

メディア工学の研究動向

田川 憲 男^{†2}, 田良島周平^{†2}, 吉田周平^{†3}, 河村 圭^{†4}, 多田昌裕^{†5}, 新井啓之^{†6}

1. まえがき

メディア工学に関わる近年の研究においては、基盤技術の急速な発展に加え、社会実装が進みつつある。その応用分野は、産業、医療、社会インフラ、エンタテインメントなど大きな裾野を見せている。本稿ではこういった動向の中から、いくつかの注目すべき技術を取り上げ、その概要および今後の展望を紹介していく。

Deep Learning(深層学習)を中心とするAI技術の発展はめざましく、基盤技術としてのみならず、自動運転等の有望な応用分野への実装が進みつつある。一方、空間の3次元情報を計測し、利用することは従来から重要な技術課題となっているが、通常のカメラで撮影した画像を用いる技術の着実な進展に加え、デブスカメラやLIDAR(Light Detection and Ranging)などの奥行き方向の距離を計測可能なセンサやカメラを活用する技術も整備されつつある。

AI技術の研究動向として、まず主として画像・映像を対象としたパターン認識技術について最近の注目技術を紹介する。また画像生成を中心に新たな裾野を見せているGAN(敵対的生成ネットワーク, Generative Adversarial Networks)を取り上げ、その研究動向や課題について述べる。次に、画像系列から対象の3次元形状を計測する技術として多視点ステレオ、Visual SLAMの動向、および3次元形状を表現する方式の一つである点群データについて、その符号化技術の標準化動向を紹介する。そしてこれら基盤技術の応用先として、運転支援のためのドライバモニタリング、道路などの社会インフラのモニタリング技術をとあげ、その動向を紹介する。

2. AI技術の動向(1) パターン認識

画像・映像のパターン認識を中心としたAI研究コミュニティの規模拡大は極めて著しく、CVPRやICCVといったトップカンファレンスの規模はこの2年で実質倍増した¹⁾²⁾。当研究会でも当該技術に関連する内容の発表は非常に多く、年次大会や冬季大会では常に全発表の3割～5割を占めている。研究動向やベンチマークのキュレーションを目的とした取り組み^{3)～6)}が国内外で数多く台頭したが、このことは最新研究を俯瞰的にキャッチアップすることの困難さが増し続けていることを端的に示していると言えよう。本章では、パターン認識技術の研究動向のうち、この2年で注目度が大きく向上した三つのトピック(Attention, Graph Neural Network, Self-Supervised Learning)について取り上げる。

2.1 Attention

Attentionは入力データの中で重視すべき部分を推定する機構の総称である。深層学習への適用は自然言語処理(Natural Language Processing: NLP)分野で先行して進められていた⁷⁾⁸⁾が、最近では画像・映像分野の深層学習へも広く応用されるようになった。典型的な用途の一つとして、畳込みニューラルネットワーク(Convolutional Neural Network: CNN)中間出力のリスケージングが挙げられる。例えばSqueeze-and-Excite⁹⁾やGather-Excite¹⁰⁾では、CNN中間出力の3次元テンソルを空間ないしはチャンネル方向に並べられた特徴集合とみなし、プーリングと多層パーセプトロンを組み合わせたモジュールへ入力して得られた出力で各特徴を重み付けする方法が提案されている。BAM¹¹⁾やCBAM¹²⁾では、同一モジュール内で空間・チャンネル方向のAttentionが共に考慮されるとともに、畳込み処理を導入することで性能向上が計られている。一般に、CNNにおける畳込みフィルタの受容野(receptive field)は入力に近い層ほど小さく、また各フィルタは独立している。上記のAttentionモジュールは、データの大域的な構造やフィルタ間の相互作用に基づき中間特徴を変換するアプローチの一つであり、既存のCNNアーキテクチャに容易に導入可能で、かつコンパクトなモジュールとして設計されることが特徴

^{†1} 首都大学東京 システムデザイン学部

^{†2} NTTコミュニケーションズ/東京大学 大学院情報理工学系研究科

^{†3} NEC データサイエンス研究所

^{†4} 株式会社KDDI総合研究所

^{†5} 近畿大学 理工学部

^{†6} 日本工業大学 先進工学部

"Media Engineering" by Norio Tagawa (Tokyo Metropolitan University, Tokyo), Shuhei Tarashima (NTT Communications, Tokyo/Graduate School of Information Science and Technology, The University of Tokyo, Tokyo), Shuhei Yoshida (NEC, Kanagawa), Kei Kawamura (KDDI Research, Inc., Saitama), Masahiro Tada (KINDAI University, Osaka) and Hiroyuki Arai (Nippon Institute of Technology, Saitama)

的である。すでにこれらのモジュールはセグメンテーション¹³⁾、超解像¹⁴⁾、再同定¹⁵⁾をはじめとするさまざまなタスクへ導入されており、性能向上が報告されている。

また最近では、NLP分野で大きな成功を収めているSelf-Attention⁸⁾を適用した報告^{16) 17)}も目立つ。Self-Attentionは、画像処理分野におけるノイズ除去フィルタとして代表的なNon-Local Meansの一形態であることを示す報告¹⁸⁾もあり、異分野をつなぐ架け橋としても大変興味深い。

2.2 Graph Neural Network

Graph Neural Network (GNN) はグラフデータを対象とする深層学習を指す。点群、骨格、物体間の関係性等、グラフを用いたデータ表現が自然であるケースはパターン認識分野に数多く存在するが、最近ではこれらのタスクに対しGNNを適用し、エンドツーエンドにモデルを学習する事例も多く見られるようになった。例えば、映像中の人物の動作を分類するタスクに対し、各時刻のフレームから推定された人物の関節点をノードとするグラフを構築し、関節点に対応する領域特徴から、グラフ構造を考慮した畳込み (Graph Convolution¹⁹⁾) を用いて動作分類のための大域特徴を抽出する方法が提案されている²⁰⁾。他にも、映像中の各フレームで検出された物体をノードとするグラフからGraph Convolutionで特徴抽出を行うことで、映像分類やシーングラフ生成 (ノードを接続するエッジの分類) といったタスクに対する性能向上も報告されている^{21) 22)}。点群²³⁾を入力としたタスクへの適用も多く、例えば、文献²⁴⁾では点群のセグメンテーションに対しGraph Convolution²⁵⁾およびGraph Pooling²⁶⁾が適用されている。

Graph Convolutionのアルゴリズムは数多く提案されているが、画像・映像を入力とするパターン認識タスクに対しては、GCN¹⁹⁾やGraph Attention²⁵⁾といった手法を直接適用するケースが現状多い。これらの手法は適用が容易である一方、グラフ構造に基づく情報伝播を単純化しているため、グラフが内包する高次の相互作用を必ずしも捉えきれていない。分野に適応した新たなアプローチを含め、高次のグラフ構造を捉える手法の導入が進めば、より複雑なタスクを精度良く解く手法の開発につながる可能性がある。

2.3 Self-Supervised Learning

Self-Supervised Learningは教師なし学習 (Unsupervised Learning) の一形態であり、実質的な人手のコストを要せず得られた教師情報を用いてモデルを学習する方法の総称である。このような手法の研究自体は以前から存在していたが^{27) ~ 30)}、深層学習モデルの学習に、人手でアノテーションされた大規模な教師データの構築が必要であるという課題が広く認知されたことに伴い、Self-Supervised Learningという呼称が近年急速に浸透した。

Self-Supervised Learningの特徴の一つとして、データへ人為的に (かつ、実質コストフリーに) 適用した変換自体を推定することでモデルの学習を行うという点が挙げられる。例えば、ランダムに画像を回転させ、その回転を推定するようCNNの事前学習を行う手法が提案されており^{31) 32)}、タ

スクによってはラベル付きデータでの学習に匹敵する性能が得られることが報告され話題になった。また、グレースケール変換した画像を入力として元の色を推定するようモデルを学習するというアプローチも存在し²⁸⁾、時系列性を考慮して色復元を行うよう学習されたモデルが、物体追跡タスクへ直接転用可能であることを示唆する報告もある³³⁾。さらには、入力された映像と音声が同期しているか否かを分類するよう学習したモデルを事前学習し、マルチモーダル行動認識や映像中の音源特定といったタスクに適用する手法も提案されている^{34) ~ 36)}。CNNの事前学習の他にも、カメラポーズや奥行きを推定するモデルを教師なし学習することを目的として、推定されたパラメータや3次元情報から画像ペアの一方を変換し、他方との見えの差異を損失関数として用いる方法^{37) ~ 39)}が複数提案されている。

これらのSelf-Supervised Learning手法はいずれも非常にシンプルであり、故に適用可能な余地も依然多く残されていると言えよう。さまざまなタスクへの適用が進むことで、その有効性は今後より広く認知されることになるかもしれない。

3. AI技術の動向 (2) GANの研究動向

Generative Adversarial Networks (GAN)⁴⁰⁾は深層学習に基づく生成モデルである深層生成モデルの一種であり、主に画像生成の手法として発展してきた。本章では、GANの基本アーキテクチャを概観した後、最近の発展を①敵対的学習の安定化、②生成画像の高解像度化、③画像生成以外への応用の3点に分けて紹介する。

3.1 GANのアーキテクチャ

Goodfellowらは2014年の論文³⁷⁾でGANを発表した。GANの目標は、分布 $p(x)$ からサンプルされた学習データを元に、 $p(x)$ に従うデータを生成する生成器 G を学習することである。 $p(x)$ からサンプルされた「本物」のデータと G が作成した「贋作」とを見分ける識別器 D を導入し、 G と D とのmin-maxゲームにより生成器の学習を実現するのが、GANの基本アイデアである。生成器のアーキテクチャにCNNを採用したDeep Convolutional GAN (DCGAN)⁴¹⁾の登場以降、GANによる画像生成の研究が本格化する。オリジナルのGANやDCGANは単一の分布 $p(x)$ から画像を生成するが、カテゴリーや属性を指定した条件付き確率 $p(x|c)$ から画像を生成するよう発展させた手法が提案されている^{42) ~ 45)}。代表的なConditional GAN (cGAN)⁴²⁾は、生成器と識別器の両方が属性の情報を用いる。cGANをはじめ多くの手法は属性が付与された学習データを使用するが、InfoGAN⁴⁵⁾は生成画像 $G(z; c)$ と属性 c の相互情報量を最大化することで、属性そのものを教師なしで学習することを可能にしている。

3.2 敵対的学習の安定化

GANの学習が不安定であることはGANの黎明期から指摘されてきた課題である^{40) 41) 46)}が、近年、理論的にも経験的にも多くの研究がなされ、状況は改善されてきている。安

定化の手法には、分布間の距離尺度を変更するもの^{(47) (48) (50)}、新たな正則化の提案^{(51) ~ (53)}、その他モデルや最適化手法の工夫などがある。

オリジナルのGAN⁽⁴⁰⁾で行うmin-max最適化は、実質的に学習データと生成データの分布のJensen-Shannon (JS) divergenceの最小化を実行しているとみなせる。これに対し、JS divergenceを含む一般的なdivergenceを最適化するアプローチを採ったのがf-GAN⁽⁴⁷⁾である。f-divergenceは凸関数fにより定まる距離尺度であるためfの選択に応じて無数に存在するが、その中でもPearson 2-divergenceのメリットについてMaoら⁽⁴⁸⁾が論じている。一方、理論的な考察から別の分布間距離尺度を導入するアプローチの代表例がWasserstein GAN (WGAN)⁽⁵⁰⁾である。Arjovskyら⁽⁴⁹⁾はJS divergenceの問題点を分析し、二つの分布の台が一致しない場合に勾配法による学習が失敗することを示した。この問題を解消するため、Arjovskyら⁽⁵⁰⁾は生成分布をデータ分布と一致させる問題を最適輸送問題⁽⁵⁴⁾と捉え直し、Wasserstein計量を最適化するWGANを提案した。初期のWGANは識別器に対する強すぎる制約を課しており、後に改善されている^{(52) (53)}。

GAN学習の正則化は、識別器のLipschitz定数を制約するものが主流である^{(51) ~ (53)}。中でもSpectral Normalization (SN)は、損失関数の勾配計算以外に勾配計算が不要であり、ハイパーパラメータのチューニングも必要ないため、計算量が低く抑えられる。また、Zhangら⁽⁵⁵⁾は、生成器にもSNを適用することで学習が安定化することを示した。

その他、個々のサンプルではなくミニバッチ単位で識別を行うminibatch discrimination^{(46) (56)}、生成器と識別器で異なる学習率を使うTTUR⁽⁵⁷⁾、大きなミニバッチとモデルを使った学習⁽⁵⁸⁾などが提案されている。

3.3 生成画像の高解像度化

生成画像が高解像度になるほどGANの学習は識別器に有利となり、学習が困難になる。しかし、Progressive GAN⁽⁵⁶⁾、および同じ著者らによるStyle-based GAN^{(5) (9)}は、低解像度の生成から学習し、段階的に解像度を上げることで1メガピクセルの画像を生成することに成功した。Style-based GANでは生成器のアーキテクチャを大きく変更し、ノイズベクトル z を生成器の入力とするのではなく、AdaIN変換⁽⁶⁰⁾を介して生成器の中間層に挿入する。cGANの高解像度化としては、BigGAN⁽⁵⁸⁾が提案されている。BigGANは学習を安定化させる研究の集大成と言える。前節で紹介したものを含めたさまざまな方策を組み合わせることで、高解像度な条件付き画像生成を段階的な高解像度化を行わずに実現にした。高解像度化に成功したこれらの手法にも、生成画像のバリエーションが少ないなど課題があり、新しいアイデアが求められている。

3.4 画像生成以外への応用

GANの応用先は多岐にわたる。ここでは、画像変換^{(61) ~ (63)}、異常検知^{(74) (75)}、ドメイン適応^{(68) ~ (70)}に絞って代表的な例を紹介する。

画像変換への応用では、変換前後の画像ペアを使って変換を学習する従来手法に代わって、そのようなペアデータなしで変換を学習する手法^{(61) ~ (63)}が広く使われている。これらの手法では、順変換と逆変換を掛けて元に戻ること (Cycle Consistency) を課すことでペアデータなしでの学習を実現している。この考え方は音声変換⁽⁶⁴⁾、時系列表現学習⁽⁶⁵⁾、物体追跡⁽⁶⁶⁾など画像変換以外のタスクにも応用され広がりを見せている。

GANによる異常検知では、正常データのみからGANを学習する。このとき、推論時に与えられたデータが学習済みの生成器で生成できるものかどうか判定することで、異常を検知できる。AnoGAN⁽⁷⁴⁾やGANomaly⁽⁷⁵⁾はそのようなアイデアに基づいた異常検知アルゴリズムである。

一般に画像認識では、学習データの撮影環境と運用環境が異なると、運用時の精度が損なわれる。この問題に対処するのがドメイン適応 (Domain Adaptation) であり、学習データの特徴量分布と運用環境での特徴量分布を一致させる問題として定式化される⁽⁶⁷⁾。そこで、ドメイン識別器を導入して、特徴抽出器と敵対的に学習させるドメイン適応の手法が提案され^{(68) ~ (70)}、物体検知⁽⁷¹⁾やセマンティック・セグメンテーション⁽⁷²⁾にも応用されている。最近、ドメイン適応に対する理論的な考察から、ある種の敵対的学習が自然に導かれることが示される⁽⁷³⁾など理論的な進展も興味深い。

学習を安定化させる手法の発展を背景に、「GANは学習困難なもの」というイメージは払拭されつつある。実用を視野に入れた今後の研究動向、応用展開が注目される。

4. 多視点ステレオ・Visual SLAM

インターネット上にある多くの画像群から、あるいは時系列的に視点を変えた動画像から3D物体形状を復元する研究が、近年盛んに進められている。前者は多視点ステレオ (MVS: Multi-View Stereo) の枠組みで議論されることが多い。後者は自律走行技術の一つであるVisual SLAM (Simultaneous Localization and Mapping)が代表的であり、その基本原理はSfM (Structure from Motion) である。

MVSの最近の研究動向として、テクスチャレス領域の復元精度の低下を避ける二つの研究の流れがある。一つは、奥行値が同じ、あるいは滑らかな奥行値変化を持つ画素領域において、サポート領域を適切に決定することで、テクスチャ情報が不足している領域への奥行伝搬を行う考え方である^{(76) (77)}。従来は、固定された局所領域で奥行の平均化を行うことが多かったが、この方法では奥行不連続の保持が困難となるため、適応的にサポート領域を決定する必要がある^{(78) (79)}。一方で、奥行が連続的に変化する領域を適応的かつ効率的に決定することは困難な課題である。文献⁽⁸⁰⁾では木構造を利用する効率的な手法が提案されており、この考え方に基づく研究の今後の方向性を示すものと考えられる。もう一つのテクスチャレス領域への対処法は、SfS (Structure from Shading) の併用である。MVS、一般にStructure from

Stereoは、テクスチャや形状の凹凸により輝度変化が細かい領域に適しており、一方でSfSはテクスチャレス、すなわちalbedo一定で輝度変化が緩やかな領域に適用しやすい。テクスチャ情報が豊富な領域では、奥行に加えてalbedoの値も未知量とする必要があり、不良設定問題になりがちである。当初は、albedo一定^{81) 82)}、あるいはalbedo一定のいくつかの領域からなる対象に限定して⁸³⁾、輝度変化の特徴的な点にMVSを適用し、輝度変化の緩やかな領域に対してSfSを適用する手法が多く提案された。その後、albedoが空間的に変化する対象に対して、テクスチャ情報の多い画素や領域でMVSによって奥行値を決定したのち、テクスチャレスに近い領域にSfSを適用して、奥行のリファインメントを行う手法が提案されてきた^{84) 85)}。これに対し、文献⁸⁶⁾ではステレオとシェーディングの二つのキューを一つの評価関数の中に統合することで、両者を対等に扱い、自動的に相補的に機能する手法を提案している。シェーディングキューに関しては、輝度値の対数をとったものを処理することでalbedoの陽な扱いを不要にしている。形式的に美しくシンプルな手法となっており、今後の研究へのインパクトが強く感じられる。なお、当然のことながら、深層学習による方法も提案されてきており⁸⁷⁾、今後の更なる高性能化が期待できる。

Visual SLAM技術に関しては、これまでのアプローチの多くが静的な環境を対象としていたのに対し、これからは動的な環境、特に複数の移動物体が含まれるケースへの対応が必須である。自動運転での障害物検知、ヒューマン/ロボットインタラクション、協調ロボットや共同マッピング、AR応用など、多様なアプリケーションがそのようなケースに対応する。そのための技術ステージは、①ロバストなVisual SLAMの実現、②複数移動物体の分離と追跡、③複数移動物体の形状復元、などに分けることができる。①では、まず画像上での動きの分離が必要になる。そして、静的な対象の情報を使って、カメラ情報を獲得することになる。この動きの分離にもさまざまな手段が活用されている。移動体の情報が既知でありデータベース化されていれば処理しやすい⁸⁸⁾。背景モデルに基づく差分手法も提案されている⁸⁹⁾。より一般的には、フレーム間での幾何拘束の利用が適切であり、エピソード拘束からの乖離の情報⁹⁰⁾やカメラ情報の誤り⁹¹⁾、画像位置追跡後の不一致度⁹²⁾など、いくつかの評価が使用されている。オプティカルフローベースでの動き分離手法⁹³⁾、深層学習による方法も多数提案されている^{94) 95)}。3D復元での課題は、カメラ情報の推定にあたってのドリフトであり、一方で全フレームを使ったバンドル調整は計算コストの面で厳しいため、最後の少数のフレームを使う方法⁹⁶⁾や、キーフレームを検出して利用する方法⁹⁷⁾が提案されている。こちらについても、深層学習による方法が検討されているが、3D計算まではできておらず、奥行値推定にとどまっている。これらは教師あり学習と教師なし学習に分けることができ、後者はカメラ情報と奥行値から画像生成を行って計測画像との差を減ら

す学習が行われている⁹⁸⁾。②と③についての詳細は省略するが、特に重要となってくる技術は、3Dとしての物体分離と追跡の際のモデルフィッティングである。対象が一つのケースでも、この種の課題では特徴量の対応付けとそれに基づくモデルフィッティングを解く必要があり、特に動きや形状の異なる複数物体が存在する課題では非常に難しい問題となる。一般的には貪欲法、すなわちRANSAC等を使って順番に対象物のモデルを定めて除いていく方法が採用される⁹⁹⁾。これに対し、文献¹⁰⁰⁾では非常に興味深い方法が提案されている。特徴量のマッチングとモデルフィッティングを一つの評価関数にまとめて表現し、マッチングを表すバイナリー変数、モデル番号を表すラベル変数、およびモデルパラメータを一斉に決定する方法である。しかし、そのままではNP困難となるため、適切な近似最小化手法も提案している。この手法は、Visual SLAM以外にも多くの同様の課題に適用できる考え方を示唆しており、広い応用性を持っている。なお、カメラ画像に加えてIMU (Inertial Measurement Unit: 慣性計測装置) により計測される加速度・角速度を利用して、Visual SLAMの精度を高める手法 (VI-SLAM: Visual Inertial SLAM) も多数開発されている。それらは、バンドル調整等に基づきセンサ (カメラとIMU) の姿勢推定を行う最適化ベースの手法¹⁰¹⁾と、内部状態とするセンサ姿勢の時間更新にIMU情報を活用するフィルタベースの手法¹⁰²⁾に大別される。

以上、最近の多視点ステレオおよびVisual SLAMについての技術動向を整理した。この分野にも深層学習の波が打ち寄せていることはたしかである。従来からの数理モデルに基づく手法と深層学習によるデータ科学的な手法が今後どのような関係で進展していくのか、非常に興味深い。

5. 点群符号化の標準化動向

5.1 PCCの動向

本章では、Point Cloud Codingとよばれる点群情報の符号化を紹介する。Point CloudはAR/VRにおけるコンテンツ表現の一つとして注目されており、その符号化方式の検討が進んでいる。なお、Point Cloud自体の紹介は紙面の都合で割愛する。

MPEG-2やAVC、HEVCに代表される動画像符号化方式の国際標準化を検討しているMPEG (Motion Picture Experts Group) では、2017年からPoint Cloud Coding (PCC) の検討を進めており、2020年頃に規格化が完了する^{103) 104)}。具体的な対象データとしては、多視点カメラやデプスカメラなどから再構築された色情報を有する密な点群データと、LIDARデータのような大規模空間の疎な点群データが挙げられる。劣化を許容する場合 (ロッキー) と、劣化をほとんど、もしくはまったく許容しない場合 (ニアロスレス、ロスレス) に分けて異なる符号化方式が検討されている。前者はVideo-based PCC (V-PCC)、後者はGeometry-based PCC (G-PCC) と呼ばれている。いずれも符号化対象データは (x, y, z, R, G, B)、すなわち位置情報と属性 (色情報) となる。点群デー

タの例としてはMPEGとJPEGで整備しているPCCデータベースを参照されたい^{105) 106)}。

5.2 V-PCCの動向

点群データのエンタテインメント向けのユースケースにおいては、符号化にともなう誤差が許容される。そこで、点群を複数の適当な平面に射影し、テクスチャパッチとデプスパッチの集合に変換する。これらのパッチを1枚の大きなテクスチャ画像とデプス画像に統合して、通常の動画画像符号化方式によって符号化する。このように、実際にデータ量を削減する機能を、動画画像符号化方式に委譲することから、Video-based PCCと呼ばれる¹⁰⁷⁾。

以下に、変換の考え方を述べる。オブジェクトを包含する立方体を想定し、各面に点群を射影すれば大部分の点群を平面のテクスチャとデプスに変換できる。ただし、容易に想像されるように、いずれの方向からもオクルージョンとなる領域が存在しうる。そのような場合には、オクルージョン領域を含むより小さな立方体を仮定して、各面に射影すれば良い。一方で、ある点は複数の平面に射影される場合があり、冗長となる。そこで、必要最低限の平面情報への変換が必要となる。

実際に標準化されるのは、テクスチャ画像とデプス画像から点群を再構成する部分、すなわち復号方式のみである。符号化方法には自由度があり、その手順や性能はエンコーダ実装に依存する。ただし、標準化においては各社の提案を比較するために、符号化と復号の両方について共通のテストモデル(ソフトウェア)を提供している。

上記の方法により得られた符号化性能の例を紹介する。1024×1024×1024ボクセル30 fpsの精度をもつ人物点群の場合、1フレームあたり約800,000点から構成され、非圧縮時のデータ量としてはオーダとして1 Gbpsにもなる。これを上記のフレームワークにより圧縮すると、8 Mbpsにまで圧縮される。

5.3 G-PCCの動向

点群データのアーカイブ用途や先進運転システム向けのユースケースにおいては、符号化にともなう誤差がほとんど許容されない。そこで、点群の座標(Geometry)を最初に符号化し、次にその座標における属性(Attribute)を符号化する¹⁰⁸⁾。

まず、座標の符号化について述べる。浮動小数点表現の座標を十分に小さい間隔で均等量子化し、立方体内部の整数表現に変換する。得られた座標の符号化方法には2種類ある。Octree geometry codingでは、立方体を再帰的に八分木で分割し、小立方体内部の点の有無を8bitの符号語で表現する。ただし、点の個数や分布の偏りが少ない場合は、その座標を直接符号語で表現することもできる。これらの符号語は、当該小立方体のコンテキストを導出し、算術符号化によりビットストリームに変換する。Trisoup geometry codingでは、物体表面を構成する点群を三角形メッシュの集合として表現し、その頂点座標を算術符号化する。デコーダでは、メッシュが通るボクセルを座標として復号する。少なくとも

もOctree geometry codingはロスレス符号化が可能である。また、ロスレス・ニアロスレスの実現はエンコーダ制御に依存する。

次に、属性の符号化について述べる。複数の点を対象とした変換が3種類あり、領域適応階層変換(Regional Adaptive Hierarchical Transform: RAHT)と補間の階層最近傍予測変換(Predicting Transform)、リフティング付き補間の階層最近傍予測変換(Lifting Transform)である。RAHTは通常の静止画像に適用する変換と同様に、一連の属性値をDC成分とAC成分に変換する¹⁰⁹⁾。Predicting Transformは、最初に与えられた点群を、粗い粒度から細かい粒度へ便宜的に分類し、粗い粒度から細かい粒度の順に属性を符号化する。ここで、細かい粒度の属性は粗い粒度の属性から平均値か近傍値という予測方法を適用し、予測差分値を量子化して符号化する。Lifting Transformは、Prediction Transformと予測方法が異なる。リフティング構造の導入により、予測差分値を一定割合で参照値に戻し、さらに適応量子化を組み合わせる。

5.4 さらなる情報源について

本章では、点群情報の符号化方式をMPEGにおける国際標準化動向として紹介した。特にV-PCCは高い圧縮率を実現する一方で、既存の動画画像符号化方式を利用しているため、モバイル端末のような環境でもリアルタイム復号・視聴が可能である。例えば2019年IBC(国際放送機器展)などでデモが展示されていた。なお、標準化における寄書(入力文書)は基本的に非公開となっているが、一部の文書は公開されている。また、必要なソフトウェアも合わせて公開されている。関心のある読者はこれらを確認されたい。

6. ビジョンベースのドライバモニタリング

情報通信技術や人工知能技術の急速な発展に伴い、自動運転技術の研究開発が目覚ましい勢いで進められている。自動車用運転自動化システムの定義としては、SAE InternationalのJ3016¹¹⁰⁾(日本語参考訳¹¹¹⁾)が広く用いられている。SAEの定義では、運転自動化システムは、システムが担う動的運転タスクの程度によってレベル0から5まで分類されており、わが国では縦方向(速度、車間距離)、横方向(操舵)の車両運動制御をシステムが担う一方で、運転環境の監視はドライバが担うSAEレベル1、2相当の運転自動化システムの普及が進みつつある。また、官民ITS構想・ロードマップ2019¹¹²⁾では、運転環境の監視もシステムが担うSAEレベル4相当の車両の市場化を可能にすることが目標に掲げられ、社会実装に向けた運転自動化システムの実証実験¹¹³⁾が各地で実施されている。

運転自動化システムの普及により、事故原因の大半を占めるヒューマンエラー¹¹⁴⁾による交通事故減少や交通渋滞の緩和に寄与することが期待されている。運転自動化システムによる交通渋滞緩和効果を検証する取り組みとして、飯田らはSAEレベル1の定速走行・車間距離制御装置(ACC: Adaptive Cruise Control)搭載車両の混在比率が異

なる交通流をドライビングシミュレータ（DS）上に再現し、混在比率が増加することにより生じる交通状況の変化を評価した¹¹⁵⁾。その結果、SAEレベル1相当車両の混在比率が高くなると、運転者の反応遅れに伴う車間距離の開き、およびそれに伴う減速の伝播が緩和され、交通流率が高くなることを報告している。

一方、運転自動化システムの実証実験では、路上駐車車両の回避や1車線区間での対向車両との離合時に運転自動化システムが車両走行を停止し、手動運転で回避する場面が発生するなど課題も明らかになりつつある¹¹⁶⁾。SAEレベル3以上の運転自動化システムは未だ開発途上の技術であるため、国土交通省の自動運転車の安全技術ガイドラインでは、個々の運転自動化システムの性能および用途に応じた運行設計領域（ODD: Operational Design Domain）を設定して走行環境や運用方法を制限し、ODDの範囲を外れた場合にはドライバに運転権限の委譲を行うことが定められている¹¹⁷⁾。すなわち、運転自動化システムにより車両運動制御から解放されたドライバは、生まれた余裕を運転環境の監視に振り向け、自動運転の継続が困難な場面ではシステムに介入することが期待されている。しかしながら、運転自動化システムがドライバの漫然運転を誘発するなど、自動化による新たな事故リスクが生起するおそれも指摘されている¹¹⁸⁾。

以上の背景のもと、運転自動化システム使用中のドライバが、運転権限の委譲を受けられる状態にあるか把握するためのドライバモニタリング技術の研究開発が盛んに行われている。従来は、車両操舵情報¹¹⁹⁾や、ステアリングに搭載した心電図計測センサ¹²⁰⁾、装着型の脈波センサ¹²¹⁾などを用いて、ドライバの眠気や漫然運転状態を検出しようとする取り組みが多くなされてきた。一方、近年はビジョンベースのドライバモニタリング技術の開発が車載機器メーカーを中心に盛んに行われている^{122) 123)}。これらビジョンベースのドライバモニタリング技術は、ドライバの単位時間当たりの閉眼時間割合（PERCLOS）や瞬目回数などに着目して覚醒状態の推定を試みるものである。また、CNNとRecurrent Neural Network（RNN）を併用することで、ドライバの閉眼情報だけでなく姿勢や動きに関する大局的な情報を抽出し、ドライバが運転に集中している状態か否かを判定する手法も提案されている¹²⁴⁾。

ドライバモニタリング技術は、運転自動化システム使用中のドライバ状態を計測する用途だけでなく、近年社会的問題となりつつあるドライバの健康状態や疲労に起因する事故の予兆検知、交通事故原因の中で最多を占めるドライバの安全不確認¹²⁴⁾による事故の未然防止などさまざまな用途に応用が可能であり、今後ますますの発展が期待される分野であると考えられる。

7. 社会インフラのモニタリング

道路、橋、トンネルをはじめとする社会インフラの多くは高度経済成長期に建設されている。この時期に使われた

コンクリートの寿命は海岸部などの厳しい環境下では50年程度、比較的良い条件下でも100年程度と言われている¹²⁵⁾。2012年に起こった笹子トンネルの天井板落下事故は社会インフラ維持管理の重要性が認識される大きな契機となり、2014年に国土交通省はトンネルや橋の管理者に5年に1度の点検を「定期点検要領」に従い実施することを義務付けた。2019年にはこの「定期点検要領」が改正され、管理者にとって大きな作業負担となっている「近接目視」を同等の効果を持つ別の方法で行うことを認めた。これにより車両やドローンなどに搭載したカメラで対象を撮影し、画像解析技術を利用して点検を効率化するなど、新しい点検方法への期待が高まってきている。インフラの維持管理は安全の確保という社会的な意義がある一方で、新たなビジネスとしても注目されている。2030年にはカメラ・画像解析のみでも400億円を超える市場規模となると予想され^{126) 127)}、このほか各種センサやドローンの活用、さらに周辺サービスも含め今後大きく広がっていくと予想される。

このような流れの中で、実際にインフラを保有し運用する主体となる自治体、企業が大学等と連携しながらさまざまな取組を進めている。以下本学会のメディア工学研究会で行われた講演を中心に最近のトピックを紹介する。

2019年2月の研究会においては2件の特別講演を実施した。NEXCO東日本からはSMH（スマートメンテナンスハイウェイ）プロジェクトへの取り組みが紹介された¹²⁸⁾。このSMHプロジェクトは高速道路の「点検・調査」、「分析・評価」、「補修計画策定」、「補修・修繕」といったメンテナンスサイクルに対しICT技術の導入や機械化等を行うことで長期的な道路インフラの安全・安心を確保していこうとする取り組みである。画像解析による点検作業の効率化、ドローンなどの無人飛行体による画像撮影などメディア工学と関連の深い技術も積極的に利用していく計画となっている。東京メトロからは土木構造物の保守、特にトンネルの維持管理におけるデータ活用の取り組みが紹介された¹²⁹⁾。東京メトロにおいても40年以上経過した構造物が65%を超えその老朽化が進む中で、検査の効率化と補修・保全計画の最適化の二つが大きな課題となっている。保守用車を用いてトンネル壁面の画像を撮影、蓄積しているが、ここから自動的に要注意箇所を自動抽出する画像解析技術、またこの結果と打音検査結果との照合、路線全体での変状箇所分布の可視化、分析等を組み合わせることで維持管理の優先度の指標を得ることの見込みを得ている。今後、実務への導入を進めるとのことであった。

インフラを保有、管理する自治体や企業が画像解析等の技術を検討、導入するにあたっては、Deep Learning（深層学習）等の最先端の解析技術や知見を持つ研究機関との連携が重要となってくる。メディア工学研究会においても、インフラを保有、管理する企業と大学の連携により行われた研究が多数発表されている。鉄道会社から供与されたデータを用いトンネル壁面の画像からひび割れや漏水といった変状箇所を深層学習技術を用いて検出する研究^{130) 131)}や、

電力会社から提供されたデータを用いて送電鉄塔の劣化レベルを深層学習技術を用いて分類する研究¹³²⁾などを例として挙げることができる。

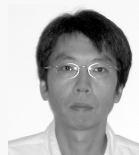
インフラの維持管理においては画像以外にもレンジセンサ、超音波センサ、歪センサ、加速度センサなど、その利用場面に応じてさまざまなセンシング手段が必要となる。上述したようなインフラを保有管理する企業と大学などの研究機関の連携に加え、情報センシング分野との交流、連携が進んでいくものと期待される。(2019年11月25日受付)

〔文 献〕

- 1) <https://github.com/hoya012/cvpr-2019-paper-statistics> (2019年10月10日参照)
- 2) <https://github.com/hoya012/iccv-2019-paper-statistics> (2019年10月10日参照)
- 3) <http://xpaperchallenge.org/cv/> (2019年10月10日参照)
- 4) <https://www.groundai.com/> (2019年10月10日参照)
- 5) <https://paperswithcode.com/> (2019年10月10日参照)
- 6) <https://sotabench.com/> (2019年10月10日参照)
- 7) D. Bahdanau, K. Cho and Y. Bengio: "Neural Machine Translation by Jointly Learning to Align and Translate", in ICLR (2015)
- 8) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin: "Attention is All You Need", in NIPS (2017)
- 9) J. Hu, L. Shen and G. Sun: "Squeeze-and-Excitation Networks", in CVPR (2018)
- 10) J. Hu, L. Shen, S. Albanie, G. Sun and A. Vedaldi: "Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks", in NeurIPS (2018)
- 11) J. Park, S. Woo, J.-Y. Lee and I.S. Kweon: "BAM: Bottleneck Attention Module", in BMVC (2018)
- 12) S. Woo, J. Park, J.-Y. Lee and I.S. Kweon: "CBAM: Convolutional Block Attention Module", in ECCV (2018)
- 13) H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi and A. Agrawal: "Context Encoding for Semantic Segmentation", in CVPR (2018)
- 14) Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu: "Image Super-Resolution using Very Deep Residual Channel Attention Networks", in ECCV (2018)
- 15) W. Li, X. Zhu and S. Gong: "Harmonious Attention Network for Person Re-Identification", in CVPR (2018)
- 16) I. Bello, B. Zoph, A. Vaswani, J. Shlens and Q.V. Le: "Attention Augmented Convolutional Networks", in ICCV (2019)
- 17) J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang and H. Lu: "Dual Attention Network for Scene Segmentation", in CVPR (2019)
- 18) X. Wang, R. Girshick, A. Gupta and K. He: "Non-Local Neural Networks", in CVPR (2018)
- 19) T.N. Kipf and M. Welling: "Semi-Supervised Classification with Graph Convolutional Networks", in ICLR (2017)
- 20) S. Yan, Y. Xiong and D. Lin: "Spatial Temporal Graph Convolutional Networks for Skeleton-based Action Recognition", in AAAI (2018)
- 21) X. Wang and A. Gupta: "Videos as Space-Time Region Graphs", in ECCV (2018)
- 22) J. Yang, J. Lu, S. Lee, D. Batra and D. Parikh: "Graph R-CNN for Scene Graph Generation", in ECCV (2018)
- 23) C. Wang, B. Samari and K. Siddiqi: "Local Spectral Graph Convolution for Point Set Feature Learning", in ECCV (2018)
- 24) L. Wang, Y. Huang, Y. Hou, S. Zhang and J. Shan: "Graph Attention Convolution for Point Cloud Segmentation", in CVPR (2019)
- 25) P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio: "Graph Attention Networks", in ICLR (2018)
- 26) C.R. Qi, L. Yi, H. Su and L.J. Guibas: "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space", in NIPS (2017)
- 27) C. Doersch, A. Gupta and A.A. Efros: "Unsupervised Visual Representation Learning by Context Prediction", in ICCV (2015)
- 28) R. Zhang, P. Isola and A.A. Efros: "Colorful Image Colorization", in ECCV (2016)
- 29) A. Dosovitskiy, J.T. Springenberg, M. Riedmiller and T. Brox: "Discriminative Unsupervised Feature Learning with Convolutional Neural Networks", in NIPS (2014)
- 30) D. Pathak, R. Girshick, P. Dollár, T. Darrell and B. Hariharan: "Learning Features by Watching Objects Move", in CVPR (2017)
- 31) S. Gidaris, P. Singh and N. Komodakis: "Unsupervised Representation Learning by Predicting Image Rotations", in ICLR (2018)
- 32) Z. Feng, C. Xu and D. Tao: "Self-Supervised Representation Learning by Rotation Feature Decoupling", in CVPR (2019)
- 33) C. Vondrick, A. Shrivastava, A. Fathi, S. Guadarrama and K. Murphy: "Tracking Emerges by Colorizing Videos", in ECCV (2018)
- 34) A. Owens and A.A. Efros: "Audio-visual scene analysis with self-supervised multisensory features", in ECCV (2018)
- 35) B. Korbar, D. Tran and L. Torresani: "Cooperative Learning of Audio and Video Models from Self-Supervised Synchronization", in NeurIPS (2018)
- 36) R. Arandjelovic and A. Zisserman: "Objects that Sound", in ECCV (2018)
- 37) T. Zhou, M. Brown, N. Snavely and D.G. Lowe: "Unsupervised Learning of Depth and Ego-Motion from Video", in CVPR (2017)
- 38) Z. Yang, P. Wang, Y. Wang, W. Xu and R. Nevatia, LEGO: "Learning Edge with Geometry All at Once by Watching Videos", in CVPR (2018)
- 39) DF-Net: "Unsupervised Joint Learning of Depth and Flow using Cross-Task Consistency", in ECCV (2018)
- 40) I. Goodfellow, et al.: "Generative Adversarial Nets", Advances in Neural Information Processing Systems 27, pp.2672-2680 (2014)
- 41) A. Radford, L. Metz and S. Chintala: "Unsupervised representation learning with deep convolutional generative adversarial networks", in Proceedings of the International Conference on Learning Representations (2016)
- 42) M. Mirza and S. Osindero: "Conditional Generative Adversarial Nets", arXiv:1411.1784 [cs.LG] (2014)
- 43) A. Odena, C. Olah and J. Shlens: "Conditional image synthesis with auxiliary classifier GANs", in Proceedings of the 34th International Conference on Machine Learning, 70, pp.2642-2651 (2017)
- 44) T. Miyato and M. Koyama: "cGANs with Projection Discriminator", in Proceedings of the International Conference on Learning Representations (2018)
- 45) X. Chen, et al.: "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", in Advances in Neural Information Processing Systems 29, pp.2172-2180 (2016)
- 46) T. Salimans, et al.: "Improved techniques for training GANs", in Advances in Neural Information Processing Systems 29, pp.2234-2242 (2016)
- 47) S. Nowozin, B. Cseke and R. Tomioka: "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization", in Advances in Neural Information Processing Systems 29, pp.271-279 (2016)
- 48) X. Mao, et al.: "Least Squares Generative Adversarial Networks", in Proceedings of the IEEE International Conference on Computer Vision, pp.2794-2802 (2017)
- 49) M. Arjovsky and L. Bottou: "Towards Principled Methods for Training Generative Adversarial Networks", in Proceedings of the International Conference on Learning Representations (2017)
- 50) M. Arjovsky, S. Chintala and L. Bottou: "Wasserstein Generative Adversarial Networks", in Proceedings of the 34th International Conference on Machine Learning, pp.214-223 (2017)
- 51) T. Miyato, et al.: "Spectral Normalization for Generative Adversarial Networks", in Proceedings of the International Conference on Learning Representations (2018)
- 52) I. Gulrajani, et al.: "Improved training of wasserstein GANs", in Advances in Neural Information Processing Systems 30, pp.271-279 (2017)
- 53) H. Petzka, A. Fischer and D. Lukovnikov: "On the regularization of Wasserstein GANs", in Proceedings of the International Conference on Learning Representations (2018)
- 54) C. Villani: "Optimal Transport: Old and New", Grundlehren der

- mathematischen Wissenschaften, Springer, Berlin (2009)
- 55) H. Zhang, et al: "Self-Attention Generative Adversarial Networks", in Proceedings of the 36th International Conference on Machine Learning, pp.7354-7363 (2019)
 - 56) T. Karras, et al.: "Progressive Growing of GANs for Improved Quality, Stability and Variation", in Proceedings of the International Conference on Learning Representations (2018)
 - 57) M. Heusel, et al.: "GANs trained by a two time-scale update rule converge to a local Nash equilibrium", in Advances in Neural Information Processing Systems 30, pp.6629-6640 (2017)
 - 58) A. Brock, J. Donahue and K. Simonyan: "Large Scale GAN Training for High Fidelity Natural Image Synthesis", in Proceedings of the International Conference on Learning Representations (2019)
 - 59) T. Karras, S. Laine and T. Aila: "A Style-Based Generator Architecture for Generative Adversarial Networks", in the IEEE Conference on Computer Vision and Pattern Recognition (2019)
 - 60) S. Huang and S. Belongie: "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization", in the IEEE Conference on Computer Vision and Pattern Recognition (2017)
 - 61) J. Zhu, et al.: "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks", in Proceedings of the IEEE International Conference on Computer Vision (2017)
 - 62) T. Kim, et al.: "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks", in Proceedings of the 36th International Conference on Machine Learning, pp.1857-1865 (2017)
 - 63) Z. Yi, et al.: "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation", in Proceedings of the IEEE International Conference on Computer Vision (2017)
 - 64) T. Kaneko and H. Kameoka: "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks", arXiv:1711.11293 (2017)
 - 65) D. Dwivedi, et al.: "Temporal Cycle-Consistency Learning", in the IEEE Conference on Computer Vision and Pattern Recognition (2019)
 - 66) X. Wang, et al.: "Learning Correspondence from the Cycleconsistency of Time", in the IEEE Conference on Computer Vision and Pattern Recognition (2019)
 - 67) S. Ben-David, et al.: "A theory of learning from different domains", Machine Learning, 79, 151 (2010)
 - 68) Y. Ganin, et al.: "Domain-Adversarial Training of Neural Networks", Journal of Machine Learning Research, 17, 59 (2016)
 - 69) E. Tzeng, et al.: "Adversarial discriminative domain adaptation", in the IEEE Conference on Computer Vision and Pattern Recognition (2017)
 - 70) K. Bousmalis, et al.: "Domain Separation Networks", in Advances in Neural Information Processing Systems 29, pp.343-351 (2016)
 - 71) K. Saito, et al.: "Strong-Weak Distribution Alignment for Adaptive Object Detection", in the IEEE Conference on Computer Vision and Pattern Recognition (2019)
 - 72) C.-Y. Lee, et al.: "Sliced Wasserstein Discrepancy for Unsupervised Domain Adaptation", in the IEEE Conference on Computer Vision and Pattern Recognition (2019)
 - 73) S. Ben-David, et al.: "A theory of learning from different domains", Machine Learning, 79, 151 (2010)
 - 74) T. Schlegl, et al.: "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", Information Processing in Medical Imaging pp.146-157 (2017)
 - 75) S. Akcay, et al.: "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training", in the 14th Asian Conference on Computer Vision (2018)
 - 76) MPEG-3DG: "MPEG Point Cloud Compression", <http://www.mpeg-pcc.org/>
 - 77) S. Sshwarz, et al.: "Emerging MPEG Standards for Point Cloud Compression", IEEE JETCAS, 9, 1 (2019)
 - 78) MPEG: "Point Cloud Coding", <https://mpeg.chiariglione.org/standards/mpeg-i/point-cloud-compression>
 - 79) JPEG: "JPEG Pleno Database", <https://jpeg.org/plenodb/>
 - 80) E.S. Jang, et al.: "Video-Based Point-Cloud-Compression Standard in MPEG: from Evidence Collection to Committee Draft", IEEE Signal Processing Mag., 36, 3 (2019)
 - 81) K. Mammou, et al.: "G-PCC codec description v2", ISO/IEC JTC1/SC29/WG11 N18189 (2019)
 - 82) R. Queiroz and P. Chou: "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform", IEEE Trans. Image Processing, 25, 8 (2016)
 - 83) S.A. Adhyapak, N. Kehtarnavaz, M. Nadin: "Stereo matching via selective multiple windows", J. Electron. Imag., 16, 1, Art. No.013012 (2007)
 - 84) H. Hirschmuller, P.R. Innocent, J. Garibaldi: "Real-time correlation-based stereo vision with reduced border errors", Int.J. Comput. Vis., 47, 1, pp.229-246 (2002)
 - 85) C. Richard, D. Orr, I. Davies, A. Criminisi, N.A. Dodgson: "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid", in Proc. Eur. Conf. Comput. Vis., pp.510-523 (2010)
 - 86) Q. Yang: "Hardware-efficient bilateral filtering for stereo matching", IEEE Trans. Pattern Anal. Mach. Intell., 36, 5, pp.1026-1409 (2014)
 - 87) W. Wu, H. Zhu, S. Yu, J. Shi: "Stereo matching with fusing adaptive support weights", IEEE Access, 7, pp.61960-61974 (2019)
 - 88) R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, A.M. Bruckstein: "RGBD-fusion: real-time high precision depth recovery", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.5407-5416 (2015)
 - 89) D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai, T.J. Cham: "Recovering surface details under general unknown illumination using shading and coarse multi-view stereo", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.1526-1533 (2014)
 - 90) L.F. Yu, S.K. Yeung, Y.W. Tai, S. Lin: "Shading-based shape refinement of RGB-D images", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.1415-1422 (2013)
 - 91) M. Zollhofer, A. Dai, N. Innmann, C. Wu, M. Stamminger, C. Theobalt, M. Niebner: "Shading-based refinement on volumetric signed distance functions", ASM Trans. Graph, 34, 4, pp.96:1-96:14 (2015)
 - 92) Q. Chen, V. Koltun: "A simple model for intrinsic image decomposition with depth cues", in Proc. IEEE Int. Conf. On Comput. Vis. (ICCV), pp.241-248 (2013)
 - 93) F. Laugguth, K. Sunkavalli, S. Hadap, M. Goesele: "Shading-aware multi-view stereo", in Proc. Eur. Conf. Comput. Vis., pp.469-485 (2016)
 - 94) J. Flynn, I. Neulander, J. Philbin, N. Snavely: "DeepStereo: Learning to predict new views from the world's imagery", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.5515-5524 (2016)
 - 95) Y.T. Wang, M.C. Lin, R.C. Ju: "Visual SLAM and moving-object detection for a small-size humanoid robot", Int.J. Adv. Robot. Syst., 7, 2, pp.133-138 (2010)
 - 96) M. Babaei, D.T. Dinh, G. Rigoll: "A deep convolutional neural network for video sequence background subtraction", Pattern Recognition, 76, pp.635-649 (2018)
 - 97) A. Kundu, K.M. Krishna, J. Sivaswamy: "Moving object detection by multi-view geometric techniques from a single camera mounted robot", in Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., pp.4306-4312 (2009)
 - 98) K.-H. Lin, C.-C. Wang: "Stereo-based simultaneous localization, mapping and moving object tracking", in Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., pp.4306-4312 (2010)
 - 99) D. Zhou, P. Tan: "CoSLAM: Collaborative visual SLAM in dynamic environments", IEEE Trans. Pattern Anal. Mach. Intell., 35, 2, pp.354-366 (2012)
 - 100) J. Klappstein, T. Vaudrey, C. Rabe, A. Wedel, R. Klette: "Moving object segmentation using optical flow and depth information", in Proc. Pacific-Rim Symp. Image Video Technol., pp.611-623 (2009)
 - 101) T. Qin, P. Li, S. Shen: "VINS-Mono: A robust and versatile monocular visual-inertial state estimator", IEEE Trans. Robotics, 34, 4, pp.1004-1020 (2018)
 - 102) M. Bloesch, M. Burri, S. Omari, M. Hutter, R. Siegwart: "IEKF-based visual-inertial odometry using direct photometric feedback", Int. Jour. Robotics Research, 36, pp.1053-1072 (2017)
 - 103) A. Dosovitskiy, P. Fischery, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van der Smagt, D. Cremers, T. Brox: "FlowNet: Learning optical flow with convolutional networks", in Proc. IEEE Int. Conf. On Comput. Vis. (ICCV), pp.2758-2766 (2015)

- 104) E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox: "FlowNet 2.0: Evolution of optical flow estimation with deep networks", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.2462-2470 (2017)
- 105) E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, P. Sayd: "Generic and real-time structure from motion using local bundle adjustment", Image Vis. Comput., 27, 8, pp.1178-1193 (2009)
- 106) G. Klein and D. Murray: "Parallel tracking and mapping on a camera phone", in Proc. IEEE Int. Symp. Mix. Augment. Real., pp.83-86 (2009)
- 107) T. Zhou, M. Brown, N. Snavely, D.G. Lowe: "Unsupervised learning of depth and ego-motion from video", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.1851-1860 (2017)
- 108) E. Vincent and R. Laganier: "Detecting planar homographies in an image pair", in Proc. IEEE Int. Symp. On Image and Sig. Process. Anal. (2001)
- 109) H. Isack and Y. Boykov: "Energy based multi-model fitting & matching for 3D reconstruction", in Proc. IEEE Int. Conf. On Comput. Vis. Pattern Recog. (CVPR), pp.1146-1153 (2014)
- 110) SAE International: "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles", J3016, 201806 (2018)
- 111) 自動車技術会: "自動車用運転自動化システムのレベル分類および定義", JASOテクニカルペーパー, JASO TP-18004 (2018)
- 112) 高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議: "官民ITS構想・ロードマップ2019" (2019)
- 113) 国土交通省: "道の駅等における自動運転サービス実証実験", <http://www.qsr.mlit.go.jp/n-michi/zidouunten/jikken.htm> (2019年10月18日参照)
- 114) 警察庁: "平成30年中の交通事故発生状況", (2019)
- 115) 飯田克弘, 和田泰明, 多田昌裕, 筑後智弘, 安時亨, 澤田英郎, 紀ノ定保礼: "ACC車両の混在比率が異なる交通流の安全性・円滑性評価", 交通工学論文集 (特集号A), 4, 1, pp.A_238-A_245 (2018)
- 116) 中山間地域における道の駅等を拠点とした自動運転ビジネスモデル検討会: "道の駅等を拠点とした自動運転サービス「中間とりまとめ」" (2019)
- 117) 国土交通省自動車局: "自動運転車の安全技術ガイドライン" (2018)
- 118) 和田脩平, 萩原亨, 浜岡秀勝, 二宮芳樹, 多田昌裕, 大廣智則: "高速道路における準自動運転車利用時の高齢および一般ドライバーの周辺認知に関する研究", 自動車技術会2018年春季学術講演会講演予稿集 (2018)
- 119) 久米拓弥, 内藤貴博, 石田健二, 河合政治, 松永真也, 西井克昌, 北島洋樹: "車載装置を利用した漫然状態の検出および解消手法の開発", 自動車技術会論文集, 45, 3, pp.567-572 (2014)
- 120) 中川剛, 河内泰司, 西井克昌: "ドライブ中の体調管理", 医療機器学, 83, 4, pp.374-378 (2013)
- 121) 中野泰彦, 佐野聡, 角中光, 岩男真由美: "心拍センサを用いた大型トラック運転中の覚醒度評価", 自動車技術会論文集, 46, 3, pp.679-685 (2015)
- 122) 大見拓寛: "画像センサによる眠気状態推定とドライバーステータスモニターの開発", Denso technical review, 21, pp.93-102 (2016)
- 123) 式井慎一, 砂川未佳, 楠亀弘一, 望月 誠, 北島洋樹, 下村義弘: "眠気検知・予測技術に基づくドライバーモニタシステム", パナソニック技報, 64, 2, pp.69-73 (2018)
- 124) 日向匡史, 木下航一, 西行健太, 長谷川友紀: "自動運転時代におけるドライバモニタリング技術", OMRON TECHNICS, 50, 1, pp.36-41 (2018)
- 125) 平成25年度国土交通白書
- 126) 渡本隆太: "連載・インフラを生かす (1) 国土省の一手で動き出す…インフラメンテ5兆円市場狙う異業種たち", 日刊工業新聞ニューススイッチ, <https://newswitch.jp/p/18774> (2019年10月1日参照)
- 127) 富士経済: "2018年版次世代インフラ維持管理技術・システム関連市場の現状と将来展望" (2018)
- 128) 水口和之: "高速道路におけるSMHの取組～データに基づく意思決定プロセスの確立へ～", 映情学技報, 43, 5, ME2019-27, pp.13-15 (2019)
- 129) 小西真治: "東京メトロにおける土木構造物の保守～トンネル維持管理でのデータ活用の取組～", 映情学技報, 43, 5, ME2019-28, pp.17-18 (2019)
- 130) 王安, 原川良介, 小川貴弘, 長谷山美紀: "畳み込みニューラルネットワークを用いた地下鉄トンネルにおける変状検出の高精度化に関する一検討", 映情学技報, 43, 5, ME2019-36, pp.121-124 (2019)
- 131) 豊田陽, 原川良介, 小川貴弘, 長谷山美紀: "レーザーデータを用いた地下鉄トンネル内の変状検出に関する検討～全層畳み込みネットワークを用いた変状領域の可視化～", 映情学技報, 43, 5, ME2019-55, pp.295-299 (2019)
- 132) 前田圭介, 高橋翔, 小川貴弘, 長谷山美紀: "異種特徴を用いた深層学習に基づく送電鉄塔の劣化レベル分類の高精度化に関する検討", 映情学技報, 43, 5, ME2019-68, pp.361-364 (2019)



田川 憲男 1989年, 東京工業大学大学院修士課程修了。同年, (株)富士通研究所入社。1992年, 東京都立大学工学部助手。1998年, 首都大学東京システムデザイン学部助教授。2007年, 同学部教授。現在に至る。コンピュータビジョン, 医用超音波工学の研究開発に従事。正会員。



田良島 周平 2009年, 東京大学工学部卒業。2011年, 同大学大学院新領域創成科学研究科修士課程修了。同年, NTT入社。現在, NTTコミュニケーションズ(株)技術開発部所属, 東京大学大学院情報理工学系研究科博士課程在籍。画像認識に関する研究開発に従事。正会員。



吉田 周平 2013年, 東京大学理学部物理学卒業。2015年, 同大学大学院理学系研究科修士課程修了。2018年, 同大学大学院理学系研究科博士課程修了。同年, NEC入社。現在, NECデータサイエンス研究所研究員。機械学習・画像認識に関する研究開発に従事。博士(理学)。



河村 圭 2004年, 早稲田大学理工学部電子・情報通信学科卒業。2005年, 同大学大学院国際情報通信研究科修士課程修了。2010年, 同大学大学院国際情報通信研究科博士課程修了。同年, KDDI(株)入社。現在, (株)KDDI総合研究所超臨場感通信グループ研究マネージャ。2017年, 本学会鈴木記念奨励賞等受賞。主に, 動画像符号化方式の研究・開発および国際標準化に従事。博士(国際情報通信学)。正会員。



多田 昌裕 2005年, 中央大学大学院理工学研究科博士後期課程修了。同年, (株)国際電気通信基礎技術研究所(ATR)入所。2013年, 近畿大学理工学部講師。2018年, 同大学准教授となり, 現在に至る。ドライバ行動解析, 運転支援システム等の研究開発に従事。博士(工学)。正会員。



新井 啓之 1989年, 東京理科大学理工学部物理学卒業。1991年, 北海道大学理学研究科物理学専攻修士課程修了。同年, 日本電信電話(株)研究所に入所。2001年～2006年, 情報通信研究機構ナチュラビジョンプロジェクト特別研究員。2017年より, 日本工業大学工学部情報工学科(現先進工学部情報メディア工学科)教授。画像処理, 画像認識技術の研究開発に従事。正会員。