

上半身識別器と動き情報を用いた映像からの受講者挙動解析

江間 雄二 中島 克人

東京電機大学大学院未来科学研究科

1. はじめに

講義内容の改善のための参考指標の1つとして、受講者による講義評価が用いられることが多い。例えば、一連の講義の後に受講者が記入する講義評価アンケートで明らかになった問題点を、次期の講義内容や講義方法にフィードバックできる。しかし、アンケートは受講者の主観的な評価であるため、そのみでは講義の改善に繋がる情報として十分とは言えず、何らかの客観的な評価が望まれる。そこで我々は、“講義中の受講者の挙動”を講義改善の参考指標に出来ると考えた。そのための現実的で最も簡易な仕組みとして、教室前方上部に設置されたカメラによる映像を元に受講生の挙動解析を行うシステムを提案する。

本稿では本システムにおける挙動解析手法の提案と評価について述べる。

2. 従来研究

服部ら[1]は講師行動、スライドや板書内容、受講者の顔上げ行動を時系列コンテキストとして表示するシステムを開発している。このシステムでは、主に講師行動に注目して詳しい解析を行っている一方で、受講者に注目した解析は顔上げ行動のような単純なものしか行われていない。

西野ら[2]は、固定カメラで撮影した講義映像をオフライン処理で解析し、受講者の演習中の移動動線を自動検出する問題に取り組んでいるが、何をしているかという挙動については扱っていない。

3. 提案手法

本システムでは図1に示すように、撮影した映像を読み込み、HOG特徴を用いた上半身識別器と肌色情報を併用して受講者の位置検出を行う。その後、受講者の肌色情報と動き情報を用いて挙動解析を行う。

受講者を教室正面から撮影すると、手前と奥の受講者同士の重なりにより、手前の受講者しか解析できない場合がある。そのため、本システムではカメラを受講者の斜め上に設置して撮影することを想定する。例えそのようにカメラを設置しても、受講者全員を手元まで含めて撮影することは困難であるが、本システムでは受講者の挙動の傾向を得ることが目的のため、受講者の過半数程度の解析を行えばよいものとする。なお、挙動解析はオフライン(講義後)に行うものとする。

以下に各処理の詳細を示す。

3.1 受講者の位置検出

受講者の位置検出はHOG(Histogram of Oriented

Video Analysis on the Behavior of Class Participants by Upper Body Classifier and the Motion

Yuji EMA, Katsuto NAKAJIMA

Graduate School of Science and Technology for Future Life, Tokyo Denki University

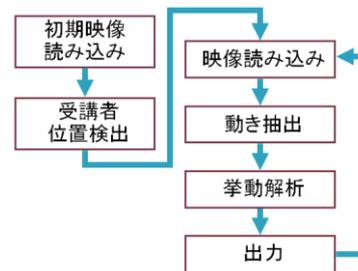


図1 システムの流れ

Gradients)特徴[3]を学習させた識別器を用いる。HOGは局所領域における勾配方向をビンとし、勾配強度を度数としたヒストグラムを特徴とする特徴量である。

識別器にはSVM(Support Vector Machine)を使用する。SVMは教師あり学習により構築される識別器の1つであり、学習データを2クラスに分離する超平面の中でマージンが最大になる超平面を選択するという特徴により、未知のデータに対する識別率の高さに定評がある。

本システムは着席型の講義に利用することを目的としており、受講者の位置は講義中に大きく動くことはないと考えられる。そのため、位置検出は映像読み込み直後に行い、それ以降は行わないものとする。

識別器で位置検出を行った後、検出領域内で肌色情報の検査を行う。この肌色量検査に合格した領域を受講者のいる領域とする。肌色量検査ではHSV空間の色相Hの値が8から36の範囲[4]を肌色の判断基準とする。ただし、輝度が極端に低い、もしくは、高い場合は、色相だけによる検査は誤検出の増大を招くため、輝度の値がある閾値を満たすことも条件に加える。

3.2 動き抽出

受講者の動き量を得るために、フレーム間差分を用いる。フレーム間差分とは、撮影時間がわずかに異なる2枚の画像の各画素位置において差分を取ることで移動体を検出する手法である。今回はRGB色空間上で1つ前のフレームと現在フレームの画素値のユークリッド距離dの値を各画素位置で求め、これを数フレーム間繰り返して蓄積する。この蓄積した値がある閾値以上かどうかで動きの有無を判断する。数フレームに亘って蓄積するのは、撮影時のノイズに対し耐性を持たせるためである。

図2は一連の入力画像(a)に対して動きがあったとする位置を白色画素で示した例を(b)に示している。



(a)入力画像の蓄積

(b)動き抽出画像

図2 動き抽出

3.3 受講者の挙動解析

受講者の位置検出後、各検出領域内で再度、肌色情報と動き情報を検査し、受講者の挙動解析を行う。講義中に受講者の挙動は数秒で大きく変わることはない想定するため、挙動解析は毎フレーム行わず、例えば 10 秒毎に行う。解析した受講者の挙動は表 1 のように分類する。なお、人間の目でも判断できない状況、例えば、顔を前方に向けて呆けていたり、下を見ながら話を聞いていたりする状況の誤判断は許容する。

表 1 受講者の挙動の分類

	下を見ている (顔上げフラグ OFF)	それ以外 (顔上げフラグ ON)
動きあり	作業	
動きなし	居眠り	聴講

3.3.1 顔上げ状態の判定

受講者が顔上げを行った場合、検出領域内の肌色量の増加が起こると考えられる。そこで 10 秒前の検出領域内と現在フレームのその肌色量を比較し、その増加量が閾値を超えると顔上げ行動を行ったと判断し、逆に肌色量の閾値以上の減少があれば下を向いたと判断する。どちらの状態にあるかは顔上げフラグで管理する。

3.3.2 居眠り状態の判定

一定時間、例えば 20 秒間動きが無く、かつ、顔上げフラグが OFF の時、居眠りしていると判断する。

4. 評価

実験は事前に撮影した 2 種類の動画を用いた。動画サイズは 2 種類とも 640×480 ピクセルである。動画読み込み直後の位置検出の結果画像を図 3 に示す。図 3(a), (b) とも実際の講義中に撮影した動画である。今回、受講者の位置検出に用いる輝度値は 5 以上 150 以下の範囲とした。

4.1 学習データ

SVM による学習を行うためには大量の正例・負例の学習データが必要となる。今回の学習には、正例として Google や Yahoo による画像検索などで収集した画像から 1,000 枚の上半身画像を切り出して用いた。カメラの設置位置を考慮し、斜め前からの上半身画像も選択している。負例としては正例収集のために集めた画像から、上半身でない部分をランダムに切り出した 1,000 枚を用いた。画像サイズは正例負例とも 64×64 ピクセルである。

4.2 位置検出の評価

受講の様子を撮影した 2 種類の動画の位置検出の結果を表 2 に示す。表 2 の検出予定数とは、検出領域内に受講者の顔と体がそれぞれ半分以上映るような受講者を著者が数えた総数である。検出数は図 3(a), (b) において位置検出を正しく行えた受講者の総数であり、検出領域内に受講者の顔と体がそれぞれ半分以上入っていればよいものとした。誤検出は人でない場所の検出数と同じ受講者の重複検出数の合計である。表 2 から、2 種類の動画とも、HOG 特徴で検出し肌色情報で絞り込んだ方法により、検出予定数の約半分を検出できたことがわかる。



(a)動画 1 (b)動画 2

図 3 評価用動画

表 2 位置検出の結果(単位: 人)

	検出予定	検出	誤検出	未検出
動画 1	25	12	8	13
動画 2	24	13	5	11

4.3 挙動解析の評価

システムによる作業、聴講、居眠りの挙動分類と、5 名の被験者の人手による分類を比較した。

各被験者にシステムが検出した受講者領域を 1 箇所ずつ示し、合計 100 秒の動画に対して 10 秒(300 フレーム)間隔で、作業、聴講、居眠りのどれかに分類して貰う。これを動画 1 では 12 箇所、動画 2 では 13 箇所の検出領域に対しそれぞれ行って貰った。その後、本システムの分類結果と比較し、その一致率を評価した(表 3)。

一致率を下げる主な理由は以下の 2 点であった。

- (1) 撮影位置や角度、パソコンなどにより受講者の手元が映らない場合、人目では受講者が作業しているように見えても、システムでは手元部分の動きが取得できずに聴講状態に誤分類する。
- (2) 人目では聴講しているように見えても検出対象付近にいる受講者の動きを取得してしまい、作業状態に誤分類する。

表 3 挙動解析の結果

	解析数(人)	一致率
動画 1	12	43%
動画 2	13	49.85%

5. まとめ

本稿では、教室前方上部から撮影した映像による受講者挙動解析システムの提案と、その受講者位置検出および挙動解析の評価を行った。位置検出および挙動解析の精度はいずれも約半分であり、これらの向上が今後の課題である。また、システムの有用性の検証も行わなくてはならない。

参考文献

- [1] 服部博憲, 他, 講義映像インデキシングのための講師・受講者の行動認識に基づく時系列コンテキストの獲得, 京都大学工学部情報学科卒業論文, 2007.
- [2] 西野博貴, 他, 講義映像における移動物体動線の自動検出, FIT2013(第 12 回情報科学技術フォーラム), H-052, 第 3 分冊, pp.207-208, 2013.
- [3] N.Dalal, et al., Histograms of Oriented Gradients for Human Detection, Proc.CVPR, vol.1, pp.886-893, 2005.
- [4] J.Sherrah, et al., Skin Colour Analysis, University of Edinburgh, 2001.