論　文

Latent SVMによる人体姿勢推定の効率的学習のためのサンプル選択

松山　洋一[†]　　浮田　宗伯[†a)]　　萩田　紀博[†]

Efficient Modeling by Selecting Samples for Human Pose Estimation by Latent SVM

Yoichi MATSUYAMA[†], Norimichi UKITA[†a)], and Norihiro HAGITA[†]

あらまし　学習サンプルを適切に選択することにより，サンプル数の少ない高速学習でも，大量サンプルにおける学習と同じ結果（異なる複数のモデルから，同じモデルが最高の認識結果を得る学習結果）を獲得することを目的とする．本論文では，人体姿勢推定を研究対象とする．提案法は，二つの選択法の組み合わせからなる．一つは一般的なクラスタリングによる選択で，重複的な学習を避けるために採用する．このクラスタリングは，人体姿勢のための特徴量の低次元化によって高速化させる．二つめは識別境界からの距離による選択で，姿勢モデルの学習に利用されるLatent SVMの特性に応じた識別境界の効率的な更新に着目した．Latent SVMにおける反復姿勢探索の効率化のため，画像中の人体姿勢探索における枝刈りも加える．実験の結果，提案手法により，複数のモデル中で最高の姿勢推定正答率を得るモデルが，全サンプル学習と提案手法間で同じになることを確認し，かつ，学習時間を79%削減できた．

キーワード　学習高速化，学習サンプル選択，人体姿勢推定，Pictorial structure models，Latent SVM

1. まえがき

機械学習によって物事を認識・判別する能力は向上の一途をたどっている．一般に統計的なパラメータに基づくモデル学習では，学習するデータが多いほどモデルの性能が向上する．しかし学習にかかる計算コストは，学習するサンプル数の増加に応じて急激に大きくなってしまう．例えば，画像に写っている物体の種類を認識する問題において，ImageNet [1] という120万枚の画像が1000個のクラスに分かれているデータセットがある．このデータセットの学習には，少なくとも250日もかかってしまう [2]．この実験が仮に一度のみの学習で良いのであれば，膨大な学習時間も許容される．しかし実際には，モデルの学習時に条件やパラメータなどを設定し，計算機実験を行い，そのフィードバックを受け条件やパラメータなどを再設定し，再度実験を行う．このように何度も実験を繰り返し，最も良い性能を得る条件やパラメータをもつモデルを選択するのが一般的である．よって，膨大な学習時間は研究活動において大きな障害となる．

認識モデルの学習において，学習時間と学習モデルによる正答率はトレードオフの関係にあり，両者のバランスが重要である．複数モデルから最高のモデルを選択する段階では，認識正答率の絶対的な高さは必要ない．このモデル選択で重要なのは，複数モデル間での正答率の大小である．すなわち，正答率最高のモデルが，全サンプル使用時と選択サンプル使用時で同じならば，選択サンプルによる学習後に，検証用データで正答率最高のモデルを選び，そのモデルを全サンプルで再学習すればよい．この際，正答率最高のモデルは2回学習されるが，図1に示すように，選択サンプルによる学習が高速なほど，またテストするモデルが多いほど，全行程に要する学習時間は全サンプル学習時と比べて短くなる．

上述した，モデル選択から最終モデル学習に至る処理を実現するためには，少数サンプルのみを学習したときでも，全サンプルを学習したときと同じモデルが最高正答率を得る必要がある．このモデル選択を保証するため，本論文で研究対象とする人体姿勢推定で

[†] 奈良先端科学技術大学院大学，生駒市
Nara Institute of Science and Technology, Ikoma-shi, 630–0192 Japan
a) E-mail: ukita@is.naist.jp
DOI:10.14923/transinfj.2014JDP7104

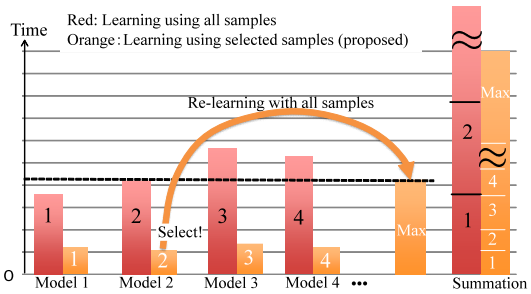

図 1　提案法による学習時間削減．赤棒が全サンプルを学習して最高のモデルを得る計算時間．橙棒が，提案法により選択サンプルのみを学習して最高モデルを選び，最高モデルを全サンプルで学習する計算時間．

Fig. 1　Effects of reducing computational cost by the proposed method. Red and orange bars show computational costs of a general approach (i.e. learning all samples) and the proposed method, respectively.

利用するモデルや特徴量の特性を考慮したサンプル選択法を提案し，複数モデル中から全サンプル学習時と変わらぬ最高正答率モデルが得られることを実験的に確認した．

2. 関 連 研 究

本論文では，研究対象として静止画の人体姿勢推定問題を選んだ．この問題では，図 2 に示すように，静止画中の人体パーツの位置・姿勢を推定する．人体姿勢からは，人間の行動認識が可能になる [3]．人の行動認識は多様な応用（セキュリティなど）の基礎技術となることが，人体姿勢推定を問題として選んだ理由である．もう一つの理由は，人体姿勢推定の複雑さにある．人体モデルは，最適化の対象である頭や手足などのパーツが多く，効率的な学習が不可欠である．

人体姿勢推定では，各パーツのアピアランスを表す特徴量と，Pictorial structure models（PSM）[4] などによる幾何学的制約と，識別的学習 [5] の組み合わせによって，高い推定性能が得られている．

アピアランス特徴量には，HOG [6] を元にした特徴量が広く用いられている．HOG を人体パーツ用に特化させるため，Yang らはパーツごとに HOG の勾配方向に重みを付けた [7]．Wang らは，遮蔽に対し頑健になるように HOG と LBP [8] を組み合わせた [9]．

人体パーツ間の位置関係を表した制約モデルでは，制約の構造や種類の検討（例：パーツ間の連結関係や階層構造 [10]，連結の見えの特徴量化 [11]）が研究されている．また，パーツの見えのクラスタ化 [7], [12],

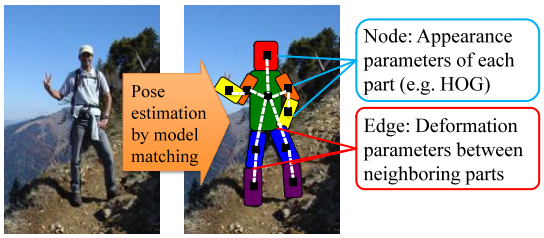

図 2　PSM による姿勢推定．画像中に重畳表示された各矩形と点線が姿勢推定結果であり，それぞれパーツ（ノード）とパーツ間の連結関係（エッジ）を表す．

Fig. 2　Pose estimation by PSM. Each node and edge depicts a part and connection between parts, respectively.

姿勢のクラスタ化 [13], [14] のように，パラメータの適切なクラスタ化によるモデル効率化も提案されている．

以上のような研究では，パラメータ（例：HOG のヒストグラム構造やパーツ・姿勢のクラスタ数）を変更しつつ実験を繰り返して最良モデルを選択するため，学習の高速化が重要になる．

学習高速化の為に少数サンプルで学習する方法として，Boosting を用いる手法がある [15], [16]．この手法ではアピアランス学習のみ少数サンプルで行われおり，幾何制約モデルの学習は効率化されてない．また，明示的にどのような種類のサンプルを選択するかという基準については検討されていない．これに対し，本提案手法はアピアランス特徴量だけでなく，幾何制約に関するパラメータも含めた全特徴量を参照したサンプル選択により，学習の高速化を実現する．

3. 姿勢推定のための Pictorial Structure Models

PSM による姿勢推定の概要を図 2 に示す．木構造 $G = (V, E)$ では，n 種のパーツ（手足，頭など）がノード $V = \{v_1, \cdots, v_n\}$ に，連結する 2 パーツ v_i, v_j の相対的な位置関係（距離，角度など）がエッジ $e_{i,j} \in E$ に対応する．画像内における各パーツの位置 $L = \{l_1, \cdots, l_n\}$，すなわち人体姿勢は，以下のコストを最小とする位置 L^* の探索により得られる．

$$L^* = \arg\min_L \left(\sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i, v_j) \in E} d_{ij}(l_i, l_j) \right) \quad (1)$$

$m_i(l_i)$ は位置 l_i におけるアピアランス特徴とモデルがもつパーツ i のアピアランスモデルとの相違度であり，$d_{ij}(l_i, l_j)$ は連結するパーツ i, j がそれぞれ l_i，l_j に位置したときの相対的な位置関係とモデルがもつ

i, j の相対的な位置関係との相違度となっている．

4. Latent SVM におけるモデル学習と学習時間に関わる問題点

人体姿勢の識別的モデル学習では「人体パーツ位置がアノテーションされたポジティブ画像」と「人が写っていないネガティブ画像」が与えられる．これらの学習画像から，式 (1) のパラメータが学習される．以降，ポジティブ画像内における人体姿勢領域をポジティブサンプルと呼び，ネガティブ画像内の任意の領域（人が写っていない領域）をネガティブサンプルと呼ぶ．

ポジティブサンプルからは，人体パーツの見え，及びパーツ間の位置関係が学習される．ネガティブサンプルの学習は，人体とパーツの見えやパーツ配置が類似している人体以外の物体（背景領域など）と人体との識別性を向上させる．

ネガティブサンプルの学習では，PSM を用いてネガティブ画像内で人体姿勢を探索する．ネガティブ画像内には人は写っていないため，探索により検出された領域は誤検出である．この誤検出領域をネガティブサンプルとして学習する．本論文では，学習のために Latent SVM [5] を使用した．Latent SVM によるモデル学習では，以下のように反復学習が必要になる．

Step1　ポジティブ画像と少数のネガティブ画像から，PSM のパラメータを Latent SVM で学習する．

Step2　新たなネガティブ画像の全領域を対象に PSM で人体姿勢推定し，それらの Latent SVM のスコア（推定姿勢の「人の姿勢らしさ」）を得る．

Step3　スコアがしきい値以上の領域を誤検出とし，ネガティブサンプルとして再学習する．

Step4　誤検出が消えるまで Step2，3 を繰り返す．

Step5　誤検出が消えると，新たな画像を得て Step2 へ．

Step6　全ネガティブ画像に対し Step2〜5 を繰り返す．

Latent SVM では，ネガティブ画像での探索と学習が画像数だけ反復される．しかも学習時には，ネガティブサンプルを一つずつ反復学習する．これらの反復が Latent SVM の学習時間の大半を占める．

この反復によって，Latent SVM の学習速度は遅くなってしまう．しかし，Latent SVM は，人体姿勢推定モデル学習において，姿勢アノテーション，すなわち，学習画像中の人体の各関節の x, y 座標を Latent 変数とすることによって，関節の x, y 座標が最適な位置に調整されるようにモデル学習できる．よって，姿勢アノテーション誤差に対する頑健なモデル化を実現できる．姿勢アノテーションの誤差が，姿勢推定に与える影響は小さくなく [14]，Latent SVM を人体姿勢推定に利用する価値は大きい．したがって，本研究では，推定性能面で Latent SVM の利点を活用しつつ，その学習速度の遅さを提案手法によって補う．

5. サンプル選択による学習モデルの変化

3.，**4.** で紹介した人体姿勢推定法における，適切なサンプル選択の必要性を示す．

異なるモデル A，B，C，D，E を用意し，それぞれ同数の学習サンプルで学習し，人体姿勢推定を行い，その正答率を求めた．モデル A には，文献 [7] の mixture part model を用いた．モデル B は，各パーツのアピアランス特徴の次元数をモデル A の 32 から 27(注1) に減らしたものである．モデル C では，各パーツの mixture 数をモデル A の 5 から 3 に減らした．モデル D では，アピアランス特徴の次元数を 27 に，mixture 数を 3 にした．モデル B，C，D が，パラメータ数を減らしているのに対し，モデル E では，各パーツのアピアランス特徴の次元数をモデル A の 32 から 41(注2) に増やした．

本論文で示す人体姿勢推定の正答率は，人体の全 10 パーツ（頭，胴，左右上腕，左右下腕，左右上脚，左右下脚）のパラメータ $l_1, \cdots, l_{10}$ が，それぞれ正しく推定できている割合 $\left(\frac{\sum_n^{N_I}\sum_i^{10} S(n,i)}{N_I \times 10}\right) \times 100$ である．ただし，$S(n,i)$ は n 枚目のテスト画像で l_i をしきい値以内の誤差(注3) で推定できたときに 1，それ以外のときに 0 を返す関数で，N_I はテスト画像の総数である．

使用したデータセットは，Leeds Sports Pose（LSP）データセット [13] と INRIA Person データセット [17] で，LSP データセットのうち 1000 枚を学習用ポジティブ画像，1000 枚をテスト画像として，INRIA データセットは 1218 枚をネガティブ画像として使用した．ポジティブサンプルは，全ての試行で等しい．ネガティブサンプルは，全て学習する場合（表 1 の「全ネガティブサンプル」）と，全ネガティブサンプル中の約 3%をランダムに選択し学習する場合（表 1 の「ランダム選択」）の 2 通りである．

全正答率は，10 回試行の平均である．表 1 に示す

(注1)：手法 [7] の特徴量から，テクスチャ特徴を除いた．
(注2)：手法 [7] の特徴量において，HOG 特徴の方位ビン数を 18 から 27 に増やした．
(注3)：評価関数 $S(n,i)$ には，文献 [18] で公開されているコードを利用した．

表 1　複数モデル間（モデル A，B，C，D，E）における最高正答率の逆転
Table 1　Change in estimation accuracy among different models.

正答率 (%)	A	B	C	D	E
全サンプル	**62.7**	62.2	61.9	59.3	62.5
ランダム選択	53.4	**53.7**	53.6	52.2	53.0

ように，全サンプルを用いた場合と少数のランダム選択サンプルを用いた場合では，正答率が逆転した．このように，ネガティブサンプルをランダムに減らしただけでは，複数モデル中において最高正答率を得るモデルが保存できない．

6. ネガティブサンプルの選択法と実験

6.1 提案法の概要

Latent SVM による人体姿勢モデルの学習では，ネガティブサンプルの反復学習が学習時間の大半を占めている．この反復回数を減らすため，学習するネガティブサンプルを適切に選択する．提案手法では，人体姿勢特徴の低次元化によるクラスタリングを用いた選択の高速化（**6.3**），識別境界からの距離に基づく効率選択と人体姿勢探索の枝刈りによる効率化（**6.4**）を組み合わせ，学習時間を全体的に削減する（**6.5**）．**5.** と同様に，実験には，LSP データセット [13] と INRIA Person データセット [17] を用いた．

6.2 選択するネガティブサンプル数

以降の実験で用いるネガティブサンプル数を決定するため，予備実験を行う．図 3 に，全ネガティブサンプルを学習した際に誤検出する人領域の数を示す．横軸には，学習順に左から右に画像がならんでいる．図 3（左）のグラフが，LSP と INRIA Person を学習した結果である．このグラフから，学習の初期と中盤以降では誤検出数が大きく異なることがわかる．これは，学習初期ではモデルが未成熟であり，誤検出が多く発生してしまうためである．異なるデータセットでも同様の傾向が得られることを確認するため，ポジティブ画像に PARSE データセット [19] 中の 100 枚の学習用画像，ネガティブ画像に SUN Attribute データセット [20] から背景のみの 2315 枚の画像を抽出したセット（以降，SUN データセットと呼ぶ）を用いて同様の実験を行った（図 3（右））．

上記傾向に従い，提案法でも，各ネガティブ画像から得るネガティブサンプル数は均等ではなく，学習初期段階で多数に，学習が進むにしたがい少なくする．

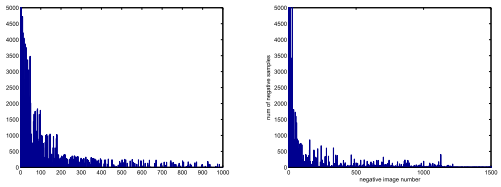

図 3　各ネガティブ画像における人体誤検出数．（左）LSP と INRIA Person データセットにおける実験結果．（右）PARSE と SUN Attribute データセットにおける実験結果．横軸が画像番号で，縦軸が各画像における誤検出/選択サンプルの数を表す．
Fig. 3　(Left) False-positives extracted at each negative image. (Right) The number of selected negative samples, which is defined by Eq. (2).

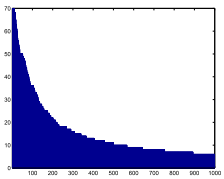

図 4　選択サンプル数 N_{NS} の例．この例は，図 2（左）のネガティブサンプル数から決定された結果である．ただし，学習進行に伴うサンプル数の変化が視覚的にわかりやすくするため，選択サンプル数が単調減少するようにスムージングさせている．
Fig. 4　Example of the number of selected samples. This example is obtained from the number of all negative samples shown in Fig. 3.

モデルの成熟度は，誤検出数に強く影響するため，誤検出数に応じたネガティブサンプル数の決定は妥当であると考える．しかし，全ネガティブサンプルを学習する結果として，学習終盤で急激にネガティブサンプルが少なくなる通常の全ネガティブサンプル学習と比べ，学習初期から限られた選択サンプルのみを学習する提案法では，学習終盤に至っても十分に学習が進んでいない恐れがある．そこで，図 3 のような傾向に類似しつつ，その傾向と比べて学習終盤にも学習サンプル数を残すことによって学習不足を避けるように，選択サンプル数 N_{NS} を決定する．

$$N_{NS} = \frac{N_{FP}}{\sqrt{N_{FP}} + 1} \tag{2}$$

N_{FP} は誤検出数である．図 3（左）の全ネガティブサンプル数を N_{FP} とした際の，式 (2) から得られる選択サンプル数 N_{NS} の傾向を図 4 に示す．この図から，目的のとおり，学習初期に急激に N_{NS} が減少し，学習終盤では減少傾向が低下している様子が確認できる．以降の実験では断りがない限り，選択するネガティブサンプル数は式 (2) に従う．

表 2 （手法 1）クラスタリングによるネガティブサンプル選択による姿勢推定モデル学習に要する時間と，そのモデルを用いた人体姿勢推定正答率．括弧内は，全サンプル学習時間に対するパーセント表記であり，値が小さいほど高速化できたことを表す．
Table 2 (Method 1) Pose estimation results using negative samples selected by clustering. A value in parenthesis indicates the percentage of the cost for learning all samples.

	正答率 %	学習時間 秒	探索時間 秒	クラスタリング時間 秒	総時間 秒
全サンプル	62.71	5911	2362	-	8273
ランダム選択	55.32	714 (12)	1674 (71)	-	2388 (29)
手法 1	57.76	653 (11)	1676 (71)	3791	6120 (74)
手法 1（特徴低次元化あり）	57.43	649 (11)	1668 (71)	669	2986 (36)

6.3 選択法 1：クラスタリングによる選択

図 5 に示すように，青や橙のネガティブサンプルを全て学習しても，識別境界に与える影響が類似していると考えられる．そこで，これらサンプルをクラスタリングし，各クラスタの代表となるネガティブサンプルのみを選択し学習する．クラスタリングする特徴量は，パーツの見え（HOG）特徴とパーツ間の接続関係を表すパラメータの連結ベクトルである．基本的な学習アルゴリズムは **4.** の Step1∼6 と同じであるが，Step3 では，全誤検出の特徴量をクラスタリングし，その代表ベクトルのみを Latent SVM で再学習する．代表サンプルは，クラスタ中心に最も距離が近いサンプルとする．クラスタ数は式 (2) で定義されるサンプル数 N_{NS} である．クラスタリングは K-means により行った．

クラスタリングに基づくサンプル選択は，古くから行われている [21], [22]．クラスタリングにかかる時間は，サンプルの次元数と個数に依存する．本実験で用いた人体姿勢推定法 [7] において，特徴量の次元数は 13489 次元である（パーツ数 26，各パーツの HOG が 512 次元，親パーツとの相対的位置を表すパラメータ 4 次元，その他バイアスなどのパラメータ 77 次元：$(512 \times 26) + (4 \times (26 - 1)) + 77 = 13489$）．図 3 で示したように，クラスタリングする誤検出の数は学習初期においては数千である．クラスタリングを用いる為には，これらを上手く削減する必要がある．クラスタリングするネガティブサンプル数の削減は次節の処理に譲り，本節では，特徴量の次元数削減によりクラスタリングの高速化を実現する．

具体的には，HOG を大幅に簡略化してクラスタリングを高速化する．ベース手法 [7] では，各パーツは 4×4 のセルから構成され，各セルが 32 次元のアピアランス特徴量をもつ．これを，セル分割を廃止して，アピアランス特徴量を 16 次元に減らすことにより，

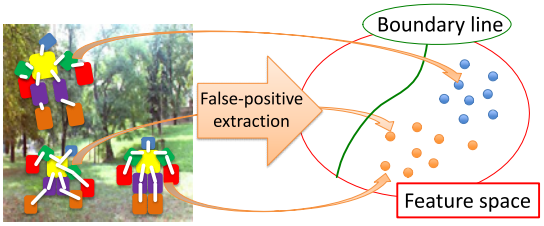

図 5 クラスタリングによるネガティブサンプル選択．特徴空間でクラスタ中心に最も近いネガティブサンプルを代表サンプルとして選択し学習する．
Fig. 5 Negative sample selection using clusterng. In each cluster, the sample that is the closest to its center is slelected for learning.

各パーツの HOG を 16 次元まで低次元化した．この HOG 特徴に親パーツとの位置関係パラメータをを加え，計 $(16 \times 26) + (4 \times 25) = 516$ 次元の特徴量ベクトルをクラスタリングした．

LSP+INRIA データセットで学習した人体姿勢推定の正答率及び学習にかかった時間を計測した（表 2 は 10 回試行の平均）．表中の学習時間，探索時間，クラスタリング時間は，それぞれ，Latent SVM による反復モデル学習（**4.** の Step1 と Step3）の総時間，Latent SVM において誤検出を反復的に得る総時間（**4.** の Step2），提案手法において誤検出を反復的にクラスタリングする総時間である．全て反復処理の総時間であるため，各ネガティブ画像で検出された誤検出数や，そこから選ばれてモデル学習に利用されるサンプル数によって反復数が変化するため，全サンプル学習や学習手法によって各処理にかかる時間が異なる．比較対象として全誤検出を学習する従来法と，クラスタと同数の誤検出をランダムに選択し学習する手法を用意した．表 2 の探索時間は，画像中で人体領域を探索する時間である．学習時間とは，探索により誤検出を得た後，Latent SVM にネガティブサンプルを与え，学習する時間である．総時間は，これらとクラス

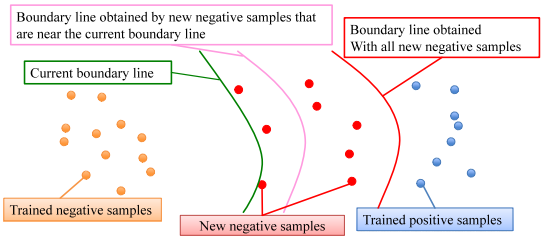

図 6　識別境界からの距離が遠いネガティブサンプルの選択．距離の遠いサンプルを選択し学習することで，ピンクの識別境界をスキップして直接赤の識別境界を学習する．

Fig. 6　Selecting negative samples that are far from a decision boundary for efficient learning.

タリングにかかった時間の和である．

クラスタリングを行う手法は，学習時間を，全サンプルを学習した場合の 11%に削減できた．しかし，HOG 特徴を低次元化しないクラスタリングでは，クラスタリングに 3791 秒（全サンプル学習の総時間の 46%）もかかってしまったため，学習総時間では全サンプルを学習した場合の 74%に留まり，ランダム選択と比べ 2.56 倍の時間がかかってしまった．一方，HOG を低次元化すると，クラスタリング時間を減らし，学習総時間を全サンプル学習の 36%に削減できた．

6.4　選択法 2：識別境界からの距離による選択

手法 2 では，識別境界からの距離が遠いサンプルを選択する．これにより，識別境界を大きく更新し，特にモデルが未成熟な段階における学習の効率化を期待する．図 6 の典型的な例では，赤色のネガティブサンプルを学習して緑色の識別境界を更新させる．識別境界近くの 3 点のネガティブサンプルを学習した場合，新たにピンクの境界が得られる．しかしその後，全ネガティブサンプルを学習すると，境界は赤色のように変化する．すると境界に近いネガティブサンプルを学習した意味はない．

このように，識別境界から離れたサンプルの学習は，一般的な SVM のバッチ学習効率化 [23] において，識別境界付近のサンプルを絞り込みサポートベクトルの決定を効率化させるのとは真逆のアプローチである．この違いは，Latent SVM では，ネガティブサンプルを一つずつ追加で学習しなければならず，しかも高速オンライン学習（手法 [24] など）で行われているように，学習済みのモデルを新サンプルで効率的に更新することができない点に起因する．すなわち，一般的な SVM のバッチ学習では全サンプルが与えられた後に，

表 3　（手法 2）識別境界からの距離による選択での結果

Table 3　(Method 2) Results of negative sample selection based on a distance from a decision boundary.

	正答率 %	学習時間 秒	探索時間 秒	総時間 秒
全サンプル	62.71	5911	2362	8273
ランダム選択	55.32	714 (12)	1674 (71)	2388 (29)
手法 2	58.83	772 (13)	1683 (71)	2455 (30)
手法 2（枝刈りあり）	58.19	798 (14)	1023 (43)	1821 (22)

全サンプルの分布から識別境界付近の点を推定して，その付近のサンプルのみを学習するのが効率的であるが，Latent SVM ではそれが不可能なためである．

手法 2 の基本的な学習法は，**4.** の Step1〜6 と同じであるが，Step3 では，Latent SVM のスコア（人の姿勢らしさ）をソートして，スコアが大きなものから式 (2) の N_{NS} だけネガティブサンプルを選択する．

識別境界からの距離には，前述の全 13489 次元の特徴量における境界から各サンプルまでの距離を用いる．この距離は，Latent SVM で学習された PSM による人体姿勢検出時に直接得られる．このように，人体姿勢推定の過程の結果をそのまま利用することにより，無駄な計算時間を省略して，高速な処理を実現する．

実験結果を表 3（10 回試行の平均）に示す．比較対象とデータセットは **6.3** と同様である．学習時間については，全サンプル学習時よりも速く，表 2 に示した手法 1 と比べても，クラスタリングにかかる時間が省略されるぶんだけ高速化されている．

学習時間と探索時間を比較すると，探索時間のほうが高速化の割合が劣る．これは，図 3（左）で示したように学習するネガティブサンプル数が学習進行に伴い激減する一方，PSM により各画像から人体姿勢探索をした際，探索時間は画像に大きく依存せずにほぼ一定であることに起因する．よって，学習総時間の更なる削減には，探索時間の削減が重要となる．

この探索時間の削減のため，PSM による人体姿勢探索の枝刈りを加える．PSM では，全ノード・リンクのコストを加算し(注4)，その総和を人体姿勢のコストとするため，加算途中でコストが大きくなった場合は枝刈り，すなわち人体姿勢評価を打ち切りできる．

枝刈りのしきい値には，「一つ前の反復で選ばれたネガティブサンプル中で，式 (1) のコストが最小の値

(注4)：文献 [4] で提案されているように，動的計画法によって画像全体における人体姿勢探索を高速に実現している．

表 4 （提案法）クラスタリングと識別境界からの距離によるネガティブサンプル選択での人体姿勢推定の結果
Table 4 (Proposed method) Results of negative sample selection based on a distance from a decision boundary and sample clustering.

	正答率 %	学習時間 秒	探索時間 秒	クラスタリング時間 秒	総時間 秒
全サンプル	62.71	5911	2362	-	8273
ランダム選択	55.32	714 (12)	1674 (71)	-	2388 (29)
手法 1（特徴低次元化）	57.33	649 (11)	1668 (71)	669	6120 (36)
手法 2（枝刈り）	58.19	798 (14)	1023 (43)	-	1821 (22)
提案手法	59.82	623 (11)	1068 (42)	46	1737 (21)

（L_{min}）」を用いる．ただし，ネガティブサンプルの反復学習によって，誤検出される人体姿勢のコストが徐々に小さくなることを期待し，枝刈りのしきい値は $0.9L_{min}$ とした．

この枝刈り込みの選択法 2 による実験結果も，表 3 に示す．この表に示すように，枝刈りなしの場合と比べて探索時間を削減できている（1683 → 1023 秒）．

6.5 提案手法：クラスタリングと識別境界からの距離によるサンプル選択の統合による効率学習

6.3 のクラスタリングに基づく手法では，大量の誤検出のクラスタリングに時間がかかりすぎてしまう．**6.4** の手法では，特徴量空間において近傍に位置する誤検出がネガティブサンプルとして選ばれてしまう．これらの問題を解決するため，誤検出を識別境界からの距離によって絞りこんだ後にクラスタリングを行う．

最終的に選ばれるネガティブサンプル数（クラスタ数）は式 (2) の N_{NS} である．クラスタリングの対象となるネガティブサンプル数 N_{SC} は，N_{NS} の特性を踏襲しつつ，これより大きい値を取るように，式 (3) のようにした．

$$N_{SC} = \frac{N_{FP}}{\sqrt[4]{N_{FP}} + 1} \tag{3}$$

結果を表 4（10 回試行の平均）に示す．比較対象として選択法 1（特徴量低次元化を行ったクラスタリング：表 4 中の「手法 1（特徴低次元化あり）」）と，選択法 2（識別境界からの距離に応じた選択法：表 4 中の「手法 2（枝刈りあり）」）の結果も再掲する．全サンプル学習と比べて，学習時間（5911 → 623）と探索時間（2362 → 1068）を削減でき，余分に必要となるクラスタリング時間（46）を短くできたため，総計算時間の割合は 21%（8273 → 1737）となった．

最後に，実運用にかかる学習時間を比較する．全サンプルを学習する従来法では，最高モデルを得るための時間は，全モデルの学習総時間の和 $\sum_i^{N_M} T_i$ となる．N_M は比較するモデルの総数であり，T_i は i 番目のモデルの学習時間である．提案法の場合，最高モデルを得るための時間は，提案法による全モデルの高速学習と，そこから得られる最高モデルを全サンプルで学習する時間の和 $\sum_i^{N_M} T'_i + T_b$ になる．T'_i は i 番目のモデルを提案法で学習する時間，b は提案法による学習で最高モデルとして選ばれるモデルの ID である．

比較するモデル数を次節の実験に倣い $N_M = 5$ とし，T_i 及び T'_i の平均を表 4 に示した全サンプル学習と提案法による学習時間とする．従来法と提案法の学習時間は，それぞれ $8273 = 41365$，$1737 \times 5 + 8273 = 16958$ となる．この例のように，比較モデル数が少ない，すなわち提案法による学習時間削減の効果が小さくても，提案法により学習時間が半分以下となった．

6.6 提案法による異なるモデル間の正答率の検証

5. で示したような異なるモデル間における最高正答率の逆転現象を，提案法によって回避できることを実験的に検証する．本論文の目的に重要なことは，提案手法によって，全サンプル学習した際と等しいモデルが最高正答率を得ることである．本節では，最高正答率を得るモデルが頑健に選択されることを検証するため，学習と姿勢推定（正答率の計算）を 10 回試行する．それら 10 回の正答率の平均を比較するとともに，正答率最高のモデルとその他モデルの間における正答率平均の間の有意差，すなわち，どれほど頑健に正答率最高のモデルが選ばれるかという点を検証する．

提案手法の有効性を安定に検証するため，ポジティブとネガティブともに，画像データセットを一つ追加する．ポジティブが 2 セット（LSP と PARSE），ネガティブが 2 セット（INRIA と SUN）で，組み合わせとして $2 \times 2 = 4$ 種類の実験を行った．結果を，図 7，8，9，10 に示す．実験には，**5.** と同様の四つのモデル（モデル A，B，C，D，E）を利用した．

いずれの組み合わせにおいても全サンプル学習（図

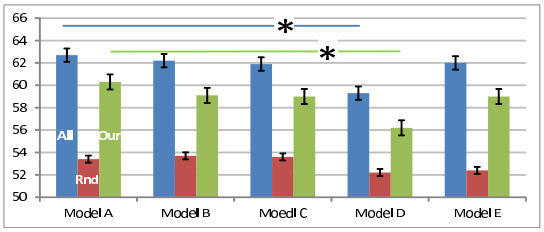

図 7　異なるモデルにおける正答率（LSP+INRIA）
Fig. 7　Accuracy in different models (LSP+INRIA).

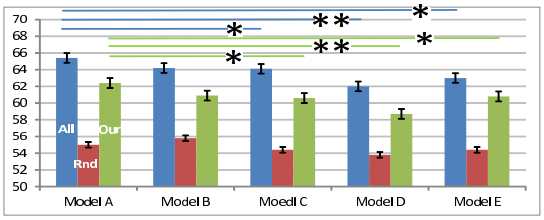

図 8　異なるモデルにおける正答率（LSP+SUN）
Fig. 8　Accuracy in different models (LSP+SUN).

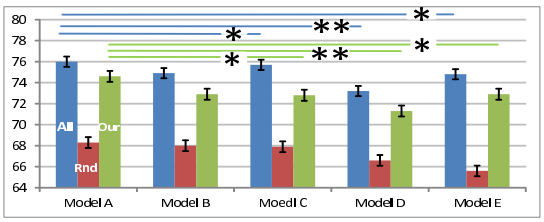

図 9　異なるモデルにおける正答率（PARSE+INRIA）
Fig. 9　Accuracy in different models (PARSE+INRIA).

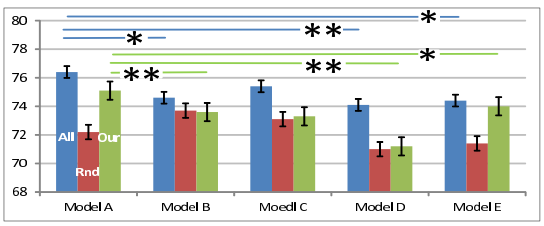

図 10　異なるモデルにおける正答率（PARSE+SUN）
Fig. 10　Accuracy in different models (PARSE+SUN).

中の青棒）では，モデル A が最高正答率を得た．同様に，提案手法（図中の緑棒）でもモデル A が最高正答率を得た．一方，ランダム選択（図中の赤棒）では，LSP+INRIA，LSP+SUN，PARSE+SUN において，モデル B が最高正答率を得てしまっている．

welch の t 検定（片側検定）により，最高正答率モデルとその他モデル間の平均正答率の有意差の大小を調べた結果を図中に示す．$p < 0.05$，$p < 0.01$ の結果を，それぞれ $*$，$**$ で示す．PARSE+SUN（図 10）のモデル A とモデル B の検定結果を除いて(注5)，全サンプル学習と提案手法では同様の結果が得られた．この結果から，提案手法により，全サンプル学習と等しいモデルが最高正答率が得られること，また，全サンプル学習と提案手法で，複数モデル間の平均正答率の大小関係（有意差）も類似していることが，実験的に検証できた．

提案法は，自動的に最高正答率を得るモデルを発見する．このように最高正答率を得るモデルは，モデル間の違いが自明であれば，人手によってもある程度予測可能である．例えば，モデル B，C，D はパラメータ数が減っているため，モデルの表現力が低下し，単純には姿勢推定の正答率が減少することが予想できた．一方，モデル E はパラメータ数が増えているため，正答率が向上することもありえた．しかし実際には，正答率は低下した．これは，モデルが複雑になりすぎ，過学習のようなことが起きていることが一因と考えられる．このように，人手による単純な予測とは異なるようなモデルの特性も，提案手法により確認できていることは，提案法の有効性を示している．

7. むすび

学習サンプルを適切に選択することで学習を高速化する手法を提案した．研究対象として静止画の人体姿勢推定を選択し，学習時間の大半を占めているネガティブサンプルの学習時に，提案法を用いることで高速化を実現し，その効果を実験的に示した．

学習画像総数 2000 程度の問題において，学習総時間を 21%にまで高速化できた．その際，全ての画像データセットにおいて，選択サンプルのみによる高速学習において正答率最高となるモデルが，全サンプル学習における正答率最高のモデルと等しいことも実験的に確認した．

更なる高速化のための今後の課題を挙げる．

- 異なるモデル間での正答率の相対的な大小関係が変化しない，最小サンプル数の検討（式 (2)，(3) の最適化.）．
- 人体姿勢の特性をより考慮したサンプル選択．
- 探索時間の高速化による総学習時間の短縮．

(注5)：全サンプル学習における p 値も 0.013 であり，$p < 0.01$ の強い有意差ありに近い．

文　献

[1] J. Deng, A.C. Berg, K. Li, and F.-F. Li, “What

does classifying more than 10,000 image categories tell us?," ECCV, pp.71–84, 2010.
[2] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, and L. Cao, "Large-scale image classification: Fast feature extraction and svm training," pp.1689–1696, CVPR, 2011.
[3] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M.J. Black, "Towards understanding action recognition," pp.3192–3199, ICCV, 2013.
[4] P.F. Felzenszwalb and D.P. Huttenlocher, "Pictorial structures for object recognition," Int. J. Comput. Vis., vol.61, no.1, pp.55–79, 2005.
[5] P.F. Felzenszwalb, R.B. Girshick, D.A. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.9, pp.1627–1645, 2010.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," CVPR, pp.886–893, 2005.
[7] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," pp.1385–1392, CVPR, 2011.
[8] L. Wang and D.C. He, "Texture classification using texture spectrum," Pattern Recognit., vol.23, no.8, pp.905–910, 1990.
[9] X. Wang, T.X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," pp.32–39, ICCV, 2009.
[10] Y. Wang, D. Tran, and Z. Liao, "Learning hierarchical poselets for human parsing," pp.1705–1712, CVPR, 2011.
[11] N. Ukita, "Articulated pose estimation with parts connectivity using discriminative local oriented contours," pp.3154–3161, CVPR, 2012.
[12] P. Ott and M. Everingham, "Shared parts for deformable part-based models," pp.1513–1520, CVPR, 2011.
[13] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," pp.1–11, BMVC, 2010.
[14] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," pp.1465–1472, CVPR, 2011.
[15] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," CVPR, pp.1014–1021, 2009.
[16] V.K. Singh, R. Nevatia, and C. Huang, "Efficient inference with multiple heterogeneous part detectors for human pose estimation," pp.314–327, ECCV, 2010.
[17] N. Ikizler and P. Duygulu, "Histogram of oriented rectangles: A new pose descriptor for human action recognition," Image Vis. Comput., vol.27, no.10, pp.1515–1526, 2009.
[18] V. Ferrari, M.J. Marín-Jiménez, and A. Zisserman, "Progressive search space reduction for human pose estimation," CVPR, 2008.
[19] D. Ramanan, "Learning to parse images of articulated bodies," pp.1129–1136, NIPS, 2006.
[20] G. Patterson and J. Hays, "Sun attribute database: Discovering, annotating, and recognizing scene attributes," pp.2751–2758, CVPR, 2012.
[21] R. Koggalage and S. Halgamuge, "Reducing the number of training samples for fast support vector machine classification," Neural Information Processing - Letters and Reviews, vol.2, no.3, pp.57–65, 2004.
[22] S. Sohn and C.-H. Dagli, "Advantages of using fuzzy class memberships in self-organizing map and support vector machines," IJCNN, 2001.
[23] Q. He, Z. Xie, Q. Hu, and C. Wu, "Neighborhood based sample and feature selection for svm classification learning," Neurocomputing, vol.74, no.10, pp.1585–1594, 2011.
[24] S.C.H. Hoi, J. Wang, and P. Zhao, "Exact soft confidence-weighted learning," ICML, 2012.

（平成 26 年 7 月 20 日受付，12 月 23 日再受付，
27 年 4 月 1 日早期公開）

松山　洋一

2014 年奈良先端科学技術大学院大学修士課程修了．在学中，人体姿勢推定に関する研究に従事．

浮田　宗伯　（正員：シニア会員）

2001 年，京都大学大学院博士後期課程修了．同年奈良先端科学技術大学院大学情報科学研究科助手．2007 年同准教授．2002 年～2006 年まで科学技術振興機構さきがけ（「情報基盤と利用環境」領域）研究員兼任．2007 年～2009 年までカーネギーメロン大学客員研究員兼任．2011 年より ATR 客員研究員を兼任．博士（情報学）．コンピュータビジョン，分散協調視覚，対象追跡，人体運動解析・姿勢推定に関する研究に従事．

萩田　紀博　（正員：フェロー）

1978 年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了．同年電電公社（現 NTT）武蔵野電気通信研究所入所．NTT 基礎研究所，ATR メディア情報科学研究所長などを経て，現在 ATR 知能ロボティクス研究所長．この間，文字認識，画像認識，コミュニケーション科学，コミュニケーションロボットなどの研究に従事．工学博士．IEEE，電子情報通信学会，情報処理学会，日本ロボット学会，人工知能学会各会員．