Skeletons from Sketches of Dancing Poses

Manuel J. Fonseca
Dept. of Computer Science and Engineering
INESC-ID/IST/TU Lisbon, Portugal
Email: mjf@inesc-id.pt

Stuart James John Collomosse
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, Surrey, UK
Email: {s.james, j.collomosse}@surrey.ac.uk

***Abstract*—The contribution of this paper is a sketch parser able to recognize the several components of a skeleton described using the drawing of a stick-man. We describe the sketch parser in detail, and briefly outline how it is applied to form the front-end of a sketch based retrieval system capable of searching for human poses in archival dance footage.**

I. Introduction

In recent years there has been a proliferation of online digital archives of the performing arts, and in particular Dance. These digital repositories promise new opportunities to search and discover historically significant performance. However current dance archives are searchable only using a text-based index of keywords, identified by the archive curators. These keywords typically focus upon authorship or time-location metadata rather than the choreographic details of the content. We report work in progress towards a system that enables choreographic search, through a sketch based interface that matches sketched human stick figures to human poses in low-fidelity dance video. Pose is the essential element of dance, and users are adept at drawing stick-figure representations of pose regardless of their artistic aptitude. Our sketch retrieval system offers an intuitive and accessible interface for choreographic search.

In this paper we outline the front-end of our sketch based pose search system. Our front-end accepts a free-hand sketch as input and transforms this into a parameterised stick-man; a set of joint angles and related skeleton limb lengths. This is a significant challenge given the intra-class variation exhibited by sketched stick-men. The joint angles of the stick-man form the representation for the subsequent pose search.

A. System Overview

We briefly outline our pose search system for context. The system parses a skeleton from the sketch using our front-end. The joint angles of the skeleton represent a point in a high dimensional feature space. We extract bounding boxes around performers in each frame of the video, and compute a shape descriptor within each box. The descriptor is computed using a variant of Zernike moments (after [1]). In an offline training process we manually mark a small number of poses in the video, and so learn a non-parametric mapping between the skeleton pose space and the descriptor space. At query-time the learned mapping is used to identify local points in the descriptor space, given a skeleton derived from the users' sketch — the related poses are displayed to the user. The skeleton therefore forms our intermediate representation for pose search, motivating a robust front-end for parsing the skeleton from the sketch.

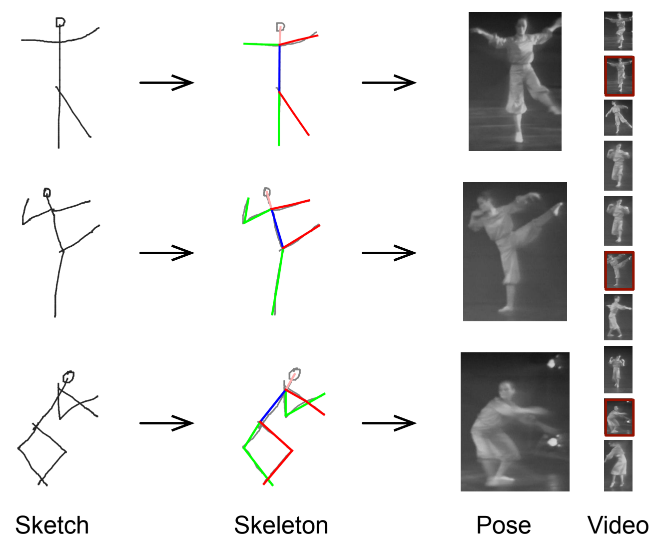

Fig. 1. Overview of the sketch based pose retrieval systems. Users draw the sketch of stick-man, which is interpreted and converted in a skeleton that is then used to locate the desired pose in the video.

II. Skeleton Parser

When we think of stick-men drawings we may be tempted to consider only simple representations like the first in Figure 1. However, when we use stick-men to represent dance poses, we can end up with complex drawings like those shown in Figure 2, where arms or legs cross, or where both limbs are on the same side of the body. In addition to the complexity of the poses, there is also the way users draw a stick-man. For instance, not all users draw the head. Some draw the neck, the torso and one leg using a single stroke, or the two limbs with a single stroke. In some cases the limbs intersect the torso, while in others they do not. A few of these cases are illustrated in Figure 2.

To overcome this, we developed a parser able to identify a skeleton from a sketch of a dancing pose. Users can draw the stick-man dancer freely and without restrictions in: the order of the strokes; the element or elements represented (for example the user can draw the head, the torso and a leg in a single stroke; or the two limbs in the same stroke); strokes do not need to intersect; users can draw or not the head. The only current restriction is that an element of the skeleton cannot be drawn using two (or more strokes). For instance, we can not draw the upper arm and then the forearm using two strokes.

Sketches consist of several points (x and y values), which are often very close, and sometimes they even overlap, especially when the user draws slowly. To eliminate this excessive number of points, and reduce noise, our algorithm starts by eliminating those that are too close (in our case 5 pixels).

Since our sketches are composed mainly by straight lines and/or two straight line segments (*e.g.* a bent arm), we applied another simplification algorithm based on the angle defined by three consecutive points, to eliminate points and straighten the lines. We also tried the Douglas-Peucker algorithm [2], but for our case the algorithm based on the angles produced better results on the identification of corners.

After simplification, we check if any of the strokes that compose the stick-man is the head. We do that by finding the stroke that is most similar to a circle or an ellipse. If we identify any, that stroke is removed from the list of strokes.

The next step is to find the stroke that corresponds to the torso of the skeleton. To do that we use a voting system composed by the following features: size of the stroke; number of intersections; position of the centre of mass; similarity to a straight line and closeness to the extreme points of the other strokes. The stroke that receives more votes is considered the torso stroke and is removed from the list of strokes.

After identifying the torso, we compute the intersections of the other strokes with it. When we have more than one intersection, we always choose the one that is closer to the top of the torso. If the strokes do not intersect the torso, we extend them until they intersect.

With all the intersections identified, we have a list of potential arms and legs. From each of these lists we select the two biggest lines as limbs and we unify the shoulders and the waist to identify the common point for the arms and the legs, respectively.

The final step is to identify the sides for each limb, by checking the position of the elbow/knee relatively to the torso. When both limbs are in the same side of the torso, we verify which is more to the right (from the user point of view) and consider it the right limb, as illustrated in the third skeleton in Figure 1.

At the end of the parsing process, we have a semantic skeleton, where we know exactly how each element are defined in terms of geometry and also what they are. This provides us with rich information for computing a descriptor to describe and compare skeletons of dance poses, allowing us to find dance poses by providing a simple sketch of a stick-man.

III. Preliminary Evaluation

To evaluate the accuracy of our skeleton parser we performed a preliminary evaluation, where two users were asked to draw sketches of stick-man poses on top of video frames. Each user drew 105 stick-men. We parsed all the 210 sketches and registered a recognition rate of 90.5%. In Figure 2 we can see various examples of recognized skeletons. All except the last were correctly identified. The majority of the incorrect recognitions were due to the wrong detection of the torso stroke. This happened mainly in poses where the dancer is in a curved position near the floor, as illustrated in the last pose from Figure 2.

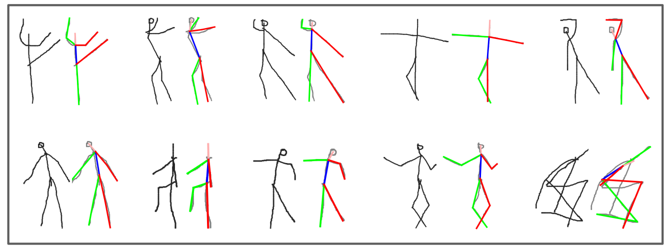

Fig. 2. Different sketched poses and the correspondent skeleton identified. We can see here the complexity of some poses and also the different ways of drawing them.

This preliminary evaluation served mainly to identify problems in the skeleton parser, so we can improve it. We plan to perform a more complete evaluation, with more users (to collect different ways of sketching) and with more poses.

IV. Conclusions

In this paper we described a parser for the identification of the various elements of a skeleton, from a sketched figure of a stick-man in a dance pose. It is based on a set of rules that uses very few parameters, offers a large drawing flexibility, allowing users to sketch freely almost without any constraints, and presents a high recognition rate. Additionally, our approach also maintains the lengths of the original sketch.

Currently we are using the skeleton parser in a system for pose retrieval, where users can search a desired pose by drawing a stick-man. The skeleton from the sketch is compared with the skeletons from a set of pre-marked frames, that then will be compared with all the frames of the video. The next step is to combine various sketches of poses in a visual narrative to provide the storyboard for the generation of new videos containing dance poses from existing videos.

Acknowledgments

This work was partially supported by Portuguese national funds through FCT (projects PEst-OE/EEI/LA0021/2011 and PTDC/EIA-EIA/108077/2008, and Manuel J. Fonseca sabbatical grant SFRH/BSAB/1221/2011) and project AHRC “Digital Dance Archives” project AH/H037926/1.

References

[1] M. Brand, “Shadow puppetry,” in *Proc. Intl. Conf. on Computer Vision*, vol. 2, 1999, pp. 1237–1245.
[2] D. Douglas and T. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” *The Canadian Cartographer*, vol. 10, no. 2, pp. 112–122, 1973.