Heterogeneous Learning を導入した Deep Convolutional Neural Network による運転手の骨格検出と顔向き推定

Pose and Face Orientation Estimation of Driver by Deep Convolutional Neural Network with Heterogeneous Learning

奥野 薫子†, 山下 隆義†, 福井 宏†, 山内 悠嗣†, 藤吉 弘亘†, 乗冨 修蔵†, 新 浩治†

†: 中部大学, {kaoruko@vision., yamashita@, fhiro@vision., yuu@isc., hf@}cs.chubu.ac.jp

概要: 自動運転のレベル 3 では, システムが何かしらの原因で自動運転を継続できない場合, システムは運転手の状況に合わせて運転手操作モードに切り替える. これを実現するには, 運転手の姿勢と顔向きから等運転可能な状態かどうかを常にモニタリングする必要がある. 従来の姿勢と顔向きの計測には, 各タスクごとに特徴抽出と識別を行うため, 処理時間を要する. そこで, 本研究は, 複数タスクを同時学習する Heterogeneous Learning を DCNN に導入することで, 特徴抽出過程を共有し, 骨格検出と顔向き推定を同時に出力する手法を提案する. 評価実験により, 骨格検出の位置精度は 98%, 顔向き推定の識別精度は 91%を実現した. また, 画像 1 枚あたりの処理時間は GPU のとき 2.6ms, CPU のとき 34.1ms であることから, リアルタイムに骨格検出と顔向き推定が可能であることを確認した.

1. はじめに

自動車の自動運転はレベル 1～5 まであり, 加速, 操舵, 制動の動作をシステムが行うレベル 3 の実用化が着期待されている. 自動運転レベル 3 では, システムが対応できない場合, 運転手が操作を行う必要がある. このとき, システムは運転手の状況に合わせて運転手操作モードに切り替える. これを実現するには, 運転手の状態を常にモニタリングすることが重要となる. このとき, 運転手の状態を認識するためには, 上半身の姿勢や顔向きを計測する必要がある. また, 運転手の不注意が原因で発生する交通事故は, 2015 年 11 月の全国交通事故統計によると, 1 位が漫然運転 17%, 2 位がわき見運転 13%, 3 位が運転操作不適 12%となっている. そこで, 自動車の内部にカメラを設置して運転手の状態を把握することで運転手の不注意を抑制することも期待できる.

従来の骨格検出や顔の向き推定は, 特徴量の設計と Regression Forest[6]など回帰推定ができる識別器を組み合わせる手法が一般的である[4][7][8]. 2012 年以降は, 一般物体認識コンテストで Deep Convolutional Neural Network (DCNN) を用いた手法が, 従来の認識方法より高精度に認識をしたことで, 骨格検出や顔の向き推定にも応用されている[2]. し

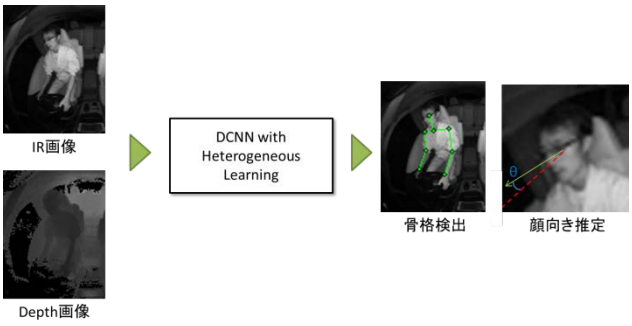


図 1 運転手の骨格検出と顔向き推定

かし, DCNN は非常に計算コストが高く, 骨格検出と顔の向き推定を行う場合はそれぞれのタスクに対して DCNN を構築する必要がある. そのため, 推定時には 2 つの DCNN を実行する必要があり, 実用化の面で大きな課題となる.

本研究では, 複数のタスクを 1 つのネットワークで学習できる Heterogeneous Learning[1]を用いることで, 図 1 のような運転手の骨格検出と顔向き推定を 1 つの DCNN で行う. Heterogeneous Learning は, 単一のネットワークで複数のタスクを学習, 識別する方法である. この Heterogeneous Learning を用いることで, 骨格検出と顔向き推定を 1 つの DCNN で実行でき, リアルタイムで処理することが可能となる. また, 自動車の内部は明暗が変化しやすいため, 入力に Infrared

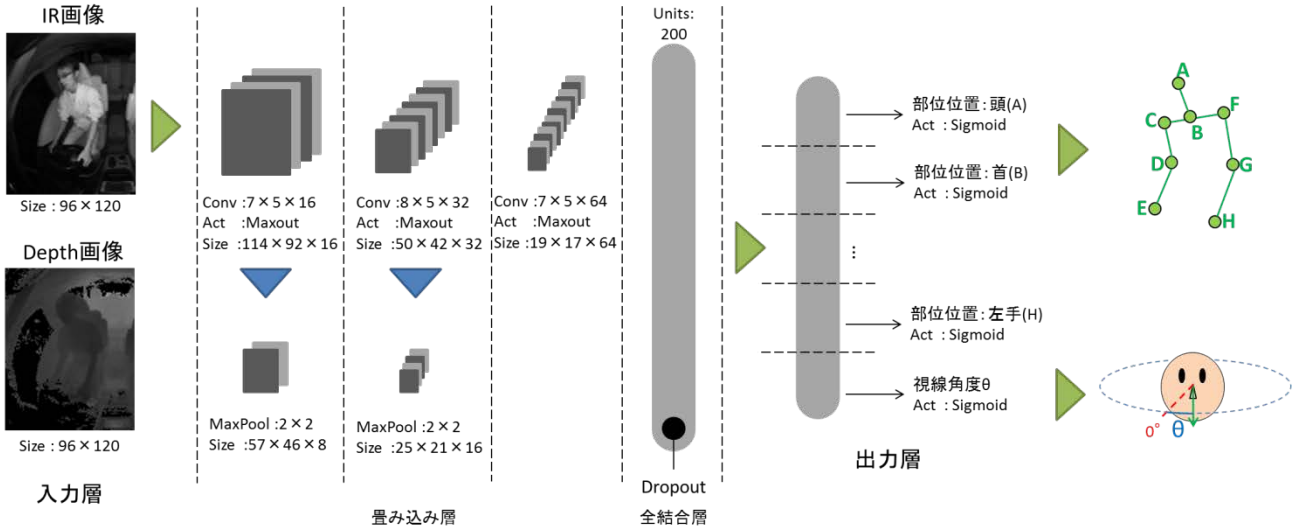


図 2 Heterogeneous Learning を導入した DCNN の構造

Radiation(IR)画像と距離(Depth)画像を用いる．これにより，自動車が屋内やトンネル等に入った時や夜間の運転時に起きる明暗の変化に対して頑健な骨格検出と顔向き推定が期待できる．

2. 関連研究

本章では，従来の骨格や顔向きの推定方法について述べる．

2.1. 機械学習を用いた骨格検出と顔向き推定

Shotton らは，Depth 画像から検出した人体に対して Random Forests[5]を用いてパーツの識別を行い，各パーツの重心位置を検出する方法を提案している[3]．Depth 画像と各パーツを色で表した正解ラベルの組み合わせを学習サンプルとして構築した Random Forests から，離れた 2 点の画素が持つ距離値の差を特徴量としてしきい値処理により左右に分岐する．そして，パーツラベルごとに重心を求めることで人物の姿勢を求める．

また，G. Fanelli らは，Depth 画像を入力とし，Regression Forests[6]を用いて顔向きを推定する手法を提案している[4]．この手法では，様々な顔向きや表情の Depth 画像を人工的に大量生成し，それを入力に用いている．Depth 画像を用いることで個人差を吸収し，さらに人工生成画像により顔向きや表情変化に頑健な手法となっている．

2.2. DCNN を用いた骨格検出

Toshev は，RGB 画像から DCNN を用いて骨格検出を行う手法を提案している[2]．人間は隠れている関節の位置も他の部位との位置関係やその対象人物の動きなどから推測することができる．DCNN の畳み込みとプーリング処理により，大局的な骨格位置関係を把握することができる．これにより，隠れている関節の位置をおおまかに推定することが可能である．この Toshev らの手法では同じ構造の DCNN を直列に繋げ，全身の大まかな骨格推定した後，各推定点を入力に用いることで更に正確な骨格推定を行うことが可能である．

2.3. 従来法の問題点

2.1.の機械学習を用いた骨格検出では，自然な動作で操作可能なユーザーインタフェースとして，ゲームなどの操作入力に利用されているが，自己遮蔽に対して脆弱であるという問題がある．また，顔向き推定は，対象とする顔画像のサイズは大きく，全身を捉えた映像からの顔向き推定は困難である．そこで DCNN を用いることは 2.2.から有用であるといえるが，複数の DCNN を扱うことは処理コストの増加に繋がる．そのため，骨格検出と顔向き推定の DCNN をそれぞれ構築することはリアルタイム処理の観点から非効率的である．

3. 提案手法

本研究では， Heterogeneous Learning を導入した DCNN により，運転手の骨格検出と顔向き推定を 1 つのネットワークで行う．以下に，運転手の撮影環境と DCNN の構成および Heterogeneous Learning について述べる．

3.1. 運転手の骨格検出と顔向き推定

本研究では，自動車が屋内やトンネル等に入った

時や夜間の運転時に起きる明暗の変化に頑健に識別するために，入力には IR 画像と Depth を用いる．使用するカメラは，車内のバックミラー付近に設置されており，運転席と助手席全体を撮影している．画像サイズは，640×480 である．提案手法では，まず，運転席付近のみを切り出し， 96×120 にリサイズして図 2 のように DCNN に入力する．DCNN の出力ユニットは，部位位置の数 8×2 と顔向き角度の 17 ユニットある．運転手の部位位置は，頭，首，右肩，右肘，右手，左肩，左肘，左手の計 8 ヶ所であり，その x 座標および y 座標が出力される．また，顔向き角度は，人が正面を見ている状態を 0 度とし，左右に首を振るときの角度(yaw 角)とする．運転手視点から，右が＋，左がーとなる．使用するデータセットには，12 人のサンプル画像が 32,914 枚あり，学習に 30,000 枚，評価に 2,914 枚使用する．運転手はよそみやスマホの使用，サンバイザーを開く，顔に触るなどの様々な姿勢変化が生じているシーンを対象としている．

3.2. Heterogeneous Learning

Heterogeneous Learning は，単一のネットワークで複数のタスクを学習，識別する方法である．一般に，複数の識別タスクを実行する場合，タスクの数に比例して識別器を構築する必要がある．そのため，実用化の面で非効率的である．しかし，Heterogeneous Learning は，単一のネットワークで複数のタスクを同時に学習，実行することが可能な為，タスクが増加しても計算コストが大きく増加することない．

提案手法における Heterogeneous Learning では，出力層で骨格検出と顔向き推定の回帰の値を出力する．

骨格検出タスクの誤差

骨格検出は骨格位置 8 ヶ所の x 座標と y 座標が出力されるため，学習誤差E_sは式(1)となる．

$$E_s = \sum_{n=0}^{N} \|\boldsymbol{L}_n - \boldsymbol{O}_n\|_2^2 \qquad (1)$$

このとき，$\boldsymbol{L}_n$は教師信号，$\boldsymbol{O}_n$は出力値，Nは骨格の部位位置の数である．

顔向き推定の誤差

顔向きに対応する出力ユニットは1つであるため，学習誤差E_gは式(2)となる．

$$E_g = (L - O)^2 \qquad (2)$$

よって，学習誤差Eは式(3)となる．

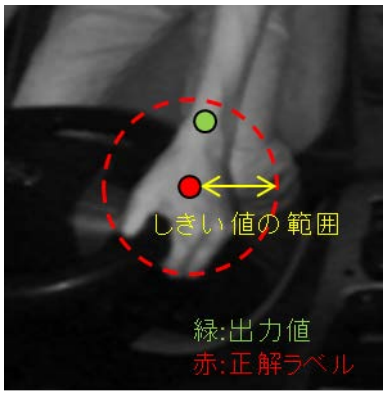


図 3 評価方法

$$E = \alpha E_s + (1 - \alpha) E_g \qquad (3)$$

αは骨格検出と顔向き推定の学習誤差に対する重みであり，どちらのタスクを優先するかを決めるパラメータである．本研究では，αを 0.5 とし，2 つのタスクを平等に扱っている．学習データセットから複数の学習サンプルを選択し，Mini-batch を作成する．Mini-batch 学習はDCNN の学習で一般的に用いられており，M 個の学習サンプルをランダムに選択して DCNN に入力する学習法である．入力した M 個の学習サンプルから学習誤差Eを求め，式(5)の誤差逆伝播法を用いて DCNN の結合重み$\mathbf{w}$を更新する．ηは学習係数であり，本研究では 0.001 としている．

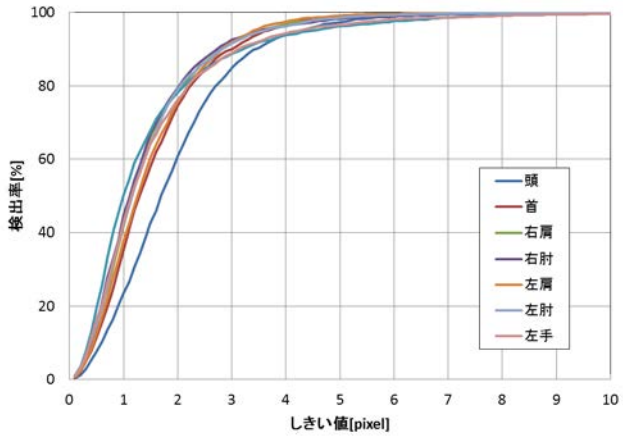
$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \frac{\partial E}{\partial \boldsymbol{w}} \qquad (4)$$

4. 評価実験

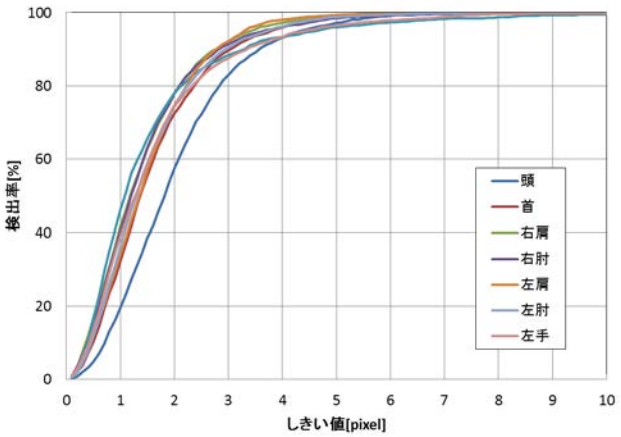
4.1. 評価方法

骨格検出の評価では，ユークリッド距離と文献[2]にて使用されている Percentage of Correct Parts (PCP)により評価をする．まず，ユークリッド距離の評価は，図 3 のように正解ラベルと出力値のユークリッド距離を求め，求めた距離がしきい値以下であれば検出成功，しきい値以上であれば検出失敗とする．ここで，しきい値は 0 から 10pixel まで変化させ，精度の変化を確認する．PCP の評価は，隣接する 2 つの骨格位置に着目し，各々の骨格に対する推定誤差が，その 2 つの骨格間のユークリッド距離の半分以下である場合に，部位の検出に成功とする．

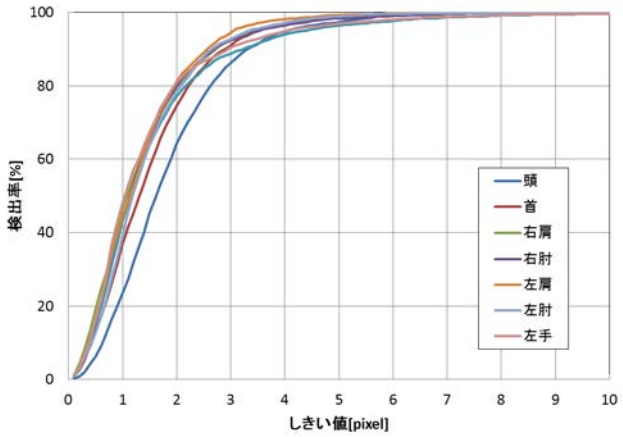
顔の向き推定では，正解ラベルと出力値の差を求め，求めた角度の差がしきい値以下であれば検出成功，しきい値以上であれば検出失敗とする．ここで，



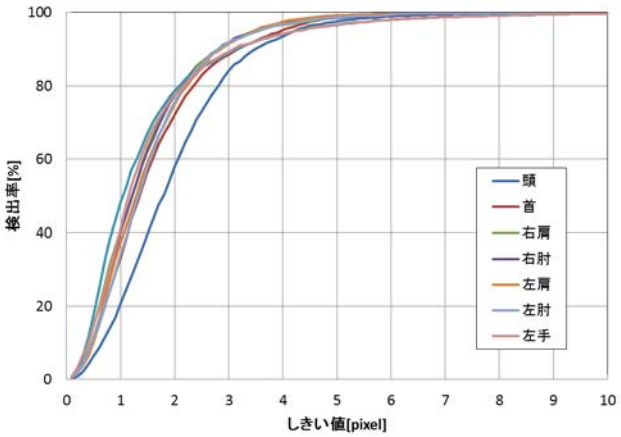
(a) Heterogeneous Learning(IR)



(b) Heterogeneous Learning(IR+Depth)



(c) 単一の DCNN(IR)



(d) 単一の DCNN(IR+Depth)

図 5 骨格検出における部位ごとの評価

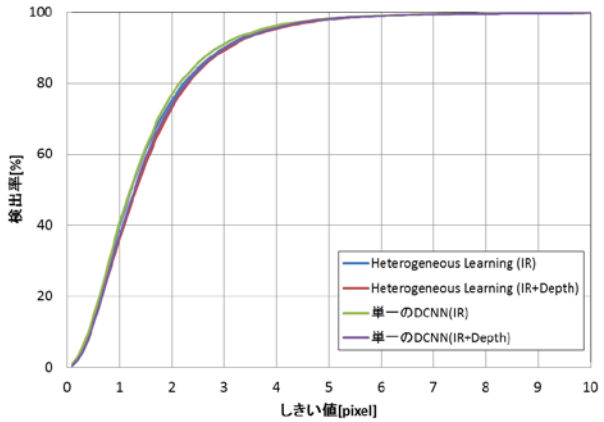


図 4 骨格検出の評価

しきい値は 0 から 20 度まで変化させ，精度の変化を比較する．

4.2. 骨格検出の性能評価

図 4 に骨格検出の評価結果を示す．図 4 より，単一の DCNN と同等の精度で検出できていることがわかる．また，各手法の入力を IR 画像または IR+Depth 画像としたとき，IR 画像を入力した方が，精度が良いことがわかる．図 5 は，部位ごとの評価結果である．全ての手法において，しきい値が 2pixel のとき，頭の精度が最も低く 60%程度である．これは頭の正解ラベルを頭部の中心にしており，正解ラベルにばらつきがあるためと考えられる．図 6 は，入力が IR 画像の提案手法の骨格検出例である．赤が正解ラベル，緑が出力値である．図 6 より，頭の正解ラベルからはずれているが，頭部に推定できているため，大きく精度が落ちることはない．また，肩，首，手のような自己遮蔽が生じる部位も，正しく推定できていることがわかる．また，図5より，全ての手法において，左側の骨格より右側の骨格の方が精度がよい．これは，撮影しているカメラの位置の関係で，左側の骨格の方が大きく写り，正解ラベルにばらつきが起きやすいためである．

カメラに映る大きさに関係なく骨格を評価するため，

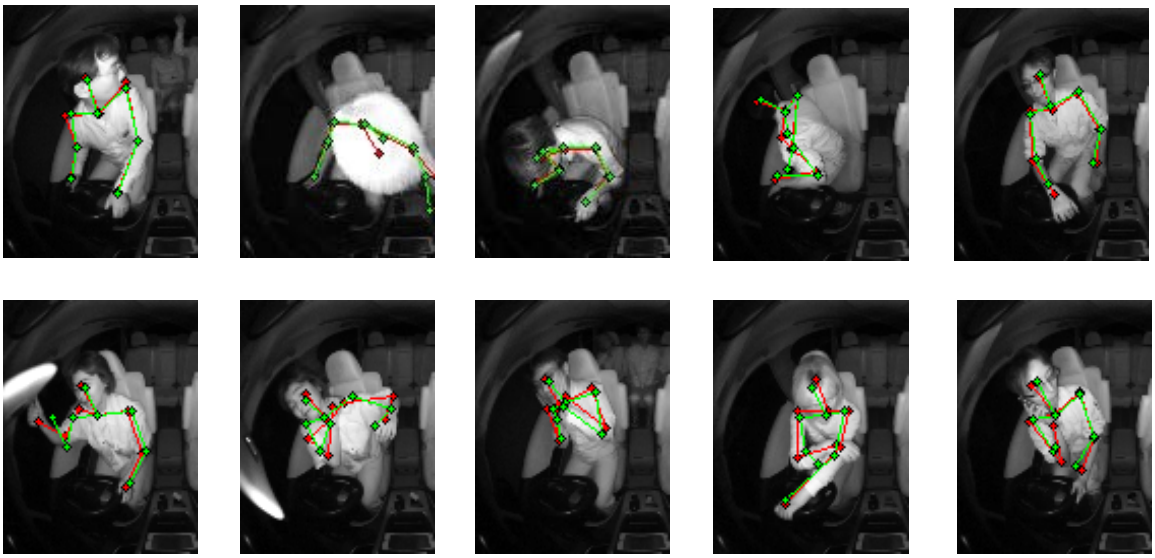


図 6 骨格検出の例

表 1 PCP による評価[%]

	右上腕部	右下腕部	左上腕部	左下腕部
Heterogeneous Learning(IR)	98.01	96.84	99.18	98.70
Heterogeneous Learning(IR+Depth)	98.80	96.84	99.59	98.73
単一骨格検出(IR)	98.66	96.91	99.49	99.18
単一骨格検出(IR+Depth)	98.97	97.29	99.38	99.11

PCP による評価結果を表 1 に示す. 表 1 より, 全ての手法において, 左の部位と右の部位の検出率は同等である. また, 全ての手法において, 下腕より上腕のほうが, 検出率が高い. これは図 5, 6 から分かるように, 手の位置は変化が大きく推定が肘や肩の位置より検出が難しいからである.

4.3. 顔向き推定の性能評価

図 7 に顔向き推定の評価結果を示す. 図 7 より, しきい値が 10 度のとき, 単一 DCNN より 10%程度認識率が低くなっている. 図 7 の結果をより詳しく確認するため. 図 8 にしきい値を 10 度としたときに, 顔角度を 5 度刻みに評価した場合の識別率を示す. 縦軸が識別率, 横軸が顔角度である. 図 8 より, 顔角度が 70 度以降, 0～30 度, －90 度以降精度が低下している. 精度が低下している角度の画像を図 9 に示す. まず, 顔角度が 70 度以降と－90 度以降の場合, 運転手は図 9 の(a), 図 9 の(c)のような体を一定の方向に大きく反らす動作をしており, 学習サンプル数が少ない. また, 顔角度が 0～30 度の場合, 図 9 の(b)のような正面付近を向いたまま頭を下げる, 体のみ前後に動かすといった動作をしており, 本実験で識別している角度(yaw 角)以外の要素が変化している. よって, 体を

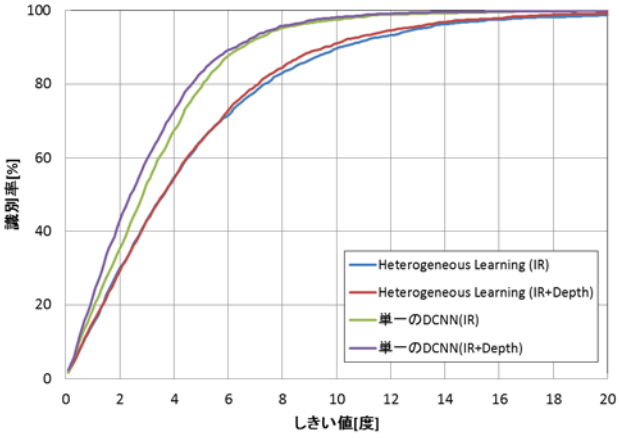


図 7 顔向き推定の評価

一定の方向に大きく反らすサンプルの追加と yaw 角以外の角度の追加を行うことで精度向上が期待できる.

4.4. 処理時間の評価

表 2 に 1 枚あたりの処理時間を示す. GPU は GTX 1080, CPU は Intel (R) Core (TM) i7-4790 CPU @ 3.60GHz を使用している. 表 2 より, 画像 1 枚あたりの処理時間は, 単一 DCNN のとき GPU で合計 4.4ms, CPU で合計 67.0ms に対して, 提案手法は, GPU で 1.8ms, CPU で 32.9ms の処理時間の短縮を実現した.

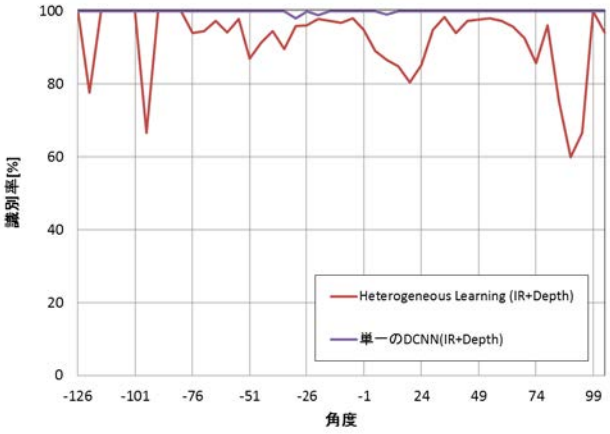


図 8 顔向きごとの識別率

また，提案手法は CPU においても，30fps によるリアルタイム処理が可能であることが分かる．

5. まとめ

本稿では Heterogeneous Learning を導入した DCNN による運転手の骨格検出と顔向き推定を提案した．提案法は，Heterogeneous Learning を導入することで，ネットワーク構成がコンパクトになり CPU でもリアルタイム処理することが可能である．今後は，推定で得られた結果を活用した運転手の行動認識に関する手法を検討する．

参考文献

[1] X. Yang, S. Kim, and F. P. Xing, ”Heterogeneous Multi-task Learning with Sparsity Constrain”, Advances in Neural Information Proce ssing Systems 22, 2009.

[2] A. Toshev, and C. Szegedy, “DeepPose: Human Pose Estimation via Deep Neural Networks Alexander”, Computer Vision and Pattern Recognition, 2015.

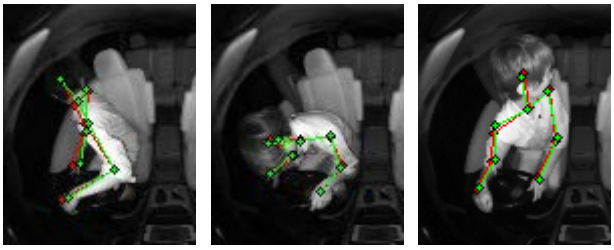
[3] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images”, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[4] G. Fanelli, J. Gall, and L. Van Gool, “Real Time Head Pose Estimation with Random Regression Forests”, Computer Vision and Pattern Recognition, 2011.

[5] L. Breiman, “Random forests”, Machine learning, 2001.

[6] M. Dantone , J. Gall, G. Fanelli, L. Gool, “Real-time Facial Feature Detection using Conditional Regression Forests”, Computer Vision and Pattern Recognition, 2012.

[7] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, ” Head Pose Estimation for Driver Assistance Systems:A Robust Algorithm and Experimental Evaluation”, IEEE Intelligent Transportation Systems Conference, 2007.



(a) 70 度以降　(b) 0 度付近　(b) -90 度以降

図 9　顔向き精度の悪い骨格例

表 2 処理時間[ms]

	GPU	CPU
単一骨格検出	2.5	33.9
単一顔向き推定	1.9	33.1
提案手法	2.6	34.1

[8] E. Murphy-Chutorian, and M. M. Trivedi, ” Head Pose Estimation in Computer Vision: A Survey”, IEEE transactions on pattern analysis and machine intelligence, 2009.

奥野薫子:現在中部大学工学部情報工学科在学中,画像を用いた骨格検出の研究に従事.

山下隆義:2002 年奈良先端科学技術大学大学院大学博士前期課程修了.2002 年オムロン株式会社入社,2009 年中部大学大学院博士後期課程修了(社会人ドクター),2014 年中部大学講師,人の理解に向けた動画像処理,パターン認識・機械学習の研究に従事.

福井宏:2016 年中部大学大学院博士前期課程修了,現在同大学大学院博士後期課程在学中,画像を用いた物体検出の研究に従事.

山内悠嗣:2012 年中部大学大学院博士後期課程修了,2010 年独立行政法人日本学術振興会特別研究員 DC.2014 年中部大学助手.コンピュータビジョン,パターン認識の研究に従事.

藤吉弘亘:1997 年中部大学大学院博士後期課程修了.1997~2000 年米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow.2000 年中部大学講師.2004 年より同大学教授.2005~2006 年米カーネギーメロン大学ロボット工学研究所客員研究員,計算機視覚,動画像処理,パターン認識・理解の研究に従事.

乗冨 修蔵:画像処理の研究に従事.

新 浩治: 画像処理の研究に従事.