

# サポートベクタ回帰を用いた骨格検出器

津田 圭一†

三菱電機株式会社 先端技術総合研究所†

## 1. はじめに

機械学習技術の進歩により、画像内から人物の関節等の座標を検出する、骨格検出技術が発展している。しかし、骨格検出器の多くは深層学習技術を用いたものである [1]ため、計算機リソース・電力を多く消費する欠点がある。そこで筆者は、深層学習に代え、省リソースなサポートベクタ回帰(Support Vector Regression, 以下SVR)を用いる骨格検出器を考案した。

## 2. 骨格検出器の構成

### 2.1. 全体構成

本検出器は、処理量の観点から現時点では上半身のみを対象を絞っており、その全体構成を図1に示す。検出器全体は8個の部位の検出器から構成されている。個々の部位の検出器は、その部位の領域および、後続部位の予測情報(予測される座標とバウンディングボックスのサイズ)を出力する。後続する部位の数は1個とは限らず、骨格の接続状況に応じ、0個~3個となっている。

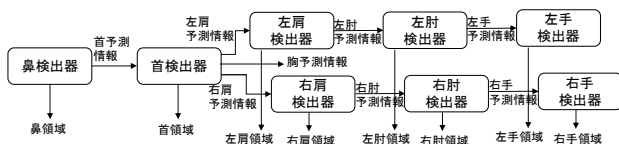


図1 上半身骨格検出器全体構成

個々の部位の検出器は、基本的に部位によらず同様の構成となっており、例えば「左肩」検出器は図2のような構成となっている。「左肩」検出器全体への入力は画像および左肩予測情報、出力は左肩領域および左肘予測情報である。

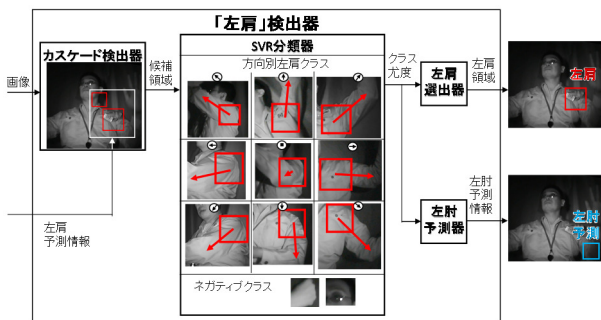


図2 左肩検出器の構成

「左肩」検出器はさらに、カスケード検出器、SVR 分類器、「左肩」選出器、「左肘」予測器からなり、以下にその機能および構成を述べる。

### 2.2. カスケード検出器

カスケード検出器は、弱検出器を複数段組み合わせた学習ベースの検出器である。検出機序は、サイズ可変のスライディングウィンドウを少しずつ動かして、その領域が左肩であるかの妥当性判定を繰り返すことである。しかし、判定回数が非常に多いため処理量が増大することがあり、処理量削減のため、判定の高速化と、判定回数の削減の双方を行う。

前者は、軽量なLBP (Local Binary Pattern)特徴量の利用によって行う。

そして後者に関しては、判定回数に影響する、候補領域として許容される位置・座標のバリエーションを減らすことで行う。具体的には、スライディングウィンドウの形状を正方形に限定し、さらにウィンドウの移動可能範囲を前段から入力された左肩の予測座標範囲周辺に、また、ウィンドウのサイズの範囲を、予測されるバウンディングボックスサイズに近いものに、各々限定する。

なお後者は、原理上、予測情報との乖離が大きな誤検出(False Positive)がなくなるため、誤検出の抑制も同時に実現できる点で有利である。

カスケード検出器は、判定された、個数可変の「左肩」候補領域を出力する。

### 2.3.SVR 分類器

SVR 分類器は、カスケード検出器の出力する候補領域の画像特徴量(ここでは精度を重視し HOG特徴量を用いた)を計算し、10 クラスの尤度計算を行う。うち9クラスはいずれも左肩画像のクラスであるが、左肩位置を基準とする左肘の向き(上下左右および斜めに加え、左肘と左肩が近接している場合の方向無し)毎に別々のクラスを立て、さらに左肘ではない画像ばかりを集めたネガティブクラスが加わり、合計 10 クラスとなる。

これは、SVR 学習モデルを構築する際のアノテーションにおいて、単に左肩領域を選択するだけでなく、さらに「左肘が9方向のどこにあるか」のラベリングを要することを意味する。しかし、現実には「左肘」検出器の学習のため左肘の学習データのアノテーションも行うので、その結

果が利用できる。すなわち、同一画像の学習データの両者の位置関係を計算することによって、このラベリングは自動化できる。

なお、SVR 分類器で用いられるカーネルには線形写像と非線形写像があるが、処理速度を重視して線形写像を用いる。

### 2.4. 「左肩」選出器

「左肩」選出器は、候補領域のうち、どの領域が真に左肩であるかを1個に絞り込む。具体的には、SVR が出力したネガティブクラス尤度が一番低い領域を、実際の左肩領域として選択する。ただし、ネガティブクラス尤度が全クラス中最も高い候補領域は棄却される。

また、複数人の骨格を検出する際、鼻選出器にのみ、この候補領域の絞り込みを行わないことで複数人の検出を行う。

### 2.5. 「左肘」予測器

「左肘」予測器は、各クラス尤度を元に、左肘の予測情報を出力する。計算に先立ち、クラス尤度からネガティブクラス分を除く補正を行う。すなわち、クラス番号 $c$ の尤度を $l_c$ 、ネガティブクラスの尤度を $l_{neg}$ とした際、クラス番号 $c$ の補正尤度 $\tilde{l}_c$ は、式(1)で得られる。

$$\tilde{l}_c = l_c / (1 - l_{neg}) \quad (1)$$

そして、左肩から左肘への予測差分ベクトル $\hat{d}$ は、左肩の候補領域の一边のサイズを $s$ 、クラス番号 $c$ の $t_c$ とした際、式(2)にて得られる。

$$\hat{d} = s \sum_c \tilde{l}_c t_c \quad (2)$$

左肘予測位置は、左肩位置、すなわち左肩バウンディングボックスの中央から $\hat{d}$ だけ移動した位置である。

ただし、学習データ左肩左肘平均ベクトル $t_c$ は、クラス番号 $c$ に分類される学習データの、から左肘へのベクトルを、左肩のボックスサイズで割った値の平均であり、学習データ収録時に計算しておく。この計算も自動化が可能である。

この予測について簡単に説明したのが図3である。ここでは簡単に、下向きと右下向き以外のクラスの尤度は0であったとしている。予測差分

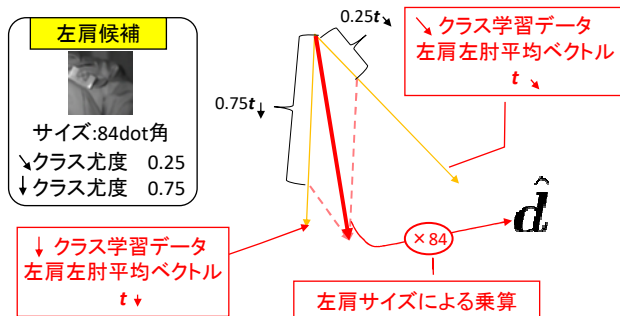


図3 左肩からの左肘予測の一例

ベクトル $\hat{d}$ は、下向きクラスの学習データ左肩左肘平均ベクトル $t_{\downarrow}$ に左肩クラス尤度0.75を乗じたベクトルと、右下向きクラスの学習データ左肩左肘平均ベクトル $t_{\searrow}$ に右下向きクラス尤度0.25を乗じたベクトルの和を84倍して得られる。この84(該当候補領域のサイズ)倍計算は、 $t_{\downarrow}$ および $t_{\searrow}$ が、「左肩」サイズが仮に1dot角であった際の長さに正規化されているため行う。

予測サイズも同様に、学習データの実績値の尤度による重みづけに、サイズ $s$ による乗算処理を施して得られる。

## 3. 適用例

前節で述べた通り、本骨格検出器の特徴は、隣接する骨格の予測による高速化と高精度の両立にあり、その効果を、鼻・首パーツの検出例にて図4に示す。

上段にはボックスが2個あり、これはカスケード検出器が出力した2個の鼻候補領域である。それぞれのボックスから伸びた矢印が、予測差分ベクトルである。正しい鼻領域から伸びたベクトルが予測差分ベクトル $\hat{d}$ であり、その先端が首予測座標である。

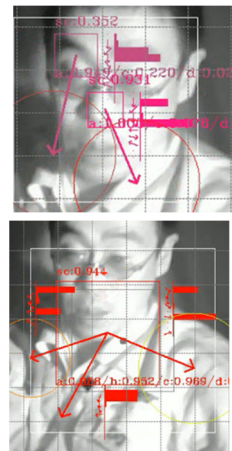


図4 鼻・首検出例

さらに、図4下段がその予測をもとに行った首の検出である。白枠が予測座標をもとに限定された探索範囲で、その中で適正と思われる位置に首領域を発見している。さらに、両肩および胸への予測ベクトルも表示されており、概ね正しいことが見てとれる。

## 4. まとめ

本稿では、一般的な深層学習に代えて、SVRを用いる骨格検出器を提案した。SVRの用途は、ある骨格に後続する骨格の予測であり、次段の候補領域探索範囲を、その予測に従って限定することで、高速性と高精度を両立している。

今後は、この手法の定量評価や課題抽出を行い、実用的な速度・精度を達成するための改善を行う予定である。

### 引用文献

- [1] e. a. Z. Cao, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, CVPR, 2017.