6Q-05

部位の重なりに応じた異なる検出器による受講者画像からの姿勢推定

相澤 将吾[1]　椋木 雅之[2]　美濃 導彦[2]　三功 浩嗣[3]

京都大学情報学研究科[1]　京都大学学術情報メディアセンター[2]　株式会社 KDDI 研究所[3]

1. はじめに

講義室に設置したカメラから得られた受講者映像の振る舞いを分析することで，講義改善を行おうとする取り組みが行われている[1]．振る舞いとは，人物の姿勢を時系列順に並べたものから意味のある区間を抽出したものである．また姿勢とは，ある瞬間での部位位置の集合であり，画像中の関節位置によって表現される．注目すべき振る舞いについては研究ごとに異なり，明確な指標が存在しない．そこで本稿では，振る舞いの構成要素である姿勢を獲得することを考える．

人物姿勢推定を妨げる要因として，人間の取りうる姿勢の多様性と，衣服や背景の変化による人物の外観の多様性が挙げられる[2]．これらの要因によって，画像中での人物とその周辺は非常に大きな多様性をもち，姿勢推定の妨げとなる．多様性を制限した状況では，比較的高い性能を発揮する検出器を作成することが可能だと考えられる．ここで検出器は，画像中の関節位置の検出を行う．特定の状況に特化した検出器を複数作成し，推定対象となる画像に対してはそのいずれかに適切に振り分ける分類器を挟むことで，全体として高性能の検出器を構成することができる．この手法では，どのように多様性を制限するのか，どのように画像を分類するのか，の 2 点が大きな問題となる．

本稿では，講義室に設置したカメラから得られた受講者画像の姿勢推定を行う．多数の受講者画像から得られた知見に基づいて 3 つのクラスを定義し，それぞれに適した検出器を作成することで推定精度の向上を図る．

2. 分類器を利用した姿勢推定

本手法は，検出器の学習と，それを用いたテストデータの姿勢推定の 2 つのステップから構成される．具体的には以下の手順となる．

学習ステップの概略を図 1 に示す．まず学習データ集合に対してクラス分類器による分類を行い，学習データ集合をクラスごとの小集合に分ける．次に小集合ごとに異なる検出器(以下，各クラス用部位検出器)を用意し，検出器に対応する小集合を学習させる．これにより，各クラスに対応した検出器を得ることが出来る．

推定ステップの概略を図 2 に示す．まず受講者画像をクラス分類器に入力する．次に分類結果に基づいて，対応する各クラス用部位検出器に受講者画像を入力し，推定姿勢データを得る．

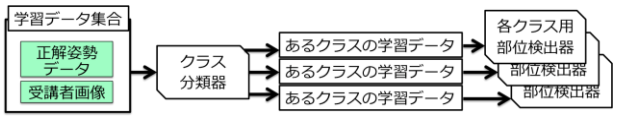

図 1：学習ステップの概略図

図 2：推定ステップの概略図

3. クラス分類器

受講者の部位の隠れに対応して，「頭が首や前腕部を隠す状態(クラス 1)」「胴体の前で一方の前腕部が他方の前腕部を隠す(クラス 2)」「その他(クラス 3)」という 3 つのクラスを定義する．クラス 1，2 は姿勢推定が難しいものの，姿勢の多様性は小さい．クラス 1，2 を分離して扱うことで多様性を制限することが可能となり，これらに特化した性能の高い検出器を作成することができる．クラス 3 については，姿勢推定が難しいクラス 1，2 と分離することで性能の向上が見込まれる．

このようにクラス分類を行う処理を図 3 に示す．受講者画像に対して，まず顔検出を行う．顔が検出されなかった画像をクラス 1 とし，顔が検出された画像では，画像中の受講者の胴体周辺における横方向エッジの強さを求める．横方向エッジが十分強い場合にはクラス 2 とし，そうでない場合にはクラス 3 とする．

Human Pose Estimation from Students' Image with Different Detector Depending on Self Occlusion

1 Aizawa Shogo, Graduate School of Informatics, Kyoto University

2 Mukunoki Masayuki, Minoh Michihiko, Academic Center for Computing and Media Studies, Kyoto University

3 Sankoh Hiroshi, KDDI R&D Laboratories

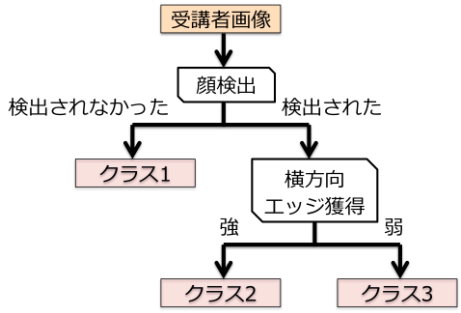

図 3：クラス分類器

4. 実験

4.1. 実験環境

提案手法と，クラス分類器を用いず単一の部位検出器で全てのクラスの学習データを学習する手法(以下，従来手法)を比較する．部位検出器には Yi Yang ら[3]の手法を用いる．Yi Yang らの手法では上半身が映った人物画像が入力されると，上半身の各部の座標値を求める．このうち頭頂部，首，両肩，両肘，両手首の座標値を対象とする．受講者姿勢は振る舞い分析に利用されるため，正解姿勢から大きく外れた姿勢が推定されることは望ましくない．そこで，画像の大きさを縦 300 ピクセル，横 240 ピクセルとし，正解姿勢と推定姿勢の誤差が 20 ピクセル以内となる確率を評価尺度として用いる．人物画像には，京都大学で実際に行われた講義を撮影した映像中の受講者を 3 人分使用する．各画像にはあらかじめ正解クラスを手動で割り当てる．

実験では第一に，クラス分類器の性能を測る．これは，各受講者について leave-one-out 交差検定を行うことで精度を求める．その後，各受講者について姿勢推定を行う．こちらも各受講者について leave-one-out 交差検定を行い，推定姿勢と正解姿勢の誤差が 20 ピクセル以内となる確率を求める．

4.2. 実験結果

クラス分類の結果を表 1 に示す．推定クラス 1 では精度が高くないものの，再現率が 100%となっている．また，正解クラス 2,3 では，80%以上の精度で正しい分類が行われていることが分かる．受講者は正面上方から撮影されており，顔が見えていても正しく顔検出が行われないことが多かったため，推定クラス 1 の精度が低くなったと考えられる．

誤差が 20 ピクセル以内となる確率を表 2 に示す．提案手法によって誤差が 20 ピクセル以内となる確率が大きくなることが分かる．推定クラスを使用した場合，正解クラスを使用するより確率は小さいが，従来手法と比較して 30%ほど改善されている．

提案手法によって，誤差が 20 ピクセル以内となる確率は従来手法のおよそ 2 倍となった．村上ら[4]は，約 80%の精度で正解姿勢から講義への興味度を正しく推定している．本稿の推定姿勢を利用した場合には精度がより低くなることが予想されるものの，分類器を利用した姿勢推定の有効性が示された．

表 1：クラス分類結果(枚)

	正解=1	正解=2	正解=3	計	精度
推定=1	113	66	97	276	0.41
推定=2	0	79	4	83	0.95
推定=3	0	16	100	116	0.86
計	113	161	201	475	-
再現率	1.0	0.49	0.50	-	-

表 2：誤差が 20 ピクセル以内である確率

手法	使用クラス	確率
従来手法	不使用	0.35
提案手法	推定クラス	0.65
提案手法	正解クラス	0.69

5. おわりに

本稿では，受講者画像の姿勢推定を対象として，画像を一定の基準で分類することで高精度な検出器を作成する手法を提案した．実験により，画像の分類を組み込むことで，性能が向上することを示した．

今後の課題として，より有効なクラス定義の模索が挙げられる．今回は受講者画像から得られた知見に基づいてクラスを定義したが，姿勢推定性能の向上により貢献するクラス定義を考える必要がある．

参考文献

[1]米谷淳，授業観察事始め-授業というフィールドにおける本格的な行動研究を目指して-京都大学高等教育教授システム開発センター編，玉川大学出版部, 2001.

[2]Sam Johnson, Mark Everingham, "Learning effective human pose estimation from inaccurate annotation", CVPR2011, pp.1465--1472, 2011.

[3]Yi Yang, Deva Ramanan, "Articulated Human Detection with Flexible Mixture-of-Parts", IEEE trans. PAMI, Vol. 35, No.12, pp.2878--2890, 2013.

[4] 村上正行,Gary Jay Coffman,正司哲朗,上田真由美,角所考,美濃導彦,"一斉型講義における受講者の姿勢情報の分析に基づく集中状態の検出",人工知能学会第 58 回先進的学習科学と工学（ALST）研究会，pp.7--10,2010.