

非同期処理がもたらす 次世代の画像処理と低消費電力 AI

東京大学 工学系研究科 特任講師・慶應義塾大学 理工学部 訪問講師
慶應義塾大学 理工学部 教授
名古屋大学 教養教育院教授・副院長
東京大学 工学系研究科 教授

芝 慎太郎
青木 義満
山里 敬也
川原 圭博

1. 緒言

画像や動画を処理する技術体系である画像処理（コンピュータビジョン）分野は、近年の人工知能（AI）の基礎として、工場自動化からスポーツ解析まで多岐に渡る産業応用を持つ（図 1）。画像処理 AI 技術による市場は 2030 年までに全世界で 20～200 兆円規模になると予測されているが[1]、画像処理研究のトップ国際会議の日本人採択率は 2.39%（2014 年）から 0.75%（2025 年）と減少傾向にあり[2]、現在の画像 AI 技術において日本は「創り出す」側ではなく「追いかける」側である。

日本からの新しいサービスや産業創出のための先端的な研究、すなわち「創り出す」リスクを取った研究がいつそう重要になる中、現在の画像 AI の発展に並行して、生物のような効率性、低消費電力性、頑健性、そして認識速度を達成する潜在性を秘めた、次世代の「非同期画像処理理論」の創出を、筆者らがリードしようと挑戦している。



図 1. コンピュータビジョン（画像処理）と人工知能（AI）の産業応用は多岐に渡る。

1.1 従来の画像処理：同期処理

2012 年[3]から現在までの画像 AI（深層学習）は、60 年以上の基礎研究、処理プロセッサ（CPU / GPU）や半導体技術の進化、カメラの質・価格などの技術革新の、様々な積み重ねによって支えられている。しかしながら、この流行によって様々な手法が成熟してきた結果、従来の画像処理システムに根本的に存在している以下のような課題の重要性が相対的に増してきた。

- 従来のカメラにおける手ブレや限られたダイナミックレンジ（カメラが捉えられるシーンの明るさの幅）による画像の劣化は、様々なシーンに対する頑健性が求められる現実世界のシステムにおいて大きな問題となる。
- 特に近年の深層学習手法は、高性能計算機による大量の電力と、メモリやデータ転送の帯域幅を圧迫し学習を肥大化させる大量のデータが必要になる。このような規模の拡大は限界に近づいており、より少ないデータや消費電力で計算する手法が必要とされている[4, 5]。
- 自動運転、AR/VR、ロボット等では、限られた消費電力や計算能力の中で高速に処理する必要がある、いかに遅延を減らし時間解像度を上げるかがシステムの精度に直結する（例えば、秒間 30 枚の動画では、1 枚の画像取得にかかる 33 ミリ秒の間に、車は 1 メートル、銃弾は 10 メートル移動する）。しかしながら、時間あたりの撮影画像数を増やすと（例：ハイスピードカメラ）、比例して処理するデータ量や電力も増大するため、時間解像度と、データ量・消費電力は本質的なトレードオフである。

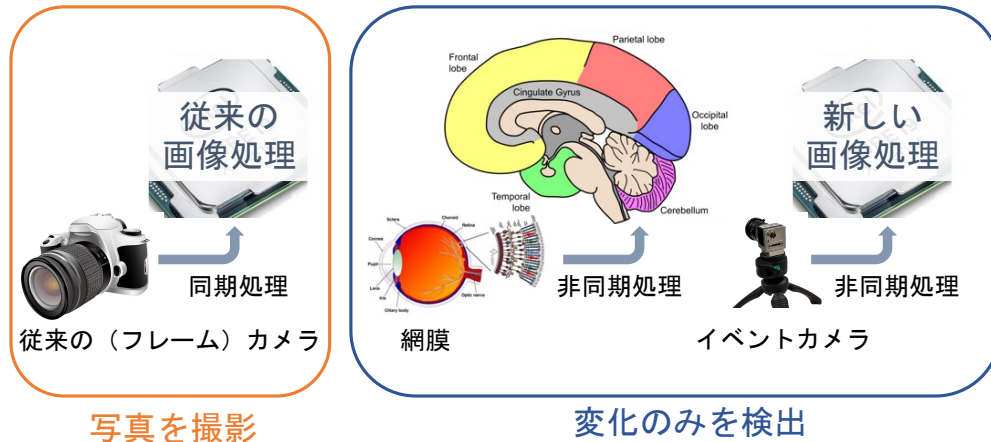


図 2. 従来のカメラは、実世界の状況に関わらず、常に決まった大きさの画像（写真）を記録する。生物の網膜やイベントカメラは、変化があった視野内の場所（画素）のみを記録する（実際のデータは図 3）。

1.2 新しい画像処理：非同期処理

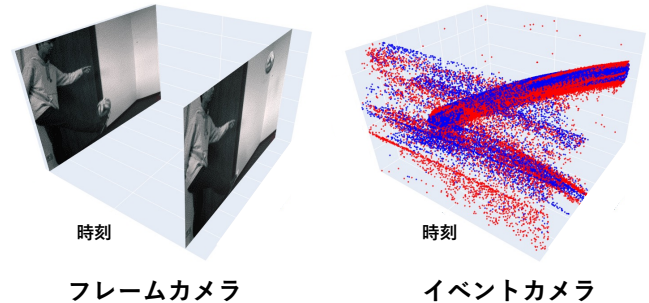
イベントカメラは、従来のカメラとその処理システムに存在する課題を根本から解決する可能性がある新しいカメラである。従来のカメラの課題は、全画素同時にデータ（フレーム）を記録する、すなわち同期処理のために生じる。イベントカメラは、フレームやシャッターという概念を持たず各画素が独立して変化（イベント）のみを検出する、非同期処理に基づくカメラである。結果として、ブレが最小になり、広いダイナミックレンジによって暗闇での撮影にも長け、電力とデータ効率が高く、時間解像度とデータ量のトレードオフを根本的に解決する、フレームベースの画像処理システムを置き換える新技术である（図 2）。

鍵となる「変化のみを捉えるため、効率的」というアイデアは、生物の網膜の仕組みに基づいている。網膜の視細胞では実際に差分しか検出しないため、我々の視覚システムでは低消費電力、省データ、高速処理を実現でき、ダイナミックレンジも広く、ブレもない。実はこの視細胞のモデル化は、ネオコグニトロンの発明で有名な福島邦彦先生が 1970 年に発表している、日本発のアイデアである[6]。その後イベントカメラは、1990 年前後にカリフォルニア工科大で半導体回路の研究として続けられ、やがてスイスに研究の中心を移して、大学発スタートアップから製品化がなされた。

しかしながら、イベントカメラから得られるデータは本質的にフレームとは異なる性質をもつため、カメラの性質を活用するために従来のコンピュータビジョンを再考しなければならない。具体的には、より連続的な（高解像度の）時刻を持つストリームをどのように扱うか（非同期性）、データが存在しない場合をどう処理するか（疎の性質）、変化のみから意味のある情報にどう変換するか（差分情報の性質）、効率的なデータに対して、アルゴリズムや処理システムとしての効率をどう実現するか（効率性）、といった新しい課題が存在する。実際に、筆者らの研究では、従来の画像処理理論を元にした手法を適用するよりも、非同期データの特性を活かす新規処理手法が圧倒的な性能向上をもたらす例が示されている（4 章）。

この論文では、このようにデータの特性と課題から考え直した、新しい非同期画像処理技術体系「イベントベースビジョン」の創出を、筆者らがいかに国際的にリードしてきたかを紹介する。2 章では、まずイベントカメラの概観と世界的な潮流について解説し、3～5 章では全て筆者らの成果を紹介する。3 章では、アプリケーションの例として、大学と企業の共同研究として進めてきた可視光通信について取り上げる。4 章では、非同期データ処理の基礎理論の魅力を、動き推定と三次元構造推定を代表的な例としながら、関連論文もまとめて紹介する。最後の 5 章では、オープンソースライブラリなどの研究以外の取り組みも紹介しながら、イベントベースビジョンの今後の発展に向けた課題と、日本の強み、将来展望について議論する。

代表者（芝）は、イベントカメラ製品化の直後、日本国内でいち早くイベントカメラの研究を始め、特許[7-10]や論文[11-28]を、大学と大企業の両方の所属から発表し、基礎研究と産業化を両輪で回してきた。3～4 章で紹介する論文は、2024～2026 年に出版・発表されたもの[8, 9, 11-22, 29-32]を中心としながらも、それ以外の近年出版したものも含めて、流れの中でまとめて紹介したい。一連の研究成果はトップ国際学会（CVPR, ICCV, ECCV）やトップ論文誌（T-PAMI）などに、口頭発表含め数多くの論文が採択され、世界的に業界をリードする若手研究者として認知されている。さらにトップ学会 CVPR, ECCV のイベントカメラワークショップや SSII（国内最大規模の画像処理学会）でも招待講演を行ったり[29, 30]、企業向け勉強会・研究会での講演[31, 32]、一般向けのニュース掲載[33, 34]などの活動も行い、様々なステークホルダーを巻き込みながら、ヨーロッパ・アメリカとの国際的な共同研究や、大学と企業両方の共同研究を主導している。この論文によって、画像 AI 研究を世界的にリードする環境を推進し、低消費電力知能システム実現のためのさらなる先進的な研究を加速する一助にできれば、幸いである。



2. イベントカメラの基礎

2.1 仕組み

イベントカメラは、画像内で一定以上の輝度変化があった画素のみが非同期的にイベントを出力する（図 3）。 k 番目のイベント $e_k = (\mathbf{x}_k, t_k, p_k)$ は、画素の位置 $\mathbf{x}_k$ 、時刻 t_k 、輝度変化の極性 $p_k = \{+1, -1\}$ （明/暗）からなる時空間の座標である。イベントはエッジ（例：サッカーボールの輪郭）の動きなどによって輝度に変化があった画素で疎に独立して発生し、発生数はシーンによって変化する。

従来のカメラと比較して、イベントカメラには次のような利点がある（表 1）。

①データ取得遅延は数マイクロ秒で、100～1000 倍の時間解像度（1000～10000 fps 相当）を持つためブレがない。また、静的なシーンはデータを出力しないため、②データ量を数 100 分の 1 に抑え、③電力消費も 10 分の 1 程度で済む。さらに、シャッターがなく輝度の差分だけに注目することから、④ダイナミックレンジが約 2 倍ある（暗闇や明るい場所での撮影に強い）。

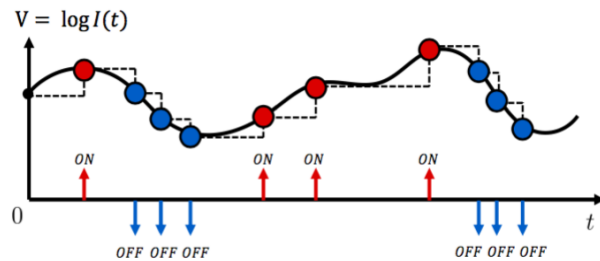


図 3. (上) 筆者がボールを蹴るシーンを撮影した、フレームカメラとイベントカメラのデータ比較。このシーンでは、ボールと人物のみが動いている。フレームカメラは、シーンによらず決まった間隔で画像（動画）を記録する。イベントカメラは、変化のあった画素（ボールと人物）のみデータを出力し、他の画素は作動しない。色が輝度の増加/減少を示す。(下) イベントカメラの撮像原理。時間（横軸）ではなく、輝度（縦軸）が一定間隔以上変化したときのみ、データを出力する。

表 1. イベントカメラと従来のカメラの性能比較。

	イベントカメラ	従来のカメラ
データ取得遅延	数マイクロ秒	数～数 10 ミリ秒
時間解像度	マイクロ秒以下	数～数 10 ミリ秒
データ量	40～180 kB/秒	10 MB /秒
消費電力	数 100 mW	数 W
ダイナミックレンジ	120～140 dB	約 60 dB

2.2 産業界と学术界のあゆみ

製品としては、スイスの大学発スタートアップ iniVation 社によって 2010 年頃に製品化したのち、2026 年現在では、フランスの Prophesee 社が市場の先駆者であり、ソニー、OmniVision、Samsung 等のイメージセンサ企業のほか、最近では米 DARPA（国防高等研究計画局）でのプロジェクトも採択されている。初期のカメラは研究試作品で、動作の安定性やデータ品質などに課題があり、解像度も低かったが、最近になるにつれて徐々に画素ピッチの小型化、高解像度化、低価格化が進んできている（図 4 上）。2026 年時点で最新のカメラは、100 万画素まで一般発売されており、価格も画素数に応じて～数 10 万円である。

現在まで、主に大学や企業の研究所で使用されており、実際の一般向け製品やサービスにはまだ搭載されていない。研究では、コンピュータビジョン、ロボティクス分野を中心に、最近ではリモートセンシングや光学的な内容も増えてきている。コンピュータビジョンでは、伝統的なフレームベースのカメラで解かれてきたような、物体認識、検出、セグメンテーション、動き推定、オプティカルフロー、深度推定、超解像などの問題に加えて、ブラー除去、画像の輝度復元、動画の高 fps 化などの新しい問題も研究されている。

応用分野としては特にロボティクス（フィジカル AI）において、高時間解像度・低消費電力が注目を集め、自己位置推定、SLAM、衝突回避、自動運転、ドローンやロボットの制御などの研究も多い。またイベントカメラは、時間解像度（マイクロ秒程度）と空間解像度（数マイクロメートル程度の画素ピッチ）の両方が高いため、システム全体として、光のパターンと組み合わせることでアプリケーションを実現する試みも多い。例えば、構造化照明や光切断法を用いると、対象の三次元構造が Lidar などを用いるよりも高速に計測でき（100fps 以上など）、空間解像度も高い。同じような仕組みで製品の外観検査に用いられた研究もある。全体としての論文数は増加傾向で、2023 年は年間 1000 本弱の論文が出ていると推計され、ヨーロッパと中国でも盛んに研究されている（図 4 下）。トップ学会では、2024 年頃から 1 分野として独立し始めており、学術誌でも特集号が盛んに組まれている。

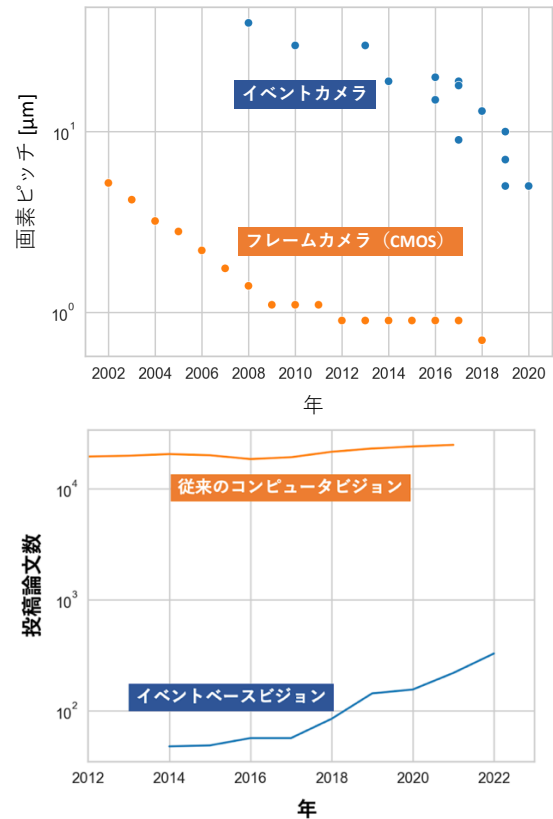


図 4. （上）従来のフレームカメラ（CMOS センサ）とイベントカメラの、ハードウェアとしての発展。ここでは画素ピッチの年次変化を取り上げる。イベントカメラの画素ピッチは従来のカメラと比べてまだまだ大きいですが、近年ムーアの法則に近い形で小型化し始め、カメラの解像度も急速に向上している（100 万画素）。（下）イベントカメラの研究論文数の変化。近年急激に増加している。

3. アプリケーション：可視光通信・光無線通信

2.2 章で紹介したアプリケーションの中でも、特に最近筆者らが取り組んできたのは、高い時間・空間解像度を活かした情報通信分野への応用である。近い将来に産業化が検討可能な技術として、大企業と大学との共同研究も推進しながら行なってきた。

可視光通信とは、LED の光を人の目には見えない程の高速に変調し、そのパターンを受信機で検知して情報を伝える技術であり、身の回りにある光を使用できる手軽さと、ITS

表 2. 可視光通信用のセンサ比較.

	時間解像度	空間解像度
フォトダイオード	～GHz	低い
フレームカメラ	～60Hz	1000 万画素
イベントカメラ	～MHz	100 万画素

(Intelligent Transport Systems: 高度道路交通システム) などを含むアプリケーションの可能性から注目されている。可視光通信の着想は以前から存在し、2000 年前後から日本を中心に盛んに研究されてきたが、フォトダイオードといった従来のセンサでは時間・空間解像度を満たすことができなかった(表 2)。可視光通信のアプリケーションとして期待される、中程度の容量の通信や、位置推定などのために、イベントカメラの時空間解像度が非常に有効であることから近年注目を集めている。

2025～2026 年にかけて行った、名古屋大学、トヨタ自動車株式会社、ウーブン・パイ・トヨタ株式会社などとの一連の共同研究では、効率的な符号化アルゴリズムや、複数の信号源からの位置推定なども含めて多くの論文を発表した[12-18]。まずは、ウォルシュ・アダマール符号化を用いることで、頑健かつ高速な通信が可能であることを示した。時速 50km の車に搭載されたカメラに対して、全体での遅延が 100 ミリ秒以下、最大で 85m 程度まで、26kbps 程度の帯域幅でエラーフリー通信が可能なが示されている[12]

(図 5)。また、複数の照明から、三角測量の要領で位置推定を行うことができるが、その性能を評価するための大規模実世界データセットも、世界に先駆けて公開し[18]、関連特許も複数申請済みである[8,9]。すでに、従来のフレームベースで、QR コードなどを使用した通信や位置推定よりも、より広範囲に、様々なシーンで通信できることが示されているが、今後のベンチマークや更なる性能改善の取り組みも期待されている。そのような照明システムを、実際の市街地の環境に応用すれば、都市の交通流最適化や、ロボットやドローンの自己位置推定に街の照明や信号機をそのまま使用することができる。これらの成果が評価され、筆者は 2025 年 6 月に、トップ国際学会である CVPR において、イベントカメラのワークショップでの招待講演や[29]、国内最大規模の画像系学会 SSII でも招待講演を行なった[30]。

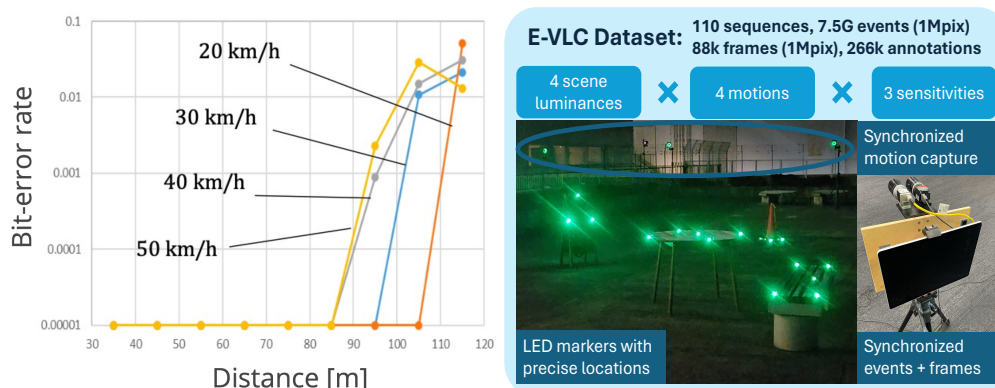


図 5 (左) 距離とカメラの速度に応じた、イベントカメラ可視光通信でのエラー誤り率[17].
(右) 位置推定の公開データセット[18].

4. 画像処理基礎理論の革新：時間情報の活用

アプリケーションではイベントカメラの利点（表 1）のいずれかに特化し際立たせるのが重要だが、理論研究ではイベントカメラ自体の汎用性を広げることを目指す。この章では特に、イベントデータの非同期性や高時間解像度に着目した、シーンやカメラの動き、さらに三次元構造の推定に関連した代表的な論文を紹介する。動き推定と言っても様々だが、代表的な推定目標であるオプティカルフローについて例示しながら、従来のフレームカメラを用いる場合と比べて高精度な推定ができる可能性があることが明らかになってきていることを示す（4.1 章）。さらに三次元再構成についても紹介し（4.2 章）、非常に様々な対象を推定するために、豊富な時間情報が鍵となっていることを示す。

4.1 動き（オプティカルフロー）推定

動物が世界を移動するとき、網膜に記録される光のパターンの動きはオプティカルフローと呼ばれる。画像処理におけるオプティカルフロー $\mathbf{v}(\mathbf{x})$ は、画素 $\mathbf{x}$ における画像平面上の速度である。フローはロボットが世界と相互作用するための基本的な情報であるほか、スローモーションビデオ合成などにも使用でき、複雑な動きも表現できるため、コンピュータビジョンにおいて非常に重要な推定目標である。

オプティカルフロー推定手法は一般に、（a）データの理論的性質から導き出すモデルベース手法と、（b）データの相関を学習させる機械学習手法（深層学習）に分類できる。従来のフレームベース画像処理における近年の（b）の成果は（a）の膨大な積み重ねの上に成り立っている。しかしながら、イベントベースのフロー推定では、その難しさから、基礎的な研究（a）が十分になされないまま、イベントを画像のように無理に変換して最近の機械学習を適用するものが多く、実世界の複雑なフローを正確に推定できなかった。筆者らの研究では、この（a）と（b）の手法にあったギャップを埋めるために詳細な分析をおこない[27]、イベントベースビジョンにおける本質的な理論貢献をおこなった[26, 28]。

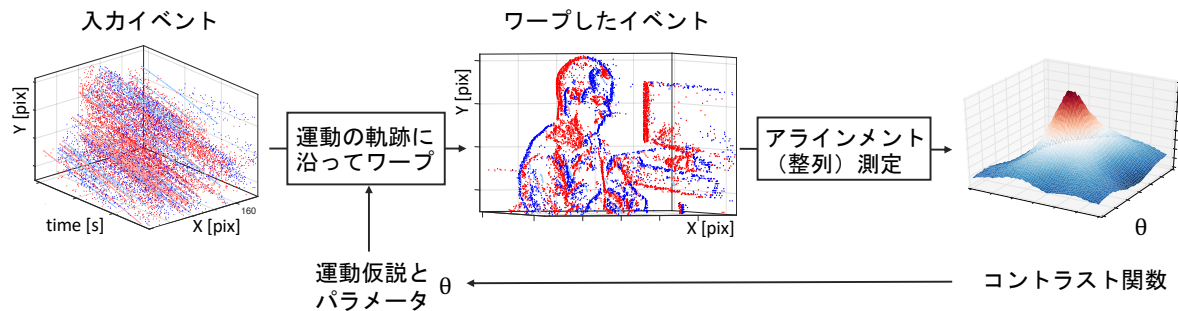


図6. コントラスト最大化の概要。入力イベントは時空間座標にまたがるストリームである。運動を仮定すると、画像平面上の画素の軌跡が計算できる。その軌跡に沿ってイベントを基準時刻にワープさせると、仮定した運動が正しければイベントが揃い（整列）、ワープしたイベントのヒストグラムは鮮明な画像になる。この鮮明さは画像コントラストによって計測できるため、コントラストを最大化することで運動を推定できる。

コントラスト最大化は、イベントデータの動き推定問題を解くための考え方である。これは、イベント $e_k = (\mathbf{x}_k, t_k, p_k)$ の集合がエッジの運動から生じると仮定し、同じエッジによって引き起こされたイベントの「揃い具合」（整列）を測定する目的関数を最大化するような最適化をおこなう（図 6）。直感的な解釈としては、イベントを引き起こしたエッジパターンの動きを補填し、鮮明な画像を復元できるような動きを推定する。動きの補填とは、推定した動きを用いて、各イベントを基準となる時刻 t_{ref} （例えば、最初のイベントの時刻）に空間的に移動（ワープ）することである。例えばオプティカルフローを推定する場合、

フロー $\mathbf{v}(\mathbf{x})$ を用いてワープしたイベント $e'_k \doteq (\mathbf{x}'_k, t_{\text{ref}}, p_k)$ は, $\mathbf{x}'_k = \mathbf{x}_k + (t_k - t_{\text{ref}})\mathbf{v}(\mathbf{x}_k)$ として得られる. 動き推定がもし正しければ, ワープしたイベントを画像平面上でヒストグラムにすると (図 6 中央), ワープしたイベントは整列し, 鮮明な画像になるだろう. 鮮明さ, すなわち画像コントラストを通してイベントの整列を測定するため, コントラスト最大化と呼ばれる.

しかし, コントラスト最大化をオプティカルフローの推定に適用しようとする, 多くのイベントを同じ画素上にワープ (縮退) させる望ましくない解に陥ってしまう. この問題のため, コントラスト最大化はフロー推定に適用できず, フロー推定の主な既存研究は, イベントを画像として表現して機械学習を適用してみる手法にとどまっていた.

この縮退問題を解決するために, 我々はまず基礎的な理論分析をおこなった[27]. この研究によって初めて, 複雑な動き推定問題において, コントラスト最大化の目的関数が多数の不良な最適解を持つことや, 機械学習モデルがデータに過適合していることが示された. さらにこの縮退問題を解決するために, 以下に示すような根本的な理論的解決を提案した[21, 26]. 特にイベントデータの非同期性と疎の性質に着目することで, 単純な手法が従来の学習手法の精度を全て上回り, また新たな学習の損失関数にも適用できるような原理的な考え方を導き出した.

1. 多時刻参照の目的関数による縮退の抑制. イベントの整列を 1 つの参照時刻 (例えば, 最初のイベントの発生時刻) でのみ計算すると, その時刻での縮退が起こる. しかし本来求めるべき動きは, 参照時刻を連続的に変化させても (例えば中間のイベントの発生時刻) 高いコントラストを生み出すべきである. イベントデータ特有の, 高い解像度の時間をより連続的に扱う 仕組みの必要性から着想を得て, 目的関数を多時刻参照にした結果, 縮退を大幅に改善した.
2. 時空間フローによる遮蔽の処理 (図 7). 従来のオプティカルフローの定義はもともとフレームベースのものであるため, フローは空間の関数 $\mathbf{v}(\mathbf{x})$ (2 枚の画像間の画素の変位) として定義される. しかし, これはイベントの非同期性と疎の性質を考慮していない. なぜなら, フレームと違ってイベントデータは基本的に同じ時刻をもたないからである. したがって, イベントベースのオプティカルフローは, 時間と空間の関数 $\mathbf{v}(\mathbf{x}, t)$ であるべきであり, イベント e_k もその時刻に合わせた時空間フロー $\mathbf{v}(\mathbf{x}_k, t_k)$ に従ってワープする必要がある. 筆者はこのような複雑なフローを偏微分方程式を介した二次元波の伝播問題として定式化し, 安定的に解けること, 遮蔽のあるような状況をよりよく処理できることを示した.

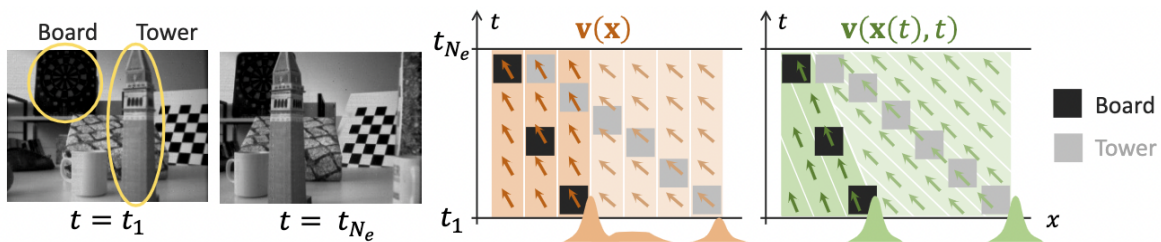


図 7. 時空間フロー. カメラが右方向に移動する例を考える. 塔 (カメラに近い方) が, 背景のダーツボードとの視差により遮蔽を起こしている. フレームベースの従来のフロー (赤) は, 画素ごとに一定のフロー $\mathbf{v}(\mathbf{x})$ を仮定し, 遮蔽を適切に処理することができない. 提案する時空間フロー (緑) は, 時空間の流線に沿って一定のフロー $\mathbf{v}(\mathbf{x}, t)$ を仮定し, より正確に遮蔽を処理できる. またデータの高い時間解像度によって, そのような複雑なフローも安定して推定可能であることを示した.

このような改善によるシンプルな提案手法は、フロー推定のベンチマークにおいて教師あり学習を含む他の 10 以上の最先端手法の中で最高の精度を達成した。

結果の図 8 の例では、イベントカメラのダイナミックレンジを活かして、従来のフレームカメラでは推定が難しかったような状況でもフローが推定できている。また、提案手法は微分可能であるため教師なし学習 (DNN) に移植可能であり、その場合にも既存手法の結果を改善した。イベントベースビジョンにおけるフローは、一定間隔で撮影されたフレームではなく、イベントのストリームの分析によって得られるのである。

フローだけでなく、自己動き (カメラ自身の動き) 推定の推定[24]や、高速な推定手法[25]、フローを活用したノイズ除去[20]などでも多くの論文を出しており、一部研究成果は複数の一般向け科学ニュース等に掲載されている[33, 34]。

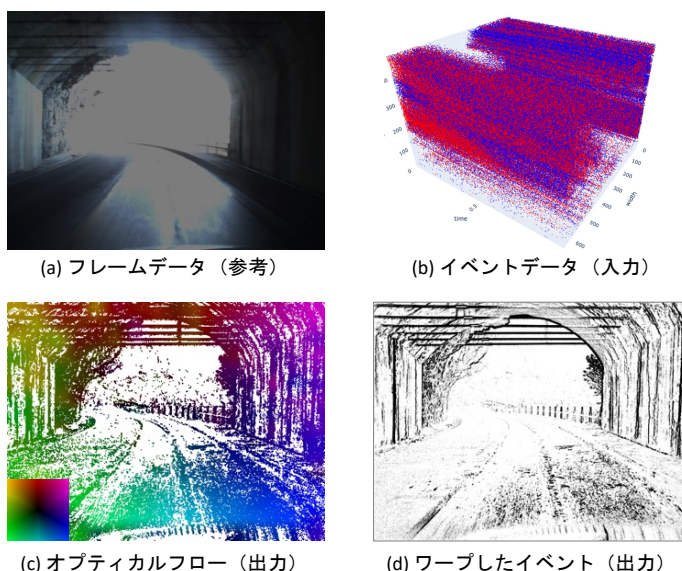


図 8. 車載カメラデータでの、イベントベースのオプティカルフロー推定結果. (a) 従来のカメラ (参考. 手法では使用されない) では、例えばトンネルでの撮影は難しい. 提案手法はイベントデータ (b) のみを入力とし、オプティカルフロー (c) を推定する (色が方向を示す). 副産物として (d) ワープしたイベント (動き補填された画像) も得る. 広いダイナミックレンジによりトンネル内部の様子や外部のポールなども見えている.

4.2 三次元構造の推定

動きが推定できると、一体何に役立つのだろうか? もちろん自動運転やドローンにとっては動きそのものの推定が重要だが、それ以外にも、近年流行している世界モデル (世界の三次元的な構造と意味、そしてその動きを理解するような基盤モデル) と、動きは実は密接に関係している。

例えば新幹線の窓から外を見た時に、遠くの山などはゆっくりと、近くの家などは早く動くように、外界の三次元構造と動きは密接に関連し、動きを正確に解くことで外界の三次元構造についての手がかりを得ることができる. この関係は、従来のコンピュータビジョンでは、Structure-from-Motion

(動きから構造へ) と呼ばれ、多視点幾何として体系的に理論化されている (図 9 上). 4.1 章で示したとおり、イベントカメラは微細な時間情報から動きを正確に捉えることができるため、連続視点幾何として

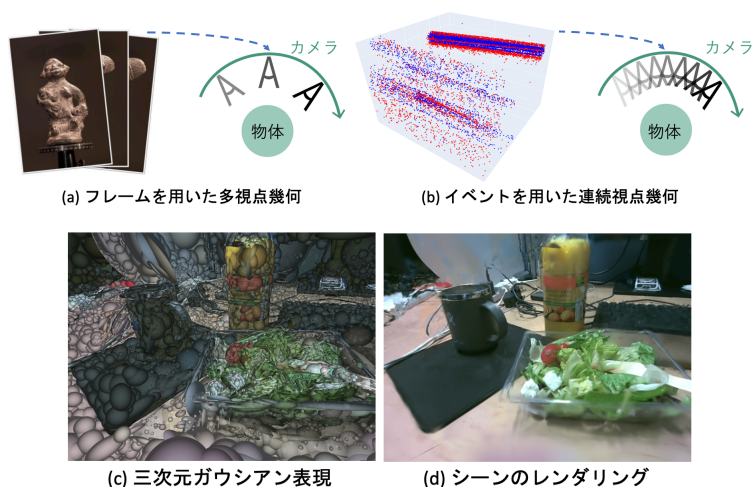


図 9. (上) フレームは時間的に離散のため、三次元再構成は多視点幾何を元に解かれるが、イベントは時間的に連続のため、連続視点幾何を考慮する必要がある. (下) 三次元ガウシアンは色や透明度のパラメータをもち、任意の三次元シーンをレンダリングできる.

理論を拡張でき、汎用性の高い三次元再構成に役立てられる。 また、モーションブレンダーが極めて少なく、ダイナミックレンジが広いため、従来のフレームベース三次元再構成における限界も克服できる。

筆者らの最新の研究 (2026) では、空間を三次元のガウシアン分布として表現し、そのガウシアンをイベントデータから直接再構成する手法を開発したことで画像 AI のトップ国際会議 CVPR に採択された [11]. ガウシアン分布を用いた三次元表現 (図 9 下) は、フォトメトリック (見た目が綺麗な) な三次元再構成および新規視点合成が高速かつスケーラブルに実現可能のため、例えば自動運転などで、世界モデルの大量の学習データを合成する際にもよく使われている。

イベントカメラの三次元ガウシアン再構成における従来の手法では、通常、あるシーンごとに2回のフレームレンダリングを実行する。これら2つのレンダリング画像の差分を取り、イベントの画素単位の集計によって得られるエッジ状の画像と比較することで、外観（見た目の）損失を算出してシーンを最適化する。しかし、この方法は全画素にわたる高密度なシーンレンダリングを2回行う必要があるため、学習速度を低下させるだけでなく、時間間隔の選択と推定精度のトレードオフという根本的な課題を生じる：2回のレンダリング間の時間間隔が短い場合、少数のイベントしか発生しないような微細な輝度変化を捉えられない。逆に時間間隔が長い場合、エッジ画像はブラーを生じ、観測されたエッジに含まれる細かい時間情報が失われてしまう。同様の問題は、全体的な照明環境と局所的な詳細構造のトレードオフとして先行研究でも指摘されていた。

そこで我々の研究では、この課題を根本的に解決し、高効率なレンダリングと高い時間分解能の活用を両立させつつ、1シーンにつき1回の画像全体の輝度レンダリングのみを行う、初のイベントベースGSフレームワークを提案した。提案手法の核心は、レンダリングにおいて、構造と外観を分離することにある(図10)。“構造”経路では、既知のカメラ動きとイベントごと(i.e., 疎)にレンダリングされた深度を活用し、コントラスト最大化(4.1章)に基づいた幾何損失を定義する。一方、“外観”経路では、レンダリングされたシーン輝度によってモデル化された瞬間的な明るさの変化と、イベントデータによって測定された変化との間の外観損失を定義する。要するに、連続時間で、空間的に疎なイベントベース3D再構成手法を提案した。

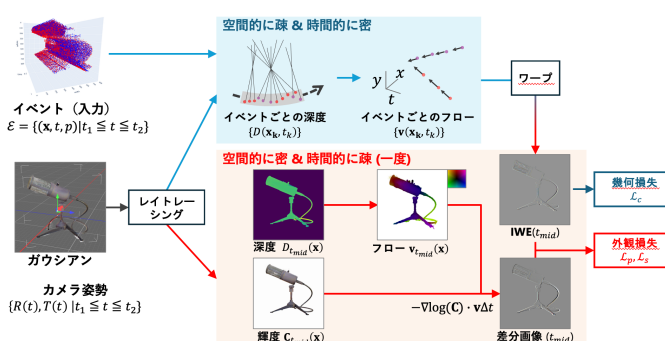


図 10. 提案手法の概要. レイトレーシングを用いて, 各イベントの深度を推定し, 補間されたカメラ姿勢からフローを計算する. イベントをワーブして, 中間時刻におけるワーブイベント画像 (IWE) を生成し, コントラスト損失を計算する (4.1 章も参照). 同時に, 画像全体の輝度をレンダリングし, 瞬時の輝度増分画像を計算することで, フォトメトリック損失に利用する.

実世界とシミュレーション両方のデータセットを用いた実験と評価によって、提案手法の以下の利点が示された．第一に、提案手法は事前知識(フレーム画像や深度・輝度の学習済みモデルなど)や外部手法による初期化を必要としない．第二に、カメラ姿勢やデータにノイズが含まれる実世界データセットにおいて、最高性能のレンダリング結果を達成し、最新の既存手法よりも学習時間が短い．第三に、1 シーンあたりに処理されるイベント数に対して頑健で、選び方によって精度を損なわない．

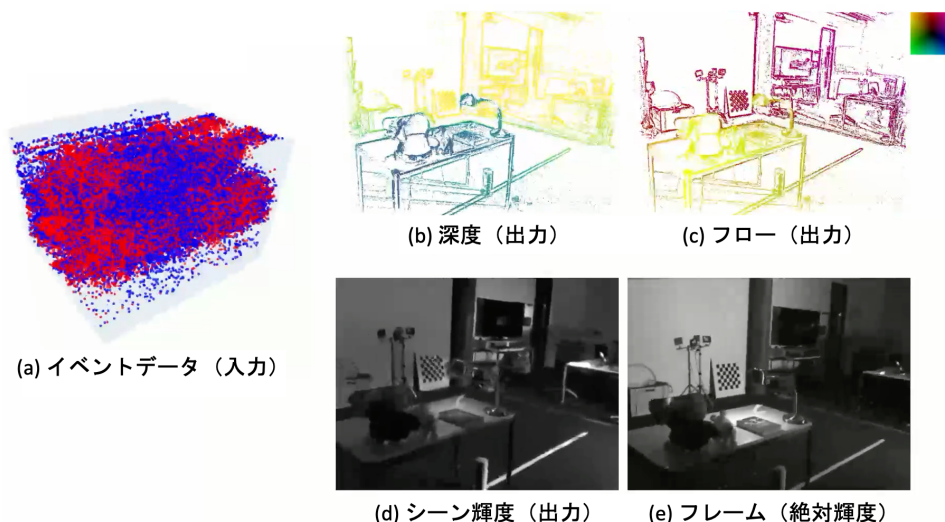


図 11. 提案したイベントベースの三次元ガウシアン再構成. イベントデータ (a) を入力とし, その時間解像度を正確に解析することで, 深度 (b) のような幾何的な情報 (色は深度), フロー (c) の動きの情報 (色はフローの向き), さらにシーン輝度 (d) の見た目の情報を再構成できることが示された. イベントという, 変化のタイミングのみからこれらが再構成できることは, カメラのデータにおける時間情報の持つ価値を示している.

図 11 に示すように, 本質的には変化のタイミングと極性 (バイナリ) のみであるイベントデータであるが, 時間情報を精密に解析することで, 深度のような幾何的な情報, フローのような動きの情報, シーン輝度のような見た目の情報まで推定できることは, 端的に言って驚きである. このような結果は, 豊富な時間情報が空間情報に変換可能であることを示唆しており, 世界モデルのような研究においても, 高い時間解像度のイベントデータが重要である可能性を示している.

5. 結論と展望

この論文では, まずイベントカメラについて紹介し, 世界的な潮流について説明した (2 章). アプリケーションは世界中で様々な研究と提案がされているが, その中でも, 筆者らが大学と企業の共同研究として取り組んできた可視光通信について紹介した (3 章). 短期的な市場拡大や事業化の上で, 大企業としての研究開発事例は非常に重要である. 一方でより長期的な視点での, 基礎理論の例として, 動き推定と, 三次元構造推定について紹介した (4 章). 特にデータの非同期性, 疎の性質, 効率性といった性質を追求し, カメラの利点を最大限活用するような理論貢献は, これまでのフレームベースシステムでは実現できなかった, イベントベースシステムの構築の基盤となる.

一方で, 研究としてのイベントベースビジョンは, まさにこれから作り上げられる段階にあるため, 様々な課題も存在する. 例えば, 汎用的で機能的なライブラリが存在せず, 研究者や開発者が新規参入しにくい. 筆者らは, 上記の研究成果のコードを公開する [35] だけでなく, イベントベースビジョン汎用オープンライブラリ [36] の開発を主導し, アメリカやドイツの学生と協業しているが, まだコミュニティ全体でのデファクト・スタンダードにはなっていない. 日本国内での研究開発コミュニティも重要であるため, 筆者らは国際的な発表だけでなく, 国内に向けた日本語での研究発表もおこなっている [13, 28]. 2025 年には, 国内学会 SSII だけでなく [30], イメージセンサ企業向けの勉強会でも登壇している [31, 32].

産業への普及という観点でも挑戦は多く, スマートフォン搭載 [37] などの試みもあるが, 実際に市場にまで至っているものはまだ少ない. その障壁の一つが, プロセッサの消費電力性能である. イベントカメラ自体の消費電力が低くても, そのデータを処理するのが GPU や CPU といった現在の汎用プロセッサである限り,

フィジカル AI のシステム全体としての消費電力を下げることはできない。したがって、本論文で示したような、非同期性などを生かすためのハードウェアを開発しながら圧倒的な低消費電力 AI を実現する取り組みも必要になるだろう。

様々な最新の研究成果から、イベントカメラは現在のカメラの一部を「ロボットの眼」として置き換え、新しいセンサ市場を開拓し始めている。製造業に強みをもつ日本で、画像処理や人工知能の基礎理論やアプリケーションまで含めた研究を行うのは、画像 AI の先端研究を垂直的に切り拓くチャンスである。そのような将来に向かって、60 年以上前に従来の画像処理理論が始まったように、イベントカメラのための理論体系「イベントベースビジョン」の構築に、今後も取り組み、世界をリードすることが非常に重要である。

謝辞

本研究の遂行に際し、研究に参加してくれた慶應・青木研究室と名古屋大・山里研究室の学生らに心より感謝申し上げます。共同研究者であるベルリン工科大学電気工学・コンピュータ科学学科 Gallego Guillermo 教授、トヨタ自動車株式会社の清水司氏、ウーブン・バイ・トヨタ株式会社の孔全氏、メルカリ株式会社の小堀訓成氏にも深く感謝申し上げます。本研究の一部は、DAAD（ドイツ学術交流会）二国間博士研究フェローシップ（57552338）、JST エッジ AI 半導体研究開発事業（JPMJES2513）の支援を受けて遂行されています。

参考文献

- [1] Markets and Markets: <https://www.marketsandmarkets.com/Market-Reports/ai-in-computer-vision-market-141658064.html>
- [2] Research Port: <https://research-p.com/column/2322>
- [3] ImageNet: <http://www.image-net.org/challenges/LSVRC/2012/>
- [4] “Revolutionizing artificial intelligence” by Geoffrey Hinton, <https://www.youtube.com/watch?v=2EDP4v-9TUA>
- [5] “From Model-centric to Data-centric AI” by Andrew Ng, <https://www.youtube.com/watch?v=06-AZXmWJjo>
- [6] Fukushima, et al. (1970). An electronic model of the retina. Proceedings of the IEEE 58.12 (1970): 1950–1951.
- [7] 芝 慎太郎. (2019). 撮像装置, 情報処理装置, 撮像方法, 情報処理方法及びプログラム, 特願 2019–103387
- [8] Shiba, S., Kong, Q., & Kobori, N. (2025). Optical wireless communication system and receiving device. US Patent US20250358005A1.
- [9] Shiba, S., Kong, Q., & Kobori, N. (2025). Receiving device. US Patent US20250379658A1.
- [10] 荒川 陸, 芝 慎太郎. (2018). 経産省 IPA 未踏アドバンス「イベントカメラ用高速画像処理ライブラリとアプリケーション開発」https://www.ipa.go.jp/jinzai/advanced/2018/gaiyou_f-1.html
- [11] Kohyama, K., Aoki, Y., Gallego, G. & Shiba, S. (2026). Geometric-Photometric Event-based 3D Gaussian Ray Tracing. CVPR 2026 (Accepted), June, 2026. (トップ会議)
- [12] Soga, R., Shimizu, T., Shiba, S., Kong, Q., Lu, S., & Yamazato, T. (2026). Real-world Latency Analysis of Vehicular Visible Light Communication with Multiple LED Transmitters and an Event-Based Camera, IEEE VTC Spring 2026 (Accepted), June, 2026. (トップ会議)

- [13] 曾我 亮太, 清水 司, 芝 慎太郎, 小堀 訓成, 路 跡, 山里 敬也. (2026). イベントカメラと複数光源を用いた車両用可視光通信システムの車載環境性能評価. 電子情報通信学会論文誌, Vol. J109-B, No. 5, pp. -, May. 2026.
- [14] Soga, R., Kobayashi, M., Shimizu, T., Shiba, S., Kong, Q., Lu, S., & Yamazato, T. (2025). Performance Evaluation of an Integrated System for Visible Light Communication and Positioning Using an Event Camera. 30th Asia-Pacific Conference on Communications (APCC), 1-6.
- [15] Soga, R., Shimizu, T., Shiba, S., Kong, Q., Lu, S., & Yamazato, T. (2025). Performance Evaluation of an Integrated System for Visible Light Communication and Positioning Using. IEICE Proceeding Series 97.
- [16] Kobayashi, M., Shiba, S., Kong, Q., Kobori, N., Lu, S., & Yamazato, T. (2025). Distance estimation in outdoor driving environments using phase-only correlation method with event cameras. 2025 IEEE Intelligent Vehicles Symposium (IV), 535-540. (トップ会議)
- [17] Soga, R., Shiba, S., Kong, Q., Kobori, N., Lu, S., & Yamazato, T. (2025). Evaluation of mobile environment for vehicular visible light communication using multiple leds and event cameras. 2025 IEEE Intelligent Vehicles Symposium (IV), 964-969. (トップ会議)
- [18] Shiba, S., Kong, Q., & Kobori, N. (2025). E-vlc: A real-world dataset for event-based visible light communication and localization. CVPR Workshop on Event Vision, 4877-4886.
- [19] Yamaki, R., Shiba, S., Gallego, G., & Aoki, Y. (2025). Iterative event-based motion segmentation by variational contrast maximization. CVPR Workshop on Event Vision, 4918-4927.
- [20] Shiba, S., Aoki, Y., & Gallego, G. (2025). Simultaneous motion and noise estimation with event cameras. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6959-6969). (トップ会議)
- [21] Shiba, S., Klose, Y., Aoki, Y., & Gallego, G. (2024). Secrets of event-based optical flow, depth and ego-motion estimation by contrast maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 7742-7759. (トップジャーナル)
- [22] Kohyama, K., Shiba, S., & Aoki, Y. (2024). 3d human scan with a moving event camera. CVPR Workshop on Computer Vision for Mixed Reality, 2024.
- [23] Shiba, S., Hamann, F., Aoki, Y., & Gallego, G. (2023). Event-based background-oriented schlieren. IEEE transactions on pattern analysis and machine intelligence, 46(4), 2011-2026. (トップジャーナル)
- [24] Shiba, S., Aoki, Y. and Gallego, G. (2023), A Fast Geometric Regularizer to Mitigate Event Collapse in the Contrast Maximization Framework. Adv. Intell. Syst. 2200251. doi:10.1002/aisy.202200251
- [25] Shiba, S., Aoki, Y. and Gallego, G. (2023), Fast Event-Based Optical Flow Estimation by Triplet Matching. *IEEE Signal Processing Letters*, vol. 29, pp. 2712-2716, 2022, doi:10.1109/LSP.2023.3234800.
- [26] Shiba, S., Aoki, Y., & Gallego, G. (2022). Secrets of event-based optical flow. In: Eur. Conf. on Comput. Vis. (ECCV), Tel Aviv, Israel, October 23-27, 2022. (トップ学会口頭発表, 採択率 2.7%).
- [27] Shiba, S., Aoki, Y., & Gallego, G. (2022). Event Collapse in Contrast Maximization Frameworks. Sensors, 22(14), 5190. doi:10.3390/s22145190
- [28] 芝 慎太郎, 青木 義満, Gallego Guillermo. (2022). イベントカメラを用いたオプティカルフロー推定: 動きとは何か?, ビジョン技術の実利用ワークショップ, 2022 (最優秀論文賞)

- [29] Shiba., S. “Real-world Event Camera Applications Towards Visible-Light Communication”, CVPR 2025 Workshop on Event-based Vision, Nashville, US, 06.2025. (招待講演)
- [30] 芝 慎太郎, “イベントカメラの研究紹介と可視光通信への応用”, 31th Synposium on Sensing via Image Information (SSII), Tokyo, 05.2025. (招待講演)
- [31] 芝 慎太郎, “イベントカメラによる動き推定”, 動体計測研究会 (企業研究会), University of Tokyo, Tokyo, 09.2024. (招待講演)
- [32] 芝 慎太郎, “Global Trends and Latest Research on Event Cameras”, 日本工業技術振興協会次世代画像入力ビジョン部会, Tokyo, 03.2024 (招待講演)
- [33] Bio-inspired robotic eyes that better estimate motion, *Advanced Science News*, Jan 27, 2023, <https://www.advancedsciencenews.com/bio-inspired-robotic-eyes-that-better-estimate-motion/> (News 掲載)
- [34] Keio Researchers Co-Develop Bio-Inspired Robotic Eyes, *The Penmark*, Feb 2023, <https://mailchi.mp/global/the-penmark-february-6264837?e=dab315f4f7> (News 掲載)
- [35] https://github.com/tub-rip/event_based_optical_flow, https://github.com/tub-rip/event_collapse, <https://github.com/e3ai/gpert>
- [36] <https://github.com/shiba24/event-vision-library>
- [37] <https://www.prophesee.ai/2023/02/27/prophesee-qualcomm-collaboration-snapdragon/>