The 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2012) *Short and Project Papers*
D. Arnold, J. Kaminski, F. Niccolucci, and A. Stork (Editors)

Towards an Affordable Markerless Acquisition of Intangible Contemporary Dance Choreographies at Large-Scaled Stages

S. Kahn[1] and J. Keil[1] and M. Zoellner[2] and B. Mueller[1]

[1]Fraunhofer IGD, Germany [2]Hof University, Germany

Abstract

While the documentation and preservation of rigid cultural heritage objects has become much easier with technologies such as 3D scanning or photogrammetry technologies, the digitalization of 3D intangible moving content is still a major issue. This concerns also the in situ creation of digital dance representations and the question of how to preserve and disseminate dance performances. In this paper, we present a generic and affordable approach for an automatized and markerless capturing of the movements of dancers, which was developed in the Motion Bank research project, as well as first application examples which analyse and visualize the captured dance data. The captured data is stored in a cloud based service and is thus made available for online and offline processing.

Categories and Subject Descriptors (according to ACM CCS): Image Processing and Computer Vision [I.4.8]: Scene Analysis—Motion, Tracking; Arts and Humanities [J.5]: Performing Arts—Dance;

1. Introduction

In the last several years, considerable efforts by leading contemporary choreographers (e.g. M. Cunningham, W. Forsythe, S. Davies, W. McGregor, E. Greco | PC) and their dance companies have been undertaken to document, analyse and disseminate both choreographic concepts and dance performances using digital technologies [Bir02] [DS06] [Ban10]. One of the earliest and best known examples of this is the Improvisation Technologies CD-Rom developed in the mid-1990s by choreographer William Forsythe in collaboration with the Center for Media Arts (ZKM) in Karlsruhe. This was followed by Forsythe in 2005 with the development of the web-based dance education project "Synchronous Objects for One Flat Thing, reproduced" in collaboration with the Advanced Computing Center for the Arts and Design, The Ohio State University [Sha11] [Syn12]. Both of these important projects used the technique of manually annotating digital video recordings of the dance to show the viewer specific patterns in the movement of a single dancer and the relationships between dancers [Syn12]. These patterns are not visible especially to the untrained observer when watching the performance. The result is not only the preservation of the choreographic concepts and dance performances in digital media, it is also made more accessible to anyone interested in developing a deeper understanding of dance and choreography. This directly addresses a central concern for cultural preservation: how to preserve not only recordings, but also additional information about context and meaning. However, the manual, very work and cost intensive approach of creating the annotations used in Improvisation Technologies and [Syn12] presents a challenge to further education in this field and its future cultural preservation using digital media technologies. The question for computer science is what might we offer the field of live dance performance to enable these future developments in ways that are more cost-effective and take advantage of the latest in technical developments. However, the core computer science research aim here is to use these latest technical developments in a way that is sensitive to the rich meaningful content and cultural context of contemporary dance. In this paper, therefore, we present a generic and affordable approach for an automatized, computer vision approach based on capturing the movements of dancers that does not disturb dancers by attaching markers or sensors to their bodies, which would disturb the qualities and the meaningful content of their movement; and also aims to avoid an intrusive preparation of the entire large-scaled environment in terms of the hardware setup that would be impossible for a dance education school or a dance company to afford or support technically. We also present early visualizations and stages of a pipeline from data to information developed in collaboration with colleagues in digital art and design and with the choreographers to ensure meaningful content that can be reviewed and replayed with web technologies. Our approach is being developed and utilized in Motion Bank [Mot12], a research project of The Forsythe Company that has the aim to develop four on-line digital dance scores with guest choreographers. This project builds on the Synchronous Objects project, but it is exactly our role in the project to explore the development of an automatized, computer vision based recording of the dancers' shapes and 3D movement paths. Interactive online visualization technologies are then used to augment the dance with additional information extracted from the captured data with the aim to reveal inherent, yet unobvious choreography and movements patterns. While the predecessor project used the proprietary Flash platform, our solution consistently focuses on open web standards.

2. Capturing Dance

While specialized high-end motion capturing solutions can capture human movements with a high accuracy and reliability, these solutions are usually very expensive. Furthermore, very specialized solutions often can only be used for a specific setup and are not easily reusable for other dances or different stages. As the goal of our research is to provide a solution based on low-cost, modular and scalable technologies, we explicitly refrain from using professional high-end motion capture setups and instead propose a low-tech approach, which is based on three main concepts: **Reusability (adaptability to other scenarios):** Particular attention is paid to develop a generic approach which is not restricted to a particular stage, but which can easily be adapted to other stages and different scenarios. This is of particular importance for capturing modern dance: The size of the stage varies strongly for different choreographies and can range from few square meters to several hundred square meters for a dance stage which has a size of 20×30 meters. **Affordability and Scalability:** To ensure the affordability of the capturing hardware, our setup is based on default video camcorders (in contrast to professional motion capture setups, which often require high-end cameras). Our solution scales with an arbitrary number of cameras. **Markerless capturing, no sensors attached to dancers:** Many state-of-the-art motion capture technologies still require the use of motion capture suits. However, motion capture suits (as well as other sensors attached to the human body) hinder the natural movement of dancers. Therefore, we thoroughly focus on markerless motion capture technology for capturing the movement of dancers without disturbing the dancers' moves.

2.1. Capturing Setup

For recording the dance without the need to connect each camera to a computer, we use standalone camcorders which record the captured videos on a local hard disk. This provides better flexibility in positioning the cameras and a much simpler hardware setup, which is easier to install. Furthermore, the distance of the cameras is not restricted by maximal cable lengths (e.g., the signal of firewire cameras can become quite unstable after a few meters, resulting in frame loss). The 2D camcorders capture HD videos with a constant framerate of 50 fps. In our basic setup, we use three camcorders which capture the stage from above the point of view of the audience (one camcorder on the left, one centered and one on the right). This setup can easily be extended by adding more cameras, for example for reconstructing a dense 3D visual hull of the dancers. For processing the 2D images (camera pose estimation and 3D reconstruction of the dancers' movements), the intrinsic parameters are estimated for each camcorder by analyzing the projection of a chessboard calibration pattern onto the camera images.

2.2. Camera Pose Estimation and Synchronization

To calculate the pose (position and orientation) of a camera, a set of 2D-3D correspondences is required. Each 2D-3D correspondence contains the 3D position of a point in the world, as well as its 2D projection onto the camera image. In order to robustly estimate a camera pose, the 2D-3D correspondences should cover a large area of the 2D camera image. Therefore, we evenly distribute spherical markers on the whole stage. The markers are composed from several nested circles, such that the center of the pattern is easily detectible from an oblique point of view. The positions of the marker centers are measured with a 1D laser scanner and the camera poses are estimated from the 2D-3D correspondences between the 3D marker centers and their projections in the 2D camera images. The captured video streams are synchronized with a filming clapboard, by the frame in which the clap sticks are shut. The latter synchronization is ensured by the fact that the cameras record the videos with a constant framerate. We decided not to use hardware triggering for the camera synchronization, because this would require to install trigger hardware and cables all around the stage. Furthermore, only cameras could be used which have built-in trigger capabilities.

3. Video Analysis and Data Distribution

The video analysis provides the basis for web based dance visualizations, e.g. by significantly reducing the required amount of storage by the factor 50-300. The main data reduction is achieved by extracting the silhouette of the dancer from the captured 2D images: While each HD image has a resolution of 1920×1080 pixel, the bounding box around a dancer typically only has a size of about 200×200 pixel. The view of the dancer on the stage can be recombined from the silhouette and a background image of the empty stage.

3.1. Silhouette Extraction

Figure 1 visualizes colored silhouettes extracted from a video of a dancer. The image pixels which form the silhou-

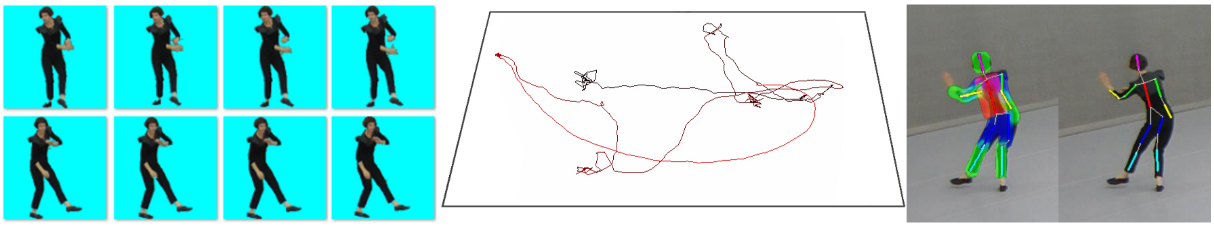

Figure 1: *Left: Colored silhouettes of a dancer. Center: Top view of 3D movement path on the stage. Right: 2D pose estimation.*

ette of the dancers are identified by background subtraction. First, a background model (storing the mean and the variance of each pixel's color) of the stage is learned from several images of the empty stage. These images are recorded before the dancers enter the stage. Then, for each pixel of each new camera image, the color value of the pixel is compared with the mean and the variance of the background model in order to decide whether the color difference is significant enough to interpret the current measurement as a foreground pixel. Then, a connected-component merging is applied to find the largest connected segment. All other foreground pixels are discarded. The background model is updated continuously in order to account for lightning changes.

3.2. 2D Human Pose Estimation & 3D Reconstruction

For estimating the 3D movement path of a dancer, the uppermost position of the head is tracked in the recorded video sequences. The head can be tracked more stable than the upper body (especially, if the dancer wears unicolored clothing as in Figure 1) and more stable than the feet (due to shadows of the dancer on the dance floor). The easiest way to approximate the uppermost position of the head in a 2D camera image is to assume that the uppermost point of the silhouette of the dancer is the uppermost position of the head. To account for the situations in which the head is not the uppermost point of the silhouette, the body pose of the dancer in the 2D image needs to be estimated (see Figure 1). For maximizing the generality and reproducibility of our data processing pipeline, a publicly available state-of-the-art tool for the 2D human pose estimation [EMJZF11] can be used. After estimating the position of tracked body parts of the dancers in the 2D images, the 3D positions of the tracked body parts are reconstructed by triangulation. Given the view rays from at least two cameras (which intersect the 2D positions of the tracked body parts in the 2D images), the 3D position of the tracked body part is reconstructed by calculating the closest intersection of the view rays. Figure 1 visualizes a snapshot of a dancer's 3D movement path, which was calculated by triangulating the uppermost points of the silhouettes.

3.3. Cloud Based Data Distribution

After processing, silhouette files and position data are stored on a cloud service. Once called, its API returns the data as JSON strings, this way making the data accessible for online and offline visualization tools.

4. From Data to Information: Early Visualizations

An important aspect of our reference dance is that it has no traditional underlying choreography as one might expect. Instead, the entire piece works with several stages that emerge through the dancer's movements throughout time. So, even if one dancer is performing several takes of the same piece, starting points, positions on stage and movements do change, although there is still a perceivable structure. In order to increase comprehension and to give an informed view on the piece's choreography and body movements, we present ways to increase understanding through data analysis and interactive data visualization techniques that add further information to the recorded dance. With acquired data from three different dancers performing the dance one after another, we explored several different visualization types: the first is web-browser based, the second is a mobile augmented reality application, and the third goes into analysis by comparing stroke and path movements of all three dancers.

4.1. Prototyping a First Web Visualization with X3Dom

The web-based visualization acts as an interactive tool, able to load and present the silhouette images and position data interactively. We employed X3DOM to generate 3D paths of each dancer on a virtual stage in addition to small plane geometries that are textured with the silhouettes. Each frame, the image textures are updated and the geometry is spatially repositioned along the trajectory path for each dancer, giving the impression as if the dance was re-performed on the virtual stage. While playing, the stage can be rotated around its center and be viewed from different viewing angles (e.g. perspective or top-view). Additionally, the pace can be controlled and each data stream is handled as a separate layer, which can be shown or hidden. This way, the online tool envisions pathways and patterns one would usually not see.

4.2. Visualizing Data in Augmented Reality (AR)

We transfer the captured data back to reality through an AR illustration where the dance can be viewed through the

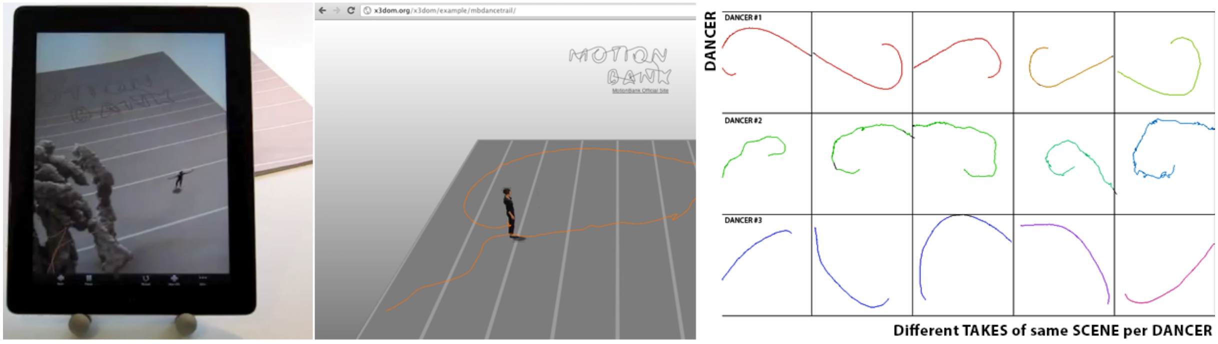

Figure 2: *From left-to-right: a) Augmented Reality application on tablet: Moving dancer augmented onto flyer. b) Screenshot of the web based X3DOM visualization . c) Plot of path variances and curve styles per dancer for a specific scene.*

screen of a recent mobile device (smartphone or tablet), with AR's video-see-through effect where one could see the dancer performing e.g. on a desk at home; this way transforming the desk into a virtual stage and the captured dancer into a ballerina/surrogate on the mobile device. We created a computer vision AR application with the instantAR framework and used a flyer from the project as tracking target. Similar to the web application, path and dancer are visualized on the flyer, allowing to access different views controlled through device motion: as soon as the user points his tablet to the flyer, the visualization is aligned to and superimposed onto it, giving the impression as if the dance was performed in front of the user (see Figure 2).

4.3. Visualizing Path Variance and Curve Analysis

In order to go deeper into data analysis, the third visualization envisions the fashion of how the dance took formally place spatially throughout time, using the position data. The intention is to illustrate how the same scene changes per dancer (intra-personal) and among all three of them (inter-personal). Although the nature of our reference dance is quite fluent, by having labels that tag intersections of scenes/steps for each recording, we took a closer look on one particular scene and compared it visually. The plot of Figure 2 c) illustrates movement styles of each dancer, where each row shows a specific scene from different takes (i.e. data acquired during different performances) of one of the three dancers. While the core choreography for this scene implies dancing a curve, the plots perfectly reveal that those might change in start-and-ending and orientation per dancer and among all of them – but in the end still define a curve. Quite impressive is also that in a formal sense the lines of each dancer (plots per row) reveal that a dancer has a very personal way of expression which gives each line some kind of style or signature, which doesn't become as obvious, when watching the real performance compared to the stroke plots.

5. Future Work

In this paper, we have described a markerless, reusable and scalable approach for capturing and visualizing dance performances on large stages. In the ongoing project, we will upload the captured and processed dance data to make it publicly available on a public website. Furthermore, we are currently extending the presented processing pipeline by incorporating full depth image based skeleton tracking for dance choreographies on smaller stages.

Acknowledgments

The Motion Bank research project is funded by the German Federal Cultural Foundation, the Hessisches Ministerium für Wissenschaft und Kunst, the kulturfonds frankfurtrheinmain and Ms. Susanne Klatten. We wish to thank all involved choreographers, dancers and project partners for their close and productive cooperation, with special thanks to our digital artists Florian Jenett and Amin Weber and to our project leader Scott DeLahunta.

References

[Ban10] BANNON F.: Dance: the possibilities of a discipline. *Research in Dance Education 11*, 1 (2010), 49–59. 1

[Bir02] BIRRINGER J.: Dance and media technologies. *PAJ: A Journal of Performance and Art 24*, 1 (PAJ 70) (2002), 84–93. 1

[DS06] DELAHUNTA S., SHAW N. Z.: Constructing memories: Creation of the choreographic resource. *Performance Research 11*, 4 (2006), 53–62. 1

[EMJZF11] EICHNER M., MARÍN-JIMÉNEZ M. J., ZISSERMAN A., FERRARI V.: 2d articulated human pose estimation software, 2011. 3

[Mot12] MotionBank, 2012. http://www.motionbank.org/. 2

[Sha11] SHAW N. Z.: Synchronous objects, choreographic objects, and the translation of dancing ideas. In *Emerging Bodies - The Performance of Worldmaking in Dance and Choreography*, Klein G., Noeth S., (Eds.). transcript, 2011, pp. 207–224. 1

[Syn12] Synchronous objects for one flat thing, reproduced, 2012. http://synchronousobjects.osu.edu/. 1