データを、C# , Processing に

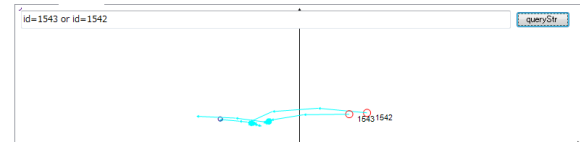


図 3 Flowline 描画

でそれぞれ作成したプログラムを用いて Flowline と HeatMap の二手法で可視化を行った。Flowline 手法は、認識したユーザの X-Y 座標、認識した時間（毎秒）からユーザの軌跡を可視化する手法で、図 3 のように、軌跡にはそれぞれ自動連番の ID が付与されているため、この付与されている ID ごとや時間ごとの軌跡検索が可能である。この Flowline 手法を用いる事で、空間内で時間ごとに発生したイベントに対しての人の動きの違いや、来場者が展示物などの解説を聞いている時に、解説者が来た事により展示物から後ずさるといった些細な動きの変化を把握できる。

一方、HeatMap は、計測空間内をメッシュで区切りその枠内に滞在が認められた場合にカウント（1 秒 1 カウント）をし、そのカウントが多くなるにつれメッシュ枠内の色が濃くなっていく可視化手法（図 4）で、時間ごとの描画も可能である。この HeatMap 手法を用いる事で、体験者がコンテンツ体験中どの程度の領域幅で動きがあったのか（体の揺れがみられたか）、その領域に何秒間いたのか、場所による空間使用率の違いを把握できる。

4. 軌跡, HeatMap を用いた展示空間の評価例

2010 年 8 月 24, 25 日に日本科学未来館にて開催された「インタラクティブ東京 2010」（以下 iTokyo）において、ResBe-T の 2 つの可視化手法を用いて、来場者の室内空間での振る舞いと展示空間の評価を公開実験として実施した。会場であるイノベーションホールは 16.0m × 10.5m の広さで、ResBe の他に 9 のインタラクティブ展示が設置されている（図 5）。来場者は、図の左側から薄暗い会場へ入退場し、自由に中を見る事ができる。図内に描画されている青い線が被験者の軌跡となっており、図では 3 分間の体験状態を可視化描画したものである。図の軌跡をみると、中心部を通り過ぎていく軌跡もみられ、必ずしも入口にある展示から見て回るわけではないという事、入口付近にある展示スペース内で体験者がよく動き回っている様子、展示ブース以外の箇所での軌跡描画ストロークが長いことから、展示ブース付近での移動より速度が速い事がわかる。なお図中にある赤い○は、オクルージョンが発生しセンサの認識が途切れた箇所となる。数秒間の遮蔽であれば復元してまた同じユーザであると認識を改めるが、今回のケースではその復元可能時間以上センサ前に人が立ってしまった、他の被験者や解説者などが遮蔽に入ってしまったなどが原因として考えられる。

次に、2 日間開催した iTokyo のうち 24 日の 16:28~16:58 の 30 分間の来場者の軌跡と空間使用率を Flowline, HeatMap を用いて描画したものを図 6 に示す。この図は、軌跡を描画した上に HeatMap を重ねて描画している。先ほどの図 5 と同じ



図 4 HeatMap 描画

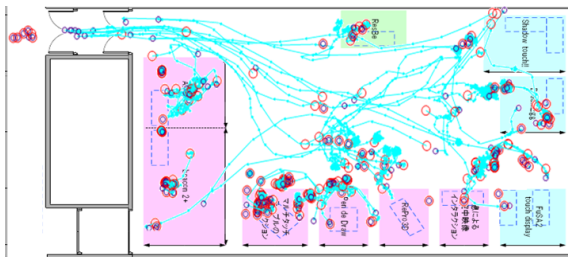


図 5 展示空間上面図+Flowline

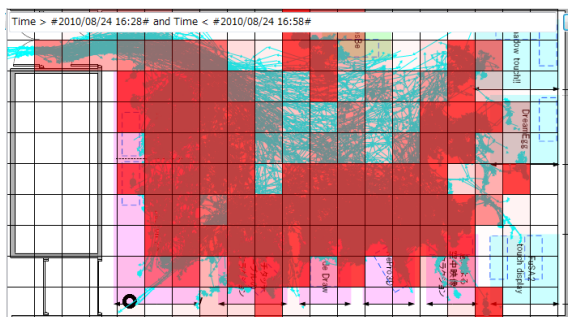


図 6 展示空間上面図+Flowline+HeatMap

計測データを使用している可視化処理であるが、時間が 10 倍と長くなっているため来場者の軌跡がとて多くなっており動きが特定しづらくなっている。しかし、HeatMap は展示ブース付近に滞在が多く、展示ブースから少し離れた空間に滞在が少ないと色の濃さではっきり出ているので空間使用率を把握するには適していると言える。色が薄かった（滞在と認められた件数が少なかった）要因としては、先程と同じく展示ブース以外の箇所での移動速度が速かったからなのではないかと考えられる。

5. Hadoop を用いたセンサデータ処理

iTokyo において取得した測域センサデータの量は少ないため、プログラム上からデータを読み込んで処理を行っており、データは全てメモリ上で扱うことができた。しかし、今後はセンサから長期間データを取得した場合の処理を行うことを視野に入れているため、HDD 上に格納されたデータに対する処理が必要となる。そこで、本稿では測域センサより取得したデータを Hadoop 上で処理することにより効率的な解析を試みる。

5.1 ResBe-T から得られるセンサーデータ

測域センサ (ResBe-T) から得られるデータは以下である。

表 1 測域センサから得られるデータの一部

Time	Num of Human	ID(1)	X(1)	Y(1)	ID(2)	X(2)	Y(2)	ID(3)	X(3)	Y(3)
1282609134	1	78	6430	8875						
1282609135	2	78	6540	8839	79	-4140	790			
1282609136	2	78	6407	8875	79	-2752	635			
1282609137	3	78	6467	8861	79	-2752	635	80	-1299	738
1282609138	2	78	6408	8880	80	-139	1284			
1282609139	3	78	6379	8888	80	1083	2280	81	-6392	3094

表 2 測域センサから得られるデータの一部 (前処理後)

Time	ID	X	Y
1282609134	78	6430	8875
1282609135	78	6540	8839
1282609135	79	-2752	635
1282609136	78	6407	8875
1282609136	79	-2752	635
1282609137	78	6467	8861
1282609137	79	-2752	635
1282609137	80	-1299	738
1282609138	78	6408	8880
1282609138	80	-139	1284
1282609139	78	6379	8888
1282609139	80	1083	2280
1282609139	81	-6392	3094

- タイムスタンプ
- センサで設定したエリア A1 に入った人数
- センサで設定したエリア A1 から出た人数
- センサで設定したエリア A2 に入った人数
- センサで設定したエリア A2 から出た人数
- センサが認識している人の数
- 認識した人間につけた連番 (ID)
- 認識した人の X 座標
- 認識した人の Y 座標
- 設定したエリア A1 内に居る人の数
- 設定したエリア A2 内に居る人の数

データの一部を表 1 に示す。ここでは今回の実験で使用した、タイムスタンプ (Time)、センサが認識している人の数 (Number of Human)、認識した人間につけた連番 (ID)、認識した人の X 座標 (X)、認識した人の Y 座標 (Y) だけを表示している。センサは新たにエリア内に侵入した人物を感知すると自動的に ID を割り当て、その人物がエリア内に居る限りは基本的には同じ ID として認識する。表の 1 行がある時間にセンサが取得したデータを示しており、認識した人物の ID、および、その人物の XY 座標を記録する。また、人物が複数居る場合、ID と XY 座標は複数登録される。

今回用いるデータは、センサが認識している人の数により 1 行の長さが異なるため、前処理として、表 2 に示すように 1 行に 1 つの ID のみの情報が記録されるようにデータフォーマットの変換を行った。

5.2 実験

実験に用いた測域データは、4 節と同様に「インタラクティブ東京 2010」において取得したデータあり、データ量は約 30Mbyte である。このデータを前節で述べた前処理を行うことで、データ量は約 80Mbyte、180 万レコードとなった。

実験で用いた Hadoop の構成は図 7 のようになっており、1

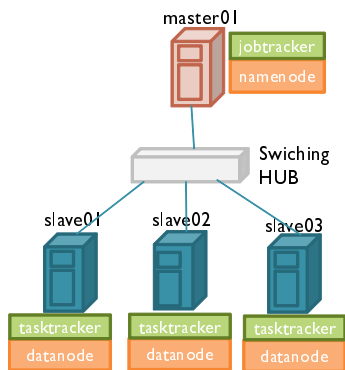


図 7 Hadoop 構成図

表 3 マシンスペック, および, Hadoop の設定

マスターノード (1 台)	OS	CentOS 6.3
	CPU	Intel Core i3-2100
	CPUコア数	2
	メモリ	8.0GB
スレーブノード (3 台)	OS	CentOS 6.3
	CPU	Intel Core i3-2120
	CPUコア数	2
	メモリ	8.0GB
Hadoop	バージョン	0.20.2
	レプリカの数	1
	各ノードのmapタスク数の最大値	2
	各ノードのReduceタスク数の最大値	2

1 台のマスターノードと 3 台のスレーブノードで構成を行った。各マシンのスペック, および, Hadoop の設定を表 3 に示す。Hadoop のレプリカ数については実験が小規模なため, 速度を重視して最小の 1 に設定した。また, map と reduce の最大値については CPU コア数と同じ値とした。

実験は前処理を行った測域データを米 Facebook 社で開発された, Hadoop 上にデータウェアハウス (DWH) を構築するための基盤ソフトである Hive 上に登録し, SQL ライクな問合せ処理を行いその処理速度の測定を行った。

5.3 実験結果

まず, データのソート処理を行い, 次に, 特定の範囲内に入ったユーザー数のカウントにかかる時間の計測を行った。また, 比較対象として既存の手法である Excel のマクロを用いた処理方法の測定も行った。データサイズと処理時間の相関を見るために, サイズは 10Mbyte から 80Mbyte まで 10M 刻みで測定を行った。なお, Excel についてはレコード数の上限の関係から 60Mbyte までの結果となる。ソート処理の結果を図 8 に示す。結果から, Hadoop においてはソートの種類と処理時間の違いはみられず, データ量が増加しても, ある一定のデータ量になるまでは処理時間に大きな差がない。一方, Excel では Hadoop と異なり, データ量が増えるほど処理時間も比例して増加していることがわかる。

次に, 特定の範囲内に入ったユーザー数のカウント処理を行った。問合せの条件として, X 座標 Y 座標の範囲を指定し, その範囲に入ったユーザの数を調べる処理を行った。今回は x 座標と Y 座標がそれぞれ 3,000 から 4,000 の範囲に入ったユーザー数のカウントを行なった。該当するユーザー数は 10Mbyte のサイズで 229 人, 40Mbyte で 1,106 人, 60Mbyte で 1,719 人である。結果を図 9 に示す。結果から, Hadoop は処理時間が非常に短く, データ量が増えても処理時間に大きな差は見られな

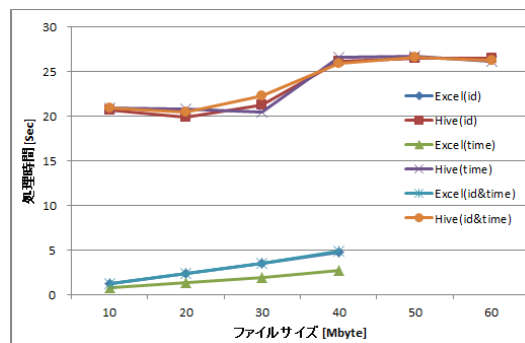


図 8 ソート処理に関する実験結果

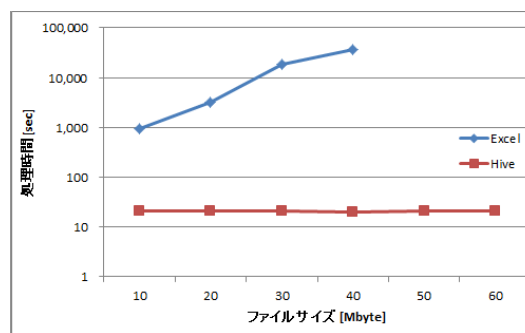


図 9 特定の範囲内に入ったユーザー数のカウントに関する実験結果

いことがわかる。一方, Excel は処理に時間がかかり, データ量が増えれば増えるほど膨大な処理時間がかかることがわかる。

5.4 おわりに

本稿では, ユーザ非装着型エンタテインメント空間評価手法 ResBe(Remote entertainment space Behavior evaluation) の研究において, ユーザの特徴的な行動抽出を効率良く行うことを目指し, Hadoop, および, Hive を用いたデータ処理法の検討を行った。今後は mysql など既存の DBMS との比較実験を行い考察を行っていきたい。

謝 辞

本研究の一部は科学研究費補助金 若手研究 (B)(課題番号 24700102) の助成による。

文 献

- [1] 「共生社会に向けた人間調和型情報技術の構築」: <http://www.ambient.jst.go.jp/>
- [2] 白井暁彦: エンタテインメントシステム, 芸術科学会論誌, Vol. 3, No. 1, pp. 22-34, 2004.
- [3] 「交通広告ナビ」: <http://www.koutsu-navi.com/search/index.html>
- [4] 岩橋翔仁, 荒原一成, 周立, 白井暁彦: ResBe:エンタテインメントシステム周囲のコミュニケーション場に対する遠隔評価手法の提案, 日本バーチャルリアリティ学会大会論文集, Vol. 15, 2010.
- [5] 白井暁彦, 岩橋翔仁, 瀬口慎人, 鈴木真一朗, 長谷川晶一: 測域センサを用いた ResBe システムと Heatmap による実世界指向エンタテインメントシステムの物理評価手法, エンタテインメントコンピューティング 2010 予稿集, 2010.
- [6] 里雄二, 北原格, 中村裕一, 大田友一: 複合コミュニティ空間における注目の共有～指示動作による注目の強調提示システム～, 日本バーチャルリアリティ学会第 6 回大会論文集, pp. 235-238, 2001.
- [7] 上岡玲子, 広田光一, 廣瀬通孝: 体験記録装置としてのウェアラブルコンピュータの研究, 日本バーチャルリアリティ学会第 6 回大会論文集, pp. 149-152, 2001.

- [8] 江田政聡, 中根傑, 横山昌平, 福田直樹, 石川博: 省エネルギー化を目指したセンサデータマイニング手法の検討, DEIM Forum (2010).
- [9] 江田政聡, 中根傑, 横山昌平, 福田直樹, 峰野博史, 石川博: センサルームにおける赤外線センサを用いた人の移動・在席状況の推定と利用, DEIM Forum (2011).
- [10] 中根傑, 江田政聡, 横山昌平, 福田直樹, 石川博: センサネットワークにおける大規模な可視化システムの開発, DEIM Forum (2010).
- [11] 森武俊: 「センシングルーム」における行動蓄積とパターン発見, 情報処理学会研究報告 (2005).
- [12] 船津聡, 齋藤豪, 中嶋正之, “自律的エージェントのための確率場に基づく動作クラスを用いた動作生成”, 電子情報通信学会大会講演論文集, 2005, 263 特殊号: 情報・システム 2. 2005.
- [13] 「IMADE-PROJECT」: <http://www.ii.ist.i.kyoto-u.ac.jp/IMADE/>
- [14] 吉田浩, 宮原浩二, 椿泰範, 米沢みどり, 高梨郁子: 「交通向けデジタルサイネージシステム」, 映像情報メディア学会年次大会 (2011) .
- [15] 椿泰範, 吉田浩, 宮原浩二, 米沢みどり, 高梨郁子: 「交通向けデジタルサイネージシステム-カメラ付端末の複数台連携によるコンテンツ制御-」, 映像情報メディア学会年次大会 (2011) .
- [16] FUJIMURA Wataru, IWDATE Shoto, SHIRAI Akihiko: CartooNect: Sensory motor playing system using cartoon actions, Proceedings of Virtual Reality International Conference (VRIC 2011), 6-8 April 2011, Laval, France.
- [17] Misumi HAJIME, Wataru FUJIMURA, Takayuki KOSAKA, Motofumi HATTORI, Akihiko SHIRAI: Development of serious game which use full body interaction and accumulated motion, NICOGRAPH International 2011, 10-11 June 2011, Kanagawa, Japan.