iview – Free-Viewpoint Video for Outdoor Sports Events

Jean-Yves Guillemaut, Joe Kilner, Jonathan Starck, and Adrian Hilton
Centre For Vision, Speech, and Signal Processing, University of Surrey, U.K.
{J.Guillemaut,J.Kilner,J.Starck,A.Hilton}@surrey.ac.uk

Summary

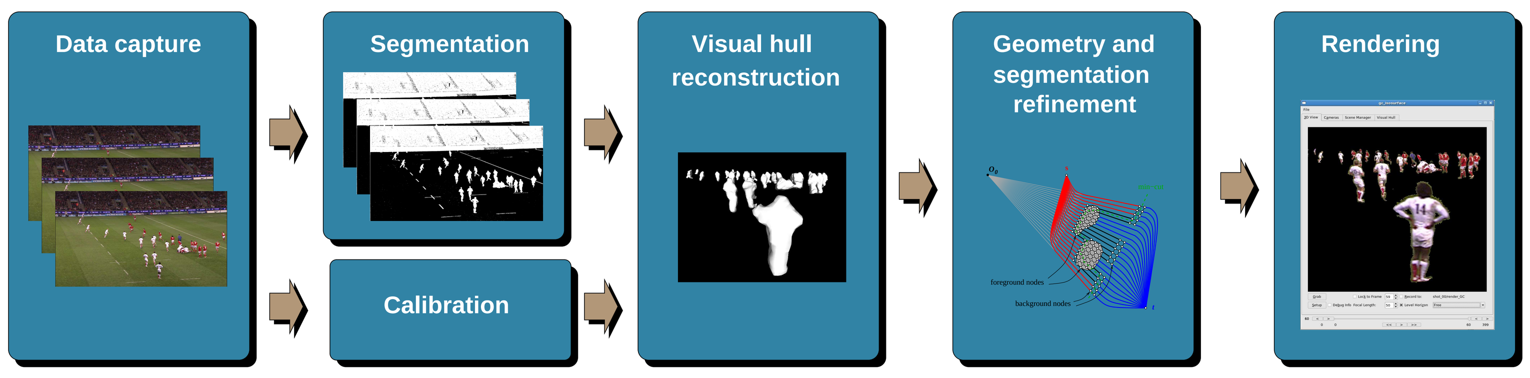
Conventional fixed-viewpoint video systems are restricted to rendering scenes from one of the viewpoints used at capture time. In contrast, free-viewpoint video systems allow the user to specify the viewpoint at rendering time. Although high quality rendering has been achieved in studio environments, broadcast quality free-viewpoint video is still an open problem in the case of scenes captured in less controlled conditions.

Aim of the project To develop algorithms for high quality free-viewpoint video for outdoor sports scenes such as a football or a rugby match. Examples of applications include giving a clearer view of an offside decision or showing the action from a viewpoint inside the pitch (e.g. goal keeper or the referee's viewpoint).

Main challenges:

- Coping with inaccurate camera calibration,
- Performing accurate foreground/background segmentation,
- Scene reconstructing from a small number of views (wide-baseline).

Overview of the Pipeline



Data capture This requires multiple synchronised cameras placed around the scene. These include some of the cameras that are used for TV broadcast as well as additional fixed cameras.

Image segmentation Chroma-keying is applied to produce an initial segmentation, which is often inaccurate as non-grass background areas such as lines, goal posts, or crowd are typically misclassified as foreground.

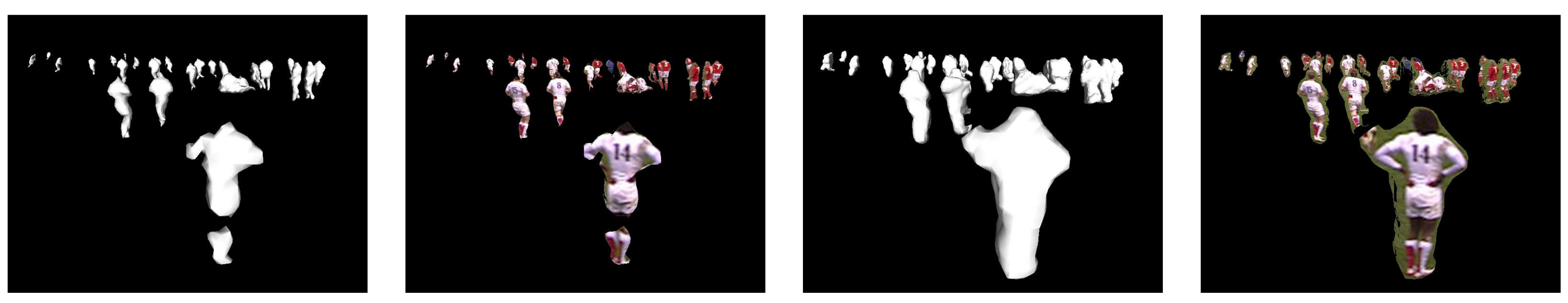
Camera calibration The pitch lines and goal posts are used as a 3D calibration pattern in order to compute extrinsic and intrinsic parameters (including radial distortion) for each frame of each camera.

Visual hull reconstruction This provides an initial global reconstruction which can be used to initialise further refinement techniques.

Geometry and segmentation refinement We use either a deformable model approach or view-dependent graph-cut optimisation algorithm.

Rendering Virtual views are synthesized by view-dependent texture mapping techniques. An OpenGL-based renderer has been implemented and allows the viewer to navigate freely in the scene.

Visual Hull Reconstruction



Visual hull | Visual hull (textured) | Conservative visual hull | Conservative visual hull (textured)

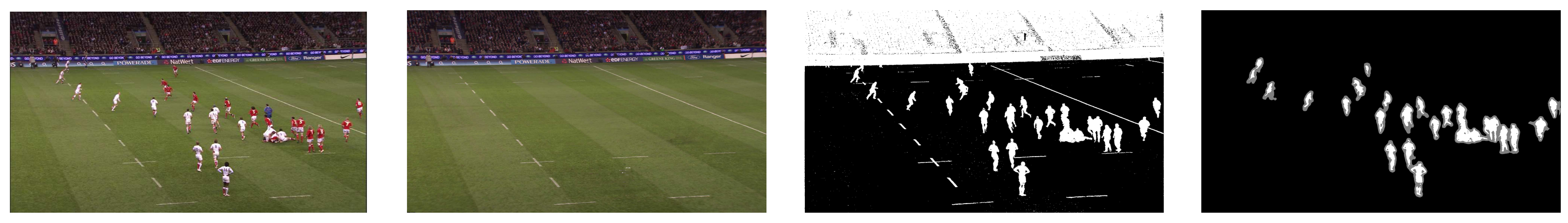
Conventional visual hull With inaccurate calibration and segmentation, this is not guaranteed to enclose the scene surface. In our case, this often results in truncated limbs.

Conservative visual hull A tolerance is allowed during intersection of the backprojected silhouettes, thus producing a dilated reconstruction which is a true upper bound on surface geometry and is suitable to initialise further refinement techniques.

Geometry and Segmentation Refinement

Global refinement is difficult in the case of inaccurate camera calibration as no satisfactory solution is simultaneously consistent with all images. We therefore use a view-dependent approach where geometry and segmentation are refined separately for each view. Two refinement approaches are proposed:

- **Deformable model** fitting (not described here due to space limitations)
- **Graph-cut** optimisation.



Input image | Background image | Chroma-key segmentation | Trimap

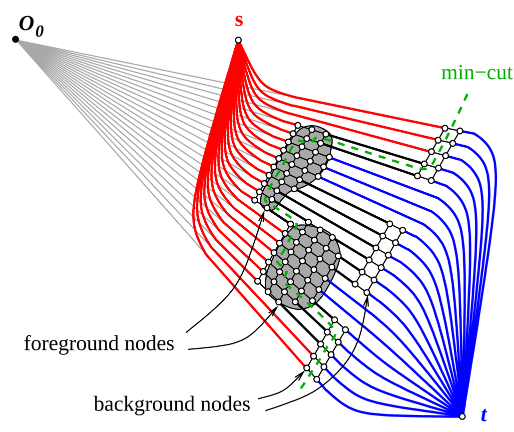
Bayesian problem formulation Given a set of N calibrated cameras indexed from 0 to $N-1$, camera 0 being the reference camera, we denote by $\mathcal{C}_i$ and $\mathcal{B}_i$ respectively the observed image and a background colour model for camera i. The background colour model is computed automatically by image mosaicking. For each pixel in the reference camera, we would like to estimate, the opacity value α (0 for background, 1 for foreground), and the depth value d for foreground pixels. We want to maximise the posterior probability, or equivalently, the log likelihood

$$L(d,\alpha|\{\mathcal{C}_i,\mathcal{B}_i\}_{i\in[0,N-1]}) = L(\{\mathcal{C}_i,\mathcal{B}_i\}_{i\in[0,N-1]}|d,\alpha) + L(d) + L(\alpha) - L(\{\mathcal{C}_i,\mathcal{B}_i\}_{i\in[0,N-1]}).$$

Description of different terms

- **Likelihood of observations** $L(\{\mathcal{C}_i,\mathcal{B}_i\}|d,\alpha)$. This term is derived from the photo-consistency score computed at the depth d in the case of a foreground hypothesis, or from the similarity to the background colour model in the case of a background hypothesis.
- **Priors on scene geometry** $L(d)$. Firstly, we require surface points to be located inside the conservative visual hull. Secondly, we impose smooth surface variations, by defining a penalty related to the distance between pixel neighbours.
- **Priors on opacity** $L(\alpha)$. Firstly, we require the solution to be consistent with an input trimap which partitions the image into regions of definite background, definite foreground and unknown value. Secondly, we penalise switches from background to foreground according to image contrast; thereby encouraging the segmentation to follow image edges.
- $L(\{\mathcal{C}_i,\mathcal{B}_i\})$ is constant and can be ignored.

Graph construction A global solution is computed via a single graph-cut. The graph is constructed by placing layers of foreground and background nodes along rays defined by the reference camera. Edges located along rays are allocated likelihood costs, while edges located across rays receive costs representing the different priors.



Production Trials

Preliminary test data - England v Azerbaijan football match: 15 fixed cameras (SD, narrow baseline)

Example input image | Virtual view from touch-line | Virtual goal keeper's view

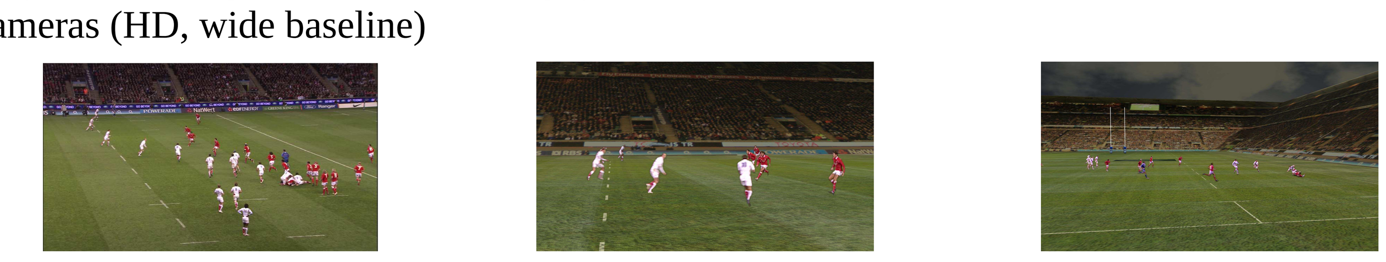
March 2007 - England v Wales (Rugby 6 Nations, Cardiff): 8 fixed cameras (HD, wide baseline)

Example input image | Virtual view from touch-line | Virtual view from try-line

October 2007 - England v Estonia (Euro 2008 qualifier, Wembley): 4 TV broadcast cameras and 2 fixed cameras (HD, wide baseline)

Example input image | Virtual offside view | Another virtual offside view

February 2008 - England v Wales (Rugby 6 Nations, Twickenham): 4 TV broadcast cameras and 6 fixed cameras (HD, wide baseline)



Example input image | Virtual close-up view | Virtual view from try-line

Acknowledgements

This work is supported by the TSB project 'iview: Free-viewpoint Video for Interactive Production' TP/3/DSM/6/I/15515 and EPSRC Grant EP/D033926/1. This work was done in collaboration with BBC Research, Snell & Wilcox and Hawk-Eye Innovations.

References

[1] J.-Y. Guillemaut, A. Hilton, J. Starck, J. Kilner, O. Grau. A Bayesian Framework for Simultaneous Matting and 3D Reconstruction. *Proc. Int. Conf. 3D Digital Imaging and Modeling (3DIM07)*, pages 167–174, Montréal, Québec, Canada, August 2007.
[2] J. Kilner, J. Starck, A. Hilton, O. Grau. Dual-Mode Deformable Models for Free-Viewpoint Video of Outdoor Sports Events. *Proc. Int. Conf. 3D Digital Imaging and Modeling (3DIM07)*, pages 177–184, Montréal, Québec, Canada, August 2007.