

人物位置関係とステレオ精度の最適化によるエレベータ内の人物追跡

田口 進也[†]

[†]三菱電機株式会社 情報技術総合研究所 〒247-8501 神奈川県鎌倉市大船 5-1-1

E-mail: Taguchi.Shinya@aj.MitsubishiElectric.co.jp

あらまし 本稿では、二台のカメラをエレベータカゴ内部に設置して人物の移動軌跡を推定する手法について述べる。提案手法では長時間にわたる頭部の2次元軌跡同士をステレオ視することでステレオ視の曖昧さを回避し、3次元軌跡を推定する。さらに人物同士の位置関係と人物数、ステレオ視精度を考慮した尤度関数をMCMC (Markov Chain Monte Carlo) で最大化し、複数の3次元軌跡候補の中から最適な組み合わせを推定する。実験の結果180cm×180cmのエレベータ内に10人の乗客が存在して遮蔽が起こる場合でも、乗客全員の乗降履歴を推定できた。

キーワード エレベータ、人物追跡、頭部検出、ステレオ、マルチカメラ、MCMC

Tracking People in Elevator by Optimizing Trajectory Interaction and Stereo Precision

Shinya TAGUCHI[†]

[†] Mitsubishi Electric Corp. Information Technology R&D Center, 5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501 Japan

E-mail: Taguchi.Shinya@aj.MitsubishiElectric.co.jp

Abstract This paper describes a novel people tracking method inside crowded elevators for efficient elevator control. First, we propose a 3D trajectory estimation method based on stereo matching of 2D trajectories in order to avoid the “phantom problem”. Second, we propose a 3D trajectory selection method which maximizes a likelihood of position and stereo precision using MCMC. We can successfully track 10 people inside a 180cm x 180cm elevator even under severe occlusions.

Keyword Elevator, Tracking, Head Detection, Stereo, Multi-camera, MCMC

1. はじめに

エレベータの効率的な運行管理には「何人が何階で乗降したか」という乗客の乗降履歴が必要とされている。従来エレベータ内部の人流計測手法としては、赤外線センサを扉上部に設置して乗客の扉通過人数を計測する方法が実用化されているが、乗客の移動軌跡を求めることができなかった。

そこで本稿では、可視カメラ2台をエレベータ内部に設置し、乗客の移動軌跡を計測する手法を提案する。以下、第2章にて関連研究について述べ、第3章にて人物追跡アルゴリズムの詳細について述べる。第4章にて実エレベータでの実験結果を示す。

2. 関連研究

近年、単眼カメラによる混雑状況での人物追跡手法が多数提案されてきた[2][3][4][5][6][7][8]。しかし単眼カメラ映像を利用した追跡手法には原理的な限界がある。例えば同じ色の服を着た人物が交差した場合に、追跡結果に曖昧さが生じる問題がある。また対象エリアに鏡が設置されている場合、鏡像の人物を誤検出してしまふ。

このため、より正確な人物追跡には多視点カメラによる3次元位置情報の利用が有効である。従来多視点カメラを用いた人物追跡手法として、背景差分で得た人物領域を利用して動的計画法[9]やParticle Filter[10]で複数人物の軌跡を推定するものがある。これらの手法は、ある視点において人物が遮蔽されている状況でも、他の視点の人物領域や時系列情報を用いて軌跡を推定している。しかし混雑しているエレベータ内ではどの視点から撮影しても常に人物が重なり合うことがあるためこれらの手法を適用できない。また用いるカメラ台数は4～8台と実用的でない。一方、Adaboostで学習した顔検出器を用いて多視点カメラ映像から人物を検出し、Particle Filterで人物追跡を行なう手法[11][12]も提案されているが、混雑して遮蔽が起こる状況での軌跡推定方法に関しては言及されていない。

そこで本研究では、エレベータ内の景観を損なわないように天井コーナーに2台のカメラを設置し、混雑した状況でも人物頭部を検出・追跡して正確に乗降履歴を求める手法を開発したので以下に報告する。

3. アルゴリズム

3-1. 概要

提案手法は「人物頭部追跡」と「2次元軌跡のステレオ視」および「MCMCによる3次元軌跡推定」の処理から構成される。遮蔽が発生して人物頭部の追跡結果に曖昧さが生じた場合でも、人物同士の位置関係と人物数、ステレオ視精度を考慮した尤度関数をMCMCで最大化することで、正確な3次元軌跡推定を可能とした。以下、各処理の詳細について述べる。

3-2. 人物頭部追跡

まず各映像から人物頭部を検出・追跡し、図1(a)に示すように、2次元軌跡を求める。本手法は頭部検出と、カルマンフィルタ同様に予測と推定処理を繰り返すことで頭部追跡を行なう。

頭部追跡アルゴリズム

- (1) 初期化 ($t=0$) まず初期画像から頭部検出器により頭部候補の矩形領域を算出。
- (2) 追跡
for $t=1$ to T
 - 予測) 時刻 t の矩形領域をテンプレートとして、画像テンプレートマッチングにより時刻 $t+1$ の矩形領域の位置を予測。
 - 検出) 時刻 $t+1$ の頭部の矩形領域を検出。
 - 推定) 時刻 $t+1$ の予測矩形と検出矩形の重なる面積が一定以上の場合に、これらの矩形座標平均を計算して矩形座標を更新し、時刻 $t+1$ の矩形領域とする。また予測矩形の近傍に検出矩形が無い場合、予測矩形を時刻 $t+1$ の矩形領域とする。

end

ここで頭部検出器は、色やエッジの相関係数に基づいた判別器をAdaboostで学習することで得た[15]。また頭部の誤検出を削減するために、以下の条件を設けた。

追跡開始条件) 矩形座標が3回連続で更新されたら(つまり3回連続で頭部が検出されたら)対象矩形を頭部として受理し、そうでない場合は対象矩形を棄却する。
追跡終了条件) 受理された対象矩形を追跡中に、3回連続で矩形座標が更新されなければ追跡を終了する。

3-3. 2次元軌跡のステレオ視

ある瞬間の人物の3次元位置をステレオ視で求める場合、「phantom problem」[10]として知られている問題により、人物の位置を誤って推定してしまうことがある。そこで図1(b)に示すように、一定の時間内の2次元軌跡の点列同士をステレオ視することで

「phantom problem」を回避し、人物頭部の3次元軌跡を求める方法を提案する。

例えば、図1(a)に示す人物H1の、カメラAの2次元軌跡 $\alpha = \{x_\alpha(t)\}_{t=T1, \dots, T3}$ とカメラBの2次元軌跡 $\beta = \{x_\beta(t)\}_{t=T2, \dots, T4}$ が得られたとする。 $x_\alpha(t)$ 、 $x_\beta(t)$ は時刻 t における2次元の正規化カメラ座標である。各軌跡の時間関係の例を図2に示す。これらの2次元軌跡 α と β が時間 $[T2, T3]$ で同時に存在する場合、 $[T2, T3]$ 内の各時刻で対応する点列同士をステレオ視して3次元軌跡を求める。例えば図1(b)に示すように、座標 $x_\alpha(t)$ と $x_\beta(t)$ に対応するカメラから物体までの直線 $V_\alpha(t)$ と $V_\beta(t)$ を算出し、二つの直線を最短距離で結ぶ線分の距離 $d(t)$ を求め、その中点を3次元軌跡位置とする。さらに $[T2, T3]$ において距離 $d(t)$ が一定の距離(ここでは15cmと設定)以下となる回数を、 $[T2, T3]$ で割った値を二つの軌跡のマッチング率とし、マッチング率が一定の閾値以下の場合には3次元軌跡を棄却する。時間的な重なりが0である場合にはマッチング率は0とする。

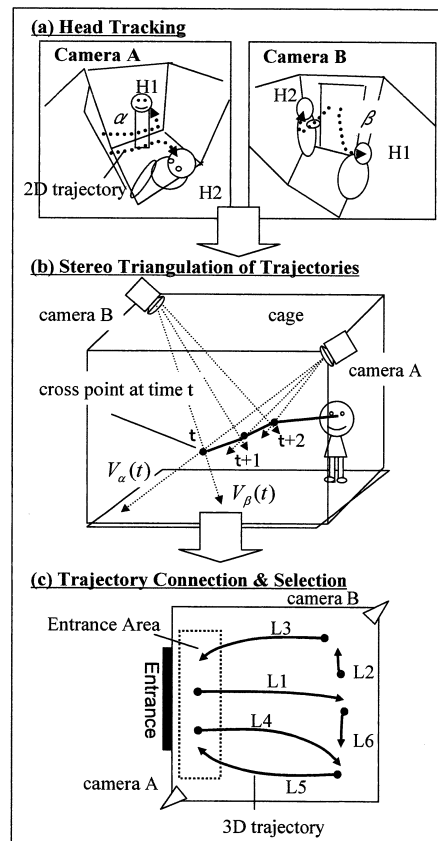


図1 提案手法概要

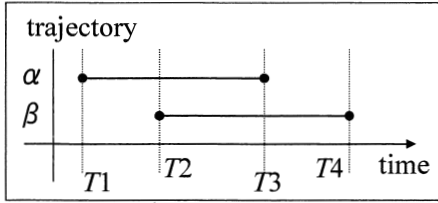


図 2 2次元軌跡の時間関係

さらにステレオ視で対象人物の高さ(身長)平均を推定し、軌跡が時間的に重なっていない時間[T1 T2]と[T3 T4]において、単眼立体視で3次元軌跡を推定する。このことで、片方のカメラに人物が写らない場合でも、その前後でステレオ視が成立していれば人物の3次元軌跡を算出可能である。

最後に、エレベータ内の入り口付近に入退場エリアを仮想的に設置し、3次元軌跡の断片を連結することで、入退場エリアから開始して入退場エリアで終了する3次元軌跡を列挙する。例えば、図1(c)は、天井から見た図であり、各3次元軌跡L1~L6が得られている。この場合、連結した軌跡{L1→L2→L3}、{L1→L6→L5}、{L4→L5}を3次元軌跡候補として列挙する。遮蔽や追跡時のドリフトが原因で、この処理段階では実際の人数よりも多くの3次元軌跡候補が列挙される。

3-4. MCMCによる3次元軌跡の推定

次に、人物同士の位置関係と、人物数、ステレオ視の精度を考慮した尤度関数をMCMCで最大化することで、列挙された3次元軌跡候補の中から最適な組み合わせを推定する。

3-4-1. モデル

前節で求めた入退場エリアから開始して入退場エリアで終了する3次元軌跡候補の集合を $Y=\{y_i\}_{i=1,\dots,N}$ とする。ここで N は軌跡の数である。また w_+ を正しい3次元軌跡の集合とし、 $w_-=Y-w_+$ 、 $w=\{w_+, w_-\}$ と定義する。目的は Y を w_+ と w_- に分割することであり、尤度関数

$$L(w|Y) = L_{ovr}(w|Y) L_{num}(w|Y) L_{str}(w|Y)$$

最大化する問題として定式化する。ここで L_{ovr} は人物位置関係に関する尤度関数、 L_{num} は人物数に関する尤度関数、 L_{str} はステレオ視精度に関する尤度関数である。以下、各尤度関数の詳細について述べる。

(a) 人物位置関係に関する尤度関数

人物の位置関係として、「3次元空間内で人物が重ならない」という条件を下記の通り定式化する。

$$L_{ovr}(w|Y) \propto \exp(-c_1 \sum_{i,j \in w_+} O(y_i, y_j))$$

ここで $O(y_i, y_j)$ は3次元軌跡 y_i と y_j の空間的重なりに関するコスト関数であり、 y_i と y_j が3次元空間で完全に重なる場合に1となり、まったく重ならない場合は0をとるものとする。また c_1 は正の定数である。ここでは $O(y_i, y_j)$ を次のように求める。 $y_1=\{y_i(t)\}_{t=T1,\dots,T2}$ 、 $y_2=\{y_j(t)\}_{t=T3,\dots,T4}$ とし、 y_1 と y_2 が時間 $F=[T3\ T2]$ で同時に存在しているとする。また

$$g(y_i(t), y_j(t)) = 1 \text{ (if } \|y_i(t) - y_j(t)\| < Th1), = 0 \text{ (otherwise)}$$

と定義する。ここで $Th1$ は適当な閾値であり、本稿では15cmに設定した。このとき、

$$O(y_i, y_j) = \sum_{t \in F} g(y_i(t), y_j(t)) / |F|$$

として空間的重なりに関するコストを求める。

(b) 人物数に関する尤度関数

人物数として「可能な限り多くの人物が存在する」という条件を

$$L_{num}(w|Y) \propto \exp(c_2 |w_+|)$$

と定式化する。ここで $|w_+|$ は w_+ の元の数である。ここで c_2 は正の定数である。

(c) ステレオ視精度に関する尤度関数

ステレオ視精度が高い軌跡が選択されるように

$$L_{str}(w|Y) \propto \exp(-\sum_{i \in w_+} S(y_i))$$

と定式化する。ここで $S(y_i)$ はステレオ視の精度に関するコスト関数であり、次のように計算する。軌跡 y_i の時間 F_i には、ステレオ視で軌跡が推定された時間 $F1_i$ と、単眼立体視で軌跡が推定された時間 $F2_i$ と、どのカメラからも観測されない時間 $F3_i$ の三つの期間が混在している。このとき

$$S(y_i) = (c_3 \times |F1_i| + c_4 \times |F2_i| + c_5 \times |F3_i|) / |F_i|$$

としてステレオコストを計算する。ここで c_3, c_4, c_5 は、 $0 \leq c_3 < c_4 < c_5$ を満たす定数である。

3-4-2. MCMCによる尤度最大化

以下では、MCMCにより尤度関数 $L(w|Y)$ を最大化する手法について述べる。アルゴリズムの概要を以下に示す。

MCMCアルゴリズム

入力: Y, w_{init}, N_{mc} 出力: w_{opt}

(1) 初期化 $w = w_{init}, w_{opt} = w_{init}$

(2) メインルーチン

for $n=1$ to N_{mc}

step1. $\zeta(m)$ に従って m をサンプリング

step2. m に従って提案分布 q を選択し、 w' をサンプリング

step3. 一様分布 $\text{Unif}[0\ 1]$ から u をサンプリング

step4. if $u < A(w, w'), w = w'$;

step5. if $L(w|Y)/L(w_{opt}|Y) > 1, w_{opt} = w;$

end

ここでアルゴリズムへの入力軌跡の集合 Y 、初期分割 w_{init} 、サンプリング回数 N_{mc} であり、また出力として最適な分割 w_{opt} を得る。初期化では、初期分割 $w_{init} = \{w_+ = \varnothing, w_- = Y\}$ とする。次に step1 では確率分布 $\zeta(m)$ に従って m をランダムにサンプリングする。ここでは $\zeta(m)$ は一様分布とした。次に、step2 では、インデックス m に対応する提案分布 $q(w'|w)$ に従って候補 w' をサンプリングする。提案分布として、「生成」、「消滅」、「交換」3つのタイプを定義した。インデックス $m=1$ は「生成」、 $m=2$ は「消滅」、 $m=3$ は「交換」に対応するものとする。次に、step3 では、一様分布 $\text{Unif}[0, 1]$ から u をサンプリングし、続く step4 では u と受理確率 $A(w, w')$ に基づいて候補 w' を受理または棄却する。ここで受理確率 $A(w, w')$ は次式で与えられる。

$A(w, w') = \min(1, q(w|w')L(w'|Y) / q(w'|w)L(w|Y))$
最後に step5 にて、尤度関数を最大化する最適な w_{opt} を保存しておく。

以下に提案分布 $q(w'|w)$ の概要について述べる。

(A) 生成 集合 w_- から一つの軌跡 y を選択して、 w_+ に加える。ここで y として、 w_+ にある軌跡と空間的に重ならないものがランダムに選択される。すなわち、 $y \in w_-$ 、 $w = \{w_+, w_-\}$ 、 $w' = \{w_+ + y, w_- - y\}$ 、とすると、提案分布は

$q(w'|w) \propto \zeta(1) \exp(-c_6 \sum_{j \in w_+} O(y, y_j))$
で与えられる。ここで $O(y, y_j)$ は前述の空間的重なりに関するコストであり、軌跡 y と y_i と完全に重なっている場合に 1 となり、まったく重なっていない場合は 0 をとるものとする。ここで c_6 は正の定数である。

(B) 消滅 集合 w_+ から一つの軌跡 y を選択して w_- に加える。ここで y として、 w_+ の内部で互いに空間的に重なる軌跡がランダムに選択される。すなわち、 $y \in w_+$ 、 $w = \{w_+, w_-\}$ 、 $w' = \{w_+ - y, w_- + y\}$ とすると、提案分布は

$q(w'|w) \propto \zeta(2) \exp(c_7 \sum_{j \in w_+} O(y, y_j))$
で与えられる。なお w_+ が空集合の場合は、
 $q(w'|w) = 1$ (if $w' = w$)、 $q(w'|w) = 0$ (otherwise)
とする。ここで c_7 は正の定数である。

(C) 交換 集合 w_+ から前述のステレオコスト S が高い軌跡 y を優先して選択し、また集合 w_- からステレオコスト S が低い軌跡 z を優先して選択し、 y と z を交換する。すなわち $y \in w_+$ 、 $z \in w_-$ 、 $w' = \{w_+ - y + z, w_- + y - z\}$ 、 $p(y|w) \propto \exp(c_8 S(y))$ 、 $p(z|w, y) \propto \exp(-c_8 S(z)) \exp(c_9 O(z, y))$ 、とすると、提案分布は

$$q(w'|w) \propto \zeta(3) \times p(z|w, y) p(y|w)$$

で与えられる。ここで c_8 、 c_9 は正の定数である。

4. 実験

180cm×180cmのエレベータ内に2台のカメラを設置し、合計15人(10人、1人、4人)が乗降する状況で人物追跡を試みた。評価映像はQVGA、15fpsで撮影し、総フレーム数は5070(338秒)である。図3(A)、(B)は魚眼レンズで撮影した各カメラ映像の例である。カゴ内部では人物が近接しており、またカメラを天井コーナーに設置しているため、例えば図4の矢印に示す人物のように、ある人物が他の人物に遮蔽されてしまうことが多い。

評価映像をパッチ処理して列挙された3次元軌跡候補の数 $|Y|$ は1138個であり、この中からMCMCにより15個の3次元軌跡が選択された。10人乗車時の2次元追跡結果を図3(A)、(B)に示す。図中の矩形は追跡中の人物頭部を示しており、誤検出(false positive)も存在している。同時刻における3次元軌跡推定結果を図3(C)に示す。図3(C)はエレベータ内を天井から見た図であり、MCMCにより選択された10人の3次元軌跡と頭部位置、各カメラの位置、入口(Entrance)、エレベータサイズ(Cage)、入退場を判断するエリア(Entrance Area)が示されている。図中スケールはmmである。また図3(D)に10人の頭部位置を円柱で示す。頭部誤検出や遮蔽により複数の3次元軌跡候補が列挙される場合でも、尤度関数の最大化により正確な人数分の3次元軌跡を選択することができた。

推定したエレベータ内の人数の時間変化を図5に示す。横軸は映像フレーム番号、縦軸は人数であり、点線が正解人数(映像目視で算出)、実線が推定人数を示している。推定人数の1フレームあたりの平均誤差は0.05人であり、ほぼ正確な人数の時間変化を推定できている。また乗客の乗降履歴を表1に示す。ドア開インデックス(Door Open Index, 以下DOIと略す)とは、映像中でドアが開閉した順番である。例えば、表1の1列目は、DOI=1で入場(Enter)しDOI=2で退場(Leave)した人数の推定値(Estimated)と正解値(Correct)を示している。提案手法により各DOIで正確な移動人数が推定できた。

5. まとめ

エレベータ内に最大10人の乗客が存在して遮蔽が起る場合でも、乗客の3次元軌跡を正確に推定できた。この推定結果と各時刻のカゴ停止階情報を組み合わせれば、ビル内の乗客乗降履歴情報を得て、効率的なエレベータ運行管理が可能となる。

本手法の課題は、乗客数が増えると、入場から退場までの3次元軌跡列挙処理に時間がかかることである。このため人物追跡を3次元追跡断片のラベリング問題に帰着させて、最適なラベリングをMCMCにて推定することで、実用的な時間内で処理を行なう手法についても検討中である。

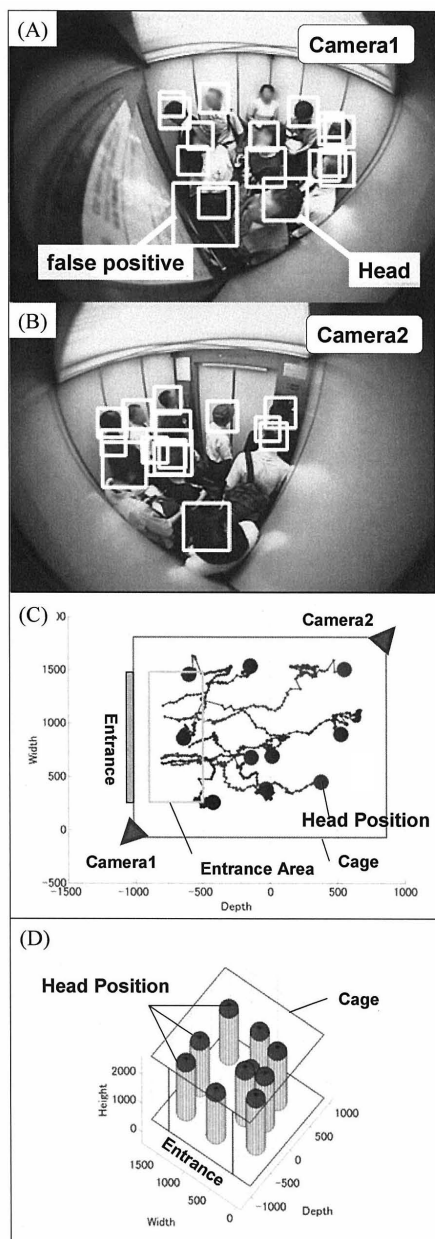


図 3 軌跡推定結果



図 4 評価映像における遮蔽の例

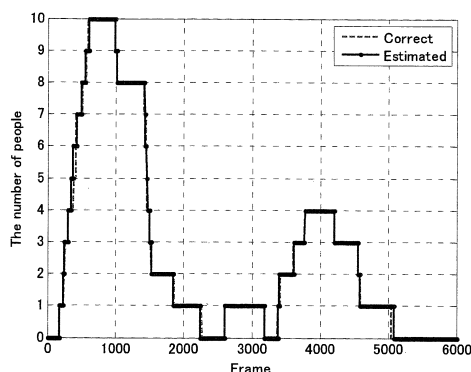


図 5 エレベータ内人数の推定結果.

表 1 乗降履歴.

Door Open Index		The number of People	
Enter	Leave	Estimated	Correct
1	2	2	2
1	3	6	6
1	4	1	1
1	5	1	1
6	7	1	1
8	9	1	1
8	10	2	2
8	11	1	1

文献

- [1] O. Tuzel, F. Porikli, and P. Meer, "Human Detection via Classification on Riemannian Manifolds," CVPR, pp.1-8, 2007.
- [2] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on Lie Algebra," CVPR, vol.1, pp.511-518, 2006.
- [3] Z. Tao, and R. Nevatia, "Bayesian Human Segmentation in Crowded Situations," CVPR, vol.2, pp.459-466, 2003.
- [4] G. J. Brostow, and R. Cipolla, "Unsupervised Bayesian Detection of Independent Motion in Crowds," CVPR, vol.1, pp.594-601, 2006.
- [5] V. Rabaud, and S. Belongie, "Counting Crowded Moving Objects," CVPR, vol.1, pp.705-711, 2006.
- [6] Yu Qian, G. Medioni, and I. Cohen, "Multiple Target Tracking Using Spatio-Temporal Markov Chain Monte Carlo Data Association," CVPR, vol.17, pp.1-8, 2007.
- [7] O. Songhai, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," Decision and Control, vol.1, pp.735-742, 2004.
- [8] K. Ishiguro, T. Yamada, and N. Ueda, "Simultaneous clustering and tracking unknown number of objects," CVPR, pp.1-8, 2008.
- [9] J. Berclaz, F. Fleuret, and P. Fua, "Robust People Tracking with Global Trajectory Optimization," CVPR, vol.1, p.744-750, 2006.
- [10] K. Otsuka, and N. Mukawa, "A particle filter for tracking densely populated objects based on explicit multiview occlusion analysis," ICPR, vol.4, pp.745-750, 2004.
- [11] 小林 貴訓, 杉村 大輔, 平澤 宏祐, 鈴木 直彦, 鹿毛 裕史, 佐藤 洋一, 杉本 晃宏, "パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡," 信学論 vol.J90-D. no.8, pp.2049-2059, 2007.
- [12] Z. Zhenqiu, G. Potamianos, L. Ming, and T. Huang, "Robust Multi-View Multi-Camera Face Detection inside Smart Rooms Using Spatio-Temporal Dynamic Programming," Proc. of the 7th International Conference on Automatic Face and Gesture Recognition, pp.407-412, 2006.
- [13] R. Schapire, and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, vol.37, no.3, pp.297-336, 1999.
- [14] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," CVPR, vol.1, pp.511-518, 2001.
- [15] 田口進也, 田中昭二, "頭部検出器と多視点カメラを用いた人物検出方式の検討," 信学技法 (IE2007-262), vol.107, no.538, pp.7-10, 2008.