社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

クラウドロボティクスのための画像認識エンジンの提案

山内悠嗣† 加藤 優† 山下隆義†† 藤吉弘亘††

†, †† 中部大学 〒 487-0027 愛知県春日井市松本町 1200
E-mail: †{yuu,u_kato}@vision.cs.chubu.ac.jp, ††{yamashita,hf}@cs.chubu.ac.jp

あらまし 少子高齢化に伴う労働人口の減少から，障がい者や高齢者を支援するためにロボットを活用することが期待されている．このようなサービスを実現するためのフレームワークとして，クラウド上の端末で情報処理を担うクラウドロボティクスの考えが普及している．本稿では，我々が提案するクラウドロボティクスのための人を観る技術を中心とした画像認識エンジンと，そのプロトタイプについて述べる．クラウドロボティクスでは，認識性能や計算量の他に，データ通信量とプライバシーについて配慮する必要がある．特に，顔画像は個人を特定できる情報になり得るため，特段の注意が必要である．本システムは，プライバシーに配慮しながらも，ロボット側の計算量とデータ通信量を制御可能なフレームワークである．

キーワード クラウドロボティクス，人画像解析，Deep Convolutional Neural Network

Proposal of Image Recognition Engine for Cloud Robotics

Yuji YAMAUCHI†, Yu KATO†, Takayoshi YAMASHITA††, and Hironobu FUJIYOSHI††

†, ††Chubu University 1200, Matsumoto-cho, Kasugai, AICHI 487-8501, Japan.
E-mail: †{yuu,u_kato}@vision.cs.chubu.ac.jp, ††{yamashita,hf}@cs.chubu.ac.jp

Abstract As the workforce continues to shrink due to the aging population, robots are expected to be utilized for the care of persons with disabilities and the elderly. As a framework for realizing such a service, cloud robotics is gaining popularity for processing the data on cloud servers. Cloud robotics require privacy and low data traffic in addition to high recognition performance and low computational cost. In particular, the robots need to be capable of personally identifying facial images. Our proposed system considers privacy while controlling the amount of calculations and data traffic on the robot side.

Key words Cloud robotics, Face image analysis, Deep Convolutional Neural Network

1. まえがき

労働力人口の減少に伴い，労働力不足を補うためにロボットの活躍が期待されている．特に，障がい者や高齢者の生活支援・人の生活の質 (QoL : Quality of Life) の向上のためのロボットが望まれている [1]．人の生活を支援するロボットに望まれるタスクとしては，人とロボットのコミュニケーションや物を拾い渡す等の短期間における自立支援から，病気の早期発見，安否確認，服薬支援等の長期間にわたる自立支援が挙げられる．

このようなタスクを実現するためには，ロボットはセンサから多くの情報を獲得し，膨大な計算を実時間で処理することが要求される．そのため，ロボットは高性能な計算機を搭載することが必要となり，ロボットの大型化，高価格化，高電力化に繋がるため，生活支援ロボットの実用化を妨げる要因となる．

この問題を解決する方法として，2013 年に Kuffner 等が提案したクラウドロボティクス [2] がある．クラウドロボティクスとは，ロボットとクラウド上の認識エンジンや情報処理技術を連携させ，大規模な計算やデータベースを要求するような情報処理をクラウドで担うという考え方である．特に，近年ではネットワークの高速化し，さらにネットワーク機器の急激に増加したことから，センサネットワークや Machine to Machine(M2M), Internet of Things(IoT) 社会が到来している．このような背景もあり，クラウドロボティクスの研究 [2]〜[6] が盛んに取り組まれている．

本稿では，我々が取り組んでいる生活支援ロボットのためのクラウドロボティクスとそのプロトタイプシステムについて述べる．生活支援ロボットには，タスクの多様性から様々な認識技術が必要であるが，認識の対象の観点から人を観る技術と物を観る技術の 2 つに分けることができる．本稿では，人を観る技術を中心としたシステムについて述べる．人とロボットの自然なコミュニケーションを実現するためには，人間の状態を観測する必要がある．また，人の状態の観測は病気の早期発見や

表 1　ロボットに要求されるタスクと必要技術の例

観る対象	必要な技術
物 (日用品)	物体認識，姿勢推定，空間認識，把持点検出
人 (本人，他者)	顔照合，表情認識，器官点検出，年齢推定，性別推定，人種推定

安否確認にも必要な技術である．このようなタスクを実現するために，我々は人の顔画像から人物の照合及び属性推定 (器官点，人種，年齢，性別，表情等) を行う認識エンジンをクラウドサーバに実装する．

人の顔画像を扱う上で問題となるのはプライバシーへの配慮である．顔画像は個人を特定できる情報になり得るため，顔画像の漏洩等には注意が必要である．本システムでは，ロボットからクラウドサーバーに顔画像を送る代わりに，顔画像から Deep Convolutional Neural Network(DCNN) [7] により抽出した特徴マップを送る．DCNN の構造や重み係数が既知でない場合には，特徴マップから顔画像を復元することはできない．そのため，特徴マップは暗号化されたデータと言えるため，プライバシーに配慮することができる．

しかしながら，ロボットが DCNN の処理の全てを担うと大きな負担となる．そこで，ロボット側の計算負荷を軽減するために，DCNN の構造を 2 分割する．ロボット側では，撮影した画像から顔を検出し，畳み込み及びプーリングを繰り返し処理する．そして，畳み込み，もしくはプーリング後の特徴マップをクラウドサーバに送信し，クラウドサーバでは受信した特徴マップを再度 DCNN に入力する．これにより，ロボットは DCNN の全てを計算する必要がないため負担が少ない．また，DCNN の後半の層から出力した特徴マップは，データ量が小さくなるため，通信という観点においても負担が小さくなる．本稿では，顔照合及び属性推定のタスクにおいて，DCNN の第何層で区切るとロボットの計算負荷，及び通信負荷が小さいか検証する．

2. 生活支援ロボットのためのクラウドロボティクス

本章では，生活支援に必要なタスクと技術について整理し，提案するシステムの特徴について述べる．

2.1 生活支援に要求されるタスク

短時間的なタスクと長時間的なタスクを整理し，その目的と必要となる技術について述べる．

短時間的な自立支援 · 見守り

短時間的な自立支援 · 見守りのタスクとしては，物を拾い渡すような “物を対象としたタスク”，人とロボットの対話コミュニケーションのような “人を対象としたタスク”がある．“物を対象としたタスク”では，まず画像中の指定された対象物をロボットが認識する必要がある．次に，認識した物を拾うためには，物体周辺の空間の認識し，物体の姿勢の推定及び物を掴むための把持点を検出する必要がある．認識する対象物体を日用品に限定しても，膨大な種類の物体が認識の対象となる．

“人を対象としたタスク”では，例えば人とロボットの対話コミュニケーションにおいて，ロボットが人物の特定及び状態を理解することが重要である．コミュニケーションに必要不可欠な要素は，対話する相手が誰であるかを理解することであり，ロボットは人物を正確に照合する必要がある．また，初対面もしくは面識がある場合においても人の状態を理解することも重要となる．例えば，顔の表情から心理状態を推定し，その状態に合わせてロボットが振る舞うことで，人に対してストレスを与えないコミュニケーションを実現できる．このようなタスクを実現するためには，時間にわたり安定して認識する必要があるため，高精度な顔画像解析が要求される．

長時間的な自立支援 · 見守り

長時間的な自立支援 · 見守りのタスクとしては，病気の早期発見，安否確認，服薬支援等を対象とする．長時間にわたり人の行動解析及び経年変化を分析することで，健康の維持 · 増進に繋がる見守り技術を実現できる．特に，人の顔は状態を把握するうえで重要な情報発信源であるとされ，顔の部位の変化から臨床症状及び，疾患が随伴する可能性を推測することができると報告されている [8], [9]．これをロボットで実現するためには，顔の器官点を検出，検出した器官点を解析することで，顔画像からの健康状態の自動診断が期待できる．また，人と共生するロボットは長時間にわたり人を観察することが可能であるため，日常的に変化する顔画像やコミュニケーションの取り方を総合的に判断することで，早期の臨床症状の発見が期待できる．

上記の各タスクでは，ロボットが観る対象物により必要となる技術が大きく異なる．観る対象で必要となる技術を表 1 に整理する．これらの高度な画像処理技術は，多くの計算量とデータベースを必要とする．ロボット単体での情報処理，データベースを保持する場合には，ロボットの大型化，高電力化，高価格化の問題が発生する．これらの問題を一挙に解決するためにもクラウドロボティクスによる情報処理が適している．

2.2 本システムの特徴

我々は，2.1 で挙げた観る技術を実現するクラウドロボティクスのシステムを開発している．現在，開発しているクラウドロボティクスのシステムの特長を下記に示す．

- リソース負担の軽減

高度な画像認識は，一般的に多くの計算量が必要となるため，ロボット側の情報処理能力を占有する問題がある．また，画像認識には膨大なデータベースや学習したモデルを必要とするため，ロボット側で全ての情報を保持することが難しい．クラウドロボティクスを用いることでロボット側のリソース負担を軽減できる．

- プライバシーへの配慮

人の顔画像は個人を特定できる情報になり得る．そこで，本システムでは，ロボットから顔画像をクラウドサーバーに送らず，顔画像から Deep Convolutional Neural Network により抽出した情報を送る．これにより，サーバーに送る情報からは顔画

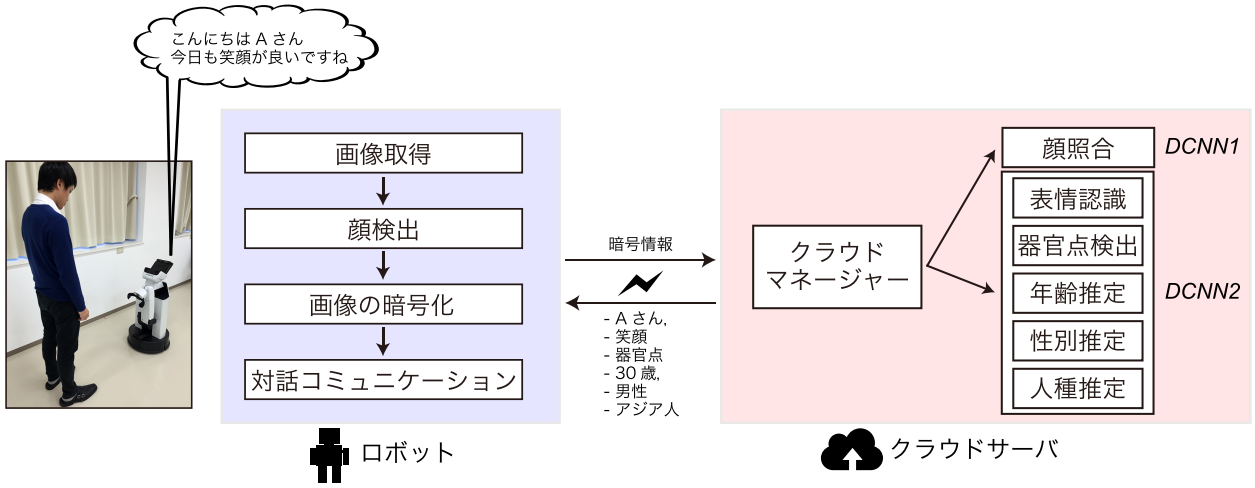

図 1　クラウドロボティクスによる顔画像解析を用いた対話コミュニケーションの例

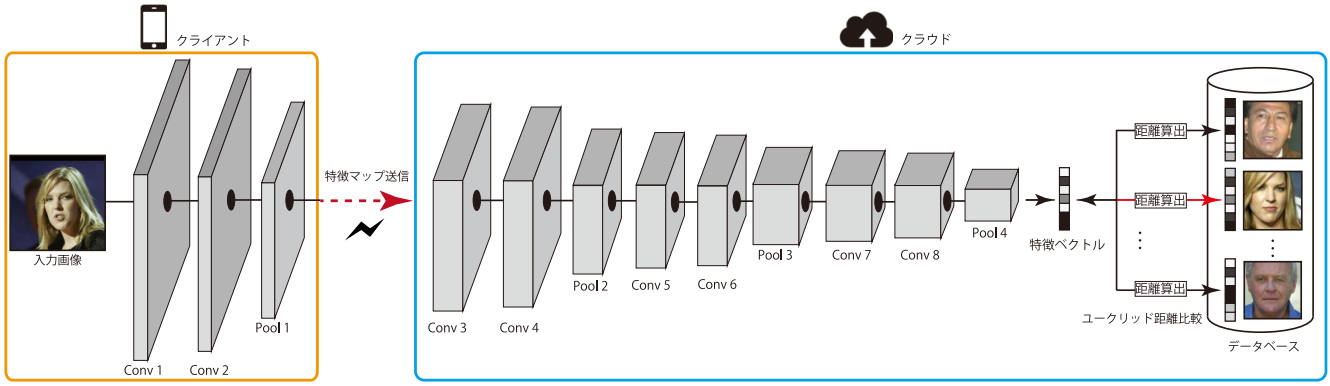

図 2　DCNN による照合の流れ.

像を復元できないため，プライバシーに配慮したシステムとなる.

- 画像認識エンジンの高精度化

ロボットから送信される画像をクラウドサーバに蓄積し，蓄積した画像を用いて逐次的に学習することで，画像認識エンジンを高精度化することができる．また，様々なロボットから送信される画像を追加学習に用いることで，画像の質，撮影環境等の違いを吸収できる認識エンジンを学習できる.

- ROS への対応

Robot Operating System (ROS) に対応したクラウドロボティクスを構成する．これにより，ROS に対応したロボットであれば本システムを使用できる構成とする.

3.　クラウドロボティクスを実現する人を観る認識エンジン

図 1 に人を観る技術による対話コミュニケーションの例を示す．まず，ロボットが取得した画像から顔を検出し，後述する方法により顔画像を暗号化する．そして，ロボットからクラウドサーバへ暗号化した情報を送り，顔画像を解析する．そして．その結果をロボットに送信し，顔画像解析結果に基づきロボットが行動する．以上の処理は，半リアルタイムで処理可能であるため，ロボットは円滑な対話コミュニケーションを実現することが可能となる．我々は，様々な分野で有効性が確認されている Deep Convolutional Neural Network(DCNN) [7], [10]〜[12] により顔画像を解析する．本認識エンジンでは，顔画像が誰であるかという顔照合，人の属性推定 (器官点検出，性別，年齢，人種，笑顔度) の 2 つを処理する.

3.1　顔　照　合

顔照合では，ロボットに搭載されたカメラの画像から検出した顔画像と，クラウドサーバに保存してある顔画像が同一人物であるかを照合する．図 2 に DCNN による顔照合の流れを示す．ロボット側では，撮影した画像から顔画像を検出し，畳み込み及びプーリングを繰り返し処理する．畳み込み，もしくはプーリング後の特徴マップをクラウドサーバに送信し，クラウドサーバでは受信した特徴マップを再度 DCNN に入力し，畳み込みとプーリング処理を行う．そして，全結合層前の特徴マップを特徴ベクトルとして使用する．抽出した特徴ベクトルと，予めクラウドサーバに保存された特徴ベクトルのユークリッド距離を計算し，距離が最も小さい特徴ベクトルの人物を出力する.

このように DCNN の構造を分割することで，ロボットは DCNN の全ての計算を負担する必要がないため，ロボット側の計算コストを低減することができる．さらに，インターネットを介したデータ通信には，画像ではなく特徴ベクトルを送信するため，プライバシーに配慮できる．DCNN の係数が既知でない場合，特徴ベクトルから元画像を復元することは不可能

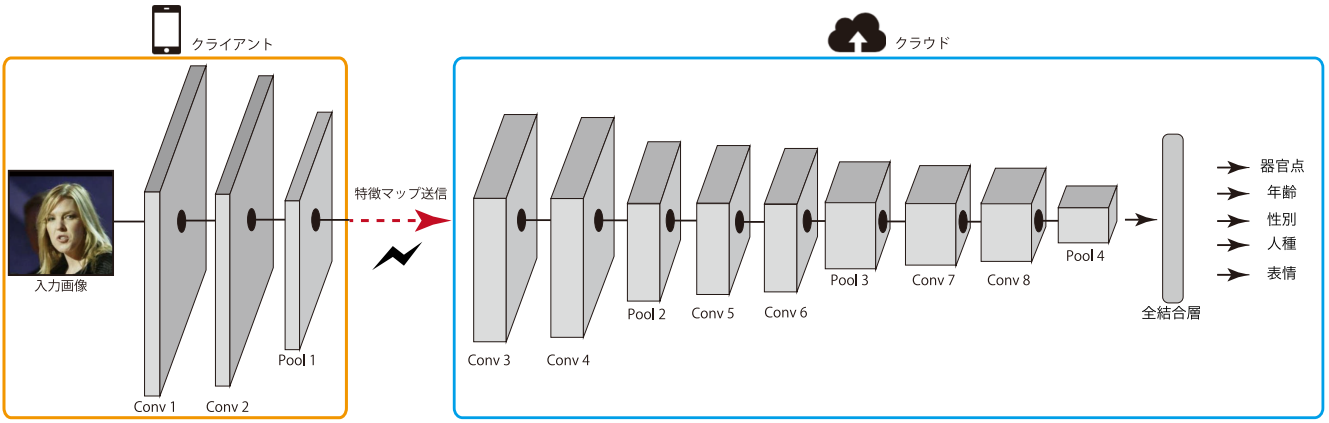

図 3　DCNN による属性推定の流れ.

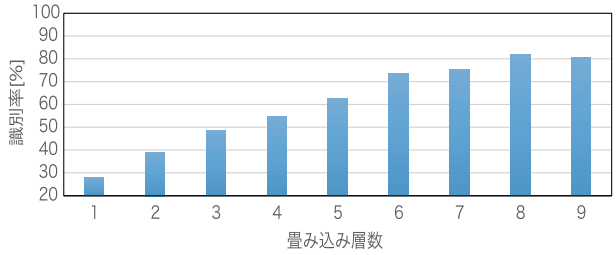

図 4　畳み込み層の数を変えた際の精度.

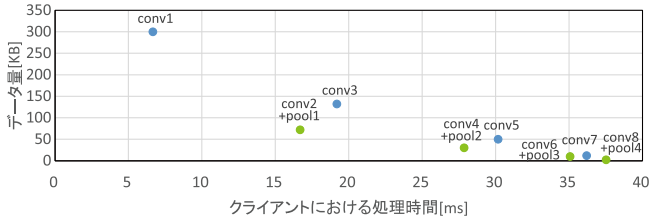

図 5　クライアントの情報送信量と処理時間.

である.

3.2　属 性 推 定

顔属性 (器官点，人種，年齢，性別，表情等) の推定法として，複数のタスクを単一の DCNN で学習，識別できる Heterogeneous Learning [13] を採用する．Heterogeneous Learning は，識別タスクと回帰タスクを同時に解く手法である．しかし，識別タスクと回帰タスクを同時に学習するため，各タスクにおける学習誤差の振幅の差により，学習に悪影響を与える場合がある．このような問題を解決するために，我々は Heterogeneous Learning に重み付き誤差関数 [14] を導入し，学習時の悪影響を抑制する方法を提案している [15]．各タスクに対して重みを与えることで，認識タスクの突発的な学習誤差を低減させ，安定した学習が可能となる.

図 3 に DCNN による属性推定の流れを示す．属性推定においても顔照合と同じように，DCNN の構造を分割することでプライバシーに配慮したシステムとなる.

4.　検 証 実 験

本章では，顔照合と属性推定において，DCNN の構造を分割した際のデータ通信量及び，計算量について検証する.

4.1　顔照合の有効性の検証

まず，顔照合に必要となる DCNN の畳み込み層の層数を調査した後に，DCNN の第何層で分割するとロボット及びクラウドサーバの負荷が少ないか検証する.

ロボット側の端末のスペックは CPU : Intel(R) Corei7-3520，メモリ : 8GB，クラウドサーバのスペックは CPU : Intel(R) Corei7-4790，メモリ : 8GB である．検証実験では，畳み込み層を 1 層から 9 層まで増やし，各構成における照合率を比較する．活性化関数には Maxout を使用し，プーリング層は畳み込み層の 2 層目，4 層目，6 層目及び 8 層目に配置する．各層のフィルタ枚数は 1，2 層目が 16 枚，3，4 層目が 32 枚，5，6 層目が 64 枚，7，8 層目が 128 枚，9 層目が 256 枚である．DCNN の学習サンプルは 4,031 枚，評価サンプルは 143 枚である.

図 4 に DCNN の畳み込み層数を変化させた時の識別率を示す．畳み込み層数を増やしていくと照合精度が向上しており，畳み込み層を 8 層とした場合の精度が最も高い．これ以降は，畳み込み層を 8 層とした DCNN を用いて検証実験を行う.

図 5 に，ロボット側で処理する畳み込み層数を変化させた時のロボット側の処理時間とデータ通信量を示す．ロボット側の端末はクラウドサーバと比べて計算能力が低いため，ロボットで処理する層が多い場合，総処理時間は長くなる．しかし，切り離す層が浅い場合，データ通信量が膨大となるため，通信時間が長くなる．図 5 より，2 層目の畳み込み層 (Conv2 + Pool1) で切り離した場合，ロボット側での処理時間が短く，かつデータ送信量が少なくなることがわかる.

図 6 は，切り離す畳み込み層を変えた時の総処理時間を示す．総処理時間は，ロボット及びクラウドサーバにおける DCNN の処理時間，クラウドサーバにおける顔照合時間，通信時間を総和したものである．顔照合時間はクラウドサーバに保存されている 10 人分のデータ(注1) と照合するための時間である．通信時間は通信速度を 100Mbps(12.5KB/ms) と仮定した際の値である．図 6 より，2 層目の畳み込み層で切り離した構成の総処

(注1)：1 家庭の構成人数は多くても 10 人と考えているため.

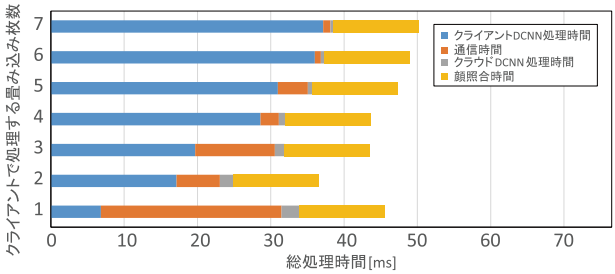

図 6　顔照合における切り離す畳み込み層の変化による総処理時間の変化.

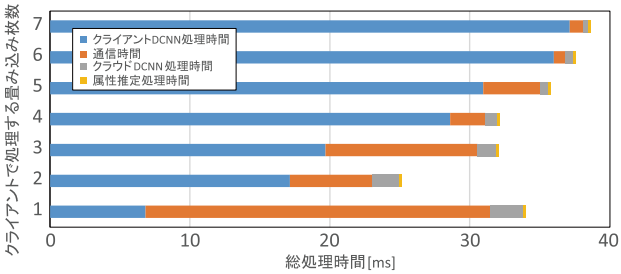

図 8　属性推定における切り離す畳み込み層の変化による総処理時間の変化.

理時間が最も短い．クラウドサーバで行う顔照合時間は，どの層で分割しても変わらないため，総処理時間はロボットの処理時間と通信時間が大きく影響していることが確認できる．

4.2　属性推定の有効性の検証

顔照合と同様に，属性推定に必要となる DCNN の畳み込み層の層数を調査する．実験に使用する端末の性能及び DCNN のパラメータは，顔照合の実験と同じである．属性を推定する DCNN の学習サンプルは 4,031 枚，評価サンプルは 143 枚である．

図 7 に DCNN の畳み込み層数を変化させた時の平均識別率を示す．器官点は真値から左右の目の距離の 10% 以内を正解，年齢は真値から ±5 歳以内を正解，笑顔度は真値から ±10 以内を正解とする．なお，笑顔度は他の手法で得られた笑顔度の値を真値とする．

属性推定では層が極端に少ない，もしくは層が多い際に精度が低下している．本検証実験から畳み込み層を 7 層とした場合の精度が最も高いことがわかる．これ以降は，畳み込み層を 7 層とした DCNN を用いて検証実験を行う．

属性推定におけるロボット側で処理する畳み込み層数を変えた際のデータ通信量とクライアントの処理時間は，顔照合と同じ DCNN のパラメータを使用しているため，図 5 と同じ結果となる．顔照合と同様に，2 層目の畳み込み層 (Conv2 + Pool1) で切り離すとロボット側での処理時間が短く，かつデータ送信量が少なくなる．

図 8 に切り離す畳み込み層を変えた時の総処理時間を示す．顔照合の際には，クラウドサーバにてユークリッド距離の計算と比較の処理時間を含んでいたが，属性推定では距離計算と比較の代わりに認識処理の時間を含む．図 8 より，属性推定においても顔照合と同じ傾向が得られていることがわかり，2 層目の畳み込み層で切り離した構成の総処理時間が最も短い．クラウドサーバで行う認識処理時間は，どの層で分割しても変わらないため，総処理時間はロボットの処理時間と通信時間が大きく影響していることが確認できる．図 9 に提案手法により属性推定した結果を示す．顔照合や属性推定の結果を用いることで，ロボットは人に対して自然なコミュニケーションを図ることが可能となる．

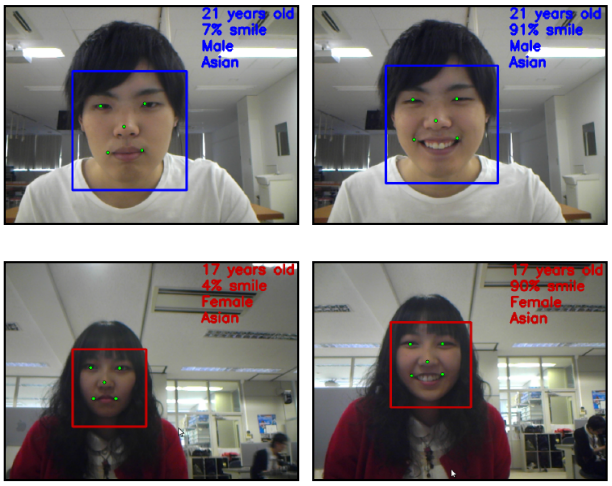

図 9　顔画像解析の結果.

5.　おわりに

本稿では，生活支援ロボットのためのクラウドロボティクスとそのプロトタイプシステムについて述べた．生活支援ロボットに期待されているタスクと必要技術について整理し，そのタスクを実現するための人を観る，物を観る技術について述べた．そして，クラウドロボティクスのための人を観る認識エンジンとして，人の顔画像から人物の照合及び属性推定 (器官点，人種，年齢，性別，表情等) を実装した．本認識エンジンでは，ロボットからクラウドサーバへ送るデータとして，顔画像の代わりに DCNN から出力される特徴マップを送信する．特徴マップは暗号化されたデータといえるため，プライバシーに配慮することができる．さらに，DCNN をどこで区切るとロボット側の計算負荷及び通信負荷が小さいか検証した．

本稿では，人を観る技術として顔照合と属性推定を担う認識エンジンについて述べたが，今後は物を観る技術を担う認識エンジンの開発を行う予定である．

文　献

[1] 山本貴史，齋藤史論，橋本国松，池田幸一，“生活支援ロボット HSR の開発”，日本ロボット学会学術講演会，2012.
[2] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg, “Cloud-Based Robot Grasping with the Google Object Recognition Engine”, IEEE International Conference on Robotics and Automation, 2013.
[3] O. Zweigle, R. van de Molengraft, R. d’ Andrea, and K. Haussermann, “Roboearth: connecting robots worldwide”,

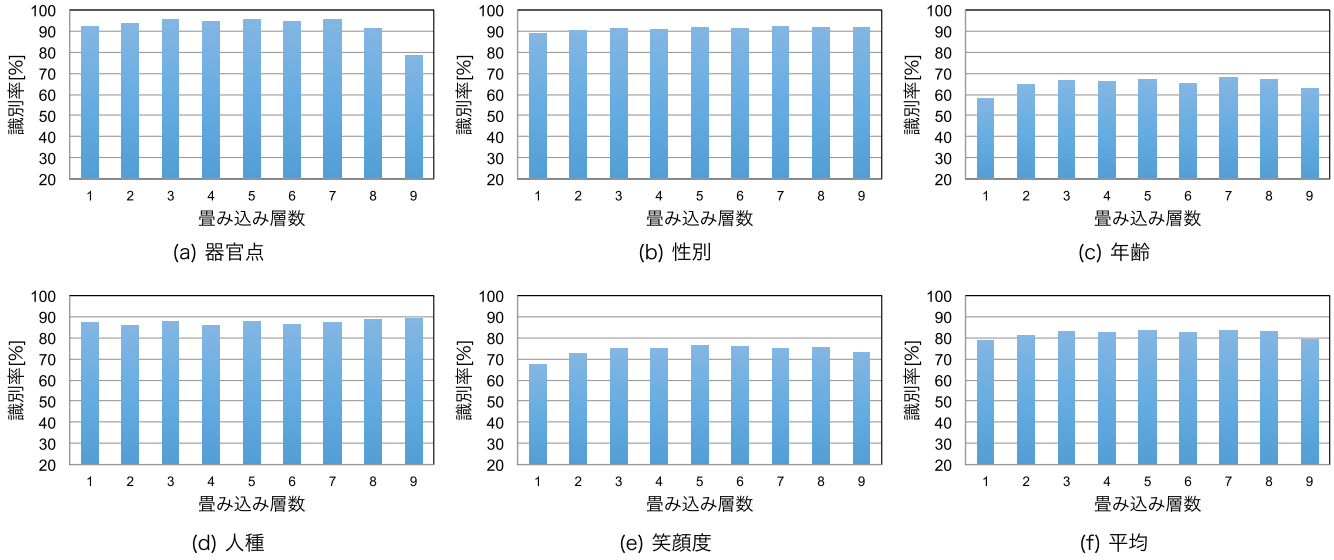

図 7　畳み込み層の数を変えた際の属性推定の認識精度.

International Conference on Interaction Sciences: Information Technology, Culture and Human. ACM, pp. 184-191, 2009.

[4] Markus Waibel, Michael Beetz, Javier Civera, Raffaello dAndrea, Jos Elfring, Dorian Galvez-Lopez, Kai Haussermann, Rob Janssen, J.M.M. Montiel, Alexander Perzylo, Bjoern Schiessle, Moritz Tenorth, Oliver Zweigle and M.J.G. (Rene) Van de Molengraft. RoboEarth - A World Wide Web for Robots. In Robotics and Automation Magazine, IEEE, vol 18, no 2, pp 69-82, June 2011.

[5] W. J. Beksi, J. Spruth and N. Papanikolopoulos, "CORE: A Cloud-based Object Recognition Engine for Robotics", International Conference on Intelligent Robots and Systems, 2015.

[6] K. Sugiura and K. Zettsu, "Rospeex, A Cloud Robotics Platform for Human-Robot Spoken Dialogues", International Conference on Intelligent Robots and Systems, 2015.

[7] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 1998.

[8] 北岡哲子，工藤千秋，"Clinical Facial Expression Scale(C-Face) による早期認知症患者の表情特徴抽出 ー FACS の併用の有用性ー"，日本早期認知症学会論文誌，Vol. 6，No. 2，2013.

[9] 北岡哲子，岸太一，"臨床的顔表情評価法 C-Face(Clinical Facial Expression Scale) の開発 ー認知症患者における検証ー"，日本早期認知症学会論文誌，Vol. 6，No. 1，pp. 84-89，2013.

[10] M. Kimura, T. Yamashita, Y. Yamauchi, H. Fujiyoshi, "Facial Point Detection Based on a Convolutional Neural Network with Optimal Mini-batch Procedure", International Conference on Image Processing, 2015.

[11] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf. "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", Computer Vision and Pattern Recognition, 2014.

[12] Z. Zhang, P. Luo, C. C. Loy, X. Tang, "Facial Landmark Detection by Deep Multi-task Learning", European Conference on Computer Vision, 2014.

[13] X. Yang, K. Seyoung and E. P. Xing, "Heterogeneous Multitask Learning with Joint Sparsity Constraints", Advances in Neural Information Processing Systems, pp. 2152-2159, 2009.

[14] T. Yamashita, T. Nakamura, H. Fukui, Y, Yamauchi, H. Fujiyoshi, "Cost-Alleviative Learning for Deep Convolutional Neural Network-based Facial Part Labeling", Information Processing Society of Japan Transactions on Computer Vision and Applications, Vol.7, pp.99-103, 2015.

[15] 加藤優，福井宏，山下隆義，山内悠嗣，藤吉弘亘，"Heterogeneous Learning と重み付き誤差関数の導入による顔画像解析"，画像の認識・理解シンポジウム，2015.