JSiSE Research Report
vol.34,no.5(2020-1)

受講生画像からの表情・姿勢推定に基づく学習状態判定機能を備えた遠隔講義システムの開発

府馬 央昂，鷹野孝典
神奈川工科大学 情報学部 情報工学科

Development of a Remote Lecture System with a Function of Estimating Student's Learning Condition based on Prediction of Facial Expressions and Postures from Student Learning Images

Hiroaki Fuma，Kosuke Takano
Department of information and Computer Sciences,
Faculty of Information Technology, Kanagawa Institute of Technology

This paper presents a remote learning system with a function of estimating student's learning state based on the prediction of the facial expressions and postures from student learning images. In order to predict the facial expressions and postures from the student learning images, Convolutional Neural Network (CNN) is applied to build a prediction model. Since continuous capturing images of student's face and posture in a classroom would give some stress to the students, we consider setting a video camera away from student positions. However, when a video camera is set away from students, the resolution of captured image of student's face and posture would be lower according to the distance, and it would cause to decrease the recognition accuracy as well. In the experiment, we confirm the feasibility of our system by measuring the recognition accuracy of face and posture according to the distance between students and a camera.

キーワード: 遠隔講義，学習状態推定，解像度，講義者，受講者，表情・姿勢推定，CNN

1. はじめに

講義では, 教師が学生の理解具合や集中度を把握し, それに応じて授業を展開することが望ましい. しかし, 大教室で大人数の学生が受講する場合は，授業中に個々の学生の様子を知ることは困難である．また，Webinar形式の講義を含めた遠隔講義システムの利用も，授業提供の機会を広めるとともに，教師の授業負担の軽減できる点で有用であると考えられる. しかし, 遠隔授業では，受講者の学習状態を把握することは一層困難となる.

本研究では, 受講生画像から, Convolutional Neural Network (CNN) を用いて個々の学生の表情と姿勢を判定し，判定された表情と姿勢の組み合わせにより，その学生の学習状態を推定する機能を備えた遠隔講義システムを示す．受講生画像を取得するには，撮影カメラが必要となる．ただし，受講生と設置カメラの距離が近いと受講生に緊張感や圧迫感を与えてしまう場合があることと，ノート PC を利用しない授業もあるため，ノートPCに内蔵されるカメラ等ではなく，講義室の天井に撮影カメラを設置することを想定する．しかし，設置カメラと受講生の距離が遠くなるほど画像の解像度が低下するため，画像中の表情や姿勢の認識

精度が低下するという問題が生じる．

このため，本研究では，設置カメラと受講生の距離に応じた表情や姿勢の認識精度の変化を測定し，提案システムの実現可能性を検証する．

2. 関連研究

センサーから取得したデータを対象として機械学習により学習状態を推定することに関する研究がなされている[(1)(2)(4)(5)(6)]．文献(1)では，聴講者の状態推定課題をパターン認識系と解釈し，機械学習の枠組みで特徴量獲得を行う Convolutional Neural Network (CNN) を用いた聴講者の状態推定システムを提案している．また，文献(2)では，マルチモーダルラーニングアナリティクス(MMLA)を「単一または複数のローレベルインタラクションリソースを用いて認知的領域・情動的領域・技能運動的領域に関する学習支援を行う」ことと捉え，実効性の高い MMLA の実現への課題を示している．また，手塚らは，椅子の座面上部に設置した圧力センサーと机の天板裏側に設置した赤外線距離計測センサーを用いて推定した受講者の状態から，取り組んでいるタスクを推定するシステムを提案している[(4)]．文献(5)において，鈴木らは，教室内で多数のグループが活動する際の学習同士の協調性を推定し，指導者にフィードバックする一手段として Kinect を用いた動作分析を行い，抽出情報と主観評価との相関傾向および他手段との組合せによる精度向上可能性について考察している．さらに，文献(6)では，人によって体格の違いがある影響により，検出できなかった姿勢・行動を検知するために複数のフォトレジスタを用いる方法について検討した．

e-Learning システムでの受講生の集中度や学習状態などの推定手法についていくつかの研究が見受けられる[(3)(7)]．文献(3)では，個人を対象とした e-learning システムで顔の向き情報と上半身姿勢情報を利用して集中度を推定する手法を提案している．また，文献(7)において， 安彦らは，一般的なパソコンに容易に設置が可能な Web カメラを用いて，e-Learning システムの受講者の顔画像を取得し，検出した視線情報から教材学習時の注意点を取得・蓄積することで，学習態度を把握するための方式を検討している．

提案システムは，遠隔講義などで講師がその場にいなくても受講生がどの程度理解や集中しているかといった学習状況を把握できることを目的とする．提案システムは，遠隔授業だけでなく，通常の講義においても利用することができる．例えば，講義で説明しているときに受講生全体の 9 割が理解できていると判定できるならば，講師が説明したことがきちんと伝わっているとみなすことができる．

また，撮影カメラは受講生との距離を置いて設置することを想定するため，受講生は撮影されているという圧迫感や緊張感などを受けずに講義に取り組める．演習時間のような場合においても学習状態を把握することにより，全体として演習が円滑に進んでいるかなどを確認できるため演習指導の効率が向上すると期待される．

3. 提案システム

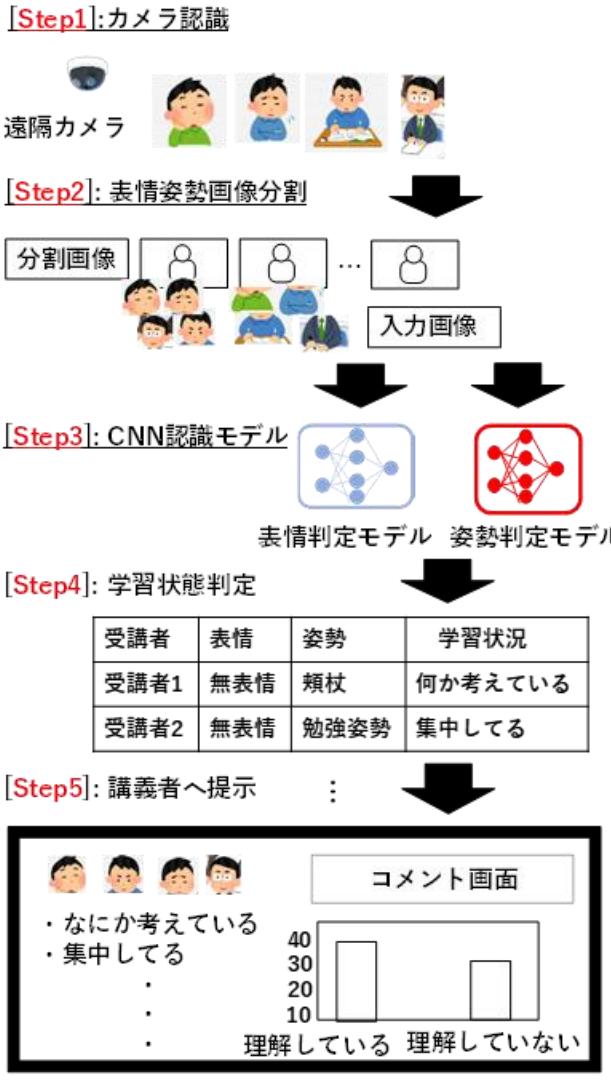

図 1 提案システムの概要図

図 1 に提案システムの概要図を示す．提案システムでは，受講生画像から個々の学生の表情と姿勢を判定し，判定された表情と姿勢の組み合わせにより，その学生の学習状態を推定し，講義者に提示する．実行手順を下記に示す．

Step-1: 天井などの設置カメラから，教室内の受講生の学習状況を撮影する．

Step-2: 撮影したカメラ動画を n 枚の画像に変換し，画像において表情と姿勢を認識する．表情と姿勢と認識された画像領域を矩形画像として切り取る．

Step-3: Step-2 で切り取った表情と姿勢の矩形画像を Convolutional Neural Network(CNN)等で構築した判定モデルにより表情と姿勢の状態を推定する．

Step-4: Step-3 で推定した結果から受講生の学習状態を推定する．

Step-5: Step4 で推定した個々の受講生の学習状態に基づき，教室全体における受講者の理解度や集中度を講義者に提示する

4. 実験システム

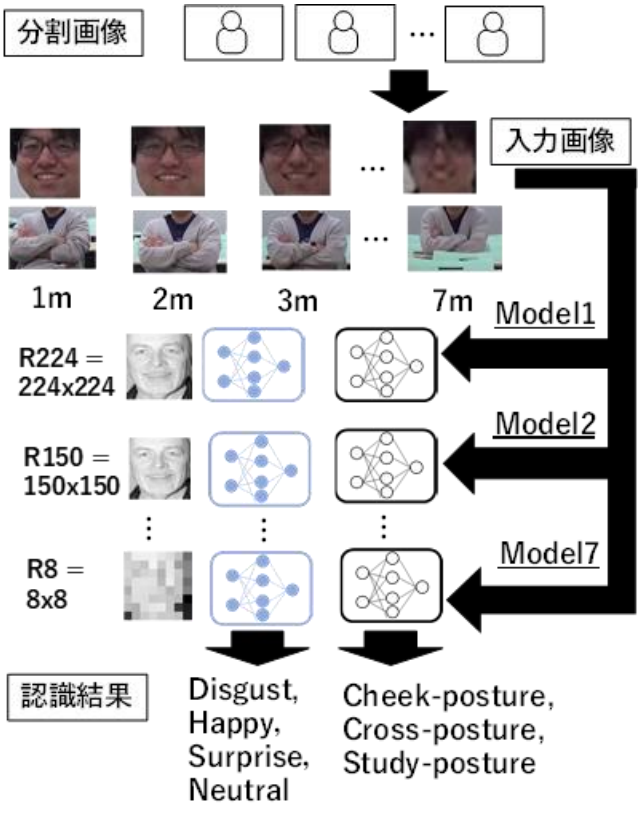

図 2 実験システムの概要図

実験システムとして，(1) 受講生画像から表情と姿勢を抽出する機能，および(2) CNN を用いて表情と姿勢の内容を推定する機能を実装した．

受講生画像から表情と姿勢を抽出する手順を下記に示す．

Step-1: 画像から顔認識により，顔の抽出領域座標を取得する．ここで，顔認識には Face Recognition[11] と OpenCV[12] の顔カスケード分類器を用いた．

Step-2: 顔の抽出領域座標をもとに姿勢の座標を決定し，姿勢の抽出領域座標を取得し，それぞれの抽出領域を切り取る．切り取った画像の例を，5 章の図 4 に示す．

また，CNN を用いた表情と姿勢の判定手順を下記に示す．

Step-3: 上記 Step-2 の表情・姿勢画像をグレースケール化，および 224x224 にリサイズする．

Step-4: 表情画像を表情判定モデル，姿勢画像を姿勢判定モデルに指定された解像度 r で入力する．

Step-5: Step-4 で判定結果を提示する．

Step-4 において，切り取った顔・姿勢画像から表情と姿勢判定を行うために，CNN を適用して表情判定モデルおよび姿勢判定モデルを構築した．事前学習済み CNN モデルとして xception を用いた．また，転移学習用画像データとして，Helem dataset[8]および Dfhq dataset[9]を用いた．さらに，crawler[10]や Google Images，Bing Images，Baidu Images などの画像検索エンジンを利用して独自に画像収集した．また，独自で撮影した画像も収集した．

CNN の判定モデルは，設置カメラと受講生の距離に応じた表情や姿勢の認識精度の変化を測定するために，7 種類の解像度(224x224，150x150，128x128，64x64，32x32，16x16，8x8)の画像でファインチューニングした 7 つの判定モデルを作成した．判定モデルの作成方法を下記に示す．

Step-1: 上記の Step-1～Step-3 の要領で，顔画像と姿勢画像を切り取る．

Step-2: 画像の色をグレースケール化し，画像サイズを 224x224 にリサイズする．

Step-3: 224x224 サイズの画像から 7 種類の画像サイズにそれぞれリサイズして画像を荒くする．7 種類の画像の例を表 1 に示す．

Step-5: 表情や姿勢の分類ラベルは手作業で付与する．

表情と姿勢の分類ラベルをそれぞれ表 2 と表 3 に示す．Step-6: Step-5 から Step-6 の画像データおよび分類ラベルを用いて，7 種類の解像度(224x224，150x150，128x128，64x64，32x32，16x16，8x8)に対応した 7 つの判定モデルを作成する．

表 1　7 種類の解像度例

解像度	表情	姿勢	解像度	表情	姿勢
224x224 (R244)			150x150 (R150)		
128x128 (R128)			64x64 (64x64)		
32x32 (R32)			16x16 (R16)		
8x8 (R8)					

表 2　切り取った顔画像とラベル例

			
Disgust	Happy	Neutral	Surprise

表 3　切り取った姿勢画像とラベル例

		
Cheek-posture	Cross-posture	Study-posture

5.　実験

設置カメラと受講生の距離に応じた表情や姿勢の認識精度の変化を測定し，提案システムの実現可能性を検証する．

5.1　実験環境

表 4 に示す撮影環境で，動画を撮影した（図 3）．動画から画像に変換するために 1 秒間に 30 枚を変換した．これらの画像は，4 章で示した 7 種類の解像度画像に対応した CNN による判定モデルの学習画像データとテスト画像データの一部として用いる．なお，表情は Neutral, Disgust, Happy, Surprise の 4 種類，姿勢は頬杖，腕組，勉強姿勢の 3 種類の画像に分類される．また，判定モデルは(i) Disgust と Neutral の表情判定モデル，(ii) Surprise と Neutral の表情判定モデル，(iii) Happy と Neutral の表情判定モデル，(vi) 頬杖，腕組，勉強姿勢の姿勢判定モデルの 4 種類の判定種別ごとに構築した．実験に用いた画像データ数を表 5，6 に示す．

	
1m で撮影した画像	3m で撮影した画像
	
5m で撮影した画像	7m で撮影した画像

図 3　撮影画像の例

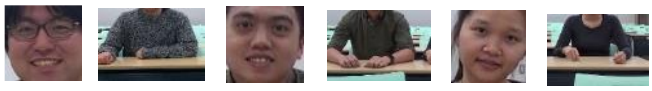

図 4　表情・姿勢画像を切り取った例

表 4　撮影環境

	概要
距離	1-7m (1m 間隔で 7 地点)
被験者	10 人
撮影時人数	1-2 人
撮影解像度	(a) 1920x1080, (b) 1280x720

表 5　表情・姿勢判定モデルの学習枚数

モデル種類	表情・姿勢種類	学習用
表情判定モデル(i)	Disgust	1595 枚
	Neutral	2024 枚
表情判定モデル(ii)	Happy	8612 枚
	Neutral	8545 枚
表情判定モデル(iii)	Neutral	2055 枚
	Surprise	2024 枚
姿勢判定モデル(iv)	頬杖	10285 枚
	腕組	10413 枚
	勉強姿勢	10433 枚

表 6　実験データ数

		解像度(a)	解像度(b)
テスト	表情	2800 枚 (= 4 種類 x 100 枚 x 7 地点)	2800 枚 (= 4 種類 x 100 枚 x 7 地点)
	姿勢	2100 枚 (= 3 種類 x 100 枚 x 7 地点)	2100 枚 (= 3 種類 x 100 枚 x 7 地点)

5.2　実験方法

カメラから 1～7m 離れた 7 地点で撮影した画像を用いて，解像度ごとに構築した 7 つの判定モデルによる認識精度の変化を比較・考察する．

5.3　実験結果

(A) 表情判定モデル(i)

表情判定モデル(i)を用いた場合の表情認識の判定精度結果を図 6～9 に示す．図 7,図 9 の結果から，解像度の一番高い R224 では,「Neutral」の判定において，設置カメラと被験者との距離が遠くなるに従い，認識精度が低下する傾向が見られた．しかし，図 6, 図 8 の結果から，「Disgust」の判定においては，「Neutral」とは逆に, カメラと被験者との距離が遠くなるに従い，認識精度が向上する傾向が見られた. 図 6~図 9 から解像度(a)と(b)の違いによる表情認識精度を比較すると，「Neutral」では解像度(a)が，「Disgust」では解像度(b)の方が，認識精度が高くなる結果であった．また図 6 の結果から，R64 では，設置カメラと被験者の距離が遠くなるに従い, 認識精度が向上する結果となった.

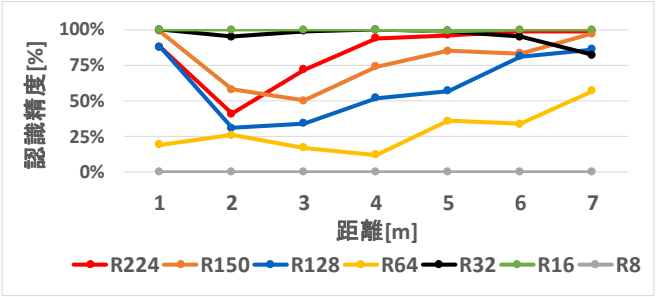

図 6　解像度(a)画像の認識精度(Disgust)

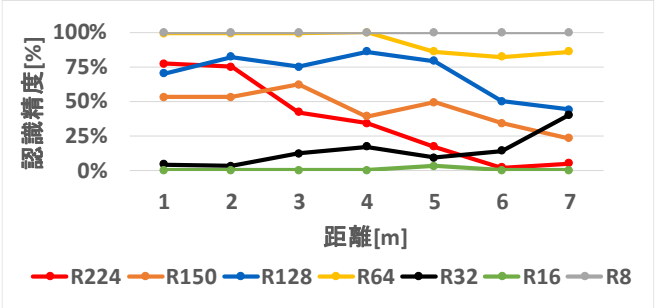

図 7　解像度(a)画像の認識精度(Neutral)

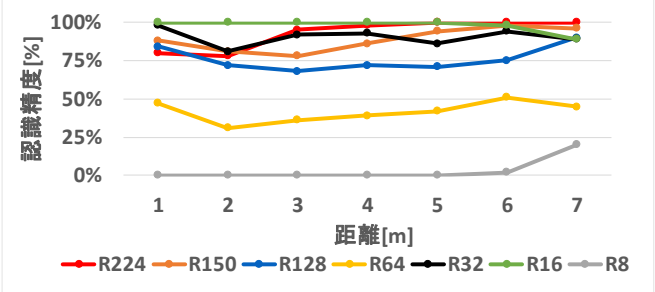

図 8　解像度(b)画像の認識精度(Disgust)

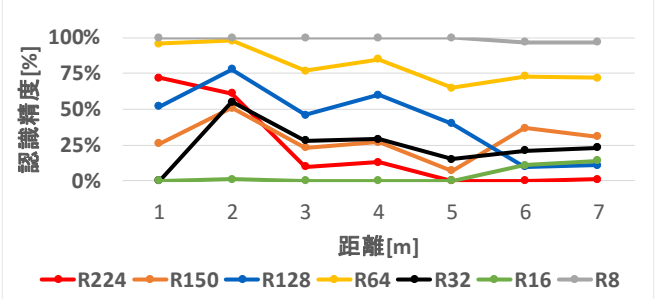

図 9　解像度(b)画像の認識精度(Neutral)

(B) 表情判定モデル(ii)

表情判定モデル(ii)を用いた場合の表情認識の判定精度結果を図 10～13 に示す．図 11，図 13 の結果から，R224 では「Neutral」の判定において，設置カメラと被験者との距離が遠くなるに従い，表情認識精度が低下する傾向が見られた．しかし，図 10，図 13 の結果から,「Happy」の判定においては,「Neutral」とは逆に，カメラと被験者との距離が遠くなるに従い，認識精度が向上する傾向が見られた．図 10~図 13 から解像度(a)と(b)の違いによる表情認識精度を比較すると，「Neutral」では解像度(a)が，「Happy」では解像度(b)の方が，認識精度が高くなる結果であった．また図 13 の結果から，R16 では，設置カメラと被験者との距離が遠くなるに従い，表情認識精精度が向上する傾向が見られた.

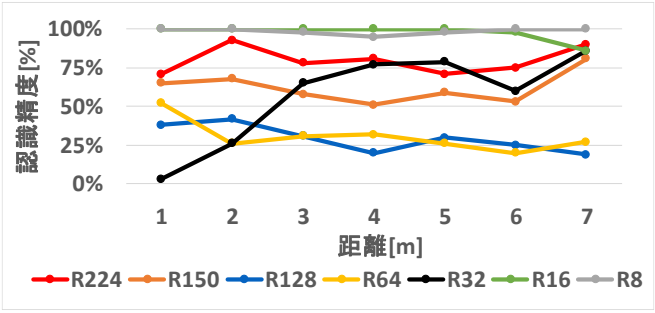

図 10　解像度(a)画像の認識精度(Happy)

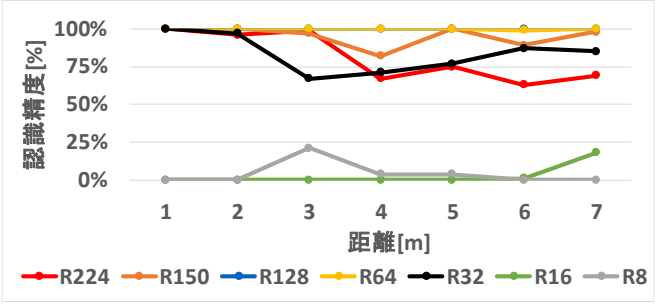

図 11　解像度(a)画像の認識精度(Neutral)

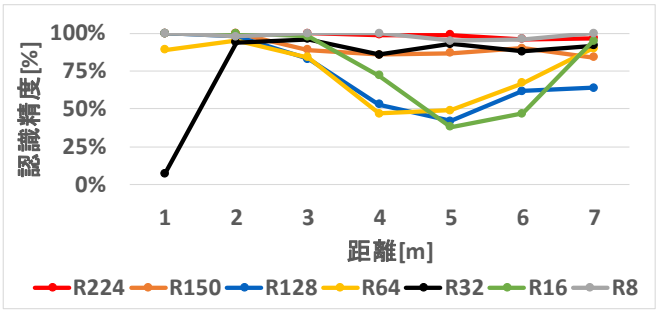

図 12　解像度(b)画像の認識精度(Happy)

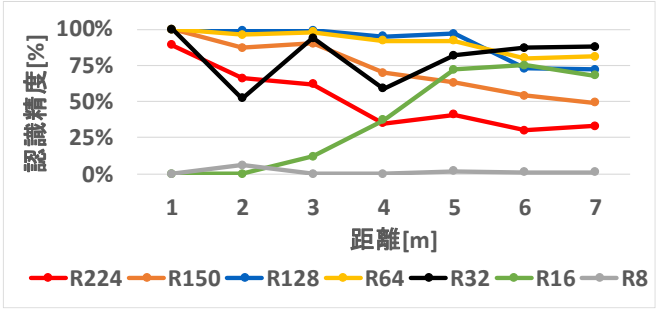

図 13　解像度(b)画像の認識精度(Neutral)

度が低下する傾向が見られた．しかし，図 15，図 17 の結果から，「Surprise」の判定においては，「Neutral」とは逆に，カメラと被験者との距離が遠くなるに従い，認識精度が向上する傾向が見られた．図 14~図 17 から解像度(a)と(b)の違いによる表情認識精度を比較すると，「Neutral」では解像度(a)が，「Surprise」では解像度(b)の方が，認識精度が高くなる結果であった．また図 15，図 17 の結果から，R64,R128 では，設置カメラと被験者との距離が遠くなるに従い，認識精度が向上する傾向が見られた．

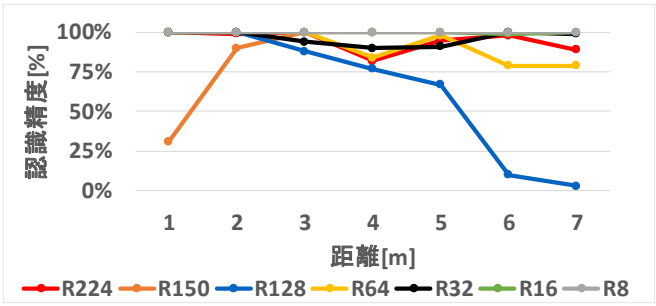

図 14　解像度(a)画像の認識精度(Neutral)

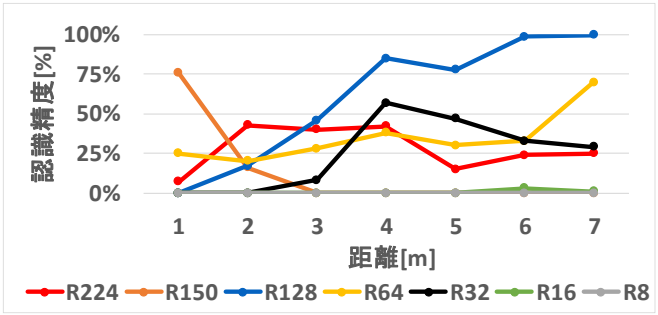

図 15　解像度(a)画像の認識精度(Surprise)

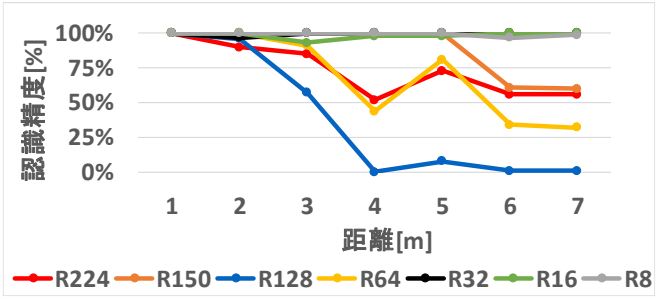

図 16　解像度(b)画像の認識精度(Neutral)

(C) 表情判定モデル(iii)

表情判定モデル(iii)を用いた場合の表情認識の判定精度結果を図 14～17 に示す．図 16 の結果から R224 では解像度(b)画像の「Neutral」判定において，設置カメラと被験者との距離が遠くなるに従い，表情認識精

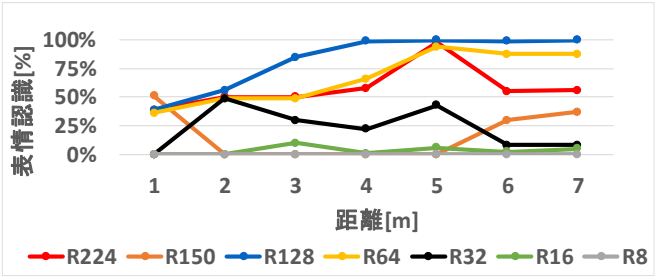

図 17　解像度(b)画像の認識精度(Surprise)

(D) 姿勢判定モデル(iv)

姿勢判定モデル(iv)を用いた場合の姿勢認識の判定精度結果を図 18～23 に示す. 図 22 の結果から, R224 では解像度(b)画像の「Cross-posture」判定において, 設置カメラと被験者との距離が遠くなるに従い, 表情認識精度が低下する傾向が見られた. 図 18, 図 21-23 の結果から, R16,R8 以外のモデルにおける認識精度に変化は見られなかった. また, 図 19, 図 22 から解像度(a), (b)を比較すると, 解像度(b)の方が, 認識精度が高くなる結果となった

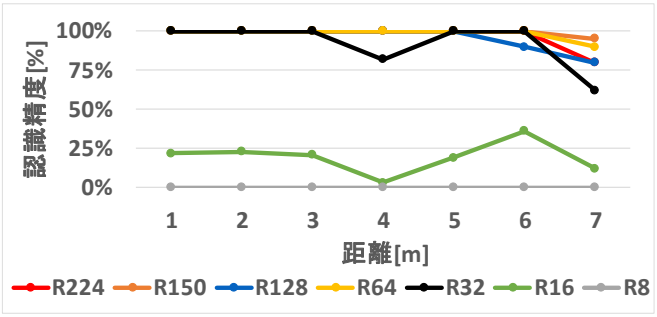

図 18　解像度(a)画像の認識精度(Cheek-posture)

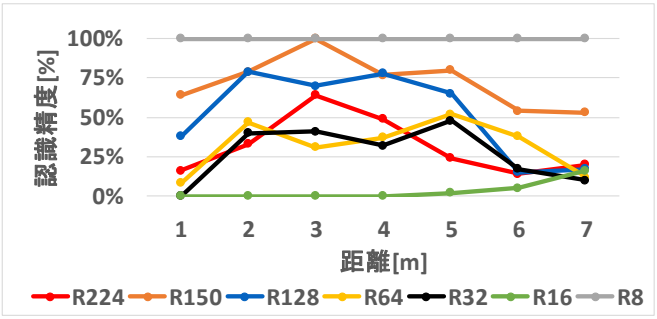

図 19　解像度(a)画像の認識精度(Cross-posture)

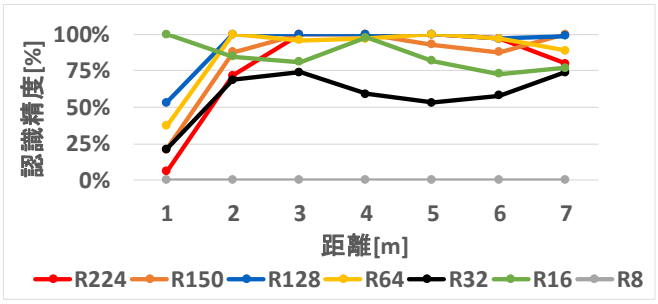

図 20　解像度(a)画像の認識精度(Study-posture)

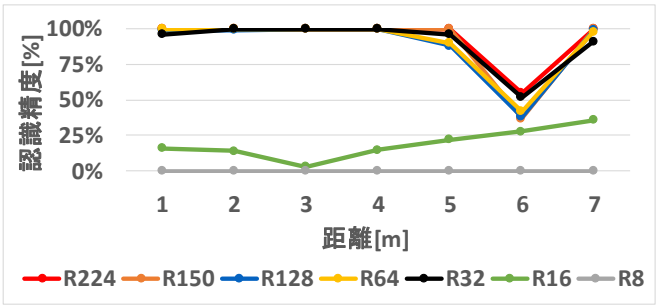

図 21　解像度(b)画像の認識精度(Cheek-posture)

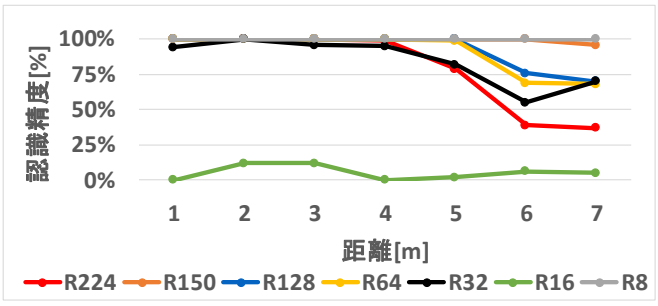

図 22　解像度(b)画像の認識精度(Cross-posture)

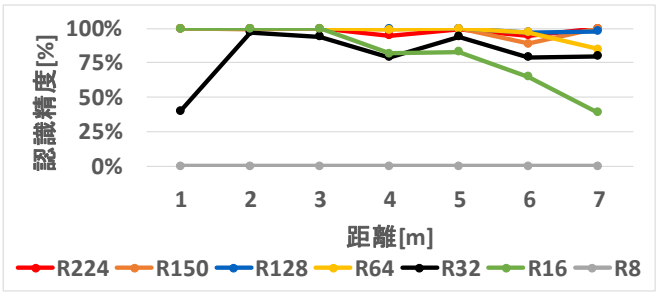

図 23　解像度(b)画像の認識精度(Study-posture)

[考察]

実験結果から, 構築した CNN モデルは解像度が(a), (b)で撮影したときの解像度に適応可能であると考えられる. 姿勢判定の結果から解像度(a), (b)において認識精度の差があまり見られなかった結果から、画像枚数が多い場合には高い認識精度で解像度(a), (b)で認識できると考えらえる. R8, R16, R32 の認識精度が不安定になった原因として撮影した解像度とそ

れぞれの CNN モデルの解像度が適応しなかった可能性が考えられる．

6. まとめ

本研究では，受講生画像中の表情と姿勢を CNN を適用した判定モデルを用いて分類し，表情と姿勢の組み合わせから学習状態を推定して受講生全体の理解度として提示する機能を備えた遠隔講義システムを提案した．実験結果から，画像の解像度が 64x64 ピクセル～224x224 ピクセルであれば 1-7m の距離で撮影した場合において表情と姿勢の推定が可能であることが確認できた．また，表情と姿勢を推定することができたことから，表情と姿勢推定の組み合わせによって学習状態も推定できることが分かった．これらの実験結果から，受講生全体の理解度や集中度を判定することができる見込みを得ることができた．

現時点では，動画像中の受講生の表情・姿勢を時系列データとして捉えてないため，今後の課題として，時間変化を考慮した表情・姿勢判定を行うことが必要だと考えられる．また，構築した判定モデルでは，表情・姿勢の認識パターン数がともに数種類程度にとどまっているため，より多くのパターンの表情・姿勢を認識可能な判定モデルを構築していく予定である．

参 考 文 献

(1) 島田大樹,彌富仁: “畳み込みニューラルネットワークを使った授業映像中の聴講者の状態推定システムの構築と特徴量獲得に関する検討”, 知能と情報(日本知能情報ファジィ学会誌), Vol.29, No.1, pp.517-526 (2017)

(2) 松居辰則: “マルチモーダルラーニングアナリティクス”,情報処理, Vol.59, No.9, pp.810-814 (2018)

(3) 立花優斗,今井順一: “e-learning 学習者の上半身姿勢情報を利用し集中推定”,FIT2016 第 15 回情報科学技術フォーラム,N-018,pp.329-330(2016)

(4) 手塚太郎,清野悠希,古谷遼平,佐藤哲司: “姿勢計測による e-learning 受講者の行動推定”,知能と情報(日本知能情報ファジィ学会誌),Vol.28,No.6,pp.952-962(2016)

(5) 鈴木雅実，張諾，木村寛明，高木正則: “学習者の行動分析に基づく協働学習支援に向けて-Kinect を用いた協調性の判定-”，第 30 回全国大会(2016)，2D3-2

(6) 森章汰，佐々木皓平，高瀬治彦，川中普晴，北英彦: “複数のフォトレジスタを用いた講義中の学生の行動推定の試み”，2019 PC Conference

(7) 安彦智史，池辺正典，丸山広，長谷川大: “PC 内蔵カメラを用いた学習態度把握方式の検討”，情報教育シンポジウム 2015 論文集，pp.103-108(2015-08-10)

(8) Helen-dataset, http://www.ifp.illinois.edu/~vuongle2/helen/

(9) ffhq-dataset, https://github.com/NVlabs/ffhq-dataset

(10) icrawler, https://pypi.org/project/icrawler/

(11) face_recognition, https://github.com/ageitgey/face_recognition

(12) OpenCV, https://opencv.org

(13) Keras https://keras.io/ja/