DHQ: Digital Humanities Quarterly
2021
Volume 15 Number 1

Comparative K-Pop Choreography Analysis through Deep-Learning Pose Estimation across a Large Video Corpus

Peter Broadwell <broadwell_at_stanford_dot_edu>, Stanford University
Timothy R. Tangherlini <tango_at_berkeley_dot_edu>, University of Calfornia, Berkeley

Abstract

The recent advent of deep learning-based pose detection methods that can reliably detect human body/limb positions from video frames, together with the online availability of massive digital video corpora, gives digital humanities researchers the ability to conduct "distant viewing" analyses of movement and particularly full-body choreography at much larger scales than previously feasible. These developments make possible innovative, revelatory digital cultural analytics work across many sources, from historical footage to contemporary images. They are also ideally suited to provide novel insight to the study of K-pop choreography. As a specifically non-textual modality, K-pop dance performances, particularly those of corporate and government-sponsored "idol" groups, are a key component of K-pop's core mission of projecting "soft power" into the international sphere. A related consequence of this strategy is the ready availability in online video repositories of many K-pop music videos, starting from the milieu's origins in the 1990s, including an ever-growing collection of official "dance practice" videos and fan-contributed dance cover videos and supercuts from live performances. These latter videos are a direct consequence of the online propagation of the "Korean wave" by generations of tech-savvy fans on social media platforms.

In this paper, we describe the considerations and choices made in the process of applying deep learning-based posed detection to a large corpus of K-pop music videos, and present the analytical methods we developed while focusing on a smaller subset of dance practice videos. A guiding principle for these efforts was to adopt techniques for characterizing, categorizing and comparing poses within and between videos, and for analyzing various qualities of motion as time-series data, that would be applicable to many kinds of movement choreography, rather than specific to K-pop dance. We conclude with case studies demonstrating how our methods contribute to the development of a typography of K-pop poses and sequences of poses ("moves") that can facilitate a data-driven study of the constitutive interdependence of K-pop and other cultural genres. We also show how this work advances methods for "distant" analyses of dance performances and larger corpora, considering such criteria as repetitiveness and degree of synchronization, as well as more idiosyncratic measures such as the "tightness" of a group performance.

1. Introduction

The ability to derive accurate information about human body poses and movements from arbitrary still images and videos introduces considerable new opportunities for digital humanities scholarship, especially in the realm of dance choreography analysis. Most such inquiries have previously occurred within visual media studies and among what might be considered "DH-adjacent" communities of dance and performance, with scholars using motion-capture systems to record the movements of small numbers of live dancers in controlled environments. The advent in the past few years of powerful deep learning-based models capable of accurately estimating poses directly from digital images and video footage greatly expands the scope and variety of questions researchers can pursue. We consider the particularly exciting prospect of being able to conduct studies of massive amounts of recorded choreography as another facet of the emergent practice of "distant viewing" of visual materials — a development that is itself analogous to the foundational digital humanities practice of "distant reading" of large collections of texts [Arnold and Tilton 2019] [Wevers and Smits

2019].

Deep learning-based approaches tend to be faster and more accurate than prior computer vision methods for estimating human poses in standard visual-spectrum single-camera images and videos [Elhayek et al. 2017]. The new methods also rival the accuracy of dedicated motion-capture systems and far exceed their potential scope given the physical demands of dedicated motion capture, raising the prospect of applying these deep learning methods to large recorded corpora [Blok et al. 2018]. We present the initial stages of such an inquiry, focusing on a domain that is an excellent match for the capabilities of deep learning-based pose estimation: K-pop dance choreography. 2

2. Why K-pop?

K-pop gained considerable traction in the South Korean domestic entertainment market in the aftermath of the financial downturn that rocked the South Korean economy in the late 1990s [Kim and Ryoo 2007] [Choi and Maliangkay 2014]. The pop music genre began to dominate the regular and virtual airwaves in tandem with the rise in online social networks, video and music sharing platforms, and the broader cultural phenomenon of the Korean Wave (Hallyu), a wave that gained its initial impetus with the immense popularity of Korean television dramas throughout East Asia [Lee and Nornes 2015]. It was into this well-primed social media environment that K-pop was launched, and the genre quickly evolved to include several distinguishing features, including: (i) an emphasis on individual songs (as opposed to larger "albums") promoted via music videos; (ii) the development of individual "idols" and largely single-sex musical/dance groups; (iii) a coherent musical style based largely on non-antagonistic Europop and American hip-hop styles; (iv) a heavily produced visual style that emphasized costumes, sets, dramatic lighting, and a kinetic shot vocabulary; and (v) tightly choreographed, frequently energetic, dance. 3

The global ubiquity of online video sharing and streaming services, which accelerated with the launch of YouTube in 2005, helped make K-pop an international phenomenon, with videos attracting many millions of views and solidifying fan bases throughout the world. Because of the potential for significant financial gain, the K-pop industry quickly began to attract substantial funding from the private sector and public agencies eager to promote South Korean cultural products globally [Lie 2012]. As the genre developed and the music market pivoted almost entirely away from album-based sales to video-singles and ad-based revenue, the industry began to internationalize. Consequently, it is not uncommon for producers, videographers, choreographers, music composers, lyricists, musicians, and even the idols and K-pop group members themselves to come from countries other than Korea. This internationalization has resulted in a remarkably productive collaborative environment with creative input coming from people with diverse musical, choreographic and visual backgrounds and traditions. In turn, this creative melting pot feeds a productive tension between the expectations of the broad consensus of what constitutes "K-pop" developed over the past decade by the industry, performers and their fans on the one hand, and the individualistic creative desires of the various individuals contributing to the production of new K-pop videos on the other. While aspects of K-pop production [Unger 2015], economics [Oh 2015] [Messerlin and Shin 2017], musical collaboration [Kim 2015], fandom and international reception [Han 2017] [Jang and Song 2017] [Epstein 2016] [Lee and Nornes 2015] [Otmazgin and Lyan 2014] as well as broader considerations of K-pop in the contexts of gender [Laurie 2016] [Manietta 2015] [Ota 2015] [Oh 2015], political philosophies [Kim 2017b], body aesthetics [Elfving 2018] [Oh 2015], and hybridity and cultural appropriation [Jin and Ryoo 2014] [Oh 2014a] have received considerable scholarly attention, far less attention has been paid to the kinesthetic dimensions of the genre [Saeji 2016]. In particular, a typology of K-pop dance movements and considerations of the overall choreography of K-pop have not been subjected to rigorous analysis, possibly because of the overwhelming size of the ever-growing K-pop corpus. 4

K-pop dance is marked by the integration of a broad range of popular dance styles, most notably American hip hop genres including b-boying (breakdancing), popping and locking, and other street dance styles; Indian popular dance genres such as bhangra; and borrowings from other coordinated dance traditions such as American cheer and stepping. While not all K-pop videos are dominated by dance, or can even be considered "dance forward," those that are tend to include either an individual solo dancer, or highly coordinated, often same-sex, groups featuring 4–9 dancers with break-out solo dances, occasionally set against much larger coordinated ensemble dances. Psy's satirical "Gangnam Style" music video provides an excellent sampler of the different types of dances that characterize the genre [Howard 5

2015]. Ironically, the video far exceeded the international popularity of any previous K-pop song while parodying the genre's conventions along with the superficial lifestyles of Seoul's nouveau riche.

6 In the "official" music video for a K-pop song, dances are often interspersed with narrative video scenes, and presented in a fragmentary form. Such fragmentary dance visualizations are challenging for automated analysis. Fortunately, many groups also release "dance practice" videos that present the entire dance choreography for the song, supplementing renditions of the choreography in concert and in "comeback" (new release) performances on the Korean networks' weekly live music television broadcasts. These sources allow fans to learn the dance moves and, ultimately, to record their own "dance cover" of a song. As a result, there is a considerable and growing corpus of variant forms of entire dances, shorter dance sequences, and dance moves that, taken together, represent an intriguing opportunity to explore aspects of K-pop dance reproduction, stability and variation, including considerations of borrowing from other periods, genres, styles or artists; inter-artist influence; and incremental change in dance moves and sequences.

7 K-pop dance videos provide excellent material for developing approaches to understanding and describing dance moves and sequences in a consistent manner at scale. The growing corpus offers a unique opportunity to develop pose- and movement-oriented analytical techniques and data models that in many ways parallel previous research with text corpora: producing, for example, a kinesthetic search engine that would allow dance poses, moves, or larger sequences to act as the search input, as opposed to text descriptors, and return time-stamped results from the broader corpus. Such a search engine would, in turn, facilitate the study of dance evolution, influence, borrowing, and innovation across not only K-pop but, if scaled to include other dance traditions, potentially across many different dance and movement domains, from those mentioned above to other popular Korean music genres such as trot (트로트), and even martial arts and folk dances. We limit the corpus considered in this study to K-pop music videos produced in Korea or by Korean management groups and production houses between 2004 and 2020, resulting in a full-size corpus of over 10,300 videos, primarily sourced from YouTube. Our analytical case studies draw from a subcorpus of approximately 220 official choreography demonstration/dance practice videos posted to YouTube since 2012.

3. Related/foundational computational approaches: motion capture

8 The primary contribution of deep learning-based pose estimation to the K-pop research envisioned here, and to similarly AV-oriented digital humanities agendas, derives from its speed and accuracy when detecting and estimating the poses of potentially unlimited numbers of human figures from images and video captured "in the wild" with a single visual-light camera. Earlier, non deep-learning approaches to this type of pose estimation tended to perform less well overall in the same way and for the same reasons that deep learning-based approaches to automated image analysis tasks such as semantic segmentation, captioning, and object detection and masking decisively surpassed previous computer vision approaches [Elhayek et al. 2017].

9 It is important to note that motion capture-based pose detection methods have been — and remain — capable of capturing pose and movement data at greater levels of accuracy and comprehensiveness than deep-learning approaches. For example, these methods typically record positions in three dimensions natively, rather than inferring 3-D positions (if done at all) from single-camera 2-D images, as in the case of most deep learning methods. Motion capture, though, tends to work only with a small number of dancers (often just one), imaged live in controlled conditions using dedicated hardware and software systems. Such "mocap rigs" have over the years included wearable wireless (or wired) tracking devices, setups in which one or more cameras detect reference markers worn on the body and the face, and "markerless" systems that combine a visual-light camera with infrared laser ranging sensors to build a 3-D map of the objects in front of them. The second of these technologies drove the proliferation of performance capture-based characters in popular films of the early 2000s, and the first and third were the signature innovations of the Nintendo Wii and the Microsoft Kinect interactive gaming systems, introduced in 2006 and 2010 respectively [Reilly 2013] [Sutil 2015]. Despite the broad adoption of such methods and their potential benefits, traditional dance motion capture assumes that one has access not only to the equipment, but also to dancers capable of reproducing the desired moves, sequences and complete choreography.

10 The partial equivalence of deep learning pose estimation output to motion-capture data means that researchers using

deep learning-based techniques can derive inspiration and potentially even analytical techniques from prior motion capture-based dance studies. A rich lineage of motion-capture inquiries exists, with many studies exploring how to record, analyze, and communicate the nuances of full-body choreography — as opposed to previous systems for notating foot movements — that prompted Rudolf Laban and his colleagues to develop Labanotation in the 1920s, to be followed by Benesh Movement Notation in the 1940s as well as several other schemes [Watts 2015]. Yet technology-assisted dance analysis efforts to date have been typically somewhat narrow in scope, often focusing on just a single dancer, and primarily intended to contribute to dance pedagogy [Rizzo et al. 2018], close analysis of the micro-scale nuances of a specific dance or movement [Dong et al. 2017], or oriented towards dance entertainment systems and video games.

11 One noteworthy motion-capture dance study, highly relevant to the present discussion, was a large-scale, multi-year project at the Korean Electronics and Telecommunication Research Institute (ETRI), which also focused on K-pop dance [Kim 2017a]. From 2014 to 2017, the ETRI researchers used a Kinect-type system to generate motion-capture recordings of professional dancers re-enacting choreography from a large set of popular K-pop numbers. They subsequently developed techniques for characterizing and comparing the recorded poses, eventually assembling a large database of K-pop dance poses. Potential uses of the comparison techniques and database as envisioned in the project documentation included helping to adjudicate choreography copyright disputes and serving as source data for a mocap-based K-pop "dance karaoke" platform or a similarly featured K-pop dance pedagogy system [Kim 2017a].[1] The project's methods for pose characterization are broadly similar to the distance matrix-based approach used in the present study [Kim et al. 2018]. The ETRI researchers' technique for comparing short dance moves uses an extended version of a previous study's metric based on "dynamic time-warping" analysis [Raptis et al. 2011]. This method, which involves comparing the frames of a given motion to a labeled "reference" version of the motion, is not applicable to the present study, which endeavors to extract descriptors of K-pop poses and gestures in a largely unsupervised manner (i.e., directly from video sources). Nevertheless, it may prove relevant to future expansions of this work.

4. Getting a leg up with deep learning

12 The deep learning methodologies underpinning recent advances in pose estimation are fundamentally the same as those that drove the earlier breakthroughs in object detection and segmentation for still images. In brief: large sets of potentially meaningful image features, often derived by applying certain filters, overlays, and "convolutions" to the images, are fed to an interlinked system of data structures ("neurons"), which repeatedly applies fairly straightforward mathematical calculations to "learn" which features are the most effective at helping the entire network discern between different labels for the input images, e.g., "dog," "cat," "arm," "leg." Large quantities of such labeled data are needed to train the models. Although this input can be derived from existing digital resources, massive amounts of novel manual efforts, often obtained through low-wage or entirely uncompensated labor, are needed to train state-of-the-art models, (e.g., Google's reCAPTCHA service) [Hara et al. 2017]. The resulting models tend to perform quite well at tasks related to isolating, identifying and describing the objects they have been trained to detect. For pose estimation, the first of these tasks often consists of deciding which "regions of interest" on an image (established by dividing the input image into a grid of regions of varying sizes and dimensions and then applying an "ROI" evaluation sub-model to each) seems likely to contain a human figure. After finding these ROIs, the pose estimation model is employed to identify and localize discrete portions of the detected figure.

13 Many pose estimation models operate by collapsing the detected probabilistic "field" for, say, an arm into discrete keypoints (wrist, elbow, shoulder) until eventually a full pose "skeleton" of keypoints and connecting linkages is obtained. This approach was used for the first major open-source deep learning-based pose estimation project, OpenPose [Cao et al. 2018]. Deep learning-based face detection methods also follow a similar process, and it is not difficult to see how the pose comparison approaches described below, predicated as they are on the notion of a pose "fingerprint," resemble some facial recognition/matching techniques. It is also not hard to envision the mass surveillance applications to which both methods lend themselves [Kumar et al. 2020] [Byeon et al. 2016]. Importantly, deep learning-based pose detection has grown to encompass a much wider variety of use cases beyond its most obvious applications in security and retail surveillance and driver-assist technology. Examples include the development of "in-bed pose

estimation" for monitoring of hospital patients using low-cost cameras [Liu et al. 2019], as well as the DeepLabCut software, which facilitates the training of non-human pose estimation models to aid observational studies of a virtually limitless range of animals [Mathis et al. 2018].

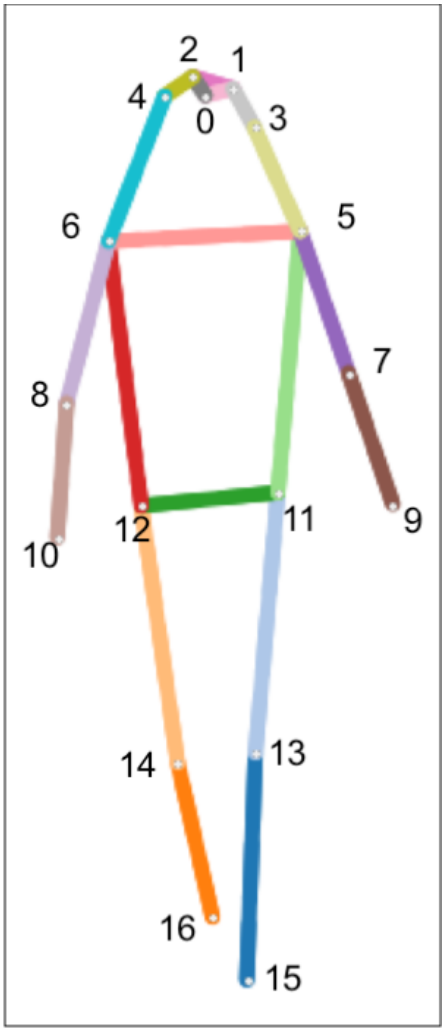

Figure 1. The COCO “Common Objects in Context” pose keypoint set. The numbering of the keypoints may vary between software implementations. The numbers here apply to the other figures in this paper.

When providing output coordinates for detected figures, human pose estimation models usually adhere to a standard set of keypoints, such as the 17 keypoints of the COCO (Common Objects in Context) library [Lin et al. 2015]. This particular set truncates the figure's arms at the wrists and the feet at the ankles, which is not ideal for choreographic analysis nor for many other potential applications (Figure 1). Accordingly, developers have developed expanded body keypoint standards (such as BODY_25) or, in the case of OpenPose, simply superimposed other models for detecting hand and face landmarks [Cao et al. 2018]. The resulting figures can have as many as 130 keypoints, though one motivation for using smaller keypoint sets is that detecting and subsequently comparing more keypoints usually requires more processing time, storage and power. The "model zoo" provided by the developers of a given pose estimation package likewise typically includes a range of tuned models, each prioritizing or deprioritizing speed, accuracy, the number of output keypoints, model size, and resource requirements based upon its expected use. For example, a faster but lower-resolution model might be deployed in a smartphone application offering real-time body tracking. A slower, "heavier," higher-resolution model could be the right fit for a well-resourced digital humanities research team seeking to resolve fine details of inter-frame movement within a set of pre-recorded dance videos, especially if the team has ample time to run the pose estimation software on the videos.

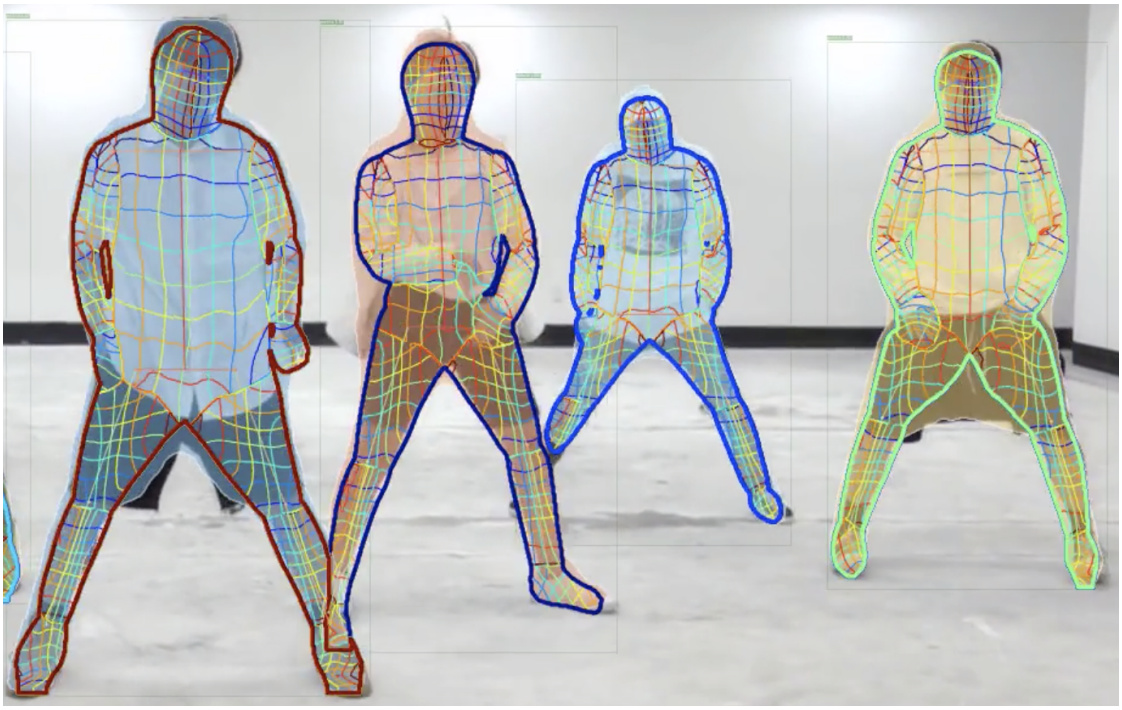

Figure 2. A visualization of the full DensePose output: body masks (including hair and clothing), segmented body parts with contours, and keypoint skeletons.

Other pose estimation models actually project the detected probabilistic body part fields onto a 3-D model of the body, so that instead of a simplified skeleton, the output consists of a much larger set of points, regions and connectors that define a full 3-D body surface, similar to how a fishing net thrown over a person would form a body-shaped "mesh" (Figure 2). This mesh output option is available from Facebook Research's DensePose project [Güler et al. 2018]; it gives researchers access to a much less reductionist model of the body (including full hands and feet) which is certainly appealing in some applications. One practical concern is that the resulting per-frame body meshes for a single video can use an enormous amount of storage space if not properly compressed. 15

Another practical consideration is that even the most heavyweight, non time-constrained pose estimation models may not perform well with imagery that is visually distorted, shot from oblique angles, poorly lit, or involves figures whose features are obscured by costumes and walls, or are simply truncated by the frame. That the previous list resembles a primer on music-video cinematography should suggest one reason why most of the present study's initial investigations used supplemental "dance practice" videos or fan-produced "cover" dance videos rather than the original music videos. At the root of these problems is the very limited ability of most deep learning models to extrapolate beyond their training data, so if a pose-estimation model is primarily trained via labeled, segmented images of people in brightly lit environs with visible faces performing mundane activities like walking or standing, its ability to resolve, for example, figures wearing masks or swinging their arms vigorously above their heads is likely to be quite limited. 16

Most pose estimation projects at present aspire to excel at figure detection and pose estimation during the first "pass" across an image, which has obvious relevance to core time- and resource-constrained applications such as real-time pedestrian detection systems for automobiles. This emphasis on first pass methods means that, in general, alternative methods involving multi-pass processing and smoothing tend to receive less attention, despite their potential to improve model performance in ways that would be especially beneficial to digital humanities researchers working with recorded media. The PoseFix package, for example, managed to achieve state-of-the-art accuracy simply by applying its statistical "pose refinement" model to the output of other methods [Moon et al. 2018]. Such corrective calculations can be as straightforward as setting limits on how distant a figure's head can possibly be from the shoulders in a non-catastrophic scenario. Similarly promising but heretofore backgrounded efforts involve expanding models to incorporate the causal implications of a figure's previous position (and its future position, if known) to its current one [Pavllo et al. 2019]. Developers acknowledge the importance of tracking multiple poses across frames — especially so when figures 17

may pass each other in the same shot — but this is often, and perhaps erroneously, relegated to a "post-processing" step, something to be considered after the actual pose estimation has been done [Andriluka et al. 2018].

18 The accuracy of pose estimation models continues to improve as more varied training sets and clever algorithms are developed. Recent advances in the speed and accuracy of three-dimensional object detection from single-camera sources promise to offer researchers an even greater wealth of information about figures recorded on video and their relationship to their environment [Ahmadyan and Hou 2020]. Furthermore, software platforms and cloud services continue to emerge that make it much easier to provision and configure the software "stack" and computing resources necessary to run these models (for example, by providing access to cloud-based graphical processing units, or GPUs, which greatly accelerate deep learning tasks). As a consequence of this rapid pace of development, the decision whether to run pose estimation on raw music videos or on their accompanying dance demonstration videos is already more a question of focusing exclusively on dance choreography versus also examining computationally the many other types of performative uses of the human pose that appear in music videos.

5. Getting down to the features: analytical methods

19 This section describes the fundamental approaches to analyzing deep learning pose estimation output that we have applied to data from K-pop dance videos in the early phases of the research agenda outlined above, and also outlines some of the more elaborate techniques we may pursue in future work. These methods primarily concern pose characterization and comparison and the exploratory and interpretive affordances they offer when applied to a large number of videos. We also outline potential approaches to pose clustering and time-series analysis of movement and synchronization.

Pose representation, comparison and correction

20 The raw output of deep learning-based posed estimation software is not fundamentally different from motion-capture data, so many of our techniques may have appeared in prior mocap-based analyses. Because of the reliance of these earlier studies on closed-source systems and the lack of technical details in their associated publications, such implementation-level aspects are difficult to ascertain fully. In any case, our methods by no means encompass the available techniques. Yet seeking comprehensiveness would undercut the central message of this paper: that the ability to run deep learning-based pose detection at scale across large video corpora empowers researchers to pose new questions and to develop methods for addressing them that probably have never been used before — at the very least, not in the domain of choreography analysis. One need only consider the history of computational text studies as an analogue: methods of linguistic examination and structural analysis certainly existed in the pre-digital era, but the advent of digital texts, and particularly the availability of massive quantities of digital texts, prompted an explosion of computational methods, especially at the level of large-scale, "distant" reading: topic models (LDA), semantic embeddings, named entity detection, network analysis to name but a few. Choreographic analysis has the potential to follow a similar trajectory.

21 Despite the relatively porous boundaries of K-pop vis-a-vis other forms of Korea-based popular music and the paucity of meaningful descriptive metadata for YouTube videos, in previous work we showed that it is possible to discover thousands of official K-pop music videos on YouTube by querying the videos uploaded to channels run by K-pop production companies identified via online knowledge bases (Wikidata, MusicBrainz) [Broadwell et al. 2016]. For reasons discussed above, we supplemented this list with a smaller number of official dance practice, demonstration and "dance cover" videos to facilitate development and evaluation of our choreographic pose analysis techniques.

22 We obtained the pose estimation data used in the case studies below by processing videos with software from the Open PifPaf project [Kreiss et al. 2019], which we used due to its accuracy and relative ease of setup. The 17-keypoint COCO pose output data has modest storage requirements — an average of 7 megabytes of uncompressed data for a 4-minute video — and is fairly straightforward to process as CSV or JSON (Figure 3). We also ran pose estimation on the entire video corpus using a DensePose model, a process that took several weeks on a dedicated multiple-GPU system.[2] The full DensePose "mesh" output, which is appealing because it allows for the calculation of additional features such as

hands and feet positions and even clothing, can require 30 gigabytes or more of storage per 4-minute video (highly reactive to the number of figures in most shots), and often is neither straightforward to compress nor to interpret.

frame_timecode	frame_id	figure_index	nose_x	nose_y	left_eye_x	left_eye_y	right_eye_x	right_eye_y
0.48	12	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.52	13	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.56	14	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.6	15	1	319.29153	138.8482	322.26694	134.86655	315.3243	133.87114
0.64	16	1	318.25806	136.51443	321.20477	132.52725	314.32913	132.52725
0.68	17	1	318.20584	137.45535	321.1731	133.47325	314.2495	133.47325
0.72	18	1	318.25464	138.43123	322.20602	134.4362	315.29108	134.4362
0.76	19	1	319.12305	137.1793	322.0857	134.19124	315.17285	133.1952
0.8	20	1	318.90002	136.19379	322.87646	133.20044	314.92358	133.20044
0.84	21	1	319.1521	137.93892	323.13556	133.94016	315.16864	133.94016


```
,{"0.48": {"boxes": [], "keypoints": [], "frameID": "12"}}
,{"0.52": {"boxes": [], "keypoints": [], "frameID": "13"}}
,{"0.56": {"boxes": [], "keypoints": [], "frameID": "14"}}
,{"0.6": {"boxes": [["291.02515", "117.44682", "343.5907", "319.51562"]], "keypoints": [{"right_hip": ["309.3735", "218.48123"], "right_wrist": ["297.47186", "218.48123"], "l
,{"0.64": {"boxes": [["292.22882", "121.06415", "346.25177", "330.39053"]], "keypoints": [{"right_hip": ["308.4357", "215.26102"], "right_wrist": ["310.40018", "177.38292"],
,{"0.68": {"boxes": [["289.0279", "115.05602", "345.40564", "320.13428"]], "keypoints": [{"right_hip": ["308.315", "215.10634"], "right_wrist": ["314.2495", "155.3748"], "lef
,{"0.72": {"boxes": [["290.101", "114.960495", "343.4447", "325.6977"]], "keypoints": [{"right_hip": ["308.37616", "220.3291"], "right_wrist": ["293.55847", "217.33282"], "le
,{"0.76": {"boxes": [["289.00278", "115.76479", "346.28067", "322.93777"]], "keypoints": [{"right_hip": ["307.27246", "218.85327"], "right_wrist": ["293.44675", "220.84532"],
,{"0.8": {"boxes": [["286.59143", "116.73698", "344.24982", "329.26526"]], "keypoints": [{"right_hip": ["307.96478", "217.0144"], "right_wrist": ["293.05316", "219.00998"], "
,{"0.84": {"boxes": [["289.77405", "115.44587", "344.5467", "335.37784"]], "keypoints": [{"right_hip": ["308.19757", "216.91449"], "right_wrist": ["290.27197", "220.91325"],
,{"0.88": {"boxes": [["291.7315", "115.63084", "346.9701", "333.56155"]], "keypoints": [{"right_hip": ["308.00717", "217.09859"], "right_wrist": ["313.92557", "155.11829"], "
,{"0.92": {"boxes": [["290.49365", "119.504036", "347.3943", "328.68646"]], "keypoints": [{"right_hip": ["308.9614", "216.62445"], "right_wrist": ["325.93176", "200.68674"],
,{"0.96": {"boxes": [["292.43842", "120.14389", "347.15155", "331.0218"]], "keypoints": [{"right_hip": ["308.85236", "218.58688"], "right_wrist": ["295.92017", "220.58572"],
,{"1.0": {"boxes": [["291.35144", "119.84539", "346.75262", "328.7471"]], "keypoints": [{"right_hip": ["309.65363", "216.3"], "right_wrist": ["296.79263", "220.29813"], "left
```


Figure 3. Keypoint pose estimation output as tabular data (top) and JSON (bottom). Mesh data consists of nested arrays and is not immediately comprehensible outside of visualizations.

23

At the fundamental level, an estimated pose is defined by its labeled keypoints and the extrapolated linkages between them, expressed as x,y coordinates on the visual plane of the screen (with a third spatial coordinate, z, if depth is also estimated). One obvious method of quantifying the difference between, for example, two 17-keypoint COCO poses is simply to sum the distance between the two instances of each keypoint using a metric such as Euclidean distance. This metric also can be a proxy for the amount of motion between two poses if they derive from the same figure at adjacent time intervals. These methods may suffice in a controlled, single-dancer motion capture studio environment, but generalizing pose characterization and comparison to any "in the wild" video footage requires more sophisticated approaches. Simply summing raw paired keypoint distances can be an incredibly inaccurate measure when, for example, we wish to compare poses in two different contexts, such as when the figures being compared are viewed at varying distances from the camera, or belong to different-sized people.

24

Solutions typically require devising a different "feature set" to describe the pose numerically, which in turn calls for different inter-pose comparison techniques. One rudimentary approach is to consider only the angle of certain linkages, such as in the arms or legs, because their angles are not affected by changes in scale. This approach, however, discards all potentially useful information about the positions of unconnected keypoints. Another, more promising class of solutions involves representing the pose not as a set of keypoints and the skeletal linkages between them, but rather as a "distance matrix" of the distances from each keypoint to every other keypoint (Figure 4). Comparing two such pose representations to quantify similarity or movement can then be accomplished by applying statistical tests designed to measure the degree of correlation between two matrices — a process that ignores differences in scale between the two poses. For our initial studies, we used the Mantel test, which provides a measure both of the strength of the correlation between the input matrices (this correlation can be expressed as a number between 0 and 1) and the computed probability that this correlation is due to random fluctuations in the data [Giraldo et al. 2018].

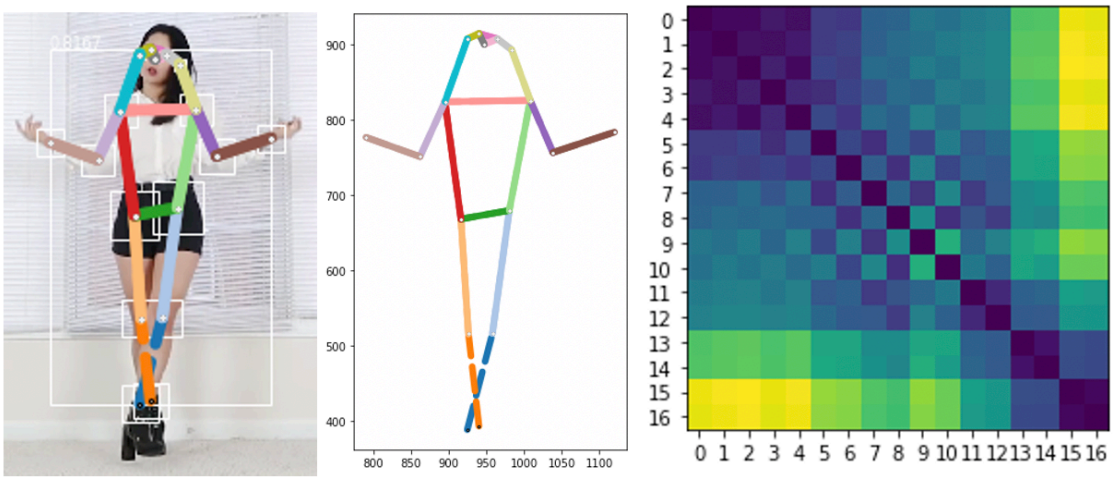

Figure 4. A source image with keypoint overlay, a plot of the detected keypoints plotted separately (center), and the corresponding normalized inter-keypoint distance matrix (right). In the distance matrix, pairs of points that are close together are represented by dark squares, while those that are far apart receive light squares.

Another class of enhanced pose representation and comparison methods considers only whether keypoints are closest to each other. In the simplest form of such an "adjacency" matrix, the cell for [right elbow, right ear] would record a 1 if the right elbow is closer to the right ear than to any other keypoint, and a 0 otherwise. This approach obviously disregards much of the estimated pose data and sacrifices accuracy as a consequence, but retains the overall spatial organization of the pose and has the advantage of being quite fast, requiring few calculations to characterize a pose and to compare two poses to each other. Our implementation of this method expands it further by calculating the Delaunay triangulation around the detected keypoints [Delaunay 1934]. This technique provides an alternative set of keypoint linkages to the standard body skeleton model, one in which every keypoint is connected to at least three others, producing a set of triangles that covers the shape of the pose in a geometrically efficient manner (Figure 5). We then represent these connections via a graph Laplacian matrix, which quantifies both the adjacency and degree of each point [Chung 1997]. Although the distance matrix and graph Laplacian matrix for a pose have the same dimensions (17x17 for the COCO keypoints), because a graph Laplacian contains only positive or negative integers, comparing two poses represented in this manner involves simply subtracting one from the other and summing their differences — a very computationally "lightweight" operation compared to the Mantel test for the distance matrices. It is also worth noting, however, that a standard Delaunay triangulation discards the right/left labels of the keypoints, meaning that frames of a pose with the figure facing the camera would be scored as identical to the mirror-image of the same pose with the figure turned 180 degrees away from the camera — which may or may not be desirable.

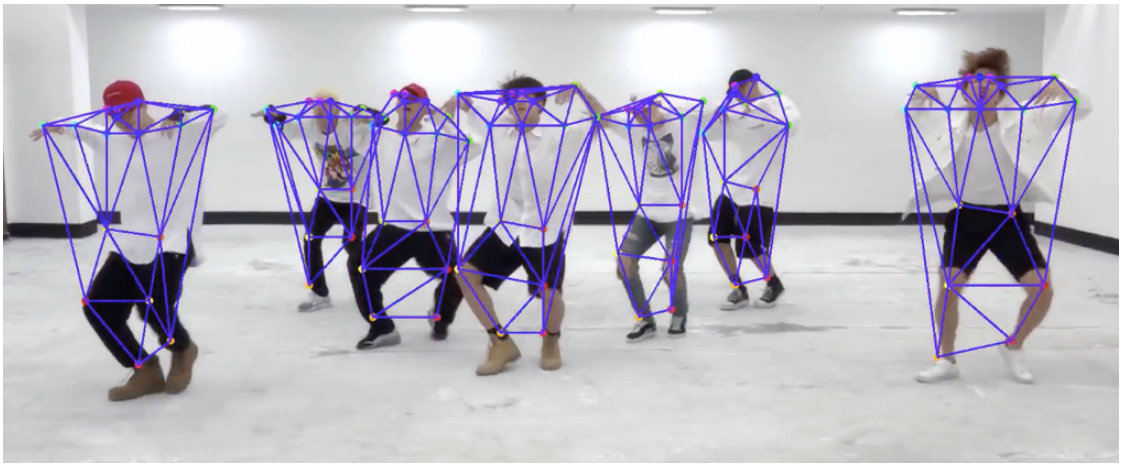

Figure 5. Delaunay triangulations of detected figure keypoints.

Pose comparison methods, including those described above, must accommodate the near-certainty of incomplete pose and keypoint data. Even sophisticated pose detection software can lose track of body landmarks and sometimes entire figures for multiple frames. Often, the software detects part of the pose, but its "confidence" value for a keypoint or the entire figure drops low enough that the data points are removed from the output to avoid spurious results. Especially problematic with single-camera "in the wild" videos are cases in which even a human observer could only speculate as to the true coordinates of a keypoint, such as when a limb or facial landmark is obscured from view. Most pose comparison methods, including the matrix-based methods used here, do not easily accommodate missing data values. Our software therefore falls back on both spatial and temporal interpolation to fill in missing keypoints. The relatively high frame rates of modern video assists in the latter: the position of a missing keypoint often can be placed somewhere between its last and next known position. Failing that, spatial interpolation often allows us to place with high confidence, for example, an eye that is obscured by a hat brim between its adjacent ear and nose. A last-resort option is to position a missing keypoint at the center of the pose. Both the distance matrix technique and the Delaunay triangulation-based graph Laplacian approach, due to their addition of extra linkages to the base keypoint set, are still able to produce usable results when undefined values are replaced with such a default. 26

A final obstacle is that most pose detection packages do not attempt to track the figures in a scene, simply numbering the figures in a shot via an arbitrary ordering, e.g., left-to-right, or sorted by the size of their bounding box, regardless of identity. This practice leads to discrepancies in the movement data when, for example, one figure changes places with another. To ameliorate these errors, we also run our corrected pose detection output through the AlphaPose project's "PoseFlow" software, which attempts to generate a single movement track for each distinct figure throughout the duration of the video [Xiu et al. 2018]. Even when PoseFlow fails to "keep track" of a dancer, our current choreography analysis methods benefit from considering each figure only within the frame of reference of its own body landmarks, excluding lateral and backwards/forwards movement relative to the camera and other dancers, and from the fact that we generally examine the simultaneous movements of multiple dancers in aggregate rather than individually. Therefore the failures of the pose tracking software to resolve correctly the trajectories of two overlapping poses generally do not introduce significant discrepancies into our analytical results. 27

Pose clustering, sequence detection, and "distant" movement characterization

Having chosen a set of methods that can represent a pose and quantify the similarity and difference between two poses, such that the degree of difference between two poses from the same dancer can serve as a reasonable proxy for motion, it is then possible to pursue a variety of more aggregate, "distant" analyses that consider groupings and sequences of poses over time. The case studies below describe the techniques that we find particularly promising and relevant for the study of K-pop choreography and its influences, but they are by no means an exhaustive set. Ultimately, we believe that a macroscopic approach that includes a series of these analytical techniques will support researchers as they move toward the "thick" analysis of dance at scale [Geertz 1973] [Börner 2011]. 28

The computational considerations germane to these analyses involve at their fundamental levels many standard details of statistical and numerical computation. These include sampling (e.g., whether to consider every pose in every frame), or whether a subset of poses (chosen randomly or by weeding out repetitions) can suffice to produce significant results at much lower computational cost in time and resources. Of further concern are methods for smoothing and interpolation, already discussed above regarding pose correction: how best to reduce the influence of transient data errors on the results without disregarding legitimate phenomena, and to "fill in" missing values with a suitable degree of confidence. At the higher levels are issues such as the choice of algorithm for clustering or time-series comparison and their various hyperparameters (e.g., how many clusters we expect to find in the data set). 29

The example analyses below illuminate how the foundational pose characterization and comparison methods described in the previous section build progressively towards one of our primary long-term goals: isolating a typography not only of K-pop poses but of gestures and ultimately dance "moves." Although this work is ongoing, the examples below indicate the planned trajectory: using pose similarity and clustering methods to identify "key" poses that occur frequently with minor variations; the sequences of key poses and interstitial movements that recur frequently across the time series of one or multiple videos thus identify themselves as significant "moves" within the dance vocabulary of K-pop. 30

6. Case studies with K-pop

31 The following case studies highlight some of the potential applications of the pose characterization and comparison techniques described above. Specifically, the first of these examples involves applying pose similarity calculations in a variety of ways to a single-dancer video performance: comparing time-adjacent poses to establish the movement profile of the dance, and comparing poses across the duration of the dance either exhaustively or selectively (via clustering) to highlight repeated poses and pose sequences. The second example applies many of the same techniques to a video of multi-person choreography, with the addition of inter-dancer comparison of poses and motion to detect and quantify the degree of synchronized posing and movement present across the video. The third study calculates per-video and aggregate values of each previously discussed metric for a test corpus of 20 K-pop choreography videos, divided into equal halves according to the gender of the performing group.

Solo dance: repeated pose sequence discovery

32 For a single-person video, our goal was to use the pose comparison methods described above to detect when a dance video is repeating certain poses and motions. Being able to find repeated poses illustrates the suitability of these techniques to building a kinesthetic database of poses that can be searched via a "query" pose to find similar poses. Given this, detecting repeated motions can be as straightforward as noticing when a "query" sequence of poses matches a reference sequence of poses within some similarity threshold and across some time window. As this example will show, groups of repeated motions tend to correspond to repeated formal sections and suggest the potential of performing automated computation-based formal analyses across a large dance video corpus.

33 For this case study, we used an instructional recording by dance cover specialist Lisa Rhee of Blackpink member Jennie Kim's debut single "Solo," which was choreographed by the prolific New Zealand-based choreographer Kiel Tutin.[3] One way of testing how computational pose comparisons can contribute to detecting dance repetition is via the brute-force expedient of comparing the pose in every frame of a single-dancer video to every other frame, and plotting the results on a correlation heatmap matrix (a type of visualization commonly used to explore patterns of repetition in time-series data). Such heatmaps tend to be difficult to read at first, but provide the initiated with a wealth of visual cues about patterns of correlation and similarity; if desired, these features also can be described numerically by applying established analytical techniques to the correlation matrix.

34 Figure 6 shows a correlation heatmap for the entire performance, which lasts just over 2 minutes and 43 seconds, captured in 4,079 frames (25 frames per second). The poses were represented as normalized distance matrices as explained above, and therefore the comparison between any two poses returns a Mantel correlation value from 0 (least similar) to 1 (most similar). On the heatmap, these similarity values are visualized via the color scale, with lighter colors indicating lower similarity. The x and y axes of the heatmap represent the progression of frames of the video, with the start time 0:00 at bottom left, so that the cell at position x, y represents a comparison of the pose at frame x to the pose at frame y. The matrix is therefore symmetric around the diagonal x=y axis, which naturally always has a similarity value of 1 and appears as a dark diagonal line dividing the heatmap into two right triangles.

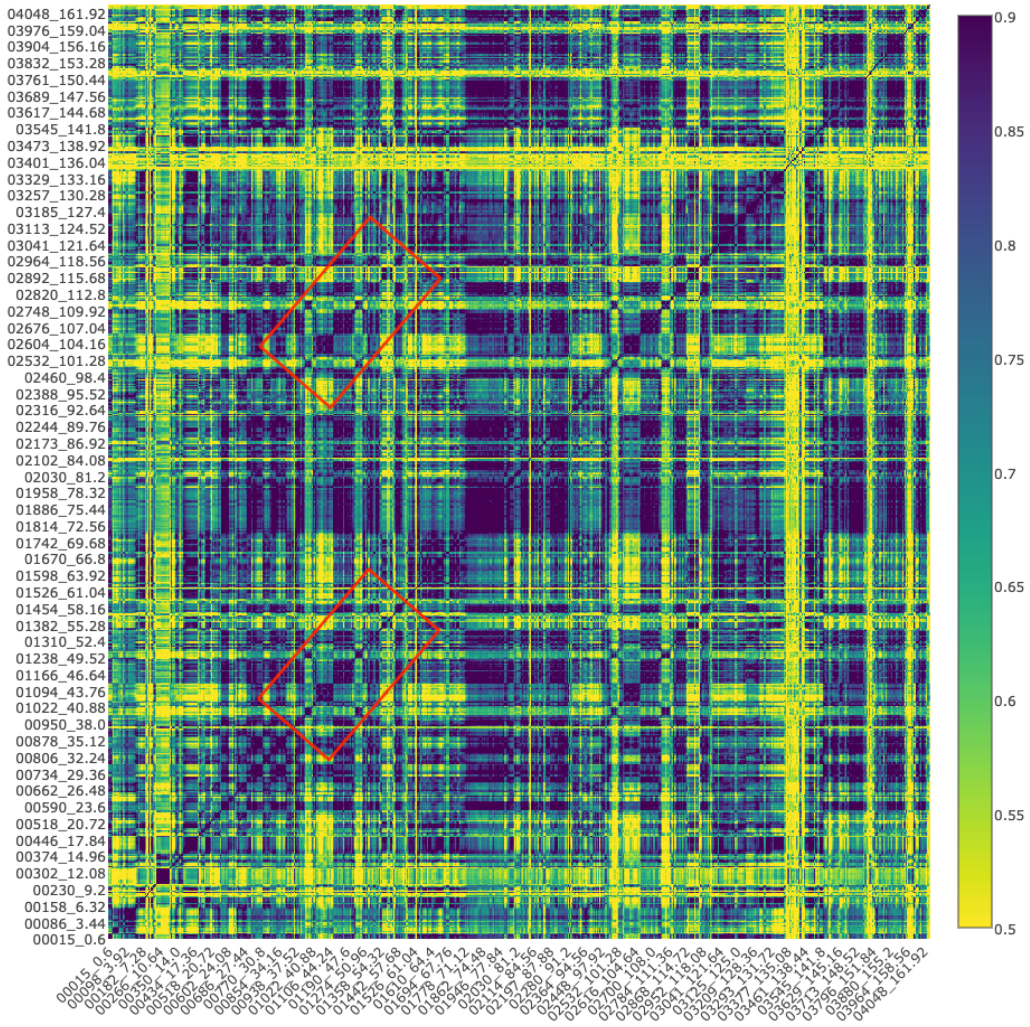

Figure 6. Time-series correlation heatmap of Lisa Rhee's dance cover of Jennie's "Solo." Darker colors indicate higher degrees of pose similarity. The repeating dance "chorus" sections at 0:38–0:60 and 1:38–2:00 are highlighted.

On a time-series correlation heatmap, repeated sections typically appear as dark lines running parallel to the central diagonal x=y axis, and such a feature is indeed visible here (highlighted in Figure 6). Closer inspection reveals that it does in fact indicate a very close repetition of the choreography originally seen at 0:38 to 0:60 almost exactly a minute later, at 1:38 to 2:00. Not surprisingly, the repeated choreography corresponds exactly to the appearances of the song's chorus, hinting at the likely interpretive payoffs of a truly multimodal audiovisual analysis. 35

Exhaustively comparing every pose to every other pose across an entire video, let alone multiple videos, quickly becomes a prohibitively cumbersome computational task. A more scalable method is to apply similarity-based clustering to all or to a representative sample of the poses in one or more videos, and to identify when poses in the same similarity cluster, or "family," reoccur, and moreover when members of clusters reoccur in sequence — a major step towards identifying both formal sections within a choreographic plan, and also smaller, segmentable dance sections (i.e., moves). 36

Clustering necessarily sacrifices some precision, and requires judicious selection of parameters to produce a useful partitioning of the entire pose space. For this example, we applied the OPTICS hierarchical density-based clustering algorithm [Ankerst et al. 1999]. Many other types of clustering algorithms may be applied fruitfully to this task, but we found OPTICS suited to the exploratory nature of our analysis because it does not require one to specify how many clusters are to be found in advance, but rather builds clusters based on a user-supplied minimum number of members per cluster. This hyperparameter can be set to a value with some intuitive justification; we chose to use the number of frames per second in the video recording (approximately 25, in this case), reasoning that a pose that is held for longer 37

than one second, or that appears cumulatively for at least this length of time, may be significant and should be considered for cluster membership by the algorithm.

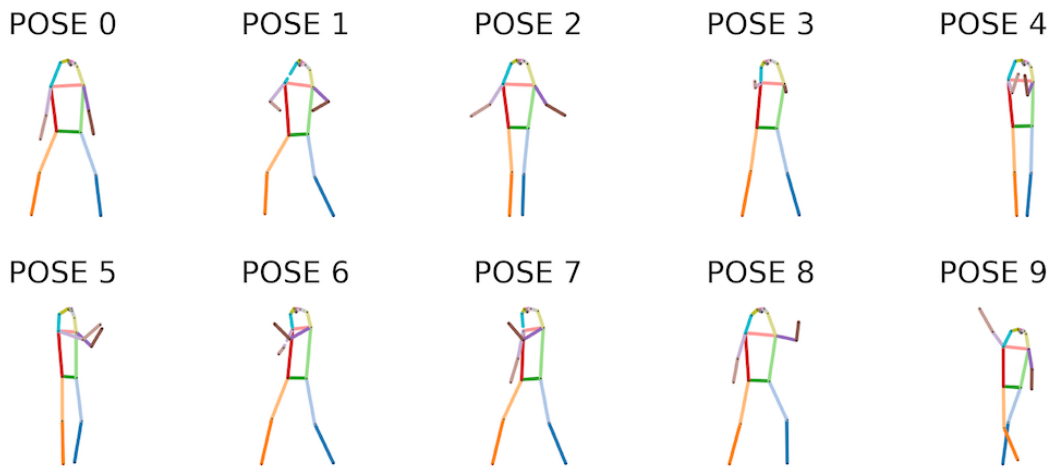

Figure 7. The representative “key" poses of the 10 pose clusters found by the OPTICS algorithm when it is configured to find a minimum of 25 poses per cluster. Each key pose was constructed by taking the average keypoint locations of all members of a cluster.

The clustering analysis for this video found ten groups of poses using the settings discussed above. To aid in visualization and comparison, we computed a representative "centroid" pose for each cluster by averaging the relative keypoint positions of each pose in the cluster (Figure 7). The OPTICS algorithm leaves a potentially large number of samples unassigned to any cluster, so we elected to assign each of these to the cluster with a representative centroid pose that was most similar to the unassigned pose. Visualizing the occurrences of these cluster groups on a timeline (Figure 8) makes it possible to detect repeated segments of choreography. The chorus sections identified in the pose correlation heatmap above, as well as the bridge section leading to them, are easily discernible. Key pose #9, with the upraised right arm, is particularly recognizable as a signature pose of the song, even though the COCO 17-keypoint set is unable to resolve the raised index finger that evokes the song's title.

38

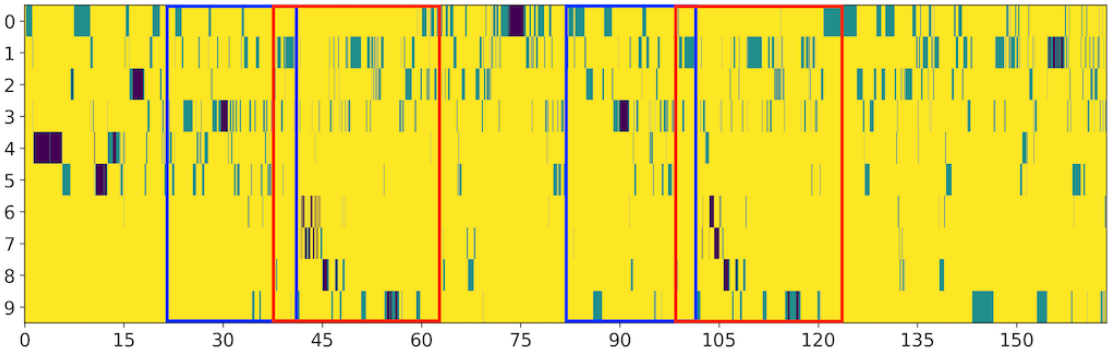

Figure 8. The pose distribution heatmap of the key poses from Figure 7 throughout the duration of the song, with the two appearances of the chorus outlined in red and the two “bridge" sections outlined in blue. Both section types repeat their previous choreography when they reoccur, which is apparent in the similar patterns of their sections on the heatmap. Occurrences of poses that the OPTICS clustering algorithm explicitly assigned to one of the 10 groupings are represented by blocks with a darker shade, while the unmatched poses that we subsequently assigned to their “nearest neighbor" key pose are in a lighter shade.

Thinking more broadly while looking more closely, it might also be appealing to be able to search through a much larger corpus of K-pop videos for occurrences of pose #9, as well as others from the directly mimetic gestures that populate the first 15 seconds of the choreography. Several of these are well enshrined in K-pop iconography such as, for example, the two-handed heart-shaped pose at 0:04, which is quickly and dramatically split in two at the six-second mark — an obvious nonverbal declaration that this is a breakup song — while others are relative newcomers, such as

39

the sharp dismissal of items (messages, in this case) from a smartphone screen at 0:10 to 0:12.

Figure 9 visualizes the analytical methods described above via accompanying graphics as well as overlays of an excerpt of the video, with displays of summary values and a progress indicator (the moving red vertical line) superimposed on the time-series components. Viewing the changes in the distance matrix visualized this way alongside the actual poses imparts a more intuitive understanding of how it is used to generate similarity and movement values. The general resemblance of the distance matrix-based movement and the graph Laplacian-based movement series is apparent here, as is the graph Laplacian series' comparatively “noisier” representation of the amount of inter-frame movement.

Comparative K-Pop Choreography Analysis: Video 1

from **Peter M. Broadwell**

02:43

Figure 9. Deep learning-based choreography analysis of an instructional recording by dance cover specialist Lisa Rhee of Blackpink member Jennie's debut single “Solo.”

Group dance: synchronized movement detection

Multi-person performances are by far the most common type in K-pop, just as multiple-member idol groups are much more numerous than solo performers. Even solo songs are likely to incorporate backup dancers in live performance as well as in the official music video. Reflecting the eclectic combination of genres that are incorporated into K-pop, the choreography for an idol group single may feature several different types of dance sections, including solo interludes or *pas de deux* by the group's primary dancer(s) as well as gestural punctuations during a showcase of the group's main rapper. Yet the signature so-called "point" dance segments, which are most often staged with the group performing the

same moves while facing towards the camera, tend to receive the most attention, to the extent that mechanistic mass synchronization is arguably the most salient feature of K-pop choreography in the global public imagination. This is largely by design; the didactic nature of these sections' staging, in addition to serving other functions within a music video's narrative, furthers the choreography's main contribution to the carefully crafted viral appeal of K-pop singles, which is that fans and sometimes the general public in Korea and occasionally even the global population (as in the case of Psy's "Gangnam Style") derive social capital from knowing and being able to re-enact these moves.

Group synchronization is straightforward to detect and to quantify using the pose comparison methods described above, through the simple expedient of observing the computed difference between each pair of figures in a single frame (note that the number of comparisons required per frame is the answer to the well-known "handshake" problem: n*(n - 1)/2 for n figures), then examining the mean and standard deviation of these per-frame values over time. We would expect that the mean degree of similarity would increase while the standard deviation would decrease during episodes of synchronization, which is exactly what we see when this technique is applied to the official "dance practice" video for the boy group BTS's single "Fire," choreographed (as are several of BTS's other hits) by the American choreographer Keone Madrid.[4] 42

As visualized in Figure 10, the mean intra-frame pose similarity increases markedly and the standard deviation ranges shrink around 0:10 as the frenetic, popping and locking-influenced group dance begins and the song launches into the introductory hook section, with its accompanying EDM "hoover" synth effects. Note that this is also the first dance choreography that appears in the official music video for "Fire," following its introductory narrative imagery, reinforcing the importance of synchronized group dance sections to the multimedia appeal of K-pop singles. Pose synchronization levels remain high for much of the choreography, with the highest sustained values occurring during the returns of the main hook at 1:07 and 2:04, and again leading to the coda at 2:43. These sections, and particularly the first two, share several moves and poses, although it is somewhat indicative of BTS's unorthodox approach to K-pop conventions that the choreography is generally more varied among these recapitulations than might be expected. 43

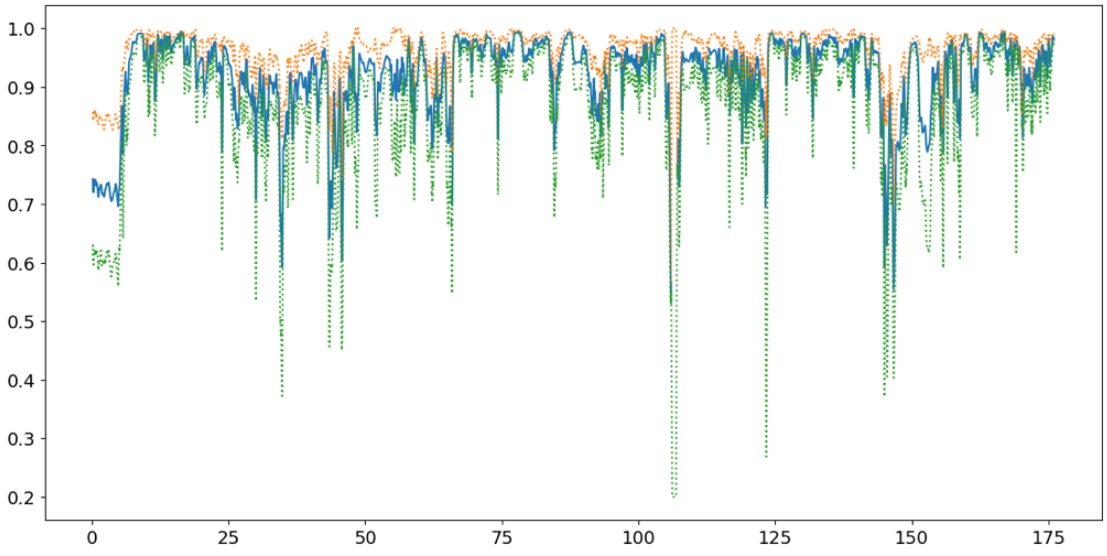

Figure 10. Mean (blue line) intra-frame pose similarity values with the standard deviation plotted above and below (orange and green dotted lines) for each frame of BTS's official dance practice video for "Fire." Similarity values were smoothed by computing the moving average of a one-second sliding window around each time value. Sections of high similarity (> 90%) indicate dancing with synchronized movements and poses among the group members.

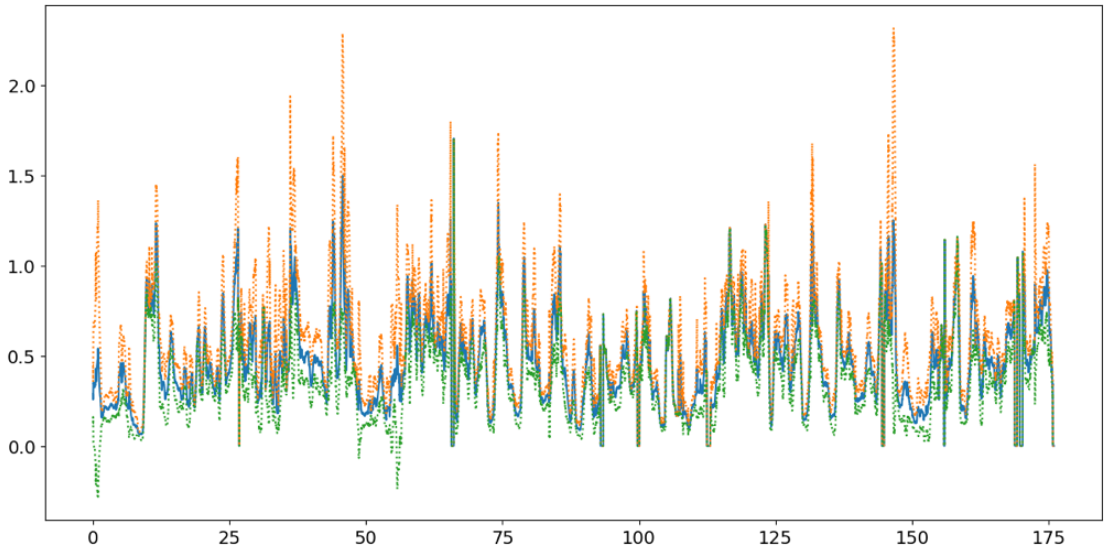

Figure 11. Mean (blue line) movement values averaged across all dancers for adjacent frames with the standard deviation plotted above and below (orange and green dotted lines). The movement values were smoothed by computing the moving average of a one-second sliding window around each time value.

Figure 12 animates the time-series analyses from Figure 10 and Figure 11 with progress indicators alongside playback of the dance practice video, further accompanied by a visualization of the average inter-frame movement values for each body keypoint computed across all figures in the frames, as well as a time-series heatmap and visualization of the most prominent key poses detected in the video. The "individual pose movement" timeline superimposes the inter-frame movement values for each of the detected dancers, highlighting some details that the averaging visualization (Figure 11) elides. Note that the analysis terminates prior to the conclusion of the dance practice video itself; the arrival of dozens of backup dancers at 3:00 makes it difficult to compare the conclusion section to what came before.

Comparative K-Pop Choreography Analysis: Video 2

from **Peter M. Broadwell**

Figure 12. Deep learning-based choreography analysis of the dance practice video for the boy group BTS's single “Fire.”

“Distant” movement analysis: early steps

There is a temptation to apply the techniques for pose characterization and synchronization detection described above in an evaluative manner, aggregating numbers to support the conclusion that idol group A is "more synchronized" than group B, or that group A employs a greater variety of dance poses than their presumably less talented and less dedicated fellows. Our intention, however, is to employ such "distant" analyses to provide material for a more in-depth consideration of the influences and factors shaping the parameters of creativity and production in K-pop dance. Just as establishing a typology of K-pop dance poses and gestures will aid in identifying influences from other genres, pose and motion analysis can help to investigate some of the cultural and artistic practices being enacted (or subverted) through K-pop dance performances. 45

As an illustration of the interpretive potential of "distant" aggregate analyses of choreographic corpora, we inspect the movement-based signatures of one of the most foregrounded structural factors in K-pop, namely the gender divide that results in the vast majority of idols being segregated into groups consisting solely of young men or young women. As mentioned above, a great deal of K-pop scholarship investigates the degree to which the performances, fashion, makeup, comportment, marketing and reception of K-pop idols either conforms to or seeks to blur notions of gender roles and modes of masculinity and femininity, in Korea or internationally. A computationally driven inquiry into dance choreography potentially can contribute much to this discussion. Our initial case study involved selecting an equal 46

number of K-pop dance practice videos from boy and girl idol groups over a limited time period and examining how the aggregate data produced from the time-series analyses described above can be used to answer, and more importantly, to raise questions about the role of gender in K-pop performance.

47 Our selection method involved surveying the available dance practice videos from boy and girl idol groups from the recent past and excluding videos with aspects that would complicate automated pose detection-based choreography analysis, such as camera angles that obscure dancers or studio mirrors and costumes that might confuse the pose-detection software. Of the remaining videos, we sought, albeit informally, to select a range of group sizes and song types that would be generally representative of the population of recent official dance practice videos. For this initial study, our data set consisted of the 20 videos (10 each from girl and boy groups) from 2017 to the present listed in Table 1. We ran each video through the pose detection, correction, tracking, interpolation, smoothing, movement and synchronization analyses discussed above, producing the summary statistics for each video presented in the table.

48 Also present in Table 1 as well as Table 2 is an additional statistic that we were interested in investigating: the degree and significance of the correlation between a group's averaged inter-frame movement time series and its average intra-frame pose similarity time series, as visualized in Figure 11 and Figure 10, respectively. One could reasonably expect these to be inversely correlated, i.e., a group's inter-pose similarity might become more difficult to maintain when its members are all moving quickly. Following this line of reasoning, we sought here to investigate the intuitive hypothesis that a reduced negative correlation between movement and synchronization might indicate what musicians and dancers (and others) refer to as a "tight" performance, i.e., one that maintains group cohesion amidst increased difficulty. To assign this notion a quantitative metric, we computed the Pearson correlation coefficient between the movement and pose similarity time series for each video. Under our hypothesis, a "tight" performance would have a higher correlation value than others. This inquiry is an early exemplar of our aspirations to introduce more discerning analytical criteria into our analytical methods than the overtly reductive quantification of synchronization and movement, as even a casual overview of a few K-pop choreography videos will reveal that there is a lot more going on than rote synchronization: different sub-groupings of the members may perform different moves simultaneously, or enact the same movements in cascading sequence over time, among many other patterns.

BOY GROUPS	**Video release date**	**Frames per sec**	**Group members**	**Avg movement per 1/6 sec per dancer**	**Avg intra-frame pose similarity**	**Pct of video with frame pose sim >.9**	**Movement: similarity correlation (Pearson)**
TXT - "Can't We Just Leave the Monster Alive?"	2020-04-12	29.97	5	5.31 +/- 4.01	0.93 +/- 0.09	72%	-0.29
TXT - "Angel or Devil"	2019-12-01	29.97	5	4.85 +/- 3.68	0.93 +/- 0.08	71%	-0.32
BTS - "DNA"	2017-09-24	29.97	7	4.94 +/- 3.58	0.9 +/- 0.11	55%	-0.14
BTS - "Idol"	2018-09-02	59.94	7	5.66 +/- 3.95	0.92 +/- 0.08	73%	-0.15
X1 - "Flash"	2019-09-04	29.97	10	4.74 +/- 3.78	0.88 +/- 0.14	56%	-0.37
Cravity - "Break All the Rules"	2020-04-20	29.97	9	5.71 +/- 3.43	0.93 +/- 0.07	72%	-0.28
Dongkiz - "Lupin"	2020-03-20	23.98	5	3.94 +/- 2.67	0.94 +/- 0.08	78%	-0.3
EXO - "Electric Kiss"	2018-01-12	29.97	8	5.74 +/- 3.74	0.93 +/- 0.07	77%	-0.26
SF9 - "Good Guy"	2020-01-09	23.98	9	4.35 +/- 3.37	0.92 +/- 0.09	70%	-0.35

Stray Kids - "Levanter"	2019-12-11	23.98	8	4.66 +/- 3.76	0.86 +/- 0.18	44%	-0.32
GIRL GROUPS							
AOA - "Excuse Me"	2017-01-10	23.98	7	2.59 +/- 1.99	0.96 +/- 0.04	87%	-0.11
Blackpink - "As If It's Your Last"	2017-06-24	23.98	4	5.9 +/- 3.97	0.96 +/- 0.05	68%	-0.04
EXID - "I Love You"	2018-11-26	29.97	5	2.6 +/- 2.37	0.96 +/- 0.06	88%	-0.08
GFriend - "Crossroads"	2020-02-05	59.94	6	6.88 +/- 7.83	0.95 +/- 0.06	81%	-0.31
Itzy - "Wannabe"	2020-03-11	60	5	7.22 +/- 7.85	0.95 +/- 0.08	80%	-0.6
Momoland - "I'm So Hot"	2019-03-24	29.97	7	4.04 +/- 2.97	0.95 +/- 0.08	82%	-0.7
Oh My Girl - "Bungee (Fall In Love)"	2019-08-13	29.97	7	3.97 +/- 2.64	0.91 +/- 0.11	74%	-0.28
Red Velvet - "Umpah Umpah"	2019-08-29	29.97	5	2.98 +/- 1.48	0.96 +/- 0.04	91%	-0.1
Twice - "Dance the Night Away"	2018-07-10	30	9	5.88 +/- 3.82	0.94 +/- 0.07	79%	-0.18
Twice - "Knock Knock"	2017-02-25	29.97	9	3.31 +/- 2.59	0.95 +/- 0.06	77%	0.08

Table 1. Per-video statistics for 20 choreography videos, 10 each from girl and boy groups. The average per-dancer movement (calculated every 1/6 of a second to allow straightforward comparison between videos recorded at 24, 30, and 60 frames per second) and the average intra-frame similarity for all dancers are presented with standard deviation values, which help to indicate whether the movement and similarity values are generally consistent or vary widely across the video.

COMBINED VIDEO STATISTICS				
Mean movement per 1/6 sec - boy groups	5.01 +/- 3.68		Welch's *t*-test - boys movement vs. girls movement (*t*)	9.76
Mean movement per 1/6 sec - girl groups	4.5 +/- 4.61		Welch's *t*-test - boys movement vs. girls movement (*p*)	1.77E-22
Mean intra-frame pose similarity - boy groups	0.92 +/- 0.09		Welch's *t*-test - boys similarity vs. girls similarity (*t*)	-33.1
Mean intra-frame pose similarity - girl groups	0.95 +/- 0.06		Welch's *t*-test - boys similarity vs. girls similarity (*p*)	7.26E-234
Movement:similarity correlation - boy groups (Pearson's *r*)	-0.24		Movement:similarity correlation - girl groups (Pearson's *r*)	-0.28

Table 2. Aggregate statistics for the 20 choreography videos from Table 1. To facilitate comparison between the "girl group" and "boy group" categories, the movement time series and intra-frame pose similarity time series for the 10 videos in each category were concatenated together. The distributions of the movement and pose similarity values then were compared to each other via a Welch's *t*-test, with the resulting *t*-statistic and two-tailed *p*-value (the probability that the differences between the two are due to chance) provided in the table. The correlation values between the movement and similarity time series for each category's meta-video also was measured by calculating Pearson's correlation coefficient (Pearson's *r*).

A perusal of the statistics in Table 1 gives the impression that despite considerable variability among videos, the girl group performances in our sample feature a greater amount of intra-frame pose similarity, while the boy groups tend to exhibit a higher degree of overall movement. The difference in pose similarity is especially evident from the fraction of

each video in which the intra-frame similarity (calculated using the Mantel matrix correlation test described above) is above 90%. The aggregate statistics in Table 2, derived by concatenating all of each category's videos together at a sample rate of 1/6th of a second (the smallest common subdivision among videos recorded at 24, 30 or 60 frames per second, so that videos with higher frame rates do not have disproportionate effect on the results) and applying a Welch's unequal variances *t*-test to compare the two resulting "meta-videos," confirm the high statistical significance of these gender-based disparities. Given the small number of videos in this sample relative to the thousands of choreographed K-pop numbers, however, it is entirely possible that these differences are solely an artifact of our video selection process. Even so, it is difficult to resist the impulse to divine the influence of gender-based cultural notions of activity/passivity, conformance/individuality, extroversion/introversion, as well as physical differences, upon the numbers. More productive and less statistically questionable humanistic insights, however, may be derived by examining the exceptions to and outliers of these trends, as the following example briefly demonstrates.

50 The analysis of the correlation between movement and synchronization via the Pearson correlation test described above produced intriguing but not entirely conclusive results. As with the other statistics, it would benefit from a larger sample size and from controlling for song tempo (e.g., in average beats per minute) as a numerical proxy for song style. Yet also in a similar manner to the other comparisons, there is potentially even more to learn by investigating the outliers rather than merely the aggregated statistics. In this case, the dance performance of the girl group Twice's single "Knock Knock" is the only video in either category with a positive correlation between movement and synchronization. Watching the video closely reveals that this correlation (or more to the point, the lack of a negative correlation) is due to the choreography's fairly unorthodox alternation of sustained tableaux featuring multiple sub-divisions of the group's members in contrasting poses (low similarity, low movement) with more standard synchronized dance sections (high similarity, higher movement). The use in this performance of a choreographic technique that differs considerably in kind and degree from the rest serves to highlight the diverse range of aesthetic "concepts" that K-pop groups and their artistic collaborators deploy to differentiate one release from another. The further study of such outliers promises to lead to a more in-depth understanding of the role of dance in the K-pop culture industry. And returning to the question of whether it might be possible to quantify the "tightness" of a performance, this early experiment suggests that the metric also would need to take into account the speed of the movements (i.e., motion over time), assigning more significance to faster motions.

51 As the discussion above indicates, a sizable set of other "distant" analyses remains to be explored, in addition to expanding the size of the analyzed video corpus. Further metrics might include more summational measures, such as calculating an entire dance's average pose-based entropy (roughly indicating the "diversity" of poses) or its degree of movement autocorrelation (whether classes of similar poses, or sections featuring high-velocity movement, tend to appear in quick succession, or else tend to alternate regularly with contrasting materials), or quantifying the "fluidity" or "jerkiness" of movements individually or collectively. The brief analyses presented thus far illustrate some of the potential of these methods.

7. Conclusions and next steps

52 Our application of pose detection techniques and methods for comparison of the resulting matrices offers a visually rich approach to understanding dance similarity both within single K-pop videos and across multiple videos. While we are still quite far from being able to select a dance move or dance sequence and find all other instances of it in the ever-expanding K-pop video corpus, our experiments have shown that this is no longer a distant hope. The development of increasingly accurate pose detection algorithms coupled to relatively fast algorithms for calculating distances and, thus, similarities across detected poses allows us to identify a clear path forward for dance move and sequence comparison. Already, our work has produced intriguing discoveries, such as the relative infrequency of absolute group coordination in K-pop videos despite the general consensus that this is a key feature of K-pop dance performances.

53 These initial experiments are a foundational beginning to the complex problem of devising methods for the consistent discovery of dance moves and dance patterns at scale. There is significant room for development, and these preliminary engagements with the complex material of K-pop music videos suggest several productive avenues for future inquiry. The nascent field of "big dance" studies is already becoming more active: recently, the Google Arts &

Culture group has engaged in high-profile collaborations with the choreographer Wayne McGregor with the goal of devising a motion-based search engine for a large, curated corpus of dance recordings. Their recurrent neural network-based next-pose generation project at the very least indicates one direction in which this type of research can be developed.[5] More relevant to the present study is the "Living Archive," an experimental UMAP embedding visualization web app, which is an excellent example of a big-data choreography analysis project, albeit one focused on a single choreographer's work.[6] It is worth noting that the latter project is not more sophisticated technically than our pilot experiments, which thus suggests that relatively small DH research groups such as ours (approximately four people) are able to move the needle considerably without the massive resources of an internet giant.

While our work has focused exclusively on K-pop videos, the encouraging results suggest that these methods can help us address a series of complex questions not only in this corpus, but across many dance traditions that have been filmed. These questions include the discovery and documentation of stylistic similarity across large dance corpora. Given the ability to align videos with dates of production, these discoveries of similarities and, equally importantly, variations in similar dances could help with the characterization of large-scale dance trends over both time and, in the case of geographically linked traditions such as West Coast Hip Hop or K-pop, space. Through the application of neighborhood detection and clustering algorithms, it should also be possible to use the similarity of dance sequences as part of a classification system, allowing for the computational characterization of sub-genres [Abello et al. 2012]. Intriguingly, given the ability of pose estimation to label body parts automatically, it may be possible to devise a typology of poses and, with an analysis of sequencing, dance moves, leading to an understanding of the vocabulary of moves and sequences in any video or group of videos. Such a computationally derived morphology of K-pop dance could, in turn, lead to investigations of influence, homage and borrowing (inadvertent or intentional) from other domains such as hip hop, Bollywood, Latin dance, martial arts and many other traditions, including traditional Korean dance genres such as p'ungmul performance. 54

In short, our approach confirms the need for a macroscopic approach to the complex domain of popular dance. In addition to comparing poses, motions, moves, and sequences across thousands of videos, researchers are eager to analyze these dances at many different scales, from the broad domain of an entire genre such as K-pop all the way down to the individual performances of a specific dancer at a specific time. We believe these preliminary investigations and refinements of various techniques, given the productive results derived from their application, provide a clear roadmap for the further development of these methods and will help us answer open questions regarding dance development not only for K-pop but for many other genres as well. 55

Acknowledgements and notes

We are grateful to Dr. Paul Chaikin for suggesting the Delaunay triangulation approach to graph Laplacian pose characterization and comparison. Dr. Francesca Albrezzi and the students of the Winter 2019 Digital Humanities capstone seminar at UCLA inspired us to expand the range of research questions and corpora treated in this paper. We also thank the contributors to the K-Pop Database site (dbkpop.com) for maintaining a detailed listing of available K-pop dance practice videos on YouTube. 56

Source code for the examples presented in the text is available at https://github.com/broadwell/choreo_k 57

Notes

[1] Such systems have yet to appear, though similar motion mimicry-based dance games exist for the now-discontinued Microsoft Kinect and Nintendo Wii consoles. The dance move data for specific K-pop singles remains for sale on an online portal associated with the project, http://shop2.mocapkpop.cafe24.com.

[2] Computational image analysis techniques proceed at speeds much closer to real-time human viewing than, say, computational text analysis tasks.

[3] https://www.youtube.com/watch?v=Zu3hBEZ0RvA

[4] https://www.youtube.com/watch?v=sWuYspuN6U8

[5] https://artsandculture.google.com/story/1AUBpanMqZxTiQ

[6] https://artsexperiments.withgoogle.com/living-archive/

Works Cited

Abello et al. 2012 Abello, James, Peter Broadwell, and Timothy R. Tangherlini. “Computational Folkloristics.”*Communications of the ACM* 55.7 (2012): 60–70.

Ahmadyan and Hou 2020 Ahmadyan, Adel, and Tingbo Hou. “Real-Time 3D Object Detection on Mobile Devices with MediaPipe.”*Google AI Blog* (blog). Accessed March 11, 2020. https://ai.googleblog.com/2020/03/real-time-3d-object-detection-on-mobile.html.

Andriluka et al. 2018 Andriluka, Mykhaylo, Umar Iqbal, Anton Milan, Eldar Insafutdinov, Leonid Pishchulin, Juergen Gall, and Bernt Schiele. “PoseTrack: A Benchmark for Human Pose Estimation and Tracking.”In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, IEEE, Salt Lake City, UT, USA (2018) p. 5167. https://doi.org/10.1109/CVPR.2018.00542.

Ankerst et al. 1999 Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. “OPTICS: Ordering Points to Identify the Clustering Structure.”*ACM SIGMOD Record* 28.2 (June 1, 1999): 49–60. https://doi.org/10.1145/304181.304187.

Arnold and Tilton 2019 Arnold, Taylor, and Lauren Tilton. “Distant Viewing: Analyzing Large Visual Corpora.”*Digital Scholarship in the Humanities*, March 16, 2019. https://doi.org/10.1093/digitalsh/fqz013.

Blok et al. 2018 Blok, Dylan, Jacob Pettigrew, Thecla Schiphorst, and Herbert H. Tsang. “Human Pose Detection Through Searching in 3D Database With 2D Extracted Skeletons.”In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 470–76. IEEE, Bangalore, India (2018). https://doi.org/10.1109/SSCI.2018.8628776.

Broadwell et al. 2016 Broadwell, Peter, Timothy Tangherlini, and Hyun Kyong Hannah Chang. “Online Knowledge Bases and Cultural Technology: Analyzing Production Networks in Korean Popular Music.”In Jieh Hsiang (ed.), *Digital Humanities: Between Past, Present, and Future*. NTU Press, Taipei (2016), pp. 369– 94.

Byeon et al. 2016 Byeon, Yeong-Hyeon, Sung-Bum Pan, Sang-Man Moh, and Keun-Chang Kwak. “A Surveillance System Using CNN for Face Recognition with Object, Human and Face Detection.”In Kuinam J. Kim and Nikolai Joukov (eds), *Information Science and Applications (ICISA) 2016* 376:975–84. Lecture Notes in Electrical Engineering. Springer Singapore, Singapore (2016). https://doi.org/10.1007/978-981-10-0557-2_93.

Börner 2011 Börner, Katy. “Plug-and-Play Macroscopes.”*Communications of the ACM* 54.3 (2011): 60–69.

Cao et al. 2018 Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields.”*CoRR* abs/1812.08008 (2018). http://arxiv.org/abs/1812.08008.

Choi and Maliangkay 2014 Choi, JungBong, and Roald Maliangkay. *K-Pop–The International Rise of the Korean Music Industry*. Routledge (2014).

Chung 1997 Chung, Fan R. K. *Spectral Graph Theory*. Regional Conference Series in Mathematics 92. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I. (1997).

Delaunay 1934 Delaunay, Boris. “Sur La Sphère Vide.”*Bulletin de l'Académie Des Sciences de l'URSS, Classe Des Sciences Mathématiques et Naturelles* 6 (1934): 793–800.

Dong et al. 2017 Dong, Ran, Dongsheng Cai, and Nobuyoshi Asai. “Dance Motion Analysis and Editing Using Hilbert-Huang Transform.”In *ACM SIGGRAPH 2017 Talks*, 75:1–75:2. SIGGRAPH '17. New York, NY, USA : ACM (2017). https://doi.org/10.1145/3084363.3085023.

Elfving 2018 Elfving-Hwang, Joanna. “K-Pop Idols, Artificial Beauty and Affective Fan Relationships in South Korea.”In *Routledge Handbook of Celebrity Studies*, Routledge (2018), pp. 190–201.

Elhayek et al. 2017 Elhayek, A., E. de Aguiar, A. Jain, J. Thompson, L. Pishchulin, M. Andriluka, C. Bregler, B. Schiele, and C. Theobalt. “MARCOnI — ConvNet-Based MARker-Less Motion Capture in Outdoor and Indoor Scenes.”*IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.3 (March 1, 2017): 501–14. https://doi.org/10.1109/TPAMI.2016.2557779.

Epstein 2016 Epstein, Stephen. “From South Korea to the Southern Hemisphere: K-Pop below the Equator.”*Journal of*

World Popular Music 3.2 (2016): 197–223.

Geertz 1973 Geertz, Clifford. *Thick Description: Toward an Interpretive Theory of Culture*. Basic Books, New York (1973), pp. 3–30.

Giraldo et al. 2018 Giraldo, Ramón, William Caballero, and Jesús Camacho-Tamayo. “Mantel Test for Spatial Functional Data: An Application to Infiltration Curves.”*AStA Advances in Statistical Analysis* 102.1 (January 2018): 21–39. https://doi.org/10.1007/s10182-016-0280-1.

Güler et al. 2018 Güler, Riza Alp, Natalia Neverova, and Iasonas Kokkinos. “DensePose: Dense Human Pose Estimation In The Wild.”*CoRR* abs/1802.00434 (2018). http://arxiv.org/abs/1802.00434.

Han 2017 Han, Benjamin. “K-Pop in Latin America: Transcultural Fandom and Digital Mediation.”*International Journal of Communication (19328036)* 11 (2017).

Hara et al. 2017 Hara, Kotaro, Abi Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey Bigham. “A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk.”*ArXiv:1712.05796 [Cs]*, December 28, 2017. http://arxiv.org/abs/1712.05796.

Howard 2015 Howard, Keith. “Politics, Parodies, and the Paradox of Psy's 'Gangnam Style.'”*Romanian Journal of Sociological Studies*, 1 (2015): 13–29.

Jang and Song 2017 Jang, Won Ho, and Jung Eun Song. “The Influences of K-Pop Fandom on Increasing Cultural Contact: With the Case of Philippine Kpop Convention, Inc.”지역사회학 18 (2017): 29–56.

Jin and Ryoo 2014 Jin, Dal Yong, and Woongjae Ryoo. “Critical Interpretation of Hybrid K-Pop: The Global-Local Paradigm of English Mixing in Lyrics.”*Popular Music and Society* 37.2 (2014): 113–131.

Kim 2015 Kim, Jeong Weon. “행보와 동행:≪ 월간 윤종신≫ 의 매체와 협업에 관한 고찰.”대중음악 15 (2015): 45–73.

Kim 2017a Kim, Dohyung. “생체역학적용 K-POP 댄스 안무 검색 및 자세 정확성 분석 기술 개발 (The Development of the Choreography Retrieval System from the K-POP Dance Database Including Biomechanical Information and the Analysis Technology of the Correctness of a Dance Posture).” Electronics and Telecommunications Research Institute, Daejeon, Korea (2017). http://www.ndsl.kr/ndsl/search/detail/report/reportSearchResultDetail.do?cn=TRKO201700003705.

Kim 2017b Kim, Gooyong. “Between Hybridity and Hegemony in K-Pop's Global Popularity: A Case of Girls' Generation's American Debut.”*International Journal of Communication (19328036)* 11 (2017).

Kim 2018 Kim, Suk-Young. *K-Pop Live: Fans, Idols, and Multimedia Performance*. Stanford University Press, Stanford , CA (2018). http://www.sup.org/books/title/?id=29375.

Kim and Ryoo 2007 Kim, Eun Mee and Jiwon Ryoo. “South Korean Culture Goes Global: K-Pop and the Korean Wave.” *Korean Social Science Journal* 34.1 (2007): 117-152.

Kim et al. 2017 Kim, Dohyung, Dong-Hyeon Kim, and Keun-Chang Kwak. “Classification of K-Pop Dance Movements Based on Skeleton Information Obtained by a Kinect Sensor.”*Sensors* 17.6 (June 1, 2017): 1261. https://doi.org/10.3390/s17061261.

Kim et al. 2018 Kim, Yeonho, and Daijin Kim. “Real-Time Dance Evaluation by Markerless Human Pose Estimation.”*Multimedia Tools and Applications* 77.23 (December 2018): 31199–31220. https://doi.org/10.1007/s11042-018-6068-4.

Kreiss et al. 2019 Kreiss, Sven, Lorenzo Bertoni, and Alexandre Alahi. “PifPaf: Composite Fields for Human Pose Estimation.”*CoRR* abs/1903.06593 (2019). http://arxiv.org/abs/1903.06593.

Kumar et al. 2020 S.V. Aruna Kumar, Ehsan Yaghoubi, Abhijit Das, B.S. Harish and Hugo Proença. “The P-DESTRE: A Fully Annotated Dataset for Pedestrian Detection, Tracking, Re-Identification and Search from Aerial Devices” *arXiv:2004.02782*, 2020. http://arxiv.org/abs/2004.02782.

Laurie 2016 Laurie, Timothy N. “Toward a Gendered Aesthetics of K-Pop.”In Chapman, Ian, and Henry Johnson (eds), *Global Glam and Popular Music: Style and Spectacle from the 1970s to the 2000s*, 1st ed., Routledge (2016).

Lee 2015 Lee, Seoung-Ah. “Of the Fans, by the Fans, for the Fans: JYJ Republic.”*Hallyu 2.0: The Korean Wave in the Age of Social Medial.* University of Michigan Press, Ann Arbor (2015), pp. 108–130.

Lee and Nornes 2015 Lee, Sangjoon, and Abé Markus Nornes. *Hallyu 2.0: The Korean Wave in the Age of Social Media*.

University of Michigan Press, Ann Arbor (2015). https://doi.org/10.3998/mpub.7651262.

Lie 2012 Lie, John. “What Is the K in K-Pop? South Korean Popular Music, the Culture Industry, and National Identity.”*Korea Observer* 43.3 (2012): 339–363.

Lin et al. 2015 Lin, Tsung-Yi, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. “Microsoft COCO: Common Objects in Context.”*ArXiv:1405.0312 [Cs]*, February 20, 2015. http://arxiv.org/abs/1405.0312.

Liu et al. 2019 Liu, Shuangjun, Yu Yin, and Sarah Ostadabbas. “In-Bed Pose Estimation: Deep Learning With Shallow Dataset.”*IEEE Journal of Translational Engineering in Health and Medicine* 7 (2019): 1–12. https://doi.org/10.1109/JTEHM.2019.2892970.

Manietta 2015 Manietta, Joseph Bazil. *Transnational Masculinities: The Distributive Performativity of Gender in Korean Boy Bands*. University of Colorado Boulder (2015).

Mathis et al. 2018 Mathis, Alexander, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. “DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning.”*Nature Neuroscience* 21.9 (September 2018): 1281–89. https://doi.org/10.1038/s41593-018-0209-y.

Messerlin and Shin 2017 Messerlin, Patrick A., and Wonkyu Shin. “The Success of K-Pop: How Big and Why So Fast?”*Asian Journal of Social Science* 45.4–5 (2017): 409–439.

Moon et al. 2018 Moon, Gyeongsik, Ju Yong Chang, and Kyoung Mu Lee. “PoseFix: Model-Agnostic General Human Pose Refinement Network.”*CoRR* abs/1812.03595 (2018). http://arxiv.org/abs/1812.03595.

Oh 2014a Oh, Chuyun. “Performing Post-Racial Asianness: K-Pop's Appropriation of Hip-Hop Culture.”In *Congress on Research in Dance*, Cambridge University Press (2014), pp. 121–125.

Oh 2014b Oh, Chuyun. “The Politics of the Dancing Body: Racialized and Gendered Femininity in Korean Pop.”In *The Korean Wave*, Springer (2014), pp. 53–81.

Oh 2015 Oh, Chuyun. “Queering Spectatorship in K-Pop: The Androgynous Male Dancing Body and Western Female Fandom.”*The Journal of Fandom Studies* 3.1 (2015): 59–78.

Oh and Lee 2014 Oh, Ingyu, and Hyo-Jung Lee. “K-Pop in Korea: How the Pop Music Industry Is Changing a Post-Developmental Society.”*Cross-Currents: East Asian History and Culture Review* 3.1 (2014): 72–93.

Ota 2015 Ota, Kendall. “Soft Masculinity and Gender Bending in Kpop Idol Boy Bands.” In *Cal Poly Pomona Lectures* (2015). http://hdl.handle.net/10211.3/138192.

Otmazgin and Lyan 2014 Otmazgin, Nissim, and Irina Lyan. “Hallyu across the Desert: K-Pop Fandom in Israel and Palestine.”*Cross-Currents: East Asian History and Culture Review* 3.1 (2014): 32–55.

Pavllo et al. 2019 Pavllo, Dario, Christoph Feichtenhofer, David Grangier, and Michael Auli. “3D Human Pose Estimation in Video with Temporal Convolutions and Semi-Supervised Training.”*ArXiv:1811.11742 [Cs]*, March 29, 2019. http://arxiv.org/abs/1811.11742.

Raptis et al. 2011 Raptis, Michalis, Darko Kirovski, and Hugues Hoppe. “Real-Time Classification of Dance Gestures from Skeleton Animation.”In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '11*, 147. ACM Press, Vancouver, British Columbia (2011). https://doi.org/10.1145/2019406.2019426.

Reilly 2013 Reilly, Kara. *Theatre, Performance and Analogue Technology: Historical Interfaces and Intermedialities*. Palgrave Studies in Performance and Technology. Palgrave MacMillan, Basingstoke (2013).

Rizzo et al. 2018 Rizzo, Anna, Katerina El Raheb, and Sarah Whatley. “WhoLoDance: Whole-Body Interaction Learning For Dance Education.”In *Proceedings of the Workshop on Cultural Informatics* (2018) Vol. 2235: 41–50. November 3, 2018. https://doi.org/10.5281/ZENODO.1478033.

Saeji 2016 Saeji, Cedarbough. “Cosmopolitan Strivings and Racialization: The Foreign Dancing Body in Korean Popular Music Videos.”In *Korean Screen Cultures: Interrogating Cinema, TV, Music and Online Games*, edited by David Jackson and Colette Balmain, Peter Lang Publishers, Oxford (2016), pp. 257–92.

Sutil 2015 Sutil, Nicolás Salazar. *Motion and Representation: The Language of Human Movement*. MIT Press, Cambridge, Massachusetts (2015).

Unger 2015 Unger, Michael A. “The Aporia of Presentation: Deconstructing the Genre of K-Pop Girl Group Music Videos in South Korea.”*Journal of Popular Music Studies* 27.1 (2015): 25–47.

Watts 2015 Watts, Victoria. “Benesh Movement Notation and Labanotation: From Inception to Establishment (1919–1977).”*Dance Chronicle* 38.3 (2015): 275–304.

Wevers and Smits 2019 Wevers, Melvin, and Thomas Smits. “The Visual Digital Turn: Using Neural Networks to Study Historical Images.”*Digital Scholarship in the Humanities*, January 18, 2019. https://doi.org/10.1093/llc/fqy085.

Xiu et al. 2018 Xiu, Yuliang, Jiefeng Li, Haoyu Wang, Yinghong Fang, and Cewu Lu. “Pose Flow: Efficient Online Pose Tracking.”*ArXiv:1802.00977 [Cs]*, July 2, 2018. http://arxiv.org/abs/1802.00977.