Superior Augmented Reality by Integrating Landmark Tracking and Magnetic Tracking

Authors: Andrei State et al.
Department of Computer Science
University of North Carolina at Chapel Hill

Presented by: Matt Clothier
What is Augmented Reality?

- An AR system merges computer-synthesized objects with the user’s space in the real world.
- Synthetic objects enhance the user’s interaction with, or his perception of, the real world.
AR can be useful for navigation purposes
(image courtesy: www.howstuffworks.com)

AR can also be used to help people communicate
(image courtesy: Popular Science, Feb ’02)
Ultimate Computer Science Problem!

Components of an AR system:
- Computer Vision
- Computer Graphics
- Human/Computer Interaction
- Databases
- Networking
- Security
Motivation Video
Previous Work

Sutherland, I. E. A Head-Mounted Three Dimensional Display. (1968):
Previous Work

Lowe, D.G. Robust Model-based Motion Tracking Through the Integration of Search and Estimation. (1992):
Previous Work

Previous Work

Magnetic Tracking Video
Previous Work

Difficulties of previous systems

- Magnetic trackers are inaccurate
- Mechanical trackers are cumbersome
- Vision-based trackers are computationally problematic
- Solution? Create a hybrid tracker that uses the accuracy of a vision based tracker and the robustness of a magnetic tracker
Goals for Hybrid Tracker

- Accurate registration between synthetic and real objects
- Reasonable image generation rate
- Simple initial set up
- Minimal constraint on user motion
- Low latency
How will this system work?

- Use video landmarks as the primary method for determining camera position and orientation
- Color-code landmarks to help identify them quickly
- Use vision-based tracker to calibrate magnetic tracker “on-the-fly”
Why use a magnetic tracker?

- Image analysis acceleration
- Selection from multiple solutions
- Backup tracking
- Sanity check of the vision-based tracker
System Hardware

Virtual Research VR-4 HMD

Two Panasonic CCD video cameras

Ascension Flock of Birds magnetic tracker

Silicon Graphics Onyx graphics workstation
System Assumptions/Calibration

- The system assumes that two cameras and a magnetic tracker are rigidly interconnected - the transformation between them is easily acquired.
- The world space positions of the landmarks used are precisely calibrated (through FARO mechanical arm).
System Assumptions/Calibration

- Cameras were selected due to their low distortion - thus a pin-hole camera model could be used
- Transformation between stereo cameras computed on the fly and averaged over the last 10 frames (does not need to be computed in advance)
Operation (part 1)

- System retrieves stereo image pair and attempts to determine the head pose
- If successful, an error correcting transformation between the magnetic tracker and the head pose is computed
- The error computed is used to predict the head pose in the next frame
Operation (part 2)

- The adjusted head pose is then used for re-prediction of landmark search areas
- The system then uses these improved values to find additional landmarks
- For each landmark found, a head pose solver is invoked depending on the number of landmarks detected
Operation (part 3)

- This gives us two distinct cases:
  - The number of landmarks is not sufficient to determine the head pose (under-determined case) - use heuristic adjuster
  - There are enough landmarks (well or over-determined case) - use global solver

- Results subject to sanity check
Factor in magnetic tracker

- The magnetic tracker error is preserved for head pose prediction in the next frame
- A 0\textsuperscript{th} order prediction for the tracker error is fine as long as the frame rate is below 15 Hz
- A higher order prediction is used if the application can handle higher frame rates
Overview of hybrid tracker
Vision-only Tracking

- This can be done easily by stripping out the magnetic tracker
- Primary tracking techniques:
  - Two-color objects
  - Concentric circular dots (inner dot and a surrounding outer ring)
  - Color landmarks easier to spot and having two-colors makes spurious detection unlikely
Vision Landmark predictor

- When searching for the first landmark, no head pose corrections are available
- Thus, it is important to find this as quickly as possible
- Finding the first landmark drastically improves finding the other landmarks
Vision Landmark predictor

- Main goal is to compute the expected positions and extents of landmarks
- Done through exhaustive pixel searches inside predicted search areas
- For each landmark found, the accuracy of predicted positions and extents improves
Image analyzer

- Searches for landmarks by inspecting the area defined by the predictor
- The algorithm uses pixel marking
  - A pixel is classified as belonging to one of the landmark colors or no landmark based on the ratios of RGB colors
Image analyzer

- For all potential candidate detections that match an inner and outer color:
  - The ratio of inner vs. outer pixels is computed (so 8 outer to 1 inner)
  - Are the centers of mass of the outer and inner regions close to each other? If not, the landmark may be partially occluded/clipped

- For accepted candidates, the center of mass is considered the center of the landmark
Head Pose Determination

- Under-determined
- Well-determined
- Over-determined
Under-determined case

- Until the image analyzer detects three different landmarks, the head pose cannot be completely determined.
- The magnetic tracker is used as the primary source of information about the head pose.
- Heuristic adjustments are then made in order to ensure the highest possible head pose and registration accuracy.
Under-determined case

- Heuristic adjusters are designed to improve head pose as smoothly as possible
- Hybrid tracker is characterized by reluctant degradation in accuracy when landmarks are lost
- When landmarks are reacquired, the system recovers quickly
Under-determined case

- Six different under-determined cases: (cam = camera)
  - cam 1 sees A, cam 2 sees none
  - cam 1 sees A & B, cam 2 sees none
  - cam 1 sees A, cam 2 sees B
  - cam 1 and cam 2 both see A
  - cam 1 sees A & B cam 2 sees A
  - cam 1 sees A & B, cam 2 sees A & B
Well-determined case

- The image analyzer has found three good landmarks in the image
- Three positions of collinear points are needed to determine a rigid three-space motion
- Three landmarks on both the left and right image planes provide 3 XY pairs
- These six independent values are enough to determine a 6 DOF rigid motion for the head
Well-determined case

Figure 6. Geometric relationships between three landmarks and the two stereo cameras.
Well-determined case

- The unit vectors are:

\[ V_1 = \frac{\vec{y} - \vec{c}}{\sqrt{y_1^2 - c_1^2}}, \quad V_2 = \frac{\vec{y} - \vec{c}}{\sqrt{y_2^2 - c_2^2}} \text{ and } V_3 = \frac{\vec{y} - \vec{c}}{\sqrt{y_3^2 - c_3^2}} \]

- From landmark calibration we know the positions of the landmarks and thus and compute their lengths:

\[ L_{12} = \left| \vec{L}_2 - \vec{L}_1 \right|, \quad L_{23} = \left| \vec{L}_3 - \vec{L}_2 \right| \text{ and } L_{31} = \left| \vec{L}_1 - \vec{L}_3 \right| \]
Well-determined case

- Both cameras are rigidly mounted:
  \[ \bar{T} = \bar{C}_2 - \bar{C}_1 \]

- Let \( x, y, z \) be:
  \[ \left| \bar{V}_{L_1 - C_1} \right|, \left| \bar{V}_{L_2 - C_1} \right| \text{ and } \left| \bar{V}_{L_3 - C_2} \right| \]

- Result:
  \[ L_{12} = \left| x\bar{V}_1 - y\bar{V}_2 \right| \]
  \[ L_{23} = \left| y\bar{V}_2 - (\bar{T} + z\bar{V}_3) \right| \]
  \[ L_{31} = \left| (\bar{T} + z\bar{V}_3) - x\bar{V}_1 \right| \]
Well-determined case

- Square of both sides:

\[ \begin{align*}
    a + b \cdot x \cdot y + x^2 + y^2 &= 0 \\
    c + d \cdot y + e \cdot z + f \cdot y \cdot z + y^2 + z^2 &= 0 \\
    g + h \cdot x + e \cdot z + j \cdot x \cdot z + x^2 + z^2 &= 0
\end{align*} \]

- System of equations consisting of 3 quadratic equations with 3 variables (degree 8)

- Solutions are the intersection of three ellipsoidal cylinders with inf. extents
Well-determined case

- There are 8 solutions to these systems of equations
- Reject imaginary solutions and negative solutions
- Two potential positive solutions; one must be decided upon
Well-determined case

- If more landmarks than the ones used in the equations are found, they can be used to determine the correct solution.
- If no other landmarks can be used, the magnetic tracker is used to determine the answer (only works if the values are not close to one another).
Over-determined case

- System is over-determined if the number of landmarks is at least 3 and the total number of landmark projection pairs is at least 4
- Least square error minimization is used to find an optimum solution using all the detected landmarks
Over-determined case

- Relationship between head, camera, landmark and projected image:

\[
\begin{bmatrix}
I_x \\
I_y \\
I_z
\end{bmatrix} = \begin{bmatrix}
I'_x \\
I'_y \\
I'_z
\end{bmatrix}
\]

\[
\begin{bmatrix}
I'_x \\
I'_y \\
I'_z
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
R_c \\
-R_cT_h
\end{bmatrix} \begin{bmatrix}
R_h \\
-R_hT_h
\end{bmatrix} \begin{bmatrix}
L_x \\
L_y \\
L_z
\end{bmatrix}
\]

In the above equations,
- \( \vec{T}_h \) is a 3D vector representing the position of the head in the world space.
- \( R_h \) is a 3x3 rotation matrix representing the orientation of the head in world space.
- \( \vec{T}_c \) is a 3D vector representing the position of the camera in the head coordinate system.
- \( R_c \) is a 3x3 rotation matrix representing the orientation of the camera in the head coordinate system.
- \( f \) is the focal length.
- \((L_x, L_y, L_z)\) is the position of a landmark in world space.
- \((I_x, I_y)\) is the projected position of the landmark in image space.
- \((I'_x, I'_y, I'_z)\) is the projected position of the landmark in homogeneous image space.
Over-determined case

- Initial orientation is converted to a quaternion
- Then a hyperplane is defined such that it is tangential to the unit hypersphere at the point corresponding to this initial quaternion
- Thus, 3D coordinates can be extracted from the hyperplane
Over-determined case

- If the system is over-determined, then a non-linear, least square minimization problem must be solved.
- Thus, the Levenberg-Marquardt algorithm was implemented in the system.
- With a good initial guess, an optimized solution is computed in only a few milliseconds.
Video Explanation
Results

- 3D copy and paste – a virtual copy is made of a real object
- Virtual object interacting with a real object (with light source tracked and shadow map calculated in real time)
- Medical usage – performing ultrasound-guided needle biopsies
Results
Video Results
Results
Results
More Videos

- More video results can be seen here:

  http://www.cs.unc.edu/~us/web/quicktime.htm
Limitations

- Suboptimal performance due to lack of synchronization between the magnetic tracker and vision-based system
- Slight time discrepancy between top and bottom scanlines (17msec)
- Changing lighting conditions can diminish the image analyzer’s performance
Future work (as of 1996)

- Attaching landmarks to moving objects
- Using the system at a different scale (larger environment)
- Real-time tracking of visually unobtrusive natural features
Current Augmented Reality
MARS - Mobile Augmented Reality Systems (Columbia University)

- Prototype campus information system. The user wears a backpack and headworn display, and holds a handheld display and its stylus.
ARToolkit (Univ. of Wash. ‘99)

- Vision-based system that uses “markers” to determine the camera orientation
ARToolkit in action
“Outdoor” ARToolkit

- A large marker was created so that it could be mounted on a building (San Diego Supercomputer Center)
- A 3D model of SDSC was created as our dataset
- By placing a marker on SDSC, ARToolkit could be used outdoors
Results
ARToolkit disadvantages

- ARToolkit is primarily for lab use or closed spaces where a computer is easily accessible.

- It would be difficult to create a unique physical “marker” for every possible articulated structure we are interested in.
Our approach (Clothier ’03)

- The Global Positioning System (GPS) provides useful information about where a person is
  - Knowing the location provides an easy way to determine what data to display

- Don’t use physical targets – rather use unique features of the building’s structure as the targets instead (through KLT tracking)
Results
Things I’ve found useful:

- Segmentation
- Tracking algorithms (KLT tracking)
- Thresholding
- Determining 3D coordinates from 2D images
- Optical flow
Questions/Demo Time?

Thank you for your attention!