Stylistic Analyses of Human Pose in Theatrical Performances: Computational and Historical Frameworks

Peter Broadwell[1], Michael Rau[2], and Simon Wiles[1]

[1] Research Data Services, Stanford University, Stanford, USA
[2] Theater and Performance Studies, Stanford University, Stanford, USA

Abstract

This research project employs machine learning and computer vision to analyze directorial styles and other aspects of theater performances through the lens of pose and action recognition. By applying these techniques to video recordings of theatrical performances, we compare multiple performances per director to identify distinctive patterns in choreography and staging. Our approach combines distant and close viewing methodologies, allowing for a nuanced understanding of theatrical gestures and movements. By comparing different directors' uses of pose, we aim to quantify aspects of the elusive concept of directorial style. This interdisciplinary project bridges gaps between performing arts, computer science and digital humanities, offering computational insights into theatrical analysis and refining our understanding of directorial signatures in live performance.

Keywords: pose estimation, theater studies, semantic embeddings

1 Introduction

In theater studies, pose is so fundamental that it is often taken for granted, yet together with staging, it lies at the intersection of authorial intent and directorial vision. An actor's posture and gestures are affected by design choices of the costumes, set and lights and mediated by the performer's body, but directors, performers and audiences intuitively understand the power of well-crafted tableaux or precisely choreographed movement sequences. Certain iconic poses define productions: Brecht's silent scream choreography in *Mother Courage and Her Children*, Bob Fosse's shoulder rolls and arm pops in *The Pajama Game*, or the ensemble poses in *A Chorus Line*. These indelible poses make works memorable and also serve as a shorthand for identifying directorial style.

Our research aims to address the lack of concrete discussions of pose in theater studies by devising novel ways of quantifying and analyzing actors' physical arrangements and movements. We subsequently situate these observations within theatrical traditions and dominant schools of artistic influence, focusing specifically upon a few well-known "auteur" directors and attempting to isolate their stylistic signatures computationally.

2 Methods

Advancements in deep neural network architectures have made possible the computational study of pose and gesture from "in-the-wild" video recordings of live theatrical performances across the full history of motion pictures. Specifically, vision transformer-based estimators of human body,

Peter Broadwell, Michael Rau, and Simon Wiles. "Stylistic Analyses of Human Pose in Theatrical Performances: Computational and Historical Frameworks." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 1408–1419. https://doi.org/10.63744/GxQbSDPPyqvL.

hand and face landmarks from monocular images [10], combined with further transformer-based end-to-end models to track figures and camera positions over time [13], have opened up to analysis theatrical video corpora from the early twentieth century to the present. The aforementioned technological advances in deep learning have expanded the capabilities of these methods to interpolate occluded poses and to infer three-dimensional spatial relations. Such proficiencies facilitate further efforts to quantify motion, segment tracks into gestures, and apply pretrained or de novo semantic embedding systems [9] and per-figure action classification models [12] to pose data.

2.1 Dataset

We have developed a full-stack system that applies the feature extraction tools mentioned above to generate per-frame pose and action recognition data from recordings of full-length theatrical performances. The recordings are typically unmodified, other than being upscaled to high-definition video when necessary. Their data are further augmented via the application of embedding models and visual shot transition [17], face [15] and hand [11] detection tools. All such outputs are subsequently loaded into a Postgres database running the `pgvector` extension to index the feature vectors for fast nearest-neighbor search.

For the current study, we have indexed 31 performances from three contemporary "auteur" directors (see Table 3), comprising 10-25 hours of recordings per director. This bespoke dataset contains millions of pose-related feature vectors, which we analyze via Python scripts.

2.2 The challenge: directorial stylometry in theater

Unlike stylistic analyses of texts, music, and images, human body and hand poses and gestures in video sources are not as readily segmented into discrete units of meaning. The extra dimensionality of this domain necessitates large-scale statistical analyses involving ground-truth labels even to identify the sources of indicators that may lead to the distillation of significant stylistic phenomena. Specifically, the current work applies pose estimation to identify style-related elements of the deployment of body postures and hand gestures by attempting to train classifiers to solve a director-to-performance stylometric classification problem. Applied to the corpus, we evaluate classifiers on held-out performances and identify features with the greatest impact on accuracy.

A related avenue of exploration involves the application of a nineteenth-century theory of pose and gesture developed by François Delsarte (1811-1871), a self-taught expert on oratory. Delsarte's system, part pedagogical acting manual and part philosophical exegesis, subsequently evolved into more prescriptive forms through the efforts of his devotees, among whom numbered some of the most influential progenitors of American modern dance [16]. The system ultimately came to comprise sets of pose and hand gesture archetypes, mapped to salient semantic axes of human expression and emotion. We present the current state of efforts to explore these axes, and in particular to make use of the matrices of exemplary body and hand positions that Delsarte's followers devised to occupy pivotal elements of such spectra. This approach holds considerable potential for clarifying aspects of theatrical pose stylometry and influence upon performances contemporary with the advent of the "Delsarte System" and those from later eras.

3 Related Work

Detection of poses in substantial collections of artworks has resulted in rather large-scale studies of the depiction of still human poses across periods of art history [14] [8]. Previous computer-assisted explorations of human pose and gesture in the performing arts largely have been limited to small-scale visualizations of primarily choreographic works using data collected via special-purpose motion-capture systems [4]. Any style-related signals captured in this way are more likely

	Motion & distance	Pose embeddings	3D global coordinates	Action embeddings
LOO cosine similarity	51.61% (16/31)	77.42% (24/31)	74.19% (23/31)	80.65% (25/31)
10-fold Random Forest	57.5% stdev: 37.4%	66.6% stdev: 33.7%	66.6% stdev: 33.7%	68.3% stdev: 31.3%
10-fold Gaussian NB	76.7% stdev: 26.0%	75.3% stdev: 25.5%	72.1% stdev: 25.3%	76.9% stdev: 24.5%

Table 1: Accuracies of the specified classification approaches given different pose-related features; statistics from the Random Forest and Gaussian Naive Bayes-based classifiers reflect 10-fold cross validation. Approximately 33% accuracy would be expected by chance.

to find application in "style transfer" experiments involving the synthesis of dance moves/motions via various generative methods, from heuristic algorithms to deep neural networks [2]. The development of more sophisticated methods for indexing such data has benefited from recent work in computational museology involving archives of motion capture data of cultural forms such as martial arts, which are subsequently indexed for searching via query "cues" involving motion and pose [6]. Some systems for interactive camera-based dance pedagogy also have employed relevant pose indexing and search approaches [7]; studies of K-pop dance in particular have applied temporal-convolutional methods of human pose estimation to study fairly substantial corpora of dance performances in aggregate [3].

A relevant prior inquiry involving the stylometric analysis of biomechanical data in performance was a study by Escobar Varela and Hernández-Barraza that focused on a specific aspect of Javanese dance. The research used a motion capture system to record exemplary standing motions of several character types, subsequently analyzing the observed joint motion and rotation data to determine whether and how these quantitative observations, singly or in aggregate, could differentiate the character types [5]. Our work pursues the potential of this approach to scale up to much larger corpora via emergent "markerless" motion tracking technologies. As such, this study is the first to apply modern vision transformer-based pose estimation data and embeddings to a large collection of in-the-wild theater performances.

4 Results

Our initial tests involved building leave-one-out classifiers based upon per-director averages of particular feature vectors across every pose of the directors' oeuvres (excepting the held-out work). These were then compared to the average feature vector from the held-out performance via cosine similarity. These tests indicated that the 16-element view-invariant pose embeddings and 60-element action embeddings were most effective at assigning the held-out performance to the correct director.

Classifications based upon the normalized 3D body keypoints of the poses and motion and distance features within the performances at first seemed to be more responsive to genre than to each director's style. The features included average average in-place motion and sidereal (relative to the background) motion and the average distance between figures in populated frames. These features initially were only strongly proficient at differentiating the works directed by Bill T. Jones, a primarily choreographic director, from those of Romeo Castellucci and Kryztof Warlikowski, who tend to direct operas and other stage plays with less dynamic motion and poses. Subsequent classification tests involving 10-fold cross-validation with Random Forest- and Gaussian Naive

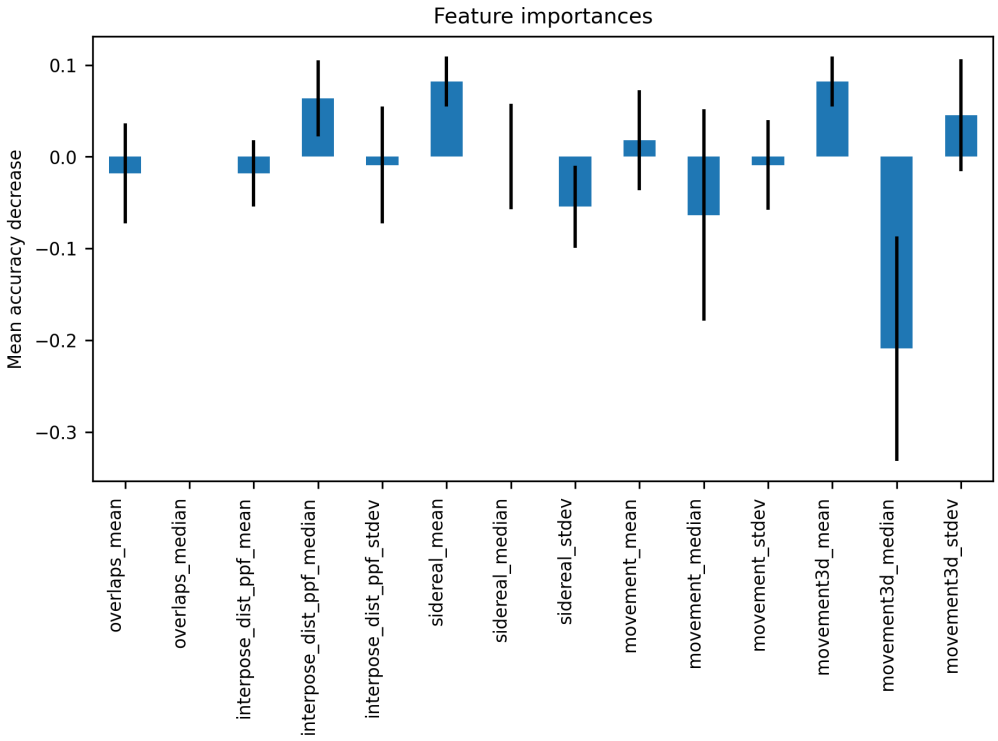

Figure 1: Directorial classification importances among motion and distance features calculated via mean decrease in accuracy when those elements are excluded during Random Forest classifier tests.

Bayes-based classifiers using the same features, however, found that in some cases the averaged motion and distance features and 3D global keypoint coordinates were nearly as effective at assigning held-out performances to the proper director (Table 1).

Feature importance tests among the motion and distance features based upon mean decrease in accuracy when particular features are excluded gave limited support to the notion that some movement features, as well as the median inter-pose distance, could be especially salient when differentiating between directors (Figure 1). A collinearity analysis indicated that all motion statistics (in-place or sidereal, 2D or 3D) were broadly collinear, as were the inter-pose distance metrics, suggesting that each of these likely could be collapsed to a single feature in future studies (Figure 2).

No strongly important features were evident in tests among vectors made up of averaged normalized 3D keypoint coordinates. We note, however, that such tests conducted with an earlier version of the corpus indicated that aspects of the positioning of the right wrist were likely to correlate to particular directors. The version of the corpus used in that phase included approximately 20% more poses whose armatures were highly extrapolated due to occlusion; they subsequently were excluded.

We did not attempt feature importance evaluations among the 60-element action recognition vectors, due to the difficulties inherent to interpreting such a large vector space.

We observed that a few elements of the view-invariant pose embeddings, which on the whole were most effective at differentiating between directors, were highly salient to such classifications (Figure 3). Yet the difficulties of interpreting such embeddings precluded drawing any immediate conclusions. These findings did, however, motivate further explorations of this embedding space. A UMAP projection of a sample of pose embeddings from the full corpus (Figure 4) suggests that

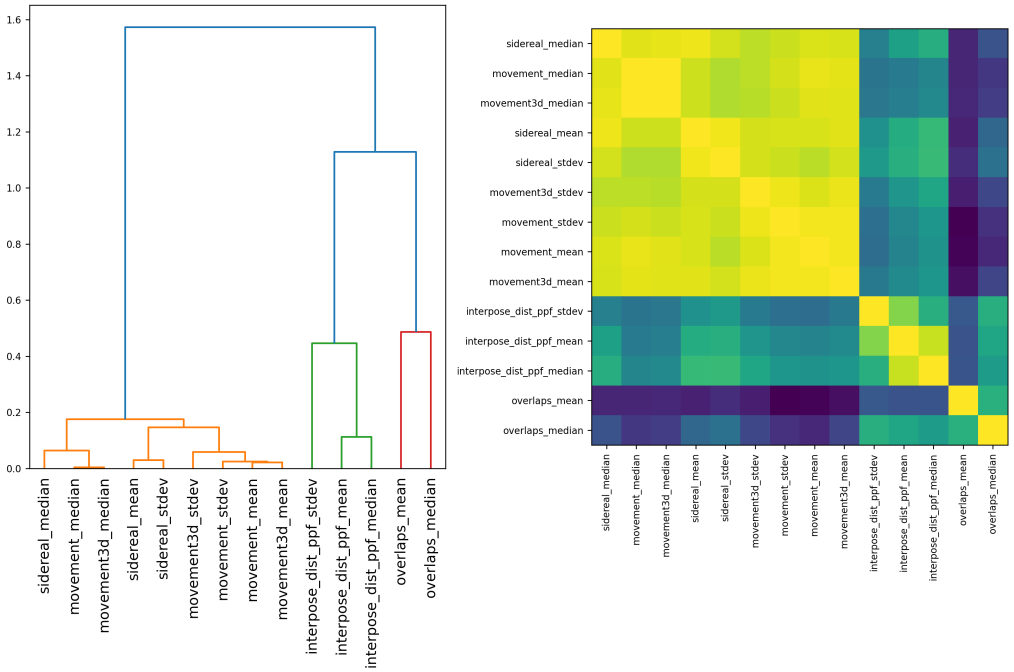

Figure 2: Feature collinearity among motion and distance features of directors' performances.

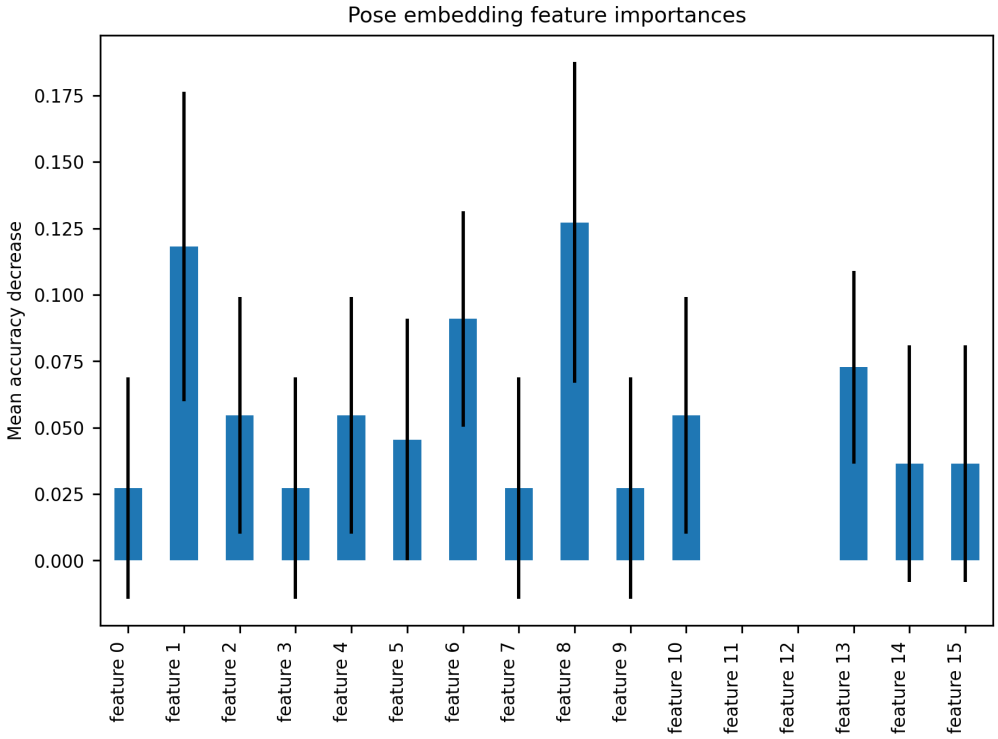

Figure 3: Feature importances of averaged view-invariant pose embedding vectors, quantified via mean decrease in accuracy when features are excluded during Gaussian Naive Bayes classification tests.

portions of the directors' pose "repertoires" occupy distinct portions of the embedding space, even

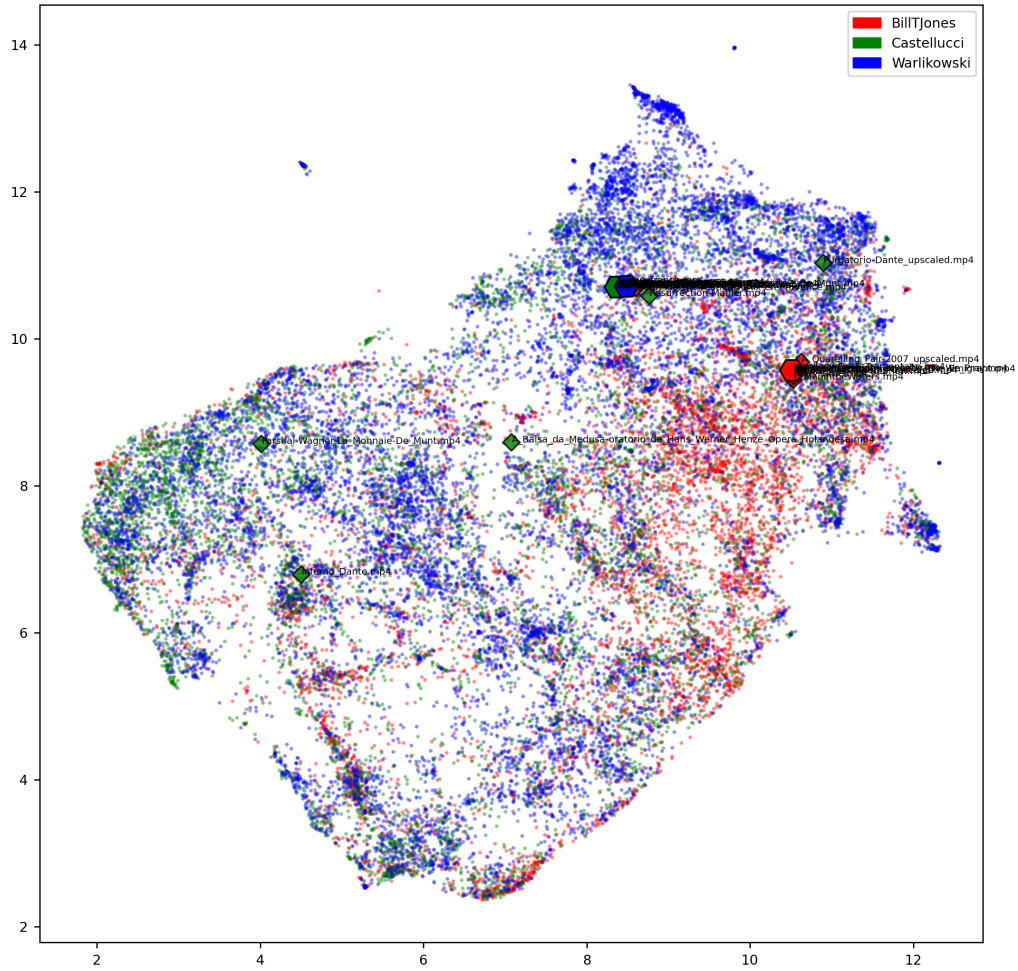

Figure 4: A UMAP projection of view-invariant pose embeddings (16-element vectors) of a 1% sample of 3 million total key poses detected in a set of 31 performance videos directed by one of three contemporary theater directors. The average embeddings of each director's entire sub-corpus are represented by large hexagons, and the average embeddings of each work are represented by diamonds.

if the works of Castellucci and Warlikowski are overall much more similar in their pose embedding projections to each other than they are to those of Jones.

It is tempting to cherry-pick the poses from our extensive pose database that are the most similar to each director's average embedding and to point to them as being representative of particular idiosyncrasies of that director's style. A more systematic inquiry into the disposition of the embedding space is likely to yield more durable insights, however. Although this work is ongoing, there are initial indications that overlaying a set of pose and perhaps of hand gesture archetypes upon such an embedding space, as a way of orienting the space along specific semantic axes, may ultimately improve its interpretability. The Deslarte archetypes described above offer one such system. Figure 5 provides visualizations of these archetypes and an indication of how they can be used to build a "fingerprint" of the average prevalence of each archetype in a single performance, as calculated via each archetype's average cosine similarity to all poses or hand gestures in the performance.

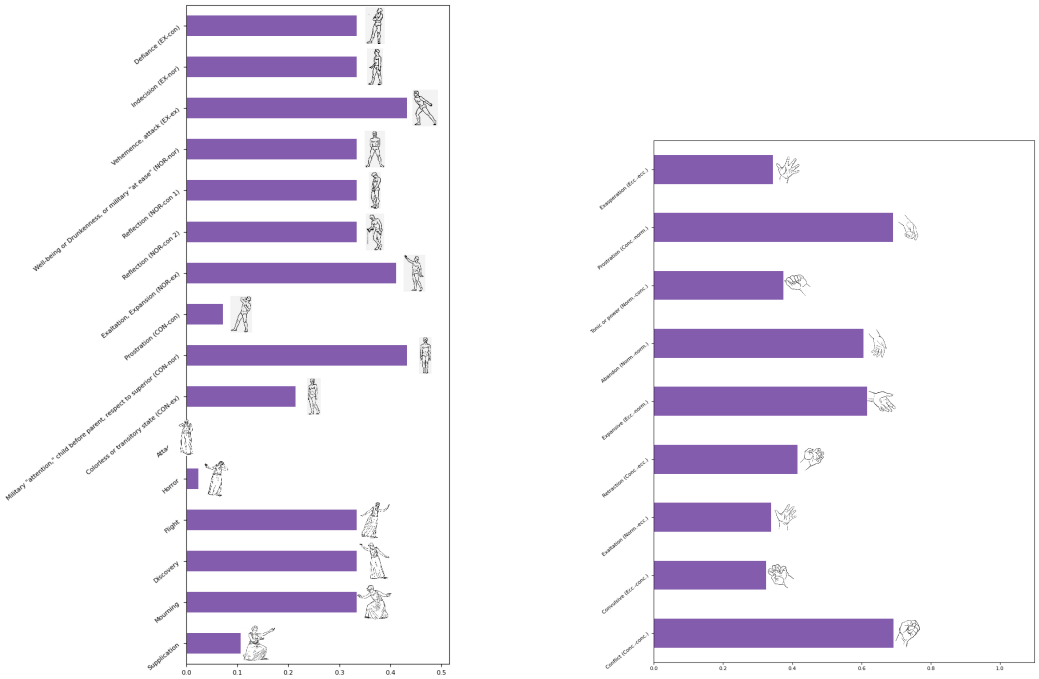

Figure 5: Sample "fingerprints" of the average similarities of the body poses (left) and hand armatures (right) from a single excerpted performance to the labeled emotional/expressive archetypes from the Delsarte system.

5 Further Work

As an indication of a potential direction for future inquiries, Table 2 provides some initial tallies of classification results when the Delsarte "fingerprints" based upon the feature classes discussed above are used to train classifiers of directorial styles. Only the classifiers using fingerprints calculated based upon the view-invariant pose embeddings consistently approach the classification accuracy of the raw features they are based upon (and some of the hands-based classifiers perform worse than chance). Yet the potential of these pose system "overlays" to improve the interpretability of the embedding space, as well as the Delsarte system's legacy as a foundational influence on modern theater, motivate further investigation along these lines.

	Pose 3D global coordinates	Pose embeddings	Hand joint angles	Hand embeddings
LOO cosine similarity	70.97% (22/31)	64.52% (16/31)	48.39% (15/31)	48.39% (15/31)
10-fold Random Forest	44.3% stdev: 28%	70% stdev: 28.1%	27.7% stdev: 27.1%	44.2% stdev: 34.1%
10-fold Gaussian NB	48.3% stdev: 30.2%	57.5% stdev: 30.4%	26.7% stdev: 35.9%	55% stdev: 29.9%

Table 2: Accuracies of the specified classification approaches given "fingerprints" based upon the average of pose and hand feature similarities to the Delsarte archetypes in the relevant latent space; statistics from Random Forest and Gaussian Naive Bayes reflect 10-fold cross validation. 33% accuracy would be expected by chance.

There is an urgent need to pursue the application of statistical profiling methods, ideally targeted towards an even more extensive corpus than the present study, which are more sophisticated than simple vector averages, descriptive statistics and cosine similarities. Furthermore, a more systematic attempt to model "gestures" as variable-length sequences of similar poses that unfold over time, rather than simply considering them to be conceptually equivalent to static poses, is arguably overdue and has antecedents in motion capture-based studies [1]. Integrating action-recognition embeddings into this effort is also worth pursing.

Other empirical tests of the effectiveness of computational pose stylometry might involve comparing the accuracy of actual human scholars versus machine classifier models when both are tasked with identifying the directors of excerpts from previously unseen performances. Similarly, comparing the likelihood of human experts to detect the presence of particular Delsarte pose archetypes to the archetype prevalence values outputted by our pose recognition system also would offer a more focused evaluation of the historically informed stylometric approaches outlined in this paper.

Another follow-on experiment, which could provide an alternative control baseline to the assumption of random guessing, would be to compare the accuracy of our directorial classifiers based on poses, actions, motion and spatial relations to those trained on specifically non-stylistic aspects of performances, such as average duration or size of the cast.

6 Conclusion

This ongoing work represents a novel attempt to apply stylistic analyses to a large bespoke dataset of pose data generated with deep-learning technologies from in-the-wild theatrical performances. The range of potential enhancements and follow-up studies is as expansive as the set of potential features extracted via these state-of-the-art tools. Yet by focusing on isolating features that are most relevant to the quantification of directorial stylistic preferences, and by situating them within relevant historical frameworks, this study highlights some of the most promising avenues for further computational exploration of the compelling topic of theatrical pose.

Performance	Duration	Poses	Shots	Tracks	Poses / pop. frame	Inter -pose dist / pop. frame (m)	Side-real move-ment (m/s)	3D move-ment (m/s)
Bill T. Jones								
A Letter to My Nephew	1:09:02	308088	698	182	3.373	2.339	0.652	1.645
Analogy/Dora: Tramontane	1:17:12	411046	2136	1794	3.269	1.851	0.534	1.361
A Quarrelling Pair	1:17:50	267257	759	729	2.144	1.722	0.753	1.716
D-Man in the Waters	0:35:10	141191	855	110	3.125	2.559	0.540	1.235

Continued on next page

Table 3 – continued from previous page

Performance	Duration	Poses	Shots	Tracks	Poses / pop. frame	Inter-pose dist / pop. frame (m)	Side-real move-ment (m/s)	3D move-ment (m/s)
Fondly Do We Hope... Fervently Do We Pray	1:23:55	435160	1001	846	2.872	2.580	1.115	2.508
Holzer Duet... Truisms	0:17:45	43206	88	65	3.158	2.009	0.539	1.379
Play and Play	1:12:50	462918	2359	230	1.547	1.211	0.462	1.500
Secret Pastures	1:33:54	419080	1470	14	3.521	1.828	0.687	1.901
Story/Time: The Life of an Idea	1:13:35	402026	1470	869	2.934	2.920	1.076	1.935
We Shall Not Be Moved	1:57:30	633482	3244	4111	3.128	2.016	0.608	1.553
Analogy/Ambros: The Emigrant	1:17:15	313735	1410	236	3.352	2.226	0.311	0.731
Romeo Castellucci								
Das Floß der Medusa (Henze)	1:21:35	114133	724	549	1.433	3.159	0.274	0.671
Democracy in America	1:43:05	219148	524	46	2.469	2.906	0.496	0.715
Don Giovanni (Mozart)	3:30:01	897915	5155	1009	2.990	1.853	0.334	0.748
Go Down, Moses	1:02:40	122960	289	78	1.818	1.626	0.268	0.692
Inferno	1:36:16	345711	2837	799	3.086	1.874	0.348	0.727
Parsifal (Wagner)	3:58:29	780960	3551	847	2.452	1.966	0.198	0.503
Purgatorio	1:13:48	93387	399	392	1.163	1.062	0.254	0.719
Requiem (Mozart)	1:38:57	507498	2874	365	4.449	2.003	0.374	0.878
Resurrección (Gustav Mahler)	1:40:15	295840	2212	610	2.709	1.990	0.476	1.095
The Magic Flute (Mozart)	2:43:15	763697	4385	863	3.592	1.827	0.193	0.496

Continued on next page

Table 3 – continued from previous page

Performance	Duration	Poses	Shots	Tracks	Poses / pop. frame	Inter-pose dist / pop. frame (m)	Side-real move-ment (m/s)	3D move-ment (m/s)
Kryztof Warlikowski								
Bluebeard's Castle (Bartók) / La voix humaine (Poulenc)	2:02:19	278984	1389	736	1.667	1.563	0.243	0.662
Die Gezeichneten (Schreker)	3:48:41	954114	5454	1345	2.962	2.025	0.262	0.603
Elektra (Strauss)	2:06:25	324719	1220	532	2.03	1.599	0.226	0.564
Iphigénie en Tauride (Gluck)	2:11:26	510703	2648	822	2.708	1.674	0.207	0.526
Lady Macbeth of Mtsensk (Shostakovich) pt. 1	1:47:13	412434	3402	774	2.533	2.017	0.295	0.719
Lady Macbeth of Mtsensk (Shostakovich) pt. 2	1:17:36	446255	3619	526	3.827	1.903	0.325	0.658
Médée (Cherubini)	2:24:41	653184	4443	1029	3.090	1.919	0.258	0.607
Les Contes d'Hoffmann (Offenbach)	3:08:39	930245	6181	1432	3.346	1.600	0.250	0.627
The French	3:45:47	906390	3410	982	2.163	1.494	0.175	0.504
Wozzeck (Berg)	1:45:20	352588	2766	887	2.294	1.398	0.309	0.744

Table 3: A listing of the performance dataset, divided into sections by director. Each row gives a performance's title, duration of the recording, number of poses, shots, and tracked motions detected in the performance, average (mean) number of poses in each frame with at least one pose, mean inter-pose distance in frames with poses, mean sidereal (relative to the background) motion, and mean motion in three-dimensional space.

References

[1] Aristidou, Andreas, Cohen-Or, Daniel, Hodgins, Jessica K., Chrysanthou, Yiorgos, and Shamir, Ariel. "Deep motifs and motion signatures". In: *ACM Transactions on Graphics 37*, no. 6 (Dec. 2018). DOI: 10.1145/3272127.3275038.

[2] Aristidou, Andreas, Stavrakis, Efstathios, Papaefthimiou, Margarita, Papagiannakis, George, and Chrysanthou, Yiorgos. “Style-based motion analysis for dance composition”. In: *The Visual Computer 34*, no. 12 (Dec. 2018), pp. 1725–1737. DOI: 10.1007/s00371-017-1452-z.

[3] Broadwell, Peter and Tangherlini, Timothy R. “Comparative K-Pop Choreography Analysis through Deep-Learning Pose Estimation across a Large Video Corpus”. In: *Digital Humanities Quarterly 15*, no. 1 (2021).

[4] delahunta, Scott, Rittershaus, David, and Stancliffe, Rebecca. “Editorial”. In: *International Journal of Performance Arts and Digital Media 17*, no. 1 (2021), pp. 1–6. DOI: 10.1080/14794713.2021.1893001.

[5] Escobar Varela, Miguel and Hernández-Barraza, Luis. “Digital dance scholarship: Biomechanics and culturally situated dance analysis”. In: *Digital Scholarship in the Humanities 35*, no. 1 (Jan. 2019), pp. 160–175. DOI: 10.1093/llc/fqy083.

[6] Hou, Yumeng, Seydou, Fadel Mamar, and Kenderdine, Sarah. “Unlocking a multimodal archive of Southern Chinese martial arts through embodied cues”. In: *Journal of Documentation 80*, no. 5 (Jan. 2024), pp. 1148–1166. DOI: 10.1108/JD-01-2022-0027.

[7] Kim, Yeonho and Kim, Daijin. “Real-time dance evaluation by markerless human pose estimation”. In: *Multimedia Tools and Applications 77*, no. 23 (Dec. 2018), pp. 31199–31220. DOI: 10.1007/s11042-018-6068-4.

[8] Kutrzyński, Marcin and Król, Dariusz. “Deep learning-based human pose estimation towards artworks classification”. In: *Journal of Information and Telecommunication 8*, no. 4 (2024), pp. 470–489. DOI: 10.1080/24751839.2024.2331866.

[9] Liu, Ting, Sun, Jennifer J, Zhao, Long, Zhao, Jiaping, Yuan, Liangzhe, Wang, Yuxiao, Chen, Liang-Chieh, Schroff, Florian, and Adam, Hartwig. “View-Invariant, Occlusion-Robust Probabilistic Embedding for Human Pose”. In: *International Journal of Computer Vision 130*, no. 1 (2022), pp. 111–135. DOI: 10.1007/s11263-021-01529-w.

[10] Pavlakos, Georgios, Choutas, Vasileios, Ghorbani, Nima, Bolkart, Timo, Osman, Ahmed A. A., Tzionas, Dimitrios, and Black, Michael J. “Expressive Body Capture: 3D Hands, Face, and Body from a Single Image”. In: *Proceedings IEEE Conferrence on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 10975–10985.

[11] Potamias, Rolandos Alexandros, Zhang, Jinglei, Deng, Jiankang, and Zafeiriou, Stefanos. “WiLoR: End-to-end 3D Hand Localization and Reconstruction in-the-wild”. 2025. arXiv: 2409.12259 [cs.CV]. URL: https://arxiv.org/abs/2409.12259.

[12] Rajasegaran, Jathushan, Pavlakos, Georgios, Kanazawa, Angjoo, Feichtenhofer, Christoph, and Malik, Jitendra. “On the Benefits of 3D Pose and Tracking for Human Action Recognition”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 640–649.

[13] Rajasegaran, Jathushan, Pavlakos, Georgios, Kanazawa, Angjoo, and Malik, Jitendra. “Tracking People by Predicting 3D Appearance, Location & Pose”. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022.

[14] Schneider, Stefanie and Vollmer, Ricarda. “Poses of People in Art: A Dataset for Human Pose Estimation in Digital Art History”. In: *Journal of Compputational Cultural Heritage 17*, no. 4 (Dec. 2024). DOI: 10.1145/3696455.

[15] Serengil, Sefik and Özpınar, Alper. “A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules”. In: *Bilişim Teknolojileri Dergisi 17*, no. 2 (2024), pp. 95–107. DOI: 10.17671/gazibtd.1399077.

[16] Shawn, Ted. *Every little movement: a book about François Delsarte, the man and his philosophy, his science and applied aesthetics, the application of this science to the art of the dance, the influence of Delsarte on American dance*. Brooklyn: Dance Horizons, 1968.

[17] Souček, Tomáš and Lokoč, Jakub. "TransNet V2: An effective deep network architecture for fast shot transition detection". 2020. arXiv: 2008.04838 [cs.CV]. URL: https://arxiv.org/abs/2008.04838.