Servingo 3D Reconstruction: Break through the Barrier of 2D Television

Dirk Balfanz, Cornelius Malerczyk

German Abstract

Das Projekt servingo hat sich unter anderem zum Ziel gesetzt, aus den Aufnahmen von gewöhnlichen Fernsehkameras dreidimensionale, animierte Szenen zu rekonstruieren. Neben einer realistischen Darstellung des Stadions werden insbesondere auch die Bewegungen der an einem Spielzug beteiligten Spieler realitätsgetreu nachgebildet. Um die Rekonstruktionsergebnisse zeitnah nach ihrer Aufnahme bereitstellen zu können, muss die Rekonstruktion weitgehend automatisch ablaufen und ein manuelles Eingreifen in die Rekonstruktion minimal sein. Durch eine 3D-Rekonstruktion der Szene wird die natürliche Barriere des zweidimensionalen Fernsehens durchbrochen. Der Anwender ist jetzt in der Lage, selbst zum Kameramann oder Regisseur zu werden und die Szene von frei wählbaren Blickpunkten aus zu betrachten. Dem Gedanken des »mobilen Fußballfans« wird in servingo auch bei den 3D-Spielszenen Rechnung getragen, indem die virtuellen Szenen nicht nur am Computer, sondern insbesondere auch auf mobilen Endgeräten wie PDAs oder Mobiltelefonen verwendet werden können.

Motivation

A main focus of the servingo project is on the 3D reconstruction of interesting scenes of a soccer match (clips of interest). This includes a system for a fast production of additional content (3D and information) as well as the detailed reconstruction of poses and motions of soccer players. These scene-graphs themselves are then used to generate virtual cameras to enable the users to look at certain situations from different virtual camera perspectives and therefore increase the engagement in and communication about the soccer match without running into the legal quandary that applications dealing with real scene reconstruction or interpolation have to face.

3D Reconstruction

The main purpose of the 3D reconstruction system is to provide essential information about short clips of interest during a soccer match. This information, the 3D reconstruction of the soccer stadium and more importantly the reconstruction of body motion information is extracted from video streams, captured by a standard TV broadcasting camera. The goal here is that neither specialized equipment for the camera systems nor sensors to be worn by the players should be used, rather the video data that is captured for live broadcasting anyway.

The approach to be followed is threefold. The reconstructed 3D model of the soccer stadium is a requirement to support the camera calibration process and is used as well for the intended application. Furthermore, the broadcasting camera needs to be calibrated: This step provides information about the camera pose and especially separated determination of camera position and camera movement (pan, tilt, and zoom). And, last but not least, 3D motions of the soccer players are reconstructed: For selected scenes (clips of interest) during a soccer match the players' body motions are extracted from the calibrated video data. A combination of computer vision-based approaches such as 3D pose adaptation and utilization of pre-knowledge in terms of kinematical motion descriptions and prediction/estimation/smoothing filters are used to finally provide the motion information in a standard animation format to be used by the rendering application on stationary and mobile devices.

Stadium Reconstruction

The reconstruction of the environment which the soccer match takes place in or which it is going to be visualized in can mainly be performed prior to the actual match. This is possible because the overall geometry of the scene does not change significantly. This approach allows, on the one hand, to spend more effort on achieving a higher quality, on the other hand, pre-existing models can be re-used at subsequent events in the same environ-

Figure 1: Pose adaptation process for scorer, goalkeeper and referee during a penalty kick.

Figure 2: Viewing the 3D reconstruction on a PDA.

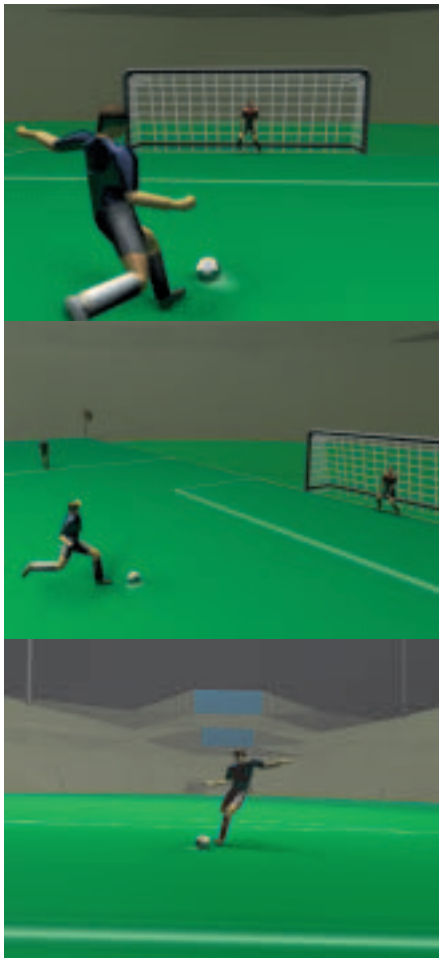

Figure 3: Different viewpoints of a reconstructed sequence: a) camera behind scorer, b) virtual closeup, c) goalkeeper's view of the penalty kick.

ment. More than a sufficient geometric accuracy, the surface color (textures) is crucial for a realistic impression of the reconstructed scene. The approach pursued here reconstructs the 3D geometry as well as the surface colors from photographic views of the scene, thereby generating a photorealistic model of the environment. In order to enhance the realistic impression of the reconstructed environment, video textures from pre-calibrated video cameras in the scene are used for texturing the 3D geometry.

Consequently, the 3D model also incorporates a semantic structure, which allows annotations, e.g. links to companies advertising on the perimeters, and encoding of the relevance of each part for a specific event. This information allows to quickly obtain a tailor-made 3D model of the environment for the composition of reconstructed scenes.

As an important by-product, the photogrammetric reconstruction of the environment incorporates also video footage from the TV cameras used for the 3D motion reconstruction, thereby yielding part of the calibration information for these cameras. As a result, the TV cameras and the 3D geometry information obtained from them are readily in the same coordinate frame as the model of the environment.

3D Motion Reconstruction

For a »clip of interest«, player tracking during a football match is performed using only image and computer vision-based technology.

With respect to the intended application and its focus on entertainment a high degree of accuracy, which is typically the focus issue of motion capturing, is considered of lower priority. Since the servingo project does not especially aim at a highly accurate motion analysis, rather at entertaining, the visual outcome perceivable for the user is a key issue. Consequently, smoothing and filtering approaches are used as post-processing steps of the motion reconstruction to ensure a realistic motion impression, usable for the rendering applications.

The reconstruction module uses a motion estimation approach combining a silhouette-based pose adaptation algorithm with a kinematics model modeling a player's typical motions like walking, running, kicking, etc. To overcome occlusion problems due to invisible limbs, a combination of inverse kinematics and a statistical kinematics model is used, taking both anatomical restrictions of possible human joint rotations and motion-specific constraints into account. Further, the kinematics model is used for predicting the player's future postures from the current and previous ones. A 3D pose adaptation module adjusts the joint rotations of an articulated synthetic body model to the captured image data.

The player's position and its calculated articulations over time is finally translated to standardized animation parameters like VRML/X3D, H-Anim 2001 or MPEG-4 Body Animation Parameters (BAP) to animate virtual soccer players on the application side. To overcome measurement errors like e.g. flickering of the avatar's movements, the reconstructed animations are post-processed aiming at eliminating jitter effects before storage.

Points of Contact

Dr. Dirk Balfanz
Dipl.-Math. (FH) Cornelius Malerczyk
ZGDV, Darmstadt, Germany
E-mail: Dirk.Balfanz@zgdv.de
Cornelius.Malerczyk@zgdv.de