イベントカメラによる人物姿勢推定の遅延補償を用いたリアルタイムエフェクトシステムの構築

大武 一平[1,a)]　北野 和哉[1]　櫛田 貴弘[4]　藤村 友貴[1]
前島 謙宣[2]　久保 尋之[3]　舩冨 卓哉[1]　向川 康博[1]

概要

人物の姿勢を使用してエフェクト出力を行うコンテンツにおいて，従来の姿勢推定をオンライン処理で実施すると，暗い舞台上で撮影した画像からは姿勢推定が困難なほか，姿勢推定の処理時間によるリアルタイム性の低下につながる．本研究では，近赤外光を用いて撮影したフレームを入力とする姿勢推定に対して，イベントカメラが持つ低遅延性と時間分解能の高さを活かし姿勢推定器の遅延を補償する手法を用い，ステージの極端な光源環境で動作可能なリアルタイムエフェクトシステムを構築する．実際のライブシーンでイベントカメラを用いた遅延補償を行い，リアルタイムなエフェクト表示が可能であることを示す．

1. はじめに

近年，動きによってエフェクトを変化させるダンスエンタテイメントや，自身の動きに応じて動的プロジェクトマッピングを行う体感型コンテンツなど，リアルタイムなインタラクティブシステムが普及しつつある．これらのアプリケーションでは，人の速い動きに対して追従するリアルタイム性の高い姿勢推定が必要不可欠である．特に姿勢推定結果に応じて視覚効果を対象に重畳表示するような，人の動きに対するエフェクト合成システムの場合，観客は対象とエフェクト出力を同時に観察することになるため，姿勢推定におけるリアルタイム性の低下はコンテンツの体験を著しく損なう．カメラフレームから逐次的に姿勢推定を行うオンライン処理においてリアルタイム性を向上させる手法として，輝度値の変化をイベントとしてストリーミングするイベントカメラ [1] を用いた遅延補償技術を開発した [2]．本稿では，ステージの極端な光源環境といった実際のダンスエンタテイメント特有の問題に対処しつつ，イベントカメラを用いた遅延補償を導入してエフェクト表示を行うことで，従来よりもリアルタイム性の高いエフェクトと一体となったダンスエンタテイメントを実現する．

1 奈良先端科学技術大学院大学
2 株式会社オー・エル・エム・デジタル，株式会社 IMAGICA GROUP
3 千葉大学
4 立命館大学
a) otake.ippei.oj2@is.naist.jp

図 1　提案手法であるリアルタイムエフェクトシステムの実施状況．演者の腕の動きとリンクした光の軌跡．

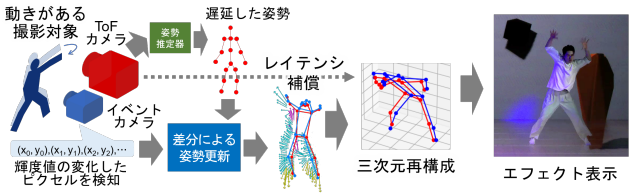

図 2　リアルタイムエフェクトシステムの概要．

2. 近赤外撮影とイベントカメラを用いた遅延補償によるリアルタイムエフェクトシステム

今回我々が構築したリアルタイムエフェクトシステムの概要を図 2 に示す．イベントカメラを用いた遅延補償技術 [2] は，イベントカメラ視点における二次元平面上での補償しか想定していない．しかし，本稿で対象とするダンスエンタテイメントにおけるエフェクト表示は，入力として三次元姿勢を必要とする．従って本稿では，三次元座標の取得が可能な ToF カメラを用いて人の動きを観測し，これに対して姿勢推定を実施する．姿勢推定処理で発生した三次元姿勢の遅延をイベントカメラ視点で補償し，再度三次元姿勢へリフトアップしエフェクト表示に用いる．また，ステージの極端な光源環境による姿勢推定への悪影響

を排除するため，ToF カメラの出力として得られる近赤外強度マップを対象として姿勢推定を行う．また，ステージの極端な光源環境と ToF が発する近赤外光の両方の影響を排除するため，イベントカメラ用に静的な近赤外照明を別途設置し，その照明の波長に適合するバンドパスフィルタをイベントカメラに装着することで，人の動きのみに関連するイベントサンプリングを行う．

2.1 近赤外光を用いて撮影したフレームを入力とする三次元姿勢推定

エフェクト表示を行うダンスエンタテイメントでは，ステージの極端な光源環境下で演者がパフォーマンスを行う．よって通常の RGB カメラで撮影したフレームでは人物を十分捉えきれず，姿勢推定が困難となる．そこで提案手法では ToF カメラの近赤外強度マップを姿勢推定器の入力として用いることで，ステージの極端な光源環境下においても安定した姿勢推定を可能とする．

本研究で構築したシステムでは，ToF カメラとして 850nm によるアクティブセンシングを行う LUCID Vision Labs 社製 Helios2+を用いた．850nm は可視光領域外であるため，センシング光はコンテンツに影響せず頑健な撮影が可能である．

深度マップに付随する近赤外強度マップ $\boldsymbol{I}_{\text{ir}} \in \mathbb{R}^{H \times W}$ を，以下の形で 256 階調で正規化し二次元の姿勢推定器に入力することで，それぞれの関節 $k = 0, \cdots, K$ に対して二次元座標 $J_k = (x_k, y_k)$ を持つ二次元姿勢 $\boldsymbol{J}$ を得る．本稿では二次元の姿勢推定器として ViTPose [3] を用いた．

$$\tilde{\boldsymbol{I}}_{\text{ir}}[x, y] = 255 \times \frac{\boldsymbol{I}_{\text{ir}}[x, y] - \min(\boldsymbol{I}_{\text{ir}}[x, y])}{\max(\boldsymbol{I}_{\text{ir}}[x, y])} \tag{1}$$

得られた二次元姿勢の各関節座標 J_k を，それぞれのピクセルに三次元座標 (X_p, Y_p, Z_p) を持つ三次元マップ $\boldsymbol{I}_{\text{depth}}$ 上で参照し，ピクセルが位置する三次元座標を得ることで，各関節座標として三次元座標 $J_k^{3D} = (X_k, Y_k, Z_k)$ を持つ三次元姿勢 $\boldsymbol{J}^{3D}$ へ再構成を行う．

$$J_k^{3D} = \boldsymbol{I}_{\text{depth}}[x_k, y_k] \tag{2}$$

2.2 イベントカメラへの視点変更

前項で得られた三次元姿勢をイベントカメラ視点に変換する．三次元姿勢 $\boldsymbol{J}^{3D}$ を，それぞれの外部パラメータ $\mathbf{R}_{\text{ToF}}, \mathbf{R}_{\text{event}}$ を用いて ToF カメラ視点 (X_k, Y_k, Z_k) からイベントカメラ視点の三次元姿勢 $J_k'^{3D} = (X_k', Y_k', Z_k')$ へ変換したのち，イベントカメラの内部パラメータを用いてイベントカメラ上のピクセル座標 $J_k' = (x_k', y_k')$ へ投影し，イベントカメラ視点の二次元姿勢 $\boldsymbol{J}'$ を得る．

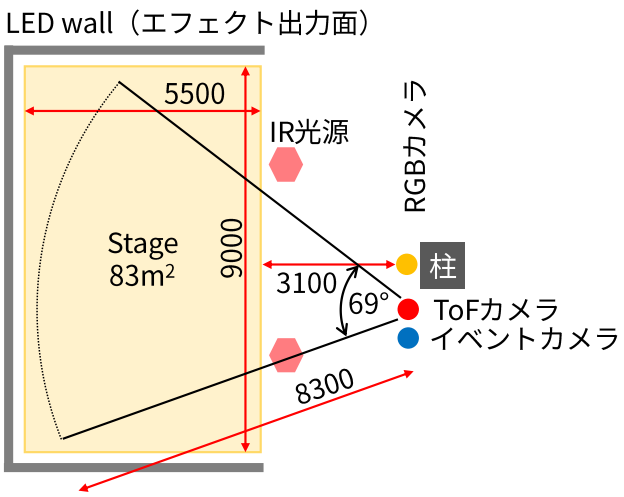

図 3 ライブステージの会場図およびカメラ配置．

$$\begin{pmatrix} X_k' \\ Y_k' \\ Z_k' \\ 1 \end{pmatrix} = \mathbf{R}_{\text{event}} \mathbf{R}_{\text{ToF}}^{-1} \begin{pmatrix} X_k \\ Y_k \\ Z_k \\ 1 \end{pmatrix} \tag{3}$$

2.3 イベントカメラを用いた遅延補償

イベントカメラを用いた多段階の姿勢更新 [2] では，姿勢推定処理時間中に一定のイベント数 N_b が得られるごとに，各ピクセルにピクセルの移動量を持つマップである Optical Flow $\boldsymbol{F}_{m-1:m}$ を導出する．処理時間中に得られた m 枚の Optical Flow 上で，遅延を含むイベントカメラ視点の関節座標 J_k' を参照し足し合わせる，以下の形の更新を m 回行うことで各関節座標は $J_{k,m}' = \left(x_{k,m}', y_{k,m}'\right)$ となり遅延補償後の二次元姿勢 $\boldsymbol{J'}_m$ を得る．

$$J_{k,m}' = J_{k,m-1}' + \boldsymbol{F}_{m-1:m}\left[x_{k,m}', y_{k,m}'\right] \tag{4}$$

イベントカメラは輝度値の変化を取得する．そのため，演者の動き以外にも，エフェクトに起因する変化を合わせて検出してしまう．エフェクトによる影響を排除するため，イベントカメラも ToF カメラと同様に近赤外光を用いた撮影とすることでこれを回避する．この際，ToF カメラと同様の波長帯，本システムにおいては 850nm 帯をイベントカメラで使用すると，ToF によるアクティブセンシング光をイベントカメラ側が取得してしまう．そのため，イベントカメラで使用する近赤外波長帯と ToF カメラで使用する近赤外波長帯は切り分ける必要がある．そこで，ToF カメラとは異なる波長帯の近赤外光源を別途設置し，その光源の波長帯のみを観測できるようバンドパスフィルタをイベントカメラへ取り付けることとした．本システムでは 800nm 帯を通過させるバンドパスフィルタ及び近赤外光源を選定し，イベントカメラへの取り付け及びステージへの光源設置を行なった．またイベントカメラには Prophesee 社製 Gen4 を使用した．

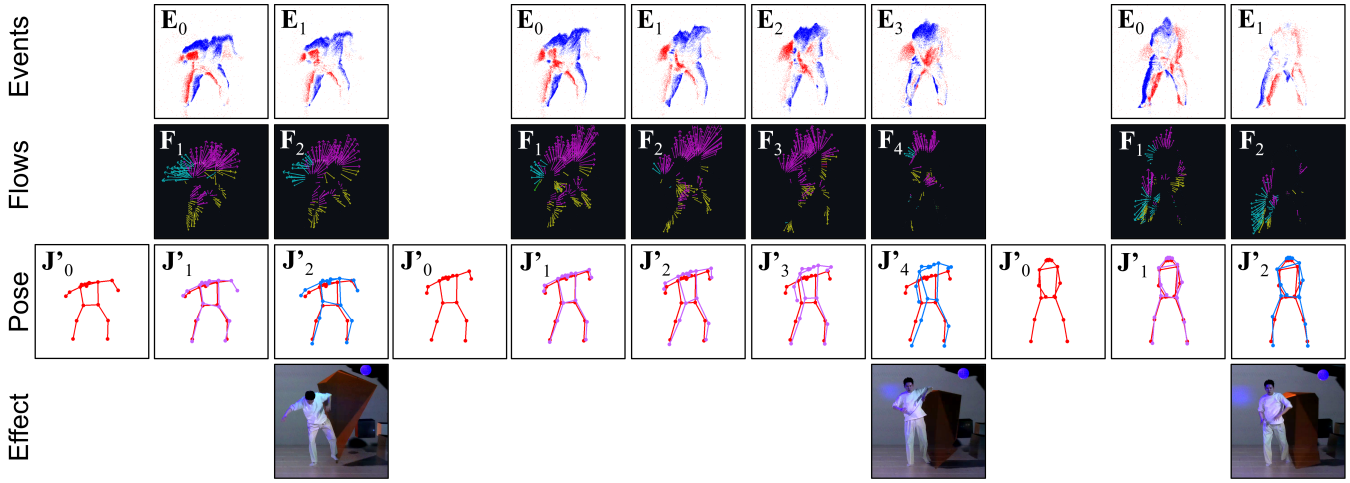

図 4 定性評価．上段から，イベント，導出された Optical Flow，入力姿勢および姿勢更新結果，最終的に遅延保証された姿勢を用いて行なったエフェクト合成結果，を示した．更新結果は，更新前の姿勢を赤色，関節位置の更新結果を青色で示した．

2.4 遅延補償した姿勢の三次元リフトアップ

最終的なエフェクト出力のため，補正後の姿勢を再度三次元へリフトアップする．ただし，姿勢推定の処理時間中に新たな ToF フレームが取得されていることがほとんどである．従って，この三次元リフトアップは新たに得られた ToF フレームを用いて行われる．これにより深度方向も最新のものに更新されることとなる．

新たに得られた ToF カメラによる三次元マップ $\boldsymbol{I}^{\text{latest}}_{\text{depth}}$ は，それぞれのピクセルに三次元成分をもつ点群である．よって 2.2 と同様に，それぞれの外部パラメータ $\mathbf{R}_{\text{ToF}}, \mathbf{R}_{\text{event}}$ を用いて ToF カメラ視点で得られた点群の座標 (X_p, Y_p, Z_p) をイベントカメラ視点の座標 $\left(X'_p, Y'_p, Z'_p\right)$ を持つ点群 $\boldsymbol{I'}^{\text{latest}}_{\text{depth}}$ へ変換する．この変換は ToF カメラからフレームが得られるたびに別プロセスで姿勢推定および遅延補償処理と並行して行われる．

$$\begin{pmatrix} X'_p \\ Y'_p \\ Z'_p \\ 1 \end{pmatrix} = \mathbf{R}_{\text{event}}\mathbf{R}^{-1}_{\text{ToF}} \begin{pmatrix} X_p \\ Y_p \\ Z_p \\ 1 \end{pmatrix} \tag{5}$$

そして，更新後の二次元姿勢の各関節座標 $J'_{k,m} = \left(x'_{k,m}, y'_{k,m}\right)$ を三次元マップ $\boldsymbol{I'}^{\text{latest}}_{\text{depth}}$ 上で参照し，各関節座標として三次元座標 $J'^{3D}_{k,m} = \left(X'_{k,m}, Y'_{k,m}, Z'_{k,m}\right)$ を持つ三次元姿勢 $\boldsymbol{J'}^{3D}_{m}$ へ再構成を行う．

$$J'^{3D}_{k,m} = \boldsymbol{I'}^{\text{latest}}_{\text{depth}}\left[x'_{k,m}, y'_{k,m}\right] \tag{6}$$

3. 実験条件

3.1 環境設定

実験として，実際のライブステージを模した環境で提案システムを実行した．プロの演者によるダンスシーンを提案システムを用いて姿勢推定し，出力される姿勢推定結果を用いてエフェクト表示をリアルタイムに行った．当日のステージ及び機材配置の概要を図 3 に示す．Helios2+は最大 8.3m までの 3 次元計測が可能である．これを加味し，視野角 69° 以内に撮影対象となる演者が収まるように配置を行なった．またイベントカメラは ToF カメラの脇に設置し，イベントカメラ用の近赤外光源はステージ前方に設置した．また遅延補償 [2] におけるイベントバッファリングは $N_b = 20,000$ ごとに行った．

3.2 評価手法

エフェクト表示に用いる更新後の姿勢 $J'^{3D}_{k,m}$ は，エフェクト表示時の現実の姿勢とどの程度一致していたかを評価する．エフェクト表示時のタイムスタンプを保存しておき，実験終了後にその時刻における姿勢と比較する．正解姿勢として，推論終了時に最も近い時刻で取得された ToF 計測結果に対して姿勢推定を行った結果を用いた．評価指標には，mean per joint position error（MPJPE），probability of correct keypoint（PCK）の 2 種類を用いた．MPJPE は各関節の出力値と正解値の平均誤差である．PCK は出力値が正解から代表長さ内にある割合で，本稿では代表長さを頭部-首の距離とする PCKh を用いた．

4. 実験および結果と考察

4.1 オンライン処理の定性評価

イベントカメラ用のバンドパスフィルタ及び光源を用いたことで，ステージの極端な光源環境および ToF カメラのアクティブセンシング光の影響を受けずに演者の動きのみに関連するイベントを取得できていることが図 4 から分かる．図に示した 1 シーケンス目と 3 シーケンス目では姿勢更新が 2 回となっているのに対し，2 シーケンス目は 4 回の姿勢更新を行っている．これは，2 シーケンス目では動

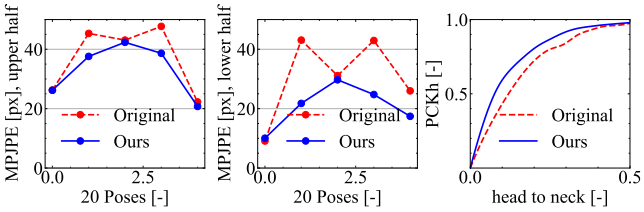

図 5 MPJPE(↓) および PCKh(↑) による定量評価．MPJPE の横軸は遅延補正出力が 20 回出力されるごとの平均である．また上半身は腰より上を，下半身は腰を含み下を評価している．

表 1 提案手法の平均処理時間．

処理内容	処理時間 [ms]
三次元姿勢推定 (2.1)	107
Optical Flow 導出 (2.3)	0.637
イベントカメラへの視点変更 (2.2 式 (3))	0.00477
姿勢更新 (2.3 式 (4))	0.126
三次元マップの視点変換 (2.4 式 (5))	0.951
補償後の三次元リフトアップ (2.4 式 (6))	0.200

きが激しく，姿勢推定処理時間中のイベント数が増加したために適応的に更新回数が増加したことを示している．

4.2 オンライン処理における処理時間の定量評価

オンライン処理における各種処理時間を表 1 に示す．各種処理時間は，ダンスシーケンス中に行った処理全てにおける処理時間を平均している．我々が遅延補償の対象としている三次元姿勢推定の処理時間は 107ms となっている．Optical Flow の導出は三次元姿勢推定の処理時間と並行し行われるため，本稿の実験条件では姿勢推定処理中に最大 150 回程度の差分導出を行うことができる．イベントカメラを用いた遅延補償のために姿勢推定器の処理完了後に行われるプロセスは，イベントカメラへの視点変更，姿勢更新，三次元リフトアップである．これらの合計処理時間は 0.331ms であり，姿勢推定に 100ms 以上の処理時間がかかっていたのに比べて，十分に速い処理時間で遅延補償処理を行うことができる．一方，補償後のリフトアップに用いる三次元マップの視点変換処理は 0.951ms と少々長い処理時間を要した．この処理時間は遅延補償によっても解決できない遅延となるが，元々の姿勢推定処理で発生していた遅延時間の 1%以下であり，大きな問題とはならない．

4.3 姿勢更新による遅延補償の定量評価

MPJPE および PCKh による評価を図 5 に示した．MPJPE において，赤で示した補正前の姿勢の MPJPE の大小は，そのまま動きの大きさと捉えることができる．よって提案手法は，動きの大きさの大小に関係なくある程度姿勢の追従性を向上させることができる．PCKh において，横軸は正解とする閾値の大きさであり，青で示した提案手法のラインは補正前の赤いラインよりも大域的にスコアが向上している．これは MPJPE による評価と同様に，動きの大きさの大小に関係なくある程度姿勢の追従性を向上させることができることを示している．これらから提案手法は適応的な姿勢更新が可能であり，動きの大きさに関係なく追従性を向上できることが見込まれる．

5. 結論

本稿では，近赤外光を用いて撮影したフレームを入力とする姿勢推定の出力に対して，異なる波長帯の近赤外光源を用いたイベントサンプリングを行い，イベントカメラを用いた遅延補償を行うことで，ステージの極端な光源環境に左右されないリアルタイムエフェクトシステムを構築した．姿勢推定処理にかかる時間に由来する姿勢推定誤差に対し，遅延を補償することで姿勢推定の精度が向上することを実験により示した．またイベント数に応じた姿勢更新が，動きの大きさに適応的なプロセスとなり，動きの大小を問わず補償が可能となることを示した．さらに実際のライブシーンでイベントカメラを用いた遅延補償を行い，リアルタイムなエフェクト表示が可能であることを示した．

謝辞

本研究におけるリアルタイムコンテンツの実験は，株式会社ピクス[*1]の坂本 立羽氏（振付，エフェクト演出），上野 陸氏（リアルタイムエフェクトのプログラミング），弓削 淑隆氏（総合プロデュース）の協力の下で実現した．ここに感謝の意を表する．本研究の一部は JST さきがけ JPMJPR2025，JSPS 科研費 JP23K16902, JP24K02953 の支援を受けた．

参考文献

[1] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A. J., Conradt, J., Daniilidis, K. and Scaramuzza, D.: Event-Based Vision: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 1, pp. 154–180 (2022).

[2] Otake, I., Kitano, K., Kushida, T., Kubo, H., Maejima, A., Fujimura, Y., Funatomi, T. and Mukaigawa, Y.: Updating Human Pose Estimation using Event-based Camera to Improve Its Accuracy, *ACM SIGGRAPH 2023 Posters*, SIGGRAPH '23, New York, NY, USA, Association for Computing Machinery (2023).

[3] Xu, Y., Zhang, J., Zhang, Q. and Tao, D.: ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation (2022).

[*1] https://www.pics.tokyo/