"THE RIGHT MOVE" – A CONCEPT FOR A VIDEO-BASED CHOREOGRAPHY TOOL

S. Zillner [a], M. Gelautz [b], M. Kallinger [b]

[a] Institute for Computer Science and Business Informatics, University of Vienna,
Liebiggasse 4, A-1010 Vienna, Austria - sonja.zillner@univie.ac.at
[b] Institute for Software Technology and Interactive Systems, Vienna University of Technology,
Favoritenstrasse 9-11/188/2, A-1040 Vienna, Austria - gelautz@ims.tuwien.ac.at

KEY WORDS: Video, Multimedia, Biometrics, Human Motion, Choreography, Dance

ABSTRACT:

In the context of a proposed project for a video-based choreography tool, we carry out video acquisition experiments with a Sony DCR-VX2000 camcorder that operates at 25 frames/sec and a high-speed HCC1000 video camera with adjustable frame rates. We perform tests with frame rates up to 462 frames/sec and observe that the performance of the Kanade-Lucas-Tomasi tracking algorithm improves significantly as the frame rate increases. Furthermore, we demonstrate the application of an available dance notation, the Labanotation, to the video-captured ballet scene.

1. INTRODUCTION AND MOTIVATION

Capturing human shape and motion from image sequences constitutes an important topic in current computer vision research. The recognition, analysis, and realistic modelling of people from videos is required in a wide range of applications, including man-machine interfaces, the creation of human avatars, medical and sports analysis, or safety applications.

The proposed project focuses on complex movements of the whole body as presented in the field of classical dance (ballet). The subtleties of the human dance movement pose special challenges during the motion capture and analysis procedure which are not encountered in studies that concentrate on, e.g., human walking. The development of a tool that supports the (semi-)automatic recording and annotation of dance sequences would be of great interest to choreographers and dance companies in order to achieve a reduction in time and cost of rehearsals. This is particularly important if an existing repertoire is going to be put on stage again with a different choreographer and/or new dancers. Beyond that, an archive of recorded dance scenes along with extracted 3D position and motion data can be utilized later by the choreographer as a computer-based creative tool as well as for animation purposes.

Video-based motion capture for dance applications faces a series of challenging questions such as

- What are the characteristic features of classical dance movements and which of them can be extracted from the video data?

- What would a suitable video acquisition set-up look like (e.g., number and type of cameras, spatial arrangement, use of markers, etc.)?

- Is it possible to classify automatically the recorded video scenes (e.g., to recognize all pirouettes?)

In order to describe the characteristics of a ballet sequence, we will make use of available movement/dance notations (e.g., Labanotation), whose design goal was to provide a concise description of the human body movement. These dance notations may, for example, be exploited in order to determine good marker positions on the human body. Section 3 describes the Labanotation (Hutchinson 1991) - named after the Austrian dancer and choreographer Rudolph Laban (1879-1958) - which is nowadays routinely utilized by choreographers and dance notators in order to annotate human body movement. In section 4, we report tests with different video cameras and examine the effect of varying frame rates on the performance of an optical tracking algorithm.

2. PREVIOUS WORK

A survey of different computer-vision techniques for human motion analysis can be found in, e.g., (Gavrila 1999). Most related research has been performed in the context of modelling human movement for realistic computer animation (Liu et al. 1999, Plaenkers and Fua 2001). Only few publications have addressed the field of dance. For example, (Suzuki et al. 2000) aim to develop an interactive dance system that takes into account the emotional state of a reference dancer by analyzing his/her body motion. (Abouafa/b 1999) deals with motion capture techniques in order to set up a dance performance in which real artists interact with virtual dancers. (Moltenbrey 2001) describes the use of an optical motion capture system to record the movements of famous pantomime Marcel Marceau in order to preserve the artist's legacy for future generations.

A relatively new area of application is the development of human-computer interfaces which take into account the human body language. A related experiment is reported by (Nakata et al. 1998), who propose to express a robot's emotion by its posture and dance. The work is based on Laban's theory on motion analysis and description. A system for generating Labanotation from motion captured data is proposed by (Matsumoto et al. 2001).

3. LABANOTATION

The approach of Labanotation (Hutchinson 1991) is based on a clearly constructed grammar and syntax of movements, which defines the relationship of movements. The base elements of this grammar of movements fall into the categories of nouns, verbs and adverbs. In the context of only one person moving, the nouns are the specific parts of the body, i.e. the right leg, the left arm, the head, etc. Only the moving body parts are mentioned. For example, the movement of walking is performed alternately by the left and right leg. Actions, which result in movements, are described by verbs. For instance, the characteristic actions of the movement of walking are its forward direction and its shifting support from one to the other leg. Adverbs specify the manner of performance, the timing, and the dynamics.

Figures 1a-c illustrate the grammar of Labanotation when describing the classical ballet jump, called Grand jeté. For this, the jump is divided into three phases - the preparation, the air, and the landing phase. The left side of the figures illustrates the appropriate phase of the jump, which is described by Labanotation symbols on the right side. Each of the Labanotation symbols describes the direction of the movement, the part of body doing the movement, the level of the movement, and the length of time it takes to do the movement. For example, the length of a movement is determined by its symbol's length, and the level of a movement is determined by its symbol's shading.

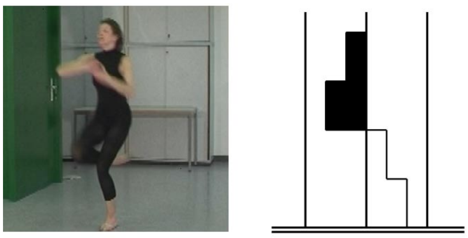

Figure 1a. The preparation phase of a Grand jeté in Labanotation.

In the preparation phase of the Grand jeté, the right leg steps forward. This is followed by the left leg stepping forward in low level, i.e. with bent knees.

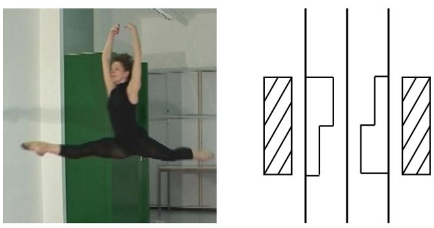

Figure 1b. The air phase of a Grand jeté in Labanotation.

This continues with the air phase, the time during which the body is without support. The right and left arm are directed high, the right leg forward and middle (horizontal), and the left leg backward middle (horizontal).

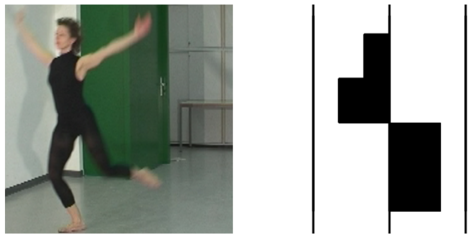

Figure 1c. The landing phase of a Grand jeté in Labanotation.

In the landing phase, the right leg goes down, i.e. the right leg is the body part which supports the weight of the body. The right leg bends in the landing. The action has no directional destination. Shortly after this, the support is shifted onto the left leg, which is moving forward.

Thus, the Labanotation provides an objective language for describing movements. On the basis of such a language, the characteristic features of classical dance movements can be described.

4. EXPERIMENTS

4.1 Video Acquisition

We captured videos of one and the same ballet scene with different video cameras and frame rates that varied from 25 frames/sec up to approximately 500 frames/sec. The goal of our experiment was to better understand which requirements on camera speed and quality are imposed by the dance application. The video shown in Figure 2 was acquired with a SONY DCR-VX2000 colour camcorder with 25 frames/sec (50 half-frames/sec). Eight consecutive frames of the Grand jeté movement are displayed. One can clearly recognize the motion blur that arises due to the fast movements of various body parts (e.g., the right foot in frame 3). This indicates that the frame rate of conventional video cameras does not suffice to capture the high-speed dance motion with sufficient quality for detailed motion analysis. Also, the pronounced differences in body posture between adjacent video frames suggest that tracking different body parts throughout the video sequence will be difficult to accomplish at this frame rate.

A further set of test data was acquired with the VOSSKÜHLER HCC1000, a monochrome high-speed camera with adjustable frame rates, which is distributed by MIKROMAK (Mikromak 2002). We increased the frame rate in several steps, varying from 25 frames/sec up to 462 frames/sec. Figure 3 shows a comparison between corresponding frames captured by the SONY camcorder at 25 frames/sec (a) and the high-speed HCC1000 at 462 frames/sec (b). The acquisition viewpoint was slightly different due to different camera positions. One can clearly recognize the reduction in motion blur that was achieved by the high-speed shutter of the HCC1000 camera (see, e.g., the region marked by a circle).

Figure 2. Grand jeté captured with a Sony camcorder at 25 frames/sec. Subsequent frames are displayed from top left to right bottom.

Figure 3. Comparison of motion blur at different frame rates: (a) SONY DCR-VX2000 camcorder at 25 frames/sec, and (b) Mikromak HCC1000 high-speed camera at 462 frames/sec.

4.2 Optical Tracking

For a quantitative analysis of the effects of different frame rates, we applied a feature-based optical tracking algorithm to the high-speed video sequence. We employed the classical Kanade-Lucas-Tomasi (KLT) feature tracker as described by (Shi and Tomasi 1994).

We utilized the original video sequence that was acquired with 462 frames/sec, and then simulated lower frame rates by selecting suitable subsets from it (e.g., every forth frame to reduce the frame rate to 116 frames/sec). The synthesized lower frame rates do not account for the additional motion blur due to the changes in shutter speed. However, we preferred this approach to using the available (real) video sequences that had been acquired of the Grand jeté movement at lower frame rates, because we wanted to exclude the natural variations in the dancer's movements between different video acquisitions from our analysis.

Figure 4 shows the number of tracked points as a function of the frame number. The different curves correspond to frame rates of 462, 116, 58, and 25 frames/sec. The tracking features ("points of interest") in the first frame are selected automatically by the algorithm. Our analysis concentrates on the moving objects in the scene (i.e., the dancer) and discards tracking features in the (static) background. The diagram clearly demonstrates how the performance of the tracking algorithm improves as the frame rate increases.

Figure 4 shows the results obtained from the KLT tracker without any further evaluation of the quality of the tracking results. We performed an additional test in which we manually post-processed the tracking results by classifying the tracked points as either "good" or "bad". Figure 5 shows the diagram that was obtained when only the successfully tracked points – as determined by a human operator - were taken into account. The differences between the frame rates are now smaller, but still present. Again, one can clearly recognize the lower performance of the tracker at conventional video rates, even without the additional motion blur that would be present in real video imagery acquired at this frame rate.

An illustration of the tracking results obtained by using 116 frames/sec is shown in the colour representation of Figure 6. The first and last frame of the tracking sequence are displayed in Figures 6 (a) and (b), respectively, with the tracking points delivered by the KLT tracker overlaid. The features of interest in the background, which were not included in the analysis, are marked in red. The blue points were rejected during the post-processing due to their low quality. The tracking results which passed the final quality control are marked in green. One can recognize that the majority of the successfully tracked points are located on the head and torso, i.e., those parts of the body which remain relatively stable during the jump. However, some points on the arms and legs are correctly tracked, too, despite their strong dynamics.

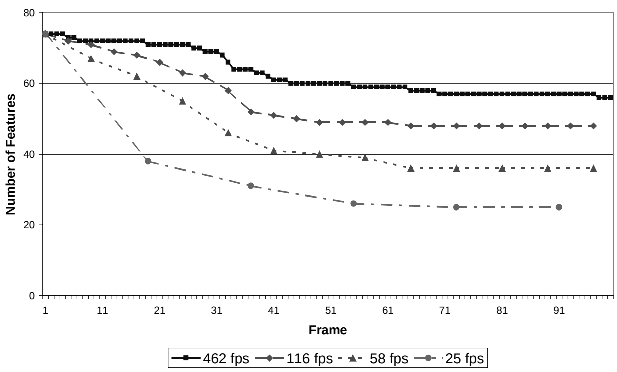

Figure 4. Tracking results obtained at different frame rates.

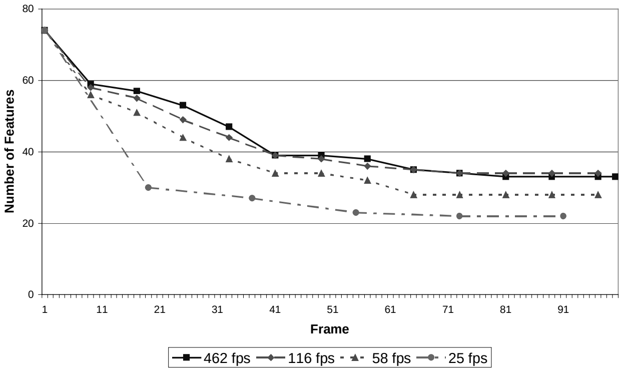

Figure 5. Tracking results obtained at different frame rates after post-processing.

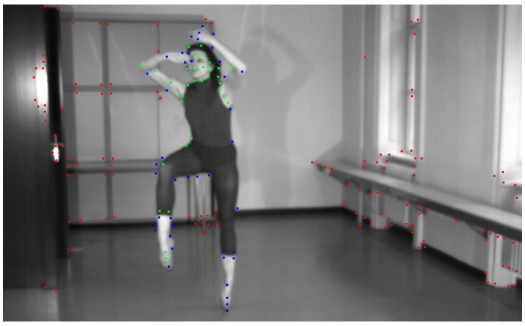

Figure 6a. First frame of the sequence (116 frames/sec) with tracking points overlaid.

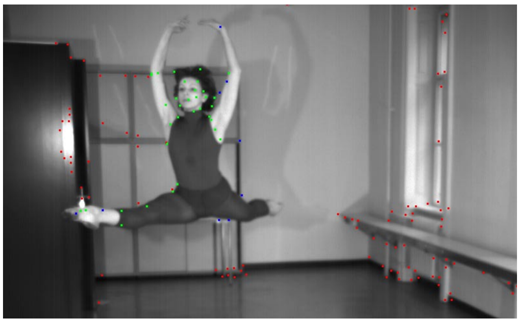

Figure 6b. Last frame of the sequence (116 frames/sec) with tracking points overlaid.

5. SUMMARY AND OUTLOOK

Our video acquisition experiments with varying frame rates indicate that a conventional frame rate of 25 frames/sec does not suffice to capture the high-speed ballet motion with sufficient quality for subsequent motion analysis. We performed tracking experiments with image sequences that were derived from the high-speed HCC1000 camera with frame rates up to 462 frames/sec and found that the performance of the Kanade-Lucas-Tomasi feature tracker improved significantly at higher frame rates.

As a next step in the proposed project, we plan to perform tests with different types of body markers in order to better understand the specific requirements of the complex ballet movements. We are particularly interested in the question whether the choreographic knowledge inherent in the Labanotation can provide useful hints on optimal marker positioning.

6. ACKNOWLEDGEMENTS

We thank Pascale Chevroton for dancing the Grand jeté.

7. REFERENCES

Abouaf, A., 1999a. Biped: A dance with virtual and company dancers, Part 1. *IEEE Multimedia*, 6(3), pp. 4-7.

Abouaf, A., 1999b. Biped: A dance with virtual and company dancers, Part 2. *IEEE Multimedia*, 6(4), pp. 5-7.

Gavrila, D., 1999. The visual analysis of human movement. *Computer Vision and Image Understanding*, 73(1), pp. 82-98.

Liu, X., Zhuang, Y., and Pan, Y., 1999. Video based human animation technique, *Proc. ACM Multimedia'99*, Orlando, Florida, November 1999.

Matsumoto, T., Hatchimura, K., and Nakamura, M. 2001. Generating Labanotation from motion-captured human body motion data. *Proc. International Workshop on Recreating the Past – Visualization and Animation of Cultural Heritage (ISPRS WG V/6)*, Ayutthaya, Thailand, pp. 118-123.

Mikromak, 2002. Mikromak GmbH, Erlangen, Germany. http://www.mikromak.com (accessed 10 Jul. 2002).

Moltenbrey, K., 2001. All mime, *Computer Graphics World*, 24(8), pp. 52-54.

Nakata, T., Sato, T., and Mori, T., Expression of emotion and intention by robot body movement. *Int. Conf. Intelligent Autonomous Systems 5 (IAS-5)*, June 1998, pp. 352-359.

Plaenkers, R., and Fua, C., 2001. Tracking and modelling people in video sequences. *International Journal of Computer Vision and Image Understanding*, 81(3), pp. 285-302.

Shi, J., and Tomasi, C., 1994. Good features to track. *IEEE Conference on Computer Vision and Pattern Recognition*, Jerusalem, Israel, pp. 593-600.

Shiphorst, T., 1997. Life Forms – Merce Cunningham: Making dance with the computer. *Choreography and Dance*, 4(3), pp. 79-98.

Suzuki, R., Iwadate, Y., Inoue, M., and Woo, W., 2000. MIDAS: MIC interactive dance system, *Proc. IEEE SMC 2000*, pp. 751-756.