

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 01/30/2024 (MM/DD/YYYY)

学科名 Department	電子物理システム	氏名 Name	小泉翔	指導員 Advisors	柳澤 政生 渡辺 裕 印
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	1W202123-0		
研究題目 Title	イベントカメラの時系列情報を用いた三次元姿勢推定 3D Pose Estimation Using Time-Series Information from Event Cameras				

1. まえがき

イベントカメラはピクセル単位で輝度変化を検出し、連続的かつ非同期的にイベントを提供するビジョンセンサーである。RGBカメラよりダイナミックレンジが広く、夜間でも被写体の動きを正確に捉えることができ、3D姿勢推定に適している。

従来のイベントベースの単眼3D姿勢推定では、フレームベースの手法が採用されている。具体的には、イベント信号ストリームを一定数のイベントパケットのシーケンスに分割し、各イベントパケットを画像に変換し、1フレームから3次元姿勢推定を行う。しかしこの方法では、ほぼすべての動き情報を保持するイベントカメラの特性を活かすことができない。例えば、腕だけを動かすような不均一な動きの場合、わずかな動きの関節を表すイベントの収集が不十分となり、全身推定の精度が不安定となる。

そこで本論文では、1フレームでは取得できない動きの少ない関節を推定するために、時系列情報を導入した姿勢推定モデルを提案する。具体的には、現在のフレームの特徴量に加えて、過去と未来のフレームの特徴量を利用することで3次元姿勢推定精度の向上を図る。

2. 従来手法とその問題点

イベントベースの姿勢推定に関する論文は数少なく、特にイベントベースの単眼3D姿勢推定モデルは、筆者の知る限り、Scarpelliniら[1]により提案されたMargi Poseモデルのみである。

具体的なモデル構造を図1の青枠に示す。まず、1フレームについてResNet34を用いて特徴量を抽出する。次に、その特徴量から限界ヒートマップ[3]と呼ばれるxy, yz, zxの各ヒートマップを作成する。このヒートマップを作成する操作はステージと名付けられており、その構造を図1の赤枠に示す。ステージを一度だけ用いるモデルと、三度繰り返すモデルが用意されており、それぞれステージ1モデル、ステージ3モデルと定義する。

この従来手法の問題点は、単一フレームから3D姿勢推定を行うことにある。各フレームは一定数のイベントから構成されており、動きの偏ったシーンのフレームでは全身の姿勢を正しく推定できないため推定精度が大幅に低下する。

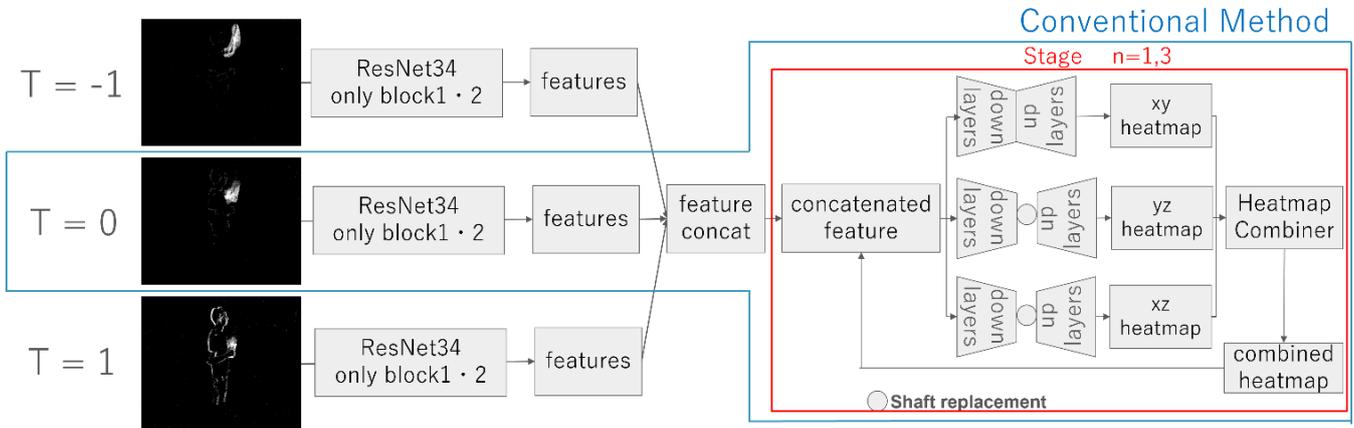


図1 提案手法のモデル構造 (IEVC 2024 で発表予定)[4]

3. 提案手法

複数フレームを用いたイベントベースの単眼姿勢推定精度を向上させる手法を提案する。イベントカメラは輝度の変化をイベントとして出力するため、単一フレームのみを用いて、動きの少ない関節を正確に検出することは困難である。

そこで、現在のフレームに加えて、直前と直後のフレームも用いて姿勢推定を行うモデルを提案する。図1に示すように、ResNet34を用いて3フレームから特徴量を抽出し、連結して時系列情報を統合する。具体的には、 $X \times Y \times C$ の大きさの特徴量を $1 \times Y \times C$ に分割し、時系列順に交互に結合する。結合された特徴量のサイズは $3X \times Y \times C$ であり、それをサイズ 3×3 のカーネルで畳み込み、限界ヒートマップを作成する。

また、5フレームを推定に使用するモデルも提案する。具体的なネットワークは3フレームのモデルと同様であるが、特徴量を結合した後はサイズ 5×5 のカーネルの畳み込み層を用いる。

4. 学習条件

提案する姿勢推定モデルの学習には従来手法と同様、DHP19データセット[2]を用いる。DHP19はイベントカメラで取得したイベントデータと、それに対応する3D姿勢推定用のアノテーションから構成される。学習用の損失には各ヒートマップのJensen-Shannon発散と、正解データの3次元座標と予測データの3次元座標の二乗誤差の合計を用いる。

5. 実験結果

従来手法と提案手法による3D姿勢推定結果を図2に、MPJPEの値を指標とした定量的評価を表1に示す。表1では、ステージ1, 3それぞれのモデルで精度を比較した。提案手法においては入力フレーム数を3, 5とした場合でそれぞれ精度を算出した。図2と表1より、時系列情報の導入による推定精度の向上が確認できた。

また、提案手法のステージ1モデルは、従来手法のステージ3モデルよりもパラメタ数が少なく、精度が良いことを確認できた。

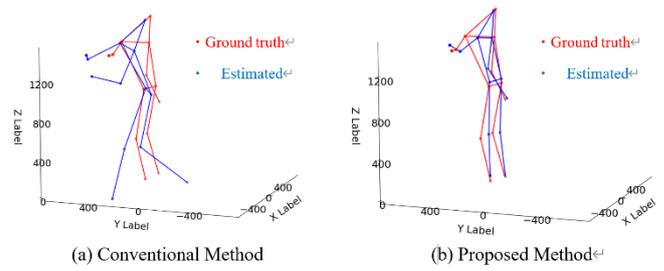


図2 3D姿勢推定結果の比較

(IEVC 2024 で発表予定)[4]

表1 各モデルにおけるMPJPE値

Method	ステージ	MPJPE [mm]
従来手法 [1]	1	84.84
提案手法(フレーム3枚)	1	68.38
提案手法(フレーム5枚)	1	67.36
従来手法 [1]	3	83.06
提案手法(フレーム3枚)	3	68.66
提案手法(フレーム5枚)	3	66.96

6. むすび

本研究では、従来のイベントベースの単眼3D姿勢推定手法であるMargi Poseモデルに対し、時系列情報を導入した手法を提案した。実験により、3D姿勢推定の精度とモデルのパラメタ数の評価により、提案手法の有効性を確かめた。今後は、より複雑な動きをするデータについて、提案手法の有効性を確かめる必要がある。

参考文献

- [1] G. Scarpellini *et al.*, "Lifting monocular events to 3d human poses," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.1358-1368, Jun. 2021.
- [2] E. Calabrese *et al.*, "Dhp19: Dynamic vision sensor 3d human pose dataset," IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1695-1704, Jun. 2019.
- [3] A. Nibali *et al.*, "3D Human Pose Estimation with 2D Marginal Heatmaps," IEEE Winter Conference on Applications of Computer Vision, pp.1477-1485, Jan. 2019.
- [4] K. Koizumi *et al.*, "3D Pose Estimation Using Time Series Data in Event-based Video," The 8th IEEE International Conference on Image Electronics and Visual Computing (IEVC 2024), Mar. 2024.

2023 年度 卒業論文

イベントカメラの時系列情報を用いた三次元姿勢推定

3D Pose Estimation Using Time-Series Information

from Event Cameras

指導教員 柳澤 政生 教授

副指導教員 渡辺 裕 教授

提出日：2024 年 1 月 30 日

早稲田大学 基幹理工学部 電子物理システム学科

1W202123-0

小泉 翔

目次

第1章	序論.....	3
1.1	研究背景.....	3
1.2	関連研究と問題点, 及び研究目的.....	3
1.3	本論文の構成.....	3
第2章	関連技術.....	5
2.1	まえがき.....	5
2.2	Margi Pose モデル.....	5
2.2.1	軸変換.....	6
2.2.2	モデルの全体構造.....	6
2.3	DHP19.....	7
2.4	むすび.....	8
第3章	提案手法.....	9
3.1	まえがき.....	9
3.2	時系列データの導入と特徴量結合手法.....	9
3.3	むすび.....	10
第4章	実験・実験結果.....	11
4.1	まえがき.....	11
4.2	実験条件.....	11
4.2.1	データセット.....	11
4.2.2	損失関数及び学習パラメタ.....	11
4.3	定量的評価.....	12
4.3.1	MPJPE による評価.....	12
4.3.2	パラメタ数による評価.....	12
4.4	定性的評価.....	13
4.5	考察.....	14
4.6	むすび.....	14
第5章	結論と今後の課題.....	15
5.1	結論.....	15
5.2	今後の課題.....	15

謝辭.....	16
參考文獻.....	17
圖一覽.....	18
表一覽.....	19
研究業績.....	20

第1章 序論

1.1 研究背景

3D姿勢推定は，ロボットの動作への活用[1]やスポーツ動作解析[2]など様々な分野で応用されている技術である．姿勢推定には一般的には RGB カメラが用いられているが，カメラが観測できる明暗の幅には制限があり，暗い場所での撮影は困難である．

イベントカメラはピクセル毎に輝度変化を検知して，イベントを連続的・非同期的に提供するビジョンセンサーである[3]．イベントカメラは時間分解能が非常に高く， μs オーダーで被写体を撮影することができる．また，ダイナミックレンジは RGB カメラよりも大きく，140[dB]に達する機種もある．これによって RGB カメラの欠点を補い，被写体のほぼ全ての動きの情報を夜間でも正確に検出することができる．さらに，輝度変化のみを検知するため，カメラを固定すれば動きのない背景情報を取得せず，画像の情報量を削減することができる．

イベントベースでの 3D 姿勢推定では，複数のカメラを用いる手法が主に採用されている．しかし，カメラ同士を精密に同期する必要があり，実際に導入するのは難しい．また，イベントカメラはコストが高く，複数カメラでの姿勢推定は実用性の障害となっている．

以上の理由から単眼イベントベースの 3D 姿勢推定は重要な技術である．

1.2 関連研究と問題点，及び研究目的

イベントベースでの単眼 3D 姿勢推定における主なアプローチには，フレームベースの推定手法がある[4]．この手法はイベント信号のストリームを定数個のイベントパケットのシーケンスに分割し，画像に変換したうえで，単一フレームから 3D 姿勢を推定する方法である．しかし，この手法では時間分解能が大きく，動きに関するほぼ全ての情報を保持できるイベントカメラの利点を活かさない．例えば，手を振るなどの局所的な動きをした場合，動きの少ない関節を表すイベントが十分に蓄積されないフレームが発生し，全身の姿勢推定精度が不安定になる．

そこで本研究では，動きが少なく単一のフレームでは取得できなかった関節の推定に，時系列情報を導入した新たな手法を提案する．実験により，イベントが偏ったフレームへの 3D 姿勢推定に対して，提案手法の有効性を示す．

1.3 本論文の構成

本論文の構成を以下に示す．

第1章 本章であり，本研究の背景，関連研究と問題点及び研究目的について述べる．

第2章 従来のイベントベースの 3D 姿勢推定及び関連技術について述べる．

第3章 本研究の提案手法について述べる.

第4章 本研究における実験の方法, 結果及び考察について述べる.

第5章 結論と今後の課題について述べる.

第 2 章 関連技術

2.1 まえがき

本章では、イベントベースの 3D 姿勢推定モデルである Margi Pose モデル[5]について述べる。また、関連技術として、イベントベースの姿勢推定に用いられるデータセット DHP19[6]について述べる。

2.2 Margi Pose モデル

Margi Pose モデルはイベントベースの単眼 3D 姿勢推定モデルである。このモデルでは、イベントデータを収集した画像 1 枚を入力として用いる。図 2.1 に示すように、まず ResNet34[7]の 1, 2 番目のブロックに画像を入力し、特徴量を抽出する。次に、特徴量から Marginal Heat maps[8,9]を作成する。これは三つのヒートマップ xy , yz , zx ヒートマップから作成される。このヒートマップを作成する操作はステージと名付けられており、その構造を図 2.1 の赤枠に示す。ステージを一度だけ用いるモデルと、三度繰り返すモデルが用意されており、それぞれステージ 1 モデル、ステージ 3 モデルと定義される。

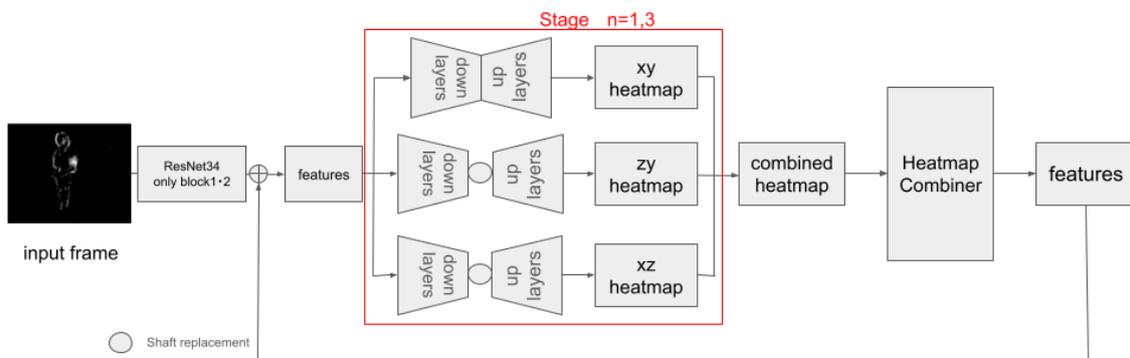


図 2.1 Margi Pose モデルのネットワーク構造

Marginal Heat Maps は単一フレームから 3D 姿勢を予測するために仮想的に z 軸を推測することによって生成された $xy \cdot yz \cdot zx$ ヒートマップのことである。以下で、Marginal Heat Maps を作成するネットワークについて、軸変換と全体のニューラルネットワーク構造に分けて説明する。

2.2.1 軸変換

xy 平面の入力特徴量から yz 平面のヒートマップの予測を考えると、一般に、z 方向の位置は x 方向に対応しない。しかし、出力である xy, yz, zx ヒートマップは入力特徴量の xy 空間に対して一つ以上空間的に一致している軸が存在する。そこで、図 2.2 のように、yz ヒートマップを作成するには xy 空間の特徴量を立方体に分割し、y 軸を中心に回転することで仮想的な yz 空間を創出する。

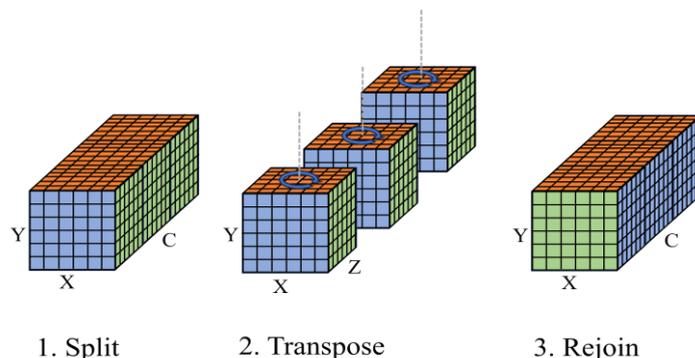


図 2.2 xy 空間から yz 空間への軸置換演算

2.2.2 モデルの全体構造

Margi Pose モデルの 一つのステージにおけるネットワークの構造を図 2.3 に示す。

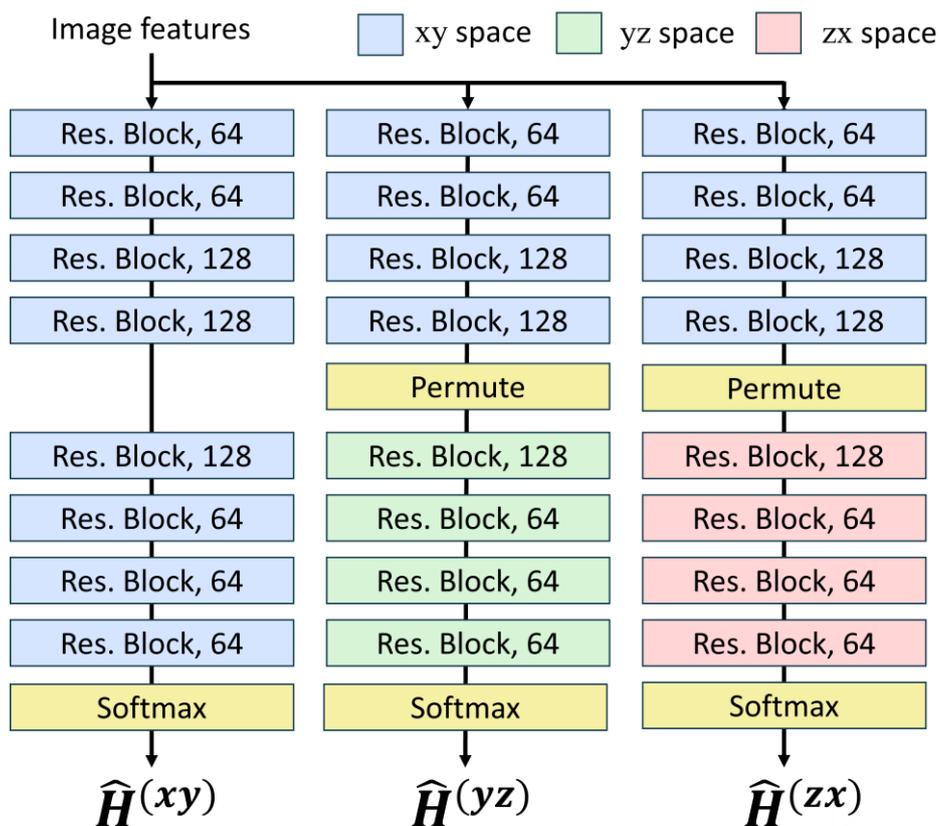


図 2.3 Margi Pose モデルにおける 1 ステージのモデル構造

上図のように、xy ヒートマップ、 $\hat{H}^{(xy)}$ は軸の置換なく 8 個の残差ブロックを通してヒートマップを作成するのに対し、yz, zx ヒートマップ、 $\hat{H}^{(yz)}$, $\hat{H}^{(zx)}$ は残差ネットワークの途中で軸変換処理を施すことで空間を移行し、作成している。そして、作成したヒートマップに soft-argmax 演算子[10]を適応することで関節の座標を導き出している。

次に、Margi Pose モデルの学習に用いる損失関数を式 (2.1) – 式(2.3) に示す。

$$L = \sum_i L_{\text{geometrical}}(\hat{p}_{xyz}^i, p_{xyz}) + JSD(H_{xy}, \hat{H}_{xy}^i) + JSD(H_{yz}, \hat{H}_{yz}^i) + JSD(H_{zx}, \hat{H}_{zx}^i) \quad (2.1)$$

$$L_{\text{geometrical}}(\hat{p}_{xyz}^i, p_{xyz}) = \|\hat{p}_{xyz}^i - p_{xyz}\|^2 \quad (2.2)$$

$$JSD(H, \hat{H}) = \frac{1}{2}KL(H||\hat{H}) + \frac{1}{2}KL(\hat{H}||H) \quad (2.3)$$

損失関数には、正解データの 3 次元座標 p_{xyz} と予測データの 3 次元座標 \hat{p}_{xyz}^i の二乗誤差を合計したものと、各ヒートマップの Jensen-Shannon 発散、 $JSD(H, \hat{H})$ を使用する。ここで、Jensen-Shannon 発散とはエントロピーの損失量を表す KL ダイバージェンス、 $KL(\hat{H}||H)$ を対称的に用いたものである。また、 H は正解データのヒートマップで、 \hat{H} は予測したヒートマップである。

2.3 DHP19

DHP19はイベントデータのみで構成された初めての3D姿勢推定用のデータセットである。これには、17人の被験者と33の動作が記録されており、被験者にマーカーを付け、同期した四つのイベントカメラを使用して作成されている。被験者は、殴る・蹴るなどの上半身・下半身のみの運動や、跳ねるなどの全身運動を10回連続で繰り返す。それらの運動をイベントカメラにより撮影し、図2.4のようなフレームに変換する。これらのフレームは、1フレームあたり7500個のイベントを含むように作成される。データセットのホームページは <https://sites.google.com/view/dhp19/home?authuser=0> であり、実験での使用許可および論文での掲載は、Creative Commons Attribution-ShareAlike 4.0のライセンスによる。

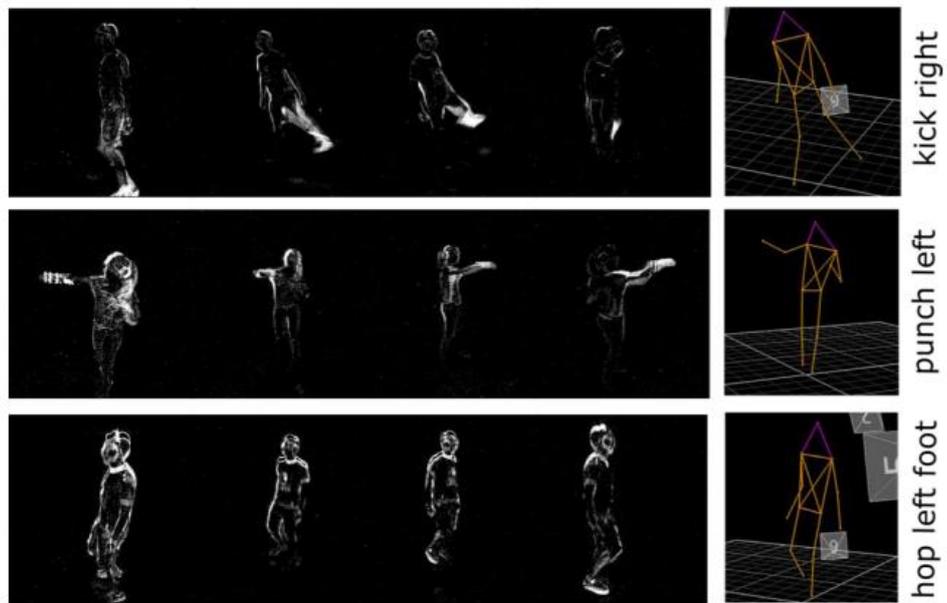


図 2.4 DHP19 データセットの一例 使用条件は CC4.0 による[6]

2.4 むすび

本章では，従来のイベントベースの 3D 姿勢推定モデルである Margi Pose モデルについて述べた．また，関連技術として，イベントベースのデータセット DHP19 について述べた．

第 3 章 提案手法

3.1 まえがき

本章では, Margi Pose モデルに, 時系列情報を導入したイベントベースの単眼 3D 姿勢推定手法を提案する. 提案手法では, 現在のフレームに加えて, 直前・直後のフレームを加えて特徴量を結合することで, より安定した 3D 姿勢推定が可能となる.

3.2 時系列データの導入と特徴量結合手法

従来のイベントベースの単眼 3D 姿勢推定では, 単一フレームのみを入力としている. この特性により, 偏った動きのイベントデータを蓄積したフレームの姿勢推定に対して精度が悪い. そこで該当フレームの直前・直後のフレームを加えた新たな姿勢推定手法を提案する.

提案手法のモデル構造を図 3.1 に示す. このモデルは, 複数フレームから特徴量を抽出する段階と, 従来手法と同様に, Marginal Heat Maps を作成する段階で構成される. 特徴量の抽出は先行研究と同様に ResNet34 の 3 番目以降の残差ブロックをカットしたものを使用し, 各フレームの特徴量を結合する方法を図 3.2 に示す.

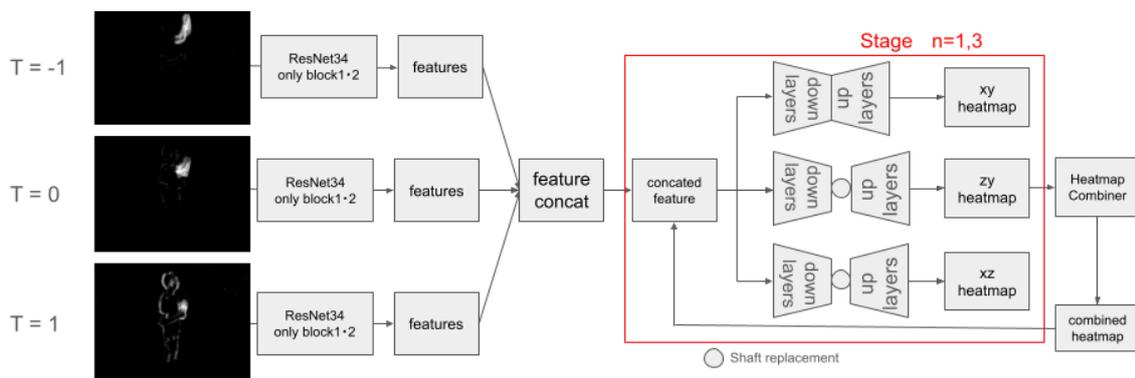


図 3.1 提案手法のモデル構造 (IEVC 2024 で発表予定)[11]

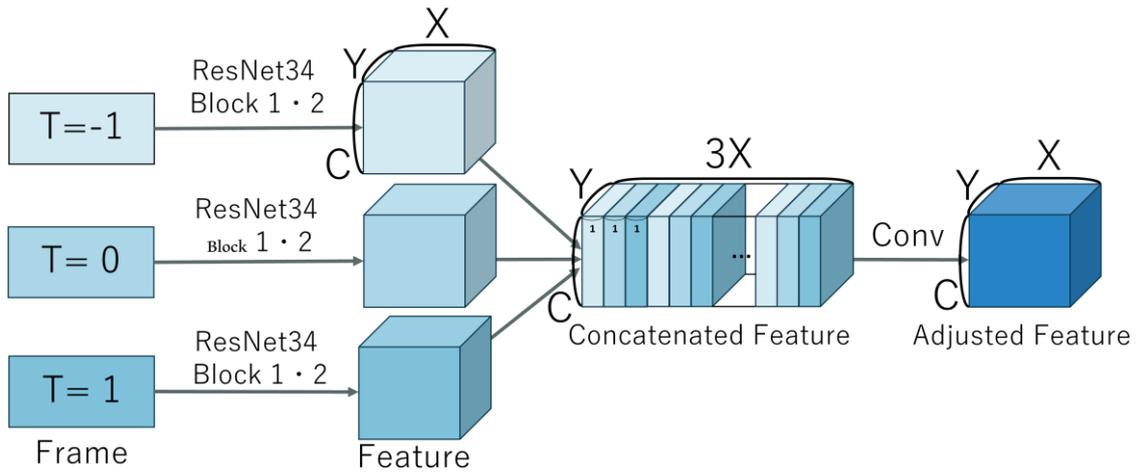


図 3.2 特徴量の結合手法 (IEVC 2024 で発表予定)[11]

まず ResNet34 を用いて、各フレームよりサイズ $X \times Y \times C$ の特徴量抽出する。それらの特徴量を $1 \times Y \times C$ に分割し、それを時系列順に交互に結合する。つまり、 X 軸の座標を x ($0 \leq x \leq X$)、フレームを過去・現在・未来の順に $T = -1, 0, 1$ と定義すると、特徴量を $[x_0^{T=-1}, x_0^{T=0}, x_0^{T=1}, x_1^{T=-1}, x_1^{T=0}, x_1^{T=1} \dots x_X^{T=-1}, x_X^{T=0}, x_X^{T=1}]$ の順に結合することである。結合された特徴量のサイズは $3X \times Y \times C$ となるが、サイズを $X \times Y \times C$ に戻すため、カーネルサイズが 3×3 、ストライドが x 方向 3, y 方向 1 の畳み込み層を用いる。

また、3枚のフレームを使用したモデルに加えて、5枚のフレームを用いて推定するモデルも提案する。具体的なネットワークや特徴量の結合方法は先に示した手法と同様である。ただし、結合された特徴量のサイズは $5X \times Y \times C$ となるため、カーネルサイズが 5×5 、ストライドが x 方向 5, y 方向 1 の畳み込み層を用いて、サイズを $X \times Y \times C$ に戻す。

3.3 むすび

本章では、本研究の提案手法である時系列情報を導入したイベントベースの3D姿勢推定法について述べた。また、提案する姿勢推定モデルの詳細な構造について述べた。

第4章 実験・実験結果

4.1 まえがき

本章では，実験で使用したデータセット，実験方法，実験結果，考察について述べる．従来手法と提案手法の性能を比較し，本手法の有効性を確認する．

4.2 実験条件

4.2.1 データセット

データセットには，第2章でも取り上げた，DHP19 データセットを使用する．このデータセットは，女性9名・男性3名の12人の被験者による33の動作を四つのカメラで撮影したデータから構成される．validation用のデータはそのデータセットの中から，5%のデータをランダムに選んで使用する．テスト用データセットは，女性3名・男性2名の33の動作からなるデータで構成されている．

ここで，図4.1のように，一つのカメラの正解ラベルがフレームとずれており，学習に影響が生じていたので，アノテーションが正しくできている残りの三つのカメラのみを用いて学習・評価を行った．ラベルがずれていた原因としては，被験者に取り付けたマーカの実座標系での値をカメラ座標系に変換するための投影行列が誤っていることが考えられる．

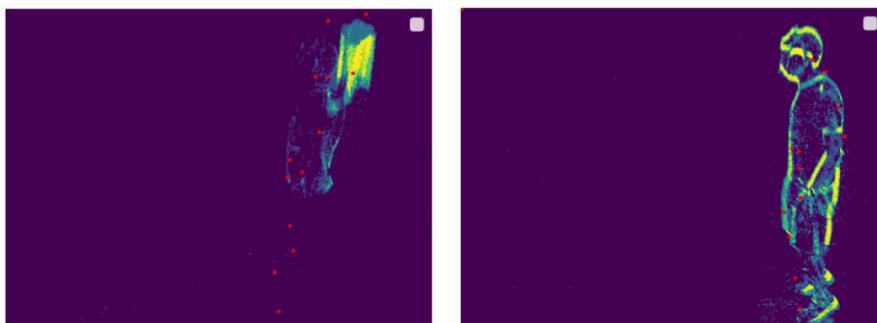


図 4.1 DHP19 データセットの正解ラベルがずれているフレームの一例

4.2.2 損失関数及び学習パラメタ

3D 姿勢推定モデルの学習には，式(2.1)の損失関数を用いた．学習には NVIDIA® GeForce RTX™ 4070 Ti を1台用い，バッチサイズを32とする．学習のエポック数は20，学習率は前半の10エポックは 3×10^{-4} に設定し，後半の10エポックは 3×10^{-5} に設定する．また，最適化関数には Adam を用いる．

4.3 定量的評価

4.3.1 MPJPE による評価

従来手法と提案手法を用いて3D姿勢推定精度の比較を行った結果を表4.1に示す。ここでは、従来手法と提案手法において、ステージ1, 3それぞれのモデルで精度を比較した。また、提案手法においては、入力に用いたフレームが3枚, 5枚の場合でそれぞれ精度を算出した。

評価指標には、すべての関節点の正解座標と推定座標の平均距離により導出されるMPJPEを用いた。これは3D姿勢推定で主に用いられており、値が小さいほど推定精度が良いことを表す。単位は[mm]で統一されている。

表4.1より、提案手法では、入力フレームを3枚用いたモデルでは平均で15.43[mm]のMPJPEの向上が確認できる。また、同様に入力フレームを5枚用いたモデルでは16.79[mm]精度が向上した。特に、従来手法で最高精度を出したステージ3のモデルに対して、5枚のフレームをステージ3のモデルに入力した手法で16.1[mm]の精度の向上を達成した。

表 4.1 各モデルに対する MPJPE 値比較

Method	Stage	MPJPE(mm) ↓
Conventional method [5]	1	84.84
Proposed method (3 frames)	1	<u>68.38</u>
Proposed method (5 frames)	1	67.36
Conventional method [5]	3	83.06
Proposed method (3 frames)	3	<u>68.66</u>
Proposed method (5 frames)	3	66.96

4.3.2 パラメタ数による評価

パラメタ数とはモデルに含まれる、最適化する必要がある変数の数を表す。パラメタ数が多いほど複雑な処理が可能となるが、テストに要する時間が増加し、必要とするデバイスのメモリ量も増大する。ここでは、現実世界での実用化の観点からモデルの軽量化を目標として、パラメタ数の比較を先と同様のモデルに対して行った。表4.2より、提案手法のステージ1モデルは、従来手法のステージ3モデルよりも、パラメタ数が少ないことを確認した。

表 4.2 各モデルに対するパラメタ数比較

Method	Stage	Parameters[M]
Conventional method [5]	1	6.99
Proposed method (3 frames)	1	7.58
Proposed method (5 frames)	1	8.10
Conventional method [5]	3	18.68
Proposed method (3 frames)	3	19.27
Proposed method (5 frames)	3	19.80

4.4 定性的評価

定性的な評価として、従来手法と提案手法の視覚的な姿勢推定結果を比較する。動きが少なく、一部の姿勢のイベントデータが取れていないフレームにおける従来手法と提案手法の結果を示す。該当するフレームと2D姿勢推定の結果を図4.2に、3D姿勢推定の結果を図4.3に示す。

従来手法では動きに偏りがあり、姿勢の一部のイベントしか蓄積されていないフレームに対しては、図4.2(a), 4.3(a)のように、正解データから大きく乖離した結果を得ていた。しかし、直前・直後のフレームも用いて推定した提案手法では図4.2(b), 4.3(b)のように視覚的に正解データに大きく近づいた推定結果を得ることができた。

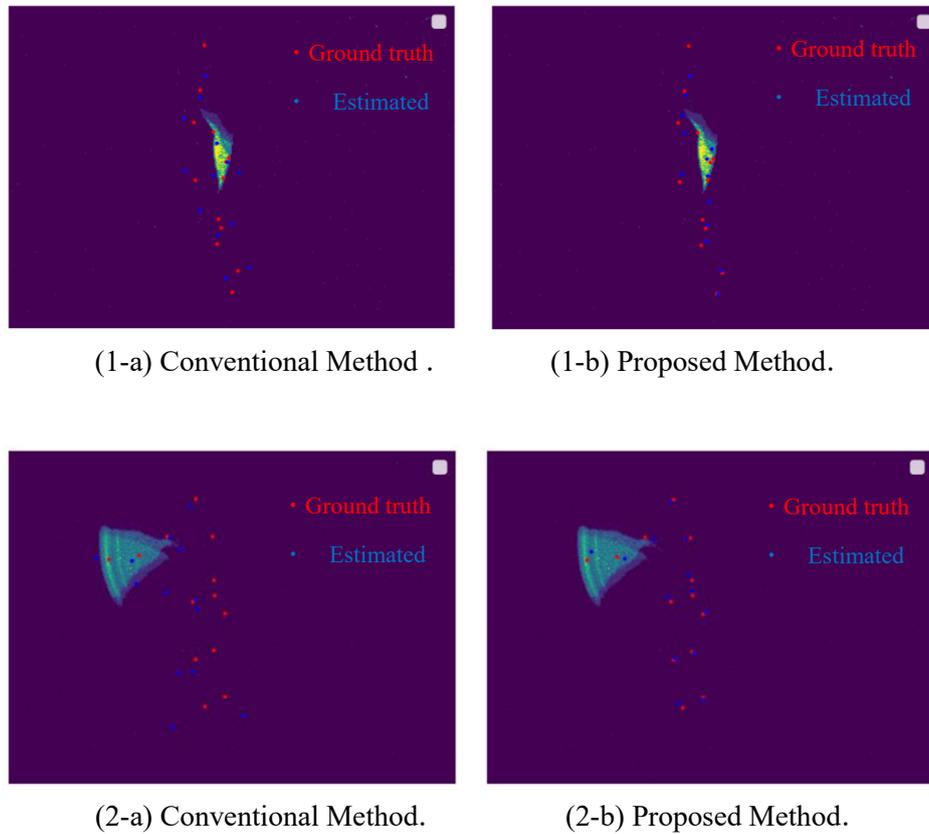


図 4.2 2D 姿勢推定結果の比較

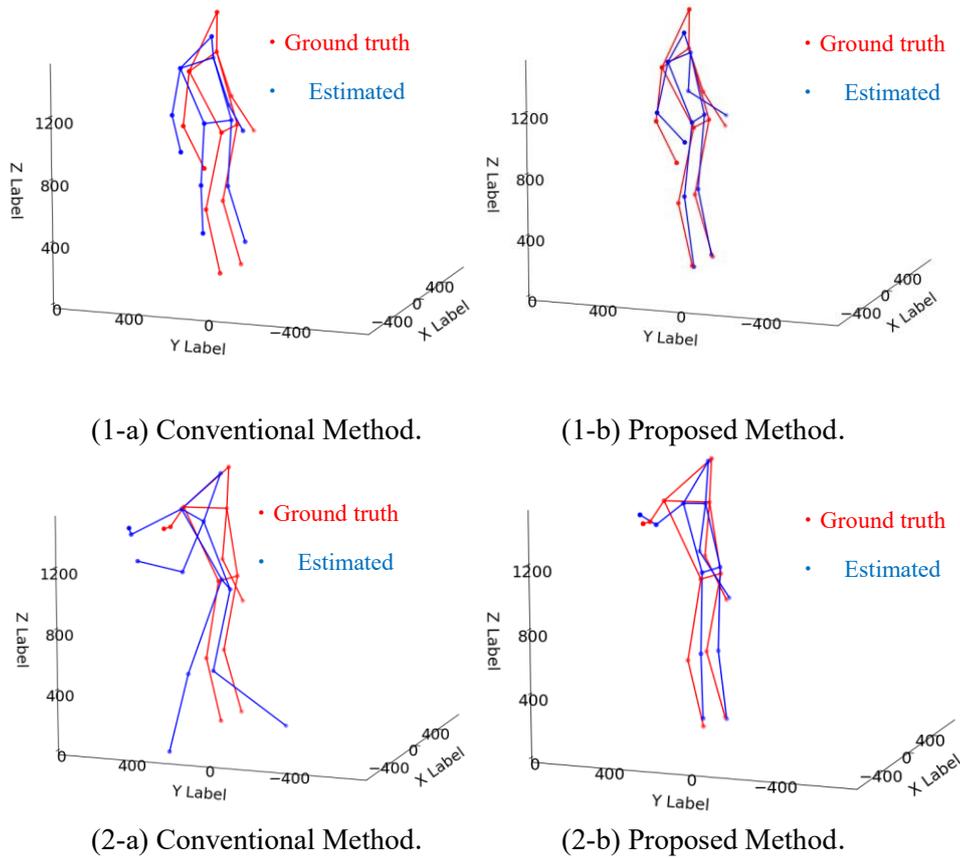


図 4.3 3D 姿勢推定結果の比較 (IEVC 2024 で発表予定)[11]

4.5 考察

4章3節より，入力フレームを1枚から3枚にした際，精度は10[mm]以上向上したのに対し，3枚から5枚に変更しても精度は1[mm]程度の向上しか見られなかった．このことから DHP19 データセットを7500点毎にフレーム化した環境において，3枚の連続したフレームを用いることで十分効果が得られるということが分かった．

また，表4.1，4.2より，3枚・5枚のフレームを入力とし，ステージを1にした提案手法では，最高精度を出す従来手法と比べて，モデルのパラメタ数を大きく抑えたまま，大幅な精度の向上を実現することが確認できた．

4.6 むすび

本章では，本研究のイベントベースの3D姿勢推定モデルにおける，学習に用いたデータセット，実験方法について述べた．また，評価実験として，同様のデータセットで従来手法と提案手法のモデルによる3D姿勢推定結果，モデルのパラメタ数などの比較を行った．提案手法では，従来手法よりも少ないパラメタ数で高い精度を実現することができた．

第5章 結論と今後の課題

5.1 結論

本研究では、従来のイベントベースの単眼3D姿勢推定手法である Margi Pose モデルに対し新たな手法を提案した。提案手法では、現在のフレームの直前・直後のフレームを用いて、それぞれの特徴量を結合することで、従来手法では正しく推定することができなかった姿勢の推定を可能にした。実験では、3D 姿勢推定の精度とモデルのパラメタ数の評価を行うことにより、提案手法の有効性を確かめた。結果として、関節の推定精度を示す MPJPE を向上させ、モデルの複雑さを示すパラメタ数を抑えることができた。また、視覚的な定性的評価からも時系列情報の導入による提案手法の有効性が示せた。したがって、提案手法により時系列情報を導入することは、画像の輝度変化を示すイベントカメラの3D姿勢推定において有効であることが確認できた。

5.2 今後の課題

提案手法では、姿勢推定精度を向上させ、計算量を抑えることができたが、実際に現実で3D姿勢推定を行うことを想定した際、より複雑な動きに対応する必要がある。また、今回は扱うことのできなかった、ラベルがずれていたデータにおいても改善を加えて学習・評価をしていきたい。

謝辞

本論文の執筆にあたり，まず電子物理システム学科からの転科を快く承諾して下さった柳澤政生教授に感謝いたします。

また，転科を承諾して下さったことに加え，大学院の研究に先立って，本研究からご指導くださり，快適な研究環境を与えてくださった渡辺裕教授に感謝申し上げます。

次に，ゼミ活動など日ごろから，研究の進捗や実験方法に関して様々なアドバイスをくださった早稲田大学情報通信研究センター招聘研究員の原潤一博士に感謝いたします。

また，日頃から素晴らしい研究内容を共有していただき，ときには研究の相談に乗ってくださり，優しく接して下さった渡辺研究室の皆様に感謝いたします。

最後に，私をここまで育ててくださり，心や生活を支えてくださっている家族に感謝いたします。

参考文献

- [1] C. Zimmermann, T. Welschehold, C. Dornhege, W. Burgard, and T. Brox, “3D Human Pose Estimation in RGBD Images for Robotic Task Learning,” 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 1986-1992, May. 2018.
- [2] J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, “Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance,” In Proceedings of the 27th ACM International Conference on Multimedia, pp. 374-382, Oct. 2019.
- [3] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, “Event-based vision: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 154-180, Jan. 2022.
- [4] S. Zou, C. Guo, X. Zuo, S. Wang, P. Wang, X. Hu, S. Chen, M. Gong, and L. Cheng, “EventHPE: Event-based 3D Human Pose and Shape Estimation,” 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10976-10985, Oct. 2021.
- [5] G. Scarpellini, P. Morerio, and A. D. Bue, “Lifting monocular events to 3d human poses,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1358-1368, Jun. 2021.
- [6] E. Calabrese, G. Taverni, C. A. Easthope, S. Skriabine, F. Corradi, L. Longinotti, K. Eng, and T. Delbruck, “Dhp19: Dynamic vision sensor 3d human pose dataset,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1695-1704, Jun. 2019.

URL: “<https://sites.google.com/view/dhp19/home?authuser=0>”
- [7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, Jun. 2016.
- [8] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, “Monocular 3d human pose estimation in the wild using improved cnn supervision,” 2017 International Conference on 3D Vision (3DV), pp 506-516, Oct. 2017.
- [9] A. Nibali, Z. He, S. Morgan, and L. Prendergast, “3D Human Pose Estimation with 2D Marginal Heatmaps,” 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1477-1485, Jan. 2019.
- [10] D. Neil, M. Pfeiffer, and S. Liu, “Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences,” NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 3889-3897, Dec. 2016.
- [11] K. Koizumi and H. Watanabe, “3D Pose Estimation Using Time Series Data in Event-based Video,” The 8th IEEEJ International Conference on Image Electronics and Visual Computing (IEVC 2024), Mar. 2024.

図一覧

図 2.1 Margi Pose モデルのネットワーク構造	5
図 2.2 xy 空間から yz 空間への軸置換演算	6
図 2.3 Margi Pose モデルにおける 1 ステージのモデル構造	6
図 2.4 DHP19 データセットの一例 使用条件は CC4.0 による[6]	8
図 3.1 提案手法のモデル構造 (IEVC 2024 で発表予定)[11].....	9
図 3.2 特徴量の結合手法 (IEVC 2024 で発表予定)[11].....	10
図 4.1 DHP19 データセットの正解ラベルがずれているフレームの一例	11
図 4.2 2D 姿勢推定結果の比較	13
図 4.3 3D 姿勢推定結果の比較 (IEVC 2024 で発表予定)[11].....	14

表一覧

表 4.1	各モデルに対する MPJPE 値比較.....	12
表 4.2	各モデルに対するパラメタ数比較.....	12

研究業績

- [1] K. Koizumi and H. Watanabe, “3D Pose Estimation Using Time Series Data in Event-based Video”, The 8th IIEEJ International Conference on Image Electronics and Visual Computing (IEVC 2024), Mar. 2024. (to appear)