3D-Reconstruction of Soccer Scenes

C. Malerczyk
Dept. Visual Computing
ZGDV Computer Graphics Center
Darmstadt, Germany

H. Seibert
Dept. Visual Computing
ZGDV Computer Graphics Center
Darmstadt, Germany

Abstract ***The aim of this paper is to demonstrate the 3d reconstruction of human body poses and motions in the field of sports events. The reconstruction does not require any technical hardware but uses originally broadcasted television images only. A game scene is reconstructed as a 3-dimensional model and the user can look at this model not only in slow motion, but also from arbitrary positions and viewing angles of his/her choice. He/she can navigate through the scene assuming various viewing angles and positions. To judge on a penalty decision, the user can e.g. assume the viewpoint of the referee; or after a goal, the position of the goalkeeper can be taken to see whether the view on the ball was obstructed.***

Keywords: 3D Body Tracking, 3D Reconstruction, Pose Estimation, Computer Vision, Image Understanding

1 Introduction

Pose and motion reconstruction for sports events is a good sample how to enhance multi modal interfaces for interaction in order to use them in immersive virtual environments for edutainment and entertainment applications.

One main focus of the Servingo project [12] is on the 3D reconstruction of interesting scenes of a soccer match (clips-of-interest). This includes a system for fast production of additional content (3D and information) as well as the detailed reconstruction of poses and motions of soccer players. These scenegraphs 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. Benefits with respect to the edutainment and entertainment characteristics of this 3D reconstructions are:

- The user becomes an active part of the game.
- The user is able to navigate freely within the 3D scene.
- The user is able to view the scene from different predefined camera positions.
- The user is able to control the playback speed of the scene (e.g. virtual slow motion).
- The user is able to switch on/off virtual enhancements like the trajectory of the ball.

Figure 1: Virtual camera during penalty shot.

2 Related Work

The outcome of many competitive sports events highly depends on very few key actions, e.g. goals in soccer games. Important part of the game experience is the discussion among the spectators on these selected scenes. TV sports coverage attends to this need e.g. by slow motion repetition.

Interactive individualised 3D game analysis brings in a completely new level of information.

The actual game scene is reconstructed as a 3-dimensional model and the user can look at this model not only in slow motion, but also from arbitrary positions and viewing angles of his/her choice. He/she can navigate through the scene assuming various viewing angles and positions. To judge on a penalty decision, the user can e.g. assume the viewpoint of the referee; or after a goal, the position of the goalkeeper can be taken to see whether the view on the ball was obstructed. The current

Figure 2: Penalty shot: original camera image (top) and view on 3d reconstructed scene (bottom).

state of the art with respect to interactive game scenes can be divided into two major domains,

1. the reconstruction of sport scenes and motion recognition and

2. the visualisation of immersive 3D environments on stationary and mobile devices.

Broadcasting of virtual or augmented sport scenes is nowadays a well-known technology to enhance the television programme during sports events. An augmented offside-line or the measurement of the distance between ball and goal during a free kick is common practice in soccer match broadcasts. Furthermore, the 3D reconstruction of a static scene (a freeze image of a video sequence) has become popular lately. Nevertheless, these reconstructions are based on one single still image of a scene sequence and the pose fitting of the athletes is mostly manual or semi-automatic work. However, 3D reconstruction of human motion is basically located in the domain of professional motion capturing studios using special hardware for the capturing itself as well as sensors to be worn be the captured actor (cp. systems like VICON [15], Ascension, etc.). A full 3D motion reconstruction using ordinary video footage demands on manual interaction. One of the first examples of a fully automatic reconstruction using television footage only is described in the IST project PISTE [7],[8] addressing the domain of athletics for interactive television broadcasts. Other systems developed mainly in the United States are Systems like Virtual Spectator [16] for sailing events, pitch trax or Tennis Proview [10] in the field of baseball and tennis and SportVision [13] in the domain of NasCAR races. Further research and development is done at SymahVision [5], ORAD [9] or in Germany by CAIROS [3]. However, most work is based on the recognition of the position in 3D space of the athletes only or the reconstruction is restricted to a small volume of interaction only (e.g. the swing at playing golf). For the rendering part of 3D reconstructed scenes standards like VRML or X3D can be used to ensure a numerousness of users that are able to view and interact with the scene. Projects like OpenSG Plus [11] are taken as a basis for the rendering on stationary devices like desktop Pcs. For mobile devices like PDAs or mobile phones Mobile Java (J2ME) technology is used. With JSR-184, the Mobile 3D Graphics API for J2ME, a standard for three-dimensional graphics on mobile devices is available that supports device independent transmission, rendering and interaction on many different mobile devices.

3 Stadium Reconstruction

The 3D reconstruction of the environment in which the soccer match takes place or in which it is going to be visualised 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 in order to achieve a higher quality, on the other hand pre-existing models can be re-used on subsequent events in the same environment. More than a sufficient geometric accuracy, the surface colour (textures) is crucial for a realistic impression of the reconstructed scene. The approach pursued here reconstructs the 3D geome-

try as well as the surface colours from photographic views of the scene, thereby generating a photo realistic 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 [7]. Consequently, the 3D model also in-

Figure 3: View on the photogrammetric reconstructed Commerzbank-Arena in Frankfurt, Germany.

corporates 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.

4 Motion Reconstruction

The main purpose of the three-dimensional reconstruction is to break through the barrier of standard two-dimensional television. For a 3d reconstruction short but important scenes of a soccer match are selected (clips-of-interest) to provide additional information like for example the trajectory of the ball or the detailed and natural reproduction of the movements of all player, which are involved in the original scene.

The reconstructed, three-dimensional game scenes are used to provide the possibility of easy and intuitive interaction for the user, to view the scene from a preselected virtual viewpoint or to move the virtual camera to an arbitrary position in the stadium. In opposite to the real camera operator or the real director of the TV station the user is now allowed to position the camera even directly on the soccer ground, which is obviously forbidden in the real world scene. The user is now able to view the scene from behind the scorer or even with the eyes of the goalkeeper or the referee. The 3D reconstruction aims at the estimation of all relevant 3D parameters of a clip of interest. For a soccer match, these clips have typically a duration between three and six seconds. Examples of typical clips are free kicks, goals, shots at the goal or even a serious foul play. An important condition for the reconstruction is the avoidance of all technical and technological accessories like special cameras or even sensors, that have to be worn by the athletes. The only input device is the originally broad-casted television footage. To achieve this aim, a new algorithm was developed, which is mainly threefold: The first step is the calibration of the recorded video sequence. As a result of this step for each frame of the video sequence position, orientation and focal length of the camera (pan, tilt and zoom) is known. The next step of the algorithm is to estimate the movements of each player involved in the scene. Therefore, a pose adaptation approach is used fitting synthetically generated silhouettes of a 3D avatar with the real silhouettes of the athlete in the camera images. Last but not least the reconstructed animations of the players are attached to 3D models of the players and combined with a 3D model of the stadium. Within this post-processing step additional 3D content like predefined viewpoints, camera flights and for example sliders to control the virtual playback speed (e.g. slow motion) can be added to the scene. As a result of the 3D reconstruction of a clip of interest the 3D scene is available in different output formats for the usage on many different output devices. Output devices are mainly divided into stationary devices like desktop computers and mobile devices like PDAs or mobile phones. Interactive 3D scenes are stored as VRML97/X3D [2] files for stationary devices, whereas for mobile devices like PDAs and mobile phones a Java application is generated, which can be transferred to the device and started locally. Therefore, the Java standard JSR-184 for mobile devices is used to ensure the applicability on a maximum of different devices. Par-

Figure 4: JSR184-based interactive 3D scene on mobile phone.

Figure 5: Viewing the 3D reconstruction on a PDA.

ticularly for technically unversed user and even for users without a 3D capable output device movie clips and still images are provided, which present the 3D scene from a predefined camera view.

5 Pose Adaptation

The core technology of the reconstruction of sports scenes is a new developed video-based algorithm for the three-dimensional reconstruction of human movements from two-dimensional camera sequences. The algorithm is not solely limited to the reconstruction of soccer sequences, but may be used in general for the pose and motion estimation of deformable objects consisting of kinematic chains. The algorithm is based on the comparison of real and synthetic generated silhouettes of the human body. Obviously, as the real silhouette the appearance of the player in the original camera image is identified. Next, a virtual three-dimensional model of the player is positioned in 3D space and by projecting this model a synthetic silhouette is generated (see figure 6). To generate correct silhouettes,

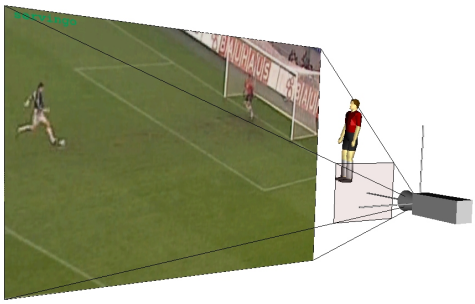

Figure 6: Principle of the silhouette based 3d pose adaptation.

a virtual copy of the original broadcasting camera with exactly the same parameters (known by the calibration process of the camera) is used. The pixel-wise comparison of both silhouettes leads to the output value of a residual error function, which is minimised using a nonlinear optimisation algorithm. When the minimum of the residual function is found, the pose (described by 3D joint rotations) of the avatar fits to the pose of the real athlete seen in the camera image (see figure 8).

5.1 Silhouette Generation

For the online generation of the synthetic silhouettes during the pose adaptation process the open source scene graph system OpenSG [11] is used. After selecting a player from a database generated prior to the soccer match a VRML model of a generic player is loaded into the scene graph. For

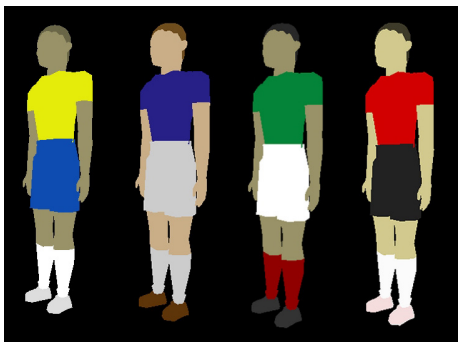

Figure 7: Low poly avatars used for synthetic silhouette creation (four different configuration files used).

the final rendering of a reconstructed 3D scene any high-end skin and bone system can be used for the visualisation of the soccer player. Due to the fact that during the optimisation process a great many of different silhouettes has to be generated as fast as possible, a low poly avatar mesh with less than 2000 polygons is used for the 3D pose adaptation. To individualise the generic player mesh with respect to the selected real player's silhouette only a few important appearance parameters are configured in the scene graph. In addition to the height of the avatar (to ensure a correct size of the rendered silhouette) only six colour attributes are read from the database and used for the mesh configuration (see figure 7):

```
Height      189
Skin        0.657 0.630 0.444
Hair        0.160 0.160 0.160
Jersey      0.859 0.078 0.078
Trousers    1.000 1.000 1.000
Socks       0.859 0.078 0.078
Shoes       0.200 0.200 0.200
```

Within the players database `Height`, `Skin`, `Hair` and `Shoes` are stored as individual parameters differing from player to player, while `Jersey`, `Trousers` and `Socks` have to be defined only once for a complete soccer team (see figure 7).

Figure 8: Comparison of real and synthetic (shifted right for visualisation) silhouette after 3D pose adaptation.

5.2 Nonlinear Optimization

Due to the numerousness of optimising parameters (3d position and orientation of the player and up to 14 rotations of the most important human joints), Simulated Annealing, a generic probabilistic meta-algorithm for the global optimisation problems [6] is used to avoid the optimisation to get stuck in local minima of the high-dimensional function space and therefore producing wrong adaptation results. To reduce the immense computational costs during optimisation and to minimise the overall pose adaptation time, each joint rotation can be constrained using standard anatomical parameters, which leads to a drastic reduction of the size of the solution space.

Simulated Annealing is a stochastic optimisation algorithm. Its purpose is the minimisation of an objective function $E(X)$ where X is a multidimensional state vector of the objective function. An initial value x_0 is randomly changed over many iterations by updating the current solution by a solution randomly chosen in its neighbourhood. The change of the variable from $x_i - 1$ to x_i may result in an increase or decrease of the function value. To avoid to get stuck in a local minimum, it is necessary not only to allow ameliorations but also suitable deterioration. This is done be introducing a temperature parameter T, which is decreased every n iterations. $E(x_i)$ is accepted as the next current solution if $E(x_i) < E(x_i - 1)$. Otherwise there are two possibilities for the state of X: Either x_i will be accepted with the probability $P(x_i)$ or rejected with $1 - P(x_i)$. P is calculated according to the Boltzmann distribution

$$P(\Delta E) = \exp\left(\frac{-\Delta E}{k_B T}\right) \tag{1}$$

with $\Delta E = E(x_i) - E(x_i - 1)$ and k_B the Boltzmanns constant. At the end of the algorithm, when T is small enough, deteriorations will be hardly accepted and most of the time only downhill steps are accepted.

For the 3D pose adaptation an arbitrary state vector x_i consists of the 3d position of the avatar followed by the components of all joint rotations (Euler angles) used for the pose definition of the player

$$x_i = (t_x, t_y, t_z, root_x, root_y, root_z, l\_shoulder_x, ...)^T. \tag{2}$$

The initial value x_0 is defined as the avatar's default pose (see figure 7) at a manually given rough position of the player in 3D space. Practically, after selecting a player the user has to define a rough position and orientation of the avatar in the first broadcasted image of the sequence only. For all following images at x_{i+1} the result pose of x_i is used as the initial pose vector for the next optimisation step.

To use Simulated Annealing for 3D pose adaptation, it is necessary to adapt the algorithm by specifying the objective function $E(X)$ and the change of the variable X from one state to another. Obviously, a random step during optimisation can easily be defined as a small change of a randomly chosen component of the current state vector. In other words: For each iteration step either the position of the avatar or a single joint rotation is changed

slightly before a new synthetic silhouette will be generated.

5.3 Residual Function

For each iteration step a synthetic silhouette is generated using the current state vector x_i and the virtual camera parameters for the projection of the avatar. To evaluate if the current 3D pose fits to the real pose of the player seen in the broadcasted image, the real and the synthetic silhouettes are compared pixel wise, which leads to the definition of the objective (residual) function used by Simulated Annealing. Within the bounding box of the projected avatar the residual is calculated by

$$E(x_i) = \sum_{p_{x,y}} (|c_r - c_s|) \cdot t + (|e_r - e_s|) \cdot (1 - t) \quad (3)$$

with a given weight factor t for colour differences between real $c_r = (R_r, G_r, B_r)^T$ and the synthetic $c_s = (R_s, G_s, B_s)^T$ silhouette pixels and between Sobel edge intensities e_r and e_s for each pixel position $p_{x,y}$. Exemplarily, figure 9 shows the residual function of the rotation of the left shoulder around the x-axis for $[-3.5; 1.0]$.

Figure 9: Residual function of x-axis of the left shoulder.

6 Web Portal

The reconstructed 3D scenes (including predefined videos and images of a scene) have been presented to the public using the Servingo web portal [12]. The Servingo web portal was developed on top of open-source software to provide a highly dynamic and interactive web based service infrastructure supporting visitors and organisers of the FIFA World Cup™ 2006 in many different ways such as guiding to local infrastructures in traffic, accommodation, entertainment and organisation as well as providing additional information on the soccer games. A detailed description of the web portal and its architecture is given in [14]. During a test period the web portal was accessible from several devices ranging from cellular phones, to personal computers including mobile devices such as PDAs. Web-servers were used to provide the media files, the upcoming DVB-H (Digital Video Broadcasting-Handhelds) technology [4] has also been tested for transmission of the media files to PDA devices. In this case the media items were provided via scheduled DVB transmissions to mobile devices with corresponding receivers. The web portal was used to provide file access and corresponding meta-data. To simplify access, appropriate meta-data on the media files was held in the databases of the Servingo portal, including information on the event itself, the location date, actual game situation, involved players as well as media type and desired target system. Cellular phones could access J2ME interactive applications, low resolution video files and low resolution images. PDAs were supplied with higher resolution for video and images, personal computers could access VRML files for the 3D reconstruction as well as videos an images in PAL standard resolution.

Figure 10: Layout for 3D scene selection on mobile devices.

Acknowledgements

Parts of the work presented here were accomplished with support of the European Commission through the SIMILAR Network of Excellence (FP6-507609, www.similar.cc) and the project Servingo [12], funded by German Federal Ministry of Economics and Labour, BMWA.

Figure 11: Different camera views on reconstructed game scenes during FIFA World Cup™ 2006.

References

[1] Balfanz, D.; Malerczyk, C.: Servingo 3D Reconstruction: Break through the Barrier of 2D Television. In: *Computer Graphics Topics, 6/2005*, ISSN 0936-2700.

[2] Behr, J; Daehne, P; Roth, M: Utilizing X3D for Immersive Environments. In: *Spencer, Stephen N. (Ed.); ACM SIGGRAPH: Web3D 2004*, Proceedings : Ninth International Conference on 3D Web Technology. New York: ACM, 2004, pp. 71-78, 182.

[3] Cairos: Highresolution 3D localization of dynamic objects, Website, Retrieved November 2006, http://www.cairos.com/.

[4] DVB-H: Homepage of DVB-H Global Mobile TV, Website, Retrieved November 2006, http://www.dvb-h.org/.

[5] Epsis by Symah Vision: Real-time technology for insertion of computer-actuated images into a video sequence, Website, Retrieved November 2006, http://www.epsis.com/.

[6] Kirkpatrick et al.: Optimization by Simulated Annealing. *Science*, 220:671-680, 1983.

[7] Klein, K.; Malerczyk, C.; Wiebesiek, T.: Creating a Personalised, Immersive Sports TV Experience via 3d Reconstruction of Moving Athletes. *International Conference on Business Information Systems (BIS)* May, 2002, Poznan, Poland.

[8] Malerczyk C. et al.: 3D Reconstruction of Sports Events for Digital TV. *Journal of WSCG Volume 11 No. 2. Proceedings (2003), pp. 306-313, International Conference in Central Europe on Computer Graphics, Visualization an Computer Vision (WSCG)*, 2003, Plzen, Czech Republic.

[9] Orad: Video and real-time image processing technologies for TV broadcasting, Website, Retrieved November 2006, http://www.orad.tv/.

[10] Tennis Pro View: Tennis Broadcast Enhancements, Website, Retrieved November 2006, http://www.questec.com/.

[11] Reiners, D.: OpenSG PLUS: Advances to Current Scenegraph Technology. In: *Federal Ministry of Education and Research: Virtual and Augmented Reality Status Conference 2004*. Proceedings CD-ROM. Leipzig, 2004.

[12] Servingo: Homepage of the Servingo project, Website, Retrieved November 2006, http://www.servingo.de/.

[13] SportVision: Enhancements for Sports Television, Website, Retrieved November 2006, http://www.sportvision.com/.

[14] Tazari, M-R and Thiergen, S.: servingo: a Service Portal on the Occasion of the FIFA World Cup™ 2006. In Proceedings: *International Workshop on Web Portal-based Solutions for Turism (IWWPST), pp. 73-93*, 2006.

[15] Vicon: Motion Capturing Systems, Website, Retrieved November 2006 from http://www.vicon.com/.

[16] Virtual Spectator: 3D Sports Anmiation, Website, Retrieved November 2006 from http://www.virtualspectator.com/.