ECE5554 – Computer Vision
Lecture 10a – Pose Estimation

Creed Jones, PhD

BRADLEY DEPARTMENT
OF ELECTRICAL & COMPUTER ENGINEERING

VIRGINIA TECH.

Course Update

- HW5 due Wednesday, 11:59 PM
- Our final quiz is tomorrow!
- Final exam is this Thursday!
  - It’s designed to be around one hour long
  - Fully online, open book and notes
- Please complete the SPOT survey

Today’s Objectives

Pose Estimation

- Concept
- Linear equations for pose estimation
- Iterative solution
- Use in augmented reality

POSE ESTIMATION

Pose estimation refers to the estimation of a 3D object's *pose* from a set of projections of 2D points

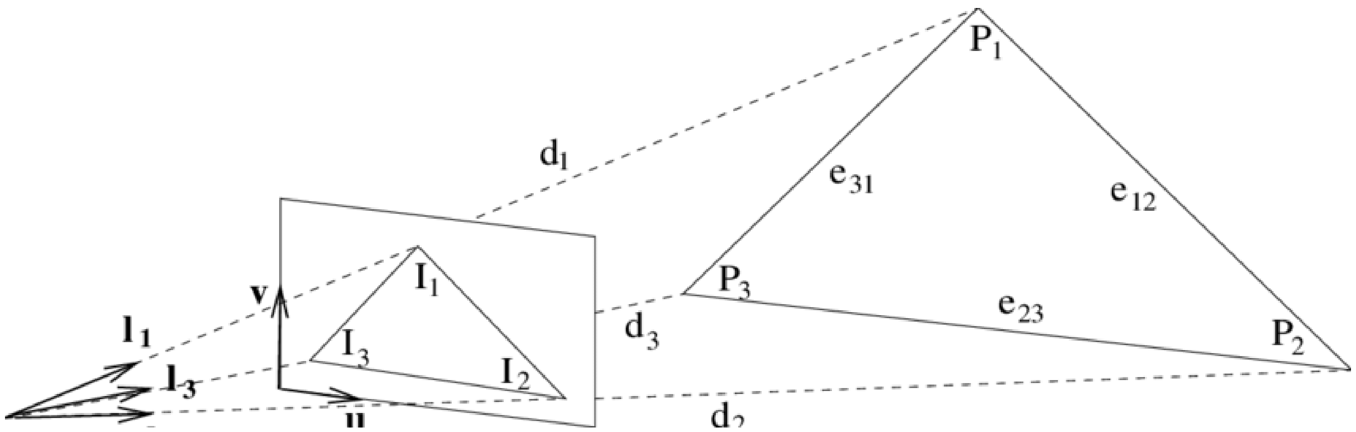
- *Pose* is the position and orientation of an object with respect to some coordinate system
  - Usually with respect to the viewer's coordinate system (the camera)

- Szeliski defines pose estimation as "determining a camera's position relative to a known 3D object or scene"

- Many people use the term "pose estimation" to refer to "<u>human</u> pose estimation" – detection of human figures in images and determining the position and articulation of their body and limbs
  - This is not that

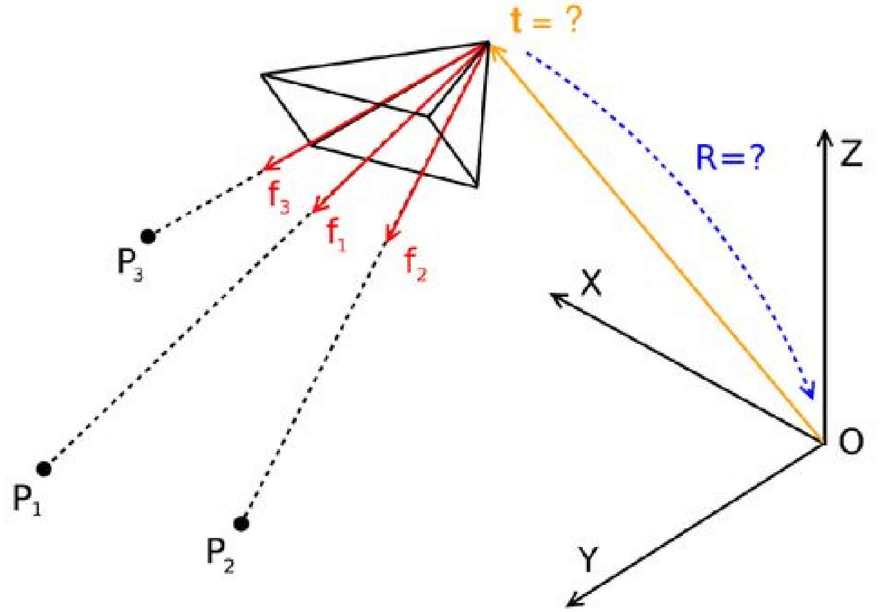
Pose estimation from a set of points is also known as extrinsic calibration

- "The problem of recovering pose from three correspondences, which is the minimal amount of information necessary, is known as the perspective-3-point-problem (P3P), with extensions to larger numbers of points collectively known as PnP."



P3P algorithm: *Perspective from 3 Points*

- Given 3 object points & their image correspondences
- Compute transformation of camera w.r.t. object frame
- Returns 4 solutions
  - Need 4th point for disambiguation
- OpenSource: (OpenGV) [Kneip&Scaramuzza, CVPR’11]



[**Kneip & Scaramuzza CVPR’11**, “A novel parameterization of the Perspective-Three-Point Problem for a direct computation of the absolute camera position and orientation]

Davide Scaramuzza – rpg.ifi.uzh.ch

A Stable Direct Solution of Perspective-Three-Point Problem

Shiqi Li

School of Mechanical Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, 430074, China
sqli@hust.edu.cn

Chi Xu

School of Mechanical Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, 430074, China
*xuchi.hust@yahoo.com.cn**

The perspective-three-point problem (P3P) is a classical problem in computer vision. The existing direct solutions of P3P have at least 3 limitations: (1) the numerical instability when using different vertex permutations, (2) the degeneration in the geometric singularity case, and (3) the dependence on particular equation solvers.
A new direct solution of P3P is presented to deal with these limitations. The main idea is to reduce the number of unknown parameter by using a geometric constraint we called "perspective similar triangle" (PST). The PST method achieves high stability in the permutation problem and in presence of image noise, and does not rely on particular equation solvers. Furthermore, reliable results can be retrieved even in "danger cylinder", a typical kind of geometric singularity of P3P, where all existing direct solutions degenerate significantly.

Keywords: Perspective-three-point problem (P3P); camera pose estimation; geometric singularity.

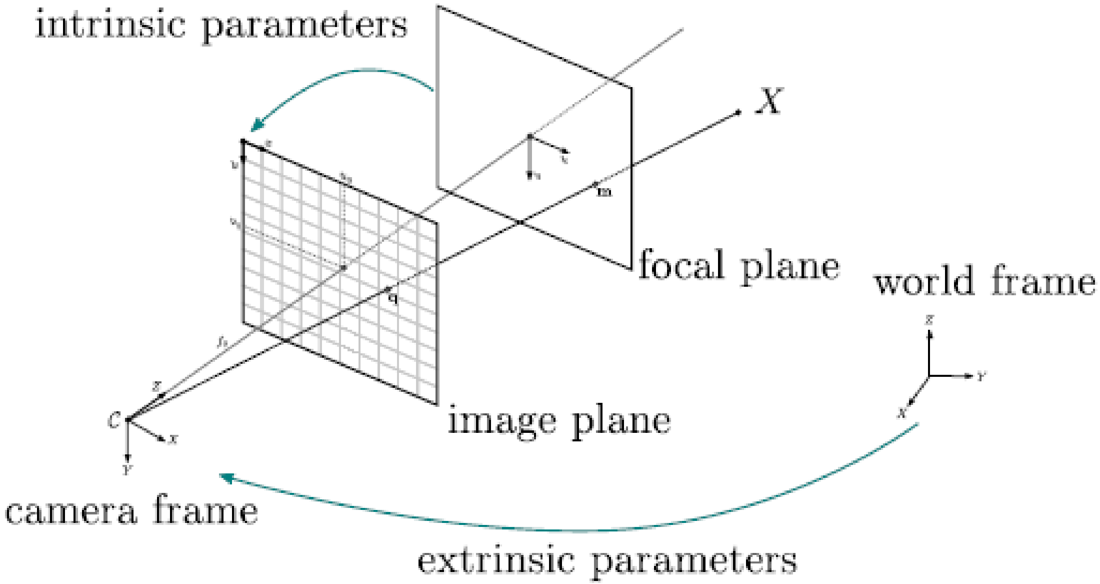
The *camera matrix* relates positions of objects in a scene to their locations in an image

- The camera matrix relates the position of a point in space to the corresponding point in an image

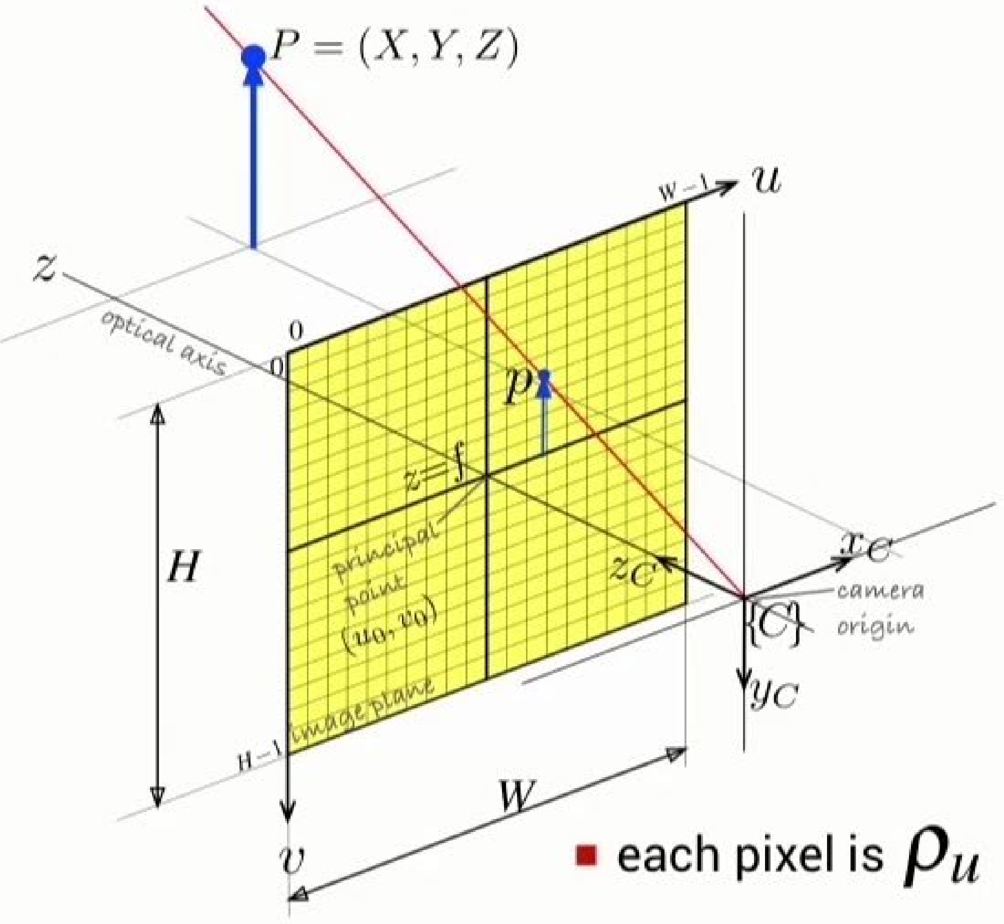
$$Imgpt = P(3Dpt)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

- Note the effect of perspective…



Change of coordinates



- each pixel is $\rho_u \times \rho_v$

With kind permission of Springer Science+Business Media

- scale point from metres to pixels
- shift the origin to top left corner

$$u = \frac{x}{\rho_u} + u_0$$

$$v = \frac{y}{\rho_v} + v_0$$

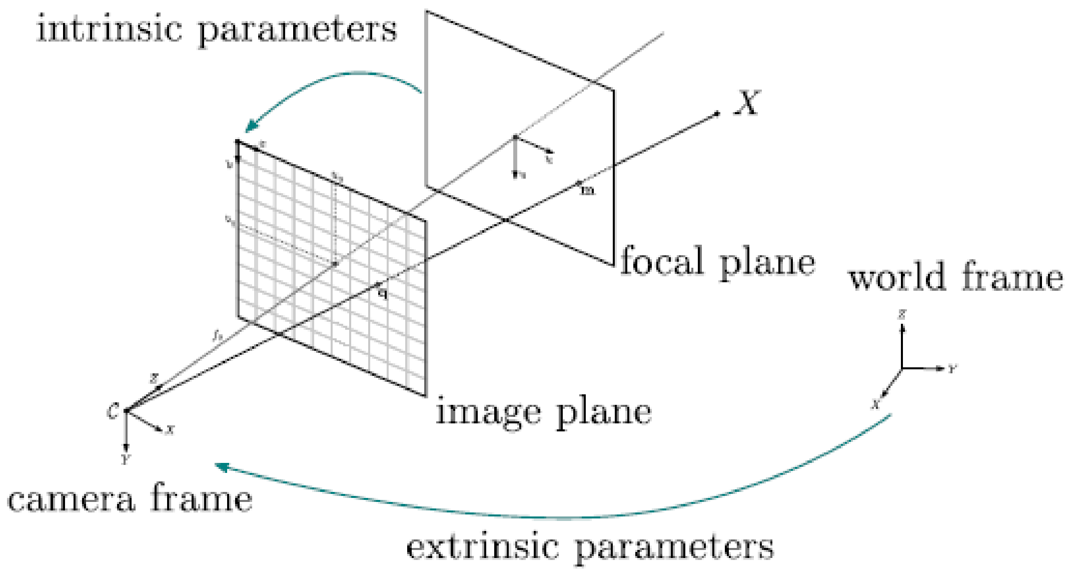
$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{pmatrix} = \begin{pmatrix} \frac{1}{\rho_u} & 0 & u_0 \\ 0 & \frac{1}{\rho_v} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix}$$

$$p = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \tilde{u}/\tilde{w} \\ \tilde{v}/\tilde{w} \end{pmatrix}$$

Camera calibration is affected by both *intrinsic* and *extrinsic* factors

- *Intrinsic* factors depend on the characteristics of the camera and optics: focal length, pixel spacing, etc.

- *Extrinsic* factors are related to the camera's position and orientation in space



The *camera matrix K* relates positions of objects in a scene to their locations in an image

"When calibrating a camera based on external 3D points or other measurements (Tsai 1987), we end up estimating the intrinsic (K) and extrinsic (R; t) camera parameters simultaneously using a series of measurements,

$$\tilde{x}_s = K[R|t]p_w = Pp_w \qquad [2.55]$$

where p_w are known 3D world coordinates and

$$P = K[R|t] \qquad [2.56]$$

is known as the camera matrix." – *Szeliski, p. 51*

Linear equations for pose estimation

- If we know where three points are actually located in space, the points in the image are related to the camera transformation matrix by:

$$x_i = \frac{p_{00}X_i + p_{01}Y_i + p_{02}Z_i + p_{03}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

$$y_i = \frac{p_{10}X_i + p_{11}Y_i + p_{12}Z_i + p_{13}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

- where (x_i, y_i) are the measured 2D feature locations and X_i, Y_i and Z_i are the known 3D feature locations

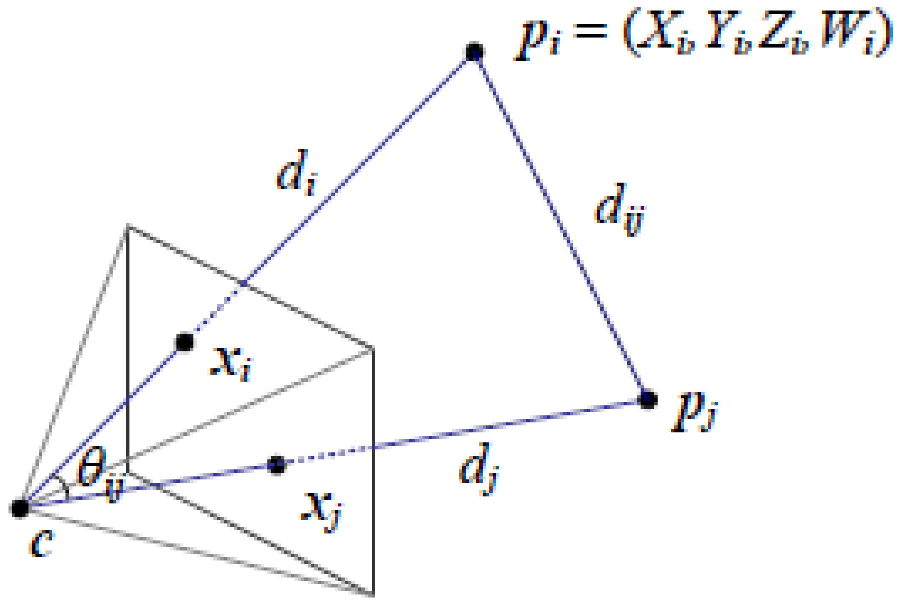


Figure 6.4 Pose estimation by the direct linear transform and by measuring visual angles and distances between pairs of points.

Note that the linear equations have twelve unknowns

$$x_i = \frac{p_{00}X_i + p_{01}Y_i + p_{02}Z_i + p_{03}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}, \qquad y_i = \frac{p_{10}X_i + p_{11}Y_i + p_{12}Z_i + p_{13}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

- Remember, point i is at $[X_i, Y_i, Z_i]$ in space and $[x_i, y_i]$ in the image
- In the general case, we need six points in our image that are at known positions in space
- If we make various assumptions (the camera sensor is not skewed, we choose the coordinate systems to be non-rotated, etc.), we can often derive the camera matrix from as few as three points

- <u>Points chosen must not be collinear!!!</u>

For a given set of points, we can solve by least squares

$$x_i = \frac{p_{00}X_i + p_{01}Y_i + p_{02}Z_i + p_{03}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}, \qquad y_i = \frac{p_{10}X_i + p_{11}Y_i + p_{12}Z_i + p_{13}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

- Often, the equations are over-determined (more equations than unknowns)
- Minimization of the squared error is usually used
  - Though outlier-aware processes like RANSAC are appropriate
- This is called *direct linear transformation* (Sunderland 1974)

- See this URL (by Kwon) for more information:
  http://www.kwon3d.com/theory/dlt/dlt.html

Non-iterative method for camera calibration

Yuzhen Hong,[1,2,*] Guoqiang Ren,[2] and Enhai Liu[2]

[1]University of Chinese Academy of Sciences, Beijing 100149, China
[2]Institute of Optics and Electronics of Chinese Academy of Sciences, Chengdu 610209, China
[*]hongyuzhen12@163.com

Abstract: This paper presents a new and effective technique to calibrate a camera without nonlinear iteration optimization. To this end, the centre-of-distortion is accurately estimated firstly. Based on the radial distortion division model, point correspondences between model plane and its image were used to compute the homography and distortion coefficients afterwards. Once the homographies of calibration images are obtained, the camera intrinsic parameters are solved analytically. All the solution techniques applied in this calibration process are non-iterative that do not need any initial guess, with no risk of local minima. Moreover, estimation of the distortion coefficients and intrinsic parameters could be successfully decoupled, yielding the more stable and reliable result. Both simulative and real experiments have been carried out to show that the proposed method is reliable and effective. Without nonlinear iteration optimization, the proposed method is computationally efficient and can be applied to real-time online calibration.

OCIS codes: (150.0150) Machine vision; (150.1488) Calibration; (150.1135) Algorithms.

For large sets of points, it can be more accurate to use an iterative process to minimize the square errors

- The projection equations are seen as a chain of matrix transformations:
- Each step can be fit by estimating the partial derivatives at that step
- Repeated iterations minimize the square error
  See section 6.2.2 in Szeliski's book

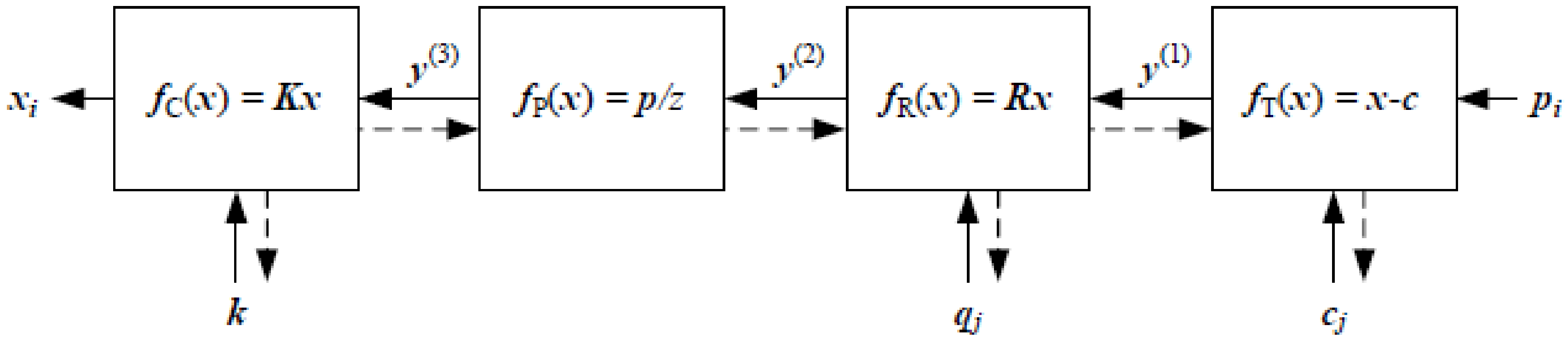
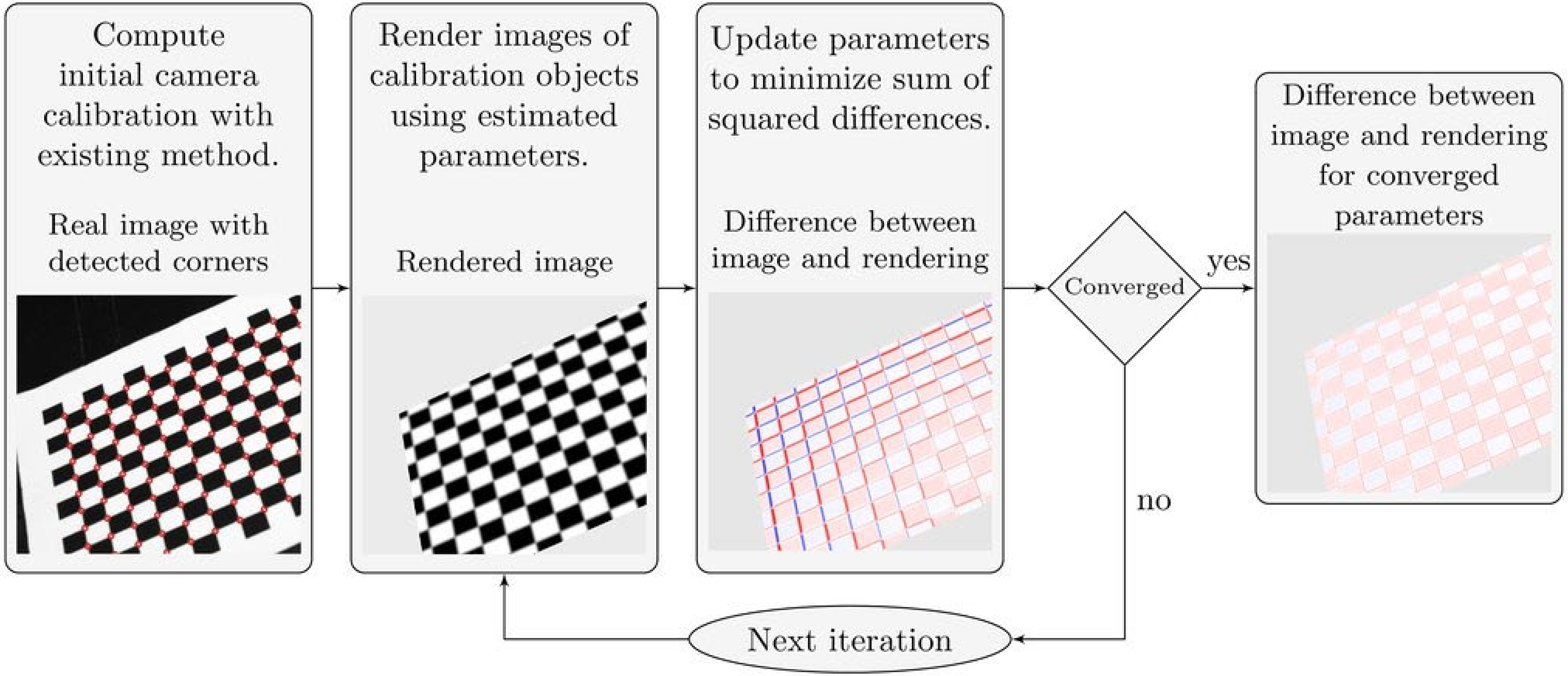


Figure 6.5 A set of chained transforms for projecting a 3D point p_i to a 2D measurement x_i through a series of transformations $f^{(k)}$, each of which is controlled by its own set of parameters. The dashed lines indicate the flow of information as partial derivatives are computed during a backward pass.



Hannemose 2020 - Superaccurate camera calibration via inverse rendering

Superaccurate Camera Calibration via Inverse Rendering

Morten Hannemose[a], Jakob Wilm[b], and Jeppe Revall Frisvad[a]

[a]DTU Compute, Technical University of Denmark
[b]SDU Robotics, University of Southern Denmark

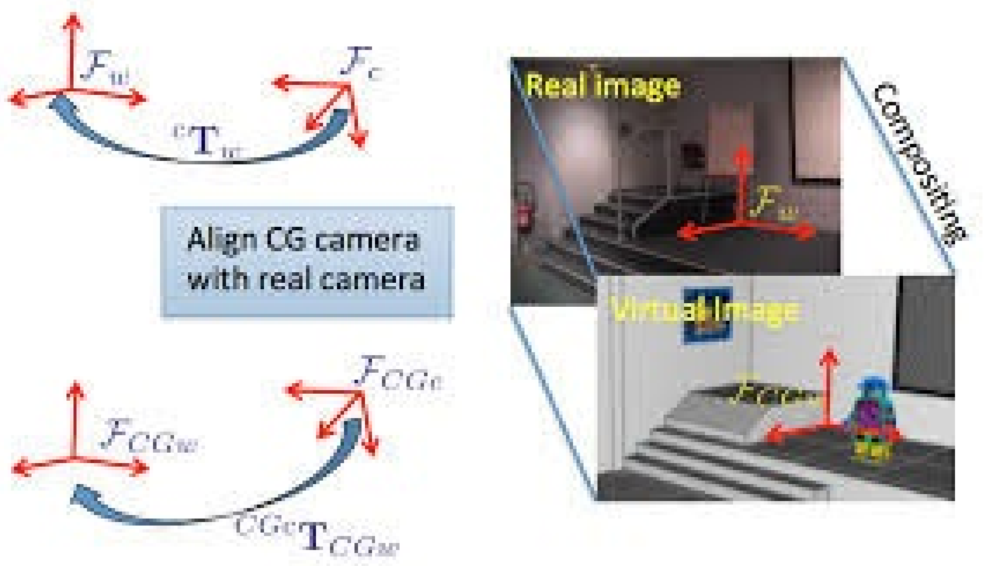
ABSTRACT

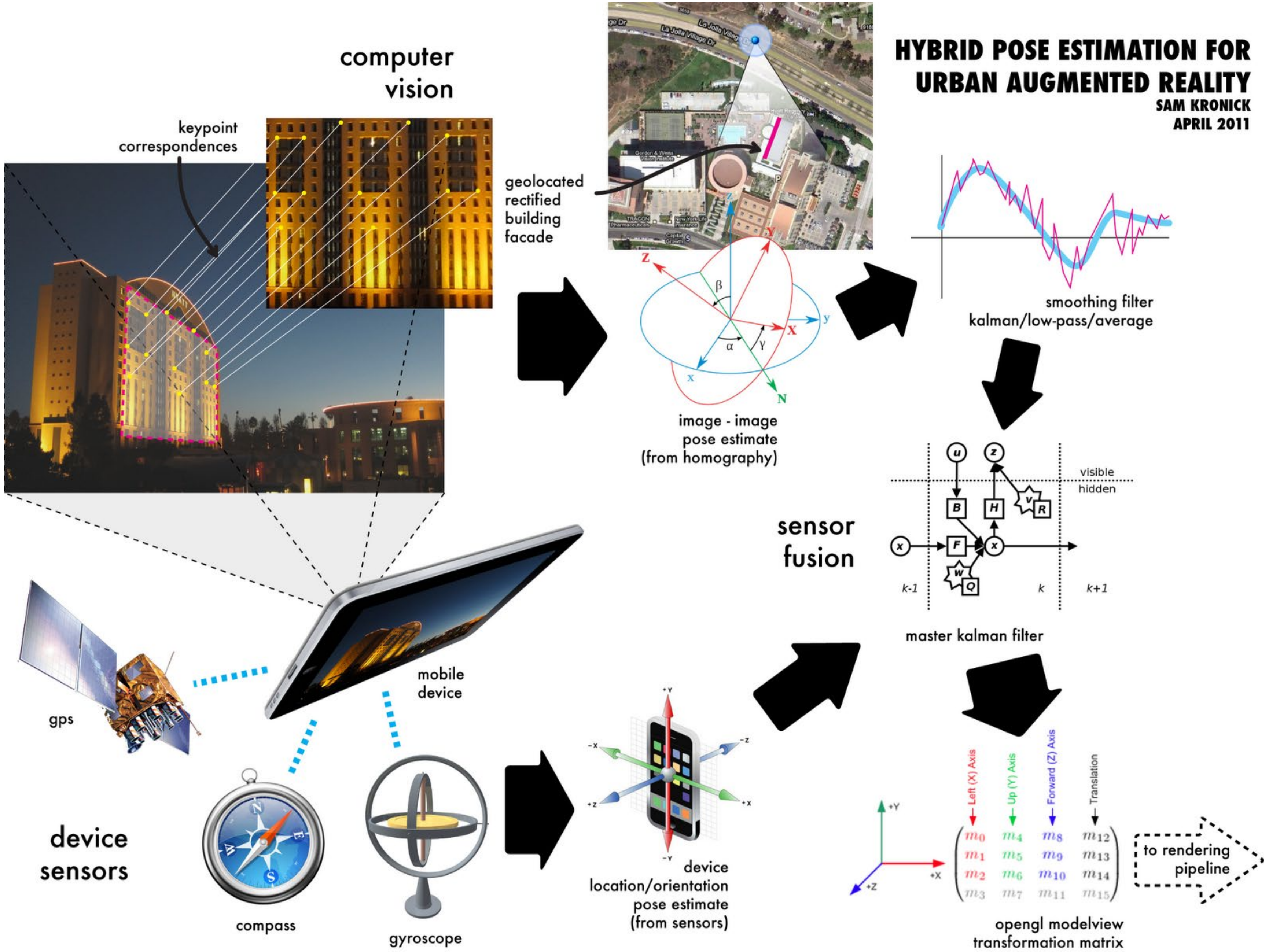
The most prevalent routine for camera calibration is based on the detection of well-defined feature points on a purpose-made calibration artifact. These could be checkerboard saddle points, circles, rings or triangles, often printed on a planar structure. The feature points are first detected and then used in a nonlinear optimization to estimate the internal camera parameters. We propose a new method for camera calibration using the principle of inverse rendering. Instead of relying solely on detected feature points, we use an estimate of the internal parameters and the pose of the calibration object to implicitly render a non-photorealistic equivalent of the optical features. This enables us to compute pixel-wise differences in the image domain without interpolation artifacts. We can then improve our estimate of the internal parameters by minimizing pixel-wise least-squares differences. In this way, our model optimizes a meaningful metric in the image space assuming normally distributed noise characteristic for camera sensors. We demonstrate using synthetic and real camera images that our method improves the accuracy of estimated camera parameters as compared with current state-of-the-art calibration routines. Our method also estimates these parameters more robustly in the presence of noise and in situations where the number of calibration images is limited.

Keywords: camera calibration, inverse rendering, camera intrinsics

Constant pose estimation is critical to augmented reality

- We need to know the camera matrix to relate created graphic content with real-world objects
- Sometimes the known points are scene features, but often markers (such as balls) are added for robustness





Thanks to: Kronick

Today's Objectives

Pose Estimation

- Concept
- Linear equations for pose estimation
- Iterative solution
- Use in augmented reality