

НЕЙРОСЕТЕВЫЕ ПОДХОДЫ К ВОССТАНОВЛЕНИЮ ТРЕХМЕРНОЙ ПОЗЫ ЧЕЛОВЕКА ПО ДВУМЕРНЫМ ДАННЫМ

Белаш В. А.

Кафедра информационных технологий в бизнесе,
факультет социально-экономических и компьютерных наук,
Национальный исследовательский университет Высшая школа экономики (Пермский филиал)
Пермь, Российская Федерация
E-mail: vabelash@edu.hse.ru

В статье исследованы современные нейросетевые подходы для восстановления трехмерной позы человека на основе двумерных данных. Рассматриваются такие нейросетевые архитектуры, как SPIN, VideoPose3D и PoseFormer. Проведен их сравнительный анализ. Выявлено, что модели, использующие временной контекст (VideoPose3D, PoseFormer), демонстрируют преимущества при анализе динамических движений по сравнению со статическими методами (SPIN). Ключевые слова: нейросетевые модели, компьютерное зрение, распознавание движений, 2D-to-3D-lifting, спортивная аналитика.

ВВЕДЕНИЕ

Современная спортивная аналитика активно использует методы распознавания позы человека с помощью компьютерного зрения. Модели, ориентированные на работу с двумерными скелетными моделями, демонстрируют значительные успехи. Однако в ходе работы, направленной на распознавание и оценку движений фигуристов [1], было выявлено явное ограничение двумерного подхода. Оно заключается в сложности двумерного представления скелетных данных для анализа пространственных характеристик движения.

Во-первых, двухмерный скелет не позволяет надежно измерить ключевые показатели (например, поворот корпуса в вертикальной плоскости или пространственную траекторию движения). Во-вторых, сложная координация в таком виде спорта, как фигурное катание требует одновременного учета симметрии обеих сторон тела спортсмена (в двумерной плоскости для этого требуется производить съемку с разных сторон и затем объединять полученные кадры). Различные вращения, прыжки, расположение свободной ноги невозможно качественно проанализировать без трехмерного понимания биомеханики движений человека.

Актуальность преобразования двумерных поз в трехмерные также обусловлена заметным дисбалансе существующих наборов данных: на сегодняшний день гораздо больше размеченных 2D-датасетов по сравнению с 3D. Процесс разметки двумерных скелетов требует меньше трудозатрат, чем создание точных 3D-аннотаций с привлечением сложного оборудования и специализированного программного обеспечения. Это усиливает интерес к нейросетевым методам трансформации двумерных изображений в трехмерные.

Переход к трехмерному анализу предполагает использование более сложных вычислительных – «тяжеловесных» – моделей, что закономерно приводит к увеличению требований к вы-

числительным ресурсам. Однако для тех задач спортивной аналитики, где ключевое значение имеют точность оценки техники выполнения элементов и контроль биомеханически корректных положений, такое повышение вычислительной нагрузки оправданно.

I. ОПИСАНИЕ ПОДХОДОВ ПРЕОБРАЗОВАНИЯ ДВУМЕРНЫХ СКЕЛЕТОВ

Основой для многих нейросетевых методов реконструкции трехмерной позы из двумерных данных является параметрическая модель SMPL – Skinned Multi-Person Linear Model [5]. Модель обучена на обширном наборе из трехмерных сканов человеческих тел и генерирует 3D меш человеческого тела с 6890 вершинами на основе двух наборов параметров: позовых (75 параметров, включающих вращения 24 суставов) и форменных (10 параметров, описывающих вариации телосложения).

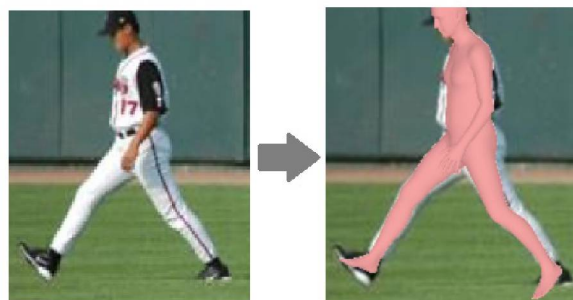


Рис. 1 – Трехмерное восстановление с помощью Synchronized Pose Image Network [2]

Так, нейросеть Synchronized Pose Image Network (SPIN) основана на интеграции нейросетевого регрессора с итеративной оптимизацией параметров модели SMPL. Особенность этого нейросетевого подхода заключается в реализации механизма самообучения, при котором 3D-модель человека устанавливается в двумерное пространство для соответствия исходным ключевым точ-

кам поданного на вход двумерного скелета (как изображение или в формате *json), чтобы последовательно уточнить параметры позы и формы тела в процессе обработки исходного одиночного изображения.

Пока модель SPIN фокусируется на статическом анализе, нейросеть VideoPose3D [3] от Facebook Research обрабатывает временные последовательности, принимая на вход последовательность 2D-поз. Архитектура строится на Temporal Convolutional Network, чье применение позволяет анализировать движение в рамках расширенного временного окна. Это дает не только повышенную точность реконструкции, но и плавность, устойчивость к шумам и естественность восстановленной 3D-анимации движения человека.

Современное развитие трансформерных архитектур нашло отражение в модели PoseFormer [4], реализующей пространственно-временной анализ последовательностей двумерных ключевых точек скелета. Архитектура использует механизм внимания на двух уровнях: пространственное внимание учитывает взаимосвязи между суставами в пределах отдельного кадра, а временное «внимание» отслеживает последовательность траектории движения отдельных суставов.

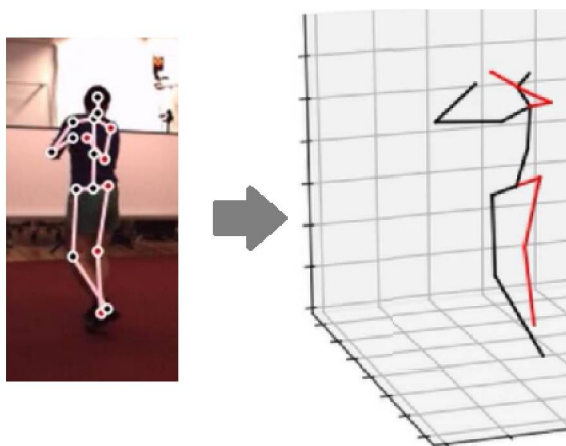


Рис. 2 – Пример преобразования с помощью PoseFormer [4]

II. СРАВНЕНИЕ НЕЙРОСЕТЕВЫХ ПОДХОДОВ

Описанные ранне подходы демонстрируют различные стратегии обработки данных: от работы с одиночными изображениями до анализа пространственно-временных видеопоследовательностей. Таблица 1 представляет собой результат сравнительного анализа.

Выбор критериев обусловлен спецификой задач спортивной аналитики (например, для анали-

за динамических движений в спорте важен учет временных зависимостей).

Таблица 1 – Сравнительный анализ

Критерий	SPIN	VideoPose3D	PoseFormer
Inputs	1 кадр	послед-ть 2D-поз	послед-ть 2D-поз
Outputs	3D-поза + меш	3D-скелет	3D-скелет
Временной контекст	–	свертки	внимание
Учет протр. контекста	огранич.	огранич.	полный
Ключевая идея	SMPL	TCN	SpTemp

ЗАКЛЮЧЕНИЕ

Перспективным направлением развития подходов к восстановлению трехмерной позы человека по двумерным данным представляется создание гибридных архитектур, сочетающих преимущества рассмотренных подходов: точность статических методов, временную согласованность видеопоследовательностей и способность учитывать пространственно-временные зависимости. Кроме того, важно адаптировать методы 3D-реконструкции для специфических требований различных спортивных дисциплин. Например, при анализе техники движений фигуристов разумно учитывать не только биомеханику тела спортсмена, но и положение коньков относительно ледовой поверхности, что требует расширения стандартных моделей скелетов.

1. Ученые НИУ ВШЭ в Перми учат искусственный интеллект разбираться в фигурном катании [Электронный ресурс] // Новости науки в НИУ ВШЭ. – 2025. – Режим доступа: <https://www.hse.ru/news/science/1034088150.html> (дата обращения: 25.10.2025).
2. Kolotouros, N. Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop / N. Kolotouros, G. Pavlakos, K. Daniilidis // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 4501-4510.
3. Pavlo, D. 3D human pose estimation in video with temporal convolutions and semi-supervised training / D. Pavlo, C. Feichtenhofer, D. Grangier, M. Auli // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 7753-7762.
4. Zheng, C. 3D Human Pose Estimation with Spatial and Temporal Transformers / C. Zheng, S. Zhu, M. Mendieta, T. Yang, C. Chen, Z. Ding // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 11656-11665.
5. Loper, M. SMPL: A Skinned Multi-Person Linear Model / M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. J. Black // ACM Transactions on Graphics (TOG). – 2015. – Vol. 34, № 6. – Article 248. – P. 1-16.