

建築分野におけるローコストモーションセンサの利活用に関する研究 その4 骨格位置による動作判別ルール

○林田和人*¹ 遠田敦*²
渡辺仁史*³

キーワード：モーションセンサ KINECT センシング 動作 骨格 判別

1. はじめに

センサを搭載した製品が大量に普及したことでセンサ自体が安価になり、人間の動作計測に有効なさまざまなセンサが開発されている。また、通信技術の進歩により、センシングされたデータはサーバリアルタイムに送信され、大量のデータを蓄積することが可能となった。このような中、身体の動きを感知しそれを反映させるテレビゲーム機^{注1)}が発売された。ゲーム機の身体の動きを感知する部分である KINECT^{注2)}は、動作する人間との距離を識別する映像を撮影し、これをもとに骨格の位置を類推する仕組みで、発売当初よりゲーム機の付属品としてだけでなく、動作検出センサとして活用されている。

このローコストモーションセンサは、人間の骨格の位置をほぼリアルタイムに取得し人間の動作を判別できるため、これまで使われてきた高価で、またキャリブレーションの手間があったモーションキャプチャシステムの代替としての可能性がある。また、非常時を判断するため身体に加速度センサを装着し転倒などの動作を識別するシステムに対しても、ローコストモーションセンサを使えば日常的な動作を撮影するだけで動作を判別し、人間の動作に合わせたさまざまなサービスが可能になる。さらに、モーションセンサが安価であれば空間に多数設置できるため、リアルタイムに、そして広範囲での人間の動作状態が取得可能となる。

2. 研究目的

ローコストモーションセンサである KINECT^{注2)}を用いて日常動作における骨格位置の変位を記録し、記録したデータをニューラルネットワークで学習させ、その学習データをもとに動作を予測することで、動作判別ルールの可能性を探ることが本研究の目的である。

3. KINECT による骨格の位置座標の記録

3.1 日常動作の種類

動作に関する資料をもとに筆者らの協議により、日常生活において骨格の位置が変位する 57 の動作を抽出した。そして、表 1 のように身体の一部のみが動く「単純仕草」、日常生活で見られる「単純な動作」、椅子と床での「立ち

座り」の 3 つの動作グループに分けた。

表 1 実験動作

単純仕草	単純な動作	立ち座り
1右へ90°	26床のダンボールを	38その場で立つ
2左へ90°	押す 27椅子を	39その場で座る
3前へ90°	28机を	40立って前へ
4後へ90°	29床のダンボールを	41前から座る
5右手を横へ肩の高さまで	引く 30椅子を	42立って右へ
6左手を横へ肩の高さまで	31机を	43右から座る
7両手を横へ肩の高さまで	置く 32床に	44立って左へ
8右手を正面へ肩の高さまで	33椅子に	45左から座る
9左を正面へ肩の高さまで	34机に	46立って後ろへ
10両手を正面へ肩の高さまで	取る 35床から	47後ろから座る
11右手を正面から真上へ	36椅子から	48その場で立つ
12左手を正面から真上へ	37机から	49その場で座る
13両手を正面から真上へ		50立って前へ
14右手だけを前後に振る		51前から座る
15左手だけを前後に振る		52立って右へ
16両手を前後に振る		53右から座る
17前へ45° 曲げる		54立って左へ
18前へ90度曲げる		55左から座る
19右へ45° 捻る		56立って後ろへ
20左へ45° 捻る		57後ろから座る
21右へ90° 捻る		
22左へ90° 捻る		
23右足を引き上げる		
24左足を引き上げる		
25両足を上げて飛ぶ		

3.2 実験概要

実験日時：2012年3月23日(金)

実験場所：縦横 10m 程度の会議室

被験者：12名(20代男性5名、20代女性5名、30代男性1名、50代男性1名)

実験機材：KINECT2台、計測用PC2台

図1の計測セットを二箇所(椅子と床)に設置し、それぞれで「単純仕草」、「単純な動作」、「立ち座り」の実験を行った。KINECTは高さ61cmの台の上に設置し、そこから正面方向に3m離れた位置で被験者に動作をしてもらい骨格の位置座標を取得した。

「単純仕草」と「単純な動作」の場合は「気をつけ」を初期姿勢として動き始め、最終姿勢はそれぞれの動作、「取る」場合のみ最終姿勢を「気をつけ」とした。また、「立ち」の場合は椅子座か床座を初期姿勢として最終姿勢はそれぞれの動作、「座り」の場合はその場や三歩各方向へ離れた位置から椅子座か床座を最終姿勢として動作してもらった。押す、引くでは、W39D19H18cmの軽いダンボ

ール、椅子は高さ 39cm、机は 72cm で、置く、取るでは B5 サイズ厚さ 3cm の書籍を使用した。

計測は、PrimeSense 社の NITE ライブラリーを用いて骨格の位置をトラッキングするプログラムを作成し、動作開始から終了まで毎秒 30 フレームで骨格の位置座標を取得した。今回用いたライブラリーによる骨格の位置は、図 2 の計 15 箇所である。なお表 1 は被験者にとって、図 2 は計測側にとっての左右の向きである。

4. 動作の指標と判別手法

4.1 動作の指標

それぞれ動作の開始から終了までの間、1/30 秒ごとに 15 の骨格データの x,y,z の座標を取得した。被験者、動作の違いによっては、動作開始から終了までの時間が異なり取得する骨格データのフレーム数が異なる (図 3) ため、これを同列に比較できない。そこで、一つの骨格データだけになるように、以下のように変換した。

- ・累計距離 (図 4) : 動作開始から動作終了までのフレームごとに取得された骨格の位置座標から、各フレーム間の移動距離を算出し累計した
- ・最遠距離 (図 4) : 動作開始時の骨格の位置座標と、動作開始から終了までのフレームごとに取得された骨格の位置座標との距離を求め、その中の最大距離

骨格同士の関係を示す角度や距離ではなく、骨格自身の移動距離としたのは、ある動作の判別では骨格同士の関係が有効だとしても、他の動作ではそれが必ずしも有効であるとは限らないので、このような矛盾を避けるため極力単純なデータとした。

4.2 動作の判別手法

取得した 57 動作それぞれに対する 12 人分のデータを、図 5 のように 6 人分のデータを学習用データとしてニューラルネットワークを用いて学習させ、残りの 6 人分を予測用データとして予測し、学習させたニューラルネットワークの判別性能を検証した。ニューラルネットワークには、そのルールに意味付けができないなど問題はあるが、ある部分では意味付けのできるルールで判別できたとしても、他の部分では矛盾を起こすと考え、ルール自体がブラックボックスであるニューラルネットワークを採用した。

具体的な手順は、まず判別したい動作の 12 人分のデータを、乱数によりそれぞれ 6 人分の学習用データと予測用データに分け、学習用データをニューラルネットワークに入力し予測するためのルールを構築し、予測用データをそのルールを用いて予測し、動作を判別できたかどうかを判定した。学習用データと予測用データの作成に偏りが起きないようにこれを 1 万回試行し、その平均値で評価した。データは、前述の「累計距離」、「最遠距離」と、各被験者の動作ごとに標準化したものの計 4 種類を使用した。

図 6 以降はニューラルネットワークによる判別の結果で、

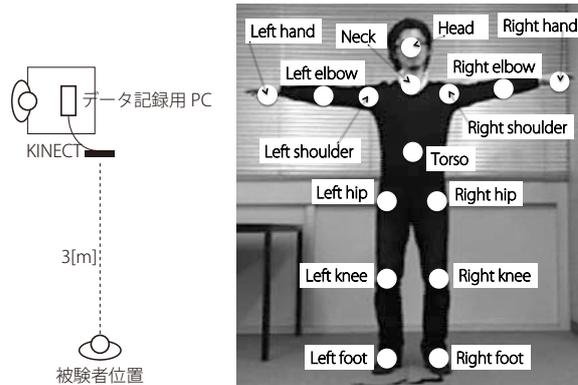


図 1 実験空間

図 2 骨格の位置

上段の左は「累計距離」、右は「累計距離を標準化」、下段の左は「最遠距離」、右は「最遠距離を標準化」を使用した結果である。予測する動作は行方向で、判別された結果が列方向となり、1 万回の試行結果を累計して試行回数で除し各セルの結果を確率として示し、「正解率」は、1 万回の試行ごとの正解率を算出し平均したものである。なお、対角線上のセルが正しく動作を判別した部分である。

ニューラルネットワークによる学習と予測は、言語環境である R の nnet ライブラリーを用いた。nnet 関数のパラメータは以下のとおりで、今回の目的がニューラルネットワークでの判別の可能性を探るためであることから、パラメータの値は一定とした。それぞれ 50 ずつのデータを持

	1	2	...	15
time	Head	Neck	...	Right foot
0	被験者Aの動作10は開始から終了まで0.3秒			
0.1				
0.2				
0.3				

	1	2	...	15
time	Head	Neck	...	Right foot
0	被験者Cの動作18は開始から終了まで0.2秒			
0.1				
0.2				

図 3 フレーム数の違い

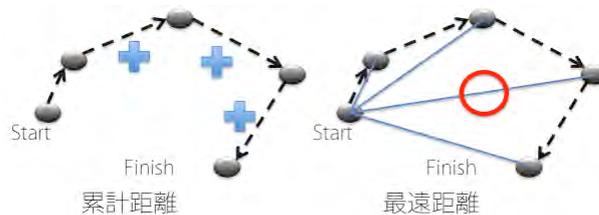


図 4 累計距離と最遠距離

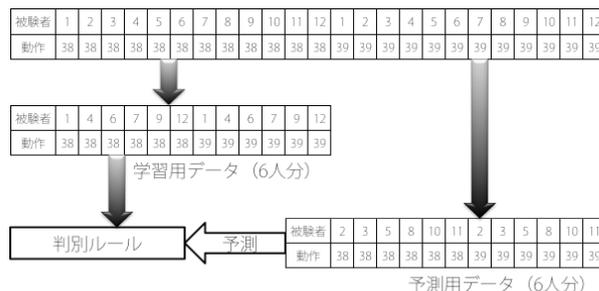


図 5 ニューラルネットワークによる判別

つ3種を判別する事例を実行している文献1を参考にして、今回のパラメータの値とした。

nnet(分析データ, data=x, size = 2, rang = 0.1, decay = 5e-4, maxit = 200)

5. ニューラルネットワークによる判別

最遠距離を各被験者の動作ごとに標準化したデータを用いて 57 動作すべてを判別したところ、その判定率は 5.2%、「単純仕草 (25 動作)」は 15.6%、「単純な動作 (12 動作)」は 9.0%、「立ち座り (20 動作)」は 10.5%と、それぞれの動作グループごとの正解率は芳しくなかった。

そこで、類似の動作ごとの判別特性を探るため、以下の分析を行った。

図6以降を見ると、おおむね最遠距離で標準化したデータの正解率が高かった。これは、個々の動作の特徴を最遠距離の方が表しており、また個人ごとに動作時の骨格の動きの幅が異なるため標準化したほうが良い結果が出たと考えられる。

5.1 単純仕草における判別

・頭の動作同士の判別 (図6)

[1 右へ90°]-[2 左へ90°]は頭の左右の判別を示すが、正解率が50%を下回ったのは、骨格の移動がほとんどなかったためであろう。頭の前後の動作を判別する[3 前へ90°]-[4 後ろへ90°]は「最遠」で7割近くの判別を見せたのは、前後に比べ骨格の移動量が大きいためであると考えられる。

	1	2
1 [右へ90°]	41.3%	58.7%
2 [左へ90°]	58.6%	41.4%
正解率	41.4%	

累計

	1	2
1 [右へ90°]	46.7%	53.3%
2 [左へ90°]	45.9%	54.1%
正解率	50.4%	

最遠

	3	4
3 [前へ90°]	54.5%	45.5%
4 [後ろへ90°]	57.0%	43.0%
正解率	48.8%	

累計

	3	4
3 [前へ90°]	66.9%	33.1%
4 [後ろへ90°]	29.5%	70.5%
正解率	68.7%	

最遠

	1	2
1	40.2%	59.8%
2	65.1%	34.9%
正	37.6%	

累計 - 標準化

	1	2
1	47.6%	52.4%
2	58.2%	41.8%
正	44.7%	

最遠 - 標準化

	3	4
3	42.6%	57.4%
4	56.1%	43.9%
正	43.2%	

累計 - 標準化

	3	4
3	59.4%	40.6%
4	34.0%	66.0%
正	62.7%	

最遠 - 標準化

図6 頭の動作同士の判別

・腕の動作同士の判別 (図7)

腕を横へ上げる動作である[5 右手を横へ肩の高さまで]-[6 左手を横へ肩の高さまで]、正面へ上げる[8 右手を正面へ肩の高さまで]-[9 左を正面へ肩の高さまで]、正面から上へ上げる[11 右手を正面から真上へ]-[12 左手を正面から真上へ]、腕を振る[14 右手だけを前後に振る]-[15 左手だけを前後に振る]

を前後に振る]の判別は、いずれも100%に近い正解率であったのは、ほとんどの骨格が移動しない状態で片方の腕の骨格だけが移動するため、移動する骨格に明確な違いがありその移動量も大きいためであろう。

それぞれに両手での動作を加えた[5 右手を横へ肩の高さまで]-[6 左手を横へ肩の高さまで]-[7 両手を横へ肩の高さまで]、[8 右手を正面へ肩の高さまで]-[9 左を正面へ肩の高さまで]-[10 両手正面へ肩の高さまで]、[11 右手を正面から真上へ]-[12 左手を正面から真上へ]-[13 両手を正面から真上へ]、[14 右手だけを前後に振る]-[15 左手だけを前後に振る]-[16 両手を前後に振る]でも、標準化したデータを使用したものは9割を超える高い正解率を示していた。これも、前者と同様の理由であろう。

	5	6	7
5 [右手を横へ肩の高さまで]	73.0%	11.6%	15.5%
6 [左手を横へ肩の高さまで]	10.3%	70.2%	19.5%
7 [両手を横へ肩の高さまで]	20.9%	24.5%	54.6%
正解率	65.9%		

累計

	5	6	7
5 [右手を横へ肩の高さまで]	85.2%	6.8%	8.0%
6 [左手を横へ肩の高さまで]	6.9%	79.2%	13.9%
7 [両手を横へ肩の高さまで]	13.0%	17.7%	69.3%
正解率	77.9%		

最遠

	5	6	7
5	100.0%	0.0%	0.0%
6	0.0%	91.5%	8.5%
7	0.1%	0.2%	99.7%
正	97.1%		

累計 - 標準化

	5	6	7
5	100.0%	0.0%	0.0%
6	0.0%	91.7%	8.3%
7	0.1%	6.6%	93.3%
正	95.0%		

最遠 - 標準化

	8	9	10
8 [右手を正面へ肩の高さまで]	67.7%	12.9%	19.4%
9 [左手を正面へ肩の高さまで]	12.5%	72.2%	15.3%
10 [両手正面へ肩の高さまで]	26.2%	23.1%	50.8%
正解率	63.6%		

累計

	8	9	10
8 [右手を正面へ肩の高さまで]	80.6%	6.6%	12.8%
9 [左手を正面へ肩の高さまで]	6.0%	86.3%	7.7%
10 [両手正面へ肩の高さまで]	13.1%	12.0%	74.9%
正解率	80.6%		

最遠

	8	9	10
8	91.9%	0.0%	8.1%
9	0.0%	100.0%	0.0%
10	0.2%	0.6%	99.2%
正	97.0%		

累計 - 標準化

	8	9	10
8 [右手を正面へ肩の高さまで]	97.7%	0.0%	2.3%
9 [左手を正面へ肩の高さまで]	0.0%	100.0%	0.0%
10 [両手正面へ肩の高さまで]	0.1%	0.0%	99.9%
正	99.2%		

最遠 - 標準化

	11	12	13
11 [右手を正面から真上へ]	70.8%	12.7%	16.5%
12 [左手を正面から真上へ]	13.4%	67.2%	19.4%
13 [両手を正面から真上へ]	26.9%	29.1%	44.0%
正解率	60.6%		

累計

	11	12	13
11 [右手を正面から真上へ]	82.1%	7.5%	10.4%
12 [左手を正面から真上へ]	8.0%	76.1%	15.9%
13 [両手を正面から真上へ]	16.0%	17.0%	67.1%
正解率	75.1%		

最遠

	11	12	13
11	92.9%	0.0%	7.1%
12	0.0%	90.0%	10.0%
13	2.7%	9.0%	88.3%
正	90.4%		

累計 - 標準化

	11	12	13
11 [右手を正面から真上へ]	99.3%	0.0%	0.7%
12 [左手を正面から真上へ]	0.0%	87.5%	12.5%
13 [両手を正面から真上へ]	0.1%	4.5%	95.4%
正	94.1%		

最遠 - 標準化

	14	15	16
14 [右手だけを前後に振る]	74.0%	9.1%	17.0%
15 [左手だけを前後に振る]	9.9%	71.2%	18.9%
16 [両手を前後に振る]	22.5%	21.9%	55.7%
正解率	67.0%		

累計

	14	15	16
14 [右手だけを前後に振る]	81.1%	5.8%	13.1%
15 [左手だけを前後に振る]	6.2%	80.9%	12.9%
16 [両手を前後に振る]	14.8%	16.0%	69.2%
正解率	77.0%		

最遠

	14	15	16
14	97.5%	0.0%	2.5%
15	0.0%	99.0%	1.0%
16	1.6%	2.7%	95.8%
正	97.4%		

累計 - 標準化

	14	15	16
14 [右手だけを前後に振る]	91.7%	0.0%	8.3%
15 [左手だけを前後に振る]	0.0%	95.4%	4.6%
16 [両手を前後に振る]	0.2%	1.1%	98.7%
正	95.3%		

最遠 - 標準化

図7 腕の動作同士の判別

・腰の動作同士の判別 (図8)

腰を前後に曲げる動作の判別である[17 前へ45° 曲げる]-[18 前へ90度曲げる]では「累計-標準化」、腰を左右に曲げる動作である[19 右へ45° 捻る]-[20 左へ45° 捻る]では「最遠-標準化」、同じく左右の角度が大きくなる[21

右へ90°捻る]-[22左へ90°捻る]では「累計-標準化」が7割近くの高い正解率であった。人による曲げる度合いが異なっていたため、標準化したほうが良い結果をもたらしたと考えられる。

累計		17	18
17 [前へ45° 曲げる]		74.3%	25.7%
18 [前へ90° 曲げる]		48.2%	51.8%
正解率		63.1%	

累計 - 標準化		17	18
17 [前へ45° 曲げる]		69.9%	30.1%
18 [前へ90° 曲げる]		34.7%	65.3%
正		67.6%	

最遠		17	18
17 [前へ45° 曲げる]		58.7%	41.3%
18 [前へ90° 曲げる]		42.2%	57.8%
正解率		58.2%	

最遠 - 標準化		17	18
17 [前へ45° 曲げる]		45.9%	54.1%
18 [前へ90° 曲げる]		58.1%	41.9%
正		43.9%	

累計		19	20
19 [右へ45° 捻る]		48.8%	51.2%
20 [左へ45° 捻る]		46.4%	53.6%
正解率		51.2%	

累計 - 標準化		19	20
19 [右へ45° 捻る]		56.6%	43.4%
20 [左へ45° 捻る]		45.3%	54.7%
正		55.7%	

最遠		19	20
19 [右へ45° 捻る]		77.0%	23.0%
20 [左へ45° 捻る]		41.6%	58.4%
正解率		67.7%	

最遠 - 標準化		19	20
19 [右へ45° 捻る]		66.3%	33.7%
20 [左へ45° 捻る]		29.6%	70.4%
正		68.4%	

累計		21	22
21 [右へ90° 捻る]		59.1%	40.9%
22 [左へ90° 捻る]		41.0%	59.0%
正解率		59.1%	

累計 - 標準化		21	22
21 [右へ90° 捻る]		62.8%	37.2%
22 [左へ90° 捻る]		27.1%	72.9%
正		67.9%	

最遠		21	22
21 [右へ90° 捻る]		57.0%	43.0%
22 [左へ90° 捻る]		25.8%	74.2%
正解率		65.6%	

最遠 - 標準化		21	22
21 [右へ90° 捻る]		56.9%	43.1%
22 [左へ90° 捻る]		27.4%	72.6%
正		64.8%	

図8 腰の動作同士の判別

・足の動作同士の判別 (図9)

足を引き上げる動作である[23右足を引き上げる]-[24左足を引き上げる]は「最遠-標準化」を用いれば7割を超える正解率で、両足の動作を含めた[23右足を引き上げる]-[24左足を引き上げる]-[25両足を上げて飛ぶ]では「累計-標準化」を使用すれば6割程度の正解率であった。

累計		23	24
23 [右足を引き上げる]		70.0%	30.0%
24 [左足を引き上げる]		30.5%	69.5%
正解率		69.7%	

累計 - 標準化		23	24
23 [右足を引き上げる]		66.8%	33.2%
24 [左足を引き上げる]		33.5%	66.5%
正		66.7%	

最遠		23	24
23 [右足を引き上げる]		76.6%	23.4%
24 [左足を引き上げる]		30.9%	69.1%
正解率		72.9%	

最遠 - 標準化		23	24
23 [右足を引き上げる]		73.0%	27.0%
24 [左足を引き上げる]		25.2%	74.8%
正		73.9%	

累計		23	24	25
23 [右足を引き上げる]		48.5%	28.6%	22.8%
24 [左足を引き上げる]		31.7%	40.0%	28.4%
25 [両足を上げて飛ぶ]		17.3%	19.4%	63.2%
正解率		50.6%		

累計 - 標準化		23	24	25
23 [右足を引き上げる]		62.3%	31.6%	6.1%
24 [左足を引き上げる]		29.0%	60.0%	11.0%
25 [両足を上げて飛ぶ]		21.0%	17.6%	61.5%
正		61.3%		

最遠		23	24	25
23 [右足を引き上げる]		54.4%	21.9%	23.7%
24 [左足を引き上げる]		22.7%	51.3%	26.0%
25 [両足を上げて飛ぶ]		21.6%	19.7%	58.7%
正解率		54.8%		

最遠 - 標準化		23	24	25
23 [右足を引き上げる]		66.0%	24.8%	9.2%
24 [左足を引き上げる]		21.2%	64.7%	14.1%
25 [両足を上げて飛ぶ]		31.9%	30.5%	37.7%
正		56.1%		

図9 足の動作同士の判別

5.2 単純な動作における判別

「最遠距離を標準化」のデータを用いて判別した正解率は、「押す」は42.3%、「引く」は36.4%、「置く」は33.0%、「取る」は42.8%と、高い正解率は得られなかった。これは、他の動作と比べて、動作間で移動する骨格の種類と距離が類似しているため、今回の「動作の指標」以外の指標を検討する必要がある。

5.3 立ち座りにおける判別

椅子での動作同士の判別である[38その場で立つ]-[39その場で座る]の正解率は「最遠」で51.3%、[40立って前へ]-[41前から座る]は「累計」で61.6%、[42立って右へ]-[43右から座る]は「最遠」で65.3%、[44立って左へ]-[45左から座る]は「最遠」で54.5%、[46立って後ろへ]-[47後ろから座る]は「最遠-標準化」で89.0%であった。また、床での動作同士の判別である[48その場で立つ]-[49その場で座る]は「最遠-標準化」で74.8%、[50立って前へ]-[51前から座る]は「最遠」で63.0%、[52立って右へ]-[53右から座る]は「累計」で52.3%、[54立って左へ]-[55左から座る]は「最遠-標準化」で75.7%、[56立って後ろへ]-[57後ろから座る]は「最遠-標準化」で87.8%であった。

椅子での動作である[38その場で立つ]-[39その場で座る]、[44立って左へ]-[45左から座る]、床での動作である[52立って右へ]-[53右から座る]以外は正解率が6割を超え、中には8割を超えるものもあった。「最遠」を用いたほうが良い結果が出る傾向があった。

6. まとめ

動作数が多い場合は、判別できると言えるまでの正解率を上げることはできなかったが、腕のように移動する骨格が少なく、移動距離が大きい動作の判別は、判別する動作の数が少なければその正解率は高く、3つ程度の動作であれば十分正解率は高いことが明らかになった。

そこで、骨格の動きの異なる少数の動作を集めてニューラルネットワークを用いた判別ルールを作成し、このような組み合わせによるニューラルネットワークを複数作成し、入力された動作を複数のニューラルネットワークによって判定し、最も判定率の良いものを予測動作とすることで、複数の動作を判別することが可能になると考える。

[注]

- 1) <http://www.xbox.com/ja-JP> (1,2とも2012年10月8日現在)
- 2) <http://www.microsoft.com/en-us/kinectforwindows/>

[参考文献]

- 1) 熊谷悦生、舟尾暢男、オーム社: Rで学ぶデータマイニング (2) シミュレーション編、2008/12

*1 早稲田大学 理工学総合研究所 客員准教授 博士(工学)
 *2 東京理科大学 理工学部建築学科 助教 博士(建築学)
 *3 早稲田大学 理工学部建築学科 教授 工学博士