行動理解とデータマイニングを適用した人物意図推定・行動予測

片岡　裕雄†　　青木　義満†

† 慶應義塾大学大学院理工学研究科
神奈川県横浜市港北区日吉 3-14-1
E-mail: †kataoka@aoki-medialab.org, ††aoki@elec.keio.ac.jp

あらまし　本稿では人物の意図を理解し次の行動を事前に取得する，”人物意図推定”や”行動予測”に関する研究を提案する．現在までの人物解析研究の提案では軌跡や姿勢，行動などを事後解析していたが，異常行動の阻止や人物の傾向を把握した自然な行動からのインテリジェント空間構築などのためにも，事前に行動を予測する意義は大きいと考える．人物の行動予測は，(1) 人物行動理解：人物が現在何をしているか，さらに (2) データマイニング：行動履歴データベースのマイニングと行動の予測 から構成される．実験では複数の人物行動理解データセットにおいて高い精度を実現し，日常の行動履歴データベースを用いて人物の意図推定・行動を予測した．いずれの状況においても標準的なラップトップ PC を適用し，20fps 前後でのリアルタイム処理を実現している．

キーワード　人物行動理解，意図推定，行動予測，HOOF，ECoHOG，Naive Bayes

1. 序　　論

1.1 研究背景

現在までの人物行動解析研究により，数多くの技術が提案されてきた [1]．人物の追跡・姿勢推定・行動理解や顔認識がその代表的な例であり，今後も発展し続けながら実利用に多大な貢献が期待される．しかしながら，ここで挙げている画像認識や機械学習技術は行動の事後解析であり,「すでに起きている」事象の解析に重きが置かれていた．一方，行動を事前に解析出来れば予期せぬ動作を事前に阻止することや人物の自然な動作からコンピュータを制御するといった高度な情報空間の構築など，多種多様なアプリケーションへの応用が期待できる．

本稿では，行動理解とデータマイニングの技術統合による人物意図推定や行動予測を提案する．人物行動理解では個別の追跡ウインドウからエッジとフロー特徴 (Edge-Flow Feature Integration) により人物の特徴を記述する．エッジ，フロー特徴共に，時系列で特徴を記述する．また，両者の組み合わせにより相補的に特徴を捉え，行動理解の識別精度を向上可能と考える．ここで，エッジ特徴としてエッジの共起性を捉える CoHOG (Co-occurrence Histograms of Oriented Gradients) [15] を拡張した ECoHOG (Extended CoHOG) を，フロー特徴としては Optical Flow を時系列で累積した HOOF (Histograms of Oriented Optical Flow) [4] を適用する．データマイニングのステップでは，Naive Bayes 識別器により行動履歴データベースを解析，未来の行動出現確率を計算し，行動を予測する．

1.2 関連研究

人物行動理解と意図推定に関する関連研究を紹介する．

人物行動理解：関連する論文としては, [2] - [10] が挙げられる．Klaser らは 3D-HOG [2] を，Marszalek らは複数の特徴量を組み合わせることで高い精度を誇る手法を提案している [3]．その中で，Chaudhry らは動き特徴を時系列で解析する HOOF(Histograms of Oriented Optical Flow) を提案し，高精度に行動を識別する手法を提案した [4]．また，最近では Satkin ら [5] や Wang [6] らが提案する手法が高い精度を誇っている．個人毎の行動理解だけでなく，人物間のインタラクションにおいても研究が進められている．Ryoo ら [7] や Perez ら [10] が提案する手法について紹介する．Ryoo らは時系列情報を詳細に解析した上での行動理解を実現した [7]．行動の始点と終点の認識，行動の主な領域から特徴を記述することや時系列行動の関係性に着目しているため，高精度な複数人物間行動の識別を実現している．Perez らは人物の位置，顔の位置や方向，人体から得られる動きを基に人物同士のインタラクションを識別している．構造的な機械学習を適用しており，識別器には SVM を適用している．しかし，個人の行動，人物間のコミュニケーションを理解し，次の行動を予測したうえでフィードバックするためには精度を向上させるだけでなく，特に処理時間の面で改善する必要があるといえる．

意図推定:Pellegrini らの LTA(Local Trajectory Avoidance) が挙げられる [11]．LTA では，追跡人物の位置/速度と，障害物や他の追跡人物から未来の位置を確率分布として予測する．また，Kiefer は，”Mobile Intention Recognition”において人物の軌跡情報や予め付加された位置情報から人物の意図，行動プランを決定する方法を紹介している [12]．しかし，これらの手法は主に位置を

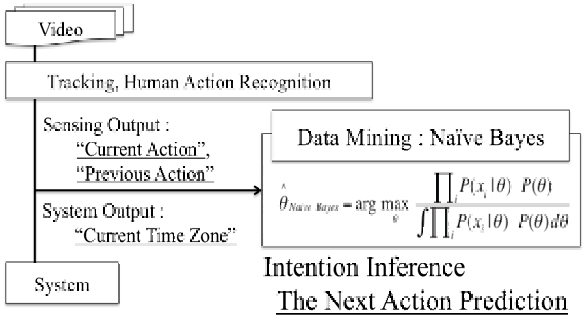

図 2　意図推定・行動予測のフレームワーク

推定する方法であり，高次な人物の行動情報を取得する手法ではない．一方，Callejas らは音声認識・自然言語処理を適用して意図を推定している [13]．主に会話を処理して心理状態を理解，さらには意図を把握する．会話から状況を認識する研究では，会話のトーンや内容から心理状況を割り出すことは可能であるが，事後解析の枠を超えていないことや，行動を認識するわけではないという点が挙げられる．

1.3　処理の流れ

図 1 に行動理解，図 2 に意図推定のフレームワークを示す．意図推定・行動予測技術は，Edge-Flow の特徴統合による人物行動理解と Naive Bayes 識別器を用いたデータマイニングから構成される．行動理解には HOOF, ECoHOG を統合した手法を適用している．行動理解により取得された行動のタグは意図推定で用いられ，Naive Bayes 識別器を適用して次の行動を予測する．

ここで，本稿の構成を示す．2 章は追跡技術，3 章で人物行動理解手法を，4 章では人物行動理解システムの実験について説明する．5 章では人物行動意図推定・行動予測について記述した上で，6 章で本稿の結びとする．

2.　人 物 追 跡

追跡手法は Particle Filter [16] と Optical Flow [17] のコンビネーションにより構成される．カラーベースの Particle Filter と階層的クラスタリングを組み合わせた Optical Flow Detection により高精度な追跡を実現した．静止状態や動作量が小さい場面においては Particle Filter による色ヒストグラム計算が効果的であるが，動きの激しい状況では Optical Flow + 階層的クラスタリングによる検出手法が効果的であり，相補的な追跡が可能である．ここで，Optical Flow と階層的クラスタリングを組み合わせた検出方法について図示する．図 3 は，Optical Flow から階層的クラスタリングを適用した結果，さらに，部分的なオクルージョンを含む場合にクラスタリングした結果である．Optical Flow から階層的クラスタリングにより部分的な領域を検出，さらには重なりを含む領域のクラスタリングにより，個人を検出可能である．

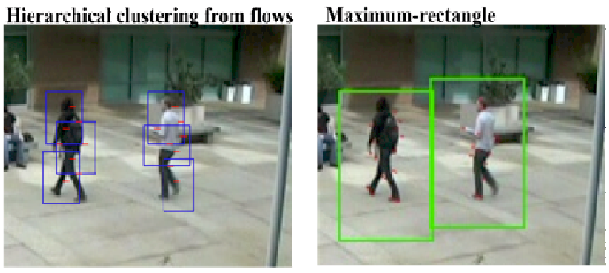

図 3　Optical Flow Detection：(左)Optical Flow によるフローと階層的クラスタリング (右) 矩形に重なりを含む場合に矩形同士を統合する．Optical Flow と階層的クラスタリングによる動物体検出．

3.　人物行動理解

人物行動はエッジとフローの特徴統合により識別する．ここではエッジ特徴として ECoHOG を，フロー特徴として HOOF 特徴を適用しているが，精度の向上を図るため，それぞれの特徴量に改良を加えている．本章では人物行動理解のための特徴量やその改良について記述する．

HOOF (Histograms of Oriented Optical Flow)：HOOF は Optical Flow を時系列で累積する特徴であり，人物のモーションをヒストグラムで表現可能である [4]．特徴量はフレーム間から取得された Optical Flow の強さ (長さ) と方向をヒストグラムに累積する．図 4 はフロー方向を 4 分割した量子化方法であり，左右対称の動作を同一行動として扱う．オプティカルフローの長さを強度として扱い，対応するヒストグラム位置に累積する．ここでは更に，個人差や動作の違いによる特徴ヒストグラムの位置ずれを補正するために，時系列の正規化処理を加えた (図 5)．取得した 5 フレームのヒストグラム値をガウシアン関数により補正した．

$$f(x) = \frac{1}{2\pi\sigma^2}(-\frac{x^2}{2\sigma^2}) \qquad (1)$$

ここで，x は正規化するフレーム数であり，σ はフレーム数における時系列の分布 $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$ を示す．5 フレームが個人間や行動間，時系列のずれを補正できる数としてヒストグラムを正規化している．

ECoHOG (Extended Co-occurrence Histograms of Oriented Gradients)：エッジ特徴，つまり人物から得られる形状特徴は CoHOG を拡張した ECoHOG により記述する．まずは CoHOG について記述する．

CoHOG は予め用意された特徴取得ウインドウからピクセルのペアで勾配を記述する特徴量である [15]．ウインドウの中心と対象ピクセル，二つの画素両方から量子化された勾配を取得し，ペアの数のカウントにより特徴を記述する．単一のピクセルのみではなく，共起性を考慮したエッジ特徴の記述により，詳細な記述を可能にす

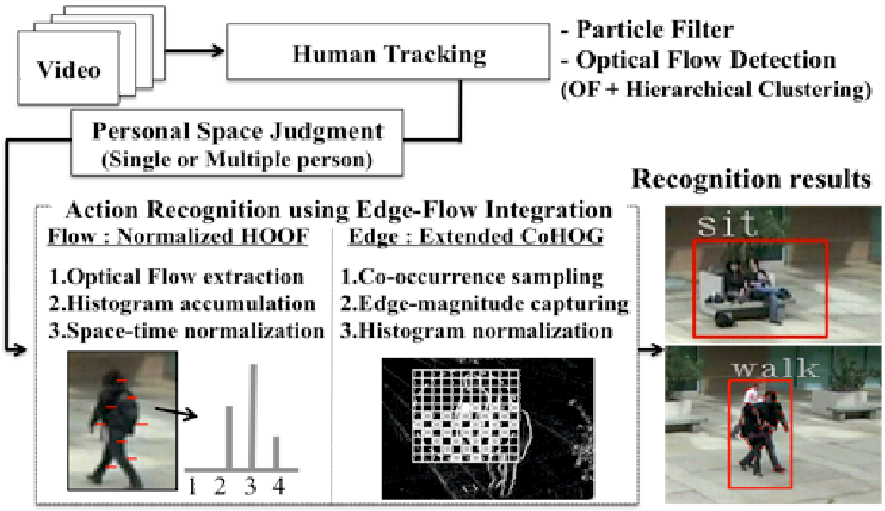

図 1　行動理解のフレームワーク

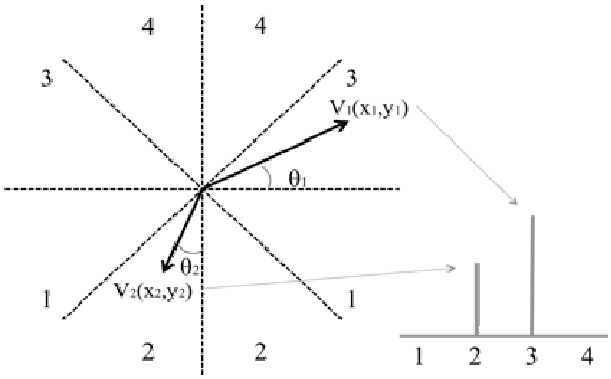

図 4　HOOF のヒストグラム：量子化方向数は 8 であるが，左右対称の動きを同一に扱う．

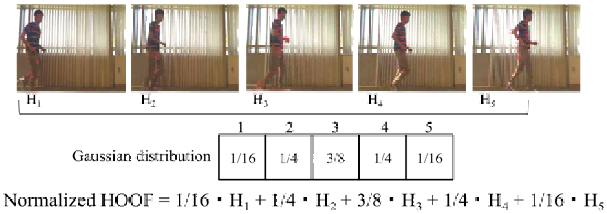

図 5　時系列の正規化：HOOF を時系列方向で正規化する．個人差，行動の違いによるヒストグラムのずれを補正する．

る．行動理解においては複数人物間の特徴記述に関して効果的であると思われる．

ECoHOG では CoHOG から以下 3 つの改良を加えた．

- エッジ強度の累積
- エッジペアのステップ取得
- 時系列特徴表現

エッジ強度の累積：CoHOG ではエッジ勾配ペアのカウントにより特徴を累積していたが，ECoHOG では強度を累積する．エッジ強度の累積により，形状だけでなく濃淡の強度や，その変化度合いを記述出来るため，人物の行動を理解する上で必要な情報を累積出来ると考える．下に特徴累積方法を示す．

$$m_1(x_1, y_1) = \sqrt{f_{x1}(x_1, y_1)^2 + f_{y1}(x_1, y_1)^2} \tag{2}$$

$$m_2(x_2, y_2) = \sqrt{f_{x2}(x_2, y_2)^2 + f_{y2}(x_2, y_2)^2} \tag{3}$$

$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} m_1(x_1,y_1) + m_2(x_2,y_2) \\ (if d(p,q) = i \\ and\ d(p+x, q+y) = j) \\ \\ 0\ (otherwise) \end{cases} \tag{4}$$

ここで f_x, f_y はそれぞれ x,y 方向における微分量であり，隣り合う画素における差分を示す．共起ヒストグラム $C(i,j)$ にはエッジ強度 m を累積する．(p,q) はウインドウ中心，$(p+x, q+y)$ はウインドウ中心位置とペアをなす特徴取得ウインドウ中の対象画素である．方向は 8 方向に量子化されており，$d(p,q)$ は 0 - 7 までの整数値である．

さらに，画像の明るさの変動に対処するために共起ヒストグラムの正規化処理を加えた．以下に正規化について記述する．

$$C'_{x,y}(i,j) = \frac{C_{x,y}(i,j)}{\sum_{p=1}^{8}\sum_{q=1}^{8} C_{x,y}(i',j')} \tag{5}$$

正規化後の共起ヒストグラム $C'_{x,y}$ (64 次元) 毎の合計が 1.0 になるように正規化されている．

エッジペアのステップ取得：ウインドウ内に空間的な距離を設け，特徴次元数の増大を防ぎながら距離の離れ

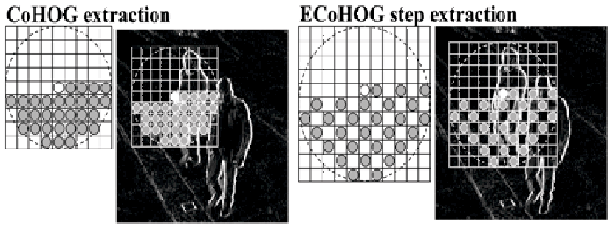

図 6　(左)CoHOG の特徴取得ウインドウ (右)ECoHOG の特徴取得ウインドウ：ECoHOG は空間的な距離を設けてエッジペアを取得している．離れた位置にある部位間や複数人物間におけるエッジの関係性を記述する．

ている画素のペアを記述する．また，隣接する画素においてほぼ特徴が同一になることから，特徴取得ウインドウに空間的な距離を設けても特徴記述性能が低下する懸念はないと考える．行動理解においてはより離れた位置に存在する複数人物間のエッジを記述し，人物同士のインタラクションを含めた行動理解において効果的である．図 6 に CoHOG と ECoHOG の特徴取得ウインドウの違いを記述する．

時系列特徴表現：画像から得られた特徴ヒストグラムを時系列で連結させることにより特徴を記述する．単純なヒストグラム連結により表現される特徴で，一フレームから取得する次元数と累積するフレーム数の積が特徴量の総次元数である．

Edge-Flow Feature Integration：特徴を統合するために，識別器の出力値を組み合わせる．本稿では AdaBoost [18] を用いて識別器の出力値を返しているが，両方の特徴が均等になるように組み合わせる．以下に出力値の統合を示す．

$$r = \alpha r_{edge} + (1-\alpha) r_{flow} \tag{6}$$

$$\alpha = \frac{r_{edge}}{r_{edge} + r_{flow}} \tag{7}$$

$$1-\alpha = \frac{r_{flow}}{r_{edge} + r_{flow}} \tag{8}$$

$$= 1 - \frac{r_{edge}}{r_{edge} + r_{flow}} \tag{9}$$

ここで，r_{edge} はエッジ特徴 ECoHOG，r_{flow} はフロー特徴 HOOF で AdaBoost を構成した際の出力値である．

4. 行動理解手法の実験

本稿で提案した人物行動理解システムの有効性を示すため，既存のデータセットを用いて実験を行った．行動理解の検証には (i) Weizmann Action Dataset [19] (図 7) (ii) Videoweb Activity Dataset [20] (図 8) (iii) Semantic Description Human Action (SDHA) 2010 Dataset [7] (図 9) と 3 種類の場面で撮影されたデータセットを適用している．ここで，それぞれのデータセットの対象と行動

図 7　Weizmann Action Dataset

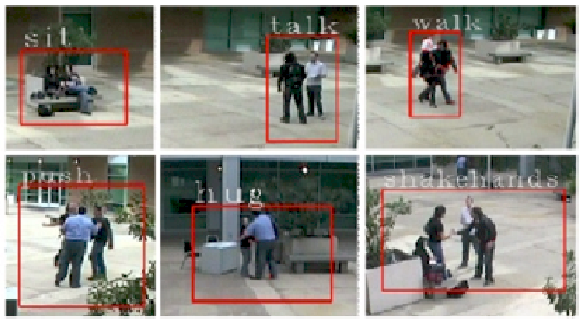

図 8　Videoweb Activity Dataset

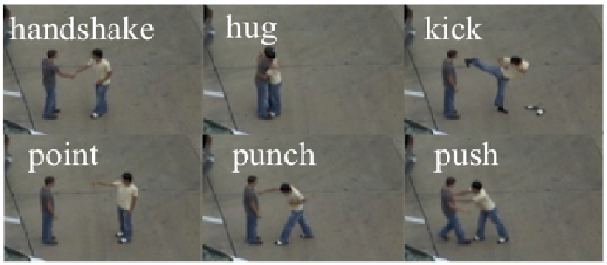

図 9　Semantic Description Human Action 2010 Dataset

の種類について説明する．(i) 対象：単一人物による行動分類を対象としている．種類：bend, jack, jump, pjump, run, side, skip, walk, wave, wave2 の 10 種類．(ii) 対象：屋外環境における複数人物のインタラクション．種類：hug, push, run, shakehands, sit, talk, walk の 7 種類．(iii) 対象：高度な人物間のインタラクション．種類：handshake, hug, kick, point, punch, push の 6 種類．

実験では標準的なラップトップ PC を適用しているが全ての実験において 20fps 前後での処理を実現した．

Weizmann Action Dataset：単一人物の行動理解精度を検証するために Weizmann Action Dataset を適用した．結果について，従来手法との比較を表 1 に示す．Weizmann Action Dataset では提案手法と複数の従来手法について比較している．

Videoweb Activity Dataset：屋外環境における複数人物間のインタラクションの精度を検証するために Videoweb Activity Dataset を適用した．結果について，従来手法との比較を表 2 に，詳細な識別結果を表 3 に示す．Videoweb Activity Dataset においては提案手法と

表 1　Weizmann Action Dataset を用いた精度検証

Framework	Accuracy(%)
Proposed method (NHOOF&ECoHOG)	98.8
Wang *et al.* [6]	97.2
Satkin *et al.* [5]	95.8
Chaudhry *et al.* (HOOF) [4]	94.4
Klaser *et al.* (HOG3D) [2]	90.7

表 2　Videoweb Activity Dataset を用いた行動理解の精度と処理時間

Framework	Accuracy (%)	Time (ms)
Proposed method	99.9 (2330/2332)	54.51
Chaudhry *et al.* [4]	76.3 (1781/2332)	39.00

表 3　The results on Videoweb Dataset

	hu	pu	ru	sh	si	ta	wa
hu	**100**						
pu		**100**					
ru			**100**				
sh				**100**			
si					**100**		
ta						**99.8**	0.2
wa						0.2	**99.8**

表 4　The accuracy on SDHA2010

Framework	Accuracy (%)	Time (ms)
Proposed method	81.8	24.4
Ryoo *et al* [7]	91.1	-
Niebles *et al* [8]	81.5	-
Dollar *et al* [9]	81.2	-

表 5　The detailed results on SDHA

	shake	hug	kick	point	punch	push
shake	**90.6**	5.9	0.1	0.9	0.3	1.8
hug	4.4	**92.4**		0.5	0.2	2
kick	6.3	6.2	**84.6**		1.3	1.4
point	1.9	1.1		**96**	0.4	0.2
punch	17	17.6	3.9		**55.4**	5.9
push	10	28	0.5		2	**59.2**

HOOF を比較した上で処理時間も計測している．表 2 には処理時間の比較も記述している．

Semantic Description Human Action (SDHA) 2010 Dataset：人物間の高度なインタラクションを認識するために SDHA2010 を適用した．行動理解の結果について，従来手法との比較を表 4 に，詳細な識別結果を表 5 に示す．

本稿の実験ではエッジ・フロー両方の特徴統合から行動を識別している．複数の特徴を統合することから多様な行動や多種の場面において強力な手法であることが期待される．実際，本稿で行った 3 つの異なる実験すべてにおいても高精度な行動理解を実現している．ここでは (i) 単一人物における多様な行動 (ii) 屋外環境における複数人物間のインタラクション (iii) 高度な人物間のインタラクション という，場面や目的の異なるデータセットを適用した．時系列で人物特徴を記述する局所特徴量により詳細な行動記述が可能になった上，特徴の統合によりエッジ・フローから相補的な行動の識別を実現したと考えられる．(e.g. (ii) における sit と talk：フローは見られないがエッジにより形状を詳細に評価可能) また，HOOF における正規化が個人差や行動間の動作のずれを補間できたことや，エッジ強度累積により濃淡変化の強さだけでなく，曲線や細部における形状の変化度合いまで含めた表現の改善も含まれると考える．複数人物間の行動について，ECoHOG における空間的な距離を設けた特徴取得方法が効果的であったと考える．エッジ特徴では，隣接する画素間には特徴の変化がほとんど見られなかったため，空間的な距離を設けても特徴記述の精度には影響が無かったことや，より広範な領域から複数人物間のエッジ共起性を捉えられることから，個人の部位毎の関係性を記述するのみでなく，複数人物間のインタラクションを高精度に捉えることができた．従来手法との精度比較からも，本稿における提案が有効であることが示せた．しかし，高度な人物間インタラクションにおいては課題が残る結果となった．SDHA2010 Dataset の動画像では，歩行を含んだ上で人物間のインタラクションしている場面が撮影されている．SDHA2010 では過半数の動作が歩行→腕部 (脚部) 伸縮という構成になっており，特に shake，punch，push においては提案手法が混同し易い特徴となった．一方で Ryoo らは，詳細なコンテキストを解析した上で行動を分離していた [7]．今後は追跡結果のフィードバックや姿勢も推定した上での総合的な行動理解が必要であると考える．

5. 人物意図推定・行動予測

4 章までで提案した行動理解手法や行動履歴データベース解析を適用した上で人物の行動を予測する．本稿では日常空間における行動に着目し，椅子に座った後，次に何をしたか，という行動を予測した．図 10 に実際の日常空間の動画像の一例を示す．図では walk - bend - sit という一連の流れである．ここで，日常空間データベースとデータマイニング技術について解説する．

日常空間データベース：本稿で適用したデータベースは”Time Zone(時間帯)”，”Previous Action(一つ前の行動)”，”Current Action(現在の行動)”，”Intention(意図)”4 つの属性から構成されている (図 11)．時間帯により人物行動のパターンが変化することや，一連の流れの中で行動が生起することから，上記 3 つの属性を入力として意図を求める構成にした．Time Zone はシステムから取得される時間を”Morning(朝)”，”Day(昼)”，”Night(夜)”に分割してデータ化した．Previous Action と Current Action は 3 章で説明した行動理解手法で求め，行動のバリエーションは”bend(屈む)”，”sit(座る)”，”stand(立

図 10　屋内空間で撮影された映像

	A	B	C	D
1	Time Zone	Previous	Current	Intention
2	morning	walk	sit	book
3	morning	walk	stand	book
4	day	walk	sit	pc
5	night	bend	sit	meal
6	night	walk	sit	pc
7	night	walk	sit	pc
8	night	bend	sit	meal
9	night	walk	stand	book
10	night	walk	sit	pc

図 11　日常空間データベース：Time Zone，Previous Action，Current Action，Intention の属性により構成される．

つ)”，”walk(歩行)”と 4 つの動作を含めた．Intention は椅子に座った後の行動で，”book(読書)”，”meal(食事)”，”PC(パソコン)”3 つの行動を予測する対象動作とした．

データマイニング：Naive Bayes 識別器を適用してデータベースを解析する．本稿では Time Zone，Previous Action，Current Action と 3 種類の属性を入力として，Intention を推定する．Naive Bayes 識別器は，属性間の独立性に着目してベイズ推定を簡略化したモデルである．

まず，MAP 推定と Bayes 式を用いて変形した式を示す．

$$\theta_{MAP} = argmax_\theta P(\theta|x_1, x_2, ..., x_n) \quad (10)$$

$$= argmax_\theta \frac{P(x_1, x_2, ..., x_n|\theta)P(\theta)}{\int P(x_1, x_2, ..., x_n|\theta)P(\theta)d\theta} \quad (11)$$

ここで，θ は予測される行動を，x はそれぞれの属性 (Time Zone, Previous Action, Current Action) を示す．

Naive Bayes では，属性間の出現に依存性はない，つまり属性間は独立と仮定することで，学習を簡単にしようという方法である．独立性を仮定した場合の $P(x_1, x_2, ..., x_n|\theta)$ は以下のように書き換えられる．

$$P(x_1, x_2, ..., x_n|\theta) = \prod_i P(x_i|\theta) \quad (12)$$

よって，Naive Bayes は，

$$\theta_{NaiveBayes} = argmax_\theta \frac{\prod_i P(x_i|\theta)P(\theta)}{\int \prod_i P(x_i|\theta)P(\theta)d\theta} \quad (13)$$

である．Naive Bayes による推定は，属性間の関係が完全に独立であるならば MAP 推定と等価となる．また，データマイニングの分野において Naive Bayes は簡易的な学習でありデータ数を増やしても高速に学習できること，非常に高い識別性能を誇ることから頻繁に使用されるアルゴリズムである．

意図推定・行動予測：時間帯，時系列行動を入力して，次に何をするか予測する．Time Zone はシステムから返却される時刻を量子化した値，Previous Action，Current Action は 3 章で提案した行動理解手法を基にして算出した．Intention は Previous Action と Current Action が切り替わる際に計算する．今回は，日常行動データベースを解析した結果，80%の確率を超える際に人物の意図として行動を予測しているが，全ての予測対象において確率を求めているため，ランク付けが可能である．

ここで，図 12 に意図推定・行動予測の結果を示す．事前に行動をセンシングして，未来の状態を事前に予測する．図の例では walk - bend - sit - meal (drink) と一連の動作が行われている場合であるが，bend から sit に行動が切り替わる際に次の行動が予測された．図 12 中では”The person will have a meal”と表示されている様子が分かる．行動が切り替わった際に確率を計算するので walk から bend に移った際にも確率は算出しているが，行動頻度が少ないために意図として認識されていない．ここで，今回の例における Naive Bayes 識別器の計算を示す．例は TimeZone = night，PreviousAction = bend, CurrentAction = sit の場合であり，Intention が book(bk), PC, meal(ml) の場合それぞれについての確率を求める．

$$[\ intention = book(bk)]$$
$$P\ (night|bk) * P(bend|bk) * P(sit|bk) * P(bk)$$

$$[\ intention = PC]$$
$$P\ (night|PC) * P(bend|PC) * P(sit|PC) * P(PC)$$

$$[\ intention = meal(ml)]$$

表 6　提案手法 (Tracking, Action Recognition, Intention Inference, Total) の処理時間計測

Process	Time (ms)	FPS
Tracking	17.3	57.5
Action Recognition	40.8	24.4
Intention Inference	5.2×10^{-5}	1.9×10^{7} (19MIPS)
Total	58.1	17.2

$$P\ (night|ml) * P(bend|ml) * P(sit|ml) * P(ml)$$

なお，ここでは book の場合の確率が 0.11，PC が 0.0，meal で 0.89 であった．よって，図 12 における行動意図は”meal”であり，次の行動が予測された．PC の確率は 0.0 なので，図中では book，meal の確率しか表示されていない．

表 6 に追跡から行動理解，意図推定までの処理時間を示す．提案システムでは追跡，行動理解，意図推定までを含めても約 17fps と高速に処理可能である．追跡ではカラーベースの Particle Filter と Optical Flow Detection を適用している．Optical Flow と階層的クラスタリングを適用して Particle Filter の重心を補間しているが，すべて単純で高速な処理を適用しているため，60fps 近い追跡を実現している．行動理解において，局所特徴の取得には比較的時間を要するが追跡により得られた矩形内で特徴を取得していることが処理時間の高速化に影響したと考える．また，本稿で適用した 2 つの特徴量は時系列で取得する特徴であるため，一フレームでは取得する特徴サイズが小さかったことも挙げられる．意図推定では 52ns と微小な処理時間になったが，Naive Bayes 識別器は単純積のみで次行動の確率を予測することが影響した．本提案では Time Zone，Previous Action，Current Action と 3 つの属性の積であるため，高速に人物の意図を推定し，行動を予測出来たと考える．

6. 結　　論

本稿では，エッジとフロー特徴の統合で人物行動を理解し，Naive Bayes 識別器を用いたデータマイニングの枠組みにより人物意図の推定，行動予測を実現した．今後の課題としてはまず，行動理解の見直しが挙げられる．人物間のインタラクションを理解するため，または日常空間における詳細な動作を認識するために，詳細なコンテキストをセンシングする必要がある．現在までの追跡手法でも軌跡，速度，姿勢など，人物行動理解に必要な情報を含んでいるので，行動理解技術を拡張するためにも付加情報の適用を検討する．また，本稿で提案した人物意図推定・行動予測技術は，位置や軌跡情報によらない，詳細な行動を含めた予測ができるという点で有意性があると考える．将来への課題を挙げるという意味でも以下に適用先の一例を示す．

異常行動の予測と防止：事前に行動を予測可能なため，異常な行動を防止可能と考える．ここでいう異常行動とは,「普段と異なる」行動も含み，行動理解手法とも組み合わせることでサービス向上を実現する．今までは事後解析であり異常が発生した後に通知するシステムであったが，今後は事前に予測する点で実利用化が期待される．

自然な動作を入力としたインテリジェント空間の構築：次に何をするかが先読み出来るので，例えば「テレビの電源を入れようと行動」しただけでスイッチを入れることが可能になる．行動予測はインテリジェント空間の構築にも適用可能となり，自然な行動から環境を操作する．現在は Time Zone，Previous Action，Current Action と 3 属性からの予測であるが，今後は詳細な場所や個人の特性を理解した上での推定が必要と考える．

文　　献

[1] T.B.Moeslund *et al*,”Visual Analysis of Humans : Looking at People”,Springer, 2011

[2] A.Klaser *et al*,”A spatio-temporal descriptor based on 3D-gradients”,BMVC2008, 2008

[3] M.Marszalek *et al*,”Actions in context”,CVPR2009, 2009

[4] R.Chaudhry *et al*,”Histograms of oriented optical flow and binet caushy kernels on nonlinear dynamical systems for the recognition of human actions”,CVPR2009, 2009

[5] S.Satkin *et al*,”Modeling the temporal extent of actions”,ECCV2010, 2010

[6] Y.Wang *et al*,”Learning a discriminative hidden part model for human action recognition”, NIPS2008, 2010

[7] M.S.Ryoo *et al*,”Spatio-temporal relationship match : video structure comparison for recognition of complex human activities”,ICCV2009, pp.1593-1600, 2009

[8] J.C.Niebles *et al*,”Unsupervised learning of human action categories using spatial-temporal words”, IJCV2008, 2008

[9] P.Dollar *et al*,”Behavior recognition via sparse spatio-temporal features”, VS-PETS2005, 2005

[10] A.P.Perez *et al*,”High five : recognising human interactions in TV shows”,BMVC2010, pp.50.1-50.11, 2010.

[11] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc van Gool, ”You’ll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking”, ICCV2009, 2009

[12] Peter Kiefer, ”Mobile Intention Recognition”, Springer , 2011

[13] Zoraida Callejas, David Griol, and Ramon Lopez-Cozar, ”Predicting user mental states in spoken dialogue systems”, EURASIP Journal on Advances in Signal Processing 2011, 2011

[14] E.T.Hall,”The hidden dimension”,1966.

[15] T.Watanabe *et al*,”Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection”,PSIVT2009, pp37-47, 2009.

[16] Michael Isard and Andrew Blake, ”CONDENSATION - conditional density propagation for visual tracking”, Int. J. Computer Vision, 29, 1, 5-28, 1998

[17] B. D. Lucas and T. Kanade, ”An iterative image registration technique with an application to stereo vision”,Imaging Understanding Workshop, pp.121-130, 1981

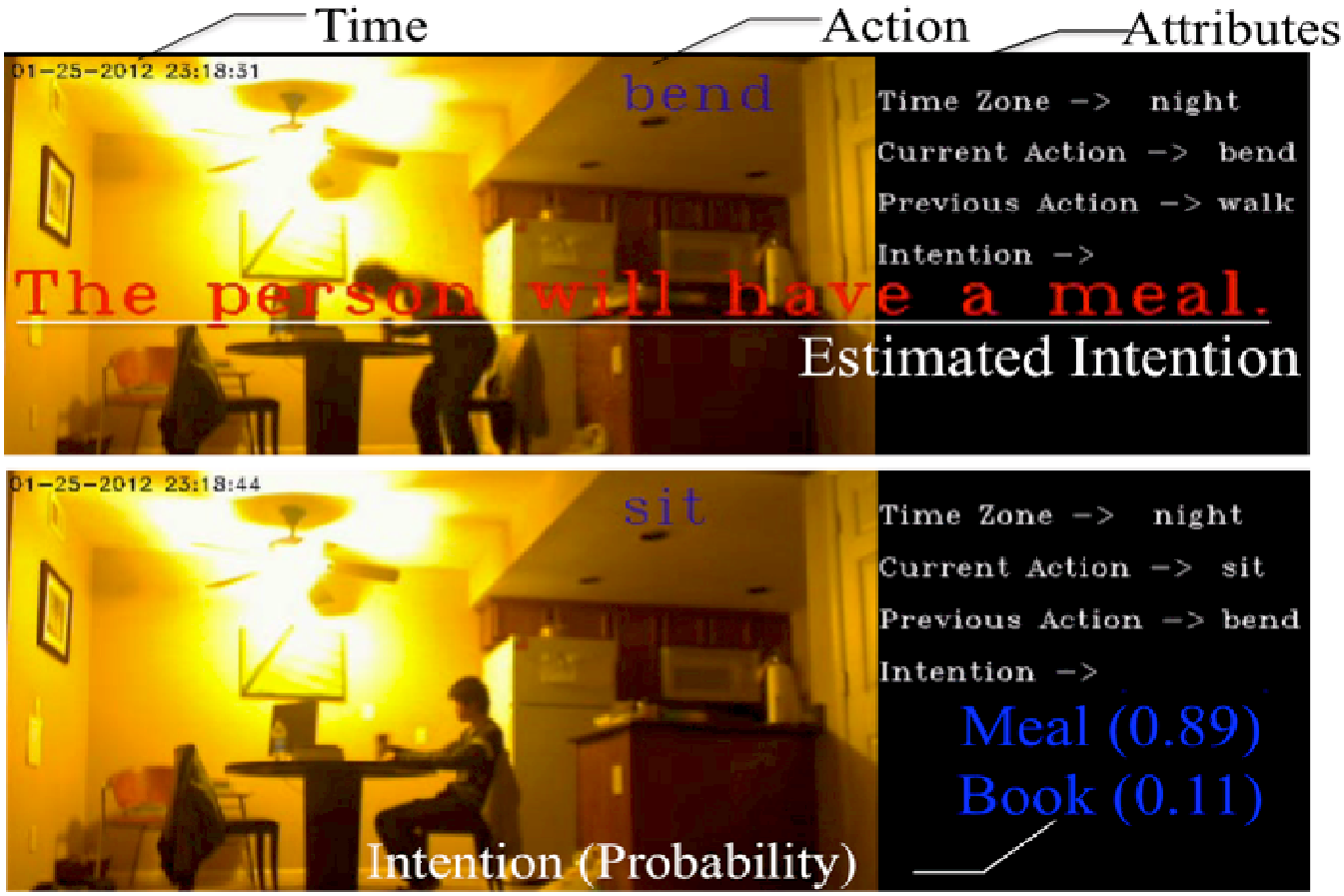

図 12 意図推定・行動予測：TimeZone=morning, day, night, PreviousAction,CurrentAction=bend, sit, stand, walk を入力として Intention=book, PC, meal を予測する．図中は TimeZone = night，PreviousAction = bend, CurrentAction = sit の場合であり，book, PC, meal の場合それぞれについての確率を求める．book の場合の確率は 0.11，meal は 0.89 であり，ランク付けの結果，”meal”が推定された意図となり，次の行動が予測された．なお，PC の確率は 0.0 のため，図中には表示されていない．

[18] Yoav Freund, Robert E. Schapire. ”A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, 1995
[19] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri, ”Actions as Space-Time Shapes”, ICCV2005, pp.1395-1402, 2005
[20] G. Denina, B. Bhanu, H. Nguyen, C. Ding, A. Kamal, C. Ravishankar, A. Roy-Chowdhury, A. Ivers, and B. Varda, ”VideoWeb Dataset for Multi-camera Activities and Non-verbal Communication”, Distributed Video Sensor Networks (Springer), 2010