

CROSS-CAMERA HUMAN MOTION TRANSFER BY TIME SERIES ANALYSIS

Yaping Zhao^{1,2,3,*}, Guanghan Li^{2,*}, Edmund Y. Lam^{1,3,†}

¹Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR

²Tsinghua University, 30 Shuangqing Road, Beijing, China

³ACCESS — AI Chip Center for Emerging Smart Systems, Hong Kong SAR

ABSTRACT

With advances in optical sensor technology, heterogeneous camera systems are increasingly used for high-resolution (HR) video acquisition and analysis. However, motion transfer across multiple cameras poses challenges. To address this, we propose an algorithm based on time series analysis that identifies motion seasonality and constructs an additive model to extract transferable patterns. Validated on real-world data, our algorithm demonstrates effectiveness and interpretability. Notably, it improves pose estimation in low-resolution videos by leveraging patterns derived from HR counterparts, enhancing practical utility. Code is available at: <https://github.com/IndigoPurple/TSAMT>.

Index Terms— Motion Transfer, Time Series Analysis, Camera Systems

1. INTRODUCTION

Multi-camera systems, driven by advancements in optical sensors, enable computational imaging by combining short-focus and long-focus lenses. These systems capture videos with a wide field-of-view (FoV) and high-resolution (HR) local-view details. By infusing HR details into low-resolution (LR) videos, a balance is achieved between breadth and fine detail [1, 2, 3]. In scenarios like smart cities and live sports, multi-camera systems are widely used to capture multi-scale human-centric videos through hybrid-camera setups. However, transferring human movement data across different cameras, particularly with motion, poses a challenge. In this paper, we tackle this complex problem, illustrated in Figure 1.

Figure 2 showcases a surveillance setup using a multi-camera system to track pedestrian movements across a large FoV video while capturing HR details with local-view cameras. However, integrating multi-scale human-centric videos into a composite video that preserves both a large FoV and HR detail presents challenges. Conventional algorithms

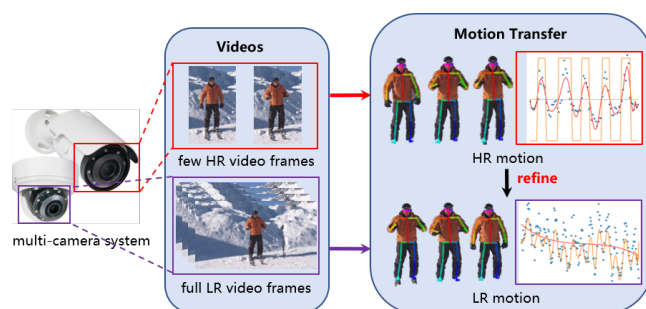


Fig. 1: Our framework utilizes a multi-camera system to capture human-centric videos at multiple scales. By transferring extracted human motion patterns from high-resolution (HR) videos, we enhance pose estimation in low-resolution (LR) videos across different camera feeds. This approach ensures superior pose estimation irrespective of the video resolution.

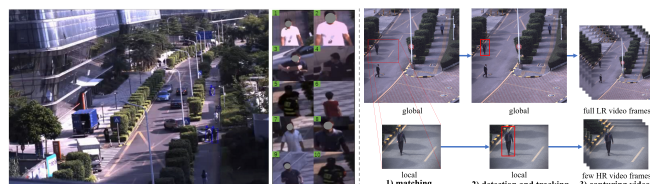


Fig. 2: For surveillance and security, a multi-camera system captures pedestrian trajectories in a large FoV video. Meanwhile, local-view cameras capture HR details of pedestrians.

struggle with nonrigid motion, subpar pose estimation on LR videos, and variations in camera settings, often resulting in blurry outputs. In contrast, deep learning approaches require intensive training data and time.

To address these challenges, we propose an efficient and interpretable motion transfer algorithm tailored for multi-camera systems. Our method, based on time series analysis, eliminates the need for training data and achieves accurate motion transfer. Experimental results on real-world data validate the effectiveness of our approach.

Our main contributions are listed as follows:

- We introduce a motion transfer algorithm based on time series analysis for multi-camera systems, complete

*Equal contribution. †Corresponding author.

This work is supported in part by the Research Grants Council (GRF 17201822), by the Research Postgraduate Student Innovation Award (The University of Hong Kong), and by ACCESS — AI Chip Center for Emerging Smart Systems, Hong Kong SAR.

with a thorough and lucid mathematical formulation.

- Our method is interpretable, avoids training procedures, and provides an algorithmic guarantee.
- Experiments on real-world data and downstream task substantiate the efficacy of our method.

2. RELATED WORK

Synthesis of Human Motion. Prior research has focused on synthesizing human motion through statistical models and learned parameters [4, 5, 6]. However, these methods are limited by the variability in the training dataset, affecting the diversity of generated motion.

Transfer of Human Motion. Deep learning approaches, such as pose-based prediction and image synthesis, have been applied to human motion transfer [7, 8, 9, 10, 11, 12, 13]. However, these methods heavily rely on training data and struggle with real-world data.

Time Series Analysis. Time series analysis has found applications in various fields [14, 15, 16, 17, 18], but its use in motion transfer between multiple cameras is lacking. We bridge this gap by introducing a motion transfer algorithm using time series analysis, with a focus on seasonality analysis [19, 20].

Multi-Camera Systems. Dual-camera systems, popular for their cost-efficiency, have been explored for image fusion and synthesis in super-resolution and denoising tasks [2, 21, 22, 23, 24, 25, 3, 26, 27]. Advanced imaging systems, such as snapshot compressive imaging systems [28, 29, 30, 31, 32, 33, 34], capture fast motion and may employ multiple cameras [35, 36]. However, the challenge of efficient, effective, and explainable motion transfer remains unresolved. Our work addresses this challenge by proposing a motion transfer algorithm for multi-camera systems.

3. METHODOLOGY

In our multi-camera setup, we propose a time series analysis-based algorithm to enhance the LR pose sequence θ^L . We focus on seasonality analysis, leveraging the repetitive patterns observed in human actions like walking, running, and exercises in HR and LR videos.

Our motion transfer algorithm comprises five steps: (1) seasonality identification, (2) additive time series model construction, (3) periodic point detection, (4) additive factor extraction, and (5) motion pattern transfer. We demonstrate the process using $\theta_{1,1}$, representing the ankle joint's first axis-angle in the SMPL model [37]. The data is normalized to the range $[0, 1]$ for analysis.

Figure 3(a) depicts an example of normalized $\theta_{1,1}$ values in a walking sequence with 80 HR frames, while Figure 3(b) displays the corresponding values for 200 LR frames.

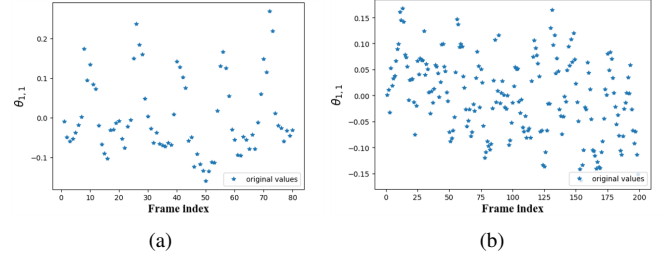


Fig. 3: HR and LR motion data. (a) and (b) represent the $\theta_{1,1}$ value of the HR and LR motion data, respectively.

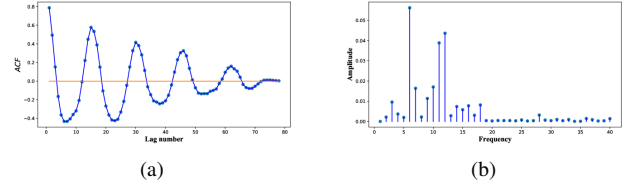


Fig. 4: (a) Auto-correlation of HR motion data shows periodic variation with a maximum value during predefined intervals, gradually decreasing to zero. (b) Fourier analysis of the HR motion data reveals a dominant response at $f = 5$, indicating a reference period of 16 (derived from $80/5$).

3.1. Identification of Seasonality

We begin by utilizing the auto-correlation function (ACF) [38] to identify the cyclical nature of HR motion data. Figure 4(a) illustrates the periodic variations of the ACF curve, reaching peak values at predefined intervals and decaying to zero.

Although the cycle length of human motion is not fixed, it typically revolves around a constant value known as the "reference period" (l). Using Fourier analysis, we estimate l based on the relationship:

$$l = n/f, \quad (1)$$

where n represents the total number of frames in the motion sequence, and f corresponds to the frequency with the strongest response. Figure 4(b) displays the Fourier series of the HR motion data. In this case, the frequency $f = 5$ exhibits the strongest response, resulting in a reference period of 16 frames ($80/5$).

3.2. Construction of Additive Time Series Model

To deconstruct the LR pose values, we employ an additive time-series model [39]. The decomposition is expressed as:

$$\mathbf{Y} = \mathbf{S} + \mathbf{T} + \mathbf{E}, \quad (2)$$

where $\mathbf{Y}$ represents the original pose values, $\mathbf{S}$ captures short-term variations, $\mathbf{T}$ models the long-term trend, and $\mathbf{E}$ accounts for noise. We estimate the long-term trend of the LR

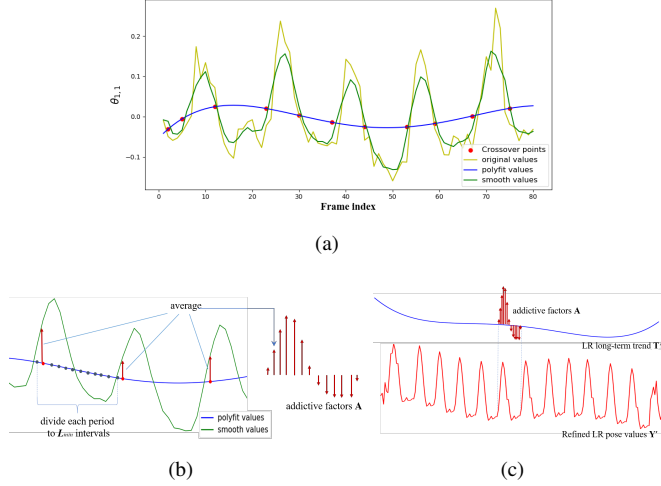


Fig. 5: (a) Crossover points of the blue and dark green curves indicate the locations of the periods, marked as red circles. (b) To refine LR pose values using HR ones, we estimate the additive factor $\mathbf{A}$ by averaging all the periods. (c) By adding the additive factor to the long-term trend $\mathbf{T}^L$, we generate the final poses for refining LR pose values with HR ones.

sequence, $\mathbf{T}^L$, by polynomial fitting using the least squares method:

$$\mathbf{T}^L = c_0 + c_1 \theta_{1,1}^L + c_2 (\theta_{1,1}^L)^2 + \dots + c_{n-1} (\theta_{1,1}^L)^{n-1} + c_n (\theta_{1,1}^L)^n, \quad (3)$$

where $c_0, \dots, c_n$ are constant parameters and $\theta_{1,1}^L$ represents the $\theta_{1,1}$ value of the LR data. The order n of the fitting function is chosen to satisfy $1^{-10} < |c_n| < f$, with c_n as the coefficient of the highest-order term and f denoting the number of periods in a pose value sequence.

3.3. Location of Periodic Points

In the next phase, we employ a moving average to smooth pose values and identify period crossover points (red circles in Figure 5(a)). Period indices within the range $[0.8l, 1.2l]$ are retained, and their averages yield additive factors $\mathbf{A}$.

By adding $\mathbf{A}$ to the long-term trend $\mathbf{T}$, we obtain refined LR poses (Figure 5(b) and Figure 5(c)). To identify frame indices corresponding to crossover points, we detect positive and negative sign changes in the difference between the smoothed pose curve and the trend curve. Imperfections in polynomial fitting are addressed by validating neighboring periodic indices. For each periodic index p , we search for a neighboring index p' satisfying:

$$\begin{aligned} p' &= \arg \min_{p'} |p' - p|, \\ \text{s.t. } ||p' - p| - l| &< (1 - \alpha) \times l, \end{aligned} \quad (4)$$

Here, we set the confidence level α to 80%, with l as the reference period. This validation process ensures neighboring periodic points fall within the confidence interval.

3.4. Extraction of Additive Factor

After identifying the periods in the time series, we calculate the minimum period number l_{min} for both HR and LR sequences and divide each period into l_{min} intervals uniformly. To extract the additive factor for motion pattern transfer, we subtract the long-term trend component from the HR pose values using Equation 2:

$$\mathbf{A} = \mathbf{Y}^H - \mathbf{T}^H = \mathbf{S}^H + \mathbf{E}^H, \quad (5)$$

Here, $\mathbf{A}$ represents the additive factor, composed of short-term variation $\mathbf{S}^H$ and noise component $\mathbf{E}^H$.

Equation 5 enables the extraction of the periodically repetitive pattern $\mathbf{A}$ for motion transfer (Figure 5(b)). To achieve comprehensive motion transfer, we compute the mean value across all periods as the mean additive factor $\bar{\mathbf{A}} = (\bar{a}_1, \dots, \bar{a}_i, \dots, \bar{a}_l)$:

$$\bar{a}_j = \frac{\sum_{i=1}^n \mathbb{1}\{\phi(i) = j\} a_i}{(f \times m)}, \quad a_i \in \mathbf{A}, \quad (6)$$

In Equation 6, a_i refers to the i_{th} value of $\mathbf{A}$ in the i_{th} interval of a period, f is the number of periods, m is the number of values in the i_{th} interval, and ϕ is a correspondence function that maps the frame index i to the j_{th} interval of a period:

$$\phi(i) = j. \quad (7)$$

3.5. Pattern Transfer

In the final phase, we integrate the additive factors $\bar{\mathbf{A}}$ with the LR long-term trend $\mathbf{T}^L$, aligning it with the extracted regular patterns from the HR motion sequence (Figure 5(c)). The optimized LR pose values $\mathbf{Y}' = (y'_1, \dots, y'_i, \dots, y'_n)$ are computed using the formula:

$$y'_i = t_i + \bar{a}_{\phi(i)}, \quad (8)$$

Here, y'_i represents the i_{th} value of the enhanced LR pose, and t_i is the i_{th} value of the LR long-term trend $\mathbf{T}^L$. This pattern transfer approach yields a refined representation of the original LR poses, providing a higher quality approximation of the initial HR sequence.

4. EXPERIMENT

To validate our proposed algorithm's effectiveness, we apply it to real-world data from a dual-camera system [2]. Using a limited number of HR frames and a complete series of LR frames, we initially estimate poses using OpenPose [40], resulting in pose sequences.

In Figure 6, the HR pose sequence is of superior quality (green curves), while the LR one suffers from severe noise. Our motion transfer approach successfully refines the LR pose values, as shown in Figure 6(b). Our results preserve

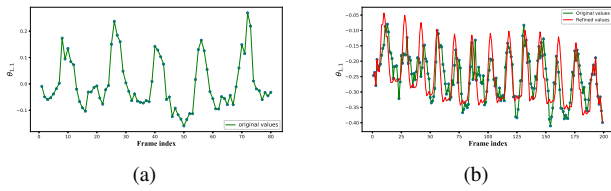


Fig. 6: HR and LR motion data. (a) and (b) represent the $\theta_{1,1}$ value of the HR and LR motion data, respectively.

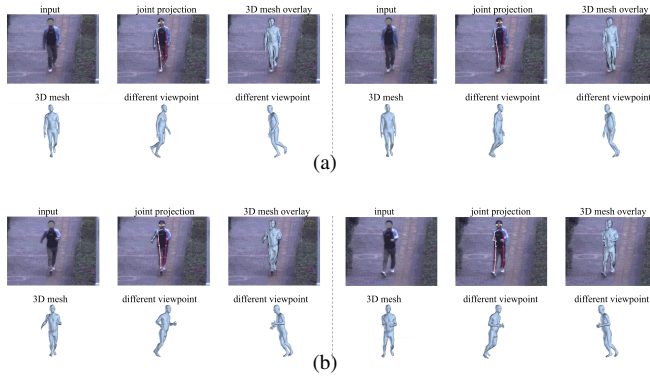


Fig. 7: The SMPL model [41] is adopted to represent 3D body and the HMR [42] is performed to reconstruct 3D human meshes from 2D video frames. (a) and (b) show successful cases and failure cases, respectively.

the long-term correlation in the LR sequence and enhance its quality by transferring motion patterns from the HR sequence.

For 3D body representation, we employ the SMPL model [41] and use the HMR methodology [42] to reconstruct 3D human meshes from 2D frames. While satisfactory results are achieved in straightforward cases (e.g., minimal leg movement, arms close to knees; Figure 7(a)), HMR struggles with accurate estimation when the subject walks briskly and swings their arms vigorously (Figure 7(b)).

To enhance 3D mesh reconstruction and demonstrate the efficacy of our motion transfer algorithm, we apply it to HR and LR real-world videos of a pedestrian walking. Without our method, the reconstructed 3D human meshes exhibit inaccuracies and anomalies due to the LR video quality (Figure 8). However, our method corrects these human poses.

Furthermore, we integrate the motion transfer method into a downstream vision task of embedding HR human details into LR videos. Without our algorithm, direct synthesis of human details is inaccurate, resulting in jittering and unnatural poses (Figure 9). However, applying motion transfer reduces motion jittering, demonstrating the utility of our approach.

Given the unavailability of ground truths of human body joints in real-world scenarios, direct quantitative evaluations are infeasible. However, by seamlessly integrating our approach with downstream tasks, we facilitate quantitative assessments. To this end, we employ video super-resolution



Fig. 8: Our motion transfer technique improves 3D human mesh reconstruction performance. In the LR video, we observed misaligned arm positions. However, the results with motion transfer effectively correct the human poses.

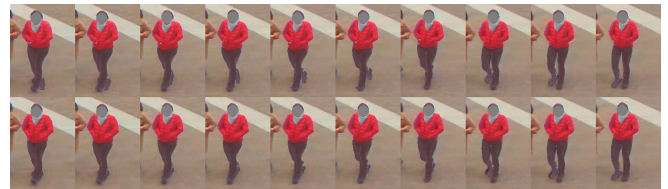


Fig. 9: Performance on synthesizing human details across heterogeneous cameras. Top: without our motion transfer method, jitters are observed on the legs. Bottom: with ours.

with texture transfer as a downstream task to quantitatively evaluate the effectiveness of our approach. For detailed experimental results, please refer to the supplementary material, which is provided due to space constraints:

github.com/IndigoPurple/TSAMT/blob/main/sm/sm.pdf.

5. CONCLUSION

We present an algorithm for motion transfer among multiple cameras, based on time series analysis. Our approach has a comprehensive and well-defined mathematical formulation and does not necessitate training data, thereby ensuring efficiency and interpretability. Through rigorous evaluations on real-world data, we showcase the effectiveness of our method.

6. REFERENCES

- [1] X. Yuan, L. Fang, Q. Dai, D. J. Brady, and Y. Liu, "Multiscale gigapixel video: A cross resolution image matching and warping approach," in *IEEE ICCP*, 2017, pp. 1–9.
- [2] G. Li, Y. Zhao, M. Ji, X. Yuan, and L. Fang, "Zoom in to the details of human-centric videos," in *IEEE ICIP*, 2020.
- [3] Y. Zhao and et al., "EFENet: Reference-based video super-resolution with enhanced flow estimation," in *CICAI*, 2021.
- [4] J. Chai and J. K. Hodgins, "Constraint-based motion optimization using a statistical dynamic model," in *ACM SIGGRAPH 2007 papers*, 2007, pp. 8–es.
- [5] J. Min and J. Chai, "Motion graphs++ a compact generative model for semantic motion analysis and synthesis," *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 1–12, 2012.
- [6] M. Lau, Z. Bar-Joseph, and J. Kuffner, "Modeling spatial and temporal variation in motion data," *ACM TOG*, 2009.
- [7] R. Villegas and et al., "Learning to generate long-term future via hierarchical prediction," in *international conference on machine learning*. PMLR, 2017, pp. 3560–3569.
- [8] R. Villegas, J. Yang, and et al., "Neural kinematic networks for unsupervised motion retargeting," in *IEEE CVPR*, 2018.
- [9] L. Ma and et al., "Pose guided person image generation," *arXiv preprint arXiv:1705.09368*, 2017.
- [10] L. Ma, Q. Sun, and et al., "Disentangled person image generation," in *Proceedings of the IEEE CVPR*, 2018, pp. 99–108.
- [11] A. Siarohin and et al., "Deformable gans for pose-based human image generation," in *Proceedings of the IEEE CVPR*, 2018.
- [12] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, "Everybody dance now," in *Proceedings of the IEEE/CVF ICCV*, 2019.
- [13] W. Liu and et al., "Liquid warping gan: A unified framework for human motion imitation, appearance transfer and novel view synthesis," in *Proceedings of the IEEE/CVF ICCV*, 2019.
- [14] W. Cleveland, "Analysis and forecasting of seasonal time series." 1973.
- [15] R. G. Brown, *Smoothing, forecasting and prediction of discrete time series*. Courier Corporation, 2004.
- [16] E. Ghaderpour, W. Liao, and M. P. Lamoureux, "Antileakage least-squares spectral analysis for seismic data regularization and random noise attenuation," *Geophysics*, 2018.
- [17] M. R. Ahmed and et al., "Introducing a new remote sensing-based model for forecasting forest fire danger conditions at a four-day scale," *Remote Sensing*, vol. 11, no. 18, p. 2101, 2019.
- [18] E. Ghaderpour and S. Ghaderpour, "Least-squares spectral and wavelet analyses of v455 andromedae time series: The life after the super-outburst," *Publications of the Astronomical Society of the Pacific*, vol. 132, no. 1017, p. 114504, 2020.
- [19] S. Hylleberg, *Modelling seasonality*. Oxford University Press, 1992.
- [20] K. F. Wallis, "Seasonal adjustment and relations between variables," *Journal of the American Statistical Association*, 1974.
- [21] Y. Zhao, H. Zheng, and et al., "MANet: improving video denoising with a multi-alignment network," in *IEEE ICIP*, 2022.
- [22] Y. J. Jung, "Enhancement of low light level images using color-plus-mono dual camera," *Optics express*, 2017.
- [23] H. W. Jang and Y. J. Jung, "Deep color transfer for color-plus-mono dual cameras," *Sensors*, vol. 20, no. 9, p. 2743, 2020.
- [24] H. Zheng, M. Ji, H. Wang, Y. Liu, and L. Fang, "Crossnet: An end-to-end reference-based super resolution network using cross-scale warping," in *ECCV*, 2018, pp. 88–104.
- [25] Y. Tan, H. Zheng, Y. Zhu, X. Yuan, X. Lin, D. Brady, and L. Fang, "Crossnet++: Cross-scale large-parallax warping for reference-based super-resolution," *IEEE TPAMI*, 2020.
- [26] Y. Zhao, H. Zheng, M. Ji, and R. Huang, "Cross-camera deep colorization," in *CAAI International Conference on Artificial Intelligence*. Springer, 2022, pp. 3–17.
- [27] Y. Zhao, H. Zheng, J. Luo, and E. Y. Lam, "Improving video colorization by test-time tuning," in *IEEE ICIP*, 2023.
- [28] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for video snapshot compressive imaging," *arXiv preprint arXiv:2201.06931*, 2022.
- [29] Q. Yang and Y. Zhao, "Revisit dictionary learning for video compressive sensing under the plug-and-play framework," in *SPIE Proceedings*, vol. 12166. SPIE, 2022, pp. 2018–2025.
- [30] Y. Zhao, "Mathematical cookbook for snapshot compressive imaging," *arXiv preprint arXiv:2202.07437*, 2022.
- [31] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for snapshot compressive imaging," in *AAAI*, vol. 37, no. 3, 2023.
- [32] Y. Zhao, Q. Zeng, and E. Y. Lam, "Adaptive compressed sensing for real-time video compression, transmission, and reconstruction," in *IEEE DSAA*. IEEE, 2023.
- [33] Z. Y. Chong, Y. Zhao, Z. Wang, and E. Y. Lam, "Solving inverse problems in compressive imaging with score-based generative models," in *IEEE DSAA*. IEEE, 2023.
- [34] Y. Zhao and E. Y. Lam, "SASA: Saliency-aware self-adaptive snapshot compressive imaging," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024.
- [35] L. Wang and et al., "Dual-camera design for coded aperture snapshot spectral imaging," *Applied optics*, 2015.
- [36] J. Hauser and et al., "Dual-camera snapshot spectral imaging with a pupil-domain optical diffuser and compressed sensing algorithms," *Applied Optics*, vol. 59, no. 4, 2020.
- [37] M. Loper and et al., "SMPL: A skinned multi-person linear model," *ACM transactions on graphics (TOG)*, 2015.
- [38] C. I. Plosser, "Short-term forecasting and seasonal adjustment," *Journal of the American Statistical Association*, 1979.
- [39] A. Maravall and D. A. Pierce, "A prototypical seasonal adjustment model," *Journal of Time Series Analysis*, 1987.
- [40] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 1, pp. 172–186, 2019.
- [41] M. Loper and et al., "SMPL: A skinned multi-person linear model," *ACM transactions on graphics (TOG)*, 2015.
- [42] A. Kanazawa and et al., "End-to-end recovery of human shape and pose," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7122–7131.