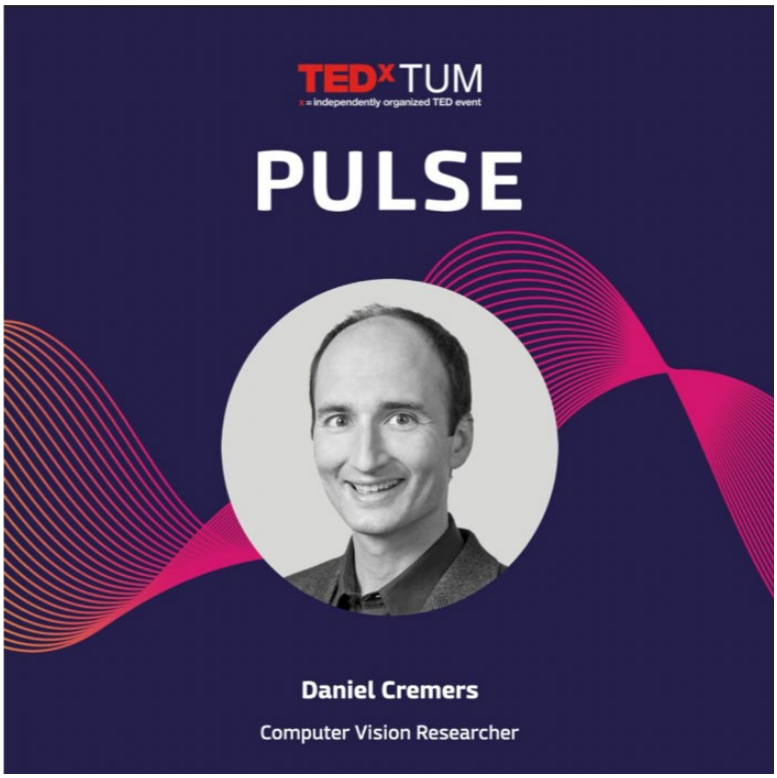
2021 年 12 月 9 日
Ｋｕｄａｎ株式会社

Artisense 創業者兼 CSO のダニエル・クレーマーズ教授が TEDxTUM にて登壇、プレゼン動画が公開されました

様々な場面で使用可能な SLAM 技術のリーティングプロバイダであるＫｕｄａｎ株式会社（本社：東京都渋谷区、代表取締役 CEO：項 大雨、以下 Kudan）は、ミュンヘン工科大学（TUM）の人工知能・コンピュータビジョン分野の首席教授であり、グループ会社である Artisense GmbH（以下 Artisense）の創業者兼 CSO（Chief Scientific Officer/最高研究責任者）のダニエル・クレーマーズ教授が TEDxTUM にて登壇し、そのプレゼン動画が公開されましたのでお知らせいたします。
なお、本動画は全編英語のため、プレゼン内容の和訳を別添資料として公開いたします。

「もし車があなたを助けてくれるほど賢かったらどうでしょう？人間と同じように世界を見ることができるとしたらどうでしょう？次に何が起こるかを理解し、予測できるとしたら？警告信号を送ってくれたり、あるいは積極的にブレーキを踏んでくれたりして、事故を回避し、命を救うことができるかもしれません。それを実現するのが、先進運転支援や自動運転車の開発です。」　－Artisense 創業者 兼 CSO　ダニエル・クレーマーズ教授

▼▼登壇動画は下記よりご確認いただけます（英語のみ）▼▼
https://youtu.be/vp8giEermfY



＜別添資料＞

どのようにして機械に 3D の世界を見せているのか

皆さん、このイベントには対話型の観客が多いと思います。では、手を挙げてください。次のようなことを想像してみてください。あなたが道を走っていると、どこからともなく何かがあなたの車の前に走ってきます。それは猫かもしれないし、犬かもしれないし、小さな子供かもしれません。皆さんの中で、このような状況に直面したことがある人はいますか？このような状況では、数分の一秒の反応時間が生死を分けることを実感していると思います。早朝でまだ目が覚めていないのかもしれません。あるいは、電話が鳴って気が散っているかもしれません。そうすると、人間はすぐには反応できないので、一瞬にして災害が起こってしまうのです。

（動画 1:02～）

しかし、もし車がそのような状況であなたを助けてくれるほど賢かったらどうでしょう？人間と同じように世界を見ることができるとしたらどうでしょう？次に何が起こるかを理解し、予測できるとしたら？警告信号を送ってくれたり、あるいは積極的にブレーキを踏んでくれたりして、事故を回避し、命を救うことができるかもしれません。それを実現するのが、先進運転支援や自動運転車の開発です。

今日は、この夢を実現するために私たちが行っているコンピュータビジョンの研究をご紹介したいと思います。コンピュータビジョンとは、画像やカメラからの情報をもとに世界を理解することです。しかし、人間と同じように世界を見ることを機械に教えるにはどうすればよいのでしょうか。あなたがこの世界を見るとき、あなたは車、歩行者、自転車などの物体を見て、その行動や関係を理解しています。

（動画 2:05～）

しかし、あなたの目が認識しているのは、実際の物体ではなく、網膜に当たっている色や輝度の値に過ぎません。ですから、人間と同じように機械が世界を見るためには、カメラだけでは不十分なのです。このギャップを埋めるためには、色のついたピクセルから、物体とその関係性という 3D の世界を理解することが必要なのです。

では、色のついたピクセルで構成された画像から、どうやって 3D の世界を理解するのでしょうか？一言で言えば、それはとてもシンプルです。

（動画 2:48～）

複数の画像の中で、色の一貫性に頼ればいいのです。ここでベートーベンを取り上げてみましょう。3D 空間の任意の点を考えて、「この点はベートーヴェンの顔の上にあるのか、ないのか？」と問いかけます。ベートーベンの目の中にある青い点のように、顔の上にあれば、異なる視点から見ても、同じ色や明るさに見えるはずです。一方、この赤い点のように、点が広い場所にある場合は、異なる視点から見たときに同じ色が見えないことが多いのです。このように、色や写真の整合性という概念を利用して、対象物にありそうな点とそうでない点を見分け

ることができるのです。あとはコンピュータに、すべての点が最大の光整合性を持つような3次元の表面を計算してもらえばいいのです。

（動画 3:50～）

簡単そうに見えますが、実はそうではありません。最も光整合性の高い曲面を見つけることは、膨大な計算上の課題につながることがわかったのです。その理由は以下の通りです。私たちのコンピュータは、子供たちがレゴブロックで何かを造形するように、観測された物体のモデルを再構築します。私にも3人の子供がいますが、レゴをやっていると必ずと言っていいほど時間がかかります。レゴブロックを組み合わせて作ることができる構造は、ほとんど無限にあります。実際、世界をたくさんの小さな体積要素、いわゆるボクセル（絵の要素を表す「ピクセル」と体積要素を表す「ボクセル」があります）に分割すると、それぞれのボクセルは「満たされている」か「空である」かの2つの状態を取ることができます。しかし、2つのボクセルを考えると、明らかに「充填」と「空」の組み合わせは4通りあります。そして、ボクセルが1つ増えるごとに、考えられる構成の総数は2倍になります。500×500×500のボクセルのグリッドを考えた場合、考えられる構成や3D再構成の総数は1に4,000万個のゼロを付けたものです。いくらコンピュータが各組み合わせを高速に評価しても、カメラの映像に最も適した組み合わせをすべて試すことはできません。仮に1つの構成をナノ秒で評価できたとしても、すべての構成を試すには宇宙の年齢以上の時間が必要になります。明らかに、このような完全な検索はダメです。では、どうすれば考えられる限りの最適な再構成を見つけることができるのか。何十年もの間、これはコンピュータビジョンの大きな計算上の課題の1つでした。

（動画 5:58～）

2009年、私たちは画期的な発見をしました。再構成問題は、凸最適化問題として表現できることを示したのです。つまり、考えられる再構成の空間を、ハイキングに行くときの風景のように思い浮かべてみてください。最適な再構成を見つけることは、その風景の中で最も低い場所を見つけるようなものです。もしこの風景が凸型であれば、最も低いところまで一歩一歩坂を下っていけばよいことになります。これと同じように、コンピュータは最初の3D形状から始めて、写真との整合性を高めるために段階的に変形させ、すべてのカメラの観測結果と最もよく一致する再構成を見つけます。

この方法により、16台のカメラを同期させて撮影した、縄跳びをする少女のアクションを再現することができました。縄跳びをしている女の子のようなアクションを、同期したカメラで撮影し、時間ごとに3Dで再構成することができます。しかも、その精度は非常に高く、ロープのような細かい部分まで忠実に再現されています。

（動画 7:38～）

さて、どうでしょう？なぜこのような画期的な技術が生まれたのでしょうか？それは、スポーツの分析から自由視点のテレビまで、さまざまな応用が可能になるということです。映画を見ているときに、視点を変えたり、アクションの中を移動したり、カメラがなかった角度から観

察したりすることができます。映画を見ているときに、視点を変えたり、アクションの中を移動したり、カメラがなかった角度から観察したりすることができます。

しかし、まだ自動運転車に使えるようなソリューションではありません。今回紹介したアプリケーションでは、通常、固定された調整済みのカメラがあり、そのカメラがどこにあるか、どの方向を向いているかを正確に把握しています。カメラを搭載した自動車やロボットが世界を移動する際には、「鶏と卵」のような厄介な問題に直面します。つまり、ある時点では、3D 再構成もカメラの位置もわからないのです。これは、3D 再構成もカメラの位置もわからないというもので、「Simultaneous Localization and Mapping（SLAM）」と呼ばれる長年の問題です。今から 100 年以上前、オーストリアの数学者アーウィン・クルッパは、2 つの画像に対応する 5 つの点（例えば、この画像には教会の塔が写っている、この画像には教会の塔が写っているなど）があれば、カメラの相対的な動きを復元できることを証明しました。この画像には教会の塔が写っている、この画像には教会の塔が写っている、などです。

（動画 9:33～）

コンピュータビジョンの分野では、何十年にもわたってクルッパの足跡を辿り、対応する点のペアという概念に基づいて SLAM 問題を解決するアルゴリズムを考案してきました。しかし、最初に点を抽出し、次に点の対応関係を計算し、カメラの動きと 3D 構造を再構築するというこの戦略では、最良の結果を得ることはできません。カメラが見ているのは、対応する点ではなく、色や明るさの値だからです。カメラが見ているのは、対応する点ではなく、色や明るさの値なのです。人が点を抽出するとき、必ず貴重な情報を捨ててしまいます。また、点の対応関係を計算する際には、SLAM 法の性能を低下させるようなミスが必ず発生します。

しかし、2014 年、私たちはこの長年の課題に新たなブレークスルーをもたらしました。カラー画像から直接、大規模な SLAM 問題を実時間で解決できることを初めて実証したのです。これは、先に述べた「写真の整合性」という概念を用いて実現しています。カメラフレーム間の色が最大限に一致するようなカメラの動きと世界の 3D マップを見つけます。このアルゴリズムを LSD SLAM（Large-scale Direct SLAM の略）と名付けました。今では最もポピュラーな SLAM 手法のひとつとなっています。

（動画 11:11～）

さて、次は何でしょう？ロボットや自動運転車などの自律システムにとって、自分自身の位置を正確に特定し、周囲の世界の大規模な 3 次元マップをリアルタイムに作成する能力は、シーンの理解、経路の計画、障害物の回避など、非常に大きな価値があります。その後、私たちはハイテク企業を設立し、より高い精度と堅牢性を実現するためにこれらの技術を進化させています。私たちは、複数のカメラ、人間の前庭システムに相当する慣性センサー、そして GPS からの情報を融合させました。トンネルや駐車場など、GPS の情報がない場所でも、数メートルの誤差で数キロメートル先の車の位置を把握できる精度を実現しています。さらに、すべてのセンサーからの情報を統合すると、誤差はほとんどなくなります。

カメラの動きを正確に推定できるようになったので、写真の整合性の概念を採用し、ニューラルネットワークを訓練して、動いている 1 台のカメラから仮想の 3D 世界を再構築することができます。

ニューラルネットワークは、人間が 1 枚の写真から物の 3D 形状を推定するように、画像から 3D 形状を生成する方法を多くの学習例から学びます。私たちは、ニューラルネットワークを訓練して、道路を 1 回走るだけで、観測された世界の忠実なコピーを生成するようにしました。これでかなり忠実なコピーができましたが、コンピュータは人間のように世界を理解することはできません。
私たち人間は、歩行者なのか、車なのか、自転車なのかなど、物の意味を理解することができます。

（動画 13:39～）
同様に、ニューラルネットワークを訓練することで、セマンティックリコンストラクションと呼ばれるものを作成することができます。ここでは、走行可能なエリア、歩道、車、歩行者、建物、植物など、さまざまなクラスのオブジェクトを区別することができます。カメラベースのシステムを全車両に導入すれば、あっという間に街のセマンティック・レコンストラクションを復元することができます。

さらに一歩進んで、この表現は高度な推論の基礎となります。環境中のさまざまな物体がどのように行動するかを予測し、人間のドライバーに警告を発することができます。あるいは、疲れていたり、気が散っていたりしても、安全に通行できるような適切な運転行動を生成することもできます。それはまた別の話ですが。ありがとうございました。

【Artisense Corporation について】
Artisense はコンピュータビジョンとセンサを融合したソフトウェア会社です。ロボット、車両、空間知能における様々なアプリケーションの自動化に向けて、カメラをリードセンサとして活用しながら、統合型のポジショニング・プラットフォームを開発しています。自律型ロボットや機械の普及の加速化に貢献することをミッションとして、Artisense は、あらゆる空間において、高精度でロバスト性に優れ、安全且つ低コストのナビゲーションを実現する製品と技術を提供しています。
詳細な情報は、Artisense のウェブサイト（http://www.artisense.ai/）をご参照ください。

【Ｋｕｄａｎ株式会社について】
Kudan（東証上場コード: 4425）は機械（コンピュータやロボット）の「眼」に相当する人工知覚（AP）のアルゴリズムを専門とする Deep Tech（ディープテック）の研究開発企業です。人工知覚（AP）は、機械の「脳」に相当する人工知能（AI）と対をなして相互補完する Deep Tech

として、機械を自律的に機能する方向に進化させるものです。現在、Kudan は高度な技術イノベーションによって幅広い産業にインパクトを与える Deep Tech に特化した独自のマイルストーンモデルに基づいた事業展開を推進しています。
詳細な情報は、Kudan のウェブサイト（https://www.kudan.io/?lang=ja）をご参照ください。

■会社概要
会　社　名：　　Ｋｕｄａｎ株式会社
証券コード：　　4425
代　表　者：　　代表取締役 CEO　項 大雨

■お問い合わせ先はこちら