

標準歩容モデルとの照合に基づく歩行者領域抽出

田上 拓弥^{1,a)} 槇原 靖^{1,b)} 村松 大吾^{1,c)} 八木 康史^{1,d)}

概要: 歩容認証は、歩き方から個人認証を行う技術で犯罪捜査などに用いられている。既存の歩容認証は、衣服の柄や照明の変化などに影響をうけない人物のシルエットに基づく特徴を用いて認証する手法が主流である。シルエットの抽出は前処理として行われ、安定したシルエットが得られているという前提がある。しかし、実際の環境下において、安定したシルエットの抽出を行うことは容易ではなく、正しいシルエットを得られない場合には認証精度劣化の原因となる。そのため、歩容認証では、さまざまな状況下で安定してシルエットの抽出を行うことが課題となっている。そこで、本稿では、歩容認証のための安定した人物の領域分割を行う手法を提案する。提案手法では、既存の背景差分に基づくグラフカットを用いた領域分割に、歩容の事前知識である歩容の標準モデルを用いることでシルエットが改善されることを示す。

1. はじめに

人の身体的・行動的特徴を用いて個人を認証する手法（生体認証）の研究が近年盛んに行われている。例えば、静脈認証や、指紋認証、虹彩認証、顔認証などが一般的によく知られている手法である。生体認証は様々な状況での個人認証手法として活用が期待されており、認証対象者の認証に対する意識の違いにより、活用が期待される状況は二種類に大別することができる。一つは携帯端末へのログオンや施設の入室管理に代表される、認証対象者が認証されることを認識しており、かつ協力的な状況であり、もう一つは、犯罪捜査や監視に代表される、認証対象者が認証に対し無意識であったり、非協力的であったりする状況である。前者の状況では、対象人物が認証用の動作（例えばセンサに指をかざす等）を行ってくれるため、認証に必要なデータを比較的理想的な環境下で取得できる。一方、後者の状況では、対象人物は認証用の動作を行うことはないため、観測できるデータのみから認証を行う必要があり、そのため、利用可能なデータが限定され、かつデータ取得環境も理想的でない場合が多い。本稿では後者の状況の一つである、路上を歩く人物を対象とし、防犯カメラ映像を用いて個人を認証する手法に焦点をあてる。

図 2 に防犯カメラにより撮影された歩行者映像の例を示す。防犯カメラ映像中の対象人物は空間解像度が低いこと

多く、鮮明な顔情報が得られない場合が多い。また顔の向きによっては顔自体が観測できない。そのため、顔認証の適用が困難となる場合が頻繁に生じる。そのような場合にも認証を可能とする手法として、人の歩く姿から歩容特徴を抽出し認証を行う歩容認証が提案されている [1]。

歩容認証手法の中でも、特に歩行時の見えに着目したアピアランスに基づく歩容特徴 [2], [3], [4] は、遠隔から撮影された歩行映像からも取得が可能であり、また空間解像度の低い映像（例えば映像中での人物高さが 30 ピクセル程度）から抽出された特徴でも個人識別能力を有するという優れた長を有しているため、防犯カメラ映像を対象とした個人認証手法として有望な手法である。しかしながら、アピアランスに基づく歩容特徴を用いた手法の認証精度は、特徴抽出に用いられるシルエットの精度に大きく影響を受けてしまうため [5]、高い認証精度を実現するためには、映像から高い精度でシルエットを抽出する手法が必要となる。シルエット抽出にはグラフカット [6] や時空間的に局所的な領域に対して色変換を適用した背景差分に基づく手法 [7] などが提案されており、比較的理想的な環境下で撮影された映像に対して精度改善を実現しているが、一般的な実環境下の防犯カメラ映像に対する有効性は限定的である。

そこで、本研究では、より一般的な防犯カメラ映像に対し、シルエット抽出精度を改善する手法を提案する。提案手法では、グラフカットに基づくシルエット抽出結果に対し、標準歩容モデルをあてはめることでシルエットを修正し、シルエット抽出精度を改善する。

¹ 大阪大学
Osaka University

a) tanoue@am.sanken.osaka-u.ac.jp

b) makihara@am.sanken.osaka-u.ac.jp

c) muramatsu@am.sanken.osaka-u.ac.jp

d) yagi@am.sanken.osaka-u.ac.jp

2. 問題設定

本研究では、異なる場所に設置された防犯カメラで撮影された歩行者の個人認証を行う。防犯カメラ映像からの歩行者認証は、「歩行者検出」、「シルエット抽出」、「特徴抽出」、「照合」の処理からなるが、本稿では、シルエット抽出処理による精度改善に焦点をあてるため、歩行者は検出済である（対象歩行者の矩形領域は事前に与えられる）ことを前提としている。また歩行者の歩行方向は一定であると仮定し説明を行う^{*1}。

3. 提案手法

3.1 グラフカット領域分割

グラフカット領域分割 [9] は、一連の画像列における領域分割対象の画素をサイト (グラフのノード) q として表現し、各サイト q に割り当てるラベル X_q (FG: 前景, BG: 背景) に応じて定義される以下の関数の最小化問題として定式化されている。

$$E(X) = \sum_{q \in Q} g_q(X_q) + \sum_{(p,q) \in E} h_{pq}(X_p, X_q) + \sum_{q \in Q} g_s(X_q) \quad (1)$$

ここで、 X はサイトの集合 Q に対する前景と背景ラベルの組み合わせを表し、 E は全ての隣接サイトの集合を表す。ここで、隣接サイトとしては、空間的に隣接している画素に加えて、時間方向においてはオプティカルフロー (実装には Lucas-Kanade 法 [10] を利用) によって対応が得られた画素間を隣接サイトとしている。上式において、第1項はデータ項、第2項は平滑化項、第3項は標準歩容モデル項であり、各々の定義は次節以降にて行う。

3.2 データ項

データ項は、ラベルが背景の場合の背景データ項 $g_q(X_q = BG)$ と、ラベルが前景の場合の前景データ項 $g_q(X_q = FG)$ で表現される。

まず、背景データ項については、通常背景差分に基づいて設定する。具体的には、複数枚の背景画像を用いて、各サイト q にする背景色の平均 $\mu_{bg,q}$ と共分散行列 $\Sigma_{bg,q}$ を算出し、各サイト q の色 c_q に対するマハラノビス距離 $d_{bg,q}$ を算出する。

$$d_{bg,q} = (c_q - \mu_{bg,q})^T \Sigma_{bg,q}^{-1} (c_q - \mu_{bg,q}) \quad (2)$$

更に、マハラノビス距離に基づいて、背景データ項を以下のように設定する。

$$g_q(X_q = BG) = w_{bg} \exp(-\kappa_{bg} d_{bg,q}) \quad (3)$$

^{*1} カメラが校正済であれば、歩行軌跡とカメラ校正パラメータから標準歩容モデルを方向変換モデル [8] により変換することで、提案手法を歩行方向が異なる認証問題にも適用可能である。



図1 標準歩モデルの例

ここで、 w_{bg} , κ_{bg} は、背景データ項の重みを制御するハイパーパラメータである。本稿では、カメラ毎に背景データ項の重みを設定しているため、これらのハイパーパラメータは実験の章で記す。

次に、前景データ項については、前景領域の色を混合ガウス分布 (Gaussian mixture model: GMM) によってモデル化することで、表現する。具体的には、まず、背景データ項に対して閾値処理を施して初期前景領域を抽出する。次いで、初期前景領域に対してクロージング、オープニング処理を含むモルフォロジー処理を適用し、前景候補領域を作成する。更に、その前景候補領域に含まれる色を前景色のサンプルとして、 k 平均法を適用することで、混合ガウス分布を学習する。ここで、混合ガウス分布の k 番目の要素に対する平均と共分散行列を $\mu_{fg,k}$, $\Sigma_{fg,k}$ とすると、 k 番目の要素に対するマハラノビス距離 $d_{fg,k}$ は以下のように定義される。

$$d_{fg,k} = (c_q - \mu_{fg,k})^T \Sigma_{fg,k}^{-1} (c_q - \mu_{fg,k}) \quad (4)$$

すると、前景データ項 $g_q(X_q = FG)$ は、最小マハラノビス距離を与える要素に対する距離を用いて、以下のように定義される。

$$g_q(X_q = FG) = w_{fg} \exp\left(-\kappa_{fg} \min_k d_{fg,k}\right) \quad (5)$$

ここで、 w_{fg} , κ_{fg} は 前景データ項の重みを制御するハイパーパラメータである。今回は実験的に $w_{fg} = 1.0$, $\kappa_{fg} = 0.3$ と設定している。

3.3 平滑化項

平滑化項は、隣接二点間での前景・背景ラベルの変化を抑制する項である。本論文では、特に、隣接二点間で色の変化が小さい場合に、ラベル変化のペナルティを大きくかけるよう、以下のように平滑化項を設計した。

$$h_{pq}(X_p, X_q) = \begin{cases} 0 & (X_p = X_q) \\ w_{sm} \exp\left(-\kappa_{sm} \frac{|c_q - c_p|^2}{|c_q + c_p|^2 + \epsilon}\right) & (X_p \neq X_q), \end{cases} \quad (6)$$

ここで、 w_{sm} , κ_{sm} , ϵ は、平滑化項の重みを制御するハイパーパラメータである。本稿ではそれぞれ、 $w_{sm} = 1.0$, $\kappa_{sm} = 0.01$, $\epsilon = 63.0$ と設定している。

3.4 標準歩容モデル項

本節では、まず、入力画像と標準歩容モデルの照合について、その概要を説明する。(詳細については、[6] を参照されたい。)

まずは、標準歩容モデルを、 m 番目 ($m = 1, \dots, M$) の学習用被験者の一歩行周期を含む高品質シルエットの歩容系列から、高さを正規化して位置合わせをしたシルエット系列 $\mathbf{g}_m(\phi) \in \mathbb{R}^{W_g \times H_g}$ として生成する (図 1)。ここで、 ϕ は歩容の位相を、 W_g と H_g は正規化した画像の高さと幅を表す。

次に、背景データ項から前景らしさを表す画像として、

$$f_q = 1 - \exp(-\kappa_{bg} d_{bg,q}) \quad (7)$$

を定義し、改めてその n フレーム目の画像を $\mathbf{f}(n) \in \mathbb{R}^{W \times H}$ として定義し直す。ここで、 W, H はそれぞれ画像の幅と高さである。また、追跡や検出の結果として与えられると想定する外接矩形の中心の座標を (c_x, c_y) 高さを h とし、その外接矩形で正規化した画像を $\mathbf{f}_N(n; c_x, c_y, h) \in \mathbb{R}^{W_g \times H_g}$ として表す。

ここで、サイズ正規化後の前景らしさの画像 $\mathbf{f}_N(n; c_x, c_y, h)$ と標準歩容モデル $\mathbf{g}_m(\phi)$ を動的計画法 (Dynamic programming: DP) により照合することを考える。

まず、各フレーム・各状態における画像の照合には、谷本距離 [11] を用いる。前景らしさの画像 $\mathbf{f}_N(n; c_x, c_y, h)$ と、標準歩容モデル $\mathbf{g}_m(\phi)$ の谷本距離を以下の以下のように定義する。

$$D_T(\mathbf{f}_N(n; c_x, c_y, h), \mathbf{g}_m(\phi)) = 1 - \frac{\sum_{(x,y)} \min\{f_N(n, x, y; c_x, c_y, h), g_m(\phi, x, y)\}}{\sum_{(x,y)} \max\{f_N(n, x, y; c_x, c_y, h), g_m(\phi, x, y)\}} \quad (8)$$

また、DP における探索空間である状態量を $\mathbf{x} = (\phi, \mathbf{s})$ と表し、 $\mathbf{s} = (s_x, s_y, s_h)$ は、それぞれ、 x 方向の並進、 y 方向の並進、スケールを各々表すものとする。そこで、そのような並進とスケールを施した前景らしさの画像を $f_Q(n; \mathbf{s})$ として定義する。すると、ある状態における入力的前景らしさ画像で標準歩容モデルの相違度に関するコストが

$$C_{sil}(n, \mathbf{x}) = D_T(\mathbf{f}_Q(n, \mathbf{s}), \mathbf{g}(\phi)) \quad (9)$$

として定義される。

まず、各フレーム n に対する状態を $\mathbf{x}(n)$ 、DP の n フレーム目における各状態に対する累積コストを $C_{DP}(n, \mathbf{x}(n))$ として定義し、DP の初期フレームにおけるコストを以下によって初期化する。

$$C_{DP}(1, \mathbf{x}(1)) = C_{sil}(1, \mathbf{x}(1)), \forall \mathbf{x}(1) \quad (10)$$

次いで、DP コストを漸化式の要領で以下のように更新していく。

$$C_{DP}(n, \mathbf{x}(n)) = C_{sil}(n, \mathbf{x}(n)) + C_{trans}(n, \mathbf{x}_p^*(n-1; \mathbf{x}(n)), \mathbf{x}(n)), \quad (11)$$

ここで、遷移コスト C_{trans} は以下のように表される。

$$C_{trans}(n, \mathbf{x}(n-1), \mathbf{x}(n)) = C_{DP}(n-1, \mathbf{x}(n-1)) + C_{smt}(\mathbf{x}(n-1), \mathbf{x}(n)), \quad (12)$$

$$C_{smt}(\mathbf{x}(n-1), \mathbf{x}(n)) = \alpha |\min\{\delta\phi, N_{gait} - \delta\phi\}|, \quad (13)$$

$$\delta\phi = |\phi(n) - (\phi(n-1) + v_\phi)|, \quad (14)$$

ここで、 v_ϕ は平均的な位相進行速度を表しており、標準歩容モデルと入力画像が同じフレームレートで撮影されている場合には、 $v_\phi = 1$ とする。また、 α はその平均位相進行速度からずれた場合のペナルティの重みを表すハイパーパラメタであり、実験的に 0.05 として設定している。

また、 $\mathbf{x}_p^*(n-1; \mathbf{x}(n))$ は、現在の状態 $\mathbf{x}(n)$ に遷移するのに最も良い 1 フレーム前の状態を表しており、以下のよう求められる。

$$\mathbf{x}_p^*(n-1; \mathbf{x}(n)) = \arg \min_{\mathbf{x}(n-1) \in X(n-1, \mathbf{x}(n))} \{C_{trans}(n, \mathbf{x}(n), \mathbf{x}(n-1))\} \quad (15)$$

ここで、 $X(n-1, \mathbf{x}(n))$ は、 $\mathbf{x}(n)$ への遷移が可能な $n-1$ フレーム目の状態の集合 \mathbf{w} を表し、一定幅の位相・並進・スケールのみが許容されるように設定するものとする。

以上により定義される累積コストについて、最終フレームにおける最適値を見つけ、それをバックトラックすることで最適パス及び最適コストを計算することができる。また、この処理を他の標準歩容モデルについても独立に実施し、最終的に最も累積コストを小さくする標準歩容モデルと、それに対応する最適パスを選択するものとする。ここで、最適な標準歩容モデルと最適パス (最適な状態の系列) が求めると、入力的前景らしさに対して、標準歩容モデルを重畳して描画できることに注意されたい (図 6(c) 参照)。

そのようにして得られた各フレームに対する歩容モデルのシルエットに対して、符号付き距離場 $d_{Q,sh}$ を計算し、標準歩容モデル項の計算を行う。

$$g_{sh}(X_q = BG) = w_{sh} \frac{1}{1 + \exp(-\kappa_{sh} d_{sh})} \quad (16)$$

$$g_{sh}(X_q = FG) = w_{sh} \frac{1}{1 + \exp(\kappa_{sh} d_{sh})} \quad (17)$$

ここで、 κ_{sh}, w_{sh} は標準歩容モデル項の重みを表すハイパーパラメタである。本稿では、 $\kappa_{sh} = 0.2, w_{sh} = 0.3$ に設定している。

4. 実験

4.1 データセット

実験には、小学校に設置された 2 台の防犯カメラで撮影した動画像を用いた。各動画像のスナップショットを図 2 に示す。カメラ A とカメラ B では歩行者の歩行方向が異

表 1 小学校の映像の諸元

| | |
|---------------|-----------------------------|
| 解像度 [画素] | 800 × 450 |
| フレームレート [fps] | 12 |
| 歩行者の高さ [画素] | 約 100 (カメラ A), 約 80 (カメラ B) |



図 2 小学校の映像のスナップショット

| ID | 11 | 12 | 24 | 33 | 44 |
|------|----|----|----|----|----|
| カメラA | | | | | |
| カメラB | | | | | |

図 3 対象となる人物の例

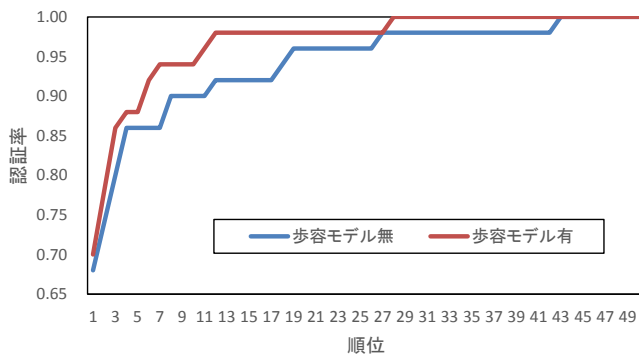


図 4 CMC 曲線

なるため、歩行方向を揃えるため、カメラ A で撮影された動画は回転させて用いた。動画の詳細は表 1 に示す。カメラ A、カメラ B で撮影された歩行者 50 人に対して 50 対 50 の認証を行う。対象となる歩行者の例を図 3 に示す。また、今回の実験で用いた標準歩容モデルは、身長や歩き方の異なる 6 人から、マニュアル操作で作成したものをカメラ A、カメラ B の合計 12 パターンを用意し使用した。また、カメラ A では背景項のハイパーパラメータを $\kappa_{bg} = 0.02$ 、カメラ B では $\kappa_{bg} = 0.01$ とカメラ毎に設定して実験を行った

4.2 性能評価

歩容モデルを用いて得られたシルエットと、歩容モデルを用いずに得られたシルエットの比較を行い性能を評価する。まず、得られたシルエット列から、標準歩容モデルの作成時と同様に、高さを正規化して位置合わせをした正規



図 5 1 対多認証結果の例

化シルエット列を作成する。次に正規化シルエット列から周期検出を行い、歩容特徴として、平均シルエットである GEI [2] を作成する。最後に、周期ごとの距離の最小値を最終的な特徴量とする。

評価の指標として、1 対多認証で一般的に使われている累積識別精度特性 (Cumulative match characteristic: CMC) 曲線を用いる。CMC 曲線は、ある入力特徴と N 人の登録特徴の照合で得られた N 個の相違度を元に順位付けを行い、正解人物が各順位までに含まれる確率を示したグラフである。

今回の実験で得られた CMC 曲線は図 4 に示す。歩容モデルを用いたものは歩容モデルを用いなかったものに比べて認証率が向上していることが見られる。認証率が 100% になる順位も標準歩容モデル有がランク 28 に対して歩容モデル無はランク 43 となり、性能の改善が確認できる。また、標準歩容モデルを用いた 1 対多認証の結果の例を図 5 に示す。

5. 考察

まず、歩容モデルを用いることによる改善部として、足の部分でのシルエット改善が大きいと考えられる。背景差分に基づいたシルエット抽出では、足などの細かい部分の抽出は難しいとされている。例えば今回のように背景の色と

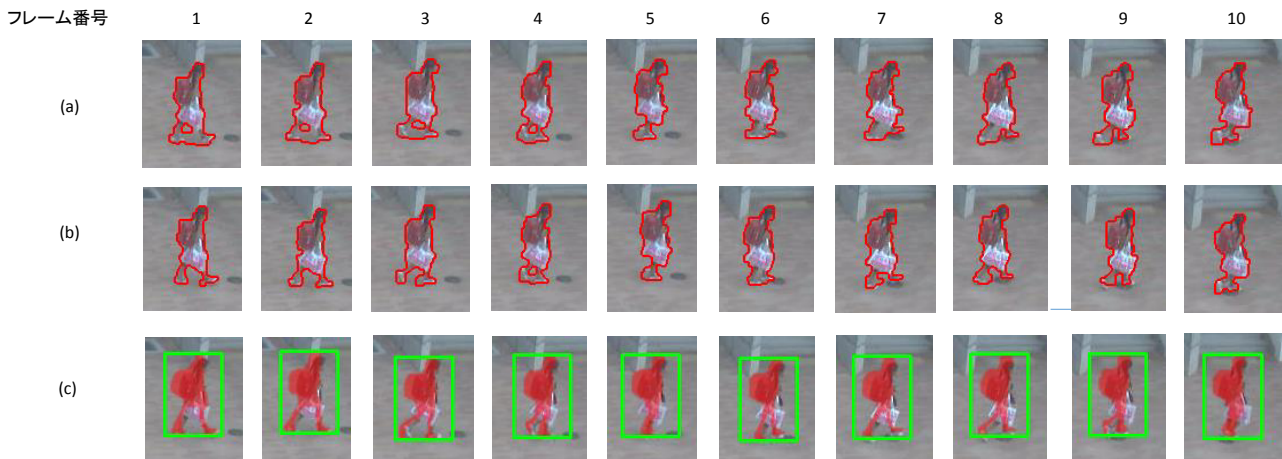


図 6 領域分割結果と照合されたモデル
(a) 歩容モデル無 (b) 歩容モデル有 (c) 照合されたモデル

足の部分の色が近い場合はアンダーセグメンテーションとなったり、カメラ B で撮影された動画のように足元に影ができたりすることで、オーバーセグメンテーションになったり。このような状況で歩容のモデルを当てはめることにより、シルエットの改善が行える。一つの改善例を図 6 で示す。図 6 では (a) は歩容モデル無の領域分割結果を表し、(b) は歩容モデル有の領域分割結果、(c) は照合された歩容モデルを原画像上に赤色で重ね合わせた画像である。このように歩容モデルを用いなかった場合、足の部分の領域分割は、オーバーセグメンテーションになってしまっているが、標準歩容モデルを用いることによって領域分割結果が改善されているのが見られる。

次に、標準歩容モデルを用いた領域分割結果の改善を定量評価を行い示す。定量評価の指標として F 値を用いる。 F 値は一般的に領域分割に用いられる指標で以下のように定義される

$$F = \frac{2PR}{P+R} \quad (18)$$

$$P = \frac{T_p}{T_p + F_p} \quad (19)$$

$$R = \frac{T_p}{T_p + F_n} \quad (20)$$

ここで、 T_p は真陽性の要素数 (例えば、ある画素で真値と入力がどちらも前景の場合、この画素は真陽性となる) を表し、 F_p は偽陽性、 F_n は偽陰性を表す。また、 P と R はそれぞれ精度と再現率を表し、それぞれトレードオフの関係にあることから、 F 値がそのトレードオフを表す一つの指標として用いられる。(高いほど領域分割の精度が良いことを意味する。)

図 6 に示す実験結果に対する F 値の遷移を図 7 に示す。 F 値が改善しておりシルエットが実際に改善していることが分かる。特に歩容モデルを用いなかった場合足部分がつながってしまっている、1, 2, 7 フレーム目では、歩容モデルを用いた場合の F 値の増加が顕著に見られる。次

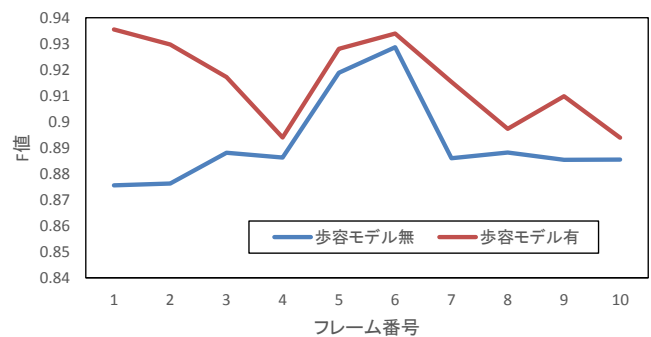


図 7 F 値

に、より精度を向上させるためには、標準歩容モデルとの照合がより正しく行われる必要がある。今回の実験では、標準歩容モデルは、6 人分のもののみを用いたが、この人数では、照合を行う時に、対応する姿勢などが含まれていないものもあるため、より多くの標準歩容モデルを用意する必要がある。図 7 の場合、適合率の高い標準歩容モデルを使用しており、シルエットの改善も大きい。そのため、標準歩容モデルの適合率は今回の提案手法において、重要であると考えられる。

6. おわりに

本論文では、グラフカットを用いた領域分割に標準歩容モデルを適用することによって、シルエットの改善、そして、歩容認証において精度が向上することを示した。特に足のような細い部分でのシルエットの改善が大きいことを示した。

今後の課題としては、より精度を上げるために、より多くの標準歩容モデルを用いての実験、様々な状況下における実験を行うことが挙げられる。

参考文献

[1] Nixon, M. S., Tan, T. N. and Chellappa, R.: *Human Identification Based on Gait*, Int. Series on Biometrics,

- Springer-Verlag (2005).
- [2] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322 (2006).
 - [3] Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T. and Yagi, Y.: Gait Recognition Using a View Transformation Model in the Frequency Domain, *Proc. of the 9th European Conf. on Computer Vision*, Graz, Austria, pp. 151–163 (2006).
 - [4] Lam, T. H. W., Cheung, K. H. and Liu, J. N. K.: Gait flow image: A silhouette-based gait representation for human identification, *Pattern Recognition*, Vol. 44, pp. 973–987 (2011).
 - [5] Hofmann, M., Bachmann, S. and Rigoll, G.: 2.5D gait biometrics using the Depth Gradient Histogram Energy Image, *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pp. 399–403 (online), DOI: 10.1109/BTAS.2012.6374606 (2012).
 - [6] Wang, J., Makihara, Y. and Yagi, Y.: Human tracking and segmentation supported by silhouette-based gait recognition, *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 1698–1703 (online), DOI: 10.1109/ROBOT.2008.4543445 (2008).
 - [7] Makihara, Y. and Yagi, Y.: Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation, *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4 (2008).
 - [8] Muramatsu, D., Shiraishi, A., Makihara, Y. and Yagi, Y.: Arbitrary View Transformation Model for Gait Person Authentication, *Proc. of the 5th IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, No. Paper ID 39, Washington D.C., USA, pp. 1–6 (2012).
 - [9] Boykov, Y. and Jolly, M.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images, *Proc. Int. Conf. on Computer Vision* (2001).
 - [10] Lucas, B. D. and Kanade, T.: An iterative image registration technique with an application to stereo vision, *Proceedings of the 1981 DARPA Image Understanding Workshop*, pp. 121–130 (1981).
 - [11] Jr., K. R. S. and Tanimoto, S.: Progressive Refinement of Raster Images, *IEEE Transactions on Computers*, Vol. 28, No. 11, pp. 871–874 (1979).