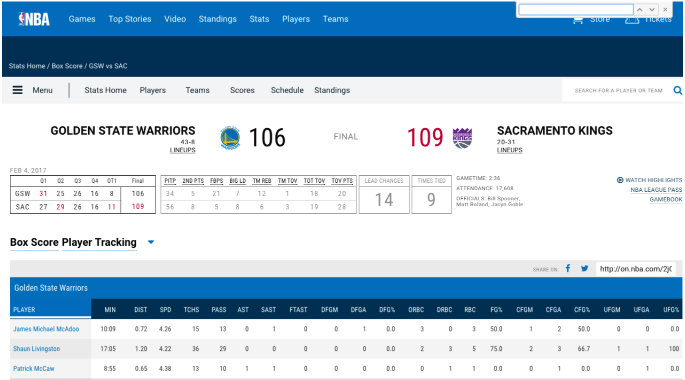
Sports Field Localization

Namdar Homayounfar

February 5, 2017

Motivation

Sports Field Localization: Have to figure out where the field and players are in 3d space in order to make measurements and generate statistics.



NBA Games Top Stories Video Standings Stats Players Teams Store Tickets

Stats Home / Box Score / GSW vs SAC

Menu | Stats Home Players Teams Scores Schedule Standings | SEARCH FOR A PLAYER OR TEAM

GOLDEN STATE WARRIORS 43-8 LINEUPS 106 FINAL 109 SACRAMENTO KINGS 20-31 LINEUPS

FEB 4, 2017

	Q1	Q2	Q3	Q4	OT1	Final
GSW	31	25	26	16	8	106
SAC	27	29	26	16	11	109

PITP	2ND PTS	FBPS	BIG LD	TM REB	TM TOV	TOT TOV	TOV PTS
34	5	21	7	12	1	18	20
56	8	5	8	6	3	19	28

LEAD CHANGES 14

TIMES TIED 9

GAMETIME: 2:36
ATTENDANCE: 17,608
OFFICIALS: Bill Spooner, Matt Boland, Jacyn Goble

WATCH HIGHLIGHTS
NBA LEAGUE PASS
GAMEBOOK

Box Score Player Tracking

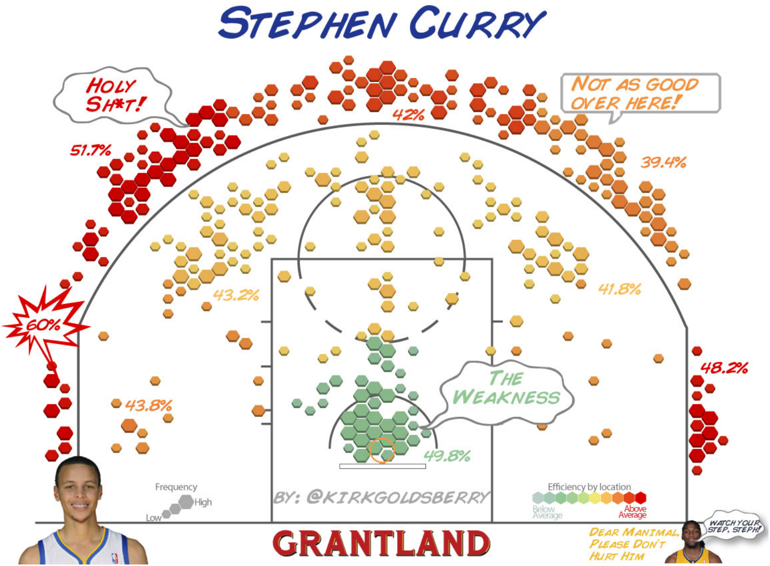
SHARE ON: http://on.nba.com/2j(

Golden State Warriors

PLAYER	MIN	DIST	SPD	TCHS	PASS	AST	SAST	FTAST	DFGM	DFGA	DFG%	ORBC	DRBC	RBC	FG%	CFGM	CFGA	CFG%	UFGM	UFGA	UFG%
James Michael McAdoo	10:09	0.72	4.26	15	13	0	1	0	0	1	0.0	3	0	3	50.0	1	2	50.0	0	0	0.0
Shaun Livingston	17:05	1.20	4.22	36	29	0	0	0	0	0	0.0	2	3	5	75.0	2	3	66.7	1	1	100
Patrick McCaw	8:55	0.65	4.38	13	10	1	1	0	0	0	0.0	0	1	1	0.0	0	1	0.0	0	1	0.0

`http://stats.nba.com/`

Motivation



a.espncdn.com

How is this done in practice?

- There are various tracking and analytics companies
- Their solutions for field localization is based mostly on hardware

http://www.stats.com/sportvu/sportvu-basketball-media/

How is this done in practice?

Real-Time Objects Tracking and Motion Capture in Sports Events

US 20080192116 A1

ABSTRACT

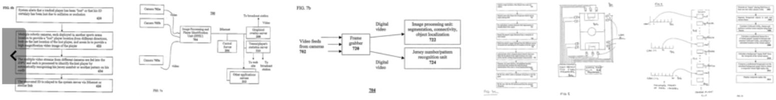
Non-intrusive peripheral systems and methods to track, identify various acting entities and capture the full motion of these entities in a sports event. The entities preferably include players belonging to teams. The motion capture of more than one player is implemented in real-time with image processing methods. Captured player body organ or joints location data can be used to generate a three-dimensional display of the real sporting event using computer games graphics.

Publication number	US20080192116 A1
Publication type	Application
Application number	US 11/909,080
PCT number	PCT/IL2006/000388
Publication date	Aug 14, 2008
Filing date	Mar 29, 2006
Priority date	Mar 29, 2005
Also published as	EP1864505A2, EP1864505A4, WO2006103662A2, WO2006103662A3
Inventors	Michael Tamir, Gal Oz
Original Assignee	Sportvu Ltd.
Export Citation	BiBTeX, EndNote, RefMan

Patent Citations (13), Referenced by (96), Classifications (8), Legal Events (1)

External Links: USPTO, USPTO Assignment, Espacenet

IMAGES (16)



http://www.google.com/patents/US20080192116

How is this done in practice?

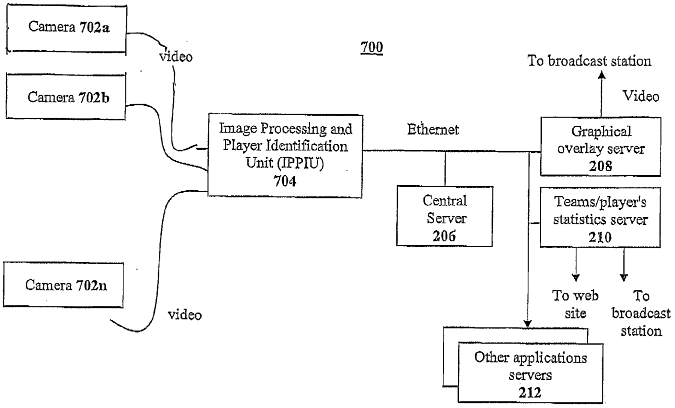
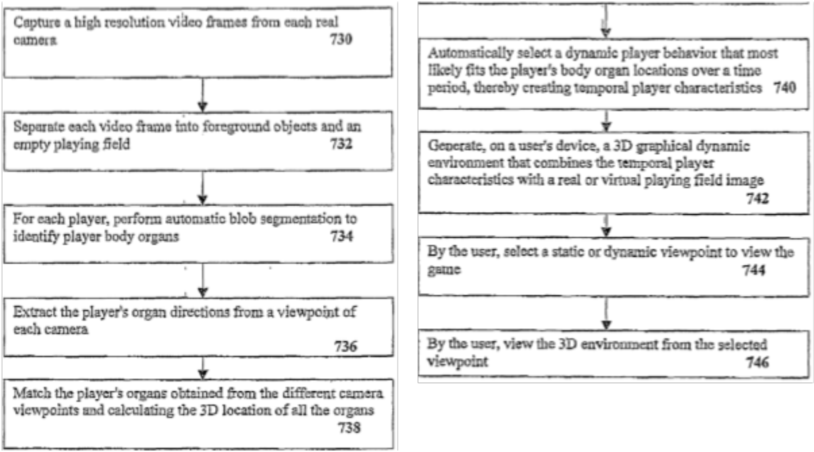


FIG. 7a

http://www.google.com/patents/US20080192116

How is this done in practice?



http://www.google.com/patents/US20080192116

How is this done in practice?

- There are various tracking and analytics companies
- Their solutions for field localization is based mostly on hardware

http://pixellot.tv/

How is this done in practice?

- There are various tracking and analytics companies
- Their solutions for field localization is based mostly on hardware

http://www.catapultsports.com/

Drawbacks

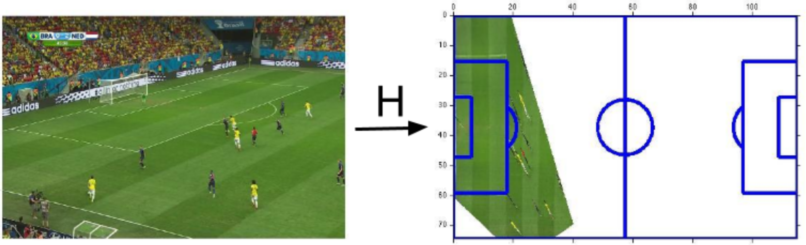
- Very expensive: e.g. Sportvue costs $>$ \$100000 per season for a team
- Only rich teams can afford them
- Have to maintain all the hardware
- Still not bulletproof. Require workers to fix mistakes

Simpler Solution?

Can we get rid of all these cameras/gps systems and just figure out where the players are by looking at a broadcast image of the field?

Simpler Solution? YES!

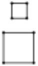
Goal: Given a single broadcast image of a sport game, such as soccer, can we localize it?



- H is a 3×3 invertible matrix with 8 d.f. Called a projective transformation/homography

Homography Matrix

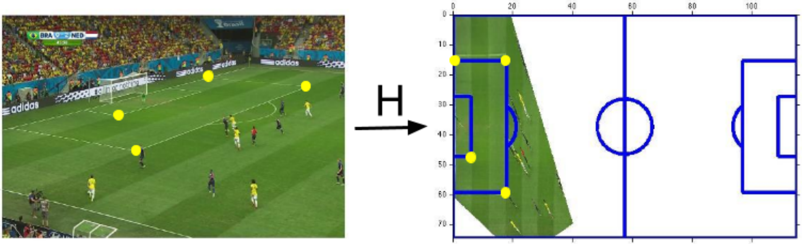
The matrix H captures all the following transformations:

Group	Matrix	Distortion	Invariant properties
Projective 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$		Concurrency, collinearity, **order of contact**: intersection (1 pt contact); tangency (2 pt contact); inflections (3 pt contact with line); tangent discontinuities and cusps. cross ratio (ratio of ratio of lengths).
Affine 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (e.g. midpoints), linear combinations of vectors (e.g. centroids). The line at infinity, $\mathbf{l}_\infty$.
Similarity 4 dof	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Ratio of lengths, angle. The circular points, **I**, **J** (see section 2.7.3).
Euclidean 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Length, area

Multiple View Geometry by Hartley and Zisserman

How to find H?

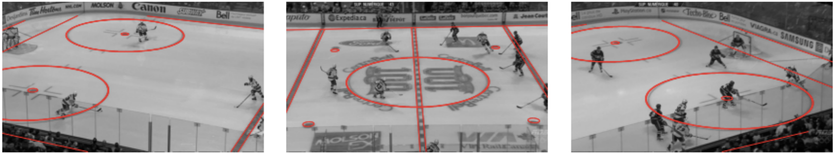
- Require 4 point correspondences



- Difficulty arises in finding the 4 corresponding points

Related Work: Academia

- Hess et al., Improved Video Registration using Non-Distinctive Local Image Features, 2007
- Gupta et al. Using Line and Ellipse Features for Rectification of Broadcast Hockey Video, 2011

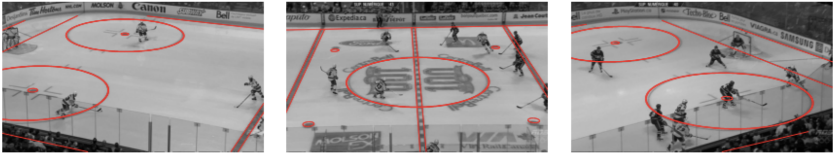


Key-frame 1 Key-frame 3 Key-frame 5

Gupta et al. 2011

Related Work: Academia

- Based on Keyframes and old school computer vision transforms (eg. SIFT)



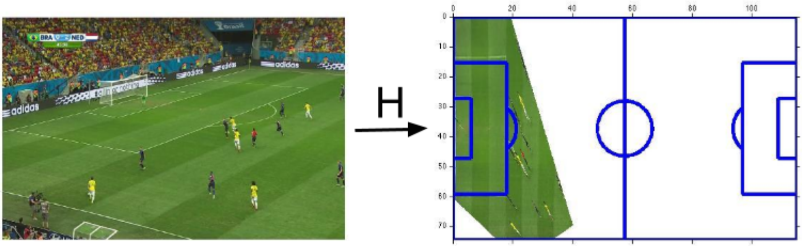
Key-frame 1 Key-frame 3 Key-frame 5

Gupta et al. 2011

- Limitation: Depends on fixed features and also requires manual annotation of keyframes for each game and

Can we do better?

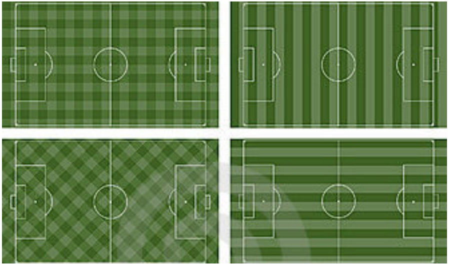
- Can we automatically localize the field from a broadcast image?



- Lets come up with a learning approach
- Based on joint work with Sanja Fidler and Raquel Urtasun submitted to CVPR

In case of Soccer

- Large dimensions and exposed to the elements
- Different grass textures and patterns

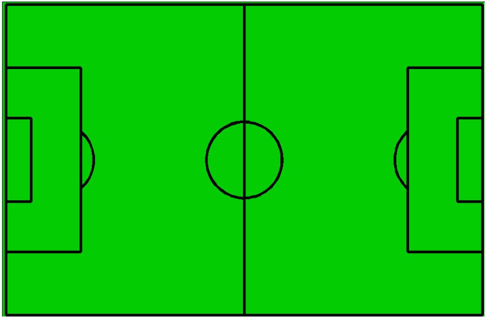


This Work

- Introduce a parametrization of the field
- Incorporate prior knowledge about the soccer field as potentials in an CRF
- Find the mapping H implicitly by doing inference in the CRF
- Single Camera, No key-frame, Fast Inference

Methodology

- Let $x \in \mathcal{X}$ be random variable corresponding to a broadcast image of a soccer field.
- A soccer field is restricted by two long sides referred to as **touchlines** and two shorter sides referred to as **goallines**



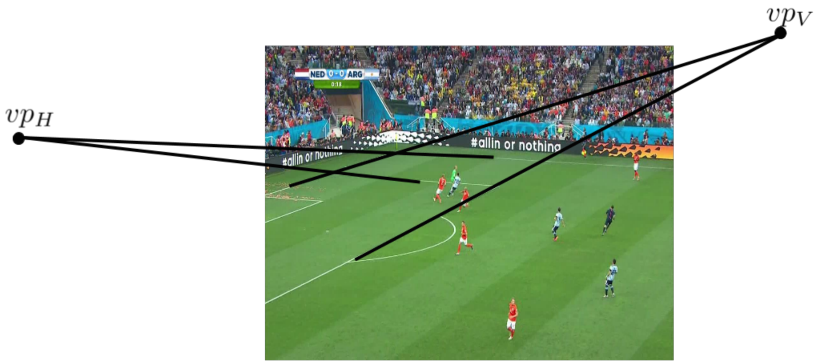
- We aim to infer the position of the touchlines and the goallines in the image x. (Not all visible at the same time)

Methodology

Methodology: Parametrization

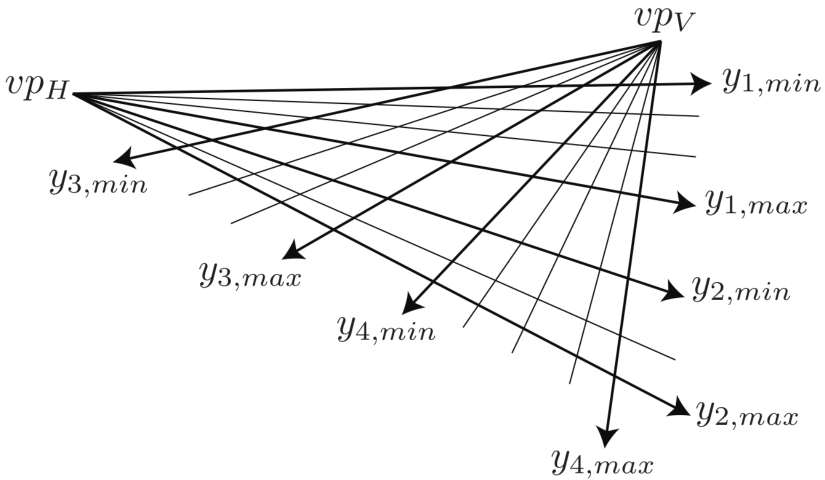
- It's very important how we parametrize this problem
- What are we trying to find?
- **Vanishing Points:** Where parallel lines meet in the image
- **Manhattan World Assumption:** Existence of three dominant orthogonal vanishing points in human-made scenes.
- In a soccer field we usually have clues for the two orthogonal vanishing point

Methodology: Parametrization

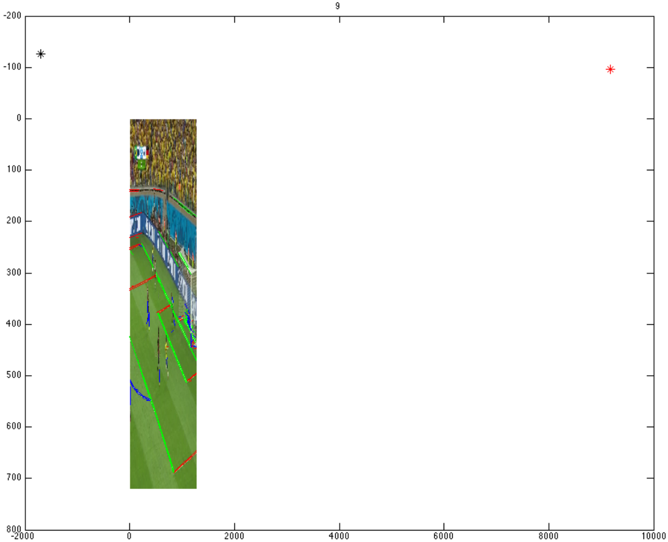


Methodology: Parametrization

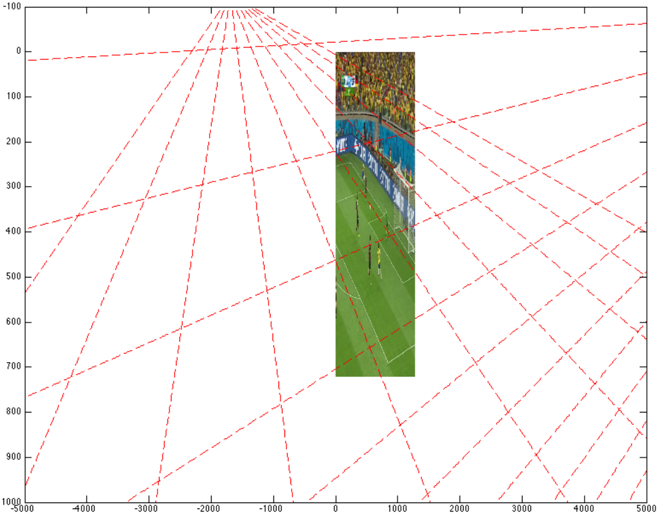
- Create a grid by emanating rays from the vanishing points



Methodology: Parametrization

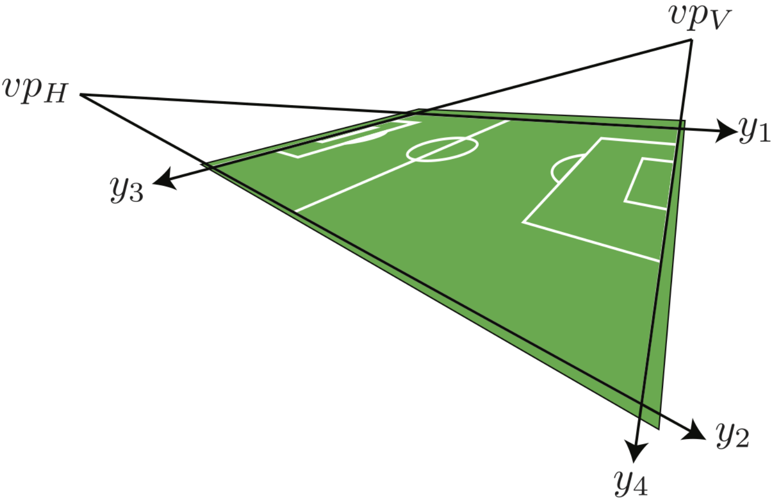


Methodology: Parametrization

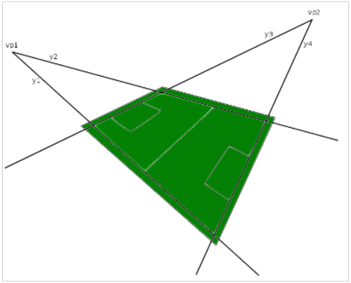


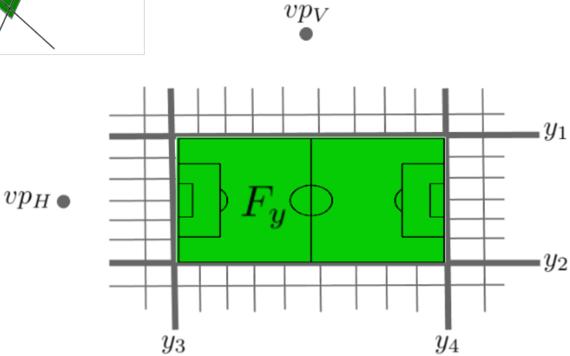
Methodology: Parametrization

- We parametrize the soccer field by four rays $y = (y_1, y_2, y_3, y_4)$ on the grid
- State space: $\mathcal{Y} = \prod_{i=1}^{4} \left\{ [y_{i,min}^{init}, y_{i,max}^{init}] \right\} \subset \mathbb{N}^4$



Methodology: Parametrization





Inference Task

Given an image x of the field, obtain the best prediction of the touchlines and the goallines by solving the following inference task:

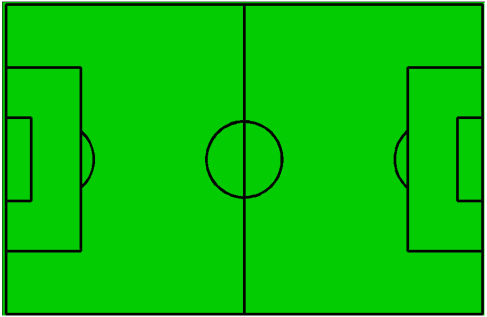
$$\hat{y} = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y)$$

- $\phi(x, y)$: feature vector
- w: weights to be learned from training data
- **Note:** $|\mathcal{Y}| \propto (\#\text{rays from } vp_H)^2 (\#\text{rays from } vp_V)^2$

We find an exact solution by using branch and bound for inference. More on that later

Model: Features

A soccer field is made up of grass and there are white marking corresponding to lines and circles with fixed dimensions



We incorporate these as features.

Model: Features

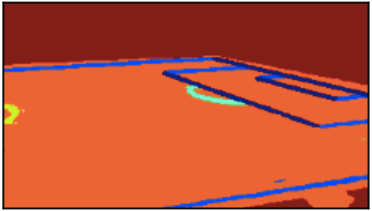
- We need good features in the presence of noise
- Different weather and lighting conditions and shadows
- Methods based on heuristics are very fragile

Model: Features - Semantic Segmentaition

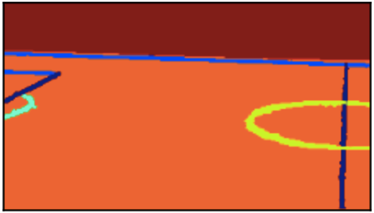
Train a semantic segmentation network to classify image pixels to either belonging to:

1. Vertical Lines
2. Horizontal Lines
3. Middle Circle
4. Side Circles
5. Grass
6. Outside

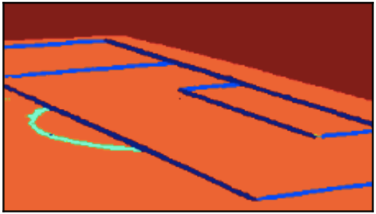
Model: Features - Some Examples



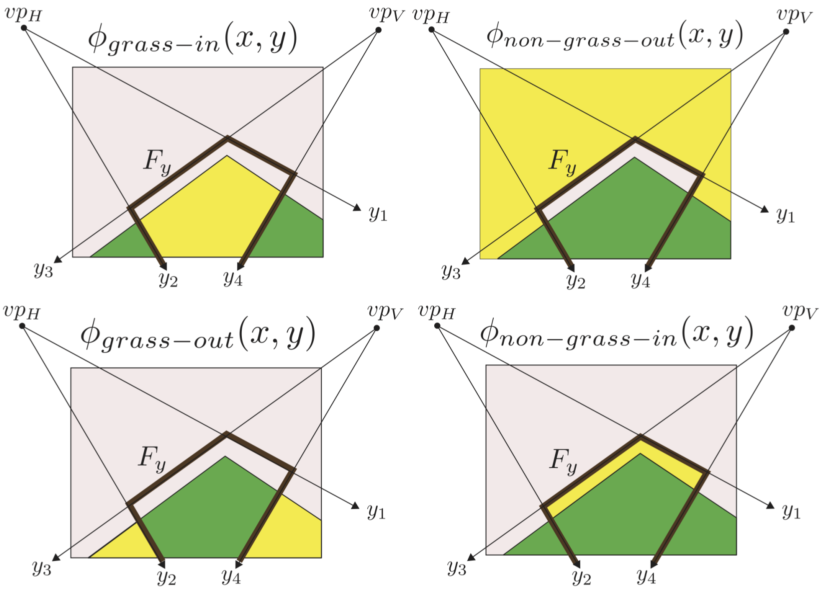
Model: Features - Some Examples



Model: Features - Some Examples

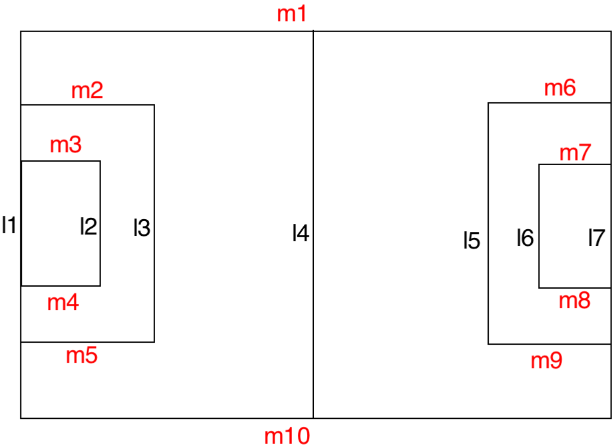


Model: Features — Grass



Model: Features — Lines

7 vertical line segments corresponding to vp_V and 10 horizontal line segments corresponding to vp_H

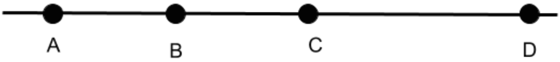


Model: Features — Lines

- Given y, need to construct a potential function that is large when the projection of each line segment in the image x is close to its ground truth.

- But given y, where does each line segment fall in the image x?
- Use **Cross Ratios:** Given 4 points A, B, C, D on a line, their cross ratio is given by:

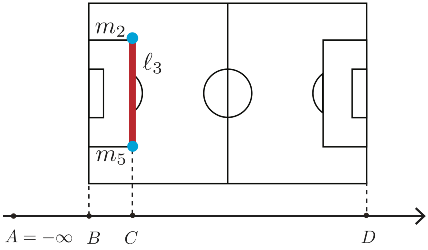
$$CR(A, B, C, D) = \frac{|A - C| \cdot |B - D|}{|B - C| \cdot |A - D|}$$

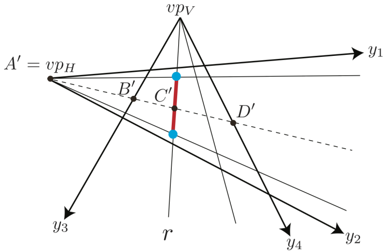


- Cross ratios invariant under any projective transformation.

Model: Features — Lines

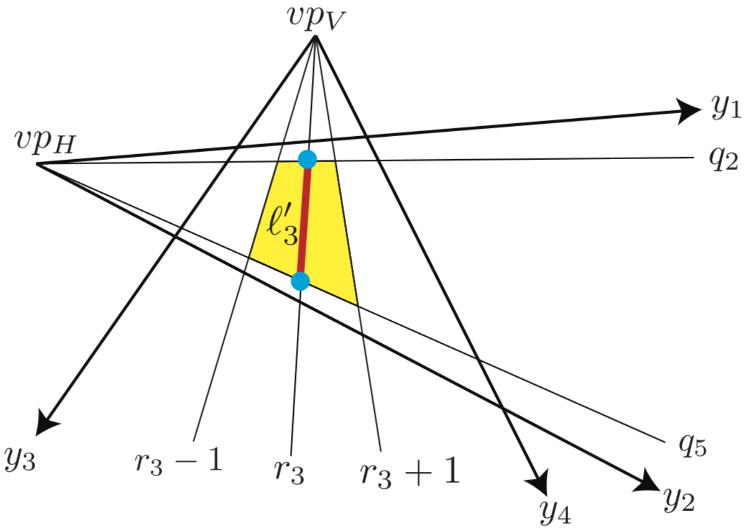
- Use cross ratios to find the position of each line on the grid





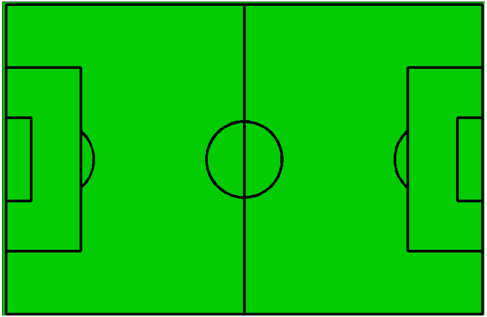
Model: Features — Lines

- For example for line ℓ_3:



Model: Features — Circles

A cemicircle on each side of the field C_2, C_3 and a circle in the middle:



Model: Features — Circles

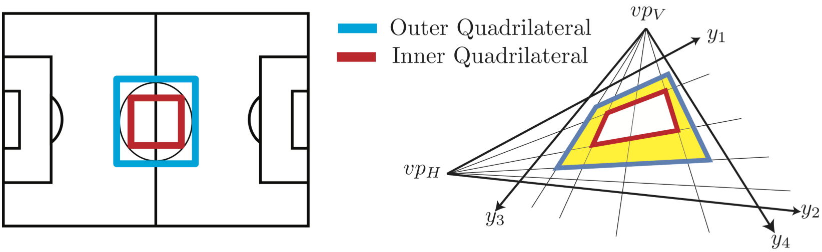
Transformed to ellipses C'_k in x

Model: Features — Circles

- Similar to line potentials, want potential functions that count the fraction of supporting pixels in the image x for each circular shape C_i given a hypothesis field y
- Unlike lines, the ellipses don't fall on the grid.
- Ellipse detection: slow and unreliable

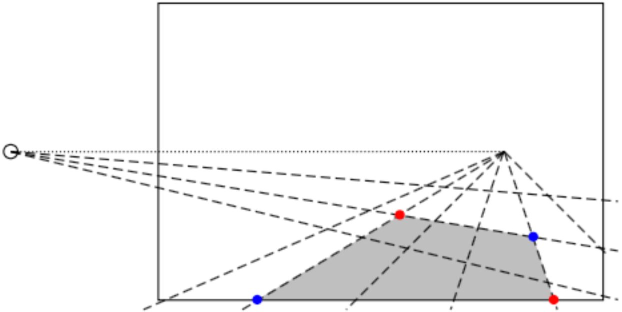
Model: Features — Circles

- For each circle there are unique **inscribing** and **circumscribing** rectangles aligned with the vanishing points.
- Similar to lines, we can find the quadrilaterals associated with these rectangles on the grid $\mathcal{Y}$.



Model: Features — Efficient Computation

- We have positive features.
- Can use 3d accumulators to compute the potentials efficiently.



Schwing et al. 2012

Branch and Bound Inference

- Inference task

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y)$$

- Aim to do it efficiently and exactly
- Exactness comes from using branch and bound
- Efficiency comes from using integral images and tight upper bounds in branch and bound

Branch and Bound Inference — 3 Ingredients

Suppose $Y \subset \mathcal{Y} = \prod_{i=1}^{4} \left\{ [y_{i,min}^{init}, y_{i,max}^{init}] \right\}$ is a subset of parametrized fields. Branch and bound needs

- A branching mechanism that divides the set Y into two disjoint subsets Y_1 and Y_2 of parametrized fields.
- A set function $\bar{f}$ such that $\bar{f}(Y) \geq \max_{y \in Y} w^t \phi(x, y)$.
- A priority queue PQ which orders sets of parametrized fields Y according to $\bar{f}$.

Branch and Bound Inference — Optimality

In order to guarantee the optimality of the converged solution:

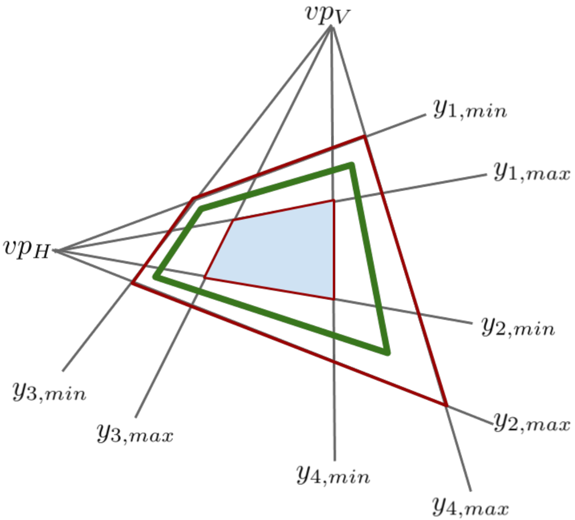
1. $\bar{f}(Y) \geq \max_{y \in Y} w^t \phi(x, y)$ for any arbitrary $Y \in \mathcal{Y}$
2. $\bar{f}(Y) = w^t \phi(x, y)$ when $Y = \{y\}$

Algorithm 1 branch and bound (BB) inference

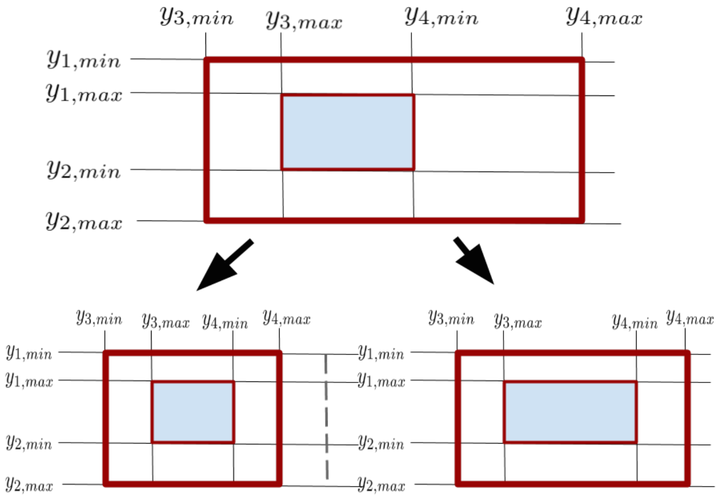
put pair $(\bar{f}(\mathcal{Y}), \mathcal{Y})$ into queue and set $\hat{\mathcal{Y}} = \mathcal{Y}$
repeat
  split $\hat{\mathcal{Y}} = \hat{\mathcal{Y}}_1 \times \hat{\mathcal{Y}}_2$ with $\hat{\mathcal{Y}}_1 \cap \hat{\mathcal{Y}}_2 = \emptyset$
  put pair $(\bar{f}(\hat{\mathcal{Y}}_1), \hat{\mathcal{Y}}_1)$ into queue
  put pair $(\bar{f}(\hat{\mathcal{Y}}_2), \hat{\mathcal{Y}}_2)$ into queue
  retrieve $\hat{\mathcal{Y}}$ having highest score
until $|\hat{\mathcal{Y}}| = 1$

Branch and Bound Inference — Branching

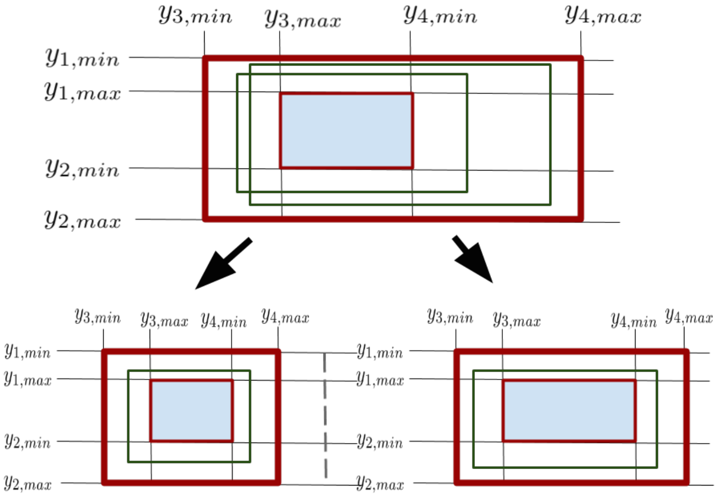
- How to branch a set $Y = \prod_{i=1}^{4} \{[y_{i,min}, y_{i,max}]\} \subset \mathcal{Y}$ into two disjoint sets Y_1 and Y_2



Branch and Bound Inference — Branching



Branch and Bound Inference — Branching



Branch and Bound Inference — Bounding

- Decompose $\phi(x, y)$ into potential with strictly positive weights and those with weights that are either zero or negative:

$$w^T \phi(x, y) = w_{neg}^T \phi_{neg}(x, y) + w_{pos}^T \phi_{pos}(x, y)$$

- Construct a lower bound set function $\bar{\phi}_{i,neg}$ and an upper bound set function $\bar{\phi}_{j,pos}$ such that

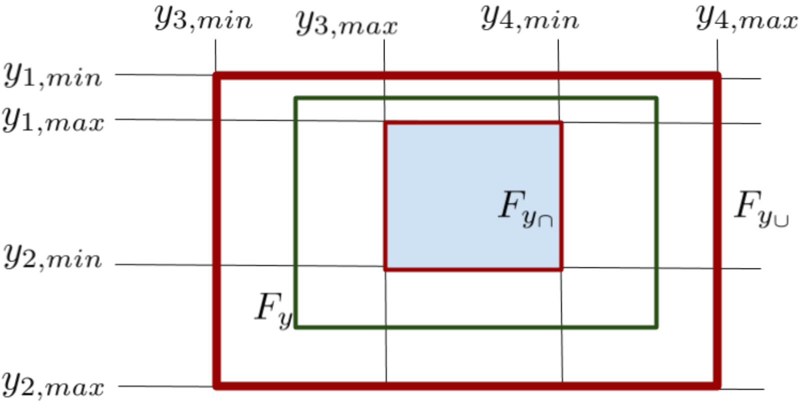
$$\begin{aligned} \bar{\phi}_{i,neg}(x, Y) &\leq \phi_{i,neg}(x, y) \qquad \forall y \in Y \\ \bar{\phi}_{j,pos}(x, Y) &\geq \phi_{j,pos}(x, y) \qquad \forall y \in Y \end{aligned}$$

Branch and Bound Inference — Bounding — Grass

Grass Potential:

$$\phi_G(x, y) = \Big(\frac{\#\text{ of grass pixels in } F_y}{\text{total }\#\text{ of grass pixels}}, \frac{\#\text{ of non-grass pixels in } F_y^c}{\text{total }\#\text{ of non-grass pixels}}\Big)$$

Branch and Bound Inference — Bounding — Grass



Branch and Bound Inference — Bounding — Grass

Note that for any field $y \in \mathcal{Y}$, we have

$$F_{y_\cap} \subset F_y \subset F_{y_\cup}$$

The above relation implies that

$$\begin{aligned} \#\text{ of grass pixels inside } F_{y_\cap} &\leq \#\text{ of grass pixels inside } F_y \\ &\leq \#\text{ of grass pixels inside } F_{y_\cup} \end{aligned}$$

Branch and Bound Inference — Bounding — Grass

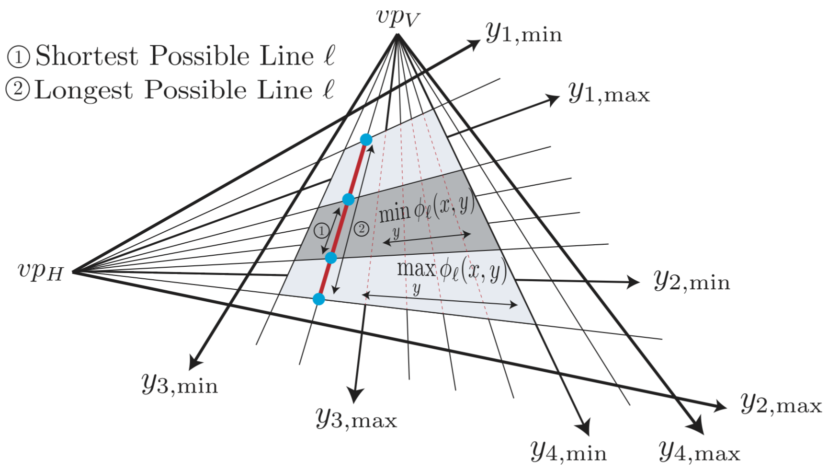
Hence, we can define the upper bound for the grass potential as:

$$\bar{\phi}_{G,pos}(x, Y) = \phi_G(x, y_\cup)$$

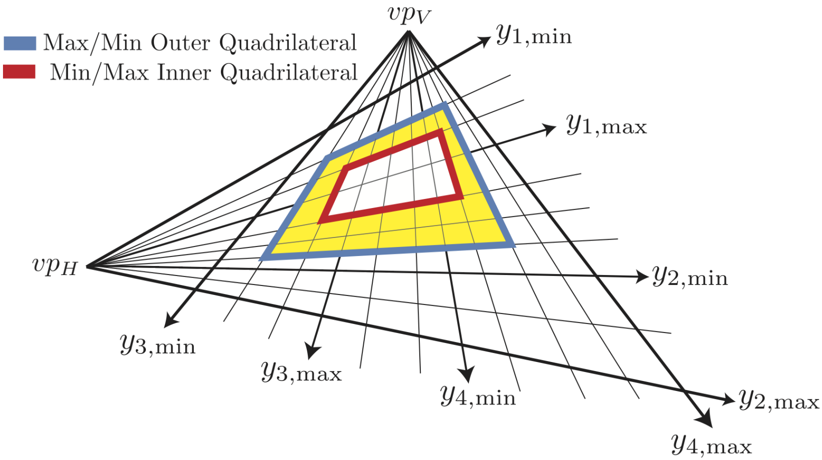
Similarly, a lower bound can be defined as:

$$\bar{\phi}_{G,neg}(x, Y) = \phi_G(x, y_\cap)$$

Branch and Bound Inference — Bounding — Lines



Branch and Bound Inference — Bounding — Ellipses



Learning — Structural SVM

- The outputs $y = (y_1, \ldots, y_4)$ of equation (1) are discrete random variable with complex dependencies,
- Use SSVM
- Given a dataset of ground truth training pairs $\left\{x^{(i)}, y^{(i)}\right\}_{i=1}^{N}$ we learn the parameters w by solving the following optimization problem

$$\min_{w} \frac{1}{2} \|w\|^2 + \frac{C}{N} \sum_{n=1}^{N} \max_{\hat{y} \in \mathcal{Y}} \left(\Delta(y^{(n)}, \hat{y}) + w^T \phi(x^{(n)}, \hat{y}) - w^T \phi(x^{(n)}, y^{(n)}) \right)$$

where $\Delta : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+ \cup \{0\}$

Learning — Structural SVM — Δ

- A hypothesis field $\hat{y}$
- $T_{\hat{y}}$ be the collection of cells in the grid $\mathcal{Y}$ corresponding to the region inside the quadrilateral defined by $\hat{y}$
- $T^c_{\hat{y}}$ be the complement of $T_{\hat{y}}$ in the grid $\mathcal{Y}$

$$\Delta(y, \hat{y}) = 1 - \frac{\#\text{ of GT cells in } T_{\hat{y}} + \#\text{ of cells NGT in } T^c_{\hat{y}}}{\text{Total number of cells in } \mathcal{Y}}$$

Experiments:

Datasets:

- 395 images from 20 games from World Cup 2014
- 209 train/val from 10 games
- 186 test from the 10 other games

- 4000 images from 10 NHL games
- 2000 train/val
- 2000 test