歩行の向きに依存しない多視点人物歩容認識におけるカメラ配置の検討

朱 暁東，高橋 友和，井手 一郎，村瀬 洋
名古屋大学 大学院情報科学研究科

カメラを用いた人物歩容認識では，学習時と認識時のカメラに対する人物の歩行の向きの違いが精度を低下させる要因となる．そこで，本稿では，人物の歩行の向きに依存しない歩容認識の実現に向けて，複数のカメラを用いた人物歩容認識手法におけるカメラの配置と認識率の関係を調査する．本研究では歩容辞書の作成に方向変換モデル VTM（View Transition Model）を用いる．はじめに学習と認識でカメラを１台ずつ用いた認識実験により，VTM の有効性を確認する．次に学習と認識でそれぞれカメラ２台と１台用いて，学習での最適なカメラ配置を分析する．最後に，学習と認識でカメラを２台ずつ用いた認識実験を行い，その結果から学習と認識でカメラを１台ずつ用いた場合と比較して，カメラ２台ずつ用てそれぞれ最適な配置を用いた場合，認識率が 52.2%から 84.8%に向上することが確認できた．

1. はじめに

現在，指紋，掌の静脈などの生体情報を用いた個人認証装置が多く設置されている．米国政府は，2004 年 1 月 5 日より，「US-VISIT プログラム」と呼ばれる新たな出入国管理システムを導入した．このプログラムにおいては，米国への渡航者は，米国出入国時に指紋のスキャン並びに顔写真の撮影が必要になる[1]．一方，2007 年 11 月 20 日から日本への入国を申請する外国人も，入国審査の際に専用の機器を使って指紋及び顔写真を提供をすることが必要となった[2]．しかし，これらは個人認証のために，専用の機械に触れる必要があるという制約がある．それに対して，非接触生体情報の１つとして，歩容（歩き方）が挙げられる（図 1）．近年，人々の安全・防犯への意識の高まりとともに，商店やデパートに設けられている監視カメラを用いた人物の識別に関わる研究や応用事例が多く報告されている[3]．歩容は人物がカメラから離れても認証が可能であるという点が特徴であり，このような監視カメラを用いた部外者の進入検知や，不審人物の認識などへの応用も期待されている．

歩容認識の分野では，モデルベースの手法とアピアランスベースの手法の２つが提案されている．前者は，主に人物の関節，足，手などの体の動き特徴を抽出し，モデルを構築することで認識を行う．Yu らは周波数領域における特徴の KFD（Key Fourier Descriptors）を用いて人物を認識している[4]．Chai らは各フレーム画像から切り出した頭と胴体，足の３つの部分の動き特徴に加え，各フレーム中の人物の高さと幅の比率を歩容特徴として人物を認識している[5]．これらのモデルベースの手法は一般的にノイズに影響されやすいとされている．一方，後者のアピアランスベースの手法として，Han らは歩容を表す特徴として GEI（Gait Energy Image）を使用した人物認識手法を提案している[6]．

これらの手法では，学習と認識でカメラに対する人物の歩行の向きが等しいという仮定の下で実験が行われている．しかしながら，学習と認識での人物の歩行の向きの違いは，認識率低下の大きな要因となる．これに対して，槇原らは VTM（View Transition Model）を用て異なる向きの歩容を生成することによって，人物認識を行っている[7]．これまで，我々は学習に VTM を用い，認識に複数のカメラを用いた人物歩容認識手法を提案してきた[8]．この中で，複数のカメラを用いる場合には，カメラの配置が認識率に大きく影響するという結果が得られている．

そこで，本稿では，学習と認識のそれぞれのカメラ配置を適切に決定することが認識率の向上に繋がると考え，カメラ配置と認識率の関係を報告する．

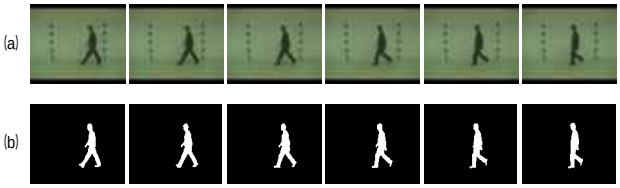
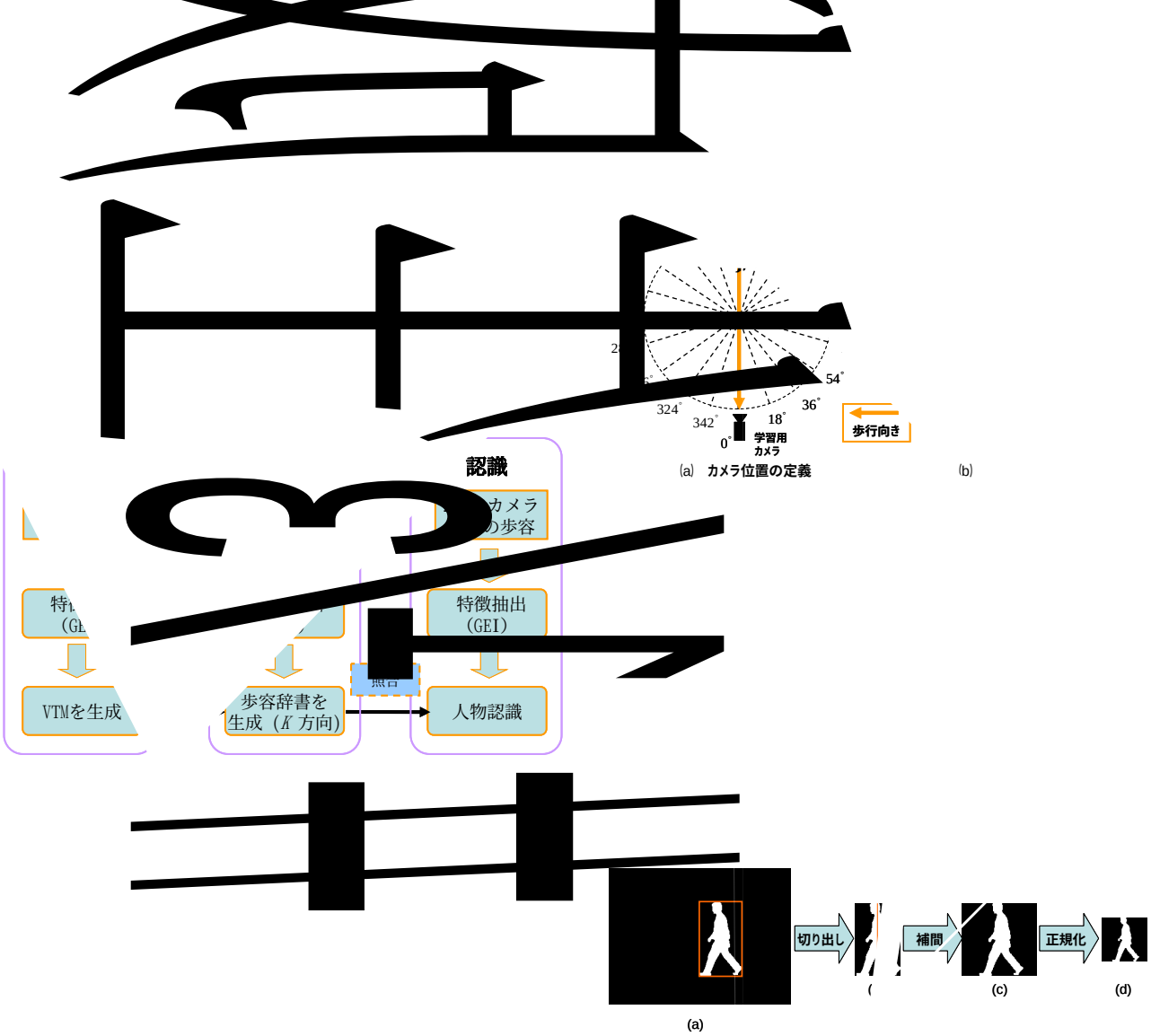
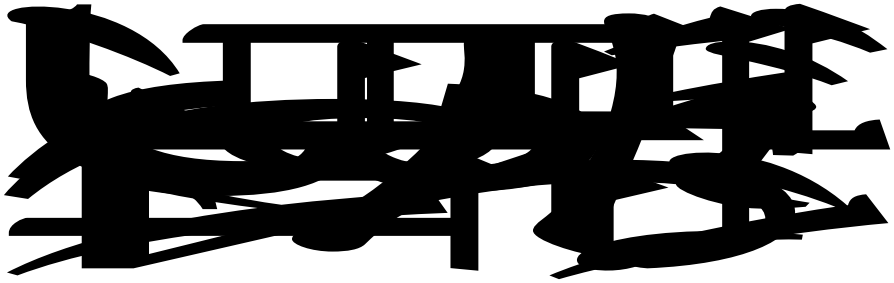


図 1．（a）実際のフレーム；（b）シルエット画像



$$N = \arg \qquad C(N)$$



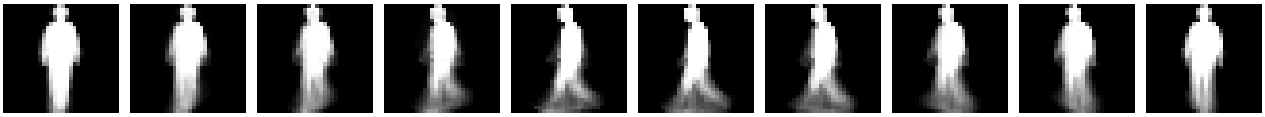


図 6． 様々なカメラ位置から得られる GEI の例

のシルエット画像中の位置(x, y)における画素値であり，N_{total}は総フレーム数である．以下では，歩行周期の下限$N_{\min}$，上限$N_{\max}$としてそれぞれ 15，30 を仮定する．

2.5 Gait Energy Image (GEI)

人物を認識する際の歩容特徴である GEI を以下の式で定義する．

$$G(x,y) = \frac{1}{N_{gait}} \sum_{n=1}^{N_{gait}} g_n(x,y) \qquad (4)$$

GEI は 1 周期分の正規化されたシルエット画像の平均画像である．様々なカメラ位置から得られる GEI を図 6 に示す．

2.6 View Transition Model (VTM)

VTM は対象の姿勢や方向変化に対応した画像生成を実現するためのモデルである．VTM を用いた歩行の向きの変換は，VTM 作成と画像生成の 2 段階に分けられる．はじめに，カメラ位置θ_iから得られる GEI から他のカメラ位置θ_jの GEI を生成する変換行列$\mathbf{T}_{\theta_i \to \theta_j}$を計算することで，VTM を作成する．次に変換行列を用いて入力された GEI から任意のカメラ位置の GEI を生成する．

- **VTM 作成**

M 人のK個のカメラ位置からの歩容を収集し，GEI をそれぞれ抽出する．それから，特異値分解によって以下の式を得る．

$$\begin{bmatrix} \mathbf{a}_{\theta_i}^1 & \cdots & \mathbf{a}_{\theta_i}^M \\ \mathbf{a}_{\theta_j}^1 & \cdots & \mathbf{a}_{\theta_j}^M \end{bmatrix} = \mathbf{USV}' = \begin{bmatrix} \mathbf{P}_{\theta_i} \\ \mathbf{P}_{\theta_j} \end{bmatrix} \begin{bmatrix} \mathbf{v}^1 & \cdots & \mathbf{v}^M \end{bmatrix} \qquad (5)$$

ここで，$\mathbf{a}_{\theta_i}^m$，$\mathbf{a}_{\theta_j}^m$は人物mについて位置θ_i，θ_jのカメラから得られた GEI から作られる画像ベクトルである．行列$\mathbf{V}'$の各列ベクトル$\mathbf{v}^m$は人物m特有の特徴を持っており，テクスチャ特性ベクトルと呼ばれている．式 (5) から，位置θ_iのカメラからの GEI のベクトルは$\mathbf{a}_{\theta_i}^m = \mathbf{P}_{\theta_i}\mathbf{v}^m$で表される．これに$\mathbf{a}_{\theta_j}^m = \mathbf{P}_{\theta_j}\mathbf{v}^m$を代入することによって，位置$\theta_i$から$\theta_j$への変換行列$\mathbf{T}_{\theta_i \to \theta_j}$は以下のように算出できる．

$$\mathbf{a}_{\theta_j}^m = \mathbf{P}_{\theta_j}(\mathbf{P}_{\theta_i}^T \mathbf{P}_{\theta_i})^{-1} \mathbf{P}_{\theta_i}^T \mathbf{a}_{\theta_i}^m \qquad (6)$$

$$\mathbf{T}_{\theta_i \to \theta_j} = \mathbf{P}_{\theta_j}(\mathbf{P}_{\theta_i}^T \mathbf{P}_{\theta_i})^{-1} \mathbf{P}_{\theta_i}^T \qquad (7)$$

- **画像生成**

$\mathbf{T}_{\theta_i \to \theta_j}$は人物に依存せず，カメラ位置（あるいは，カメラに対する歩行の向き）だけに依存するので，VTM 作成時と異なる人物nのθ_iの GEI からθ_jの GEI を生成することができる．VTM により生成される GEI ベクトル$\hat{\mathbf{a}}_{\theta_j}^n$は，次式によって得られる．

$$\hat{\mathbf{a}}_{\theta_j}^n = \mathbf{T}_{\theta_i \to \theta_j} \mathbf{a}_{\theta_i}^n \qquad (8)$$

同様に，2 つのカメラ位置θ_i，θ_jの GEI からθ_kの GEI を生成する場合，$\hat{\mathbf{a}}_{\theta_k}^n$は式 (9) のようになる．このように，生成元として複数の画像を用いることによって，変換精度の向上を図ることができる．

$$\hat{\mathbf{a}}_{\theta_k}^n = \mathbf{T}_{(\theta_i, \theta_j) \to \theta_k} \begin{bmatrix} \mathbf{a}_{\theta_i}^n \\ \mathbf{a}_{\theta_j}^n \end{bmatrix} \qquad (9)$$

- **変換結果**

VTM を用いて位置36°のカメラと$36^\circ, 126^\circ$の 2 台のカメラからの GEI からそれぞれ生成した様々なカメラ位置からの GEI と，実際の GEI を図 7 に示す．

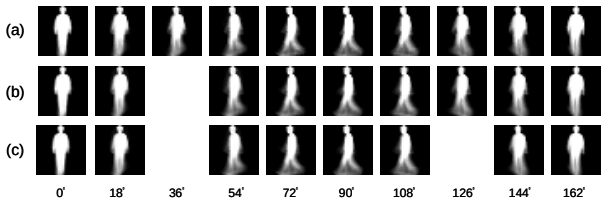


図 7． 変換結果．(a) 実際の GEI；(b) 36°からの変換結果；(c) $36^\circ, 126^\circ$からの変換結果

2.7 VTM の変換誤差

VTM の変換誤差を，実際の GEI と他の歩行の向きから生成された GEI との間のユークリッド距離で定義する．カメラを 2 台用いるときの変換誤差$e_{(\theta_i, \theta_j)}$は，以下の式により得られる．

$$e_{(\theta_i, \theta_j)} = \frac{1}{K-2} \frac{1}{N} \sum_{\substack{k=1 \\ k \neq i, j}}^{K} \sum_{n=1}^{N} \| \hat{\mathbf{a}}_{(\theta_i, \theta_j) \to \theta_k}^n - \mathbf{a}_{\theta_k}^n \| \qquad (10)$$

ここで，$\hat{\mathbf{a}}^{n}_{(\theta_i,\theta_j)\to\theta_k}$ は人物 n のカメラ位置 θ_i，θ_j から θ_k へ変換された GEI ベクトルであり，$\mathbf{a}^{n}_{\theta_k}$ は θ_k の実際の GEI ベクトルである．N と K はそれぞれ人数，歩行の向きの数を表す．

2.8　照合基準

認識段階で C 台のカメラを用いる場合を考える．各入力画像から抽出した GEI ベクトルを $\mathbf{x}_c$（$c=1,2,\cdots,C$）とし，人物 p のカメラ位置 θ の歩容辞書データを $\hat{\mathbf{a}}^{p}_{\theta}$ としたとき，次の式で認識結果 $\hat{p}$ を得る．

$$\hat{p}=\arg\min_{p}\left\{\min_{\theta}\min_{c}\left(\|\mathbf{x}_c-\hat{\mathbf{a}}^{p}_{\theta}\|\right)\right\} \qquad (10)$$

3.　実験

歩容データベース（CASIA）[9]中の水平方向に18°間隔で撮影された 85 人分×2 セットのシルエット画像列を実験で用いた．データベース中には，0°～180° の 11 通りのカメラ位置から撮影された歩容データが存在する．本実験では，180°～342° のカメラ位置からの歩容データとしてそれぞれ 0°～162° のカメラ位置から得られるシルエット画像の鏡像を用いた．はじめに，20 人分の歩容データから VTM を作成した．学習時には歩容辞書作成のために VTM 作成時とは異なる 65 人分の任意の 1 つ，あるいは 2 つのカメラ位置からの歩容データを用いた．VTM によりそれ以外のカメラ位置からの GEI を生成し，歩容辞書を作成した．本実験では，まず学習と認識でカメラを 1 台ずつ用いた実験によって，VTM の変換誤差と認識率の関係を調べた．次に学習と認識でカメラをそれぞれ 2 台，1 台用いて実験によって，学習におけるカメラの最適な配置を VTM の変換誤差を用いて分析した．最後に学習と認識でカメラを 2 台ずつ用いた場合，VTM の変換誤差を用いて得られたカメラ配置を学習で用い，認識時の最適なカメラ配置を調査した．

3.1　学習カメラ 1 台，認識カメラ 1 台

学習段階において歩容辞書を作成する際に，VTM の変換元としてカメラ位置 90° の GEI を用いた場合のカメラに対する歩行の向きと認識率の関係を図 8 に示す．学習と認識でカメラに対する歩行の向きが等しい場合，認識率が最も高くなった．また，カメラに対する歩行の向きが 90° の位置から離れるに従って認識率が低下することが分かった．一方，学習段階において様々なカメラ位置の GEI から歩容辞書を作成した際に，式（10）によって計算した変換誤差と，認識段階における平均認識率との関係を図 9 に示す．その結果から，変換誤差が最も小さい 126° のときに，認識率が 52.2%最も高くなることが分かった．このことから，VTM の変換誤差から学習時のカメラ位置を決定することは認識率の向上に有効であることを確認した．

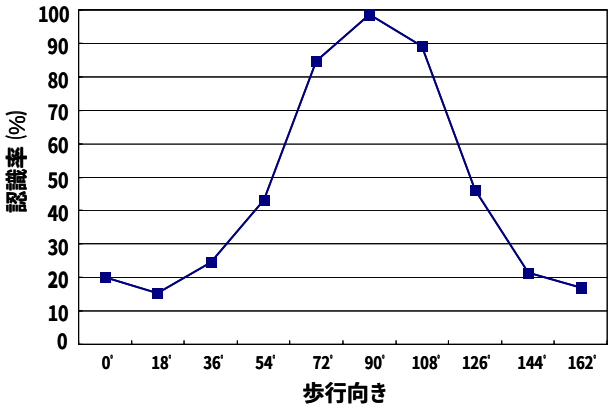


図 8．　認識結果（学習段階でカメラ位置：90°）

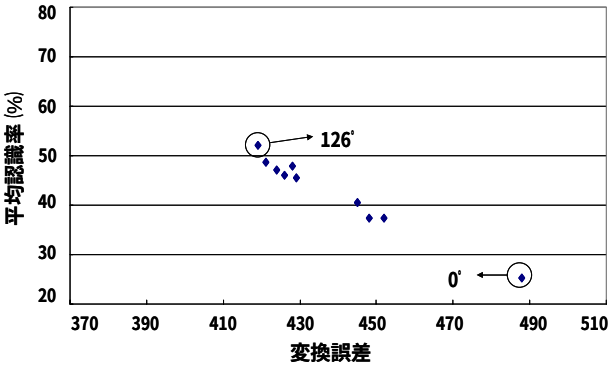


図 9．　変換誤差と認識率の関係（学習カメラ 1 台）

3.2　学習カメラ 2 台，認識カメラ 1 台

学習段階では，鏡像の冗長性を除いた 2 台のカメラの配置の組み合わせ（${}_{10}C_2$=45 通り）を考える必要がある．学習時の 2 台のカメラの位置とその間の角度差に対する変換誤差と平均認識率の関係を図 10 に示す．図の左上方にある認識率の高い点は学習時の 2 台のカメラ間の角度差が 90° の場合が多かった．このことから，学習段階でのカメラの間の適切な角度差が 90° であることが分かった（図 11）．

3.3　学習カメラ 2 台，認識カメラ 2 台

認識に 2 台のカメラ $c1$，$c2$ を用いる場合，1 台の認識カメラの配置方法が 20 通り（18° 間隔で全円 360°）あるので，2 台のカメラ配置としては 190 通り（${}_{20}C_2$）ある．しかし，認識の際にカメラに対する人物の歩行の向きが等確率で現れると仮定すると，

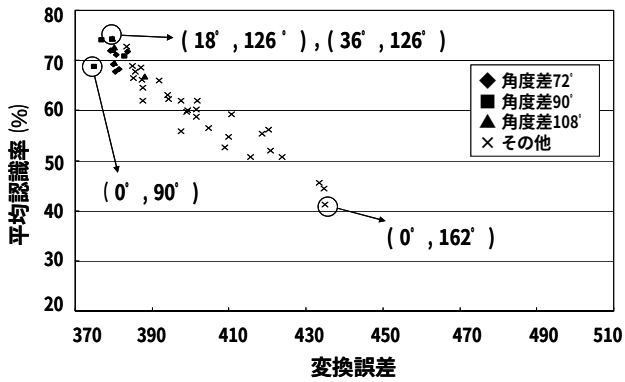


図 10.　変換誤差と認識率の関係（学習カメラ 2 台）

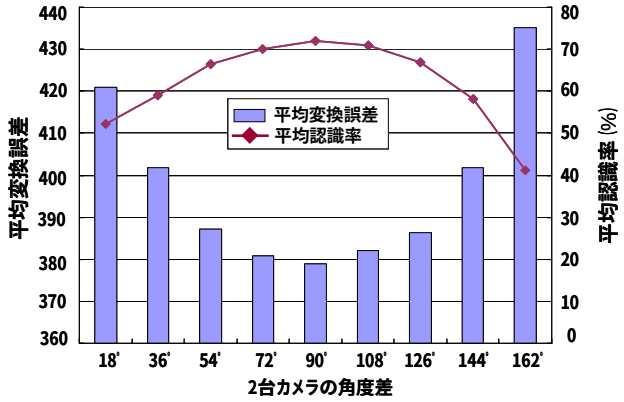


図 11.　角度差，変換誤差と認識率の関係

この 190 通りの中には，回転の冗長性があり，それに加えて本実験では歩容データの鏡像を用いていることから，実際には 2 台のカメラの角度差が 18°, 36°, 54°, 72°, 90° の場合の 5 通りとなる．

学習と認識でカメラを 2 台ずつ用いた実験について述べる．3.2 の実験結果から，学習段階では 2 台のカメラ間の適切な角度差は 90° であることが分かった．また，3.1 の学習と認識で 1 台ずつ用いた実験から，認識段階では学習で使用したカメラ位置に対する歩行の向きに近い方向の歩容データが入力された際に認識率が高いことが分かっている．学習段階で 2 台のカメラの位置を 36°, 126° とした場合，図 12（a）中に破線円で示すように認識時のカメラに対する歩行の向きが 36°，126°，及び GEI の性質からそれらの方向と類似した特徴が現れる 216°，306°，ならびにそれらの周囲の角度のとき，高い認識率が得られる．また，図 12（b）のように 2 台のカメラ間の角度差を 54° にすれば，図 12（c）のように人物のどの向きの歩行に対してもいずれかのカメラによって高い認識率が得られる．太い円は両方のカメラにとって高い認識率を得られる方向である．一方，もし 2 台のカメラ間の角度差を図 13（a），（b）のように 18° にすれば，図 13（c）のようにどちらのカメラに対しても高い認識率が得られない方向が存在する．学習段階でのカメラ間の角度を 90° にしたとき，認識段階での 2 台のカメラ間の角度差と認識率の関係を表 1 に示す．

この結果から，学習段階で 2 台のカメラ間の角度差を 90° としたとき，認識段階での 2 台のカメラ間の適切な角度差は 36°（または 54°）であることが分かった．

表 1.　平均認識率

		2 台の学習カメラの位置				
		0° 90°	18° 108°	36° 126°	54° 144°	72° 162°
2 台の認識カメラの角度差	18°	72.9 %	80 %	80.9 %	77.8 %	78.9 %
	36°	77.4 %	82 %	84.8 %	82.6 %	78.3 %
	54°	77.5 %	82.5 %	83.8 %	81.8 %	79.7 %
	72°	72.6 %	78.5 %	78.6 %	78.3 %	78.5 %
	90°	68.6 %	74.5 %	72.9 %	72.3 %	74.8 %

表 2.　認識率の比較

	学習カメラ 1 台 認識カメラ 1 台	学習カメラ 2 台 認識カメラ 2 台
学習カメラ位置	126°	36°，126°
認識カメラの角度差	---	36°
平均認識率	52.2 %	84.8 %

4.　おわりに

本稿では，歩行の向きに依存しない人物歩容認識手法において，カメラを複数使う場合の最適なカメラ配置を検討した．VTM の変換誤差と認識率の関係を調べることによって，学習時には 2 台のカメラ間の適切な角度差が 90° であることが分かり，認識段階で 5 通りのカメラ配置だけを考えれば良いことが分かった．この結果を用いて，学習段階と認識段階でカメラを 2 台用いた実験を行い，学習と認識でカメラを 1 台ずつ用いた場合と比較して，カメラを 2 台ずつ用い，それぞれ最適な配置を用いた場合，平均認識率が 52.2%から 84.8%に向上することが確認できた（表 2）．

認識にカメラを 2 台用いる場合，歩行の向きによっては，カメラ 1 台の場合より認識精度が低下する場合があるので，2 台のカメラから得られた認識結果の適切な統合方法を検討する必要があると考える．