Character Animation from 2D Pictures and 3D Motion Data
ALEXANDER HORNUNG, ELLEN DEKKERS, and LEIF KOBBEKT
RWTH-Aachen University

Presented by:
Harish
CS-525

First presentation
Abstract

This article presents a new method to animate photos of 2D characters using 3D motion capture data. Given a single image of a person or essentially human-like subject, this method transfers the motion of a 3D skeleton onto the subject’s 2D shape in image space, generating the impression of a realistic movement.

This article presents robust solutions to reconstruct a projective camera model and a 3D model pose which matches best to the given 2D image. Depending on the reconstructed view, a 2D shape template is selected which enables the proper handling of occlusions. After fitting the template to the character in the input image, it is deformed as-rigid-as-possible by taking the projected 3D motion data into account.

It works for images from arbitrary views and requires only a small amount of user interaction. The authors presented animations of a diverse set of human (and nonhuman) characters with different types of motions, such as walking, jumping, or dancing.

Introduction

In recent years, research on 2D image manipulation has received a huge amount of interest. Chuang, in his article showed that he animated passive elements, such as water and trees, which are subject to natural forces like wind. In this article, he wanted to take the idea of creating animations directly in image space one step further by making photographed persons move.

One possible approach to address this problem would be the reconstruction of a textured 3D model, and to animate this model using classical animation techniques. This, however, would require
complex, fully textured 3D models which have to be created and adapted per image. In particular, for highly detailed characters such as the Scarecrow example shown in Figures 2 and 7, the required manual model-adaptation process would be impractical. Moreover, it would be necessary to apply very sophisticated 3D rendering techniques to realistically embed the 3D model into the 2D image so as to preserve the photorealism or style of the original input image.

By contrast, even simple 2D morphing and blending often leads to more convincing results than using sophisticated 3D reconstruction and rendering. For instance, methods such as Poisson-matting or 2D image completion allow for a smooth and realistic combination of different image contents, which is much harder to achieve when trying to augment 2D images with 3D models.
Overview

As mentioned in the introduction, the central idea of this paper is to deform the 2D shape of an arbitrary character directly in image space using 3D motion data (publicly available from, e.g., CMU Graphics Lab Motion Capture Database [2007]). The 3D motion data consists of a sequence of poses of a predefined human skeleton model. Each pose contains the respective positions and
orientations of the skeleton bones and joints in 3D space. The first step maps the 3D data to the 2D image by reconstructing a proper camera model and a model pose that best fits the 2D character in the image. This can be achieved by letting the user manually specify correspondences between 2D and 3D, simply by selecting the joint positions in the 2D image that correspond to the joints of the given 3D human model.

**Joint Selection**

To establish the necessary correspondences between the subject’s 2D shape and the 3D skeleton model, we let the user manually select joint positions in the input image. This is done by a simple interface, where the user can move the joints of a stylized skeleton having a structure compatible to our 3D data (see Figure 2). We preferred a manual user interaction over automatic procedures because this step generally takes just a few minutes to complete, and leads to superior results in poses which are ambiguous for automatic human-pose estimators, or which are difficult to estimate due to occlusions, for example. Furthermore, such methods would require a diverse dataset of example poses to work for arbitrary images.
INITIAL MODEL-FITTING

To create convincing animations of a 2D character, its shape would be deformed in a plausible way, while keeping the effort for generating animations on a minimal level. The basic idea is to use a set of generic shape templates $T$, which represents the topological changes of a generalized character model for different viewpoints of the camera (see Figure 3). These templates are fitted in a semiautomatic approach to the character. In our context, we define a shape template as a 2D nonmanifold triangle mesh which allows for
the representation of different animation layers. For example, when animating characters from a side view, we have to consider different layers for separately moving parts of a body, for example, one layer for the foremost arm, one for the body and the foremost leg, and one for the remaining arm and leg, respectively (see Figure 3 (c)). Moreover, these layers cannot move independently, but have to be “stitched” together to convey the impression of a connected body when animated. Hence, each layer of a template $T$ consists of a triangulated set of vertices representing the shape boundary and skeleton joints. The different layers of each template are connected by shared boundary vertices. Additional vertices are added inside each layer, which allows us to generate more realistic animations by “inflating” the shape templates prior to animation.

Fig. 3. Set of three shape templates generated by cycling in 45 degree steps around a person; (a) frontal view with one animation layer; (b) and (c) halfprofile view and a side view with four layers each: foremost arm, body and front leg, and one for the leg and the arm at the back, respectively
CONCLUSION

In this article, the authors presented a complete, easy-to-use system for animating 2D images of arbitrary characters with 3D motion. They showed how simple user interaction, namely the selection of a few 2D joint positions, can be exploited to automatically reconstruct a geometrically plausible camera calibration and model pose from 3D motion data. They presented an initial model-fitting step using a generic set of shape templates to animate arbitrary human characters. Finally, they introduced an as-similar-as-possible shape deformation algorithm to deform these shapes in a protectively more correct manner, allowing us to generate still frames and animations of a large variety of characters from different types of 3D motion.