J Multimodal User Interfaces (2010) 4: 37–46
DOI 10.1007/s12193-010-0049-x

ORIGINAL PAPER

Browsing a dance video collection: dance analysis and interface design

Damien Tardieu · Xavier Siebert · Barbara Mazzarino · Ricardo Chessini · Julien Dubois · Stéphane Dupont · Giovanna Varni · Alexandra Visentin

Received: 26 February 2010 / Accepted: 18 August 2010 / Published online: 17 September 2010

Abstract In this article we present a system for content-based browsing of a dance video database. A set of features describing dance is proposed, to quantify local gestures of the dancer as well as global stage usage. These features are used to compute similarities between recorded dance improvisations, which in turn serve to guide the visual exploration in the browsing methods presented here. The software integrating all these components is part of an interactive touch-screen installation, and is also accessible online in association with an artistic project. The different components of this browsing system are presented in this paper.

Keywords Dance · Database · Video · Motion capture · Gesture analysis · Data visualization · Interactive installation · Browsing interface · Touchscreen

[1] www.dancersproject.com.

D. Tardieu (✉) · J. Dubois · S. Dupont
TCTS Lab., University of Mons, 31 Boulevard Dolez, 7000 Mons, Belgium
e-mail: Damien.Tardieu@gmail.com

X. Siebert
MathRO Lab., University of Mons, 9 rue de Houdain, 7000 Mons, Belgium

B. Mazzarino · G. Varni · A. Visentin
infoMus Lab—DIST, University of Genova, Viale Causa 13, 16145 Genova, Italy

R. Chessini
SEMI Lab., University of Mons, 31 Boulevard Dolez, 7000 Mons, Belgium

1 Introduction

This article describes a software system for content-based browsing of a database of short dance videos that has been gathered for the "DANCERS!"[1] project. Designed by the choreographer Bud Blumenthal, this project aims at giving dancers the opportunity to demonstrate their art without having to adapt to the vision of a choreographer. It results into a large video collection of dancers' improvisations that constitutes an overview of what can be dance today.

This collection can be accessed in two complementary ways: an interactive installation that will be touring worldwide, and a website. In each case the videos are displayed either as a list or as a map that is organized according to the similarity between the dances. The former is fairly traditional and will not be described here, but it can be viewed online.[2] The latter provides an innovative way to look at dance performances and is the main focus of this article. The specific aims of this work are as follows. First, features that are representative of the performance have to be automatically extracted from the videos. Second, an interface has to be conceived to allow features-based navigation through the video collection. This involves the design of a visual interface and of user interaction scenarios, which was achieved in collaboration with graphic designers and with the choreographer. Third, algorithms and technical solutions have to be implemented so that the features are used to organize the screen display.

The structure of this article follows the hierarchy of the three above-mentioned aims. We start by a brief survey of existing methods for visualization and browsing of databases (Sect. 2). This defines requirements that our system should fulfill. Then, in Sect. 3 we present the collection of

[2] www.dancersproject.com/browse/.

dance videos, followed by a description of the features that we used to analyze the dance videos (Sect. 4). Section 5 presents the details of the new interface we developed, and emphasizes the links between our design and the requirements pointed out in Sect. 2. Technical details about the software are grouped in Sect. 6. Finally, Sect. 7 describes the installation and the website that were built with our system.

2 State of the art on visualization and browsing

Query-By-Content (QBC) software packages (reviewed among others in [1]) are designed to retrieve items (e.g. sound, images, videos, text, ...) based on an analysis of their content. Typically, visualization is used by these software packages only to display the results of the queries in an array, which makes it difficult for the user to quickly obtain an overview of the results. However, visualization should be considered not only as a way to show the results, but also as an integral part of the software, as a means to guide the user [2]. The QBC paradigm has thus gradually been replaced by a more active exploration process, that aims at providing a context around each item to address the so-called *semantic gap* (e.g., [3, 4]).

Several papers and reports on state-of-the art visualization methods (see [5, 6] and references therein) attempt to define guidelines that a generic database exploration system has to follow, and that we kept in mind when designing ours, as emphasized in Sect. 5. First, it should give a faithful *overview* of the database distribution. Second, it should preserve the *structural relationships* among data, for example in the form of distances or clusters. Third, it should provide a visualization that is *understandable by the user*. One could also add that it should be *scalable*, which refers to the ability of the system to manage sizable databases. Several data structures have been devised (e.g. hierarchies, networks, ...), to organize the database so that retrieval or browsing is faster than with sequential access, as reviewed in [7].

The fundamental reason why the above-mentioned requirements are difficult to fulfill is that a multi-dimensional space has to be projected on a two-dimensional screen, moreover of limited size. Indeed, media analysis transforms each item of the database into a M-dimensional vector of features with M typically ranging from 5 to 100. The M-dimensional space (*information space*) has to be somehow projected onto the display (*visualization space*), which is bi-dimensional in our case. Several methods have been used to achieve this, either linear (e.g., Principal Component Analysis (PCA) [8], Multi-Dimensional Scaling (MDS) [9]) or non-linear, the latter being obviously most adapted when the *information space* has a non-linear structure. A related issue is that not all features are equally useful for measuring similarity [7] and features combination should be adapted to the user and the task [10].

The *visualization space* is in fact limited not only in the number of dimensions, but also in its size (e.g., a computer screen). Therefore, only a subset of the database can be shown simultaneously to avoid overloading the display both from a computational and from a cognitive point of view. One common method involves clustering the data and selecting one or several items representative of each cluster, so as to minimize overlapping. Interestingly, recent work has shown how to formulate these requirements as an optimization problem [5].

Browsing is an issue closely related to the visualization. Indeed, the user often starts the data exploration process with a vague idea that gets refined along the browsing, as a result of the interaction with the database [7]. For an effective data exploration, it is thus necessary to adapt the visualization according to the user's actions. Three common techniques are used in the literature and in our software. First, the projections can be changed dynamically to present other dimensions corresponding to other features. Second, the subset of data displayed on the screen can vary, by zooming in/out a specific region of the database, or by switching between different regions. Third, distortion techniques can also be used to show some portions of the data with a high level of details while other remain at a low level of details. Other techniques are reviewed in [2].

3 Dance video corpus

3.1 Setup

The corpus contains short videos of dancers' performances, that are filmed in a calibrated setup. Each dancer performs a two-minutes improvisation in a limited space of the stage (colored in red on Fig. 1), either in silence or with a piece of music. Two videos are recorded: one from the front (labeled "EX1" on Fig. 1) and one from the top. For technical reasons, the top camera (wide angle) was placed about 6 meters from the floor, and the sides of the stage are slightly cut on the top-camera video recording (see Fig. 2, right). The front videos are recorded in High Definition (HD), 25 frames per second, interlaced, using the *XDCAM-EX* codec.

At the time of writing, recording sessions already occurred in Brussels and Paris, leading to a set of about 280 videos (two trials per dancer), and more recordings are planned in the future.

3.2 Procedure

Every detail of the procedure should be the same in each recording. Dancers can enroll in the project through the website, without any pre-selection. After welcoming and debriefing, dancers go to the warm-up room where they select the music for their performance among a given set on a computer. Meanwhile a television screen allows them to watch the video of the dancer currently being recorded on stage. By groups of five, dancers go into the shooting space to get comfortable with the environment and procedures. The choreographer explains some important principles concerning the space and the camera frame, the way of entering, how to know when the 2 minutes are up, etc.... To start the performance, the dancer enters the space at “real size center” of the image. One after the other, the five dancers do their first solo and then their second. Returning to the warm-up room, the dancer views both solos on a computer and then chooses which solo would participate in the project.

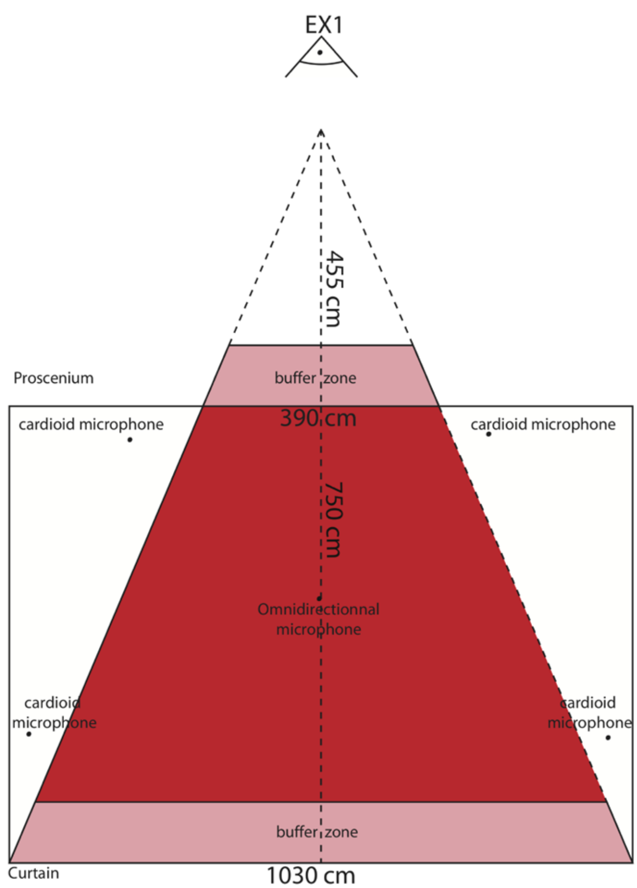

Fig. 1 Top view of the recording stage setup. The front camera is labeled “EX1”. The dancer can occupy the *red area*, delimited by buffer zones in *pink*. Microphones are placed around and above the stage to record

4 Dance motion description

4.1 Introduction

No sensors were embedded on the dancers to avoid hampering their movements, so that our dance motion description relies on raw video analysis. Using background and color subtraction techniques, we obtained the silhouette of the dancers, from which we extracted motion features. *Time-dependent* features are first computed on a frame by frame basis, and are then summarized into a small set of features (see *features analysis* below) for a concise description of the movements and the performance. This set has been defined in collaboration with the choreographer Bud Blumenthal. Time-dependent features were extracted using the EyesWeb XMI software platform (in particular its Gesture Processing Library [11]). This library follows a multilayer framework for analysis of expressive gesture in human full-body movement and, in particular, in dance performance [12]. The various steps for feature extraction are described below.

4.2 Time-dependent features

In the following sections, we make a distinction between the *General Space* and the *Personal Space*.

4.2.1 General space features

General Space features describe how dancers use the stage during the performance. For this analysis the dancer’s movement can be approximated by the movement of the center of gravity (CoG) of the dancer. We extracted the following features from each frame recorded by the top camera:

– position (x, y) of the CoG, normalized with respect to the dimension of the image;

Fig. 2 Snapshot of a front and top videos (dancer: Claire O’Neil)

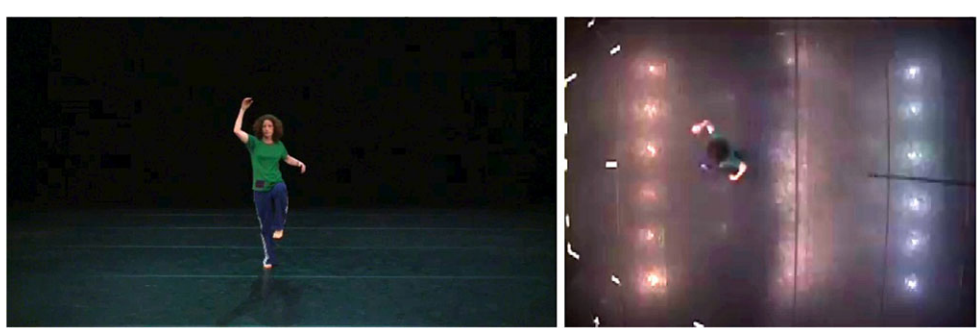

Table 1 Formulas, names and adjectives associated to the features currently used in our system. $E(.)$ is the average and $\sigma(.)$ the standard deviation. *CI* is the contraction index, *QoM* the quantity of motion, and *BBR* the bounding bow ratio, all defined in the text

Formula	Name	Adjectives
$E(v_x, v_y)$	Speed	low, middle, high
$\sigma(x, y)$	Stage use	less, middle, much
$E(x)$	Proximity	close, middle, far
$E(CI)$	Body expansion	small, medium, large
$E(QoM)$	Energy	low, medium, high
$E(BBR)$	Level	floorwork, middle, upright

– the velocity of the CoG along the x and y axes, calculated using a centered derivative based on the position at 3 consecutive frames;
– the acceleration of the CoG along the x and y axes, calculated the same way.

4.2.2 *Personal space features*

Personal Space features describe the movement inside the "Kinesphere", concept first introduced by Laban [13]:

> "Whenever the body moves or stands, it is surrounded by space. Around the body is the sphere of movement, or Kinesphere, the circumference of which can be reached by normally extended limbs without changing one's stance, that is, the place of support. The imaginary inner wall of this sphere can be touched by hands and feet, and all points of it can be reached. (...) Thus, in actual fact, he never goes outside his personal sphere of movement, but carries it around with him like a shell."

We focused on the following measurements, that were implemented in [12] based on theories from the fields of dance, cognitive science and psychology [14, 15] and available in the EyesWeb Gesture Processing Library [11].

– the *Quantity of Motion* (QoM) is an estimation of the amount of overall movement (variation of pixels) the video-camera detects [12].
– the *Contraction Index* (CI) provides information on the spatial occupation of the kinesphere by the body. More specifically, it measures the contraction/expansion of the dancer's body with respect to its centre of gravity and can be calculated as the ratio between the area of the silhouette and the area of its bounding box.
– the *Bounding Box Ratio* (BBR) of dimensions gives additional information about the dancer's posture. For example, the verticality of the dancer (standing vs. lying) can be related to the ratio between the width and the height of the bounding box.

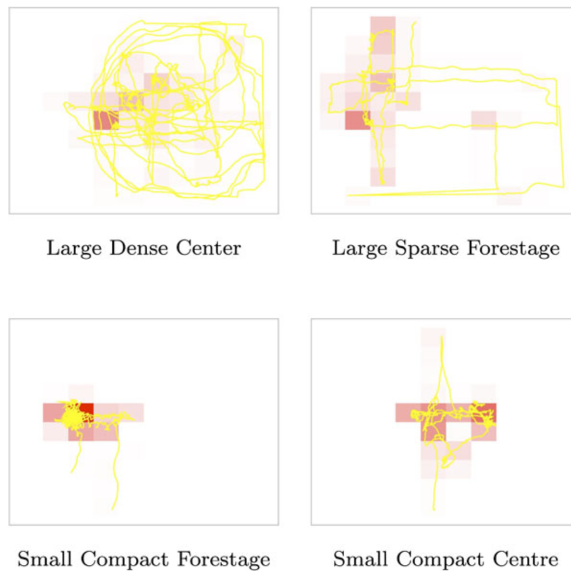

Fig. 3 Results of the General Space Occupation for four different dancers. The *yellow line* is the trajectory of the Center of Gravity of the dancer, seen from the top. The space is divided into cells, colored in *red* (the darker the *red*, the most occupied the cell). The front camera is located on the left (not shown)

4.3 Features analysis

The time-dependent (*General Space* and *Personal Space*) features represent a huge amount of data and need to be summarized to a small set of relevant features that could be further linked to a semantic description of the performance (see Sect. 5.4).

The time-dependent features are first summarized by calculating their mean and standard deviation, which provides meaningful information but fails to describe the temporal evolution of the choreography. Indeed, the average position on stage with respect to the audience ("proximity" in Table 1) contains information about some of dancer's intentions. For example some dancers develop their performance in the back of the stage, while others come very close to the camera. Similarly, the standard deviation of their trajectory relates to the amount of space covered by the dancer ("stage occupation" in Table 1). For instance, some dancers explore

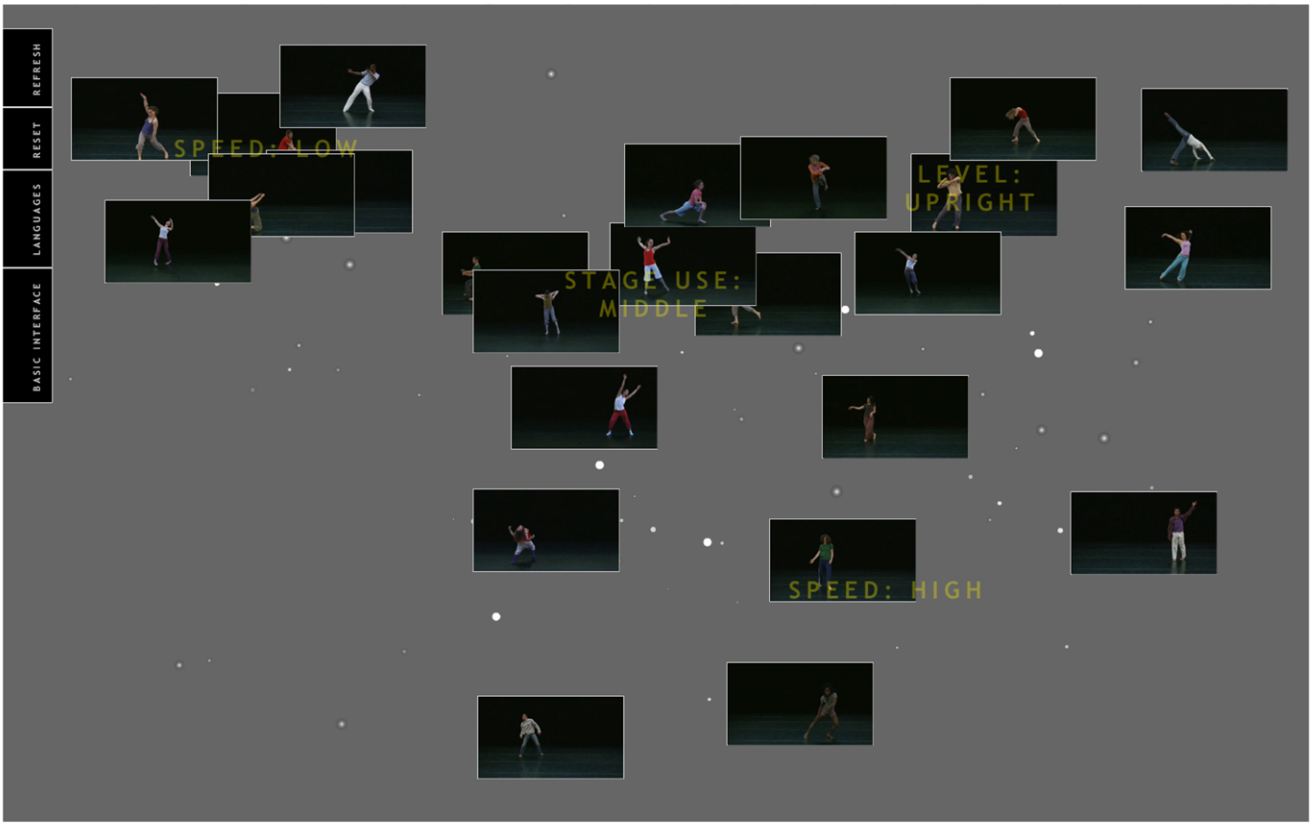

Fig. 4 A screenshot of the exploration display

the limits of the stage to cover as much as possible the red space defined in Fig. 1.

Nevertheless, the mean and standard deviation do not provide information about the temporal evolution of the performance. To overcome this limitation we studied stage occupation using the Occupation Rate concept [16]. The image was divided in a 10 by 10 grid (corresponding to 100 cells of approximately 80×70 cm each) to compute a space-occupation map (i.e. a two-dimensional histogram) representing the percentage of time that the dancer spends in each cell. From this map we extract the preferred zone used in the choreography, which we describe in terms of forestage, centre or rear of the stage on Fig. 3.

We also studied the compactness (or its opposite, the sparsity [17]) of the trajectory by converting the trajectory to a single image and calculating its contraction index as defined above. This differentiates two trajectories of similar width, one where the dancers remains in the periphery and the other where the dancers go through the middle of the trajectory, as illustrated on Fig. 3. However, discussions with the choreographer revealed the difficulty in transforming this compactness measure into a word that would be understandable for a general audience. It is therefore currently not yet included in our software interface.

5 User interface design

5.1 Design overview

The users of the software (either in the installation or online) are not necessarily dance specialists. Therefore, the software has to allow both the discovery of the various aspects of the dance performances and the focus on a particular dancer. In collaboration with two graphic designers and with the choreographer of the DANCERS! project, we elaborated a navigation scenario that led us to the design of two displays. The first one called *exploration display* provides explicit clues to help the user understand the global feature-based organization of the database. The second one, called *similarity display*, allows the user who became interested in a particular dancer to explore similar ones, which is likely to occur in later stages of the navigation process. The *exploratory display* (Fig. 4) appears in the first window when the software is launched or when the user clicks on the *reset* button. It gives an initial point of view on the collection and shows clusters of videos, following the first guideline mentioned in Sect. 2 (*overview*). These clusters are associated with a label emphasizing one prominent aspect of each cluster (e.g., “Speed:low” or “Stage use:middle” on Fig. 4). The need for labels was revealed by preliminary user tests showing that the features-based clustering alone is not intuitive enough. This concern is related to the third guideline mentioned in Sect. 2 (*understandable to the user*). More precisely, it is difficult just by looking at the thumbnails to guess which features are similar in the clusters. In our interface, the static snapshots are thus replaced by a series of snapshots of the performance, and text labels are added to characterize one of the most salient features of each cluster.

When the user clicks on the button "More Related Videos" the *similarity display* is launched, whose goal is to show simultaneously three features by which dancers can be similar, to emphasize the fact that dancers can be very similar on one feature but different on others.

Interestingly, the discussion with the choreographer and the designers led to a set of requirements that closely resemble the guidelines listed in Sect. 2. An additional practical requirement concerns user interaction. In the interactive installation, the software is displayed on a touch-screen, but it is also accessible online through a conventional web browser. In either case, the user's actions are quite limited: a single touch in the first case, and a click or a hover in the second. The user interface has been designed with these limitations in mind. In the following, we describe each display in terms of a three-steps procedure: selecting the features, the videos and the labels to display. These steps are related to the second guideline mentioned in Sect. 2 (*structural relationships*).

5.2 Exploration display

5.2.1 Features selection and projection

To encourage exploration of the database all computed features should be accessible for browsing. However, it is not efficient to show all of them simultaneously, because it becomes confusing for the user. We thus decided to show three features at any time, while giving the opportunity to change the selection of those three features. Only these three features are used for subsequent computations. To project the data on the 2D screen, we perform a principal components analysis (PCA) using the selected features and keep the first two principal components, which usually explain 90% of the variance. The features are previously centered and normalized so that the PCA is done only on correlations.

5.2.2 Videos selection

Given the screen dimensions and the thumbnails dimensions only a subset of the video collection can be displayed simultaneously. In the following, M is the size of this subset and K the number of clusters into which they are grouped. A two-level top-down hierarchical k-means clustering is performed: the first one to select the K clusters and the second one to select the M/K representatives of each group. To obtain more homogeneous clusters in the first level (i.e., small intra-cluster variance), the collection can be partitioned into $K' > K$ clusters, out of which K are randomly selected. The representative videos are then randomly selected in the clusters of the second level. This method allows the selection of videos that are not too close to each other in the feature space while avoiding some video being selected more often than others. In our settings $K = 4$, $K' = 8$ and $M = 25$.

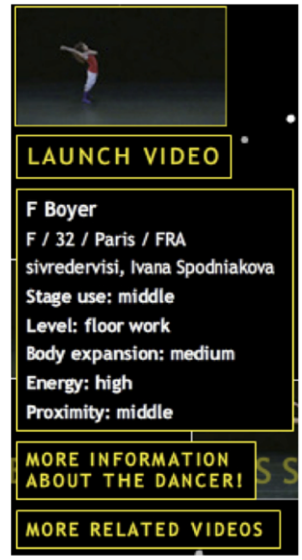

Fig. 5 The box under the thumbnail when it is clicked

5.2.3 Labels selection

The labels and their position on the screen have to be chosen to help the user understand the meaning of the features. To present a variety of features without overloading the display, we decided to show four labels. Two labels are associated to the same feature (e.g., "speed" on Fig. 4) but with different values (e.g., "low" and "high" on Fig. 4). The other two labels correspond respectively to two other features with any value (e.g., "Stage use:middle" and "level:upright"). The former two are chosen to be the two most distant clusters on the screen and are labeled with the name of the most discriminative feature. The latter two clusters are labeled using the two remaining features. The choice of the displayed adjective for each label is described in Sect. 5.4.

5.2.4 Navigation

Upon clicking on any thumbnail, a box appears below it (see Fig. 5) with additional information on the video such as the name of the dancer and the textual value of some features. Three buttons are also displayed. The first one ("launch video") allows to watch the full-size video. The second one ("more information about the dancer!") is a link to the

Fig. 6 A screenshot of the similarity display

dancer's personal page with the full-sized video and additional information. The third one ("more related videos") allows to find similar dancers using content-based similarity. When this third button is clicked, the browsing switches to the *similarity display* described in the next section.

5.3 Similarity display

5.3.1 Features selection and projection

The selection of the three displayed features among all the computed ones is the same as for the *exploratory display*, but the projection method differs. The space is divided into three angular areas labeled according to the predominant feature (e.g., "level", "proximity", "stage use" on Fig. 6). The distance from the target (Eq. 2) represents the overall degree of similarity while the angular position (Eq. 3) of a video depends on the relative importance of the three features in the similarity. Let $\mathbf{f_i}$ be the coordinates vector of any video i in the three-dimensional features space. The distance vector $\mathbf{d_i} = (d_{ij})_{j\in\{1\ldots3\}}$ between any video i and the target t (of coordinates $\mathbf{f_t}$) for each feature is given by the absolute value of the element-wise difference:

$$\mathbf{d_i} = |\mathbf{f_i} - \mathbf{f_t}| \tag{1}$$

and is thus embedded in the fist octant of the 3D space. We project this octant on the 2D plane of the screen and express the position of the video i in polar coordinates (r_i, θ_i). The radius r is equal to the Euclidean distance:

$$r_i = \|\mathbf{f_i} - \mathbf{f_t}\| \tag{2}$$

To compute the angle, we first convert the distance measure into a similarity measure:

$$\mathbf{s_i} = \mathbf{1} - \frac{\mathbf{d_i}}{\sum_{j=1}^{3} \mathbf{d_{ij}}} \tag{3}$$

Then each feature is assigned to a direction $\mathbf{v_j}$ in the 2D plane such that the plane is divided into three equal areas. For example, in Fig. 6, the directions are $v_1 = (0, 1)$, $v_2 = (-\frac{\sqrt{3}}{2}, -\frac{1}{2})$, $v_3 = (\frac{\sqrt{3}}{2}, -\frac{1}{2})$. Using $\mathbf{v_j}$, the angle is finally calculated by:

$$\theta_i = \angle \left(\frac{\sum_{j\in M_i} s_{i,j} \mathbf{v_j}}{\sum_{j\in M_i} s_{i,j}} \right) \tag{4}$$

where M_i represents the set of the two features with the highest similarity for point i. Note that in the 2D space it is impossible to show that all three features are important, but a dancer that is close to the target on all three features

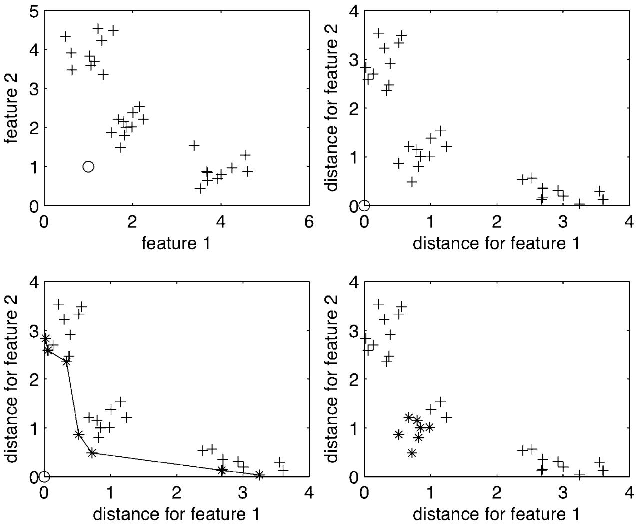

Fig. 7 Example of Pareto rank selection. Target is a *circle*, selected items are *stars*. *Top-left*: items in the feature space. *Top-right*: items in the criteria space. *Bottom left*: selection using Pareto rank. *Bottom-right*: selection using Euclidean distance

will be very close to the center, thus giving visually less importance to the angle.

5.3.2 Videos selection

To choose the video to display, we need a definition of similarity that takes into account the three selected features. The usual measures of similarity (for instance on images [10]) are based on Euclidean distance or cosine distance with equal and fixed weights. The underlying assumption is that all features are equally important for the perceived similarity, thereby discarding content that is very similar for only a subset of features. However, in this application, it is important to be able to identify dancers that are very similar for only some criteria. We therefore introduce another measure of similarity based on the concept of "Pareto ranking" borrowed from multi-criteria optimization [18]. In short, a solution c of a multi-criteria optimization problem is said to be "non dominated" if there is no solution c' at least as good on all criteria and strictly better for at least one criteria. The set of non-dominated solutions is called the *Pareto front* and the solutions belonging to the front have a *Pareto rank* of 1. If we remove all the solutions of the Pareto front from the collection, the new set of non-dominated solutions have a rank of 2 and so on. In terms of similarity, a solution belonging to the Pareto front is the closest to the target for a given set of feature weights. Figure 7 shows an example of Pareto front. In the top left figure the data are represented in the features space. In the top right figure, the same data are plotted in the feature distance space (obtained by Eq. 1), that we call criteria space. The bottom figures show a comparison between the closest point according to the Pareto front on the left and to the Euclidean distance on the right. The important point is that Euclidean distance completely ignores the two clusters that are very close for only one feature, whereas Pareto ranking does not. In this work we are using a mixture of Pareto rank and of Euclidean distance to retrieve videos that are close for all the features and videos that are close considering a subset of features. Formally, out of the collection $\mathcal{C}$ we select the videos c whose rank relative to the target is smaller than $\bar{R}$, where $\bar{R}$ is defined as follows:

$$\bar{R} = \min R \quad \text{s.t.} \quad |\{c \in \mathcal{C} | rank(c) \leq R\}| \geq M$$

In this subset the M closest videos according to the Euclidean distance with equal weights are selected.

5.3.3 Labels

The selection of the labels is simpler than for the *exploratory display* since they are directly associated to the three selected features. On the display they are preceded by the word "similar".

5.4 Feature names and adjectives

For both the *exploratory* and the *similarity display* the names of the dancer's features have to be easily understood by the user. The choice of these names (see Table 1) was made with the choreographer so that the words are simple but also do not carry any judgment about the dancer. For example, to describe the mean position with respect to the camera, we used "proximity" instead of "presence". Indeed, while the latter belongs more to a choreographer's vocabulary, a "low presence" bears a negative connotation. We also have to translate the numerical value of the feature into an adjective. We have chosen to divide the feature range into three parts in the following way. Considering a normalized feature $f \sim \mathcal{N}(0, 1)$,

$$l = \begin{cases} 0 & \text{if } f < -0.5 \\ 1 & \text{if } f > -0.5 \text{ and } f < 0.5 \\ 2 & \text{if } f > 0.5 \end{cases} \quad (5)$$

Then l is translated into an adjective using Table 1. That way, about one third of the dancers fall in each category.

6 System implementation

The system for analysis and navigation is mainly composed of two parts. First, EyesWeb[3] XMI (for eXtended Multimodal Interaction) is a platform for real-time multimodal

[3] www.eyesweb.org.

processing of multiple data streams. We used EyesWeb's libraries for the extraction of the time-dependent features from the video streams. Second, Mediacycle is a software for management and content-based browsing of multimedia collections (sound, images, videos). Mediacycle [19] is used here to analyze the features and to manage the database. It allows to perform similarity queries via a nearest neighbor search and k-means clustering. Its kernel uses plugins to compute the media features. Communication between Mediacycle and EyesWeb is done with the Open Sound Control [20] protocol. The graphical user interface was designed with Flash by the graphic designers David Coumont[4] and Nicolas Rome (http://www.speculoos.com/). It communicates with Mediacycle using the TCP protocol. The global architecture is shown in Fig. 8.

7 Results and demonstrations

7.1 Installation

The installation is composed of a life-size image (6 m × 3.50 m) projected in high definition on a wide screen (16/9 format), a control booth with an interactive touch screen, a video projector and a computer video server, on which a software is installed. The computer contains the software for navigating the database. Sound is played through 5-channel surround to create sonic perspective of solos. It is designed to be installed in a public space and was presented for the first time at the Charleroi-Danses festival, in a subway station (see Fig. 9), where many passersby could interact. As seen on the live videos of the installation on the DANCERS! website, the first impressions of the users were encouraging. Several users noted that the display encouraged them to think about dance in new ways.

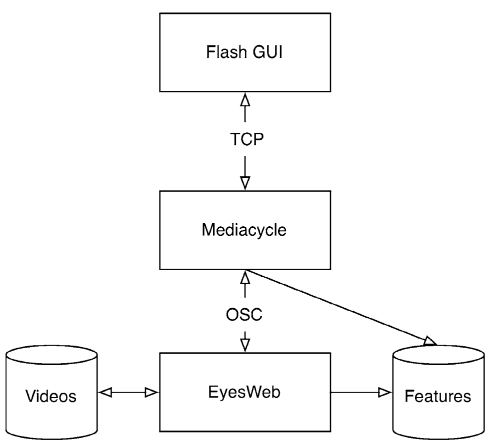

Fig. 8 Architecture of the system

[4]http://www.davidcoumont.be/.

7.2 Website

The software for navigating the database, which is installed on the computer in the above-mentioned application, has been slightly adapted for the web in two ways. First, instead of projecting the high-definition image on a separate screen, it is in this case displayed on the same screen. Second, the level of interaction was adapted, because the single-touch touchscreen has a much lower resolution than what can be achieved when browsing the website with a mouse or a trackpad.
The website keeps track of the number of times each dancer has been watched, showing that it has already attracted thousands of visitors. This is also an indication of the interest generated by each dancer. The website currently gives the possibility to send feedback in the form of a short text message. It could also be used to annotate videos and to evaluate our browsing system.

8 Conclusions and perspectives

In this project we have addressed several aspects of the development of a dance video browsing software. We proposed a set of features allowing the description of several aspects of the dance performances, in particular a compact description of the stage occupation and a description of the personal space around the dancer. Those features allow content-based navigation in the database of the DANCERS! project.

Fig. 9 Live installation at the Charleroi-Danses festival. *Left*: overview of the installation. *Right*: zoom on the touch screen

The public installation of our system as part of a dance festival as well as the associated website have shown the robustness of our system and its practical use. Several users indicated that the features-based navigation led them to think about a dance choreography in new ways.

The proposed description is still considered "low-level", because it is based essentially on the silhouette of the dancer and its bounding box. The description of the personal space could be extended to "high-level" concepts such as the smoothness or the impulsiveness of the gesture [21]. Our work also raised the issue of comparing the temporal structure of two performances, which needs to be further studied.

An objective evaluation of the proposed features has to be performed in the context of database exploration and query by similarity. A first step in this direction has been done by selecting manual annotation criteria and performing a first annotation. Future work will include the adaptation of the website for annotation purposes, as well as usability tests of the interface.

Acknowledgements We would like to thank Bud Blumenthal, the members of his collective (www.bud-hybrid.org) and the dancers for making it possible to work with their high-quality videos.

This work has been supported by the numediart research project, funded by Région Wallonne, Belgium (grant N°716631).

References

1. Smeulders A, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22(12):1349–1380
2. Keim D (2002) Information visualization and visual data mining. IEEE Trans Vis Comput Graph 1–8
3. Santini S, Jain R (2000) Integrated browsing and querying for image databases. IEEE Multimed 7(3):26–39
4. Zhang Z, Zhang R (2008) Multimedia data mining: a systematic introduction to concepts and theory. Chapman & Hall/CRC, London
5. Nguyen G, Worring M (2008) Interactive access to large image collections using similarity-based visualization. J Vis Lang Comput 19(2):203–224
6. Boujemaa N, Compañó R, Dosch C, Geurst J, Kompatsiaris Y, Karlgren J, King P, Köhler J, Le Moine J, Ortgies R et al (2007) Chorus deliverable 2.1: state of the art on multimedia search engines
7. Heesch D (2008) A survey of browsing models for content based image retrieval. Multimed Tools Appl 40(2):261–284
8. Moghaddam B, Tian Q, Lesh N, Shen C, Huang T (2004) Visualization and user-modeling for browsing personal photo libraries. Int J Comput Vis 56(1):109–130
9. Rubner Y, Guibas L, Tomasi C (1997) The Earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In: Proceedings of the ARPA image understanding workshop, pp 661–668
10. Rorissa A, Clough P, Deselaers T (2008) Exploring the relationship between feature and perceptual visual spaces. J Am Soc Inf Sci Technol 59(5):770–784
11. Camurri A, Mazzarino B, Volpe G (2004) Analysis of expressive gesture: the eyes web expressive gesture processing library. In: Lecture notes in computer science. Springer, Berlin
12. Camurri A, Mazzarino B, Ricchetti M, Timmers R, Volpe G (2004) Multimodal analysis of expressive gesture in music and dance performances. In: Lecture notes in computer science. Springer, Berlin, pp 20–39
13. Laban R (1963) Modern educational dance. Macdonald & Evans, London
14. Laban R, Lawrence FC (1947) Effort. Macdonald & Evans, London
15. Boone RT, Cunningham JG (1998) Children's decoding of emotion in expressive body movement: the development of cue attunement. Dev Psychol 34:1007–1016
16. Volpe G (2003) Computational models of expressive gesture in multimedia systems. Ph.D. Dissertation, Faculty of Engineering, Department of Communication, Computer and System Sciences
17. Hurley N, Rickard S (2008) Comparing measures of sparsity. Mach Learn Signal Process 55–60 (2008)
18. Ehrgott M (2005) Multicriteria optimization. Springer, Berlin
19. Siebert X, Dupont S, Fortemps P, Tardieu D (2009) Mediacycle: browsing and performing with sound and image libraries. In: Dutoit T, Macq V (eds) QPSR of the numediart research program, vol 2(1). Numediart research program on digital art technologies, vol 3, pp 19–22. http://www.numediart.org/docs/numediart_2009_s05_p3_report.pdf
20. Wright M, Freed A (1997) Open sound control: a new protocol for communicating with sound synthesizers. In: International computer music conference
21. Mazzarino B, Mancini M (2009) The need for impulsivity & smoothness—improving hci by qualitatively measuring new high-level human motion features. In: SIGMAP, pp 62–67