

# イベントビジョンシステムによる視対象の運動推定法の検討

## Motion Estimation of Visual Target Using Event-Based Vision and Spiking Neural Network

○ 安川 真輔 (九工大) 吉松 秀峻 (九工大)  
正 石井 和男 (九工大)

Shinsuke YASUKAWA, Kyushu Institute of Technology, s-yasukawa@brain.kyutech.ac.jp  
Hidetaka YOSHIMATSU, Kyushu Institute of Technology  
Kazuo ISHII, Kyushu Institute of Technology

We proposed a method for estimating the angular velocity of a rotating visual target robustly to changes in illumination conditions using event-based vision and spiking neural networks (SNNs). By using the output of the event-based vision directly as input to the SNN, the image processing function can be achieved by keeping the advantages of the vision sensor in the previous stage. A board printed with natural images was placed at a certain distance from the vision system as a viewing target. The vision system captured rotating objects at various angular velocities under multiple lighting conditions to generate a data set. Using this data set, we verified whether the SNN can estimate the visual target. We achieved about 80% accuracy on both the training and test datasets in both bright and dark lighting environments.

**Key Words:** Event-based vision, Motion estimation, Spiking neural network

### 1. 緒言

ロボットが動的環境で行動する際や操作対象が動物体である時、視対象の運動推定を行う必要がある。不確定性の高い動的環境やハンドリングタスク、例えば移動障害物の動きやロボットが掴んだ紐状柔軟物の変形情報を推論できれば、ロボットが知的な行動を計画する際の手掛かりとなる。

そのような外界の動き情報を捉える際、露光を前提とする汎用カメラを用いたビジョンシステムでは主にモーションブレイカーの問題とダイナミックレンジの問題が避けられない。

近年、画素ごとの光のコントラスト変化情報のみを出力するイベントベースカメラが開発されており、その高ダイナミックレンジかつ高時間分解能の画素応答特性に注目を集め、ロボティクス分野への応用も検討され始めている。

本研究ではイベントベースカメラの出力をそのままスパイクニューラルネットワーク(SNN)で受けることで、カメラの利点を継承しつつ、画像処理機能が実現できるか検証する。異なる照明下で回転体の角速度を推論する課題を例に手法の確立を目指す。

### 2. イベントビジョンシステムの構成

本研究ではイベントベースカメラの後段にスパイクニューラルネットワークを配した構成をイベントビジョンシステムと呼ぶ。

イベントベースカメラは各々の画素に入射する光の輝度の変化を非同期で検出し、変化したデータを座標や時間情報と組み合わせて出力するセンサである。イベントベースカメラ

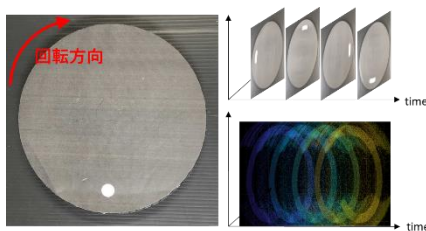


Fig. 1 Differences in output between general-purpose cameras and event-based cameras.

の特徴として一般的に汎用カメラに比べて、高ダイナミックレンジ、低遅延、低消費電力、そして高時間分解能であることが挙げられる。図1に回転する白い点に対する汎用カメラのとイベントベースカメラの出力の差を示す。汎用カメラでは撮影時に一定時間、対象からの光を露光する必要があるためモーションブレイカーが発生する(図1右上)が、イベントベースカメラではそれが生じていない(図1右下)。一方で、出力形式が特殊であるため、その特長を活かした後段の新規処理方法の研究開発が必要である。

### 3. Spiking Neural Network (SNN)による回転運動推定

本研究ではイベントベースカメラからの入力をSNNで受け、視対象の回転角度を推定する。SNN内で用いるスパイクニューロンとして、単純で扱いやすいスパイク応答モデル(SRM)を用いた[1]。ネットワーク構造はConvolution Neural Network (CNN)を採用し、具体的な層構造は図2に示すように、畳み込み層を5層、その間にプーリング層を4層、最後に全結合層とした。また学習方法としてSLAYERを用いた[2]。スパイク応答では連続値が表現できないため、全結合層に入る前に各ニューロンの応答は $\alpha$ 関数によって信号を時間方向に積分した。また損失関数は各時刻における真の角速度と推定値の2乗和誤差の和とした。

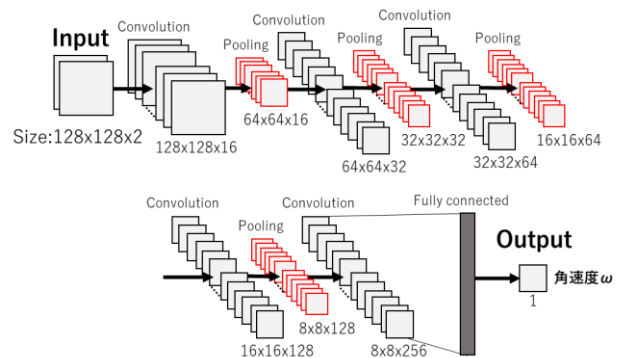


Fig. 2 Structure of a spiking neural network for rotational motion estimation.

## 4. 実験と結果

イベントビジョンシステムにより、回転する視対象の角速度  $\omega$ [rad/s]を異なる照明環境下で推定できるか検証した。

### 4.1 データ取得実験

本実験の環境を図 3 左側に、撮影対象を図 3 右上に示す。実験機材として、イベントベースカメラ (Prophesee 社、PEK3MVCD[3])、角速度の推定を行う対象として自然な視覚刺激として岩場の写真が印刷された正方形の画用紙を用いた。また、撮影対象を回転させる電動式のターンテーブル (アズワン、2-5058-01) を用いた。撮影対象が 128\*128[pixels]に収まるように画角を調整しながら、イベントベースカメラの光軸中心と撮影対象の回転中心が同じになるように設置した。撮影は照明環境として明るい場合(532.9[lx])と暗い場合(2.3[lx])で行われた。本実験では、ターンテーブルの回転速度調整ダイヤルを用いて、その回転速度を 11 段階に変更した。図 4 に撮影したイベントベースカメラの出力の例をフレーム表現として示す。カメラ出力を 10[ms]積分し、フレーム表現に変換した。

上記のように取得したデータセットの情報として入力時間は 300[ms]、学習データは 880 個、テストデータは 220 個、ラベルは各角速度を最大値であるデータ番号 11 の角速度で 0 から 1 に正規化した。ここで表 1 に明るいときと暗いときのデータ番号と角速度の対応をそれぞれ示す。

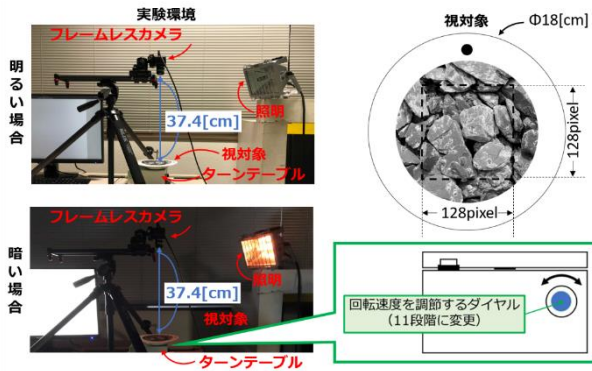


Fig. 3 Experimental environment.

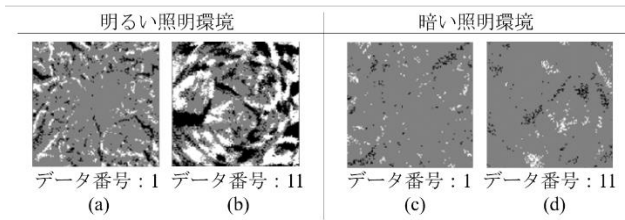


Fig. 4 Examples of data sets (frame representation). Slow (a) and fast (b) rotation speeds under bright lighting conditions. Slow (c) and fast (d) rotation speeds in dark lighting conditions.

### 4.2 運動推定結果

取得したデータセットを用いて、回転体の角速度  $\omega$  [rad/s]を学習・推定した。学習時のエポック数は 30 とした。図 5 に明るい照明環境(図 5(a))及び暗い照明環境(図 5(b))の場合の学習曲線をそれぞれ示す。横軸がエポック数で縦軸が損失である。この結果から、両方の場合でエポック数が増えていくにつれて損失が減少していることが確認できた。学習後の損失の値、絶対平均誤差、平均精度を表 2、表 3 に示す。角度推定の精度は学習データとテストデータで 8 割程度であった。

Table 1 Data sets acquired by event-based vision.

明るい照明条件		暗い照明条件	
データ番号	角速度 $\omega$ [rad/s]	データ番号	角速度 $\omega$ [rad/s]
1	2.07	1	2.04
2	3.45	2	3.39
3	4.96	3	4.61
4	6.39	4	5.95
5	7.83	5	7.41
6	9.44	6	9.15
7	10.93	7	10.59
8	12.49	8	12.35
9	14.05	9	13.61
10	16.25	10	15.66
11	17.22	11	16.84

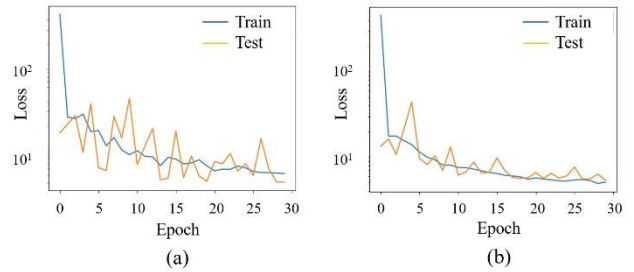


Fig. 5 Learning curve. (a) bright lighting conditions. (b) dark lighting condition

Table 2 Experimental results in a bright lighting environment.

	損失	絶対平均誤差	平均精度
学習データ	5.218	0.097	74.348
テストデータ	4.091	0.059	85.398

Table 3 Experimental results in a dark lighting environment.

	損失	絶対平均誤差	平均精度
学習データ	5.092	0.083	78.339
テストデータ	5.401	0.065	82.576

## 5. 結言

イベントベースカメラと SNN を組み合わせたイベントビジョンシステムにより、学習データとテストデータで 8 割程度の精度で回転速度を推論できた。この精度は、ネットワーク構造やスパイクニューロンのパラメータ調整の工夫を更に検討することにより向上すると考えられる。

## 参考文献

- [1] Gerstner, W., "Time structure of the activity in neural network models", Physical review E, Vol. 51, No.1, p. 738, 1995.
- [2] Shrestha, S. B. and Orchard, G., "SLAYER: spike layer error reassignment in time", In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 1419-1428, 2018.
- [3] Posch, C., Matolin, D., and Wohlgenannt, R., "A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS", IEEE Journal of Solid-State Circuits, Vol. 46, No.1, pp. 259-275, 2010.