見えに基づく判別的重み付き積分動き特徴による行動認識

松川　徹†　　栗田　多喜夫††

† 東京大学生産技術研究所　〒 153–8505 東京都目黒区駒場 4-6-1
† 広島大学大学院工学研究院情報部門　〒 739-8521 広島県東広島市鏡山 1-7-1
E-mail: †te2@iis.u-tokyo.ac.jp, ††tkurita@hiroshima.ac.jp

あらまし　本研究では，局所的な動きパターンの積分特徴の位置不変性を保ったまま認識率向上を行うために，見えに基づき動き特徴を重み付き積分することを提案する．特徴の座標値に基づき判別的に特徴量の重み付けを行うフィッシャー重みマップでは，積分特徴の利点である位置不変性を損なってしまう．提案アプローチでは，見えをフィッシャー重みマップの座標値と見なす．この重みが有効である理由は，何が動きているかやどのような姿勢における動きであるかという見えの文脈に基づく動きは認識対象の詳細な情報を捉えた判別力の高い特徴となるからである．具体的に局所領域における見えとフレーム内における見えという 2 種類の見え重み付けを検討する．また，認識精度向上のために動きに関しても重み付けを行う．公開データベースを利用した実験により提案手法の有効性を確認した．

キーワード　行動認識, 2 次元判別分析, 重みマップ, CHLAC 特徴，共起

1. はじめに

映像中の人物の行動認識は，映像の検索や監視，HCI 等の広範な用途に応用することの可能な重要な研究対象である．行動認識は対象カテゴリの行動の参照データとの照合により行うことができるが，この照合処理の前に特徴抽出が行われる．一般的な映像中の行動認識に用いられている局所的に算出された特徴量のパターンの大域的積分表現 ([1] 等) は，時間や場所に対して不変であるという初期的な表現であるため，行動の起こっている時間や位置の変動に頑健である．このような位置不変性を有する積分特徴を用いれば，対象の切り出しが不要な極めて単純な照合処理により，認識を行うことができる [2]．そこで本研究では位置不変性を保存しつつこれら局所パターンの積分特徴による行動認識精度を向上することに着目する．

局所的なパターンは予め設定されたパターンを利用するもの ([2] [3] 等) や局所特徴のクラスタリングによりパターンを決定するもの ([1] 等) がある．それらの局所的な領域のパターンは見えと動きに大別される．従来の研究の多くは動きの特徴を利用しているが，一枚の画像のみからも行動認識が行えることも確認されている ([4] 等)．すなわち見えと動きは共に重要な特徴量である．そこで複数の個別の特徴の大域的表現において見えと動き特徴の重み付き統合を用いて認識が行われる ([4] [5] 等)．例えば，Ikizler-Cinbis らは，人物や物体を中心とした見えや動き，背景の形状、色特徴をそれぞれ個別に Multiple Instance Learning (MIL) 表現を行い，重みづけを行っている [5]．これらは主に，水泳の飛び込みやテニス等の一般的な映像認識にプールやテニスコート等の背景情報を利用することを意図したものであるが，同一背景における行動認識においても特徴統合は有効である ([6] 等)．

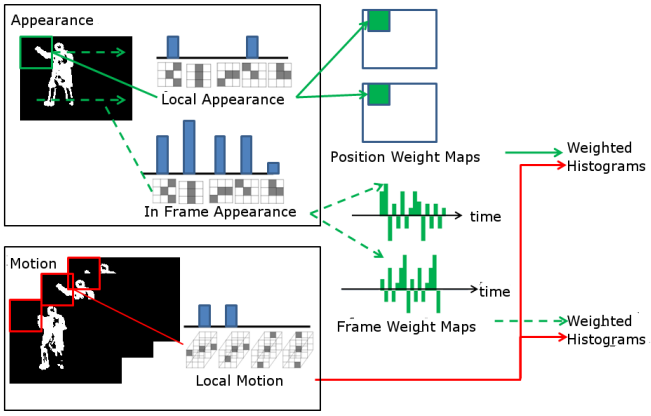

図 1　提案アプローチ．見えに基づき動き特徴を判別的に重み付き積分することが本研究の着想である．実際には，認識性能向上のため，動きの重み付けも行うが，これもフィッシャー重みマップに基づき自然に実現される．

これらの特徴は個別に算出した見え特徴と動き特徴を認識の段階で統合して利用しているが，それに対して何が動きているかやどのような姿勢における動きであるかという見えと動きの共起情報はより判別力の高い特徴となると思われる．特に一般動画像認識では，背景の動きなど認識に関係のない部分の動きも含まれる．それらの動きを見えの重要度に応じた区別を行うことが有効であると考えられる．そこで本研究では，動き特徴の判別性能を位置不変性を保ちつつ高めるために動き特徴を見えに基づき重み付き積分することを提案する．具体的に図 1 に示すような局所的な見えに基づく局所動きパターンの重み付けとフレーム全体における見えに基づいた局所動きパターンの重みという 2 種類の見え重みを検討する．さ

らに認識性能を高めるために動き特徴に関しても重み付けを行う．この特徴の重みの学習には構造化データの重み付けに適する 2 次元判別分析 [7](または，フィッシャー重みマップ [8]) を利用する．

提案法の同一位置の異なる特徴に基づき特徴を重み付き積分することは Top-Down Color Attension [9] から着想している．ただし，この研究の重み付けは本研究と異なり判別的ではない．また，判別的な重み付き積分はフィッシャー重みマップ [8] に基づいている．ただし，従来のフィッシャー重みマップは位置毎の重みを利用する．関連研究との比較の説明は 2 節に行うが，従来の共起表現手法に比べ 2 次元判別分析を重み学習に利用する利点を簡単にまとめる．まず，判別的基準を用いている点，多数の判別的な重みを獲得することが出来る点があり，これらは高い認識精度を達成するために有効である．判別的に特徴を重み付けや選択している手法との比較としては，まず AdaBoost に比べ，重み学習の計算量が高速であり多クラスの識別問題を自然に扱える点，通常の判別分析と比較して高次元の共起ベクトルを分割することなく，一度の学習により扱える点である．

2. 関 連 研 究

2.1 共起による特徴統合

共起による統合手法は，特徴量の共起の表現や選択の手法であり，MSF, Apriori, 出現頻度に基づく重み等がある．Markov Stationary Feature (MSF) [10] は，特徴の共起頻度を特徴間の遷移確率とみなし，その定流状態を特徴ベクトルとする手法であるが，提案手法と異なり同一次元数の特徴の共起に適用対象が限定され，また，クラス間判別のための基準は利用されない．APriori アルゴリズム [11] はクラス毎の頻出する特徴ベクトルの組を探索し，選択するが，提案手法と異なり判別的な情報は用いていない．出現頻度に基づく重みでは，クラス毎の特徴量の出現頻度に基づき重み付けが行われる．例えば，Top down color attension [9] では，色に基づいたクラスの事後確率分布に基づいた形状特徴のヒストグラムがクラス数個作成されるが，その重みは各クラスの色の特徴の出現頻度のみから決定されるため，形状特徴と合わせた際に判別的となっている保証はない．提案手法は特徴共起表現に判別的な重み付けを導入している．それゆえ，従来手法よりも識別性能が高いことが期待される．

行動認識における共起では，特徴の時空間的位置関係性を利用する認識手法の有効性も示されている ([12] [11] [13] 等) が，これらは基本的には離れた位置での同一種類の特徴共起を扱う手法であり，見えと動きの共起を扱った研究は少ない．3D-SIFT [14] のような見えと時間情報を含んだ時空間特徴があるが，一般に見えと動きのそれぞれの情報は個別に算出した特徴量に劣る．また，HOG/HOF 記述子 [14] 等の動きと見えの記述子を連結させた特徴記述子も存在するが，各特徴の重要度の重み付けは行われず，k-means クラスタリング等でパターン化される．人検出において見えと動き特徴の共起を検討しているものもある [15] が，検出器をスライドさせる必要があり位置不変特徴でない．また，固有値問題に基づき数分で重み学習が行える提案手法と異なり，誤認識サンプルに基づく重み付き弱識別器学習を複数ラウンド繰り返すことにより，サンプル数等の条件により異なるが，一般的に学習に数時間単位の時間を必要とする AdaBoost を利用している．

2.2 2 次元判別分析

本研究で特徴の重み付けに利用する 2 次元判別分析 [7] は，通常ベクトルに対して適用される判別分析を行列をベクトル化せずに直接適用するように拡張したものである．即ち，同じ列（見えの種類) に対して同一の重みを与える．顔認識等では，これは高次元，少数のサンプル問題に適することが示されている．2 次元判別分析は多数の論文で提案されているが，文献 [7] で初めて提案された手法である．詳しくは [16] 等が参考になる．

2 次元判別分析と同一な手法に Shinohara らの特徴ベクトルを位置毎に重み付けするフィッシャー重みマップがある [8]．この手法は特徴ベクトルを位置毎に並べた行列に 2 次元判別分析を利用している．Harada らは，これを領域ごとの特徴量の重み付けに利用し [17]，森下らはフーリエ基底に対して用することで時間重みを実現した [18]．しかし，座標値に基づく位置重みは本研究の目標である位置不変性の保存とは相反する考えである．また，歩容認証手法では，異なるスケールと方向のガボール特徴にテンソル判別分析が利用されている [19] が，やはり位置不変性を有せず，対象の切り出しを必要とする．彼らの手法が位置・時間の座標に依存した重み付けであるのに対して，提案アプローチは見えに依存した位置・時間の重み付けであるため，位置と時間に対する不変性が保存される．

3. 提案アプローチ

3.1 基 本 特 徴

本研究では，見え特徴と動き特徴の統合のみに着目する．そこで，簡単のためカメラが静止している状況を仮定する．また，人物の大きさや視点変動は少ないものとする．こうした状況下において動き特徴として CHLAC [2]，見え特徴として HLAC を利用する．これらを利用する理由は，k-means クラスタリング等を利用して特徴パターンを算出する手法に比較して，特徴抽出処理が高速に行えるため提案手法の基礎的な評価が容易になるため，処理の簡易さに比べ認識精度が高いためである．また，データセット毎に特徴パターンを学習する必要がなく，新しいデータセットに即座に対応できるため，実応用上，望ましいことも付記しておく．

まず，入力動画像をフレーム間差分と閾値処理によって

2 値画像列に変換する．本研究では, CHLAC において 2 値画像を算出する閾値を下げることにより多くの特徴点検出を利用することにより認識精度が向上すること [13] に基づいて, 大津の 2 値化を用いず, 3x3 pixel の SSD が 500 以上の場所を 1, それ以外では, 0 とする処理により実装した．もしカメラが静止していれば, このような 2 値画像列は動きいている領域のシルエットを表す．時空間の領域 $D : X \times Y \times T$ において, 点を$\boldsymbol{r}= (x, y, t)^t$，2 値動画像の各点の値を $\mathrm{f}(\boldsymbol{r}) \in \{0, 1\}$ とする．ここで, X と Y は画像フレームの幅と高さであり, T はフレーム長である．ここで $\mathrm{f}(\boldsymbol{r}) = 1$ は動いた点, $\mathrm{f}(\boldsymbol{r}) = 0$ は静止している点を表す．

本研究では $\{\boldsymbol{r}|f(\boldsymbol{r}) = 1\}$ を満たす全ての点において, HLAC と CHLAC [2] の自己相関関数を算出する．

HLAC の自己相関関数は以下のように定義される．

$$v(\boldsymbol{r}) = f(\boldsymbol{r})f(\boldsymbol{r} + \boldsymbol{a}_1) \cdots f(\boldsymbol{r} + \boldsymbol{a}_N), \qquad (1)$$

ここで, パラメータ$\boldsymbol{a}_n= (a_{nx}, a_{ny}, 0)^t, n = 1, ..., N$ は画像平面における変位ベクトルである．これらのパラメータは $a_{nx}, a_{ny} \in \{\pm\Delta r, 0\}$, $N \in \{0, 1, 2\}$ に制限する．一つのパラメータ $(\boldsymbol{a}_1, ..., \boldsymbol{a}_N)$ の組みに対応する自己相関関数 $v(\boldsymbol{r})$ の値が HLAC の一つの次元である．CHLAC と HLAC のパターンを組み合わせる際に独立となるため, 本研究では, 平行移動によって生じる重複したパラメータの組を除去しない．この場合, HLAC のパラメータの組のパターンは 37 個となる．空間的にさまざまなスケールの特徴を捉えるため，HLAC は複数個のスケールで算出され，各スケールの特徴を連結させることが，しばしば行われる．空間間隔数の個数は任意である．本研究では, 5 つの空間間隔 $(\Delta r = 1, 2, 4, 8, 12)$　によって算出された HLAC 特徴のパターンを連結させる．$N = 0$ $(\mathrm{v}(\boldsymbol{r}) = \mathrm{f}(\boldsymbol{r}))$ のパターンは 5 つのスケールで共通であるため, その重複を除くと, 各点$\boldsymbol{r}$に対して $(36\times5 + 1) = 181$ 次元の見えパターンの出力値を得る．

HLAC の自然な時空間拡張である CHLAC の自己相関関数は以下のように定義される．

$$h(\boldsymbol{r}) = f(\boldsymbol{r})f(\boldsymbol{r} + \boldsymbol{a}'_1) \cdots f(\boldsymbol{r} + \boldsymbol{a}'_N), \qquad (2)$$

ここで $\boldsymbol{a}'_n= (a_{nx}, a_{ny}, a_{nt})^t, n = 1, ..., N$ は画像平面と時間における変位ベクトルである．これらのパラメータは $a_{nx}, a_{ny} \in \{\pm\Delta r, 0\}$, $a_{nt} \in \{\pm\Delta t, 0\}$, $N \in \{0, 1, 2\}$ に制限する．この場合, 352 個の $(\boldsymbol{a}'_1, ..., \boldsymbol{a}'_N)$ が得られる．本研究では, CHLAC の空間間隔と時間間隔を $\Delta r = 4$ と $\Delta t = 1$ にそれぞれ設定したが，その理由はこれらのパラメタが実験において最も良い性能を示していたからであり，他のパラメータに設定しても提案手法の本質的な問題はない．また，CHLAC も時間情報のパターンを含んでいるが, 同一次数において HLAC のマスクパターンにおける画像情報よりも自己相関点数が少ない．従って, CHLAC と HLAC の組合わせは画像平面に対してより詳細な情報を抽出することが可能である (図 2.).

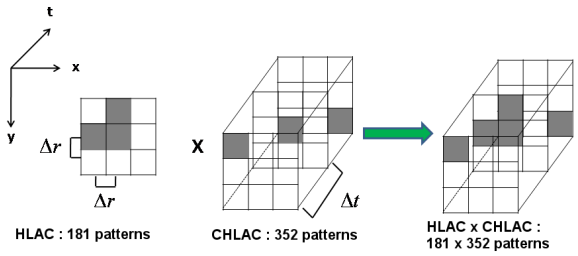

図 2　局所的な見えパターンと動きパターンの統合．本図は HLAC と CHLAC の組合わせは各々の特徴より詳細なパターンを捉えることを示すものであり，提案法は最終的に各特徴の複数個の重みづけで統合される (3.5 節).

3. 2　局所的な見え重み付き積分動き特徴

本節では，局所的な見えに基づく動き特徴の概要を説明する．参照点$\boldsymbol{r}$における d_2 次元の動き特徴を $\boldsymbol{H}(\boldsymbol{r}) = (h_1(\boldsymbol{r}), ..., h_{d_2}(\boldsymbol{r}))^t$ としたとき，位置に基づく重み付き積分特徴として, 画像領域における研究としてフィッシャー重みマップ [8] [17] がある．これを直接動き特徴の重みに利用した場合，重み付き動き特徴は，$\boldsymbol{y}$ $= \int_{\boldsymbol{r}} w(\boldsymbol{r})\boldsymbol{H}(\boldsymbol{r})d\boldsymbol{r}$ となり, 行動認識対象の存在する位置の変動を受けてしまう．そこで見えに基づく動き特徴として, $\boldsymbol{y} = \int_{\boldsymbol{r}} w(\boldsymbol{V}(\boldsymbol{r}))\boldsymbol{H}(\boldsymbol{r})d\boldsymbol{r}$ を提案する．ここで，$\boldsymbol{V}(\boldsymbol{r}) = (v_1(\boldsymbol{r}), ..., v_{d_1}(\boldsymbol{r}))^t$ を参照点$\boldsymbol{r}$における d_1 次元の見えパターンとする．本研究では，見え重みと位置毎の見えパターンの線形性を仮定する．すなわち，見え重み係数 $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{d_1})^t$ を用いて，見え重みを $w(\boldsymbol{V}(\boldsymbol{r})) = \boldsymbol{V}(\boldsymbol{r})^T\boldsymbol{\alpha}$ と定義する．具体的な見え重み係数の学習方法は, 3.4 節で説明するが，以下では，重み係数の学習に必要となる，重み付き動き特徴を共起行列と重み係数から計算する方法について述べる．

まず, 同一位置における見えパターン $\boldsymbol{V}(\boldsymbol{r})$ と動きパターン $\boldsymbol{H}(\boldsymbol{r})$ を算出する．そうしたら, $\boldsymbol{V}(\boldsymbol{r})$ と $\boldsymbol{H}(\boldsymbol{r})$ を以下の共起行列に組み合わせる．

$$\boldsymbol{X}_R(\boldsymbol{r}) = \boldsymbol{V}(\boldsymbol{r}) \otimes \boldsymbol{H}(\boldsymbol{r})^T, \qquad (3)$$

ここで $\otimes$ はクロネッカー積であり, $\boldsymbol{X}_R(\boldsymbol{r})$ は (i, j) 成分が $v_i(\boldsymbol{r})$ と $h_j(\boldsymbol{r})$ の積で構成される $d_1 \times d_2$ 次元の行列である．

はじめに述べたように提案手法は，行動の起こっている時間や位置に対する不変性を持ちつつ認識精度を向上させることを目的としている．映像の一定区間内で時間や位置に不変な特徴を算出するため, 3 次元のボリューム D において特徴量 $\boldsymbol{X}_R(\boldsymbol{r})$ の積分を行う．即ち,

$$\boldsymbol{X}_R = \int_{\boldsymbol{r}\in D} \boldsymbol{X}_R(\boldsymbol{r})d\boldsymbol{r}. \qquad (4)$$

を算出する．そうしたら, 見え重み係数 $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{d_1})^t$ を用いて d_2 次元の特徴量 $\boldsymbol{y}_R$ を次式のように定義する．

$$\boldsymbol{y}_R = \boldsymbol{X}_R^T\boldsymbol{\alpha}. \qquad (5)$$

このように定義した特徴は，以下のように局所的な見え重み付き積分特徴量であることが解る．

$$\boldsymbol{y}_R = \int_{\boldsymbol{r}\in D} \boldsymbol{X}_R(\boldsymbol{r})^T\boldsymbol{\alpha}d\boldsymbol{r} = \int_{\boldsymbol{r}\in D} \boldsymbol{H}(\boldsymbol{r})\otimes\boldsymbol{V}(\boldsymbol{r})^T\boldsymbol{\alpha}d\boldsymbol{r} = \int_{\boldsymbol{r}\in D} w(\boldsymbol{V}(\boldsymbol{r}))\boldsymbol{H}(\boldsymbol{r})d\boldsymbol{r}.$$

ここで 見え重み $w(\boldsymbol{V}(\boldsymbol{r})) = \boldsymbol{V}(\boldsymbol{r})^T\boldsymbol{\alpha}$ はスカラー値である．この重みは見えから決定される重みであり，共起の要素毎の重みとは異なる．なお，重み係数が学習された後でテストサンプルに特徴抽出に利用する場合には，式 (5) により特徴を抽出してもよいし，$\boldsymbol{y}_R = \int_{\boldsymbol{r}} \boldsymbol{V}(\boldsymbol{r})^t\boldsymbol{\alpha}\boldsymbol{H}(\boldsymbol{r})d\boldsymbol{r}$ により特徴を抽出してもよい．

3.3 フレーム見え重み付き積分動き特徴

行動には認識に重要なフレームがあると考えられるため，本研究では局所的な見え重みを拡張し，フレーム見え重みも検討する．本研究では, フレーム見え特徴に局所的な見えパターンのヒストグラムを用いる．フレーム番号を t, フレームの領域を $D(t)$ とすると, フレーム内の見え $\boldsymbol{V}(t) = (v_1(t), ..., v_{d_1}(t))^t$ を, $\boldsymbol{V}(t) = \int_{\boldsymbol{r}\in D(t)} V(\boldsymbol{r})d\boldsymbol{r}$ とする．ここで, フレーム内の見え特徴を L1 ノルム正規化をする．即ち, $\boldsymbol{V'}(t) = \frac{V(t)}{|V(t)|}$ とする．正規化を行わない場合, 特徴点が多いフレームの重みが大きくなるため, この正規化は重要である．そうしたら, フレーム内の見えと局所動きとの共起行列を

$$\boldsymbol{X}_F(\boldsymbol{r}) = \boldsymbol{V'}(t)\otimes\boldsymbol{H}(\boldsymbol{r})^T, \tag{6}$$

とする．この場合,

$$\boldsymbol{X}_F = \int_{\boldsymbol{r}\in D} \boldsymbol{X}_F(\boldsymbol{r})d\boldsymbol{r} = \int_t\int_{\boldsymbol{r}\in D(t)} \boldsymbol{V'}(t)\otimes\boldsymbol{H}(\boldsymbol{r})^T d\boldsymbol{r}dt = \int_t \boldsymbol{V'}(t)\otimes\int_{\boldsymbol{r}\in D(t)} \boldsymbol{H}(\boldsymbol{r})^t d\boldsymbol{r}dt = \int_t \boldsymbol{V'}(t)\otimes\boldsymbol{H}(t)^T dt.$$

である．ここで，$\boldsymbol{H}(t) = \int_{\boldsymbol{r}\in D(t)} H(\boldsymbol{r})d\boldsymbol{r}$ である．即ち,CHLAC マスクパターン値の出力ヒストグラムをフレーム内で一回算出し, HLAC のフレーム内の出力値と一回掛け合わせることで共起行列 $\boldsymbol{X}_F(\boldsymbol{r})$ が算出できる．具体的に実験では$\boldsymbol{X}_F$ は，$\boldsymbol{X}_R$ と比較して約 2.5 倍高速に算出出来ることが確認された．局所的な見え重みと同様にそうしたら, 見え重み係数 $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{d_1})^t$ を用いて d_2 次元の特徴量 $\boldsymbol{y}_F$ を次式のように算出する．

$$\boldsymbol{y}_F = \boldsymbol{X}_F^T\boldsymbol{\alpha}. \tag{7}$$

この式は以下のように書き変えることが出来る．

$$\boldsymbol{y}_F = \int_t \boldsymbol{X}_F(t)^T\boldsymbol{\alpha}dt = \int_t \boldsymbol{H}(t)\otimes\boldsymbol{V'}(t)^T\boldsymbol{\alpha}dt = \int_t w(\boldsymbol{V'}(t))\boldsymbol{H}(t)dt = \int_t w(\boldsymbol{V'}(t))\int_{\boldsymbol{r}\in D(t)} H(\boldsymbol{r})d\boldsymbol{r}dt.$$

ここでフレーム見え重み $w(\boldsymbol{V'}(t)) = \boldsymbol{V'}(t)^T\boldsymbol{\alpha}$ はスカラーである．局所見え重みが場所毎に重みを算出するのに対しフレーム重みはフレーム毎に 1 回重みを算出するため, フレーム内見え重みは局所見え重みよりも高速に算出することが出来る．

また, 実験ではフレーム内見えに基づく局所的見え重み, フレーム内動きに基づく局所的動き重みも検討する．フレーム毎の特徴量の正規化を行わない場合, これらの特徴行列の上三角行列はフレーム毎の特徴を動画像で積分した場合の GLC [17] と等価となる．

3.4 2 次元線形判別分析による重み係数学習

次に重み係数$\boldsymbol{\alpha}$の決定法を説明する．3.2, 3.3 節の議論より，$\boldsymbol{X}^T\boldsymbol{\alpha}$の判別力が高くなるように重み係数を決定すれば，$\boldsymbol{y}$の判別力の高くなる見え重み $w(\boldsymbol{V}(\boldsymbol{r}))$ が求まることが分かる．そこで，本研究ではこのような問題のために提案されている 2 次元線形判別分析 [7] を利用する． 各クラス k に対して m_k 個の共起行列 $\{\boldsymbol{X}_i^{(k)} \in R^{d_1\times d_2}\}, i = 1, ..., m_k$ が学習用にあるとする．まず, 一般化されたクラス内共分散行列 $\boldsymbol{S}_W \in R^{d1\times d1}$ とクラス間共分散行列 $\boldsymbol{S}_B \in R^{d1\times d1}$ を次式で定義する．

$$\boldsymbol{S}_W = \sum_{k=1}^{C}\sum_{i=1}^{m_k}(\boldsymbol{X}_i^{(k)} - \boldsymbol{M}^{(k)})(\boldsymbol{X}_i^{(k)} - \boldsymbol{M}^{(k)})^T, \tag{8}$$

$$\boldsymbol{S}_B = \sum_{k=1}^{C} m_k(\boldsymbol{M}^{(k)} - \boldsymbol{M})(\boldsymbol{M}^{(k)} - \boldsymbol{M})^T. \tag{9}$$

ここで m_k はクラス c_k のサンプル数, C はクラス数, $\boldsymbol{M}^{(k)} = \frac{1}{m_k}\sum_{i=1}^{m_k}\boldsymbol{X}_i^{(k)}$ は クラス c_k の $\boldsymbol{X}_i$ の平均 , $\boldsymbol{M} = \frac{1}{\sum_{k=1}^{C} m_k}\sum_{k=1}^{C}\sum_{i=1}^{m_k}\boldsymbol{X}_i^{(k)}$ は全学習サンプルの平均である．そうしたら, 2 次元に拡張されたフィッシャー判別基準を以下のように定義する．

$$J(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T\boldsymbol{S}_B\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\boldsymbol{S}_W\boldsymbol{\alpha}}. \tag{10}$$

この基準を最大化するの重み係数 $\boldsymbol{\alpha}$ は, 次式の一般化固有値問題の最大固有値に対応する固有ベクトルとして得られる．

$$\boldsymbol{S}_B\boldsymbol{\alpha} = \lambda\boldsymbol{S}_W\boldsymbol{\alpha}. \tag{11}$$

この一般化固有値問題を解くため, 本研究では，式 (11) を $\boldsymbol{S}_W$ の固有ベクトル分解を用いて通常の固有値問題へ帰着させる方法 [20] を用いた．実際には，重み係数は一つに限定することなく，複数の重み係数に対する重み付き積分特徴を算出し，それらを連結させた特徴量を認識に用いることにより認識精度が向上する．そこで，式 (10) の基準を最大化する上位 s 個の重み係数を $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_s$ として, 式 (11) の一般化固有値問題の上位 s 個の固有値に対応する固有ベクトルを用いる, ここで，$\boldsymbol{A} = [\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_s]$ は $d_1 \times s$ 行列である．重み付けにより得られる行列 $\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{A}$ をベクトル化して $d_2\times s$ 次元の特徴ベクトルを得る．

ここで, [8] 等では触れられていないが, 1 次元の線形判

別分析と異なり，2 次元の線形判別分析の s は $C-1$ 次元よりも高くとることができる．具体的に 2 次元判別分析は，クラス数が Cd_2，サンプル数が d_2 倍となった d_1 次元特徴の判別分析の形になることを示すことができる [21]．

1 次元の判別分析による重み付けとの違いをさらに述べる．例えば共起の要素毎の重みとして，共起行列の要素に対して独立な重み $\boldsymbol{\Gamma} \in R^{d1\times d2}$ により $\boldsymbol{X}\cdot\boldsymbol{\Gamma}$，(ここでの $\cdot$ は要素毎の積）なる重み付けを $\boldsymbol{X}$ をベクトル化することにより行うことが考えられるが，これには重みの次元数が $d_1\times d_2$ 次元と高くなり，クラス内分散，クラス間分散行列の次元は $d_1d_2\times d_1d_2$ となる．これには重み学習に際してメモリの不足や学習サンプルの不足といった問題があり，特徴ベクトルを分割する等の不自然な処理を必要とする．2 次元判別分析では，クラス内分散，クラス間分散行列の次元は $d_1\times d_1$ 次元となり，特徴ベクトルを分割する必要なく学習が行える．

3.5 双線形重み（動きの重み係数の学習）

前節の手法で学習を行った重みは見えに関する重みであり，動きの特徴に関しては重みづけが行われていない．しかし，動きの特徴も重み付けが有効であると考えられるため，見えに基づき重み付けを行う前に動きのパターンに対しても重み付けを行うことでより高い認識性能が達成できることが期待される．この動きと見えの重みづけは，[21] で最初に提案されているように 2DLDA を行列の 2 方向に適用することにより自然に実現することができる．まず，動きに関する重み係数 $\boldsymbol{\beta}=(\beta_1,...,\beta_{d_2})^t$ を用いて次式のように行列を重み付けする．

$$\boldsymbol{X}'=(\boldsymbol{X}^T)^T\boldsymbol{B}=\boldsymbol{X}\boldsymbol{B}, \tag{12}$$

ここで，$\boldsymbol{B}=[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,..,\boldsymbol{\beta}_l]$ は $d_2\times l$ 行列である．重み係数 $\boldsymbol{\beta}$ の学習法は前節の方法と同様である．即ち，一般化クラス内共分散行列 $\boldsymbol{S}'_W\in R^{d2\times d2}$ とクラス間共分散行列 $\boldsymbol{S}'_B\in R^{d2\times d2}$ を行列 $\boldsymbol{X}^T$ に対して作成する．そうしたら，l 個の重み $\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_l$ は次式の固有値問題の上位 l 個の固有値に対応する固有ベクトルとして得られる．

$$\boldsymbol{S}'_B\boldsymbol{\beta}=\lambda\boldsymbol{S}'_W\boldsymbol{\beta}. \tag{13}$$

次に，$d_1\times l$ 行列である $\boldsymbol{X}'$ に対して重み係数を学習することにより，$d_1\times s$ の重み係数行列 $\boldsymbol{A}'=[\boldsymbol{\alpha}'_1,...,\boldsymbol{\alpha}'_s]$ を得る．最終的に次式により重み付けされた特徴量，

$$\boldsymbol{Y}'=(\boldsymbol{X}')^T\boldsymbol{A}'=\boldsymbol{B}\boldsymbol{X}^T\boldsymbol{A}', \tag{14}$$

を得る．ここで，$\boldsymbol{Y'}$ は $l\times s$ 行列であり，この行列をベクトル化して識別に用いる．この $\boldsymbol{Y'}$ の (i,j) 要素は，$\boldsymbol{y'}_{i,j}=\boldsymbol{\beta}_i^t\boldsymbol{H}\boldsymbol{V}^T\boldsymbol{\alpha'}_j$ であり，例えば局所的な動きと局所的な見えとの統合の場合，$\boldsymbol{y'}_{i,j}=\int_{\boldsymbol{r}}\boldsymbol{\beta}_i^t\boldsymbol{H}(\boldsymbol{r})\boldsymbol{V}^T(\boldsymbol{r})\boldsymbol{\alpha'}_j d\boldsymbol{r}=\int_{\boldsymbol{r}}w_j(\boldsymbol{V}(\boldsymbol{r}))\boldsymbol{\beta}_i^t\boldsymbol{H}(\boldsymbol{r})d\boldsymbol{r}$ という特徴が算出されることとなる (図 3)．またこれは，各点 $\boldsymbol{r}$ における複数個のパターンの混合から算出される動き重み $w_i(\boldsymbol{H}(\boldsymbol{r}))=\boldsymbol{H}(\boldsymbol{r})^T\boldsymbol{\beta}_i$ と見え重み $w_j(\boldsymbol{V}(\boldsymbol{r}))=\boldsymbol{V}(\boldsymbol{r})^T\boldsymbol{\alpha'}_j$ との積（共起）特徴が積分されているとも見れる．

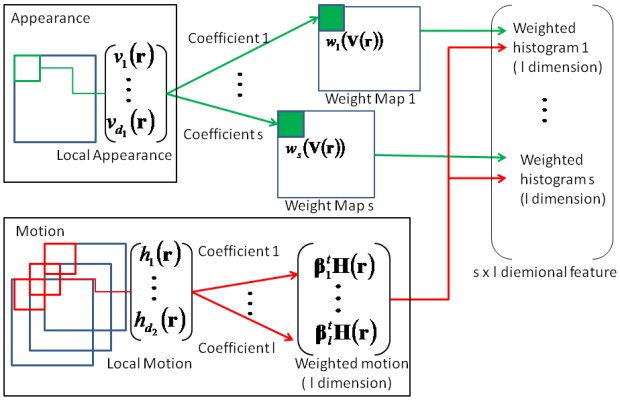

図 3　双線形重み付き積分特徴．

4. 実　　験

4.1 実験条件

提案特徴統合手法の有効性を確認するため，KTH human action dataset [22](KTH) と UT-interaction dataset(UT) [12] を用いて実験を行った．KTH は 25 人の被験者が 4 つのシナリオ (屋外の場合，屋外で撮影のスケール変動が起きる場合，屋外で異なる衣服を着ている場合，屋内の場合) で演じた 6 種類の行動を含むデータセットである (図 4 上部)．24 人の被験者のビデオを学習サンプルに用い，残りの被験者のビデオをテストサンプルに用いることを 25 回行った平均結果である leave-one-out cross validation 法による評価を行った．

UT は，6 種類の人間同士のインタラクションを含んだクラスを識別するデータセットである (図 4 下部)．このデータセットには，背景の動きやカメラの jitters/zoom 等も含まれる．このデータセットは，set1 と set2 という異なる場所で撮影された 2 つの set からなり，それぞれで提案手法を評価する．このデータセットで用意されているサンプル数は各セット内で各動作カテゴリ 10 シーケンスずつ，全動作合計 60 シーケンスである．

UT データセットにおいても leave-one-out cross validation 法を用いている．ただし，UT データセットのサンプル数が少ないため，学習段階では左右反転したシーケンスも利用し，学習サンプル数を 2 倍に増加させている．テスト段階では，左右反転していないもののみを利用している．このような方法によるサンプル数増加は，右の動きか，左の動きかを識別する場合を除き，一般性を失わず用いることができる．

識別には，線形 SVM を one-against-all 方式で用いた．識別の前に，特徴量の各次元を（全学習データ上で）各次元の平均がゼロに，標準偏差が 1 になるように正規化した [13]．学習データ上の 5 分割交差確認法を SVM の

図 4　KTH の例 (上部), UT データセットの例 (下部).

パラメータを調整するために実行した．サンプル数の少ない UT データセットでは k-NN を用いている．k-NN ではユークリッド距離を用いている．k の値は k=5 としたが，異なる値でも問題ない．

4.2　学習された重み

KTH において提案手法により獲得された局所的な見え重みの例 (双線形重みの場合) を図 5 に示す．図中の赤や青は符号の正負を示す．重みの解釈はやや困難であるが, それぞれ異なる重み付けがされていることが解る．また, 3 番目の重みで全て同一符号となったのは, 動き特徴を始めに重み付けしているからであり, 見えに関しては同一符号を用いても判別力があるからであると思われる．UT データセットにおける同様な例も図 6 に示す．KTH と同様な同一符号重みが 1 番目の重みに表れている．また，５番目の重みが特徴的であり，手や足の内部等の重みが強くなっており，これは HLAC の自己相関の次数の高いパターンの重みが強かったためと思われる．

また, KTH において提案手法により獲得されたフレーム見え重みの絶対値の例 (双線形重みの場合) を図 7 に示す．図は walking における第 2 固有ベクトルと第 4 固有ベクトルに対応するものであるが，局所的な見え重みの場合と同様に局所的な動き重みの判別力が高いことによると思わる要因により, 第 1 固有ベクトルの変化が小さかったためにこれらを例としている．この例では, 第 2 固有ベクトルでは足が閉じた状態における動きが強調され，第 4 固有ベクトルでは足をやや開いた状態での動きが強調されることが分かる．これらの傾向は running や jogging にも当てはまる傾向であった．他のカテゴリのフレーム重みでは，類似したシルエットに対して同一な重み学習されているものの，それがそのカテゴリを象徴するものであるかという直観的な理解は難しかった．

4.3　提案手法の認識結果

まず, 提案手法を CHLAC と HLAC の単純な結合法と比較を行った．単純な個別特徴の結合法の実験結果を表 1 に示す．比較手法として線形判別分析 (LDA) を利用しているが, 圧縮次元数を 5 次元に設定している．CHLAC や HLAC を判別分析にかけること, CHLAC と HLAC を両方用いることにより, 認識性能が向上していることが解る．

また, KTH における提案手法の認識率を表 2 に示す．双方向の重みを用いる BiWeight(2DLDA) が 見えのみ

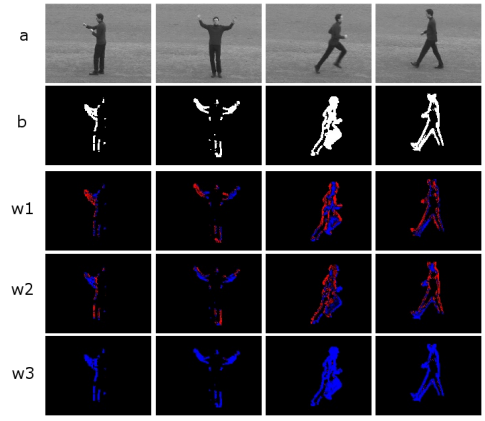

図 5　$\boldsymbol{X}'$の局所的な見え重み $w(\boldsymbol{V}(\boldsymbol{r}))$ 例 (KTH). a: 濃淡画像, b: フレーム差分と 2 値化, w1-3 : 第 1 ー第 3 固有値に対応する重みマップ.

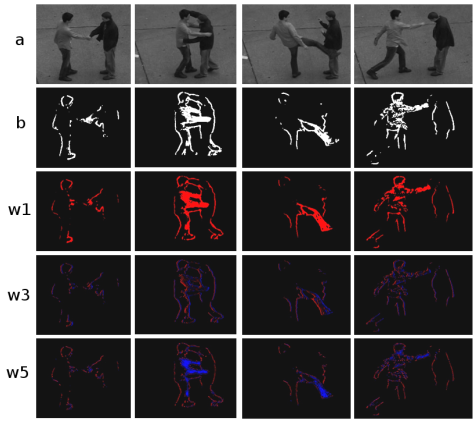

図 6　$\boldsymbol{X}'$の局所的な見え重み $w(\boldsymbol{V}(\boldsymbol{r}))$ 例 (UT). 図 5 と同様.

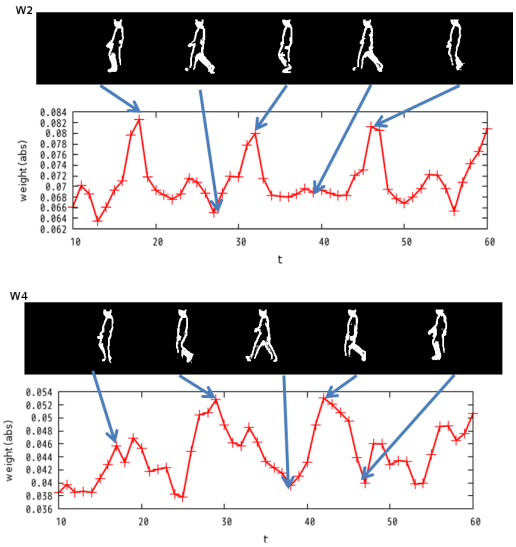

図 7　$\boldsymbol{X}'$のフレーム見え重み $w(\boldsymbol{V}'(t))$ の例 (KTH). w2,w4 : 第 2, 第 4 固有値に対応する重みマップ.

に重み付けする Weight(LDA) よりも認識精度が向上することが解る．また, 重みの数を増やすことで認識精度が向上していることが解る．また, 表 2A,B,C,D はそれぞれの特徴量の重み付けの組を示しているが, A の局所

表 1　単体特徴の単純結合法の認識結果 (%) .

Method	Dim.	KTH	UT-Set1	UT-Set2
LDA(CHLAC+HLAC)	5	89.74	60.00	70.00
LDA(CH.)+LDA(HL.)	5+5	88.65	53.33	68.33
LDA(CHLAC)	5	88.99	51.60	61.66
LDA(HLAC)	5	79.38	45.00	58.33
CHLAC+HLAC	352+181	87.70	33.33	46.66
CHLAC	352	87.56	38.33	45.00
HLAC	181	73.61	35.00	45.00

的な見えに基づく重み付け局所動き特徴が最も認識精度が高く, 次に B のフレーム見えに基づく重み付け局所動き特徴の認識精度が高い．また, C のフレーム見えに基づく重み付け局所見え特徴の認識率は 84.58% であり, HLAC に LDA を適用した LDA(HLAC) の 79.38%よりも認識率が高いが, D のフレーム動きに基づく重み付け局所動き特徴の認識率 86.94%は, CHLAC に LDA を適用した LDA(CHLAC) の 88.99%よりも認識率が低い．また，UT データセットにおける提案手法の認識率を表 3 に示す．UT データセットにおいては局所的な見え重みの結果のみを報告しているが，これはこのデータセットにおいてフレームに基づく重み付けの結果が不安定であったためである．その要因としては，データに縞の服の人物が含まれており，縞をシルエットとして抽出することによる画像全体特徴の不安性や背景に関連のない人物が存在するデータがあること，学習データが少ないことなどが考えられる．

重み次元数毎の認識率の違いを図 8 に示す．これは局所的な見え重みにおける比較である．双方向の重みにおいて, $C-1$ の次元数である 5 よりも多い重みを用いることで認識率が向上していることが解る．LDA が重みの数が最大 $C-1$ 個に限定されるのに対して, 2DLDA はより高い数の重みを得ることが出来る．これが, 2DLDA による重み付けの利点の一つである．

また，計算時間を Xeon 2.66GHz の 1 コア使用，2GB RAM の PC を利用し，C++による実装で計測を行った．KTH における特徴抽出に要した時間は局所共起特徴$\boldsymbol{X}_L$が，1 フレームあたり 21msec, フレーム共起特徴$\boldsymbol{X}_F$が 7.9msec であった．HLAC 特徴のみを算出した場合においても 7.3msec 掛かっており，フレーム共起特徴による特徴算出時間の増加は HLAC と CHLAC を算出した場合に比べても少ないことが確認された．1 回の評価セット毎の判別的な重み係数の学習時間は$\boldsymbol{X}_L$に対する係数が 67.05(動き係数)+8.12(見え係数)sec, $\boldsymbol{X}_F$ に対する係数が 45.68(動き係数)+3.1(見え係数)sec と重み係数の学習も現実的な計算時間 (約 1 分) で行うことが出来た．

4.4　提案手法の結合の認識結果

局所見えとフレーム見え重みは別種な物であるので, それらを結合してを利用して精度が向上するかを調べた．認識結果を表 4 に示す．表において 5,10,..,100 は局所見えまたはフレーム重みの数であり, その数の重みで重み付けされた特徴量の結合を意味する．(I) が局所見え重みを利用した局所動き積分特徴である．局所見え重みとフレーム見え重みを両方用いること (II) により, 局所見え重みのみから認識精度が向上することが解る．さらにフレーム見え重みを導入した局所見え特徴を加えること (III) により, 認識率が更に若干上昇することが解る．また, フレーム動き重みを導入した局所動き特徴を加えた場合 (IV) では, 認識率が向上しなかった．これは, LDA(CHLAC) に比べて, 認識精度が向上していない, つまりフレーム動き重みは見え重みよりも有効でなかったためであると考えられる．

表 2　共起特徴の KTH の認識結果 (%)．A: 局所見え重み × 局所動き, B: フレーム見え重み × 局所動き, C：フレーム見え重み × 局所見え, D: フレーム動き重み × 局所動き.

Method	Dim.	A	B	C	D
BiWeight(2DLDA)	80×100	**95.38**	**93.94**	**84.58**	**86.94**
BiWeight(2DLDA)	80×5	91.46	90.67	80.01	82.06
BiWeight(2DPCA)	80×100	93.47	92.23	81.90	82.72
BiWeight(2DPCA)	80×5	87.01	86.17	80.10	83.52
Weight(2DLDA)	352×5	89.83	86.76	81.86	83.15
MSF	181or352	-	-	77.19	84.32

表 3　共起特徴の UT データセットの認識結果 (%)：局所見え重み × 局所動き.

Method	Dim.	Set1	Set2
BiWeight (2DLDA)	50× 20	**76.66**	**71.66**
BiWeight (2DLDA)	50× 5	71.66	66.66
BiWeight (2DPCA)	50× 50	38.33	56.66
BiWeight (2DPCA)	50× 5	33.33	56.66
Weight (2DLDA)	352×5	56.66	63.33

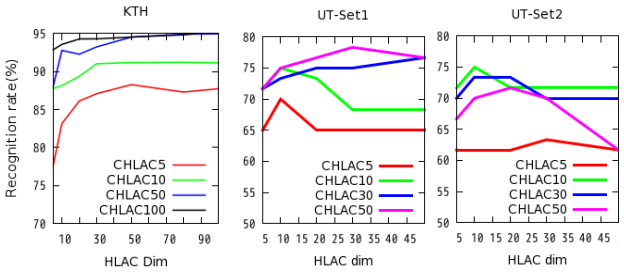

図 8　見え重みの数毎の認識結果.

表 4　異なる重み付け特徴の結合による KTH の認識結果 (局所動きは 50 個の重みで圧縮)．I-IV: 本文参照.

Methods	5	10	30	50	80	100
I	91.62	91.91	93.63	93.56	93.84	93.64
II	92.17	93.38	94.72	**95.18**	95.31	**95.22**
III	**93.22**	**93.52**	**94.77**	95.06	**95.35**	95.14
IV	92.56	93.31	94.69	94.85	95.31	95.10

4.5　他手法との比較

従来の論文のおいて離れたピクセルのコードブックの共起や遷移を特徴量化する方法として用いられている統合法を提案手法と同様に同一位置における見えと動きの

共起として利用して比較する. 比較対象となる手法は以下の通りである.

2DPCA. 共起行列の重み付けに判別的な重みである 2DLDA の代わりに 2 次元主成分分析を適用した場合.

MSF. Markov stationary feature [10] による共起表現. $\boldsymbol{X}$ の行を正規化したものを $\boldsymbol{P}$ とする. $A = \frac{1}{n+1}(I + \boldsymbol{P}^1 + \boldsymbol{P}^2 + \cdots \boldsymbol{P}^n)$ を列方向に足したベクトルが MSF の特徴ベクトルである. 従来研究に従い $n = 50$ と設定した. 提案手法と同一の識別手法により識別を行う. MSF は P が正方行列の場合しか計算できないため, 見えと見え, 動きと動きの統合のみ利用した.

実験結果は表 2, 3 の中に示している. 提案手法がこれら, 判別的な重みを利用していない方法よりも認識精度が高いことが確認できる.

また, KTH の最新の他手法との認識率の比較を行う. 判別的な重みを利用した局所的な動き特徴の内の最良の結果として, 95.38%がある (表 2). 単一の特徴を利用する共起の学習を行う手法 ([12] 93.8%) と比較しても提案法の認識性能が高く提案重み付け手法が比較的有効であることが分かる. 現在までに提案手法よりも認識率が高い結果 ([11] 等) も報告されているが, これらは局所特徴パターンの学習を必要としていることが多い. また, 局所パターンを学習しない特徴の一つとして LTP [3] を実装し提案法と同一な識別条件で識別を行ったところ位置分割を利用しない場合 89.41%, [3] と同様な位置分割による最良結果が 94.66%であった. また, CHLAC の改良 ([13] 93.85%) よりも認識率が高い. これより, これまでに報告されている研究の中で, 本論文で検討した CHLAC と HLAC の組合わせと判別的重み付け積分特徴が局所パターンを学習しない手法の中では最も認識性能が高いと考えられる.

5. ま と め

本研究では, 見えに基づいて判別的に動き特徴量を重み付き積分する手法に 2 次元判別分析 (または, フィッシャー重みマップ) を利用することにより自然な形で重み係数を学習した. 従来の共起表現手法に比べ提案手法の利点は. 判別的基準を用いている点, 多数の判別的な重みを獲得することが出来る点, AdaBoost や MKL と異なり重み学習が高速な点である. KTH データセットを利用した実験により, 個別特徴量の単純な統合方法や判別的な重み付けを用いていない他の特徴共起手法よりも認識率が高いことが確認された.

また, 共起表現手法の欠点として, 以下の点が確認された. まず, 共起表現の行列を作成する全手法に共通して, メモリ消費量が大きいこと, 特に局所見えに基づく動き特徴に関しては, 計算時間も CHLAC や HLAC に比べて遅くなることである. フレーム見えに基づく重み付けでは, 背景情報が有効な場合においては, その情報を反映することもできるが, そうでない場合, 認識に関係のない背景の見えの影響も受けやすくなると思われる. この問題に対応するために, 領域重みを利用することが考えられる. また, フレーム重みとしては, 局所見えの線形重みを利用しているが, 全体の姿勢は局所見えの加法ではなく, 組み合わせで表現することが適切であると考えられるため, 今後はフレーム見え重み付けの特徴量を変更することや非線形重みを利用することも検討したい.

文 献

[1] P,Dollar et.al, "Behavior recognition via sparse spatio-temporal features", in *VS-PETS*, 2005.

[2] T.Kobayashi and N.Otsu, "Three-way auto-correlation approach to motion recognition", *Pattern Recognition Letters*, vol.30, issue 3, pp.212–221, 2009.

[3] L.Yeffet, and L.Wolf, Local Trinary Patterns for Human Action Recognition, In ICCV, 2009.

[4] K.Schindler et.al, "Action snippets: how many frames does human action recognition require ?", *CVPR*2008.

[5] N.I.-Cinbis, S.Sclaroff, "Object, Scene and Actions: Combining Multiple Features for Human Action Recognition", in *ECCV*, 2010.

[6] X.Sun, M.Chen, and A.Hauptmann, "Action Recognition via Local Descriptors and Holistic Features", in *CVPR workshop*, 2009.

[7] K.Liu, et.al, "Algebraic feature extraction for image recognition based on an optimal discriminant criterion", *Pattern Recognition*, vol.26, no.6,1993.

[8] Y.Shinohara and N.Otsu, "Facial expression recognition using fisher weight maps", in *FG*, 2004.

[9] F.S.Khan. J.v.d.Wijer, M.Vanrell, "Top-Down Color Attenstion for Object Recognition", in *ICCV*, 2009.

[10] J.Sun, X.Wu, S.Yan, L.-F. Cheog, T.-S. Chua and J.Li, "Hierarchical Spatio-Temporal Contex Modeling for Action Recognition", in *CVPR*, 2009.

[11] A.Gilbert, J.Illingworth, R.Bowden, "Fast Realistic Multi-Action Recognition using Mined Dense Spatio-temporal Features", in *ICCV*, 2009.

[12] M.S.Ryoo and J.K.Aggarwal, "Spatio-temporal relationship match: video structure comparison for recognition of complex human activities", in *ICCV*, 2009.

[13] 松川徹, 栗田多喜夫"局所的な動き属性の相互相関特徴による行動認識 ",PRMU, 2010.3.

[14] H.Wang, M.M.Ullah, A.Klaser, I.Laptev, C.Schmid, "Evaluation of local spatio-temporal features for action recognition", in *BMVC*, 2009.

[15] 山内雄嗣, 藤吉弘宣, H. B.-Woo, 金出武雄, "アピアランスと時空間特徴の共起に基づく人検出", MIRU, 2007.

[16] S.Yan, et.al, "Multilinear discriminant analysis for face recognition", *IEEE trans. on Image Processing*, vol.16, no.21, pp.212–220, 2007.

[17] T.Harada, H.Nakayama, and Y.Kuniyoshi, "Improving Local Descriptors by Embedding Global and Local Spatial Information", in *ECCV*, 2010.

[18] 森下雄介, 小林匠, 森崎巧一, 大津展之, "時間重みと外的規準を用いた動作評価手法 ", PRMU, 2008.3.

[19] D.Tao, et.al, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition", *IEEE trans. on PAMI* , vol.29, no.10, pp.1700–1715, 2007.

[20] X.He et.al, "Locality preserving projections (LPP)", *Technical Report*, The University of Chicago, 2002.

[21] J.Yang, et.al, "Two-dimensional discriminant transform for face recognition", *Pattern Recognition*, vol.38, issue7, pp.1125–1129, 2005.

[22] C.Schuldt, I.Laptev, and B.Caputo, "Recognizing human actions: a local svm approach", in *ICPR*, 2004.