PA198
Augmented Reality Interfaces

Lecture 7

Augmented Reality Registration and Calibration

Fotis Liarokapis
liarokap@fi.muni.cz

07th November 2016

Image and Video Registration

Image and Video Registration

- Image and video registration is a deeply studied area from Computer Vision
- It is concerned with the alignment of image sequences with respect to each other
  - i.e. in a common 3D coordinate system
- Many applications ranging from ubiquitous panorama creations to sophisticated interfaces for interaction with media

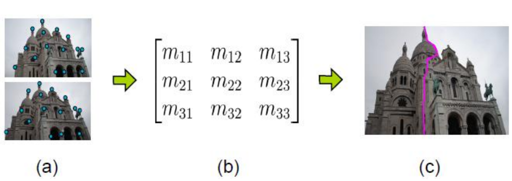
http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Image Registration Common Pipeline

- Suppose we have a sequence of images we want to put in the same coordinate system
  - First, we need to identify features in the images and find correspondences between pairs of images
  - Then, we can estimate the transformation relating the image planes
  - This estimation must be robust to outliers due to mismatched features
  - Finally, we optimize the position of the images taking the whole set into account, for better results

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Image Registration Common Pipeline .



(a) Feature detection and matching
(b) Transformation estimation between two images
(c) Image stitching with seam optimization (illustrative only)

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection and Matching

- Consider two frames of a video and the task of creating a panorama from them
- While there is much coherence in the two images, the information about how they would fit together is not explicit
- The similarity of a part of the first image with a part of the second image needs to be formalized in a way that a computer can understand it
- This lecture will focus on the detection and matching of two types of features:
  - Regions
  - Lines

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Points and Patches

- Points and patches are the most common features used when trying to find correspondences between images
  - They provide the easiest algorithms
- Two approaches can be used in point matching:
  - Finding features that can be tracked in a set of images
    - Recommended when the scene does not change much
  - Finds the most prominent features of each image and then tries to match the sets

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Points and Patches .

- Four main steps:
  - Feature detection
  - Feature description
  - Feature matching
  - Feature tracking

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection

- Detect which parts of the image make the best features
  - To avoid ambiguities
- The question that comes up when confronted with the task of selecting a set of features from an image is what makes a good feature?
- The most distinct patches should be selected

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection .

- Different patches of the image
  - (b) good feature
  - (a), (c) bad features

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection ..

- The simplest way to match two features form different images is to take the sum of squared differences between them
- Can be described as:

$$Error(u) = \sum [I_1(x_i + u) - I_0(x_i)]^2$$

- where
  - I_0 and I_1 are the images being compared
  - u is a displacement vector

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection ...

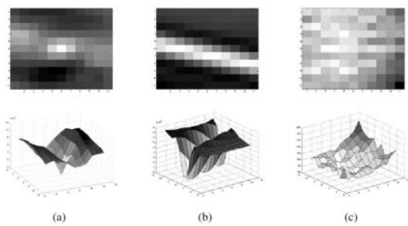
- Next, evaluate how different a patch is from the patches around it using an autocorrelation function

$$AutoCorrelation(u) = \sum [I_0(x_i + u) - I_0(x_i)]^2$$

- Note that the displacement vector u here would represent how the second patch is displaced from the original one

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection ....



Three types of features and the respective autocorrelation values (error is plotted in terms of the displacement) (a) a good feature with a distinct minimum (b) a feature suffering from aperture (c) a very bad feature for which many points in the neighbourhood are minima

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection .....

- We can approximate the function by using the first terms of a Taylor Series expansion to generate an autocorrelation matrix that will then be used to define how distinctable a feature is
- After simplifications:

$$\begin{aligned} AutoCorrelation(u) &= \sum [I_0(x_i + u) - I_0(x_i)]^2 \\ &\approx \sum [I_0(x_i) + \nabla I_0(x_i) u - I_0(x_i)]^2 \\ &= \sum [\nabla I_0(x_i) u]^2 \\ &= \sum u^T A u \end{aligned}$$

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection ......

- The matrix A is called the autocorrelation matrix and its properties can be used to decide whether a feature is distinct or not

$$A = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}$$

- How the properties of the autocorrelation matrix are used to determine whether a feature should be used or not depends on the type of detection we do

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Detection ....

- Some examples:
  - Shi and Tomasi
  - Förstner-Harris
  - Adaptive non-maximal suppression
  - SIFT

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Shi-Tomasi Example 1

https://www.youtube.com/watch?v=jKW2y5NhO_k

Shi-Tomasi Example 2

Feature Description

- Once we have decided which parts of the image would generate good features, we need to match them
- A simple approach would be minimizing the sum of squared differences
- Unfortunately, this would only work if the transformation applied to the patch was a simple translation
- In most cases, the patch will undergo arbitrary affine transformations that will require more sophisticated methods for comparison
  - See next slide

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Description .

- MOPS (Multi-Scale Oriented Patches)
  - Used in applications that do not require invariance for transformations
  - It consists of sampling a lower frequency version of the image around the feature
  - Patch intensities are re-scaled such that their mean is zero and the variance is 1
- SIFT (Scale-invariant feature transform)
  - Descriptors are formed by computing the gradient in a 16x16 window around the pixel
  - The contribution of each gradient is weighted by a Gaussian centered at the pixel
  - The 16x16 window is divided into sixteen 4x4 regions, and represents the gradient in these regions by adding them to 8 bins
  - The 128 resulting values (the values in each of the 8 bins, in each of the 16 regions, 16x8 = 128) are the SIFT descriptor
- GLOH (Gradient Location-Orientation Histogram)
  - Extension of SIFT that uses polar bins instead of square ones
  - Divides the space into 3 radial bins and 6 angular bins - with one additional bin for radius = 0

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Matching

- There are two important steps when matching features
  - Define a matching strategy
  - Combine data structures and algorithms to efficiently perform the matching evaluations

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Matching .

- A simple approach is to immediately reject matches that are further away from each other than a threshold value (using Euclidean distance)
  - In this case, we must of course observe that the threshold is consistent with our expected camera motion, to avoid false positives and false negatives
  - The problem with this strategy is that the threshold depends on each case and is difficult to optimize
- A different approach is to use nearest neighbors in feature space
  - Since some features would not have matches, a threshold approach is still used to avoid extreme false positives

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Tracking

- An alternative approach to detecting candidate features in each image and then matching them is to find features in the first image and track them throughout the image sequence
- This approach is most commonly used in videos - where the scene is not expected to change much from one frame to the next
- The definition of a good feature to track is the same as before
- If features are being tracked in long image sequences, it is important to consider the problem of big changes in the patches, as well as the loss of patches by occlusion

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

Feature Tracking .

- To solve this problem, features are matched in different images along the sequence using an affine motion model
  - Instead of a simple translational one
- Since affine matching is more expensive than translational, a first estimation is performed using simpler models
- The resulting tracker is called Kanade-Lucas-Tomasi, or KLT

http://www.inf.ufrgs.br/~rgschneider/research/reports/videoRegistration.pdf

KLT AR Example

https://www.youtube.com/watch?v=8VbylCRn3iI

AR Registration

AR Registration

- Registration is the accurate positioning of virtual information into the real environment
- Mis-registration:
  - Breaks the illusion that the two coexist
  - Prevents acceptance of many serious applications

Video Key Method

- Video Keying is a process that is widely used in television, film production and CG. (weather report)
- When using video keying to design AR scenes, one signal contains the foreground image and the other one contains the background image
- The 'keyer' combines the two signal to produce a combined video which is then sent to the display device

Video Key Method .

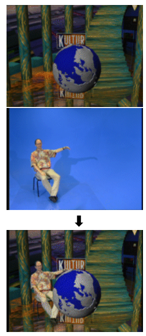
- Keying can be done using composite or component video signals
  - A composite video signal contains information about color, luminance, and synchronization, thus combining three piece of information into one signal
  - With component video, luminance synchronization are combined, but chroma information is delivered separately

Video Key Method ..

- Chroma keying involves specifying a desired foreground key color
- Foreground areas containing the keying color are then electronically replaced with the background image
- This results in the background image being replaced with the fore ground image in areas where the background image contains chroma color
- Blue is typically used for chroma keying (Chromakey blue) rarely shows up in human skin tones

Video Key Method ...

- If a video image of the real world is chosen as the foreground image, parts of the scene that should show the computer-generated world are rendered blue
- In contrast, if video of the real world is chosen as the background image, the computer generated environment will be located in the foreground



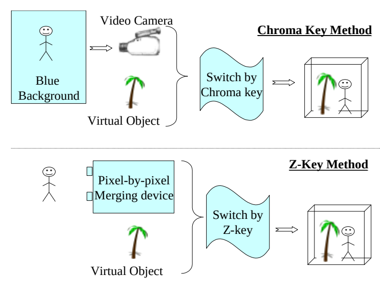
Chroma Keying AR Example

https://www.youtube.com/watch?v=FOeVKSC2XA8

Z-key Method

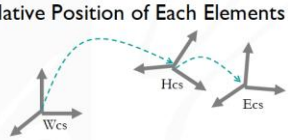
- The z-key method requires images with both depth information (depth map) as inputs
- The z-key switch compares depth information of two images for each pixel, and connects output to the image which is the nearer one to the camera
- The result of this is that real and virtual objects can occlude each other correctly
- This kind of merging is impossible by the chroma-key method, even if it is accompanied with some other positioning devices such as magnetic or acoustic sensor, since these devices provide only a gross measurement of position

Chroma Key vs Z-Key



Spatial Registration

- Defining Relative Position of Each Elements of a Scene



- Elements are: User, User's eye, Environment (Table, Room, Building), Objects, etc.
- Coordinate Systems (Euclidian System)
- Initially: Calibration
- Temporally: 3D/6D **Tracking**

Billinghurst, M. COSC 426: Augmented Reality, July 26th 2013.

Sources of Registration Errors

- Static errors
  - Optical distortions
  - Mechanical misalignments
  - Tracker errors
  - Incorrect viewing parameters
- Dynamic errors
  - System delays (largest source of error)
    - 1 ms delay = 1/3 mm registration error

Billinghurst, M. COSC 426: Augmented Reality, July 26th 2013.

Reducing Static Errors

- Distortion compensation
- Manual adjustments
- View-based or direct measurements
- Camera calibration

View Based Calibration (Azuma 94)

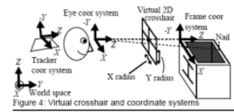
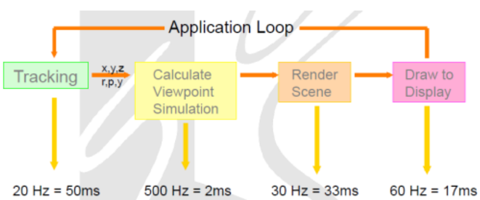


Figure 4: Virtual crosshair and coordinate systems


Figure 1: Wooden frame for calibration and registration


Figure 2: View seen in HMD, virtual axes on real frame

Dynamic Errors

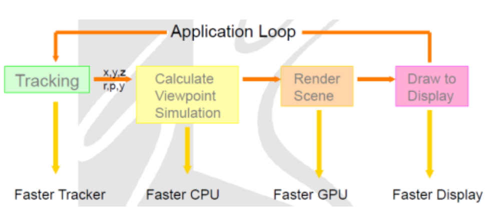


- Total Delay = 50 + 2 + 33 + 17 = 102 ms
  - 1 ms delay = 1/3 mm = 33mm error

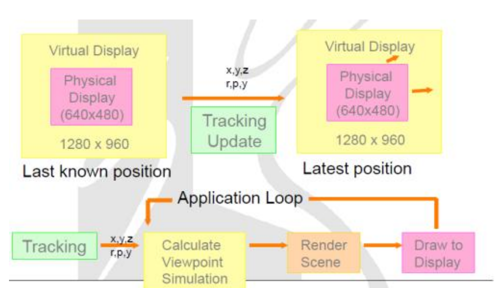
Reducing Dynamic Errors

- Reduce system lag
  - Faster components/system modules
- Reduce apparent lag
  - Image deflection
  - Image warping

Reducing System Lag

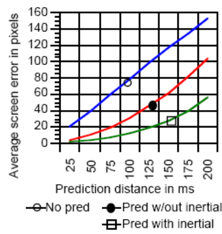


Reducing Apparent Lag



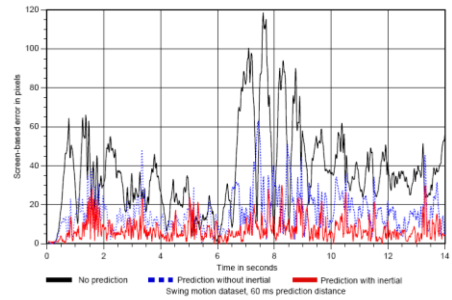
More on Reducing Dynamic Errors

- Match input streams
  - Delay video of real world to match system lag
- Predictive Tracking
  - Inertial sensors helpful



Azuma / Bishop 1994

Predictive Tracking (Azuma 94)



AR Calibration

Calibration Requirements

- "Ideally, the calibration methods should be statistically robust, there should be a variety of approaches for different circumstances, and metrology equipment should be sufficiently accurate, convenient to use, and not too expensive"
  - (Hollerbach and Wampler, 1996)

Additional Calibration Requirements

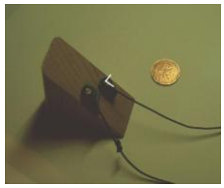
- Additional Requirements of Experimental Subsystems:
  - Independent
    - Not rely on each other
  - Subject-specific
    - Account for individual differences
  - Avoid residual cues
    - To prevent subjects using them in unanticipated ways

Typical Calibration Components

- Workspace
- Point of view
- Physical objects

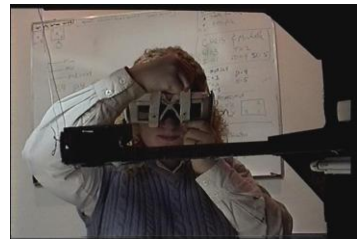
Workspace Calibration

- Markers are aligned with virtual crosses
- Exactly one position in 3-space eliminates "swim"
- do NOT need stereo to calibrate



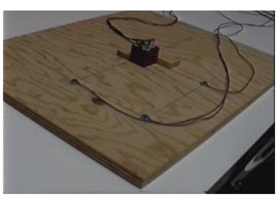
Point of View Calibration

- Subject placing eye calibration bars



Physical Object Calibration

- Markers placed anywhere on object
- Place object in frame so XYZ orientations match



Measuring ARToolKit's Tracking Error

- In wide area applications, the positioning accuracy of ARToolKit is not very robust
- In distances between 1m and 2.5m the error in the x and y values increases proportionally with the distance from the marker
- Calculate error in distances ranging between 20 cm and 80 cm under normal lighting conditions

Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

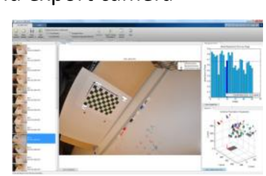
Camera Calibration with MATLAB

- Computer Vision System Toolbox™ provides an app and functions to perform all essential tasks in the camera calibration workflow, including:
  - Fully automatic detection and location of checkerboard calibration pattern including corner detection with subpixel accuracy
  - Estimation of all intrinsic and extrinsic parameters including axis skew
  - Calculation of radial and tangential lens distortion coefficients
  - Correction of optical distortion
  - Support for single camera and stereo calibration

https://www.mathworks.com/videos/camera-calibration-with-matlab-81233.html?requestedDomain=www.mathworks.com

Camera Calibrator App

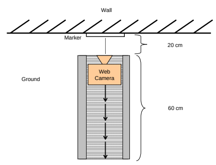
- Used to select and filter calibration images, choose the number and type of radial distortion coefficients, view reprojection errors, visualize extrinsic parameters, and export camera calibration parameters



https://www.mathworks.com/videos/camera-calibration-with-matlab-81233.html?requestedDomain=www.mathworks.com

Measuring ARToolKit's Tracking Error .

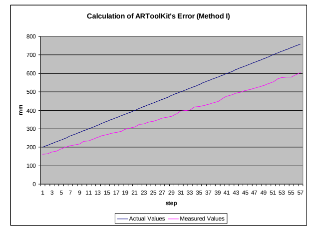
- The optimal area, which contains the least error, is the one that is perpendicular to the marker card
- A rigid path is set so that the camera can not loose its direction while moving backwards
- Numerous measurements of the location of the web camera in a local co-ordinate system



Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Measuring ARToolKit's Tracking Error ..

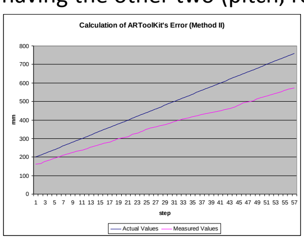
- Error is proportional to the distance



Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Measuring ARToolKit's Tracking Error ...

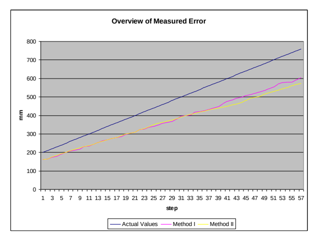
- Camera facing the marker at variable angle (yaw) having the other two (pitch, roll) stable



Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Measuring ARToolKit's Tracking Error ....

- Differences in the error produced from the experiments compared with the actual values



Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Calculating Camera Parameters

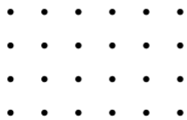
ARToolKit's Calibration Method

- ARToolKit provides two software tools called calib_dist and calib_param that can be used to calculate these camera properties
  - calib_dist is used to measure the lens distortion and the image centre point
  - calib_param is used to compute the focal length of the camera

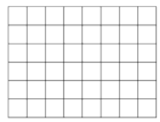
Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

ARToolKit's Calibration Method .

- In calib_dist program, an image of a pattern 6x4 dots spaced equally apart is used



Camera calibration for lens distortion

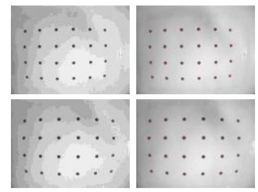


Camera calibration for focal length

Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

KLT Tracking Method

- A feature-based vision method is known as the KLT (Kanade-Lucas-Tomasi) algorithm based on a model or affine image changes



Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

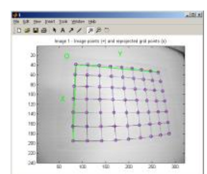
Camera Calibration Toolbox for Matlab

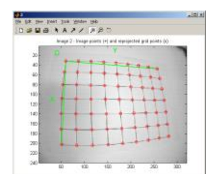
- Offers an automatic mechanism for counting the number of squares in each grid
- All calibration images are searched and focal and distortion factor are automatically estimated

	Estimated focal (pixels)	Estimated distortion factor (kc)
Image 1	201.2503	-0.2
Image 2	224.1415	-0.2
Image 3	231.7859	-0.2
Image 4	114.2254	-0.055
Image 5	265.8596	-0.28
Image 6	211.5925	-0.2

Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

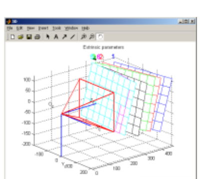
Calibration Without Lens Distortion

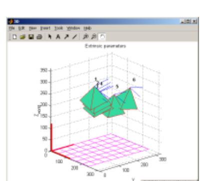




Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

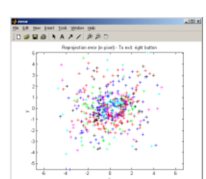
Calculation of Extrinsic Parameters



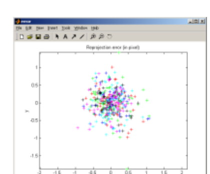


Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Error Calculation of Extrinsic Camera Parameters



Initial Error



Corrected Error

Liarokapis, F., Augmented Reality Interfaces - Architectures for Visualising and Interacting with Virtual Information, Sussex theses S 5931, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, UK, 2005

Outdoor Markerless Tracking and Registration in AR

https://www.youtube.com/watch?v=hRRAGuL8lVk

Questions