

大規模人間行動センシングコーパス HASC2012corpusの概要とその応用

河 口 信 夫^{†1} 渡 辺 穂 高^{†1} 楊 天 輝^{†1}
小 川 延 宏^{†1} 岩 崎 陽 平^{†1} 梶 克 彦^{†1}
寺 田 努^{†2} 村 尾 和 哉^{†2} 羽 田 久 一^{†3}
井 上 創 造^{†4} 川 原 圭 博^{†5}
角 康 之^{†6} 西 尾 信 彦^{†7}

人間行動の理解は、情報システムの究極の目標の一つである。我々は装着型センサを用いた人間行動の大規模コーパスの構築を行うために、HASC Challenge を 2010 年,2011 年と開催した。これらの成果として、HASC2011corpus(被験者数 116, 行動データ数 4898), HASC2012corpus(被験者数 136, 行動データ数 7668) を構築した。本稿では、HASC2012corpus の概要と、新たな実環境データの応用について述べる。

HASC2012corpus: Human Activity Corpus and Its Application

NOBUO KAWAGUCHI,^{†1} HODAKA WATANABE,^{†1}
TIANHUI YANG,^{†1} NOBUHIRO OGAWA,^{†1}
YOHEI IWASAKI,^{†1} KATSUHIKO KAJI,^{†1}
TSUTOMU TERADA,^{†2} KAZUYA MURAO,^{†2}
HISAKAZU HADA,^{†3} SOZO INOUE,^{†4}
YOSHIHIRO KAWAHARA,^{†5} YASUYUKI SUMI^{†6}
and NOBUHIRO NISHIO^{†7}

Understanding of human activity through wearable sensors is one of the ultimate objective of human-oriented computing. To overcome the situation, we have collected large scale human activity corpus, named HASC2011corpus(116 subjects, 4898 data files), and HASC2012corpus (146 subjects, 7668 data files). In this paper, we describe about HASC2012corpus and application of its real world commuting data.

1. はじめに

人間行動の理解は、情報システムの究極の目標の一つである。行動認識技術の高度化には大規模な行動コーパスの存在が大変重要であるにも関わらず、これまで研究者が利用できるパブリックなコーパスは存在しなかった。そこで我々は、大規模データ収集のために、HASC Challenge と名付けた技術チャレンジを 2010 年 2011 年に開催した (HC2010,2011)。HASC Challenge は、データやツールの共有と、センサ情報の共有・利用のノウハウ、行動認識に関する知見の向上を目的とした技術チャレンジであり、複数の参加者が協力して行動センシング情報の収集を行った。これらの成果として我々は HASC2011corpus (被験者数 116, 行動データ数 4898), HASC2012corpus(被験者数 136, 行動データ数 7668) を構築した。HASC2011corpus は、利用許諾に同意すれば、登録者は誰でも利用できる。また、HC2011 の参加者と同等のデータを提出して頂ければ、HASC2012corpus も利用可能である。

HC2011 では、従来の 6 種類の基本的な行動 (歩く、走る、階段昇降、スキップ、停止) の収集に加えて、実環境の経路移動における行動センシング情報の収集を行った。経路移動行動は、駅や店舗、オフィスといったランドマーク間を移動する行動である。実環境データであるため、基本的な行動に加え、エスカレータやエレベータなどの情報や、ドアの開閉といったイベントの情報のラベル付けも行った。これらの情報を活用した行動認識技術の新たな応用が期待される。結果として HASC2012corpus には、40 の経路が収録されている。

本稿では、HASC2012corpus の経路移動行動データの応用事例についても紹介する。一般に、ある場所から始まる移動経路は複数存在する。しかし、無限に存在するわけではな

^{†1} 名古屋大学大学院工学研究科, Graduate School of Engineering, Nagoya University

^{†2} 神戸大学大学院工学研究科, Graduate School of Engineering, Kobe University

^{†3} 東京工科大学メディア学部, School of Media Science, Tokyo University of Technology

^{†4} 九州工業大学工学研究科基礎科学研究系,
Graduate School of Engineering, Kyusyu Institute of Technology

^{†5} 東京大学大学院情報理工学系研究科,
Graduate School of Information Science and Technology, University of Tokyo

^{†6} 公立はこだて未来大学システム情報科学部,
School of Systems Information Science, Future University Hakodate

^{†7} 立命館大学総合理工学院・情報理工学部,
College of Information Science & Engineering, Ritumeikan University

い．事前に初期位置と利用する経路の知識があり，現在の経路情報との間で類似性が計算できれば，どの経路を通過したかを識別することが可能になる．HASC2012corpusには，同じ経路の情報が記録されているわけではないが，経路に手で付与されたラベルと，行動認識技術を用いて自動識別した行動ラベルとの間で類似性が計算できるかを確認した．DP マッチングを用い，また距離やラベルの不一致に関するコストを定めて類似性を計算した．行動認識率が高かった 20 種の経路を利用し，適切なパラメータを付与すると，上位 3 位までに当該経路が 90%の確率で含まれることを確認した．これにより，行動認識技術を用いて，加速度センサだけで経路識別を実現できる可能性を示した．

以下，2 章では，これまでの HASC Challenge の経緯を簡単に述べる．3 章では HASC2011corpus，HASC2012corpus の概要について述べ，4 章では HASC2012corpus における実世界データを活用した経路推定と，追加実験の内容について述べ，5 章でまとめる．

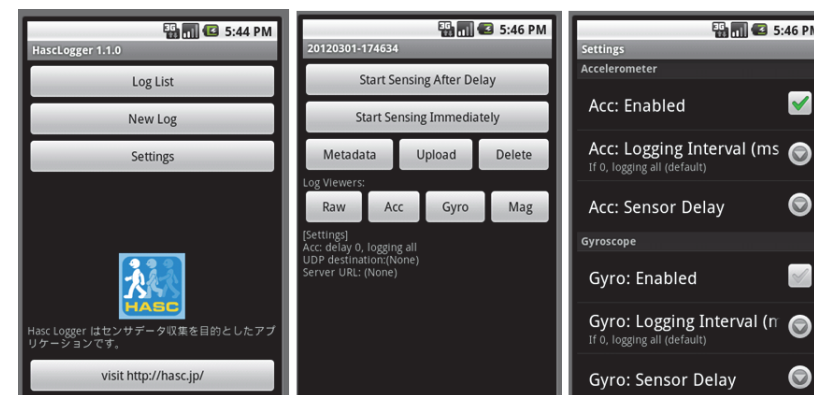


図 2 Android 用の HASC Logger
Fig. 2 HASC Logger for Android



図 1 iPhone/iPod Touch 用の HASC Logger
Fig. 1 HASC Logger for iPhone/iPod Touch

2. HASC Challenge 2010/2011

音声処理，画像処理，自然言語処理，といった分野においては，システムの性能向上に，大規模なデータに基づく学習やモデル構築の重要性が高いことは広く知られている．しかし，行動処理，行動理解という分野では，まだ十分な規模のデータベースは整備されておらず，個々の研究も小規模にならざるを得ない．

我々は，HASC Challenge という枠組みを構築し，2010 年より共有して利用できるデータベースの構築を進めている¹⁾．HASC Challenge 2010(HC2010) は 2010 年 8 月にアナウンスを行い，11 月を締切りとして参加者からデータを収集した．HC2010 ではデータ収集に 24 チームが参加登録し，21 チームがデータを提供した³⁾．また，HASC Challenge 2011(HC2011) は 2011 年 8 月にアナウンスを行い，12 月をデータ締切りとした．HC2011 では，29 チームが参加登録し，19 チームがデータを提供した．最終的に構築したコーパスについては，次章で説明する．

2.1 HC2011 における改善事項

HC2011 では，HC2010 の反省を生かし，いくつかの改善を行った．まず，センサ取付位置についてはより詳細なラベル付けを依頼した．また，シーケンスデータの時間を 2 分から 5 分へ延長し，各行動の継続時間も 5 秒から 10 秒に延長した．また，HC2010 では，センサ種別は加速度センサのみの収集であったが，HC2011 では収集可能なセンサ情報はすべて

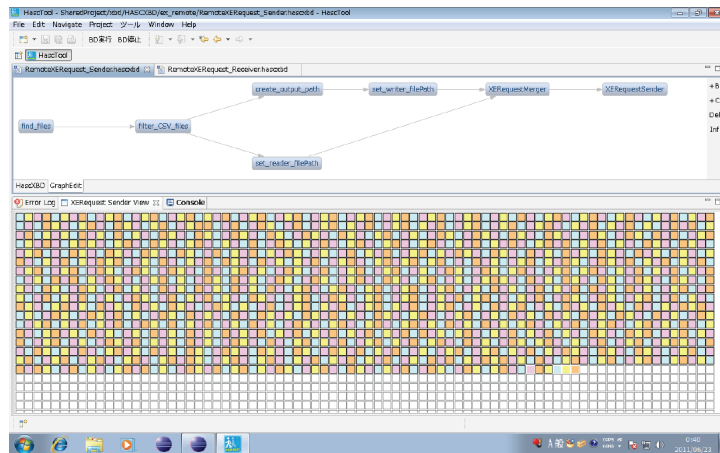


図 3 HASC Tool 上での分散データ処理の可視化

Fig. 3 Visualization of Distributed Data Processing on HASC Tool

集める、ということを行った。さらに、あらたに実世界データとしてランドマーク間の移動データを収集することとした。

参加者の増加を目指すために、チュートリアルを名古屋・東京・大阪で開催した。参加登録が伸びたのは、チュートリアルの効果だと考えている。さらに、HC2010 ではメールベースでデータ収集・登録を行ったのに対し、HC2011 では、専用の Web システムを開発し、チュートリアルの登録から、データの投稿、修正、配布までを一貫して行うことを可能にした。これによって、事務局側の手間が大幅に削減された。

2.2 メタデータ、ラベルの拡張

メタデータについては、特に端末の位置の詳細化を行った。HC2010 では、メタデータの規定は充分に行わずにデータ収集を行い、各チームの記載を参考に標準を定めた。HC2011 では、HC2010 の経験を生かし、さらに詳細な端末位置（体に対し、前後左右上下）の情報を記入するようにした。また、実世界データにおいては、天気の情報や経路の開始点、終了点の場所情報を追加した。

ラベルデータとしては、実世界データのラベル付与を行うための拡張を行った。特に大きな追加はイベント情報の追加である。ドアの開閉、エレベータ、自動販売機、屋内外の移動、改札などの情報を追加できるようにしている。

表 1 HASC2011corpus の基礎情報
Table 1 Basic Information of HASC2011corpus

性別	被験者数	ファイル数
男性	102	4464
女性	14	434
合計	116	4898

2.3 HASC Logger の拡張

実世界データを効率的に収集するために、HASC Logger(図 1, 図 2) の拡張を行った。まず、複数のセンサデータを選択的に取得できる拡張を行った。さらに、音声データを同時に録音できる機能を追加した。これにより、実世界データを収集する際に、被験者が喋りながら行動すれば、後からのラベル付与の負担を軽減することができる。

2.4 HASC Tool の拡張

行動情報処理の高度化を行うためには、共通して利用可能なツールの存在も重要である。我々はオープンソースで開発を進めている HASC Tool⁸⁾ の拡張も進めている。これまでの主な改良・拡張点を以下に挙げる。

- (1) データ処理が複数スレッド、及び、複数デバイス間で分散実行可能 (図 3)
- (2) 音声再生機能を付加。これにより、HASC Logger で音声と同時に収集したデータのラベル付が容易になる。
- (3) イベントラベルの付与機能。実世界データでは、時間区間のラベルに加え、ドアの開閉といったラベルを付加している。その可視化機能を追加した。
- (4) ラベルの可視化の際の色指定機能。ラベルへの色の付与はこれまでは自動で行われていたが、存在するラベルによって色が異なるのは良くないため、色指定を可能にした。

今後も HASC Logger, HASC Tool の改良を続け、研究者にとって利用しやすい環境の構築を目指したい。

3. HASC2011corpus, HASC2012corpus の概要

HC2010 の成果から、性別等のメタデータが不明なデータを削除し、さらに、我々で新たに収集した 20 人分のデータを加えたものを HASC2011corpus⁶⁾ とした (表 1)。HASC2011corpus は、現在 Hub HASC⁷⁾ にユーザ登録をしてもらえれば、フリーでダウンロード・利用が可能である。HASC2011corpus を利用して、被験者数の増大が認識率

表 2 HASC2012corpus の基礎情報
Table 2 Basic Information of HASC2012corpus

種別	被験者数	センサ種別	ファイル数
従来データ	96(男性 85, 女性 11)	加速度	4495
		角速度	2521
		GPS	240
		地磁気	310
実世界データ	40(男性 34, 女性 6)	加速度	40
		角速度	30
		GPS	27
		地磁気	5
合計	136		7668

表 3 HASC2012corpus に含まれるセンサデバイス
Table 3 Sensor Devices in HASC2012corpus

メーカー/デバイス名	ファイル数	割合 (%)
ATR/WAA-00X	1854	40.9
Apple/iPhone, iPodTouch	1759	38.8
Samsung	662	14.6
HTC	118	2.6
Sharp	95	2.1
Sony	32	0.7
LG	15	0.3
合計	4535	100

の向上に有用である事などが示されている⁴⁾⁻⁶⁾。

一方、HC2011 の成果から、HASC2012corpus の整備を行った (表 2) . HASC2012corpus は現在、HC2011 の参加者のみを対象に公開されているが、今年度開催予定の HASC Challenge2012 の参加者には公開予定である。HASC2012corpus では、複数のセンサ種別を収集した。デバイス毎に利用できるセンサが異なるため、表に示すように、4 種類 (加速度、角速度、GPS、地磁気) のセンサのファイル数は異なっている。

また、HC2011 から新たに追加した実世界データについては 40 の経路データ (合計時間約 390 分) を収集することができた。これまでの人工的な行動シーケンスデータと比較して、実環境で取得されたデータであり、ラベル付の精度やイベントの付与などの新しい問題も生じているが、多様な利用が期待できる。我々が行った応用の検討については、次章で説明する。

HC2011 では、多様なセンサデバイスが利用されている。表 3 は、HASC2012corpus に含まれるデバイス種別を各計測データのメタデータから取得したものである (従来、実世界の両方の合計) 。

このように、本コーパスは人の多様性だけでなく、センサデバイスに多様性がある点にも特徴がある。実際、異なる加速度センサデバイス間では、同じ場所でも、重力値を異なる値で返したり、ノイズの発生の仕方が異なり、分散が違うことなどが確認できている。

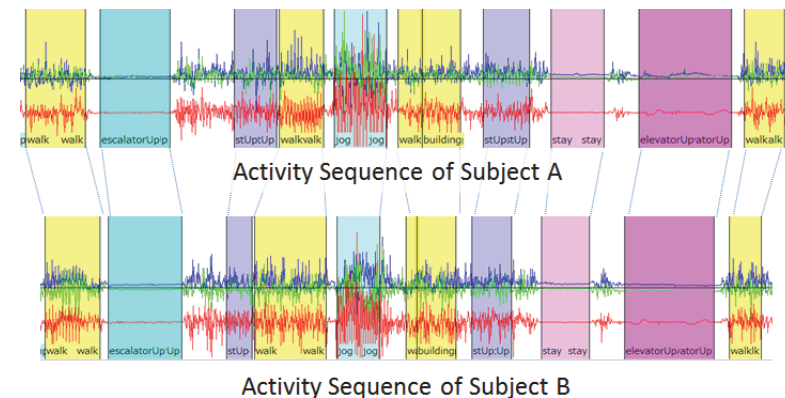
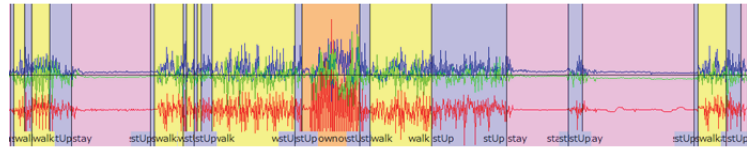


図 4 異なる被験者による同一経路の行動シーケンス
Fig. 4 Activity Sequence of Same Route

4. 実世界データの応用：類似性に基づく経路推定

HC2011 では、実世界データとして、ランドマーク間の移動を収集した。我々は、このデータを用いてどのような研究が行えるかの検討を行った。

図 4 は、同一の経路を異なる被験者 A, B が移動した際の人手によるラベル付けの結果を示している。同じ経路であるため、階段やエスカレータといった部分が共通であることが確



Recognition Result of Subject B's Activity

図 5 行動認識によるラベル付結果

Fig. 5 Labelling by Automatic Activity Recognition

認できる．すなわち，経路間のラベルの順序が時間は異なるが同一であることがわかる．もちろん，エスカレータを使わずに階段を利用したり，同じように歩くのではなく，すこし駆け足で移動するといったこともありうるため，同じ経路であるからといって行動系列が常に同じになるとは言い難いが，ある程度の類似性が存在することは期待できる．

我々は行動データの自動認識を行い，経路の類似性を評価できるかの実験を行うこととした．HASC2012corpus の実世界データ収集では，ユーザの経路を指定していないため，共通の経路を収集したものは多く存在しない．そこで，手でラベル付されたデータと自動推定されたデータ間の類似性を比較することを試みることにした．

図 5 は，被験者 B の行動を自動識別させた結果である．図 4 と異なり，各行動の間に隙間が存在しない．これは，現在の認識手法が必ず何らかのラベルを付与する仕組みになっているからである．このデータ間のマッチングは DP マッチングを利用することとした¹²⁾．図 6 に，2 つの経路間のマッチングの様子を示す．マッチングの際には，移動コストの計算を認識率に比例して行った．今回は，HASC2011corpus を用いて学習した行動認識モデルを利用し，40 経路中から比較的行動認識率の高かった 20 経路について，手でつけたラベルと，自動認識したラベルとの類似性を計算した．表 4 に示すように，3 位以内には 90% の割合で当該経路が含まれることが確認できた．この結果から，経路の類似性を行動認識のラベルを用いて判断できる可能性が示せた．

我々はさらに簡単な追加実験を行った．5 人の被験者に 3 種類の経路を移動してその行動データを計測した．今回は人手でのラベリングは行わず，自動認識のみでラベル付を行い，そのラベル系列間の距離を計算した．図 7 は，各被験者・経路の相互の距離を DP マッチングで計算し，図示したものである．被験者が異なっても，同じ経路の行動系列間が比較的近くなっていることが読み取れる．

table	stay	stay	walk	stay	stay	stay	walk	walk	stUp	stDown	stay	stUp	stay	stay
stay	1	32	154	185	216	247	369	491	613	735	766	888	919	950
walk	45	41	34	103	172	241	249	280	349	418	487	556	625	694
walk	89	85	40	74	143	212	243	251	320	389	458	527	596	665
stay	95	91	137	42	73	104	226	336	344	413	391	513	529	560
stay	101	97	184	48	44	75	197	319	429	437	397	484	515	531
stay	107	103	190	54	50	46	168	290	412	522	403	490	486	517
walk	151	147	105	98	94	90	48	79	148	217	286	355	424	493
walk	195	191	111	142	138	134	54	50	119	188	257	326	395	464
walk	239	235	117	151	182	178	60	56	90	159	228	297	366	435
walk	283	279	123	157	191	222	66	62	96	130	199	268	337	406
walk	327	323	129	163	197	231	72	68	102	136	170	239	308	377
walk	371	367	135	169	203	237	78	74	108	142	176	210	279	348
walk	415	411	141	175	209	243	84	80	114	148	182	216	250	319
walk	459	455	147	181	215	249	90	86	120	154	188	222	256	290
stay	465	461	244	149	180	211	187	183	179	213	156	278	224	255
walk	509	505	250	193	189	220	193	189	223	219	200	196	265	264
walk	553	549	256	237	233	229	199	195	229	263	244	240	236	305
stay	559	555	353	243	239	235	296	292	288	322	250	337	242	238
walk	603	599	359	287	283	279	237	268	332	328	294	290	286	282
stUp	653	649	409	337	333	329	287	283	270	345	344	296	336	332
stUp	703	699	459	387	383	379	337	333	276	316	391	302	342	382
stay	709	705	556	393	389	385	434	430	373	369	318	399	304	335
stay	715	711	653	399	395	391	478	527	470	466	324	411	310	306
stay	721	717	750	405	401	397	484	571	567	563	330	417	316	312
stay	727	723	810	411	407	403	490	577	664	660	336	423	322	318
walk	771	767	725	455	451	447	405	436	505	574	380	376	366	362

図 6 DP マッチングによる経路推定の例

Fig. 6 Example of Activity Sequence DP-matching

表 4 類似経路の計算結果

Table 4 Results of Similar Route

順位	1 位	1~3 位	1~5 位
割合	60%	90%	95%

この図からも、行動認識を利用した経路推定が行える可能性が示された。今回の実験は、まだ基礎的なものであるが、このような可能性を低いコストで確認するために、HASC2012corpusの示す役割は大きい。経路間の距離がラベルのマッチングで計算できるかどうか、という仮説を検証するために、データ収集を行うことなく実験ができた。ある程度必要な事項が確認できた後、実際のデータを収集して、さらに仮説の検証を行うことができる。今回は、行動データから行動の系列を自動的に認識し、その距離を用いて系列間の距離を計算できることを示した。この成果からは、様々な応用の可能性が期待できる。

	A-1	B-1	C-1	D-1	E-1	A-2	B-2	C-2	D-2	E-2	A-3	B-3	C-3	D-3	E-3
A-1	2	7	20	25	16	24	27	20	30	29	58	30	30	39	39
B-1	9	2	18	26	24	27	38	20	26	30	54	29	30	55	42
C-1	30	29	1	45	31	16	19	15	17	21	36	34	23	30	31
D-1	12	7	19	2	19	27	35	20	23	25	65	33	32	54	55
E-1	24	23	25	26	2	28	34	22	24	31	58	30	30	40	42
A-2	68	58	34	100	69	1	10	24	34	10	38	35	35	35	40
B-2	64	58	32	71	62	8	1	46	35	20	25	28	35	36	31
C-2	49	51	25	71	42	26	28	1	26	26	35	38	33	36	40
D-2	71	60	27	95	48	9	10	20	1	8	25	33	29	32	37
E-2	77	62	38	99	75	31	10	30	32	1	35	37	32	37	44
A-3	74	69	46	80	92	40	39	38	43	40	1	17	20	14	24
B-3	57	55	45	75	70	28	26	38	40	19	8	1	25	10	23
C-3	57	59	44	70	70	28	27	37	38	21	10	24	1	10	14
D-3	52	56	52	70	72	41	40	39	40	37	7	21	19	1	20
E-3	63	58	44	74	66	39	31	39	47	29	7	15	16	9	1

図 7 5 被験者,3 経路による経路間 DP マッチング
Fig.7 Route DP-matching of 5 subjects and 3 routes

5. ま と め

HASC Challenge 2010, 2011 を通じて、装着型センサによる大規模行動コーパス HASC2011corpus, HASC2012corpus を構築し、のべ被験者 250 名以上、データファイル数 12000 以上の収集を達成した。また、実世界データの応用として、経路推定の可能性を検証し、基礎的な実験によってその実現可能性を示した。

HASC Corpus をより有用なコーパスとするためには、被験者数を伸ばすだけでなく、年齢や身長・体重といったパラメータに対してもバランスの良いデータ収集が必要となる。

また、よりアプリケーションを意識したデータ収集も必要と考えている。

今年度も HASC Challenge2012 を開催する予定である。多くの参加者を期待している。皆様のご協力を願いたい。

参 考 文 献

- 1) 河口信夫. 人間行動理解のための装着型センサによる大規模データベース構築にむけて. 情報処理学会マルチメディア分散協調とモバイルシンポジウム (DICOMO 2010), pp.579-581(2010).
- 2) Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio, HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings. Proc. of ACM Augmented Human 2011, pp.27:1-27:5(2011).
- 3) 河口信夫, 小川延宏, 岩崎陽平, 梶 克彦, 寺田 努, 村尾和哉, 井上創造, 川原圭博, 角 康之, 西尾信彦, HASC Challenge2010 :人間行動理解のための装着型加速度センサデータコーパスの構築, 情報処理学会マルチメディア分散協調とモバイルシンポジウム (DICOMO 2011),pp.69-75(2011).
- 4) 小川延宏, 梶克彦, 河口信夫, HASC2010corpus を用いた被験者数と人間行動認識率の相関分析, 情報処理学会 マルチメディア分散協調とモバイルシンポジウム (DICOMO 2011), pp.76-82(2011).
- 5) Nobuhiro Ogawa, Katsuhiko Kaji, and Nobo Kawaguchi, Effects of Number of Subjects on Activity Recognition - Findings from HASC2010corpus. Proc of 1st International Workshop on Frontiers in Activity Recognition using Pervasive Sensing (IWFAR2011), pp.48-51(2011).
- 6) Nobuo Kawaguchi, Ying Yang, Tianhui Yang, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, Nobuhiko Nishio, HASC2011corpus: Towards the Common Ground of Human Activity Recognition, Proceedings of the 13th International Conference on Ubiquitous Computing(UbiComp2011), pp.571-572(2011).
- 7) HASC corpus portal : <http://hub.hasc.jp/> (2012).
- 8) HASC Tool:SourceForge : <http://sourceforge.jp/projects/hasc> (2012).
- 9) Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Distributed Human Activity Data Processing Using HASC Tool, Proceedings of the 13th International Conference on Ubiquitous Computing(UbiComp2011), pp.603-604(2011).
- 10) 袁 穎, 楊 天輝, 小川 延宏, 梶 克彦, 河口 信夫, HASC2011corpus を用いたサンプリング周波数と人間行動認識率の相関分析, 平成 23 年度 電気関係学会東海支部連合大会, (2011).

- 11) 渡辺穂高, 梶克彦, 河口信夫, 加速度・角速度を用いたセンサデバイスの装着位置・方向推定手法, 情報処理学会第 74 回全国大会, 6X-6(2012).
- 12) 小川延宏, 梶克彦, 河口信夫, 装着型センサを活用した行動情報によるユーザ移動経路推定, 情報処理学会第 74 回全国大会, 6X-7(2012).
- 13) Nobuo Kawaguchi, Hodaka Watanabe, Tianhui Yang, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Hisakazu Hada, Sozo Inoue, Yasuyuki Sumi, Yoshihiro Kawahara, Nobuhiko Nishio, HASC2012corpus: Large Scale Human Activity Corpus and Its Application, Proceedings of the Second International Workshop of Mobile Sensing: From Smartphones and Wearables to Big Data(Held with IPSN2012 and CPSWeek 2012),pp.10-14(2012).