

歩容・顔・身長によるマルチモーダル個人認証のための 時空間解像度に適応的なスコア統合

木村 卓弘^{1,a)} 榎原 靖^{1,b)} 村松 大吾^{1,c)} 八木 康史^{1,d)}

概要：一つのカメラ映像だけで実現可能な歩容，顔，身長によるマルチモーダル認証は精度とコストの両方の点で有効な方法である．その一方で，その精度は空間解像度や時間解像度といった撮影環境による影響を受け，また，歩容，顔，身長それぞれの特徴によってその影響は異なる．そのため，より高い精度を実現するためには，空間解像度と時間解像度に応じてそれぞれの特徴から得られるスコアに適切な重み付けをすることが重要である．本研究では，まず 1,935 人の公開歩行映像データベースを用いて，それぞれの特徴について様々な解像度のスコアデータベースを作成した．そしてそのデータベースを用いて性能評価を行い，歩容，顔，身長それぞれについて時空間解像度に応じた性能の変化を解析した．さらに，構築したスコアデータベースに基づいて，時空間解像度に応じた最適な重み付けの値を線形ロジスティック回帰によって計算した．また，学習データに含まれない時空間解像度の組み合わせに対するテストデータに対しては，ガウス過程回帰による重みの推定を行い，精度評価を行った．精度評価の結果，学習データと同じ時空間解像度の組を用いたテストデータの精度，つまりその時空間解像度における性能の上限とほぼ同等の結果が得られた．

1. はじめに

個人認証の方法としては IC カードやパスワードなどの方法が存在するが，このような方法の場合では紛失，忘却，盗難のリスクがある．そこで注目されている方法として個人の生体情報を用いる生体認証 (バイオメトリクス) [1] がある．生体認証には DNA，指紋，静脈，虹彩，顔，署名，キーストローク [2–8] など多くの方法が存在し，個人の身体情報を用いるため，これらの方法では紛失・忘却・盗難といったリスクを考える必要がなくなる．

また，認証精度を向上させる方法として，マルチモーダル個人認証 [9] が注目を集めている．具体的には，顔と指紋 [10] や顔と虹彩 [11, 12] などを組み合わせて行う個人認証手法が提案されている．しかし，マルチモーダル個人認証には各モダリティに応じた複数のセンサーが必要であるためコストがかかる点や，使用者に複数回の試行を要求することによる負担の増加といった問題がある．

そこで，一つのカメラ映像で実現可能な歩容と顔によるマルチモーダル個人認証 [13–17] が有効な手法として提案

されている．更に，事前にカメラキャリブレーション [18] をしておくことで，身長の情報も同じように一つのカメラ映像だけで得ることができる．つまりこの方法では一つのカメラだけでマルチモーダル個人認証を行うことができる点が大きな利点である．また，村松ら [19] は，歩容・顔・身長によるマルチモーダル個人認証の手法に加え，被験者の様々な撮影角度の映像の統合も行っている．

しかし，歩容・顔・身長によるマルチモーダル個人認証の精度は，空間解像度 (画像サイズ) と時間解像度 (フレームレート) という，二つの重要な解像度によって大きく影響される．人物に対する空間解像度は，画像全体の解像度やカメラからの人物までの距離に依存する．カメラから離れた歩行者の場合，顔が鮮明には映っていないかったり，身長を精度良く計算できなくなるため，顔や身長による認証を行う場合に空間解像度が大きく影響する．

一方，時間解像度は，防犯カメラの通信帯域や記憶装置の保存容量に依存する (これらは空間解像度にも影響を及ぼすが，現在の犯罪捜査では顔や服装を撮影することが重要であるため，時間解像度を犠牲にしても高空間解像度であることが重要とされている)．顔や身長は静的な特徴であるため時間解像度にはあまり影響しないと思われるが，歩容は 1 周期を特徴として用いるため大きく影響する．

そこで，本論文では，歩容・顔・身長によるマルチモー

¹ 大阪大学
Osaka University

a) kimura@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



(a) 原画像 (b) シルエット画像

図 1 歩行状況の原画像 (a) と、レンズの歪みを取り除いた後、壁に正対するように変換を行ったシルエット画像 (b)。シルエットのサイズは被験者によるが、90×180 画素程度である。

ダル個人認証を、空間解像度や時間解像度といった撮影条件を考慮したスコアレベル統合の枠組みで実現する。より具体的には、入力映像の時空間解像度に応じて、歩容・顔・身長によるスコアへの重み付けを適切に変化させる方法を提案する。代表的な時空間解像度を含む学習データを用いて最適な重みを算出しておき、それ以外の時空間解像度のテストデータに対しては、学習データから算出した重みを内挿することによって重みを推定する。

このような目的の下、本論文の貢献は、次の三つの点にまとめられる。

1. 様々な時空間解像度におけるマルチモーダルスコアデータセットの構築

大規模歩容データベース The OU-ISIR Gait Database, the large population data set [20] の 1,935 人のデータを用いて歩容・顔・身長スコアデータセットを構築する。ここでは様々な時空間解像度の組み合わせに対して、スコアを計算する。

2. 歩容・顔・身長の認証精度の時空間解像度に対する感度評価

構築したデータセットを用いて歩容・顔・身長のそれぞれの認証精度の評価を行う。具体的には、1 対 1 認証問題における等価誤り率 (Equal Error Rate, EER) が時空間解像度に応じてどのように変化するかを解析する。

3. 時空間解像度に適応的なスコア統合

まず、時空間解像度に適応的な重み付けをするにあたって、学習データのスコアデータセットに対して代表的なスコアレベル統合手法である線形ロジスティック回帰 (Linear Logistic Regression, LLR) を用いて最適な重み付けを計算する。更に、学習データ以外の時空間解像度に対しては、学習データに対する重みを内挿により、重みを推定する手法を定式化する。また、その定式化に基づいて、様々な時空間解像度に対するテストデータで精度評価を行う。

2. 様々な時空間解像度におけるマルチモーダルスコアデータセット

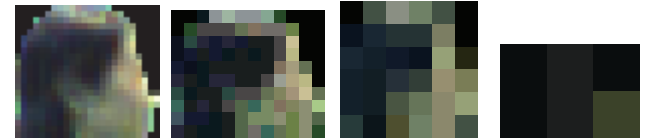
2.1 歩行映像データベース

様々な時空間解像度におけるマルチモーダルデータセットを作成するため、The OU-ISIR Gait Database, the large population data set を用いた。このデータベースでは 1 歳



(a) 30, 640×480 (b) 30, 80×60 (c) 3, 640×480 (d) 3, 80×60

図 2 GEI の例 (時間解像度 [fps] と空間解像度 [画素] の組を下に示す)。



原画像:640×480 原画像:320×240 原画像:160×120 原画像:80×60

顔画像:24×22 顔画像:12×10 顔画像:6×6 顔画像:3×2

図 3 様々な空間解像度の原画像から作成した顔の画像。顔画像のサイズは被験者に依存するが、ここでの画像のサイズ [画素] を示している。

から 94 歳までの幅広い年代層の男女合計 4,016 人の歩行映像からなる。映像の空間解像度は 640×480 画素で、時間解像度は 30fps である。顔の画像は図 1(a) の画像列から作成する。シルエット画像は、レンズの歪みを取り除いた後、壁に正対するように変換を行い、背景差分に基づくグラフカットセグメンテーションによって作成した。このシルエットは歩容の特徴抽出や身長計算に用いる。

2.2 スコアの算出

2.2.1 歩容

本研究では、歩容特徴として、歩容認証において幅広く利用されている GEI [21] を用いる。まず、図 1 のシルエット画像から 88×128 画素に正規化した歩容シルエットボリューム (Gait Silhouette Volume, GSV) を作成し、GSV を 1 周期で平均化することにより図 2 のような GEI を作成する。プローブ (入力データ) とギャラリー (登録データ) の GEI をそれぞれ G_p 、 G_g とすると、プローブ、ギャラリーの相違スコアはユークリッド距離として次のように計算される。

$$S_{\text{gait}} = \|G_p - G_g\|_2, \quad (1)$$

ここで、 $\|\cdot\|_2$ は L2 ノルムを表す。

2.2.2 顔

顔認証を行う際、髪や輪郭を取り除いた顔の部分を用いて行うことが一般的であるが、空間解像度が極端に低い場合 (25×25 画素より小さい場合) には高い認証精度を保つことができないため、髪や輪郭も含めた全体的な特徴を利用した方が良く考えられる。また、図 3 から分かるように、頭の領域のカラー画像はシルエットのマスクを加えたものとなっている。この頭の領域の画像のサイズは被験者によって異なるが、空間解像度を下げる前の 640×480 画素の画像では 18×20 画素から 31×25 画素の範囲である。

本論文ではこの頭の領域を顔と呼ぶものとする．二つの顔画像の相違度はテンプレートマッチングを用いて正規化相互相関で次のように計算を行う． F_{p_i} をプローブの i 番目のフレームの特徴とし， $F_{g_{j,k}}$ をギャラリーの j 番目のフレームの，テンプレートマッチングの探索範囲での k 番目の特徴とする．そのとき，プローブとギャラリーの相違度スコアは次のように計算される．

$$S_{face} = \min_{i,j,k} [1 - NCC(F_{p_i}, F_{g_{j,k}})], \quad (2)$$

ここで， $NCC(F_{p_i}, F_{g_{j,k}})$ は F_{p_i} と $F_{g_{j,k}}$ の間の正規化相互相関である．

2.2.3 身長

身長は，シルエットの頭頂点と足下点に基づいて計算する．2.1 節で述べたように，シルエット画像は壁面に正対するように変換されており，また，歩行者が壁面と平行な一定のコースを歩いていることから，壁面と正対するカメラから見た奥行きは，一定であると考えられる．よって，シルエット領域の外接矩形の高さ [画素] から，カメラ校正によって得られるカメラパラメータと奥行き情報に基づいて，世界座標における身長 [m] に変換することが可能である．また，身長は静的な特徴であるためフレーム毎に計算しているが，歩行動作による上下動により多少の変化が存在する．そのため，身長 h は以下のように画像列に対して平均を取ることで計算している．

$$h = \frac{1}{N_f} \sum_{i=1}^{N_f} Z_i, \quad (3)$$

ここで， Z_i は i 番目のフレームの身長， N_f は画像列のフレーム数である．

h_p , h_g をそれぞれプローブ，ギャラリーの身長とすると，相違度スコア S_{height} は絶対値を用いて以下のように計算する．

$$S_{height} = |h_p - h_g|, \quad (4)$$

2.2.4 時空間解像度のダウンサンプリング

原画像とシルエット画像を時空間解像度に関してダウンサンプルした上で，歩容・顔・身長それぞれについてスコアデータセットを作成する．時空間解像度のバリエーションは，表 1 に示す通りである．また，本論文では，表中などで，空間解像度を SR，時間解像度を TR と略記する．

まず，空間解像度について説明する．本論文では原画像の 680×480 画素を最大とした合計 13 通りを用いた．これにより，様々な人物サイズの歩容・顔・身長特徴が得られるため，カメラ固有の画像サイズの違いに加えて，カメラから人物までの距離に応じた人物サイズの変化についても考慮に入れた解析を考えることができる．空間解像度をダウンサンプルした GEI と顔の画像の例を図 2 と図 3 に示す．

表 1 空間解像度 [画素] と時間解像度 [fps] の組．

SR	TR
640 × 480, 480 × 360, 320 × 240, 213 × 160,	30, 15, 10,
160 × 120, 128 × 96, 106 × 80, 91 × 68,	7.5, 6, 5,
80 × 60, 64 × 48, 53 × 40, 40 × 30, 20 × 15,	3.75, 3, 2, 1

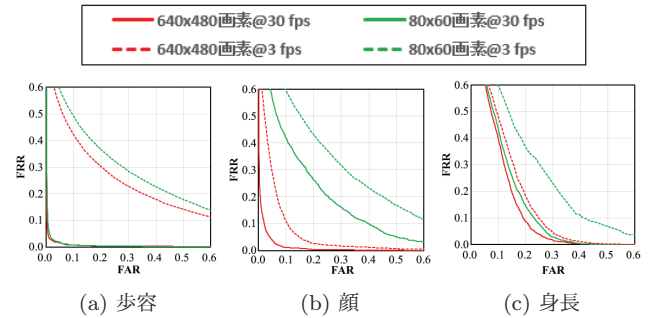


図 4 各モダリティの ROC 曲線

次に，時間解像度について説明する．本論文は 30fps を最大とした 10 通りを用いた．30fps の画像列から時間解像度に応じて一定の間隔で画像を間引くことにより，低時間解像度の画像列を作成する．

また，時間解像度をダウンサンプルした場合，開始フレームによって画像列が変化する．例えば 15fps の場合を考えると，奇数番目の画像を使用するか，偶数番目の画像を使用するかで結果が異なる．そのため，15fps, 10fps, 7.5fps, 6fps, 5fps, 3.75fps, 3fps, 2fps, 1fps に対して，それぞれ 2, 3, 4, 5, 6, 8, 10, 15, 30 種類の開始フレームに対応する個別の画像列を作成して使用する．以下ではこの開始フレームの種類を N_{TR} として表す．

2.2.5 スコアデータセット

本研究で構築するスコアデータセットには，大規模歩容データベースから 1,935 人のサブセットを選択し，それぞれの被験者について 85 度の観測方向（ほぼ側面方向）から撮影された，プローブとギャラリーの組み合わせを使用した．ギャラリーに関しては 1,935 通り，プローブに関しては 1,935 N_{TR} 通りのデータについて，空間解像度が 13 通り，時間解像度が 10 通り，つまり合わせて 130 通りのスコアを計算する．このスコアデータは，ギャラリー，プローブの被験者 ID リストのすべての組み合わせに関する相違度スコア行列の形で表される．つまり，本人同士のスコアは 1,935 N_{TR} 通り，他人とのスコアは 1,935 N_{TR} × 1,934 = 3,742,290 N_{TR} 通り存在する．この行列は歩容・顔・身長それぞれについて，時空間解像度のすべての組み合わせについて計算する．

3. 歩容・顔・身長の認証性能評価

3.1 1 対 1 認証に対する結果

ここでは 1 対 1 認証における歩容・顔・身長の個別の認証性能の結果を示す．

評価指標として，他人受入誤り率 (False Acceptance Rate, FAR) と本人拒否誤り率 (False Rejection Rate, FRR) のト

表 2 すべての空間解像度 [画素] と時間解像度 [fps] の組み合わせにおける EER [%] (歩容).

SR\TR	30	15	10	7.5	6	5	3.75	3	2	1
640×480	2.4	2.5	2.6	3.7	7.4	9.6	22.0	23.9	39.4	39.5
480×360	2.4	2.5	2.6	3.7	7.5	9.8	22.1	24.1	39.6	39.4
320×240	2.4	2.4	2.7	3.9	7.5	9.9	22.1	24.3	39.7	39.4
213×160	2.3	2.5	2.8	4.2	8.0	10.4	22.2	25.0	39.8	39.8
160×120	2.5	2.6	3.1	4.4	8.3	10.7	22.6	25.9	40.0	39.6
128×96	2.9	3.2	3.7	4.9	8.9	11.4	23.2	26.7	40.4	40.1
106×80	2.6	2.9	3.9	5.9	9.7	12.0	23.4	27.4	39.8	39.6
91×68	2.2	2.6	5.1	6.5	9.9	12.9	23.7	28.2	39.9	39.9
80×60	2.6	4.2	6.9	6.3	11.3	14.1	23.3	28.8	40.1	39.4
64×48	2.9	11.2	5.0	12.5	12.0	16.4	24.3	29.9	39.8	39.0
53×40	3.8	8.0	9.0	10.9	19.4	17.8	26.0	31.8	40.8	40.1
40×30	5.7	7.7	23.7	17.3	18.5	23.9	28.9	31.3	42.0	38.9
20×15	18.1	20.2	25.3	25.8	28.0	24.8	30.4	32.7	36.3	32.0

レードオフを示す受信者操作特性 (Receiver Operatorating Characteristic, ROC) 曲線を用いる。図 4 に (1) 高時空間解像度, (2) 高空間解像度と低時空間解像度, (3) 低空間解像度と高時間解像度, (4) 低時空間解像度の 4 通りに対する ROC 曲線を示す。さらに, 時空間解像度のすべての組み合わせにおける FAR と FRR の等価誤り率 EER を表 2-4 に示す。空間解像度が極端に低い場合, 顔画像が 1×1 画素になってしまう場合が存在する。そのような場合には, 正規化相互相関でスコアを計算できなくなってしまうため, チャンスレベルとして扱い, EER を 50%としている。

この結果から分かるように, 顔認証の性能は空間解像度が下がったときに大きく低下していることが分かる。その一方で歩容認証の性能は空間解像度が下がってもあまり落ちず, 時間解像度が下がったときに大きく低下していることが分かる。また, 身長による認証は, 同じ身長の人が多く存在することから歩容や顔と比較すると全体的に性能は低い。そして, 時空間解像度のどちらかが低くなってもあまり性能が落ちず, 両方が低くなったときに性能が低下していることが分かる。

3.2 個々のモダリティに対する考察

それぞれの特徴の傾向を分析するために時間解像度, 空間解像度をそれぞれ固定した場合の EER の変化を図 5 と図 6 に示す。

高時間解像度 (30fps, 図 5) における空間解像度に対する EER の変化を見ると, 160×120 画素より空間解像度が下がってくると, 顔認証の性能が大きく悪化していることが分かる。図 3 (c) を見ても分かるように, 顔の画像は 160×120 画素からサイズがかなり小さくなっている。その一方で, 歩容と身長は, 空間解像度の中程度までの低下に対してあまり大きくは性能が悪化していないことが分かる。53×40 画素を下回るような解像度が極めて低い場合には, 人物サイズが約 8×15 画素ほどになってしまうため,

表 3 すべての空間解像度 [画素] と時間解像度 [fps] の組み合わせにおける EER [%] (顔). “-”はチャンスレベルを表し, 50%であることを表す。

SR\TR	30	15	10	7.5	6	5	3.75	3	2	1
640×480	4.5	5.8	7.2	8.0	7.7	10.0	10.1	10.3	10.2	10.5
480×360	3.1	3.8	5.2	5.9	6.2	8.0	8.0	8.1	8.1	8.3
320×240	4.6	5.8	7.7	8.7	8.7	10.7	10.8	10.8	10.8	11.1
213×160	3.2	3.9	4.8	6.4	7.4	9.7	9.8	9.9	9.7	10.5
160×120	4.6	5.4	6.6	8.4	9.8	12.2	12.6	12.8	12.5	13.7
128×96	7.2	8.1	9.5	11.7	12.9	15.3	15.8	16.2	16.2	17.6
106×80	10.0	11.1	13.1	16.0	17.8	20.3	20.4	21.1	21.0	23.0
91×68	13.6	15.0	16.9	20.3	21.8	25.1	25.1	25.6	25.3	27.6
80×60	22.7	23.6	25.7	27.3	29.1	30.2	31.0	30.8	31.6	32.9
64×48	23.3	24.7	27.1	28.4	32.3	32.8	33.3	34.2	33.9	34.4
53×40	38.3	36.3	36.8	37.0	38.2	39.5	39.8	40.0	39.7	40.1
40×30	-	-	-	-	-	-	-	-	-	-
20×15	-	-	-	-	-	-	-	-	-	-

表 4 すべての空間解像度 [画素] と時間解像度 [fps] の組み合わせにおける EER [%] (身長).

SR\TR	30	15	10	7.5	6	5	3.75	3	2	1
640×480	16.2	16.5	17.0	17.8	17.6	19.4	19.4	19.4	19.4	19.4
480×360	16.3	16.6	16.9	18.0	17.8	19.6	19.6	19.6	19.6	19.6
320×240	16.5	16.9	17.0	18.0	17.8	19.9	20.0	20.0	19.8	20.0
213×160	16.3	16.8	17.4	18.3	18.4	20.0	20.1	20.1	20.1	20.1
160×120	16.5	17.2	17.6	18.6	18.2	20.9	21.0	21.0	21.0	21.0
128×96	17.0	17.7	18.1	19.3	19.1	21.7	21.7	21.7	21.7	21.7
106×80	17.3	18.2	18.8	20.1	20.3	23.1	23.2	23.2	23.3	23.2
91×68	15.8	16.8	17.6	19.8	19.9	22.9	23.0	23.0	23.0	23.0
80×60	18.2	19.4	21.7	23.2	23.3	27.3	27.3	27.3	27.1	27.3
64×48	15.7	18.3	21.4	24.6	24.4	29.7	29.7	29.4	29.7	29.7
53×40	18.0	21.8	24.9	28.2	28.9	33.2	33.2	33.2	33.1	33.2
40×30	19.6	24.8	30.0	33.6	33.4	38.7	38.9	38.8	38.8	38.8
20×15	31.6	37.2	41.0	42.2	41.4	43.1	43.2	43.0	43.0	43.0

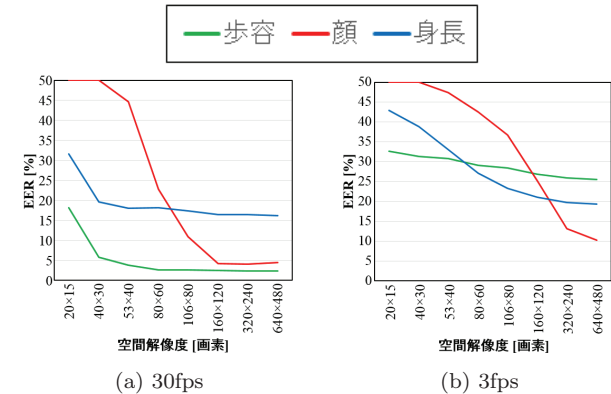


図 5 時間解像度を固定したときの空間解像度による EER の変化

性能が悪化している。

低時間解像度 (3fps) における空間解像度に対する EER の変化においても, 高時間解像度のときと同じような傾向が見られる。空間解像度が下がるにつれて顔認証の性能が大きく悪化しているが, 高時間解像度 (30fps) の時よりも高い空間解像度から性能悪化が見られる。これは高時間解

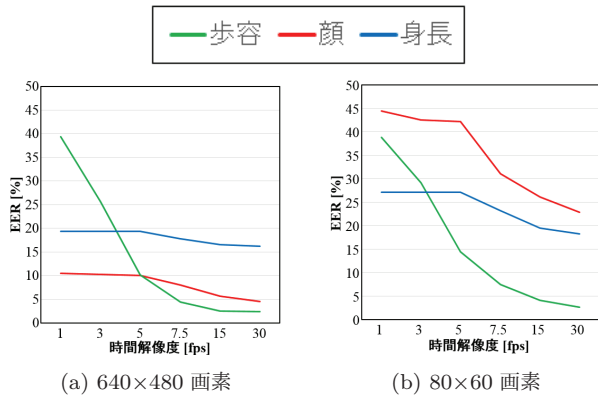


図 6 空間解像度を固定したときの時間解像度による EER の変化

像度の場合、画像のサンプリング枚数が多いことから空間解像度の低下に対して多少の性能悪化の抑制につながるものの、3fps のような低時間解像度のときには、サンプリング枚数の少なさから、空間解像度の低下がより直接的に性能悪化につながるためと考えられる。

また、身長についての結果を見てみると、3fps のときは空間解像度による影響を大きく受けているが、30fps のときは空間解像度が下がってもあまり性能が悪化していないことが分かる。このことは、身長が空間解像度によって大きな影響を受けるという事前の予想に反しているが、身長は平均をとって計算を行っているため、歩行の上下動に伴う高さの変動によりある種の超解像のような効果が得られ、結果としてサブピクセルオーダーの身長が得られていることが原因と思われる。これにより、高時間解像度の場合は空間解像度が低くても性能を高く保つことができていると思われる。

次に、高空間解像度 (640×480 画素) における時間解像度による EER の変化 (図 6) を見ると、顔と身長は時間解像度の低下に応じて緩やかに性能が悪化しているのに対して、歩容は 7.5fps を下回ると大きく性能が悪化している。

低空間解像度 (80×60 画素) における時間解像度に対する EER の変化を見ても、歩容と身長については同じような傾向が見られる。しかし、顔は高空間解像度のときと比べて、時間解像度による影響が大きいことが分かる。

4. 時空間解像度に適応的なスコア統合

4.1 スコア統合の概要

本節では、構築したスコアデータベースや各モダリティに対する性能解析の結果に基づいて、適応的なスコア統合手法を導入する。まず、各時空間解像度に対する最適な重みを、LLR によって学習する。詳細は 4.2, 4.3 節に示すが、LLR は学習データを必要とすることから、1,935 人の被験者をランダムに学習セットとテストセットに分割する、2 分割交差検定を用いて評価を行う。ランダムな分割法によって精度が異なるため、この評価を 50 回繰り返し行い、その平均によって評価を行う。

表 5 学習データとテストデータの空間解像度 [画素] と時間解像度 [fps] の組.

データセット	SR	TR
Training	640 × 480, 320 × 240, 160 × 120, 106 × 80, 80 × 60, 53 × 40	30, 15, 7.5, 5, 3, 1
Test	480 × 360, 213 × 160, 128 × 96, 91 × 68, 64 × 48	10, 6, 3.75, 2

また、このような重みはいくつかの代表的な時空間解像度の組 (学習データ) に対して求めておき、それ以外の時空間解像度の組 (テストデータ) に対しては内挿による推定を行う。そのような学習データとテストデータの時空間解像度の組は表 5 に示した通りである。ここで、3.1 節で述べたように、顔認証において空間解像度が 40×30 画素以下になると、チャンスレベルとなることから、スコア統合の実験においては、空間解像度は 40×30 画素、20×15 画素を除いた 11 通りとなっている点に注意されたい。

4.2 スコアレベル統合の枠組み

まず、スコアレベル統合を行う前にそれぞれのモダリティのスコアの正規化を行う。 $s_m(i, j)$ を i 番目のギャラリー、 j 番目のプローブにおける、モダリティ $m \in \{face, gait, height\}$ のスコアとする。そのとき正規化スコア $\bar{s}_m(i, j)$ は次のように計算する。

$$\bar{s}_m(i, j) = \frac{s_m(i, j) - \mu_m(i)}{\sigma_m(i)}, \quad (5)$$

ここで、 $\mu_m(i)$ と $\sigma_m(i)$ は平均と標準偏差である。

そして、二つの画像列が与えられたときにそれが本人同士である事後確率を、それぞれの正規化スコア $\bar{s} = [\bar{s}_{face}, \bar{s}_{gait}, \bar{s}_{height}]^T$ を用いて計算する (本人同士の事象を $X = 1$ とする)。さらに、SR q_S と TR q_T はスコアに影響を与えるため、 $\mathbf{q} = [q_S, q_T]^T$ を考慮に入れる必要がある。そこで、時空間解像度に適応的な事後確率 $P(X = 1 | \bar{s}; \mathbf{q})$ を LLR [22] を用いて表現し、ロジット関数を次のようなスコアの重み付け和で表す。

$$\log \left(\frac{P(X = 1 | \bar{s}; \mathbf{q})}{1 - P(X = 1 | \bar{s}; \mathbf{q})} \right) = \sum_{m \in \{face, gait, height\}} \alpha_m(\mathbf{q}) \bar{s}_m + \alpha_c(\mathbf{q}), \quad (6)$$

ここで $\alpha_m(\mathbf{q})$ はモダリティ m の重み、 $\alpha_c(\mathbf{q})$ は定数項である。重みの値である α_{face} , α_{gait} , α_{height} , α_c は学習データを用いて、時空間解像度の値に応じて計算を行う。

4.3 重みの推定

1 節で述べたように、すべての時空間解像度の組についてあらかじめ重みを計算しておくことは現実的な方法ではない。そのため、代表的な時空間解像度に対する重みを計算しておき、テストデータの時空間解像度 $\mathbf{q}_*$ に対する重み α_* を推定することを考える。有限の学習データ $D = [Q, \alpha]^T$ を用い、 N 個の時空間解像度の組 $Q = \{\mathbf{q}_i\} (i = 1, \dots, N)$

と学習データを用いて計算した重み $\alpha = [\alpha_1, \dots, \alpha_N]^T$ が事前に与えられるものとして、各モダリティ独立について重みの推定を行う。

図 5, 6 に示したように、それぞれのモダリティの精度は非線形に変化するので、重みの値も非線形な枠組みにて計算する。具体的には、非線形のカーネル関数によるガウス過程回帰 (GPR) を用い、学習データ D と時空間解像度 q_* から重みを推定する。

まず、二つの時空間解像度 q_i と q_j の高次元特徴空間での内積を表す、ラジアル基底関数 (radial basis function, RBF) k を以下のように定義する。

$$k(q_i, q_j; \theta) = v \exp \left(-\frac{\|q_i - q_j\|^2}{2r^2} \right), \quad (7)$$

ここで $\theta = [v, r]^T$ はカーネル関数のパラメータである。そのとき、事後確率分布 $P(\alpha_* | \bar{q}_*, D)$ は以下のような平均 μ_* 、標準偏差 σ_*^2 のガウス分布となる [23]。

$$\mu_* = k_*^T (K + \Sigma)^{-1} \alpha \quad (8)$$

$$\sigma_*^2 = k(q_*, q_*; \theta) - k_*^T (K + \Sigma)^{-1} k_* + \sigma_{o,*}^2, \quad (9)$$

ここで K は (i, j) の要素を $k(q_i, q_j; \theta)$ とする $N \times N$ の正行列、 k_* は第 i 行を $k(q_i, q_*; \theta)$ 列ベクトル、 Σ は (i, i) の要素を σ_i^2 とする $N \times N$ の対角行列である。 $\sigma_{o,*}^2$ は観測ノイズである。

これより、ある時空間解像度 q_* に対し、平均 μ_* を重み α_* として用いる。

4.4 学習データの評価

テストデータ評価に先立って、学習データに対する EER を表 6, ROC 曲線を図 10 に示す。ここでは、単純なスコア統合手法である Sum, 時空間解像度に関係なく固定した重み付けを用いる LLR (Fixed) と、時空間解像度の組に対して個別に学習した LLR の結果を比較する。ここでは、重みの学習と性能評価に用いる時空間解像度が一致していることから、LLR は、次節のテストデータの評価における LLR (GT) に相当する。

LLR の性能は、SR と TR のどちらも高いときや低いとき (図 10(a), (d)) は Sum とほぼ同じ精度であるが、どちらかが低いとき (図 10(b), (c)) は Sum より良くなっていることが分かる。

LLR (Fixed) では顔の重みが歩容や身長と比べて大きく、SR が低いときは顔の精度はかなり低くなるため、LLR (Fixed) は SR が低いときは最も精度が悪くなっている (図 10(b), (d))。

さらに、EER の、SR を固定したときの TR による変化と TR を固定したときの SR による変化を図 8 に示す。表 6, 図 8 から分かるように、ほぼすべての時空間解像度の組で、特に SR と TR のどちらか、または両方低いときに LLR の精度が良くなっていることが分かる。

表 6 学習データのすべての空間解像度 [画素] と時間解像度 [fps] における EER [%]。太字は Sum, LLR (Fixed), LLR の中で最も精度が高いものを示す。

SR	TR	Fusion rule		
		Sum	LLR (Fixed)	LLR
640×480	30	0.8	0.8	0.8
	15	0.9	0.9	0.9
	7.5	1.5	1.4	1.5
	5	3.6	3.5	3.3
	3	9.6	9.1	6.0
	1	16.6	15.4	7.6
320×240	30	0.8	0.7	0.7
	15	0.8	0.8	0.8
	7.5	1.5	1.4	1.5
	5	3.9	3.9	3.6
	3	9.8	9.4	6.8
	1	16.6	15.8	8.6
160×120	30	0.8	0.8	0.8
	15	1.0	1.0	1.0
	7.5	2.1	2.3	2.0
	5	6.2	6.9	5.5
	3	12.1	12.7	10.6
	1	19.7	19.8	14.2
106×80	30	1.3	1.4	1.3
	15	1.6	1.8	1.5
	7.5	4.2	4.8	3.3
	5	11.8	13.3	8.3
	3	18.2	19.4	14.9
	1	26.4	27.2	20.0
80×60	30	2.5	2.9	1.8
	15	3.8	4.2	2.7
	7.5	7.4	8.2	5.3
	5	16.2	17.7	12.1
	3	23.0	24.4	19.7
	1	31.2	31.5	25.5
53×40	30	5.1	13.6	3.3
	15	9.8	18.1	7.0
	7.5	15.2	23.5	10.2
	5	21.3	28.1	16.3
	3	26.7	33.1	25.5
	1	32.2	42.0	31.0

4.5 テストデータの評価

ここではテストデータの 2 分割交差検定は 5 回繰り返し行った。また、ガウス過程回帰において、時空間解像度は対数で表し (たとえば半分サイズであれば $q_s = \log(0.5)$)、カーネル関数のパラメータは $r = 0.2$, $v = 1$ とした。

また、2 通りの時間解像度 (10fps と 2fps) と 2 通りの空間解像度 (3/4, 1/10) の組み合わせ 4 通りに対する ROC 曲線を図 10 に示す。

これより、Sum, LLR (Fixed) と比較して、提案手法である LLR(GPR) の結果が明らかに良くなっていることが分かる。また、SR と TR のどちらか、あるいはどちらも低いときもほぼ LLR (GT) と同等の精度になっていること

表 7 テストデータのすべての空間解像度 [画素] と時間解像度 [fps] における EER [%]. 太字は Sum, LLR (Fixed), LLR (GPR) の中で最も精度が高いものを示す。

Fusion rule	SR	480×360				213×160				128×96				91×68				64×48			
		10	6	3.75	2	10	6	3.75	2	10	6	3.75	2	10	6	3.75	2	10	6	3.75	2
Sum		0.9	2.5	8.6	17.1	1.2	3.4	9.7	19.0	2.8	6.2	14.0	23.6	5.2	9.3	18.9	29.3	8.3	13.8	24.0	33.5
LLR (Fixed)		1.0	2.3	8.3	16.9	1.3	3.7	10.0	19.4	3.7	7.7	15.4	25.3	6.7	11.5	21.3	31.8	9.7	15.3	25.7	35.6
LLR (GPR)		1.0	2.2	6.1	8.3	1.3	3.4	8.7	12.2	2.3	4.9	12.7	18.0	3.4	6.8	15.9	23.2	4.6	10.7	21.5	30.6
LLR (GT)		1.0	2.2	5.7	8.1	1.1	3.4	8.4	11.8	1.9	4.9	12.5	17.3	3.3	6.8	15.7	21.9	4.4	10.7	21.3	28.7

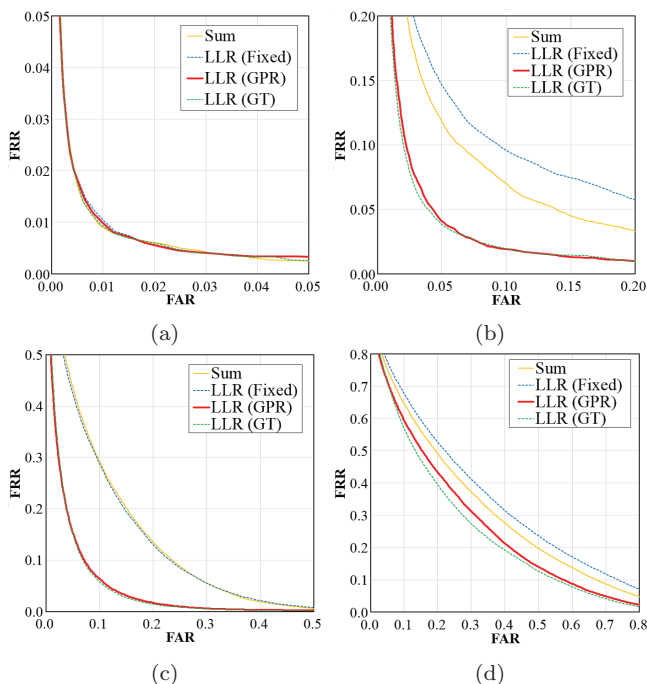


図 7 空間解像度 (左:480×360 画素, 右:64×48 画素), 時間解像度 (上:10fps, 下:2fps) における, テストデータの ROC 曲線

が分かる。

また, EER の結果を表 7 に示す。表 9 には典型的な時空間解像度での結果を示す。SR と TR の低下によって Sum や LLR (Fixed) の精度が低下していても, 重み付けによって LLR (GPR) の精度はあまり低下していないことが分かる。

5. おわりに

本論文では, 時空間解像度に適応的な歩容・顔・身長によるマルチモーダル個人認証のスコアレベル統合について述べた。最初に, 様々な時空間解像度における歩容・顔・身長の各モダリティに対する大規模スコアデータセットの作成について説明した。次に, そのデータセットを用いて歩容・顔・身長それぞれについて 1 対 1 認証による性能評価を行った。さらに, 構築したスコアデータセットに基づいて, 時空間解像度に適応的なスコア統合方法を提案した。結果として, 歩容, 顔, 身長は時空間解像度によって受ける影響が異なることを明らかにした。そして, そのことから時空間解像度に適応的に重み付けの必要性を明らかにし, 最適な重み付けをすることによって精度が向上する

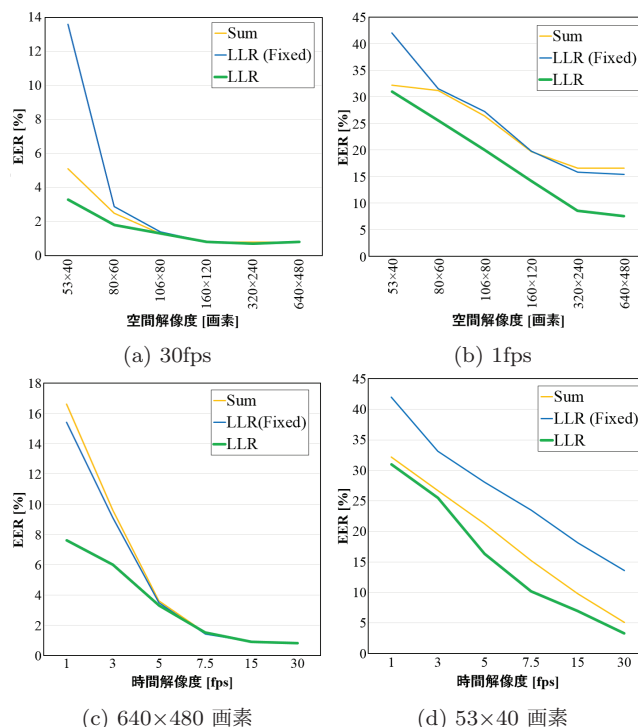


図 8 学習データにおける EER の, TR 固定での SR による変化 (上) と SR 固定での TR による変化 (下)。

ことを実験により確認した。

今後の課題として, まず, 今回用いたデータが室内で撮影されたものであることから, 屋外を含めたより実環境での性能評価が必要である。また, 本研究ではほぼ側面方向から撮影した歩行映像のみを用いて実験を行ったが, 観測方向によって歩容・顔・身長に対する重み付けが変化することも考えられるため, 観測方向の変化を考慮に入れた重み付け手法も必要である。

参考文献

- [1] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.
- [2] J. Wambaugh, *The Bleeding*. HarperCollins Publishers, 1989.
- [3] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd ed. Springer Publishing Company, Incorporated, 2009.
- [4] D. Zhang, *Palmprint Authentication*, ser. International Series on Biometrics. Springer Publishing Company, Incorporated, 2004, vol. 3.
- [5] M. J. Burge and K. W. Bowyer, *Handbook of Iris Recog-*

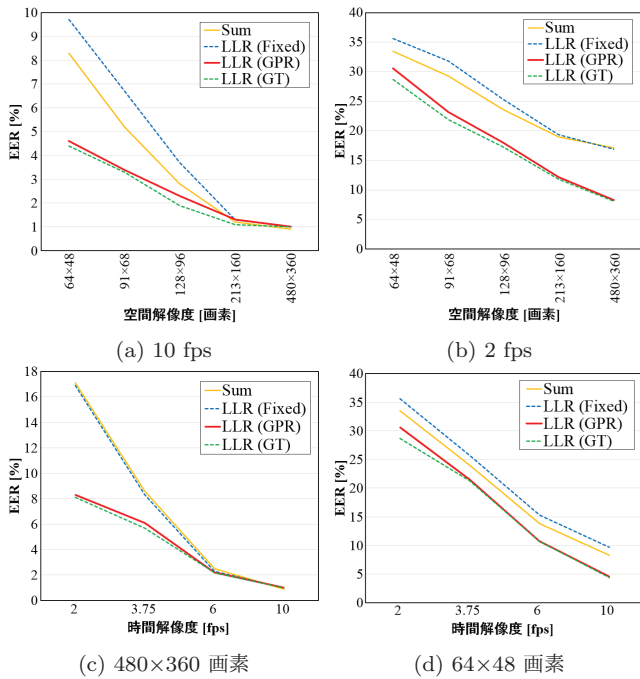


図 9 テストデータにおける EER の, TR 固定での SR による変化 (上) と SR 固定での TR による変化 (下)

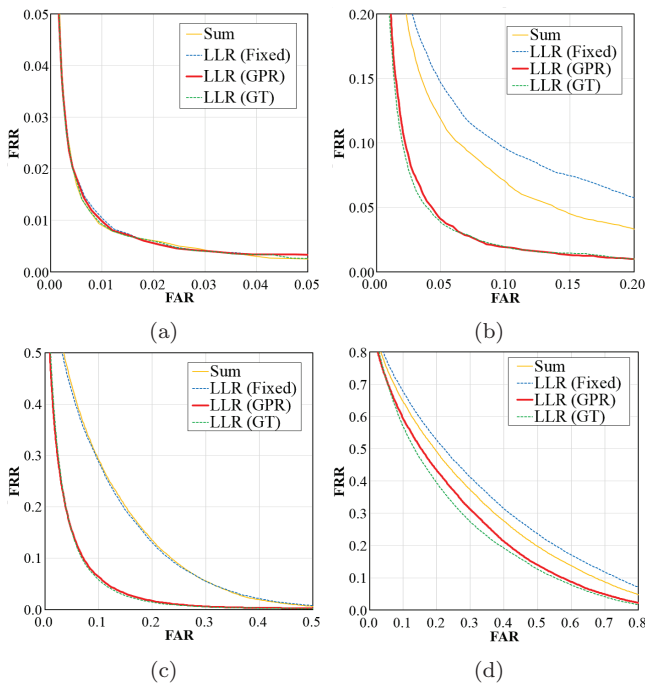


図 10 SR (左: 480x360 画素, 右: 64x48 画素) と TR (上: 10 fps, 下: 2 fps) におけるテストデータの ROC 曲線。

tion. Springer Publishing Company, Incorporated, 2013.

- [6] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
- [7] A. Osborn, *Questioned Documents*, 2nd ed. New York: Boyd Printing Company, 1929.
- [8] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, vol. 2013, no. 408280, pp. 1–24, 2013.
- [9] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook*

- of Multibiometrics*, ser. Int. Series on Biometrics. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [10] J. Basak, K. Kate, V. Tyagi, and N. Ratha, "Qplc: A novel multimodal biometric score fusion method," in *IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010*, San Francisco, CA, USA, Jun. 2010, pp. 1–7.
- [11] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognition*, vol. 38, no. 5, pp. 777–779, May 2005.
- [12] C. Boehnen, D. Barstow, D. Patlolla, and C. Mann, "A multi-sample standoff multimodal biometric system," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, 2012, pp. 127–134.
- [13] A. Kale, A. Roy-Chowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2004 (ICASSP'04)*, vol. 5, 2004, pp. 901–904.
- [14] X. Zhou and B. Bhanu, "Feature fusion of side face and gait for video-based human identification," *Pattern Recognition*, vol. 41, no. 3, pp. 778–795, 2008.
- [15] T. Zhang, X. Li, D. Tao, and J. Yang, "Multimodal biometrics using geometry preserving projections," *Pattern Recognition*, vol. 41, no. 3, pp. 805–813, 2008.
- [16] X. Geng, K. Smith-Miles, L. Wang, M. Li, and Q. Wu, "Context-aware fusion: A case study on fusion of gait and face for human identification in video," *Pattern Recogn.*, vol. 43, no. 10, pp. 3660–3673, Oct. 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2010.04.012>
- [17] M. Hofmann, S. M. Schmidt, A. Rajagopalan, and G. Rigoll, "Combined face and gait recognition using alpha matte preprocessing," in *Proc. of the 5th IAPR Int. Conf. on Biometrics*, New Delhi, India, Mar. 2012, pp. 1–8.
- [18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [19] D. Muramatsu, H. Iwama, Y. Makiyama, and Y. Yagi, "Multi-view multi-modal person authentication from a single walking image sequence," in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1–8.
- [20] H. Iwama, M. Okumura, Y. Makiyama, and Y. Yagi, "The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
- [21] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2006.
- [22] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia, "Dealing with sensor interoperability in multi-biometrics: the upm experience at the biosecure multimodal evaluation 2007," in *Proc. of SPIE 6994, Biometric Technologies for Human Identification IV*, Orlando, FL, USA, Mar. 2008.
- [23] C. K. I. W. Carl Edward Rasmussen, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.