Visual Marker Detection and Decoding in AR Systems: A Comparative Study

Matthew Walmer
Project 17: Robotic Ultrasound Needle Placement and Tracking
Project Background

- CAMP lab has designed a multi-robot surgical system.
- This mobile platform provides flexibility in an operating room environment.
- For multi-robot surgical procedures, precise coordination is key.
- Base to base calibration must be done frequently, because the platform is mobile.
- We need an efficient method to precisely calibrate multiple robots.

Objective

Explore a variety of robot-to-robot calibration methods and validate their efficacy for use in dual-robotic surgeries and experiments.


<table>
<thead>
<tr>
<th>Background</th>
<th>Criteria</th>
<th>Efficiency</th>
<th>Accuracy</th>
<th>Reliability</th>
<th>Usability</th>
<th>Summary</th>
</tr>
</thead>
</table>
The Paper


Goal
Assess the strengths and weaknesses of four marker tracking systems.

<table>
<thead>
<tr>
<th>Background</th>
<th>Criteria</th>
<th>Efficiency</th>
<th>Accuracy</th>
<th>Reliability</th>
<th>Usability</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Marker Systems

- ARToolKit (ATK)
- Hoffman marker system (HOM)
- Institut Graphische Datenverarbeitung (IGD)
- Siemens Corporate Research (SCR)
### Assessment Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Details</th>
</tr>
</thead>
</table>
| **Usability** | - How easily users can integrate the system into their applications.  
- What platforms does the system run on?  
- Scaling for applications using hundreds of markers. |
| **Efficiency** | - Running time to detect and decode a marker or multiple markers. |
| **Accuracy** | - Error in finding feature positions (marker corners) in the 2D image, measured in pixels.  
- Correctness in identifying markers in multi-marker trials. |
| **Reliability** | - Performance for non-ideal image conditions.  
- Wide angles, many markers, far away markers, poor focus. |
Efficiency

Table 1: Average processing time for marker recognition (ms/frame).

<table>
<thead>
<tr>
<th>Size</th>
<th>ROM/MPF</th>
<th>Atk</th>
<th>Hom</th>
<th>Igd</th>
<th>Scr</th>
<th>ScrT</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>68 × 68/1</td>
<td>4.1</td>
<td>5.1</td>
<td>6.2</td>
<td>11.6</td>
<td>3.5</td>
</tr>
<tr>
<td>×</td>
<td>61 × 70/1</td>
<td>4.1</td>
<td>4.9</td>
<td>6.4</td>
<td>11.9</td>
<td>3.1</td>
</tr>
<tr>
<td>240</td>
<td>188 × 148/3</td>
<td>7.1</td>
<td>10.3</td>
<td>—</td>
<td>14.9</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>257 × 207/10</td>
<td>23.9</td>
<td>35.5</td>
<td>—</td>
<td>21.9</td>
<td>—</td>
</tr>
<tr>
<td>640</td>
<td>200 × 200/1</td>
<td>13.1</td>
<td>13.6</td>
<td>19.8</td>
<td>58.2</td>
<td>22.1</td>
</tr>
<tr>
<td>×</td>
<td>514 × 414/10</td>
<td>41.6</td>
<td>51.0</td>
<td>—</td>
<td>72.9</td>
<td>—</td>
</tr>
<tr>
<td>480</td>
<td>258 × 218/10</td>
<td>33.3</td>
<td>41.3</td>
<td>—</td>
<td>58.5</td>
<td>—</td>
</tr>
</tbody>
</table>

- ROM = region of markers (pixels)
  - “The smallest rectangular region that contains all the markers in the image.”
- MPF = markers per frame
- Technical difficulties with IGD for multiple markers
- ScrT is a special “tracking mode” for SCR. Only works for single marker.
Accuracy

- Did not perform tests for accuracy of 3D poses.
  - It’s very difficult to determine a ground truth for this.

- For 2D pixel error, they established two methods to create “ground truths” for marker corner positions:
  - OpenCV corner detection (OCV)
  - Edge detection, least square line fitting, and intersection (LIT)
Defining two different ground truths gives ambiguous results.

- SCR was best with respect to the LIT points.
- IGD was best with respect to the OCV points.
- ATK had the highest error under both ground truths.
  - They theorize that this is a result of ATK’s binary image processing.
Reliability

Projective Distortion

Multiple Markers

Small Region of Marker

Poor Focus
Reliability - Projective Distortion

Table 4: Marker recognition rate under prospective distortion (%).

<table>
<thead>
<tr>
<th>angle</th>
<th>Atk</th>
<th>Hom</th>
<th>Igd</th>
<th>Scr</th>
</tr>
</thead>
<tbody>
<tr>
<td>90°</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>75°</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>60°</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>45°</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>30°</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>15°</td>
<td>71/(cf\geq0.50)</td>
<td>16/(cf\geq0.67)</td>
<td>8/(cf\geq0.75)</td>
<td>100</td>
</tr>
</tbody>
</table>

- Comparable performance up to 30 degrees.
- ATK has confidence threshold that can be configured.
- HOM has a similar scale from 0 to 6.
- SCR uses HOM’s confidence scale but only accepts high confidence values.
Reliability – Multiple Markers

- IGD omitted due to technical difficulties.
- ATK has a tendency to incorrectly identify similar markers.
- The confidence value can be high for misidentified markers.
- This is a result of ATK’s fast but simple template matching system.
- The other systems did not misidentify any markers.

Table 5: Marker recognition rate with multiple markers (%).

<table>
<thead>
<tr>
<th>Size</th>
<th>ROM/MPF</th>
<th>Atk/cf</th>
<th>Hom</th>
<th>Scr</th>
</tr>
</thead>
<tbody>
<tr>
<td>640 ×</td>
<td>(514 × 414)/10</td>
<td>90/(cf≥0.50)</td>
<td>100</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59/(cf&gt;0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>46/(cf≥0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>480 ×</td>
<td>(258 × 218)/10</td>
<td>83/(cf≥0.50)</td>
<td>100</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38/(cf&gt;0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29/(cf≥0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>320 ×</td>
<td>(257 × 207)/10</td>
<td>83/(cf≥0.50)</td>
<td>100</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39/(cf&gt;0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14/(cf≥0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>240 ×</td>
<td>(188 × 148)/3</td>
<td>100/(cf≥0.50)</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>86/(cf≥0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58/(cf≥0.75)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reliability – Small Region of Marker

- Gradually zoomed out camera until each marker could not be recognized.
- ATK had the best performance.
- HOM and SCR performed comparably.
- IGD needed a much larger region to detect the marker.

Figure 18: Marker recognition with small region of interest (image size $320 \times 240$ pixels).
Reliability – Poor Focus

ATK’s confidence values were very questionable.
• Showed higher confidence for the most unfocused images.
• HOM’s confidence metric was very reliable.

Table 6: Recognition rate with poorly focused videos (%).

<table>
<thead>
<tr>
<th>Focus</th>
<th>Atk</th>
<th>Hom</th>
<th>Igd</th>
<th>Scr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect</td>
<td>100 (cf=0.79)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Good</td>
<td>100 (cf=0.81)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Bad</td>
<td>100 (cf=0.63)</td>
<td>100</td>
<td>28</td>
<td>97</td>
</tr>
<tr>
<td>Worse</td>
<td>100 (cf=0.56)</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Worst</td>
<td>100 (cf=0.73)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Usability

• ATK was ranked highest for usability:
  – open-source
  – works on most platforms
  – very well documented.

• A downside to ATK:
  – Custom markers require extra pattern registration.
  – The other marker systems use systematic grid patterns.
  – They can generate thousands of distinct markers with no extra steps.
  – ATK does not scale well to applications requiring hundreds of markers.

• The other programs have limited availability, and do not have good multi-platform support.

• They encountered difficulties with IGD’s multi-marker tracking, but humbly attributed it to their “own unfamiliarity to the IGD system.”
Qualitative Discussion and Summary

- **ATK**
  - Open source, well documented and widely compatible.
  - Fastest detection and decoding, but for a cost.
  - Lower accuracy and misidentified markers.
  - Custom markers require extra registration step.
  - Good for prototyping and simple AR applications.

- **HOM**
  - Good speed and accuracy.
  - Excellent detection and decoding.
  - Reliable confidence metric.

- **IGD**
  - Good speed and high accuracy.
  - Inconvenient to run in Windows.

- **SCR**
  - Slowest system, but much faster in tracking mode.
  - High accuracy.
Assessment of the Paper

Pros:

• Established clear criteria for assessing the systems.

• The wide range of experiments succeeded in bringing out the strengths, weaknesses and quirks of these systems.

• Clearly demonstrated issues with ATK’s marker identification and confidence metric.

• Gave a detailed, qualitative summary of each system’s performance.

Cons:

• Some further tests are needed:
  – 3D pose error
  – Variable lighting
  – Noisy and cluttered images

• Needs a more thorough exploration of the usability criteria for HOM, IGD and SCR.

• More analysis of ATK’s systematic corner position errors.

• Two “ground truths” in accuracy testing.
Last Comments

• For our application, we use ARToolKit because it is the most accessible.
  – Good for prototyping.
  – We only need a few unique markers.
  – We may try other systems if ARToolKit is too inaccurate in practice.

• ARToolKit is by far the most frequently used of these systems, largely because it is open source and well maintained.

• The other three systems are less easy to find.

• There is a modified version of ARToolKit called ARToolKitPlus which adds systematic markers like those used by the other systems.
Questions?