Specialized depth extraction for live soccer

Specialized Depth Extraction for Live Soccer Video

Luc Vosters

Axon Digital Design

Eindhoven, University of Technology

November 18, 2010

3D

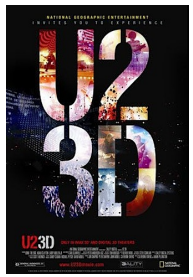
- 3D Cinema.

- 3D TV sets.

- 3D broadcast.

sky 3D

- 3D Live Events.



3D Productions

- 3D cameras.

- 2D-to-3D conversion.
  - Offline (semi-) automatic.
  - Real-time.

3D Productions

- 3D cameras.

- 2D-to-3D conversion.
  - Offline (semi-) automatic.
  - Real-time.

Why 2D-To-3D conversion?

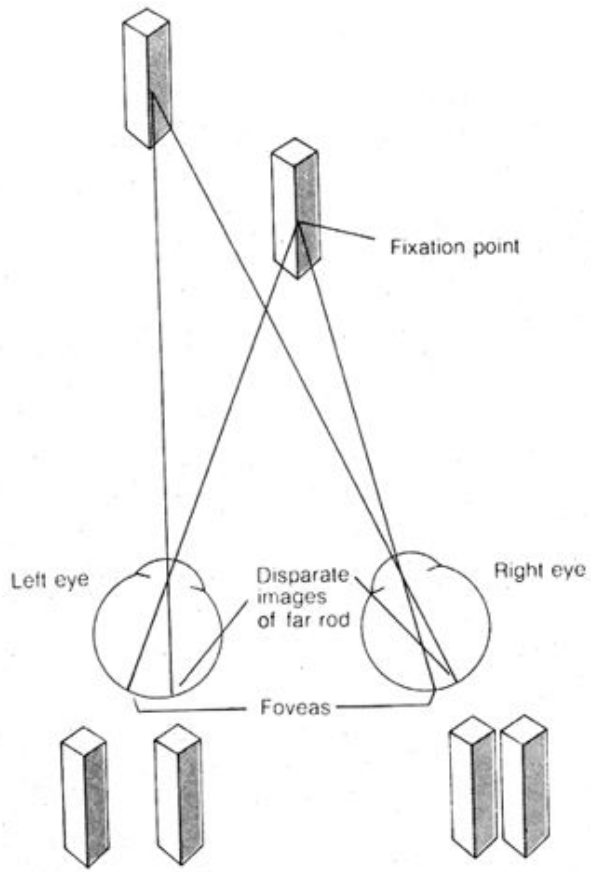
2D-to-3D Conversion

- *10,000$*
- Widely available.
- No camera rig.
- No stereographer.

3D Recording

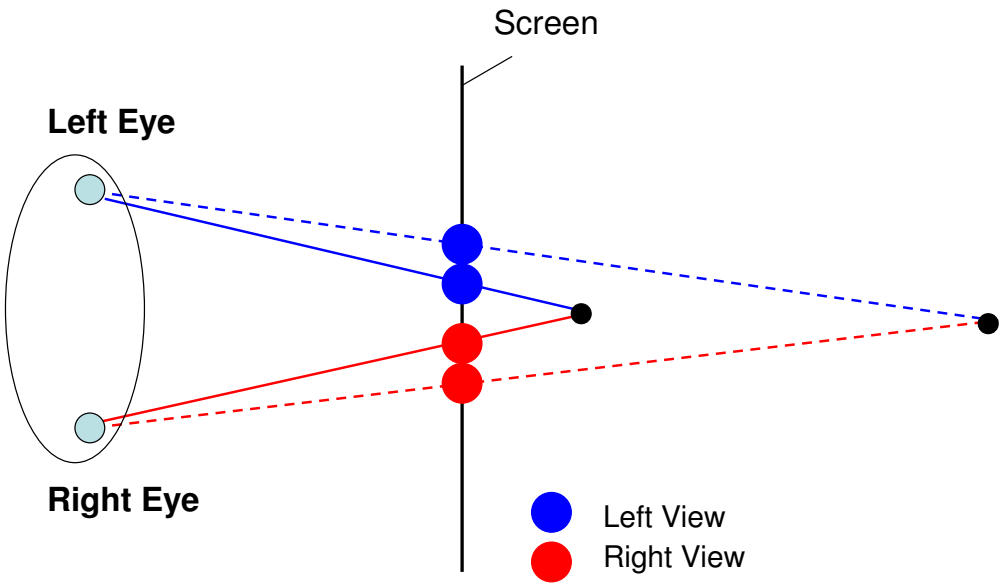
- *80,000$*
- Investment.

Binocular Disparity



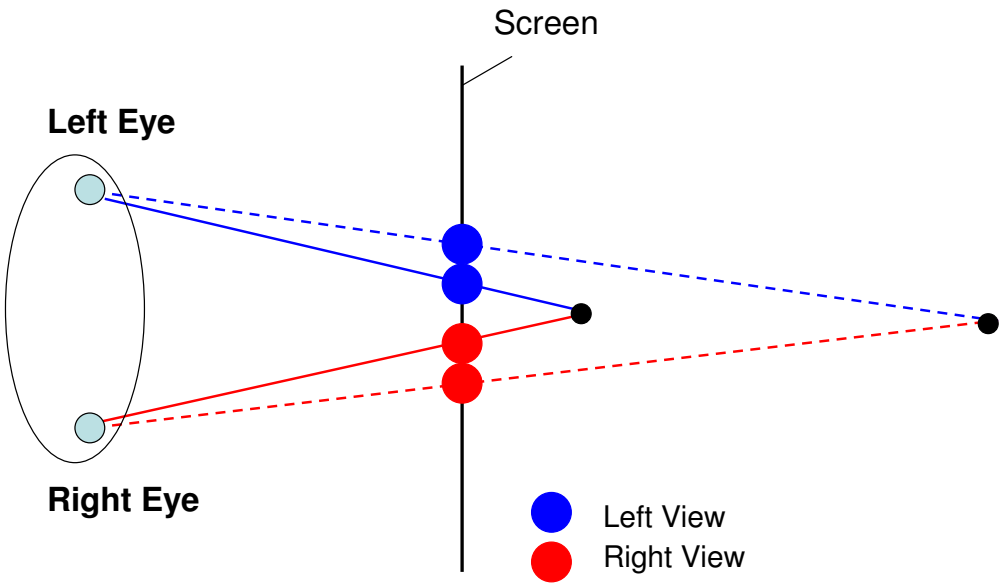
From Depth to Stereo 3D

- Depth $\leftrightarrow$ disparity.
- 2D-To-3D:
  1. Extract depth.
  2. Calculate disparity.
  3. Render Left/Right image.
- Occlusion handling.



From Depth to Stereo 3D

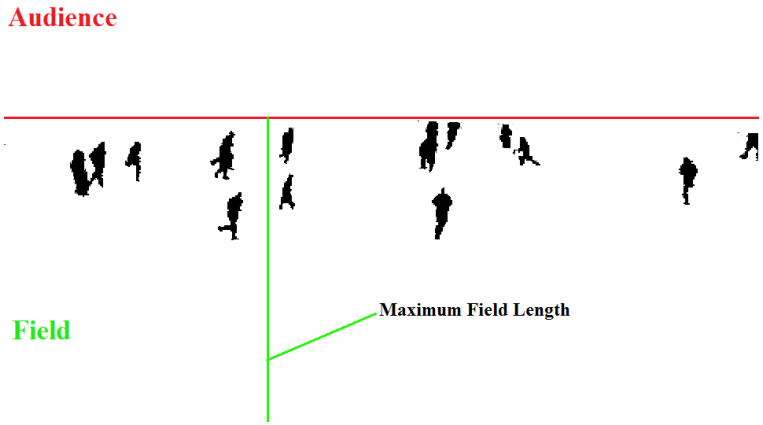
- Depth $\leftrightarrow$ disparity.
- 2D-To-3D:
  1. Extract depth.
  2. Calculate disparity.
  3. Render Left/Right image.
- Occlusion handling.

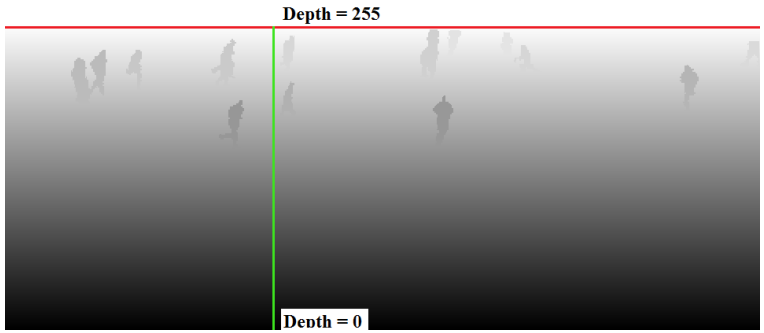


Long Shot Images in Live Soccer.

Jung et al. [1]


DEN 1-0 GENK
10:20
ALVODAK
VERHELST
RANTOUR
RANTOUR
RANTOUR


Audience
Field
Maximum Field Length


Depth = 255
Depth = 0

Jung et al. cont.

Advantages

- No occlusions.
- Few computations.

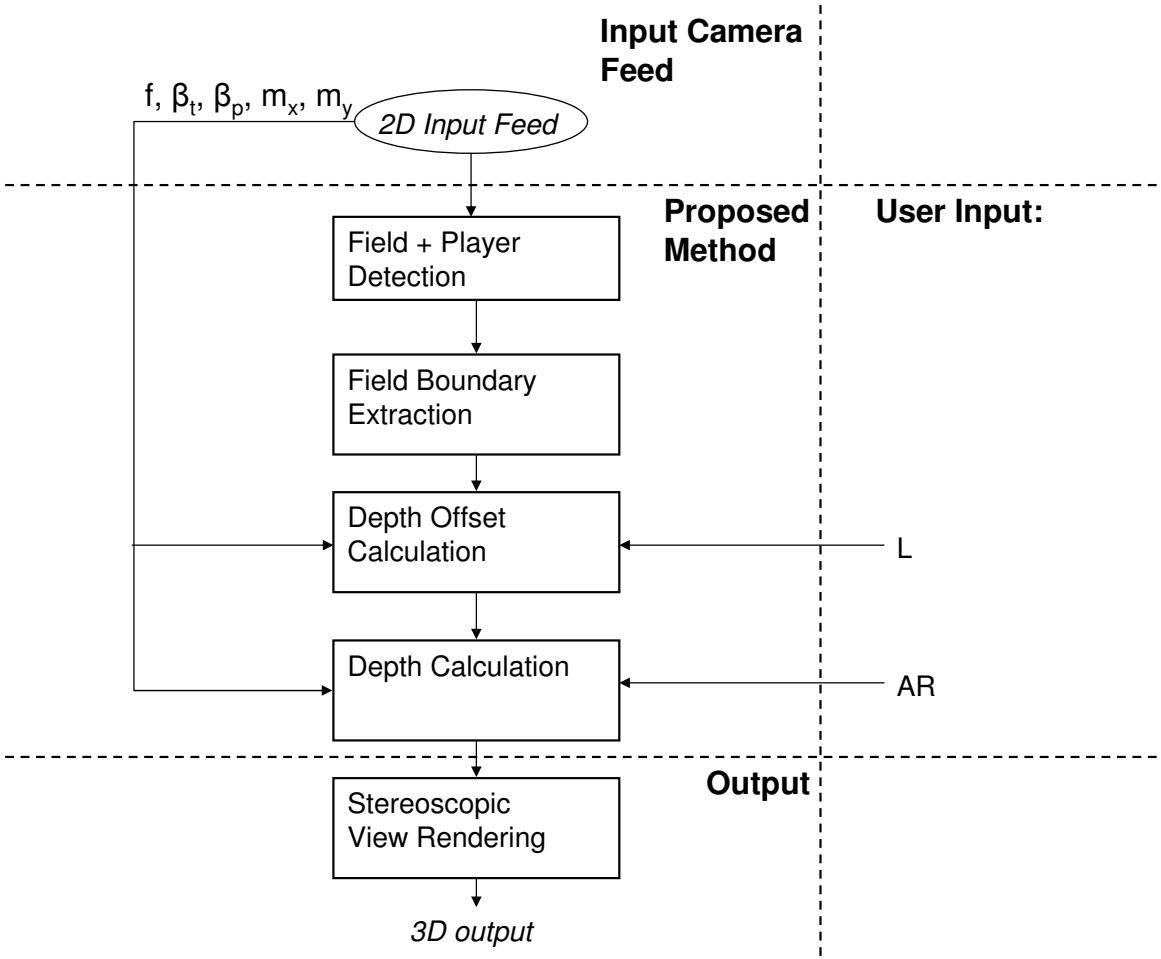
Disadvantages

- Pan, tilt, zoom not modeled.
- Audience depth constant.
- No depth offset.

Proposed Approach

- Field and audience depth model.
- Exploit *Camera* + *Scene* information.
  - Focal length
  - Tilt/Pan angle
  - Image sensor size
  - Average player length

Image Model

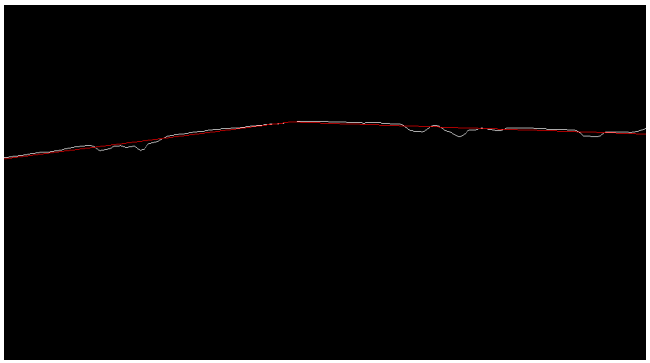

Input Camera Feed
f, β_t, β_p, m_x, m_y
2D Input Feed
Proposed Method
User Input:
Field + Player Detection
Field Boundary Extraction
Depth Offset Calculation
L
Depth Calculation
AR
Output
Stereoscopic View Rendering
3D output

Field Detector (Seo et al. [2])

- Train Hue, Sat, Val histograms.
- Extract *PeakValueIndex*, *SaturationMean*.

$$\textit{Field}, \quad \text{if} \begin{cases} G > 0.95 \cdot R \\ R > 0.95 \cdot B \\ V < 1.25 \cdot \textit{PeakValueIndex} \\ S > 0.8 \cdot \textit{SaturationMean} \end{cases} \tag{1}$$

Field Boundary Extraction

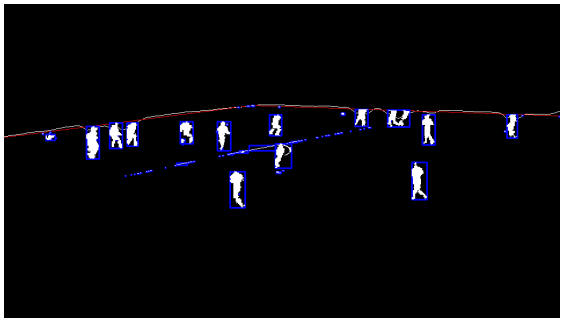


- Piecewise linear curve fit. (Cantoni [3]).
  - Specify start, end and line intersection.

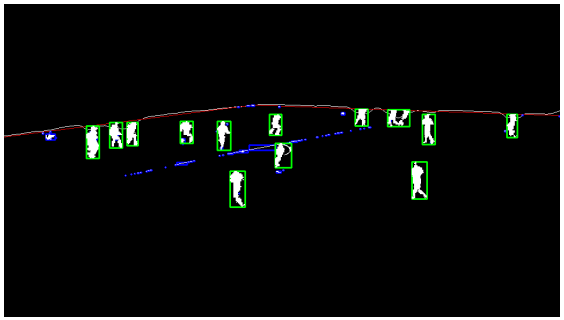
Player Detection

Input image

Connected Components



$$\textit{Player} \quad \textit{if} \begin{cases} 0.4 < & \textit{AspectRatio} & < 3 \\ & \#\textit{Pixels} & > 0.05MN \\ & \frac{BBwidth \cdot BBheight}{\#Pixels} & > 0.25 \end{cases}$$

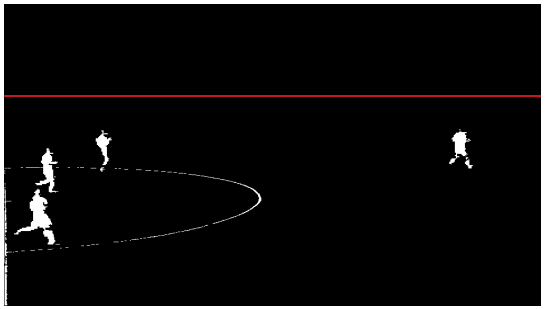


Player Detection Cont.

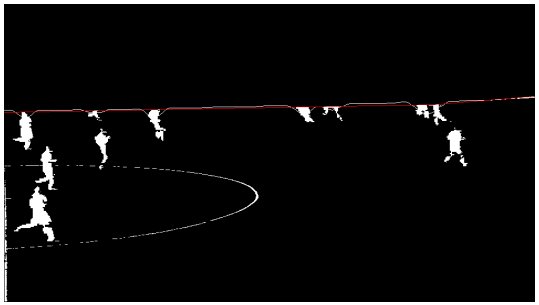
- Field Boundary Extraction aides Player Segmentation.

Input

Jung et al. [1]

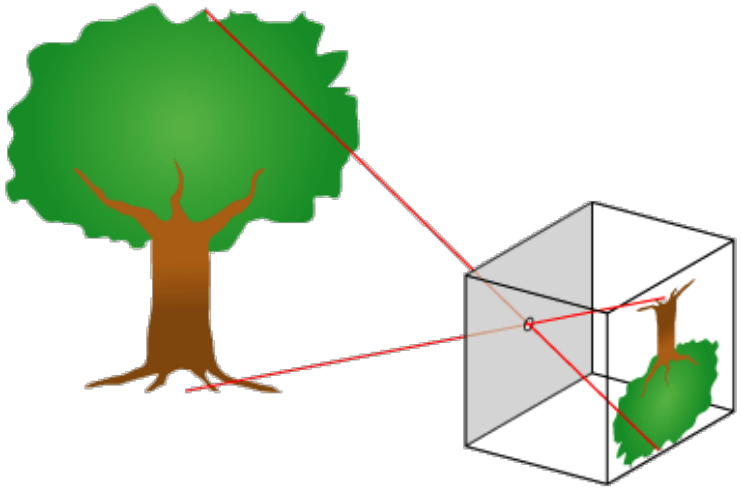


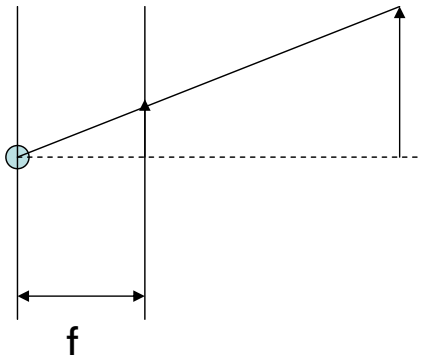
Proposed



Camera depth model

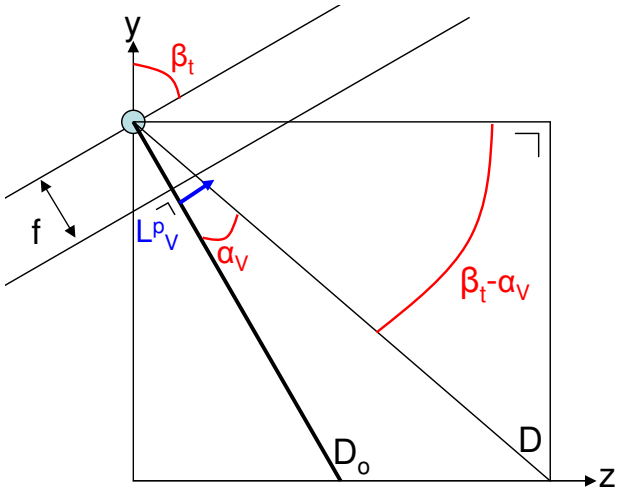
- Long shot camera
  - Small lens aperture.
  - High depth of field.
- Pin-hole camera.





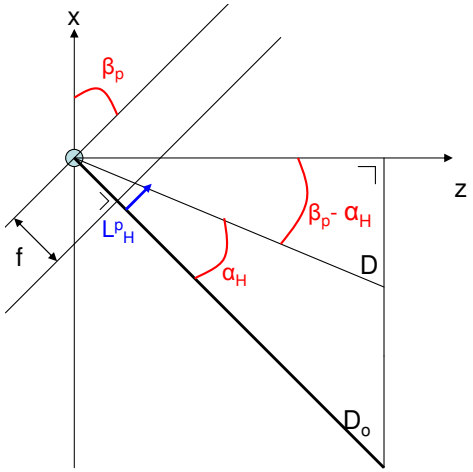
Field Depth

Depth in YZ-plane:



$$D = D_o \frac{\sin|\beta_t|}{\sin(|\beta_t| - \alpha_V)}$$

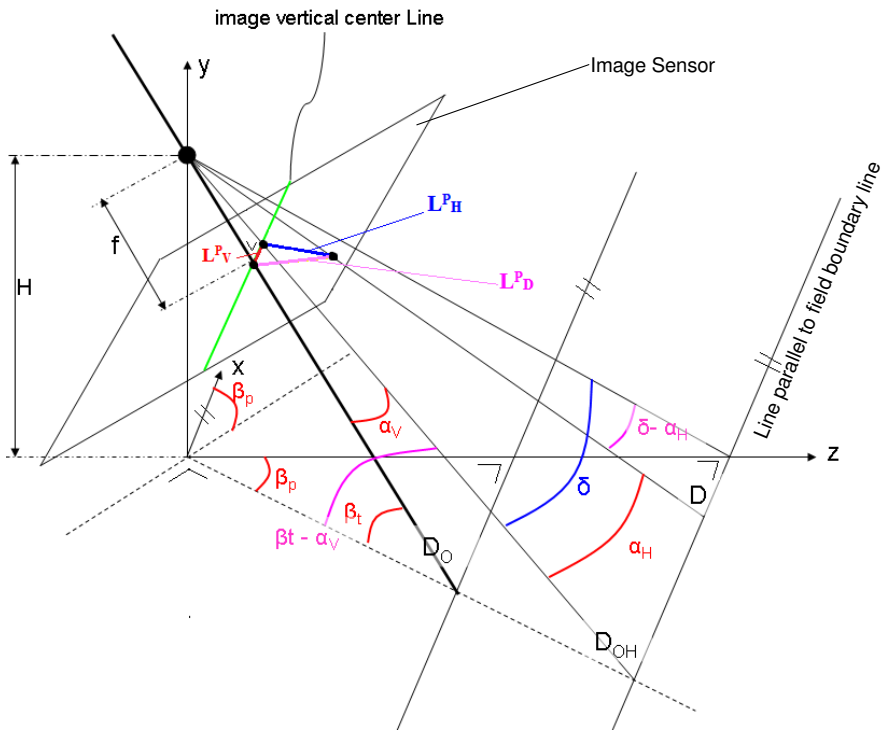
Depth in XZ-plane:



$$D = D_o \frac{\cos|\beta_p|}{\cos(|\beta_p| - \alpha_H)}$$

Field Depth cont.

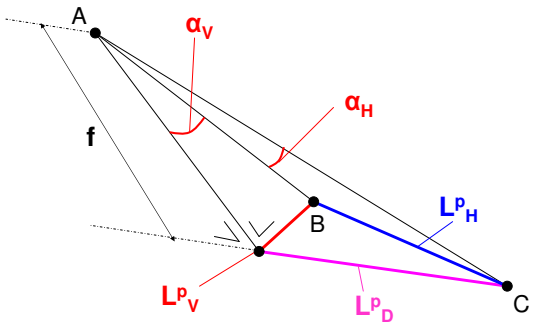
Depth in XYZ-Plane.



$$D_{field}(\alpha_V, \alpha_H) =$$

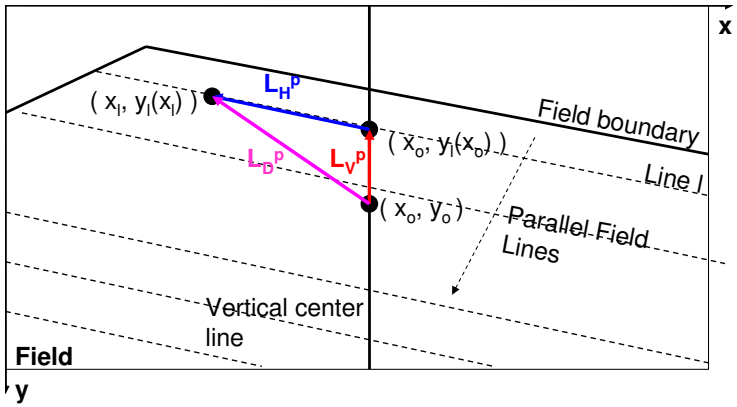
$$D_o \frac{\sin|\beta_t|}{\sin(|\beta_t| - \alpha_V)} \frac{\cos|\beta_p|}{\cos(|\beta_p| - \alpha_H)}$$

Field Depth Calculation



$$\alpha_V = \arctan\left(\frac{L_V^P}{f}\right)$$

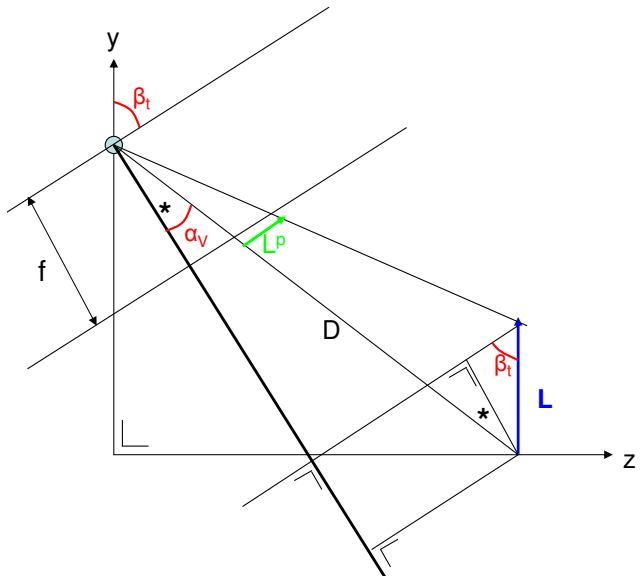
$$\alpha_H = \arccos\left(\frac{2f^2 + L_V^{P2} + L_D^{P2} - L_H^{P2}}{2\sqrt{(f^2 + L_V^{P2})(f^2 + L_D^{P2})}}\right)$$



- Depth offset unknown.

Depth Offset

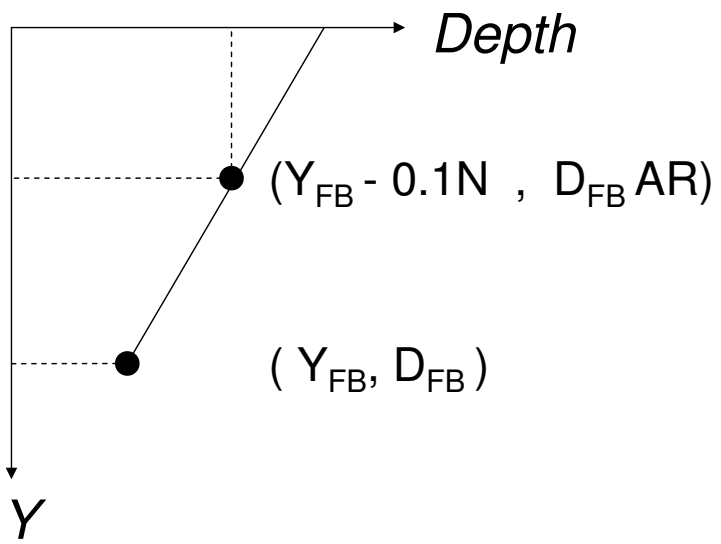
- Player length $\approx 1.80m$
- $L^P \leftrightarrow$ Tilt, f, Depth, L.
- $D_{o,i}$ depth offset player i.
- $D_o = median\{D_{o,i} | \forall i\}$



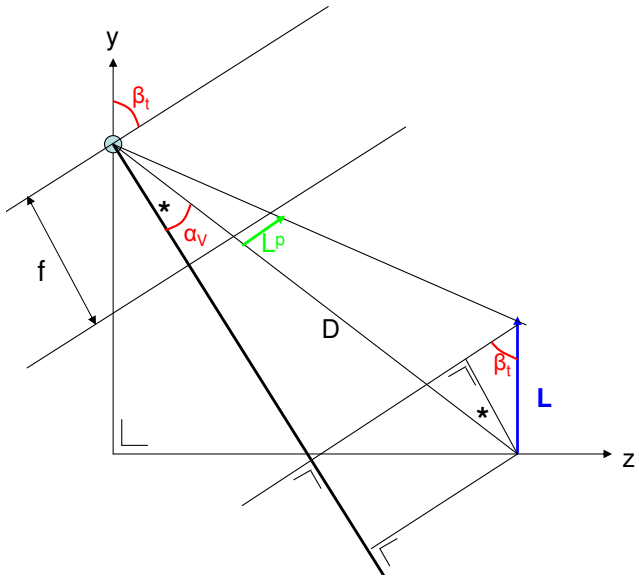
$$D = L \cdot \frac{\sin \beta_t \left[L^P \cdot m_y + f \tan \alpha_V \right] + f \cos \beta_t}{\cos \alpha_V [L^P \cdot m_y + f \tan \alpha_V] - f \sin \alpha_V}$$

Audience Depth

$$D_{aud}(y) = \frac{D_{FB} \cdot AR}{0.1N}(y - y_{FB}) + D_{FB} \ ,$$



Player Depth



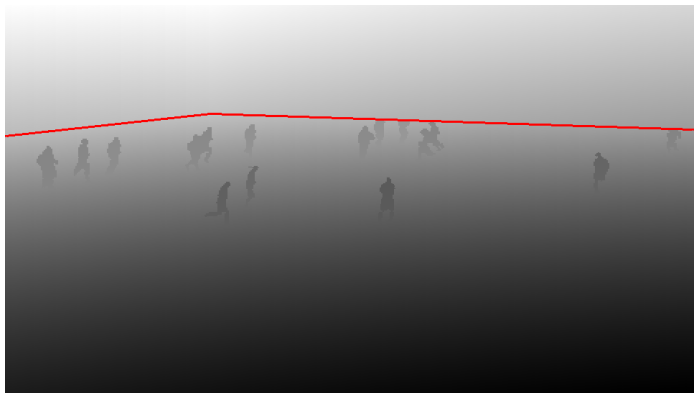
- Player approximately constant.
- $D_{player} = D_{field}(x_i, y_i)$.

Complete Camera Depth Model

- Depth Map Quantization.

$$DEPTH(x,y) = 255 \cdot \left(\frac{D(x,y) - D_{min}}{D_{max} - D_{min}} \right)$$

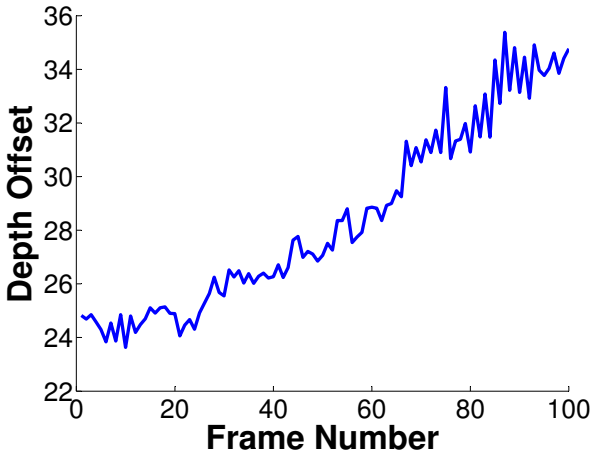
- Clip outside $[D_{min}, D_{max}]$.



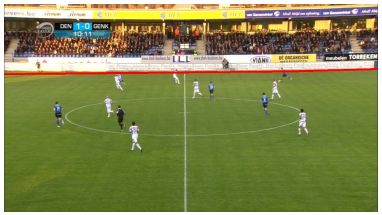
Depth Offset Calculation

- Zooming

- Keep focal length constant in depth offset calculation.



Qualitative Comparison

Input	Jung et al. [1]	Proposed
		
		
		
		

Conclusion

- Jung et al. [1] proposed specialized 2D-3D conversion of long shot images.

- We propose more advanced model:
  1. *zooming*, *panning* and *tilting* modeled.
  2. Calculates depth offset from player length.
  3. Improved player segmentation by field boundary extraction.

- Few computations.
- Hardware attractive.
- Future work: perceptual tests.

Questions

Thanks for your attention.

- Y. J. Jung, C. Kim, D. Park, Y. Kim, and J. Ko, “Method, medium, and system for generating depth map of a video image,” U.S. Patent US20 090 196 492, August 6, 2009.
- K. Seo, J. Ko, I. Ahn, and C. Kim, “An intelligent display scheme of soccer video on mobile devices,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 10, pp. 1395–1401, October 2007.
- A. Cantoni, “Optimal curve fitting with piecewise linear functions,” *IEEE Transactions on Computers*, vol. C-20, no. 1, pp. 59–67, Januari 1971.