

# 修士論文概要書

Summary of Master's Thesis

Date of submission: 01 / 30 / 2018 (MM/DD/YYYY)

|                            |                    |                              |                  |                       |                |
|----------------------------|--------------------|------------------------------|------------------|-----------------------|----------------|
| 専攻名 (専門分野)<br>Department   | 情報理工情報通信<br>専攻     | 氏名<br>Name                   | 高木 政徳            | 指 導<br>教 員<br>Advisor | 渡辺 裕 印<br>Seal |
| 研究指導名<br>Research guidance | オーディオビジュ<br>アル情報処理 | 学籍番号<br>Student ID<br>number | CD<br>5116F053-3 |                       |                |
| 研究題目<br>Title              | 姿勢情報を用いた自転車選手識別手法  |                              |                  |                       |                |

## 1. はじめに

2020年東京オリンピック・パラリンピック競技大会の開催を間近に控え、スポーツ選手の育成やファン層の拡大、ファンのコア化などを目的としたスポーツ映像処理技術の需要が高まっている。実際に映像処理技術がスポーツの放送に応用されている例として、水泳競技でのコース上への選手名や記録ラインの重畳やテニスの国際大会で採用されている The Hawk-Eye Officiating System などが挙げられる。これらのように、データを可視化することで様々な情報を視聴者に提供することが可能となる。

ツール・ド・フランスをはじめとする自転車ロードレース競技では、数百名もの選手が出場する。レースでは4-9人でチームを編成し、同一チームの選手は同一のユニフォームとヘルメットを着用している。映像から選手を特定する方法として、顔特徴量を用いた認証技術やゼッケンナンバー読み取りによる選手リストとの照合などが考えられる。しかし、大多数の選手はレース中にサングラスを着用することや、カメラ位置によっては選手の顔やゼッケンナンバーが映らない、引きで撮影した映像では選手のサイズが小さく顔やゼッケンナンバーを読み取れないなどの問題がある。そこで本研究では、Realtime Multi-Person 2D Pose Estimation[1]を用いて得られる姿勢情報を用いて、実際のレース映像での自転車選手識別手法を提案する。自転車ロードレースでは1チーム4-9人で構成されるため、チーム分類と選手識別(最大9人)の手法を提案する。なお本研究では、自転車選手の正面映像を対象とする。

## 2. 関連技術

本研究では、Zheらが提案した Realtime Multi-Person 2D Pose Estimation[1] (以下 OpenPose) を用いて自転車選手の姿勢推定を行う。Zheらの手法では Confidence Map と Part Affinity Fields (PAFs) を用いた二つの逐次予測プロセスにより、画像内人物の体のパーツ位置(肘, 肩, 足首など合計18部位)とパーツ間の結合性をボトムアップ的アプローチにより推定している。これにより、リアルタイム性を保持しつつ、画像内の複数人物の姿勢を高精度に推定し、18部位の位置座標で表される人物の関節位置情報を得ることができる。

## 3. 提案手法

### 3.1 ペダリング周期変換(時間正規化)

姿勢推定によって得られた自転車選手の右足首の y 座標の時間経過グラフを図1に示す。なお、左足首でも同様のグラフとなる。図1より、右足首の周期的な動きからピークを検出することでペダリング周期を求めることができる。ペダリング周期が30フレームで1回転となるように、姿勢推定により得られる各部位座標位置を補間処理する。補間処理には3次スプライン補間法を用いる。3次スプライン補間法は、3次の多項式で小区間を近似する方法であり、データ数が N+1 個の区分多項式 S は、以下の式(1)で与えられる。

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3 \quad (1) \\ j = 0, 1, 2, 3, \dots, N-1$$

式(1)の係数 a, b, c, d を求めることで3次スプライン補間の関数を決定できる。

### 3.2 スケール変換・位置補正

レース映像は移動カメラによって撮影されたものである。したがって、レース映像中における選手位置や大きさはカメラとの距離によって変化する。そこで、自転車選手を正面から撮影した映像において、左右の肩から撮影カメラまでの距離に差がないという前提のもと、両肩間の距離が一定になるように関節位置情報をスケール変換し、さらに首の位置が原点となるように座標位置を補正する。

### 3.3 異常値検出

OpenPoseによって得られた姿勢情報には、図2に示すように姿勢推定が失敗する場合がある。本研究で用いる映像にも、図2のように姿勢推定を失敗している場合がある。そこで、これらの失敗を異常値として検出し、SVMの訓練データから取り除く。

スケール変換後の姿勢情報について、フレームごとの各パーツ間の長さ(右手首-右肘など)は概ね一定の値の範囲に収まる。各パーツ間の長さについて、フレームごとに偏差値を算出し、その値が70以上または30以下となるものを異常値として検出し、異常値を含むフレームを異常フレームとする。ペダリング1回(30フレーム)中に含まれる異常フレーム数に応じて異常値レベルを設定する。1サイクルに含まれる異常フレーム数と全データにおける異常サイクルの割合を表1に示す。

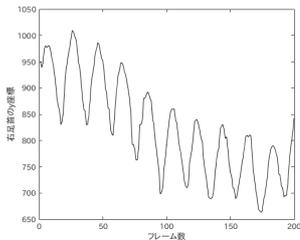


図1 右足首の座標変化



図2 姿勢推定失敗例

表1 異常値レベルごとの異常フレーム数

| 異常値レベル | 1 サイクル中の異常フレーム数 | 異常サイクル率 |
|--------|-----------------|---------|
| レベル 1  | 20-30           | 6.5%    |
| レベル 2  | 15-30           | 11.8%   |
| レベル 3  | 10-30           | 20.8%   |

### 3.4 複数サイクル多数決識別法

複数サイクル多数決識別法の概要を図3に示す。この手法では、入力データであるサイクル1からサイクルtまでの各サイクルのSVMによる予測結果から、最も多く予測された選手を予測選手と決定する。ただし、最も多く予測された選手が複数人いた場合、そのうち最近予測された選手を予測選手と決定する。

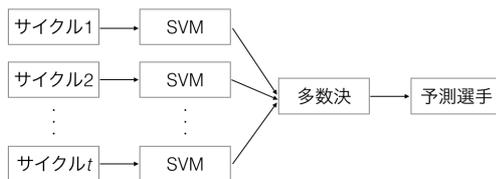


図3 複数サイクル多数決分類法

## 4. 評価実験

### 4.1 チーム識別

ツール・ド・フランス2017に出場した22チームの選手について評価用画像1100枚、学習用画像2200枚の合計3300枚の画像を用意した。学習用画像についてはデータオーギュメンテーションを行い、画像の枚数を13200枚に増やした。なお、データオーギュメンテーションでは、学習用画像に左右反転、輝度変化を施している。

本実験は、既存のCNNモデルであるGoogleNetを使用した。エポック数を50としたときの評価画像のチーム識別実験結果を表に示す。

表2 チーム識別結果

| Accuracy | Precision | Recall |
|----------|-----------|--------|
| 99.80%   | 97.87%    | 97.82% |

表2より、Accuracy, Precision, Recall はいずれの値も97%以上と、非常に高い結果となった。これは各チームのユニフォームの色や模様など特徴点となり得る箇所が多かったからであると考えられる。

### 4.2 選手識別

ペダリング1回分の姿勢情報を1つの特徴ベクトルとし、SVMによる選手の識別実験を行った。被験者数は4-9名

とし、1人あたり6シーンの映像(合計60秒程度)を使用する。なお、SVMの訓練データおよび学習データは各選手につき2パターン用意し、用意した全パターンについての識別結果正解率の平均値を算出した。

OpenPoseによって得られた姿勢情報に対し、スケール変換・位置補正、ペダリング周期変換(時間正規化)および異常値検出処理をし、SVMで学習し識別した結果を表3に示す。ここで、レベル0とは異常値サイクルの除去を行わなかった場合を表す。また、被験者4-9人について、サイクル $1 \leq t \leq 12$ までの各サイクルの分類結果を考慮し、最終的な選手を予測した場合の結果を図4に示す。

表3 各異常値レベルでのSVMによる識別結果

|    | レベル0   | レベル1   | レベル2   | レベル3   |
|----|--------|--------|--------|--------|
| 4人 | 82.10% | 80.02% | 81.70% | 81.71% |
| 5人 | 82.95% | 83.85% | 84.12% | 85.58% |
| 6人 | 81.80% | 82.17% | 83.67% | 83.89% |
| 7人 | 78.24% | 81.51% | 80.37% | 80.28% |
| 8人 | 72.97% | 73.07% | 74.70% | 76.08% |
| 9人 | 66.02% | 64.06% | 67.64% | 69.17% |

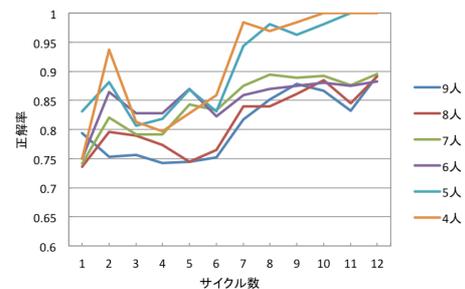


図4 複数サイクル多数決識別法による識別結果

## 5. まとめ

本研究では、自転車選手の正面映像に対してOpenPoseを用いて姿勢推定を行い、得られた姿勢情報から自転車選手を識別する手法を提案した。

CNNを用いた22チームの識別実験では、Accuracyで99.8%の精度が出ることを確認し、非常に高い精度でチーム識別可能であることがわかった。また、OpenPoseによって得られた姿勢情報に対し、スケール変換・位置補正、ペダリング周期変換(時間正規化)および異常値検出処理をし、SVMで学習し識別することで4人の場合で82.10%、9人の場合で69.17%の精度が出ることを確認した。

さらに、複数サイクル多数決分類法により、1サイクルだけではなく複数サイクルでのSVM分類結果を考慮することで、4,5人の場合で6サイクル、6-9人の場合で10サイクルで90%程度の精度で選手識別が可能であることを確認した。

## 参考文献

- [1] Zhe Cao, Tomas Simon, Shin-En Wei, and Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv preprint arXiv:1611.08050, 2016.

2017年度 修士論文

# 姿勢情報を用いた自転車選手識別手法

A Method of Bike Racer Identification  
Using Posture Information

提出日：2018年1月30日

指導教員：渡辺 裕教授

早稲田大学理工学術院 基幹理工学研究科  
情報理工・情報通信専攻

学籍番号：5116F053-3

高木 政徳

# 目次

|              |   |           |
|--------------|---|-----------|
| <b>第 1 章</b> | <b>序論</b>   | <b>1</b>  |
| 1.1          | 研究の背景 . . . . .                                     | 1         |
| 1.2          | 研究の目的 . . . . .                                     | 2         |
| 1.3          | 本論文の構成 . . . . .                                    | 2         |
| <b>第 2 章</b> | <b>関連技術</b>   | <b>4</b>  |
| 2.1          | OpenPose . . . . .                                  | 4         |
| 2.1.1        | Simultaneous Detection and Association . . . . .    | 4         |
| 2.1.2        | Confidence Maps for Part Detection . . . . .        | 6         |
| 2.1.3        | Part Affinity Fields for Part Association . . . . . | 6         |
| 2.1.4        | Multi-Person Parsing using PAFs . . . . .           | 7         |
| 2.2          | サポートベクターマシン . . . . .                               | 9         |
| 2.2.1        | 線形 SVM . . . . .                                    | 9         |
| 2.2.2        | 非線形 SVM . . . . .                                   | 11        |
| 2.2.3        | 多クラス分類 . . . . .                                    | 14        |
| <b>第 3 章</b> | <b>提案手法</b>   | <b>16</b> |
| 3.1          | 選手抽出 . . . . .                                      | 16        |
| 3.2          | ペダリング周期変換 (時間正規化) . . . . .                         | 17        |
| 3.2.1        | ピーク検出 . . . . .                                     | 17        |
| 3.2.2        | 3 次スプライン補間 . . . . .                                | 19        |
| 3.3          | 位置補正 . . . . .                                      | 22        |
| 3.4          | スケール変換 . . . . .                                    | 23        |
| 3.5          | 異常値検出 . . . . .                                     | 23        |
| 3.6          | 複数サイクル多数決識別法 . . . . .                              | 24        |

|            |                    |           |
|------------|--------------------|-----------|
| 3.7        | むすび                | 25        |
| <b>第4章</b> | <b>評価実験</b>        | <b>30</b> |
| 4.1        | チーム分類              | 30        |
| 4.1.1      | データセットの生成          | 30        |
| 4.1.2      | 畳み込みニューラルネットワークの学習 | 30        |
| 4.1.3      | 評価実験               | 31        |
| 4.2        | 選手識別               | 33        |
| 4.2.1      | ペダリング周期を考慮しない場合    | 33        |
| 4.2.2      | ペダリング周期を考慮した場合     | 34        |
| 4.2.3      | 複数サイクル多数決識別法       | 38        |
| 4.3        | むすび                | 38        |
| <b>第5章</b> | <b>結論</b>          | <b>40</b> |
|            | <b>謝辞</b>          | <b>41</b> |
|            | <b>参考文献</b>        | <b>42</b> |
|            | <b>研究業績</b>        | <b>44</b> |

# 目次

|      |  |    |
|------|--|----|
| 2.1  | OpenPose の構造 . . . . .                 | 6  |
| 2.2  | Part Affinity Fields のベクトル値例 . . . . . | 8  |
| 3.1  | 右足首の座標変化 . . . . .                     | 18 |
| 3.2  | 右足首の座標変化 ( $D = 5$ ) . . . . .         | 18 |
| 3.3  | 右足首の座標変化 ( $D = 15$ ) . . . . .        | 18 |
| 3.4  | 右足首の座標変化 ( $D = 30$ ) . . . . .        | 18 |
| 3.5  | 姿勢推定失敗例 . . . . .                      | 24 |
| 3.6  | 複数サイクル多数決識別法 . . . . .                 | 25 |
| 3.7  | パーツ間の長さ (首) . . . . .                  | 26 |
| 3.8  | パーツ間の長さ (右上腕) . . . . .                | 26 |
| 3.9  | パーツ間の長さ (右前腕) . . . . .                | 26 |
| 3.10 | パーツ間の長さ (左上腕) . . . . .                | 26 |
| 3.11 | パーツ間の長さ (左前腕) . . . . .                | 26 |
| 3.12 | パーツ間の長さ (右大腿) . . . . .                | 26 |
| 3.13 | パーツ間の長さ (右下腿) . . . . .                | 27 |
| 3.14 | パーツ間の長さ (左大腿) . . . . .                | 27 |
| 3.15 | パーツ間の長さ (左下腿) . . . . .                | 27 |
| 3.16 | パーツ間長さのヒストグラム (首) . . . . .            | 28 |
| 3.17 | パーツ間長さのヒストグラム (右上腕) . . . . .          | 28 |
| 3.18 | パーツ間長さのヒストグラム (右前腕) . . . . .          | 28 |
| 3.19 | パーツ間長さのヒストグラム (左上腕) . . . . .          | 28 |
| 3.20 | パーツ間長さのヒストグラム (左前腕) . . . . .          | 28 |
| 3.21 | パーツ間長さのヒストグラム (右大腿) . . . . .          | 28 |
| 3.22 | パーツ間長さのヒストグラム (右下腿) . . . . .          | 29 |

|      |                               |    |
|------|-------------------------------|----|
| 3.23 | パーツ間長さのヒストグラム (左大腿) . . . . . | 29 |
| 3.24 | パーツ間長さのヒストグラム (左下腿) . . . . . | 29 |
| 4.1  | 学習時の精度推移 . . . . .            | 32 |
| 4.2  | 異常値レベルと正解率 . . . . .          | 36 |
| 4.3  | 鼻位置の変化の様子 . . . . .           | 37 |
| 4.4  | 半身データ . . . . .               | 37 |
| 4.5  | 複数サイクル多数決識別法 . . . . .        | 38 |

# 表目次

|      |                                 |    |
|------|---------------------------------|----|
| 2.1  | 検出パーツ一覧 . . . . .               | 8  |
| 3.1  | 異常値レベルごとの異常フレーム数 . . . . .      | 24 |
| 4.1  | チームごとの学習用画像と評価用画像 . . . . .     | 31 |
| 4.2  | Confusion Matrix . . . . .      | 32 |
| 4.3  | チーム分類結果 . . . . .               | 32 |
| 4.4  | 訓練・テストデータ数 . . . . .            | 33 |
| 4.5  | ペダリング周期を考慮しない場合の識別結果 . . . . .  | 34 |
| 4.6  | 処理手順比較 . . . . .                | 34 |
| 4.7  | 各異常値レベルでの SVM による分類結果 . . . . . | 35 |
| 4.8  | 訓練・テストデータ数 . . . . .            | 35 |
| 4.9  | 鼻位置データ除去 . . . . .              | 37 |
| 4.10 | 半身データを用いた SVM による分類結果 . . . . . | 38 |

# 第 1 章

## 序論

### 1.1 研究の背景

2020 年東京オリンピック・パラリンピック競技大会の開催を間近に控え、スポーツ選手の育成やファン層の拡大、ファンのコア化などを目的としたスポーツ映像処理技術の需要が高まっている。実際に映像処理技術がスポーツの放送に応用されている例として、水泳競技でのコース上への選手名や記録ラインの重畳、メジャーリーグベースボールにて全球団の本拠地球場に導入されているボールトラッキングシステム PITCHf/x[1] や、テニスの国際大会で採用されている The Hawk-Eye Officiating System[2] などが挙げられる。これらのように、データを可視化することで様々な情報を視聴者に提供することが可能となる。

自転車ロードレースは団体競技であり、主に一般道を走り、ゴールの着順や所要時間を争う競技である。レースの規模によって出場する選手の数異なるが、毎年ヨーロッパで行われる世界三大自転車ロードレース大会(グランツール)の一つであるツール・ド・フランスは、約 3 週間で 3000 キロメートル以上を走破する長距離レースであり、22 チーム(1 チームあたり 9 人)のおよそ 200 名の選手が出場し一斉に走る。レースの様子は全世界 190 ヶ国で放送され、大会期間中の総視聴者数は 35 億人に上る [3] とも言われている。

ツール・ド・フランスをはじめとする自転車ロードレースでは、4-9 人でチームを編成し、同一チームの選手は同一のユニフォームとヘルメットを着用している。また、大多数の選手は紫外線や風などから目を保護する目的でサングラスを着用している。そのため、各選手の見た目にさほど差がなく、レース映像の視聴者が優勝候補の選手など、特定の選手を見分けるのが困難である。

映像から選手個人を特定する方法として、顔特徴量を用いた認証技術 [4, 5] や各選手が

両腰後ろに着用するゼッケンナンバーの読み取り [6] による選手リストとの照合などが考えられる。しかし、上記で述べたように大多数の選手はレース中にサングラスを着用することや、カメラ位置によっては選手の顔やゼッケンナンバーが映らない、引きで撮影した映像では選手のサイズが小さく顔やゼッケンナンバーを読み取れないなどの問題がある。これらの理由から、上記手法によって、レース映像から特定の選手を見つけ出すのは困難であると考えられる。

これらの他に、体の動きから個人識別を行う手法として歩容認証技術 [7, 8] が知られている。歩容認証は腕の振りや歩幅、姿勢の違い等の歩き方の特徴から個人識別を行い、その手法はシルエットベース手法とモデルベース手法に大別されている。歩行者のシルエット情報を用いたシルエットベース手法では、低解像度映像でも識別が可能であるが、撮影角度やフレームレートの影響を大きく受ける。一方、人体モデルへの当てはめを行うモデルベース手法は、低解像度映像での識別は困難であるが、防犯カメラの画質向上や姿勢検出技術の向上により今後の発展が見込まれている [9]。いずれの手法も固定カメラで撮影された入力映像を前提としており、人物領域の抽出に背景差分法を用いる。したがって、カメラ位置が非固定の自転車ロードレース映像への応用は困難である。これに対して、姿勢情報を用いた個人識別はレース映像に対しても有効であると考えられる。

## 1.2 研究の目的

自転車ロードレースには数多くの選手が出場する。そのためレース映像には多くの選手が映り、顔認証などによりその中から特定の選手を見つけるのは前節で述べた理由から困難であると考えられる。そこで本研究では、OpenPose[10] を用いて得られる姿勢情報を用いて、実際のレース映像での自転車選手識別手法を提案する。自転車ロードレースでは1チーム4–9人で構成されるため、チーム識別した後、選手識別(最大9人)をする手法を提案する。なお本研究では、自転車選手の正面映像を対象とする。

## 1.3 本論文の構成

本論文は、5章で構成されている。

第1章「序論」は本章であり、本研究の背景と目的を述べている。

第2章「関連技術」では、本研究で用いる姿勢推定手法である OpenPose および、パターン認識モデルの一つである SVM について説明する。

第3章「提案手法」では、自転車選手正面映像に対する選手識別手法について述べる。

第4章「評価実験」では、畳み込みニューラルネットワーク (Convolutional Neural Network, 以下 CNN) を用いたチーム分類の評価実験と、第3章で提案した手法の評価実験について述べる。

第5章では、全体のまとめを述べる。

## 第 2 章

# 関連技術

本章では、本研究で用いる姿勢推定手法である OpenPose および、パターン認識モデルの一つである SVM について説明する。

### 2.1 OpenPose

Zhe ら [10] が提案した Real time multi person 2D pose estimation(以下, OpenPose) は、画像中の複数人物の 2D 姿勢を効率良く検出する手法である。画像内の人物を検出し、検出された各人物に対して姿勢推定を行うという従来 방식 [11, 12] とは異なり、Confidence Maps と Part Affinity Fields を用いた二つの逐次予測プロセスによって、画像内の体のパーツ位置およびパーツ間の関係性をボトムアップ的アプローチで推定している。これにより、リアルタイム性を保持しつつ高精度に姿勢情報を推定することができる。OpenPose によって推定できるパーツ位置を表 2.1 に示す。

#### 2.1.1 Simultaneous Detection and Association

図 2.1 に OpenPose の構造図を示す。OpenPose はサイズ  $w \times h[pe]$  のカラー画像を入力画像とし、画像内の各人物のパーツ位置座標を出力として生成する。入力画像は VGG-19[13] の最初の 10 層で初期化され、特徴マップ  $F$  として Stage 1 の入力となる。各 Stage のネットワークは二つに分かれており、図 2.1 の Branch 1 でパーツ位置を示す 2D Confidence Maps  $S$  を予測し、Branch 2 ではパーツ間の関係性を示す 2D ベクトル場を表す Part Affinity Fields  $L$  を予測する。Confidence Maps  $S$  および Part Affinity Fields  $L$  はそれぞれ次の式 (2.1), 式 (2.2) で表される。

$$\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_J), \mathbf{S}_j \in \mathbb{R}^{w^*h}, j \in (1, \dots, J) \quad (2.1)$$

$$\mathbf{L} = (\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_C), \mathbf{L}_c \in \mathbb{R}^{w^*h*2}, c \in (1, \dots, C) \quad (2.2)$$

なお,  $J$  はパーツ数,  $C$  はパーツペア数である.

Stage 1 で Confidence Maps  $\mathbf{S}^1 = \rho^1(\mathbf{F})$  および Part Affinity fields  $\mathbf{L}^1 = \phi^1(\mathbf{F})$  を生成する. ここで,  $\rho^1$  および  $\phi^1$  は, Stage 1 における CNN の推定結果である. 各 Stage における入力, 1 つ前の Stage での結果および元の Stage 1 への入力  $\mathbf{F}$  を合わせたものであり, 以下のように表される.

$$\mathbf{S}^t = \rho^t(\mathbf{F}, \mathbf{S}^{t-1}, \mathbf{L}^{t-1}), \forall t \geq 2 \quad (2.3)$$

$$\mathbf{L}^t = \phi^t(\mathbf{F}, \mathbf{S}^{t-1}, \mathbf{L}^{t-1}), \forall t \geq 2 \quad (2.4)$$

各 Stage の終わりにそれぞれの Branch に損失関数を適用する. ステージ  $t$  の各 Branch の損失関数は以下の式で示される.

$$f_{\mathbf{S}}^t = \sum_{j=1}^J \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{S}_j^t(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\|_2^2 \quad (2.5)$$

$$f_{\mathbf{L}}^t = \sum_{c=1}^C \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{L}_c^t(\mathbf{p}) - \mathbf{L}_c^*(\mathbf{p})\|_2^2 \quad (2.6)$$

ここで,  $\mathbf{S}^*$ ,  $\mathbf{L}^*$  は正解の Confidence maps および Part Affinity Fields,  $\mathbf{W}$  はバイナリマスクであり, 場所  $\mathbf{p}$  にアノテーションがない場合  $\mathbf{W}(\mathbf{p}) = 0$  となる.

全体の目的関数は以下で表される.

$$f = \sum_{t=1}^T (f_{\mathbf{S}}^t + f_{\mathbf{L}}^t) \quad (2.7)$$

ここで,  $T$  は最大 Stage 数を表す.

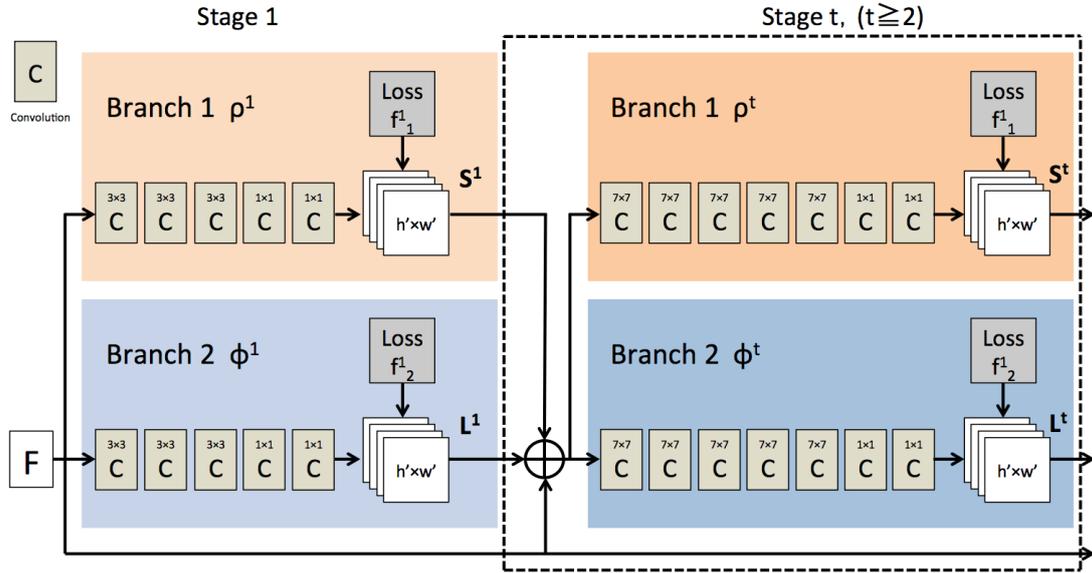


図 2.1 OpenPose の構造

### 2.1.2 Confidence Maps for Part Detection

式 (2.7) での評価のためには，正解の Confidence Maps  $\mathbf{S}^*$  をアノテーションされた 2D キーポイントから生成する必要がある．人物が複数人いて，各パーツが見えるとき，各パーツ  $j$ ，人物  $k$  に対応する Confidence Maps のピークを求める．まず，個々の Confidence Maps  $\mathbf{S}_{j,k}^*$  (正解値) を各人物  $k$  について生成する．

$$\mathbf{S}_{j,k}^*(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{x}_{j,k}\|_2^2}{\sigma^2}\right) \quad (2.8)$$

ネットワークから予測される形式に合わせて，全ての人に関して Map を結合する．このとき，各パーツに関する正解の Confidence Maps は以下のように表される．

$$\mathbf{S}_j^*(\mathbf{p}) = \max_k \mathbf{S}_{j,k}^*(\mathbf{p}) \quad (2.9)$$

### 2.1.3 Part Affinity Fields for Part Association

任意の人数の人物に関して，検知されたパーツ同士をどのように繋げるかが問題となる．図 2.2 に Part Affinity Fields のベクトル値の例を示す．図 2.2 において， $x_{j_1,k}$ ， $x_{j_2,k}$

を人  $k$  の四肢  $c$  における部位  $j_1, j_2$  の正解値としたとき、四肢  $c$  上のポイント  $p$  に関して、 $\mathbf{L}_{c,k}^*(p)$  は  $j_1$  から  $j_2$  への単位ベクトルとなる。定式化すると以下の式 (2.10) で表される。

$$\mathbf{L}_{c,k}^*(p) = \begin{cases} \mathbf{v} & (\text{if } p \text{ on limb } c, k) \\ \mathbf{0} & (\text{otherwise}) \end{cases} \quad (2.10)$$

ネットワークから予測される形に合わせて、全ての人に関して Part Affinity Fields を合わせて四肢ごとの Fields を生成する。

$$\mathbf{L}_c^*(p) = \frac{1}{n_c(p)} \sum_k \mathbf{L}_{c,k}^*(p) \quad (2.11)$$

テスト時は、対応する Part Affinity Fields の線積分を候補パーツの位置を結ぶ線分に沿って計算することで、候補パーツ間の関連度を算出する。

$$E = \int_{u=1}^{u=0} \mathbf{L}_c(\mathbf{p}(u)) \cdot \frac{\mathbf{d}_{j_2} - \mathbf{d}_{j_1}}{\|\mathbf{d}_{j_2} - \mathbf{d}_{j_1}\|_2} du \quad (2.12)$$

$$\mathbf{p}(u) = (1 - u)\mathbf{d}_{j_1} + u\mathbf{d}_{j_2} \quad (2.13)$$

#### 2.1.4 Multi-Person Parsing using PAFs

各パーツに関して複数検知される場所があり、その分関連づけられる可能性のある四肢のパターンは多くなる。Affinity Fields 上の線積分計算により各候補四肢にスコアをつける。最適な組み合わせを見つける問題は、NP-Hard 問題である  $K$  次元マッチング問題に対応する。

$$\max_{Z_c} E_c = \max_{Z_c} \sum_{m \in D_{j_1}} \sum_{n \in D_{j_2}} E_{mn} \cdot z_{j_1 j_2}^{mn} \quad (2.14)$$

$$\max_Z E = \sum_{c=1}^C \max_{Z_c} E_c \quad (2.15)$$

表 2.1 検出パーツ一覧

| 検出部位 |                |
|------|----------------|
| 1    | Nose           |
| 2    | Neck           |
| 3    | Right Shoulder |
| 4    | Right Elbow    |
| 5    | Right Wrist    |
| 6    | Left Shoulder  |
| 7    | Left Elbow     |
| 8    | Left Wrist     |
| 9    | Right Hip      |
| 10   | Right Knee     |
| 11   | Right Ankle    |
| 12   | Left Hip       |
| 13   | Left Knee      |
| 14   | Left Ankle     |
| 15   | Right Eye      |
| 16   | Left Eye       |
| 17   | Right Ear      |
| 18   | Left Ear       |

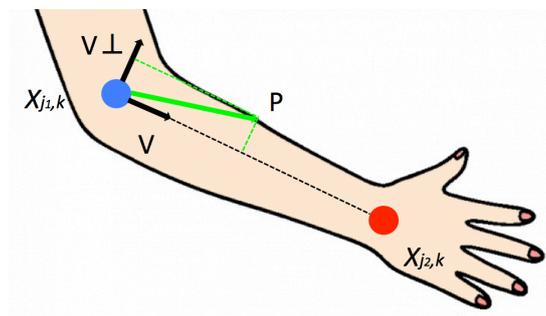


図 2.2 Part Affinity Fields のベクトル値例

## 2.2 サポートベクターマシン

サポートベクターマシン [14, 15](以下, SVM) は教師あり学習モデルの 1 つであり, 画像, 音声, 言語など様々な分野で利用されている [16, 17, 18]. 基本となる 2 クラス分類では, 2 クラスを識別する超平面と識別データとの距離を表すマージンを最大化することにより高い汎化能力を得ることができる.

### 2.2.1 線形 SVM

$n$  個の事例からなる訓練集合  $\{(\mathbf{x}_i, y_i)\}_{i \in [n]}$  が  $d$  次元実数ベクトル  $\mathbf{x}_i \in \mathbb{R}^d$  およびラベル  $y_i \in \{-1, 1\}$  で構成されているとする. このとき, 決定関数を  $f$  とし, 分類器  $g$  を以下のように定義する.

$$g(\mathbf{x}) = \begin{cases} 1 & (f(\mathbf{x}) > 0) \\ -1 & (f(\mathbf{x}) < 0) \end{cases} \quad (2.16)$$

さらに  $f(\mathbf{x})$  については, 以下の関数を考える.

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R} \quad (2.17)$$

$g(\mathbf{x})$  の定義では  $f(\mathbf{x}) = 0$  が分類結果が変化する境目となっているため,  $f(\mathbf{x}) = 0$  となる  $\mathbf{x}$  は二つのクラスを分ける境界となっており, これを分類境界と呼ぶ.

#### 2.2.1.1 ハードマージン

訓練集合  $\{(\mathbf{x}_i, y_i)\}_{i \in [n]}$  が分離可能の場合, つまり, 訓練集合内のすべてのデータが正しいクラスに分類可能な場合を考える. ある  $y_i$  について, 分類に成功しているなら  $y_i$  と  $f(\mathbf{x}_i)$  の符号が一致するので  $y_i f(\mathbf{x}_i) > 0$  が成り立つ. したがって, 分離可能な訓練集合にはすべての  $i \in [n]$  に対して  $y_i f(\mathbf{x}_i) > 0$  が成り立つような  $f(\mathbf{x}_i)$  が存在し, 一般に訓練集合を分離する分類境界は複数存在する. そこで SVM では, 各クラスのデータ集合が分類境界からなるべく遠くなるように分類境界を定め, この距離 (マージン) を最大化することを考える.

ある  $\mathbf{x}_i$  から分類境界までの距離は, 点と平面の距離の公式より以下の式 (2.18) のよう

に表現できる.

$$\frac{\mathbf{w}^T \mathbf{x}_i + b}{\|\mathbf{w}\|} \quad (2.18)$$

訓練集合が分離可能のとき, すべての  $i \in [n]$  に対して  $y_i f(\mathbf{x}_i) > 0$  なので, ある正の実数  $M > 0$  に対して,  $y_i f(\mathbf{x}_i) \geq M$  がすべての  $i \in [n]$  に成立しているとする, マージンの最大化は以下の最適化問題で表現できる.

$$\max_{\mathbf{w}, b, M} \frac{M}{\|\mathbf{w}\|} \quad (2.19)$$

なお, 制約条件は  $y_i f(\mathbf{x}_i) \geq M, i \in [n]$  である. 分離境界に最も近い  $\mathbf{x}_i$  を  $\mathbf{x}_{i'}$  とすると,

$$\frac{M}{\|\mathbf{w}\|} = \frac{y_{i'}(\mathbf{w}^T \mathbf{x}_{i'} + b)}{\|\mathbf{w}\|} = \frac{|\mathbf{w}^T \mathbf{x}_{i'} + b|}{\|\mathbf{w}\|} \quad (2.20)$$

ここで,  $\tilde{\mathbf{w}} = \mathbf{w}/M, \tilde{b} = b/M,$  とすると, 式 (2.19) は以下に変形でき,

$$\max_{\tilde{\mathbf{w}}, \tilde{b}} \frac{1}{\|\tilde{\mathbf{w}}\|} \quad (2.21)$$

さらに,  $1/\|\mathbf{w}\|$  の最大化は  $\|\mathbf{w}\|^2$  の最小化と等価であるので, 式 (2.19) は次の式と等価である.

$$\min_{\mathbf{w}, b} \|\mathbf{w}\|^2 \quad (2.22)$$

なお制約条件は  $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i \in [n]$  である.

### 2.2.1.2 ソフトマージン

訓練集合  $\{(\mathbf{x}_i, y_i)\}_{i \in [n]}$  が分離可能と仮定したハードマージンとは異なり, 一部の訓練データにノイズが乗るなどして分類可能ではない場合, ソフトマージンと呼ばれる拡張を考える. 式 (2.22) の制約条件を以下のように緩和する.

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, i \in [n], \xi_i \geq 0 \quad (2.23)$$

式 (2.23) より右辺が 1 以下となる. これは,  $\mathbf{x}_i$  が境界に近づくこと, さらには境界の反対側に食い込むことも許容する.

データ  $\mathbf{x}_i$  が誤分類されたとき,  $y_i(\mathbf{w}^T \mathbf{x}_i + b) < 0$  となるので, 式 (2.23) の制約条件が成立するには  $\xi_i > 1$  である必要がある. このため, 誤分類されたデータが  $K$  個の場合,  $\sum_{i \in [n]} \xi_i > K$  となり, これの値を小さくすることで誤分類を抑制できることが分かる.

そこで、SVM の最適化問題を以下に変更する。

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \quad (2.24)$$

制約条件は、式 (2.23) と同じである。なおここで  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^T$  であり、係数  $C$  は正則化係数である。この係数  $C$  は  $\xi_i$  に対する抑制なので、 $C$  が大きくなるにつれて誤分類への許容はされにくくなり、 $C = \infty$  ではソフトマージンはハードマージンに一致する。

## 2.2.2 非線形 SVM

線形 SVM では、訓練集合の実空間上に線形な境界を設けることにより訓練集合を二つのクラスに分類している。しかし、訓練集合が線形な境界では分離できない場合がある。そこで、訓練集合を超平面で分離できる空間に写像した後、線形な境界で分離することを考える。

### 2.2.2.1 双対問題

上記で定式化した式 (2.22)(2.24) は線形 SVM における主問題と呼ばれ、ここからラグランジュの未定乗数法を用いて双対問題と呼ばれる問題を導出する。主問題の目的関数と制約条件を以下のように表現し、

$$\begin{aligned} f(\mathbf{w}, \boldsymbol{\xi}) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \\ g(\mathbf{w}, b, \xi_i) &= y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i \geq 0 \\ h(\xi_i) &= \xi_i \geq 0 \\ & i \in [n] \end{aligned} \quad (2.25)$$

ラグランジュ乗数  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ 、 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$  を用いて、新しい目的関数を以下のように定義する。

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) &= f(\mathbf{w}, \boldsymbol{\xi}) - \sum_{i \in [n]} \alpha_i g(\mathbf{w}, b, \xi_i) - \sum_{i \in [n]} \mu_i h(\xi_i) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \\ &\quad - \sum_{i \in [n]} \alpha_i (y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i \in [n]} \mu_i \xi_i \end{aligned} \quad (2.26)$$

ここで、 $\mathbf{w}, b, \boldsymbol{\xi}$  は主変数、 $\boldsymbol{\alpha}, \boldsymbol{\mu}$  は双対変数であり、ラグランジュ関数を双対変数について最大化した関数を  $P(\mathbf{w}, b, \boldsymbol{\xi})$  とすると、

$$\begin{aligned} P(\mathbf{w}, b, \boldsymbol{\xi}) &= \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\mu} \geq 0} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \\ &\quad - \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\mu} \geq 0} \left\{ \sum_{i \in [n]} \alpha_i (y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i \in [n]} \mu_i \xi_i \right\} \end{aligned} \quad (2.27)$$

で表され、 $y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i < 0$  または  $\xi_i < 0$  のとき、関数  $L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu})$  は無限大に発散してしまう。一方、 $y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i \geq 0$  または  $\xi_i \geq 0$  が成り立てば、式 (2.27) 第 3 項の最大値は 0 となり、式 (2.24) の目的関数となる。したがって、制約条件を満たすとき  $P(\mathbf{w}, b, \boldsymbol{\xi})$  は最適化可能であり、 $P(\mathbf{w}, b, \boldsymbol{\xi})$  を主変数について最小化した

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} P(\mathbf{w}, b, \boldsymbol{\xi}) = \min_{\mathbf{w}, b, \boldsymbol{\xi}} \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\mu} \geq 0} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) \quad (2.28)$$

は、式 (2.24) と等価である。

次に、ラグランジュ関数  $L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu})$  の主変数についての最小化を考える。

$$D(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \min_{\mathbf{w}, b, \boldsymbol{\xi}} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) \quad (2.29)$$

さらに  $D(\boldsymbol{\alpha}, \boldsymbol{\mu})$  を双対変数について最大化する。

$$\max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\mu} \geq 0} D(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\mu} \geq 0} \min_{\mathbf{w}, b, \boldsymbol{\xi}} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\mu}) \quad (2.30)$$

これを整理すると、双対変数のみの問題として表すことができる。まず、式 (2.30) の右辺の内側の最小化を考える。 $L$  の  $\mathbf{w}$ ,  $b$ ,  $\xi_i$  の偏微分が 0 になるという条件を導出すると以下のようなになる。

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i \in [n]} \alpha_i y_i \mathbf{x}_i = 0 \quad (2.31)$$

$$\frac{\partial L}{\partial b} = - \sum_{i \in [n]} \alpha_i y_i = 0 \quad (2.32)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (2.33)$$

式 (2.31) より,  $L$  は  $\mathbf{w}$  に関して凸二次関数であり,  $\mathbf{w} = \sum_{i \in [n]} \alpha_i y_i \mathbf{x}_i$  で  $L$  は最小となる. 一方,  $L$  は  $b$  と  $\xi_i$  に関しては 1 次式なので, 制約のない最小化ではこれらの変数を動かすことで  $L$  をどこまでも小さくできてしまう. そのため,  $D(\boldsymbol{\alpha}, \boldsymbol{\mu})$  の最大化の際, 双対変数は式 (2.32), 式 (2.33) の条件を満たしている必要がある. 以上を満たすと式 (2.26) は以下のようになる.

$$\begin{aligned} L &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i \in [n]} \alpha_i y_i \mathbf{w}^\top \mathbf{x}_i - b \sum_{i \in [n]} \alpha_i y_i \\ &\quad + \sum_{i \in [n]} \alpha_i + \sum_{i \in [n]} (C - \alpha_i - \mu_i) \xi_i \\ &= -\frac{1}{2} \sum_{i, j \in [n]} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i \in [n]} \alpha_i \end{aligned} \quad (2.34)$$

式 (2.33) から,  $C - \alpha_i = \mu_i \geq 0$  であり, これらをまとめると, SVM の双対問題は以下のように表される.

$$\max_{\boldsymbol{\alpha}} \left\{ -\frac{1}{2} \sum_{i, j \in [n]} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i \in [n]} \alpha_i \right\} \quad (2.35)$$

なお制約条件は以下である.

$$\begin{aligned} \sum_{i \in [n]} \alpha_i y_i &= 0 \\ C &\geq \alpha_i \geq 0 \\ i &\in [n] \end{aligned} \quad (2.36)$$

### 2.2.2.2 カーネル

入力  $\mathbf{x}$  を別の特徴空間  $\mathcal{F}$  へと写像する関数  $\phi: \mathbb{R}^d \rightarrow \mathcal{F}$  を考える. このとき,  $\phi(\mathbf{x})$  を新たな特徴ベクトルだと考えると, 式 (2.17) は以下のように変形できる.

$$f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b \quad (2.37)$$

なお,  $\mathbf{w}$  は特徴空間  $\mathcal{F}$  の要素として定義される. この変形を式 (2.35) に適用すると, 以下のようになる.

$$\max_{\boldsymbol{\alpha}} \left\{ -\frac{1}{2} \sum_{i, j \in [n]} \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) + \sum_{i \in [n]} \alpha_i \right\} \quad (2.38)$$

制約条件は、式 (2.36) と同じである。

この双対問題において、 $\phi$  は内積の形 ( $\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ ) で現れている。したがって、 $\phi(\mathbf{x})$  を求めなくても  $\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$  さえ計算できればこの問題を解くことができる。そこで、この内積をカーネル関数として以下のように定義する。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \quad (2.39)$$

ある特定の性質を満たす関数を用いると、 $\phi(\mathbf{x})$  を計算することなく  $K(\mathbf{x}_i, \mathbf{x}_j)$  が計算できることが知られている。よく利用されるカーネル関数として RBF カーネル (Radial Basis Function Kernel) が挙げられ、以下の式で表される。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|), \quad (\gamma \in \mathbb{R}, \gamma > 0) \quad (2.40)$$

## 2.2.3 多クラス分類

2.2.2 節までで説明したのは 2 クラスの分類法である。しかし、実際には 3 クラス以上の多クラスを分類する場合があります。2 クラス SVM を多クラス SVM に拡張する必要がある。ここでは、2 クラス SVM を多クラス SVM に拡張する方法として 1 対他方式と 1 対 1 方式を説明する。

### 2.2.3.1 1 対他方式

1 対他方式 (one-versus-rest) は、あるクラスに属する  $\mathbf{x}$  と属さない  $\mathbf{x}$  を分ける分類器を各クラスについて学習する。いま、 $k$  番目のクラスに属する  $\mathbf{x}_i$  を正のクラス、それ以外の  $\mathbf{x}_i$  を負のクラスとみなして学習した 2 クラス分類の決定関数を  $f^k(\mathbf{x})$  とする。1 対他方式では、 $f^k(\mathbf{x})$  の最大値を選ぶことでクラス分類を行う。つまり、入力  $\mathbf{x}$  に対する分類器の出力  $g(\mathbf{x})$  は次の式で定義される。

$$g(\mathbf{x}) = \operatorname{argmax}_{k \in [c]} f^k(\mathbf{x}) \quad (2.41)$$

これは、各決定関数の出力である  $f^k(\mathbf{x})$  が大きいほど  $\mathbf{x}$  がクラス  $k$  に属している確信度が強いという解釈のもとに導かれた分類規則である。この方法は 2 クラス分類器をクラスの数だけ学習するだけで実装可能であり、実装が容易だという長所がある。

### 2.2.3.2 1 対 1 方式

1 対 1 方式 (one-versus-one) は、ある特定のクラスに入るか、また別の特定のクラスに入るかの 2 クラス分類問題を解く方式である。 $c$  個のクラスに対してクラス  $i$  とクラス  $j$

に属する事例のみを取り出して、その2クラスを分類する分類器を作成する。すべてのクラスのペアを考えると、このような分類器を  $c(c-1)/2$  個作成でき、ある入力  $\mathbf{x}$  をどのクラスに分類するかは  $c(c-1)/2$  個の分類器による多数決で決定する。

## 第3章

# 提案手法

本章では、自転車選手正面映像に対する自転車選手識別手法について述べる。

### 3.1 選手抽出

OpenPose による姿勢推定結果はフレームごとに json 形式で出力される。構造は以下のようにになっている。

Listing 3.1 姿勢推定結果の例

```
{
  "version":0.1 ,
  "bodies":[
    {"joints":[1114.15,160.396,0.846207,...]}
    {"joints":[...]}}
}
```

フレーム内に複数人の姿勢が検出された場合、前後のフレームで同一人物の姿勢情報を同定する必要がある。そこで、フレーム  $t$  とその前のフレーム  $t-1$  間において、検出人物の姿勢情報  $\mathbf{Y}$  の各パーツ位置座標の残差平方和 (Residual Sum of Squares:  $RSS$ ) を算出し、最も値の小さい姿勢情報を同一人物と決定する。RSS を求める関数を  $f$  と定義すると、入力  $\mathbf{Y}$  に対する同一人物を決定する関数の出力  $g(\mathbf{Y})$  は次式で表される。

$$\begin{aligned}
g(\mathbf{Y}) &= \operatorname{argmin}_{k \in K} f(\mathbf{Y}_k) \\
&= |\mathbf{Y} - \mathbf{Y}_{k,t-1}|^2 \\
&= \sum_{j=1}^J (x_{j,t} - x_{j,t-1,k}, y_{j,t} - y_{j,t-1,k})^2
\end{aligned} \tag{3.1}$$

ここで、 $K$  は検出人物数、 $J$  はパーツ数、 $(x_{j,t,k}, y_{j,t,k})$  はフレーム  $t$  における人物  $k$  の姿勢情報を表す。

## 3.2 ペダリング周期変換（時間正規化）

OpenPose の姿勢推定によって得られた自転車選手の右足首の  $y$  座標の時間経過グラフを図 3.1 に示す。なお、左足首でも同様のグラフとなる。図 3.1 より、右足首の周期的な動きからピーク（局所的最大値）を検出することでペダリング周期を求めることができる。

### 3.2.1 ピーク検出

Math Works 社が提供するソフトウェアである MATLAB[19] の `findpeaks` 関数を用いて図 3.1 のピークを検出する。ピーク間の最小距離である最小隔離距離を  $D$  としたとき、 $D$  の値によってピークの検出位置が変わる。 $D = 5, 15, 30$  としたときのピーク検出結果をそれぞれ図 3.2, 図 3.3, 図 3.4 に示す。なお、ピークとして検出された箇所は下向き三角形で表されている。図 3.2, 図 3.3, 図 3.4 に示したように、最小隔離距離  $D$  の値によってピーク検出位置が変化してしまうため、適切な  $D$  を設定する必要がある。本研究では、ある右足首座標変化について複数パターンの最小隔離距離  $D$  によりピーク位置を検出し、最適だと思われる  $D$  を目視により判断している。

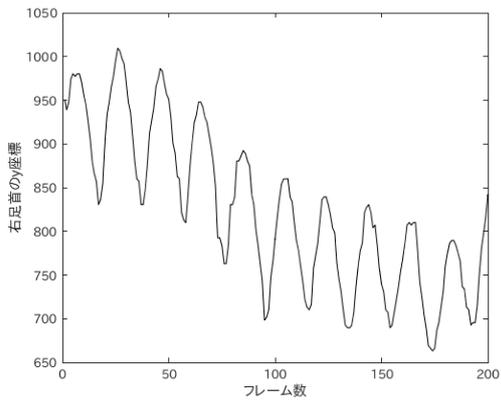


図 3.1 右足首の座標変化

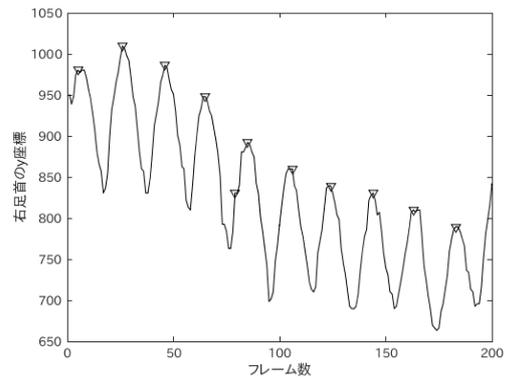


図 3.2 右足首の座標変化 ( $D = 5$ )

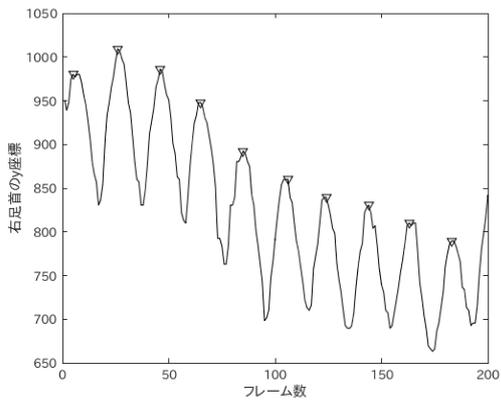


図 3.3 右足首の座標変化 ( $D = 15$ )

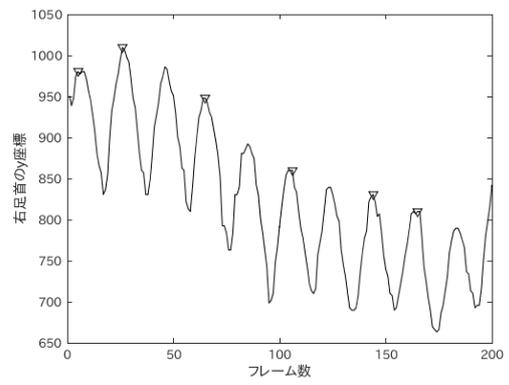


図 3.4 右足首の座標変化 ( $D = 30$ )

### 3.2.2 3次スプライン補間

図 3.3 で検出したピーク間の距離が 30 フレーム、つまり、ペダリング周期が 30 フレームで 1 回転となるように、姿勢推定により得られる姿勢情報  $\mathbf{Y}$  を補間処理する。補間処理には 3 次スプライン補間法 [20] を用いる。3 次スプライン補間法は、3 次の多項式で小区間を近似する方法であり、 $N + 1$  個データを  $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ 、区間  $[x_i, x_{i+1}]$  で補間に使う関数を  $S_i(x)$  としたとき、区間多項式  $S$  は、以下の式 (3.2) で与えられる。

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (3.2)$$

$$i = 0, 1, 2, \dots, N - 1 \quad (3.3)$$

式 (3.2) の係数  $a, b, c, d$  を求めることで 3 次スプライン補間の関数を決定できる。

$N + 1$  個のデータに対し、 $N$  個の区間多項式を求める必要がある。したがって、区間多項式の係数 ( $a, b, c, d$ ) である未知数は合計  $4N$  個あることになる。これを求めるためには、 $4N$  個の方程式が必要となる。以下の条件を満たすように  $4N$  個の方程式を作り、連立方程式を解くことで区間多項式の係数 ( $a, b, c, d$ ) を求める。

#### 条件 1

区間多項式  $S$  は全てのデータ点を通る。ゆえにそれぞれの区間多項式に対して両端での値が決まるため、 $2N$  個の方程式ができる。

#### 条件 2

それぞれの区間多項式の一次導関数は境界点において連続とする。これにより、 $N - 1$  個の方程式ができる。

#### 条件 3

それぞれの区間多項式の一次導関数は境界点において連続とする。これにより、 $N - 1$  個の方程式ができる。

#### 条件 4

両端  $x_0, x_N$  での二次導関数の値を 0 とする。すなわち、

$$S_0''(x_0) = S_{N-1}''(x_N) = 0 \quad (3.4)$$

である。

ここで、 $x = x_i$  における二次導関数の値を  $v_i$  とする。つまり、

$$v_i = S_j''(x_i), \quad (i = 0, 1, 2, \dots, N, \quad j = i - 1, i) \quad (3.5)$$

式 (3.2) から、

$$S_i''(x) = 6a_i(x - x_j) + 2b_i \quad (3.6)$$

となる。 $x = x_i$  のとき、これは式 (3.5) を用いて、

$$b_i = \frac{v_i}{2} \quad (3.7)$$

また、条件 3 より  $v_{i+1} = S_i''(x_{i+1})$  なので、

$$v_{i+1} = S_i''(x_{i+1}) = 6a_i(x_{i+1} - x_i) + 2b_i \quad (3.8)$$

したがって、式 (3.7) より、

$$\begin{aligned} a_i &= \frac{v_{i+1} - 2b_i}{6(x_{i+1} - x_i)} \\ &= \frac{v_{i+1} - v_i}{6(x_{i+1} - x_i)} \end{aligned} \quad (3.9)$$

条件 1 より、区分多項式  $S$  は全てのデータ点を通る。区分多項式の左端の  $x_i$  について、 $S_i(x_i) = y_i$  なので、式 (3.2) に代入して、

$$d_i = y_i \quad (3.10)$$

また、 $S_i(x_{i+1}) = y_{i+1}$  なので、これを式 (3.2) に代入すると、

$$a_i(x_{i+1} - x_i)^3 + b_i(x_{i+1} - x_i)^2 + c_i(x_{i+1} - x_i) + d_i = y_{i+1} \quad (3.11)$$

これに、式 (3.7) (3.9) (3.10) をこれに代入すると、

$$\begin{aligned} c_i &= \frac{1}{x_{i+1} - x_i} \{ y_{i+1} - a_i(x_{i+1} - x_i)^3 - b_i(x_{i+1} - x_i)^2 - d_i \} \\ &= \frac{1}{x_{i+1} - x_i} \left\{ y_{i+1} - \frac{v_{i+1} - v_i}{6(x_{i+1} - x_i)} (x_{i+1} - x_i)^3 - \frac{v_i}{2} (x_{i+1} - x_i)^2 - y_i \right\} \\ &= \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - \frac{1}{6} (x_{i+1} - x_i) (2v_i + v_{i+1}) \end{aligned} \quad (3.12)$$



ただし,

$$\begin{aligned} h_i &= x_{i+1} - x_i \\ w_i &= 6 \left( \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right) \end{aligned} \quad (3.18)$$

式 (3.17) を解いて各データ点での二次導関数の値  $v_i$  を求め, 式 (3.7)(3.9)(3.10)(3.12) に代入して  $a_i, b_i, c_i, d_i$  を求める. 区分局多項式  $S$  は各パーツについてそれぞれ求めるので, 最終的に区分局多項式群  $\mathbf{S} = (S^1, \dots, S^J)$  を求める.

図 3.3 で検出したピークの存在するフレームを  $T_1, T_2$ , 周期変換後のフレーム  $l$  における姿勢座標位置を  $\mathbf{Y}_l (1 \leq l \leq 30)$  としたとき,  $\mathbf{Y}_l = (\mathbf{Y}_{1l}, \dots, \mathbf{Y}_{Jl})$  であり, このとき  $\mathbf{Y}_{jl}$  は 3 次スプライン補間により求めた区分局多項式群  $\mathbf{S}$  を用いて, 以下のように表される.

$$\begin{aligned} \mathbf{Y}_{jl} &= (x_{jl}, y_{jl}) \\ x_{jl} &= \frac{T_2 - T_1}{30} l + T_1 \\ y_{jl} &= S^j \left( \frac{T_2 - T_1}{30} l + T_1 \right) \\ j &\in 1, \dots, J \end{aligned} \quad (3.19)$$

### 3.3 位置補正

本研究で用いるレース映像は, 選手の前方を走るバイクから撮影された正面映像である. そのため, カメラ位置と選手の位置関係は常に変化している. したがって, 映像中の選手位置も常に変化する. OpenPose によって得られる姿勢情報は画像中における姿勢のパーツ位置情報 ( $x, y$  座標) であるため, これらの位置を補正する必要がある. そこで, 対象人物の首の位置が原点となるように位置を補正する. 補正関数を  $f$ , 首の座標を  $\mathbf{Y}_{neck} = (x_{neck}, y_{neck})$  とすると, 各パーツ位置座標は次のように補正される.

$$f(\mathbf{Y}) = \mathbf{Y}_j - \mathbf{Y}_{neck} = (x_j - x_{neck}, y_j - y_{neck}) \quad (3.20)$$

ここで,

$$\begin{aligned} \mathbf{Y} &= (\mathbf{Y}_1, \dots, \mathbf{Y}_J) \\ \mathbf{Y}_j &\in Z \subset \mathbb{R} \\ j &\in 1, \dots, J \end{aligned} \quad (3.21)$$

である.

### 3.4 スケール変換

第 3.3 節で述べた理由と同様に、映像中の選手の大きさも常に変化する。本研究では、自転車選手の正面からの映像のみを対象としている。そのため、カメラと選手間の距離と選手の両肩間の距離は比例し、左右の肩から撮影カメラまでの距離に差がない。そこで、選手の両肩間の距離が常に一定になるように姿勢情報をスケール変換する。

変換関数を  $f$ 、右肩座標位置を  $\mathbf{Y}_{rightsholder}$ 、左肩座標位置を  $\mathbf{Y}_{leftsholder}$  とすると、姿勢情報  $\mathbf{Y}$  に対して両肩間の距離を 1 とした場合の各パーツ位置座標は次のように表される。

$$\begin{aligned} f(\mathbf{Y}) &= \frac{\mathbf{Y}}{|\mathbf{Y}_{rightsholder} - \mathbf{Y}_{leftsholder}|} \\ &= \frac{\mathbf{Y}}{|x_{rightsholder} - x_{leftsholder}, y_{rightsholder} - y_{leftsholder}|} \end{aligned} \quad (3.22)$$

### 3.5 異常値検出

OpenPose によって得られた姿勢情報  $\mathbf{Y}$  には、図 3.5 に示すように姿勢推定が失敗する場合がある。特に、複数の人物が重なっている画像に対して推定を失敗するケースが多い。このような異常値が SVM の訓練データに含まれると識別精度に影響を与えてしまう。そこで、これらの姿勢推定の失敗を異常値として検出し、除去することを考える。

スケール変換後の姿勢情報について、フレームごとの各パーツ間の長さ（右手首 - 右肘など）を算出し、この結果をもとに異常値の検出を行う。図 3.7 - 図 3.15 にフレームごとの各パーツ間の長さの図を示す。また、パーツ間長さのヒストグラムを図 3.16 - 図 3.24 に示す。図 3.7 - 図 3.15 より、各パーツ間の長さは概ね一定の値の範囲に収まっていることがわかる。また図 3.16 - 図 3.24 から、データの分布は正規分布に従うと仮定する。ここで、フレームごとの各パーツ間の長さについて、標準正規分布で平均値から標準偏差の 2 倍以上離れた値を異常値として検出し、異常値を含むフレームを異常フレームと定義する。ペダリング 1 回（30 フレーム）中に含まれる異常フレーム数に応じて異常値レベルを設定する。1 サイクルに含まれる異常フレーム数と全データにおける異常サイクルの割合を表 3.1 に示す。



図 3.5 姿勢推定失敗例

表 3.1 異常値レベルごとの異常フレーム数

| 異常値レベル | 1 サイクル中の異常フレーム数 | 異常サイクル率 |
|--------|-----------------|---------|
| レベル 1  | 20-30           | 6.5%    |
| レベル 2  | 15-30           | 11.8%   |
| レベル 3  | 10-30           | 20.8%   |

### 3.6 複数サイクル多数決識別法

上記までの節では自転車選手の 1 サイクルごとの分類手法を述べた。ここでは複数サイクルを考慮した分類法について述べる。複数サイクル多数決分類法の概要を図 3.6 に示す。この手法では、入力データであるサイクル 1 からサイクル  $t$  までの各サイクルの SVM による予測結果から、最も多く予測された選手を予測選手と決定する。ただし、最も多く予測された選手が複数人いた場合、そのうち最近予測された選手を予測選手と決定する。

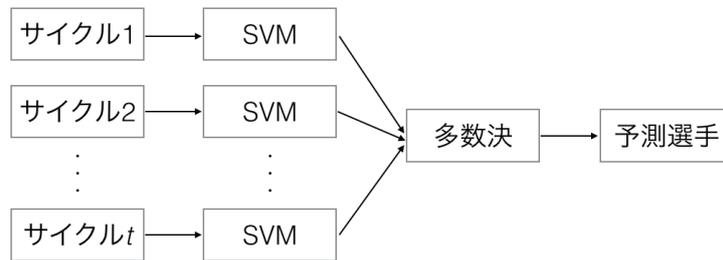


図 3.6 複数サイクル多数決識別法

### 3.7 むすび

本章では、自転車選手正面映像に対する自転車選手識別手法について述べた。OpenPoseにより得られた姿勢情報について、首位置を原点とした位置補正、選手両肩間の距離が一定となるようにスケール変換する手法、右足首に注目したピーク検出および3次スプライン補間によるペダリング周期変換、姿勢推定の失敗をデータから除去する異常値検出手法について述べた。さらに、SVMによる識別結果を複数サイクルについて考慮する複数サイクル多数決分類法について述べた。

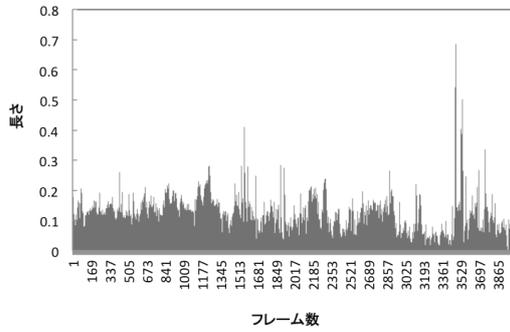


図 3.7 パーツ間の長さ (首)

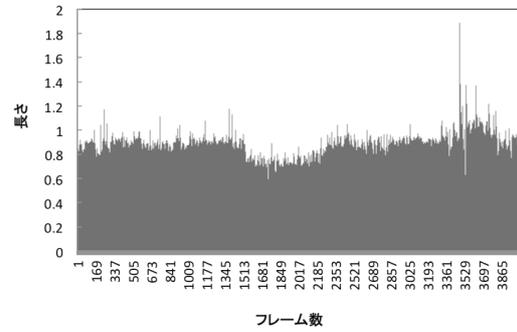


図 3.8 パーツ間の長さ (右上腕)

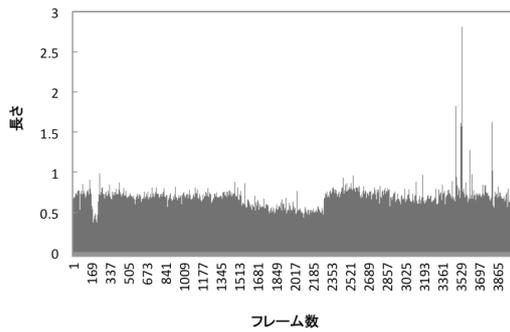


図 3.9 パーツ間の長さ (右前腕)

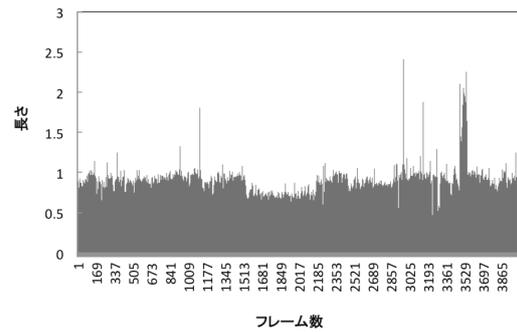


図 3.10 パーツ間の長さ (左上腕)

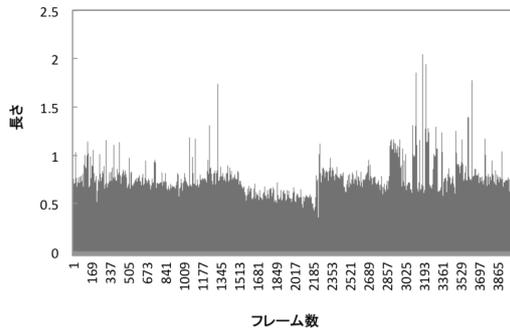


図 3.11 パーツ間の長さ (左前腕)

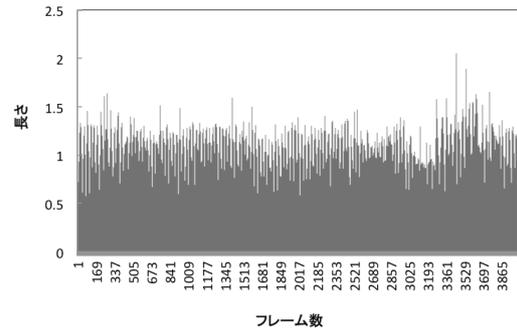


図 3.12 パーツ間の長さ (右大腿)

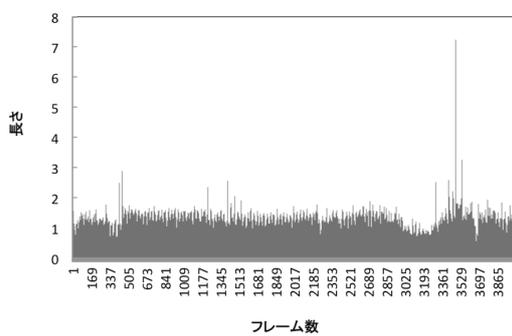


図 3.13 パーツ間の長さ (右下腿)

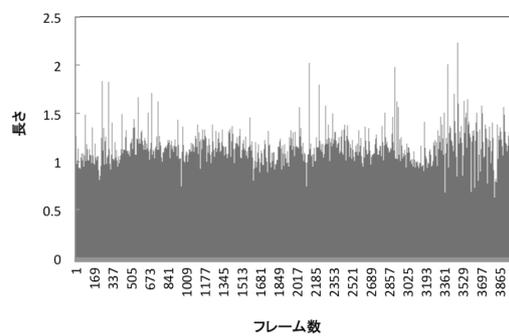


図 3.14 パーツ間の長さ (左大腿)

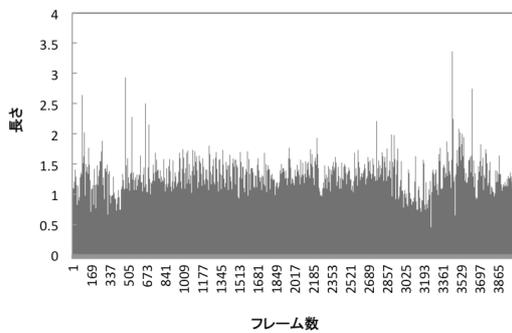


図 3.15 パーツ間の長さ (左下腿)

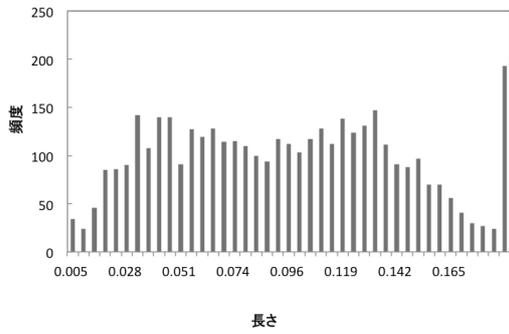


図 3.16 パーツ間長さのヒストグラム (首)

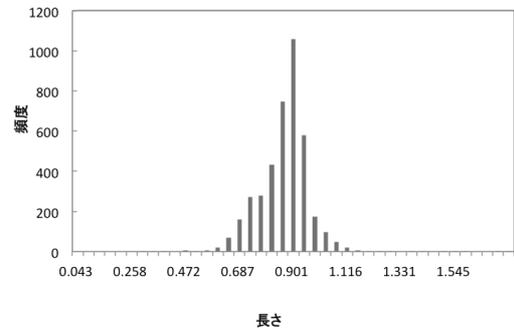


図 3.17 パーツ間長さのヒストグラム (右上腕)

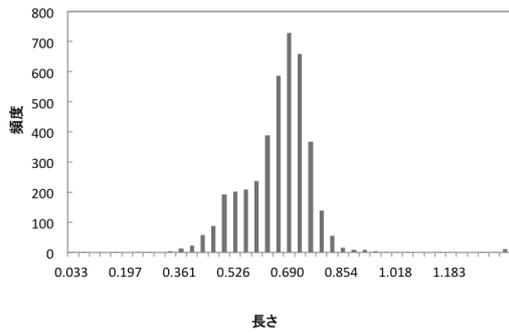


図 3.18 パーツ間長さのヒストグラム (右前腕)

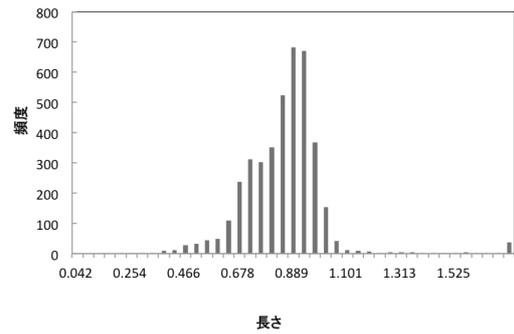


図 3.19 パーツ間長さのヒストグラム (左上腕)

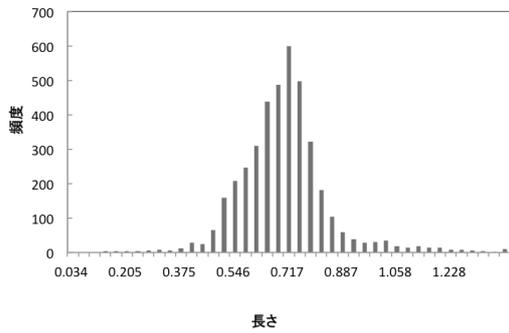


図 3.20 パーツ間長さのヒストグラム (左前腕)

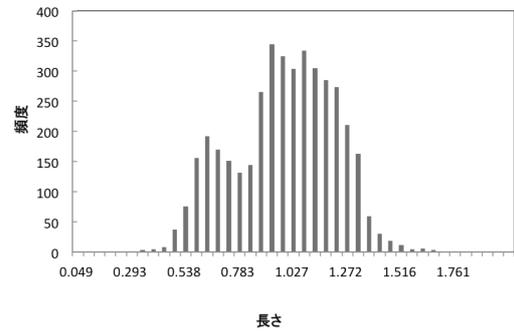


図 3.21 パーツ間長さのヒストグラム (右大腿)

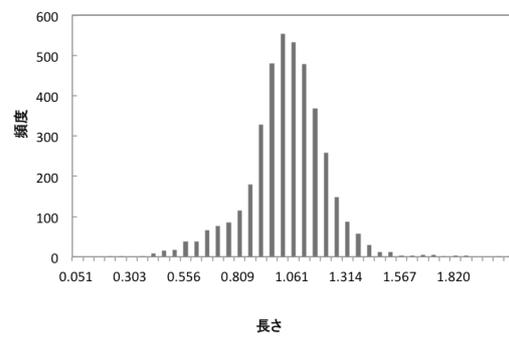
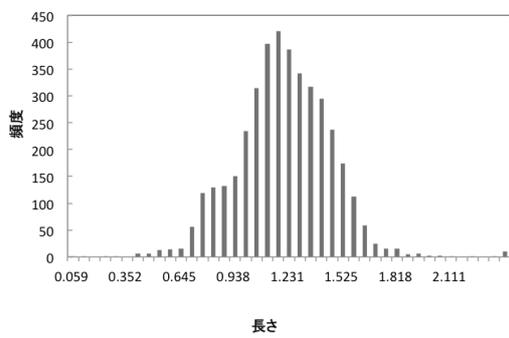


図 3.22 パーツ間長さのヒストグラム (右下腿) 図 3.23 パーツ間長さのヒストグラム (左大腿)

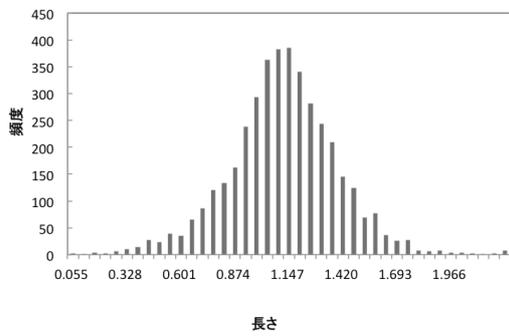


図 3.24 パーツ間長さのヒストグラム (左下腿)

## 第 4 章

# 評価実験

本章では CNN を用いたチーム識別実験および第 3 章で提案した手法の評価実験とその結果を述べる。

### 4.1 チーム分類

#### 4.1.1 データセットの生成

一般物体認識手法の一つである YOLO(You Only Look Once)[21] を用いて実際のレース映像 ( $1920 \times 1080$ [*pel*],  $30$  *fps*) から、自転車選手画像を収集した。YOLO によって 80% 以上の確率で「person」だと判定された画像領域を切り出し、ツール・ド・フランス 2017 に出場した 22 チームの選手について評価用画像 1100 枚、学習用画像 2200 枚の合計 3300 枚の画像を収集した。学習用画像についてはデータオーギュメンテーションを行い、画像の枚数を 13200 枚に増やした。なお、データオーギュメンテーションでは、学習用画像に左右反転、輝度変化を施している。ツール・ド・フランス 2017 出場チームとチームごとの学習用画像と評価用画像の枚数を表 4.1 に示す。

#### 4.1.2 畳み込みニューラルネットワークの学習

本実験は、深層学習のフレームワークである Chainer[22] を用いて、既存の CNN モデルである GoogleNet[23] を使用した。バッチサイズは 20 とし、学習時の精度推移を図 4.1 に示す。図 4.1 より、早い段階で収束していることがわかる。

表 4.1 チームごとの学習用画像と評価用画像

| チーム                                | 学習用   | 評価用  | 合計    |
|------------------------------------|-------|------|-------|
| TEAM SKY                           | 600   | 50   | 650   |
| AG2R LA MONDIALE                   | 600   | 50   | 650   |
| MOVISTAR TEAM                      | 600   | 50   | 650   |
| TREK - SEGAFREDO                   | 600   | 50   | 650   |
| BMC RACING TEAM                    | 600   | 50   | 650   |
| ASTANA PRO TEAM                    | 600   | 50   | 650   |
| UAE Team Emirates                  | 600   | 50   | 650   |
| FDJ                                | 600   | 50   | 650   |
| ORICA - SCOTT                      | 600   | 50   | 650   |
| TEAM DIMENSION DATA                | 600   | 50   | 650   |
| QUICK - STEP FLOORS                | 600   | 50   | 650   |
| BORA - HANSGROHE                   | 600   | 50   | 650   |
| TEAM KATUSHA ALPECIN               | 600   | 50   | 650   |
| LOTTO-SOUDAL                       | 600   | 50   | 650   |
| TEAM SUNWEB                        | 600   | 50   | 650   |
| COFIDIS, SOLUTIONS CREDITS         | 600   | 50   | 650   |
| TEAM LOTTO NL - JUMBO              | 600   | 50   | 650   |
| DIRECT ENERGIE                     | 600   | 50   | 650   |
| CANNONDALE DRAPAC PRO CYCLING TEAM | 600   | 50   | 650   |
| BAHRAIN - MERIDA                   | 600   | 50   | 650   |
| WANTY-GROUPE GOBERT                | 600   | 50   | 650   |
| FORTUNEO - VITAL CONCEPT           | 600   | 50   | 650   |
| 合計                                 | 13200 | 1100 | 14300 |

### 4.1.3 評価実験

エポック数を 50 としたときの評価画像の識別実験結果を表 4.3 に示す。なお評価には Accuracy, Precision, Recall を用いる。表 4.2 に示す値を用いたとき, Accuracy, Precision, Recall はそれぞれ以下の式で与えられる。

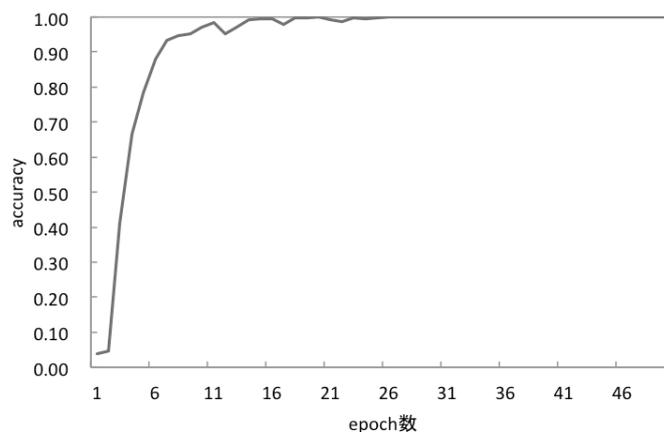


図 4.1 学習時の精度推移

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

表 4.2 Confusion Matrix

|                 |          | Actual Class        |                     |
|-----------------|----------|---------------------|---------------------|
|                 |          | Positive            | Negative            |
| Predicted Class | Positive | $TP(TruePositive)$  | $FP(FalsePositive)$ |
|                 | Negative | $FN(FalseNegative)$ | $TN(TrueNegative)$  |

表 4.3 チーム分類結果

| Accuracy | Precision | Recall |
|----------|-----------|--------|
| 99.80%   | 97.87%    | 97.82% |

表 4.3 より, Accuracy, Precision, Recall はいずれの値も 97% 以上と, 非常に高い結果となった. これは, 各チームのユニフォームの色や模様など特徴点となり得る箇所が多かったからであると考えられる.

この結果から、自転車ロードレース映像から抽出してきた選手がどのチームに属するかは高精度で識別できることが分かる。したがって、自転車選手識別はチーム内での選手識別の精度が重要となる。

## 4.2 選手識別

第3章で提案した手法について、実際の自転車ロードレースの映像データ (1920 \* 1080[*pel*], 30[*fps*]) を用いて SVM による選手の識別実験を行った。本研究で使用する SVM のカーネル関数は RBF カーネルを用い、RBF カーネルのパラメータ  $\gamma$  およびソフトマージンのコストパラメータ  $C$  の値の取り得る範囲はそれぞれ  $\gamma = \{0.01, 0.001, 0.0001\}$ ,  $C = \{0.1, 0.5, 1.0, 5.0, 10.0, 50.0, 100.0\}$  とし、SVM での多クラス分類には 1 対 1 方式を使用する。また、被験者数は 4–9 名としており、1 人あたり 6 シーンの映像 (合計 60 秒程度) を使用する。

### 4.2.1 ペダリング周期を考慮しない場合

まず、ペダリング周期を考慮しないで識別することを考える。OpenPose による姿勢推定結果  $\mathbf{Y}$  の 28 次元 (検出項目全 18 パーツのうち両目, 両耳を除いた 14 部位  $\times x, y$  座標) を 1 つの特徴ベクトルとして SVM で識別する。なお、得られた姿勢情報  $\mathbf{Y}$  に対しては、スケール変換, 位置補正の処理を行う。用いた訓練用・テスト用データ数を表 4.4 に示す。また、SVM による自転車選手識別結果を表 4.5 に示す。

表 4.4 訓練・テストデータ数

|     | 訓練データ | テストデータ |
|-----|-------|--------|
| 4 人 | 4000  | 2000   |
| 5 人 | 5000  | 2500   |
| 6 人 | 6000  | 3000   |
| 7 人 | 7000  | 3500   |
| 8 人 | 8000  | 4000   |
| 9 人 | 9000  | 4500   |

表 4.5 より、4 人での選手識別では 6 割程度の精度が出るものの、人数が増えるに従い精度が落ちていき、9 人の選手識別の精度は 4 割に満たない結果となっている。ここではパーツ位置座標 (位置関係) のみを学習しているため、選手の身体的特徴 (パーツ間の長さ

表 4.5 ペダリング周期を考慮しない場合の識別結果

|     | 正解率 [%] |
|-----|---------|
| 4 人 | 63.50   |
| 5 人 | 56.92   |
| 6 人 | 49.97   |
| 7 人 | 46.46   |
| 8 人 | 45.33   |
| 9 人 | 39.02   |

等) のみによる識別となっており、ペダリング周期を考慮しない場合では識別が困難であることが分かる。

#### 4.2.2 ペダリング周期を考慮した場合

次に、ペダリング周期を考慮した場合を考える。ペダリング 1 回 (30 フレーム) 分の姿勢情報を 1 つの特徴ベクトル (姿勢推定結果  $\mathbf{Y}$  の 28 次元  $\times$  30 フレーム = 840 次元) とし、SVM によって識別を行った。30 フレーム分のデータを 1 つの特徴ベクトルとしていることから、ペダリング周期を考慮しなかった場合に比べて訓練データ及び学習データ数が少なくなってしまう。そこで、訓練データと学習データを  $k$  人の各選手につき 2 パターン用意し、用意した全パターン ( $2^k$  パターン) についての SVM による選手識別の正解率の平均値を算出する。

まず、SVM による識別の前処理であるペダリング周期変換 (時間正規化) とスケール変換・位置補正の処理手順を考える。処理手順を 1. ペダリング周期変換, 2. スケール変換・位置補正とした場合, 順序を逆にし, 1. スケール変換・位置補正, 2. ペダリング周期変換とした場合について比較する。被験者数を 4 人としたとき, それぞれの処理手順で生成したデータにより選手識別を行った場合の結果を表 4.6 に示す。

表 4.6 処理手順比較

|                              | 正解率 [%] |
|------------------------------|---------|
| 1. ペダリング周期変換, 2. スケール変換・位置補正 | 68.3    |
| 1. スケール変換・位置補正, 2. ペダリング周期変換 | 71.4    |

表 4.6 より, 1. ペダリング周期変換, 2. スケール変換・位置補正よりも, 1. スケール変

換・位置補正, 2. ペダリング周期変換の処理手順の方が良い結果が得られた。これは, 先にペダリング周期変換 (時間正規化) をした場合, 位置とスケールの 2 つの変数に補間処理が必要であるのに対し, 先にスケール変換・位置補正をする場合は時間方向への補間処理が必要ないため, よりよい実験データの生成ができるからと考えられる。

#### 4.2.2.1 異常サイクルの除去

次に異常値検出による訓練データからの異常値サイクルの除去を考える。第 3.5 節で述べた異常値検出法によって訓練データから異常値を除去する。被験者 4-9 人に対して各異常値レベルを適応し, 異常サイクルを除去して SVM で識別した結果を表 4.7 に, またそのグラフを図 4.2 に示す。また, 各人数での異常サイクル除去前の訓練データ数およびテストデータ数を表 4.8 に示す。ここで, レベル 0 とは異常値サイクルの除去を行わなかった場合を表す。

表 4.7 各異常値レベルでの SVM による分類結果

| 異常値レベル | 4人     | 5人     | 6人     | 7人     | 8人     | 9人     |
|--------|--------|--------|--------|--------|--------|--------|
| レベル 0  | 82.10% | 82.95% | 81.80% | 78.24% | 72.97% | 66.02% |
| レベル 1  | 80.02% | 83.85% | 82.17% | 81.51% | 73.07% | 64.06% |
| レベル 2  | 81.70% | 84.12% | 83.67% | 80.37% | 74.70% | 67.64% |
| レベル 3  | 81.71% | 85.58% | 83.89% | 80.28% | 76.08% | 69.17% |

表 4.8 訓練・テストデータ数

|    | 訓練データ | テストデータ |
|----|-------|--------|
| 4人 | 240   | 60     |
| 5人 | 300   | 75     |
| 6人 | 360   | 90     |
| 7人 | 420   | 105    |
| 8人 | 480   | 120    |
| 9人 | 540   | 135    |

表 4.8 に示した通り, ペダリング周期を考慮しない場合の実験データ数 (表 4.4) と比べてペダリング周期を考慮した場合のデータ数はかなり少ないものの, 異常値検出をしない場合で 4 人で 82.10%, 9 人で 66.02% と, 高い精度が出た。また, 図 4.2 より, 除去する

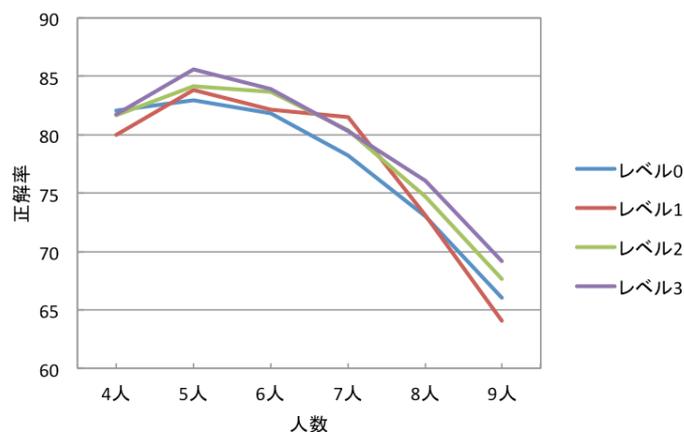


図 4.2 異常値レベルと正解率

異常値サイクルを多く除去した場合 (レベル 3) で最も識別精度が高くなることから、異常値サイクルの除去が識別精度に寄与することを確認した。

#### 4.2.2.2 一部データ除去

図 3.16 – 図 3.24 より、首の長さのばらつきが他の関節間長さに比べて非常に大きいことがわかる。これは、選手が下や横を向く動きにより、映像から検出される鼻の位置が大きく変わることに由来するものである。その例を図 4.3 に示す。図 4.3 では鼻、首の位置がそれぞれピンク、オレンジの点で表されており、その他のパーツ位置は図 4.3a, 図 4.3b でほとんど変わらないのにも関わらず、顔の向きにより鼻の位置座標が大きく変化しているのがわかる。

第 3.3 節で述べた通り、首の位置を原点にスケール変換を行っていることから、図 3.16 の結果は鼻の検出位置のばらつきが大きいということを意味している。従って、このばらつきが精度に影響を与えていると考えられる。この情報を取り除いて SVM による識別を行った場合による結果を表 4.9 に示す。表 4.9 より、鼻データを除去した場合の精度が落ちていることがわかる。このことから、鼻の位置座標のばらつきは SVM による識別精度に悪影響は与えていないと考えられる。

#### 4.2.2.3 半身データ

レース映像中、選手は風よけのために前後に連なって走行する。そのため、選手前方からの映像では、先頭にいる選手以外は体の一部が他の選手の影に隠れてしまうケースがある。これらのケースを考慮し、姿勢推定が体の一部分だけしかできていない場合での分類



(a) 正面向き

(b) 下向き

図 4.3 鼻位置の変化の様子

表 4.9 鼻位置データ除去

|     | 4人     | 5人     | 6人     | 7人     | 8人     | 9人     |
|-----|--------|--------|--------|--------|--------|--------|
| 除去前 | 82.10% | 82.95% | 81.80% | 78.24% | 72.97% | 66.02% |
| 除去後 | 75.50% | 72.08% | 75.70% | 74.39% | 67.90% | 64.06% |

を行う。具体的には、上半身、下半身、右半身、左半身のみデータ（図 4.4）を使用し、SVM で分類を行った場合の結果を表 4.10 に示す。



図 4.4 半身データ

表 4.10 より、全身データを用いた場合に比べ、いずれの半身データによる分類結果も精度が悪くなっていることがわかる。しかし、その差は  $-5\%$  程度にとどまり、下半身データでは  $-0.3\%$  と、ほぼ同程度の性能があることがわかる。

表 4.10 半身データを用いた SVM による分類結果

|     | 正解率 [%] |
|-----|---------|
| 全身  | 71.4    |
| 上半身 | 67.9    |
| 下半身 | 71.1    |
| 左半身 | 66.4    |
| 右半身 | 69.5    |

### 4.2.3 複数サイクル多数決識別法

被験者 4 – 9 人について，サイクル  $t(1 \leq t \leq 12)$  までの各サイクルの識別結果を考慮し，最終的な選手を予測した場合の結果を図 4.5 に示す．4, 5 人の場合は 7 サイクル分のデータ，6 – 9 人の場合は 10 サイクル分のデータを考慮することでその精度が 9 割程度まで向上することがわかる．

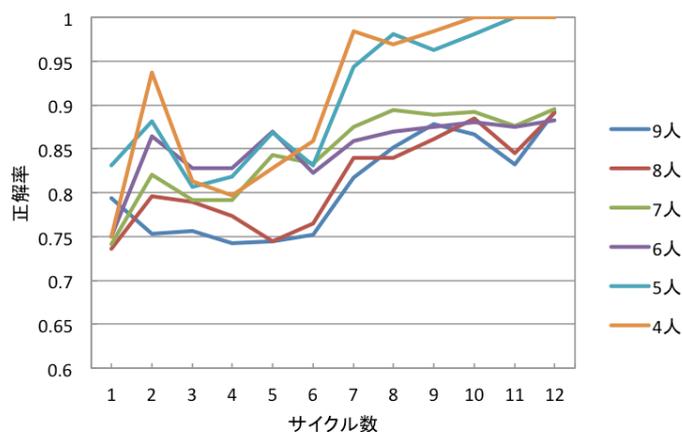


図 4.5 複数サイクル多数決識別法

## 4.3 むすび

本章では，CNN を用いたチーム識別実験および第 3 章で提案した手法の評価実験とその結果を述べた．CNN を用いたチーム識別実験では，ツールドフランス 2017 出場チーム 22 チームについて accuracy で 99.8% と，非常に高い精度で識別可能であることを確

認した。また、OpenPose によって得られた姿勢情報に対し、スケール変換・位置補正、ペダリング周期変換（時間正規化）および異常値検出処理をし、SVM で学習し識別することで 4 人の場合で 82.10%、9 人の場合で 69.17% の精度が出ることを確認した。また、複数サイクル多数決識別法により、1 サイクルだけではなく複数サイクルでの SVM 分類結果を考慮することで、4、5 人の場合で 6 サイクル、6 ～ 9 人の場合で 10 サイクルで 90% 程度の精度で選手識別が可能であることを確認した。

## 第5章

# 結論

本研究では、自転車選手の正面映像に対して OpenPose を用いて姿勢推定を行い、得られた姿勢情報から自転車選手を識別する手法を提案した。

CNN を用いた 22 チームの識別実験では、数十回程度の学習で Accuracy で 99.8% の精度が出ることを確認し、非常に高い精度でチーム識別可能であることがわかった。また、OpenPose によって得られた姿勢情報に対し、スケール変換・位置補正、ペダリング周期変換（時間正規化）および異常値検出処理をし、SVM で学習し識別することで 4 人の場合で 82.10%、9 人の場合で 69.17% の精度が出ることを確認した。また、複数サイクル多数決識別法により、1 サイクルだけではなく複数サイクルでの SVM による識別結果を考慮することで、4、5 人の場合で 6 サイクル、6–9 人の場合で 10 サイクルで 90% 程度の精度で選手識別が可能であることを確認した。

# 謝辞

本研究を遂行するにあたり、素晴らしい研究環境を与えてくださり、また貴重な時間割いてご指導と助言をいただきました渡辺裕教授、そして、研究に関する相談に乗って頂き、様々な助言をいただきました石川孝明さんに心から感謝いたします。

また、これまでの研究生生活を支えて下さった渡辺研究室の皆様、友人の皆様に感謝の意を表します。

## 参考文献

- [1] SMT : formerly Sportvision Inc. Thechnology : Pitchf/x.  
<http://www.sportvision.com/baseball/pitchfx>.
- [2] Hawk-Eye Innovations Ltd. Hawk-eye. <https://www.hawkeyeinnovations.com>.
- [3] Tour of Japan ORGANIZING COMMITTEE. ロードレース q & a.  
<http://www.toj.co.jp/roadrace>.
- [4] T.Abonen, A.Hadid, and M.Pictikainen. Face recognition with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, pp. 2037–2041, 2006.
- [5] Y.Sun, L.Ding, X.Wang, and X.Tang. Deepid3: Face recognition with very deep neural networks. *ArXiv e-prints*, feb 2015.
- [6] 上野将義, 南保英孝, 木村春彦, 上田芳弘. Ocr スコアを利用した情景画像内の文字列抽出. *科学・技術研究*, Vol. 5, No. 1, pp. 53–58, 2016.
- [7] 東山侑真, 榎原所靖, 西野恒, 八木康史. 様々な歩行状況下における歩容認証手法の性能評価. *研究報告コンピュータビジョンとイメージメディア (CVIM)*, 2013-CVIM-187, Vol. 10, pp. 1–8, 2013.
- [8] 武村紀子, 白神康平, 榎原靖, 村松大吾, 越後富夫, 八木康史. 畳み込みニューラルネットワークを用いた視点変化に頑健な歩容認証. *電子情報通信学会論文誌 A*, Vol. J99-A, No. 10, pp. 440–451, 2016.
- [9] 井元大輔, 黒沢健至, 土屋兼一, 黒木健郎, 秋葉教充, 角田英俊. 身体部位の特徴点と形状情報に基づくモデルベース歩容認証の検討. *精密工学会誌*, Vol. 83, No. 1, 2017.
- [10] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016.
- [11] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolu-

- tional pose machines. *CoRR*, Vol. abs/1602.00134, , 2016.
- [12] Adrian Bulat and Georgios Tzimiropoulos. Human pose estimation via convolutional part heatmap regression. *CoRR*, Vol. abs/1609.01743, , 2016.
- [13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, Vol. abs/1409.1556, , 2014.
- [14] Christopher M. Bishop. Pattern recognition and machine learning. 科学・技術研究, 2016.
- [15] 竹内一郎, 鳥山昌幸. サポートベクトルマシン. 機械学習プロフェッショナルシリーズ. 講談社サイエンティフィク, 2015.
- [16] T. Habib, J. Inglada, G. Mercier, and J. Chanussot. Speeding up support vector machine (svm) image classification by a kernel series expansion. In *2008 15th IEEE International Conference on Image Processing*, pp. 865–868, Oct 2008.
- [17] M. Gupta, S. S. Bharti, and S. Agarwal. Support vector machine based gender identification using voiced speech frames. In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 737–741, Dec 2016.
- [18] J. Savigny and A. Purwarianti. Emotion classification on youtube comments using word embedding. In *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, pp. 1–5, Aug 2017.
- [19] Inc. The MathWorks. Matlab. <https://jp.mathworks.com/products/matlab.html>.
- [20] 松本英敏. 3次スプライン補間法. <http://www.civil.kumamoto-u.ac.jp/matsu/spline.pdf>.
- [21] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, Vol. abs/1506.02640, , 2015.
- [22] Seiya Tokui, Kenta Oono, Shohei Hido, and Justin Clayton. Chainer: a next-generation open source framework for deep learning. *In Proc. Neural Information Processing Systems*, 2015.
- [23] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, Vol. abs/1409.4842, , 2014.

# 研究業績

- [1] 高木, 渡辺, 河村, 内藤: “レンズ歪み補正付き動き補償方式の検討“, 電子情報通信学会総合大会, D-11-53, Mar. 2016.
- [2] 高木, 渡辺, 河村, 内藤: “レンズ歪ブロックを用いた動き補償方式の検討“, FIT2016(第15回科学技術フォーラム), No.I-033, Sep. 2016.
- [3] 高木, 渡辺, 河村, 内藤: “レンズ歪補正動き補償方式に関する一検討“, PCSJ/IMPS 2016, No. P-3-19, Nov. 2016.
- [4] 高木, 石川, 渡辺: “姿勢情報を用いた自転車選手識別の基礎検討“, 映像情報メディア学会年次大会, 14D-4, Sep. 2017.
- [5] 高木, 石川, 渡辺: ”姿勢情報を用いた自転車選手識別手法に関する一検討“, 映像情報メディア学会メディア工学研究会 (スポーツ情報処理時限研究会合同開催), ME2017-121, pp.25-27, Nov. 2017.
- [6] 高木, 石川, 渡辺: ”関節位置情報を用いた自転車選手識別手法に関する一検討“, 映像情報メディア学会冬季大会, 25C-5, Dec. 2017.