CIS II Seminar Presentation

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

Matthew Walmer

Project 17: Robotic Ultrasound Needle Placement and Tracking

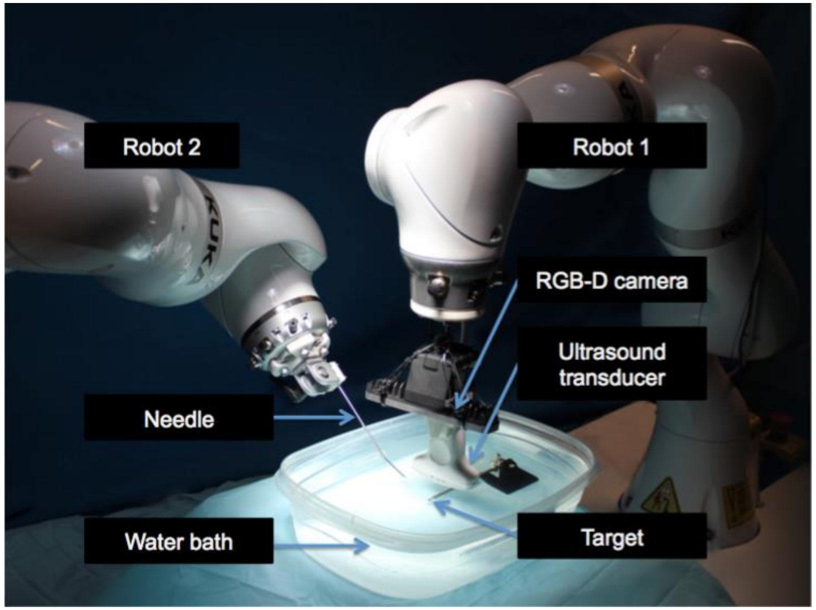
Technische Universität München
+
Johns Hopkins
Whiting School of Engineering

Project Background

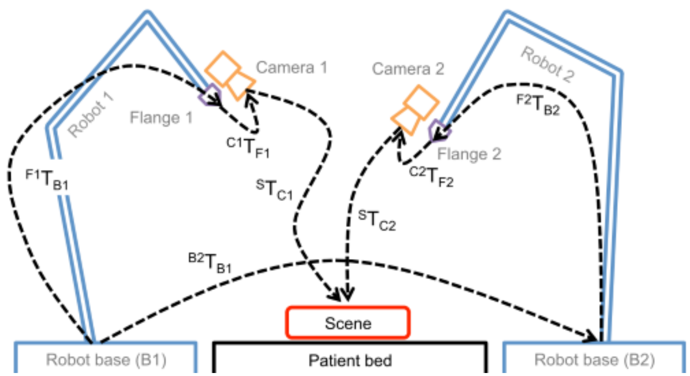
- CAMP lab has designed a multi-robot surgical system.
- This mobile platform provides flexibility in an operating room environment.
- For multi-robot surgical procedures, precise coordination is key.
- Base to base calibration must be done frequently, because the platform is mobile.
- We need an efficient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



From: R. Kojcev, B. Fuerst, O. Zettinig, J.Fotouhi, C. Lee, R.Taylor, E. Sinibaldi, N. Navab, "Dual-Robot Ultrasound-Guided Needle Placement: Closing the Planning-Imaging-Action Loop," Unpublished Manuscript.



$$ {}^{B2}T_{B1} = {}^{F2}T_{B2}^{-1}\ {}^{C2}T_{F2}^{-1}\ {}^{S}T_{C2}^{-1}\ {}^{S}T_{C1}\ {}^{C1}T_{F1}\ {}^{F1}T_{B1} $$

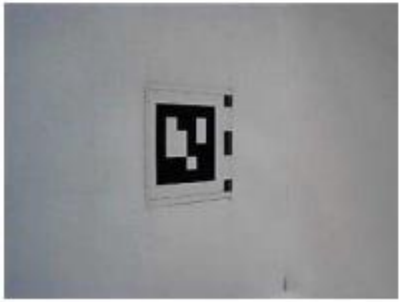
The Paper

Zhang, Xiang, Stephan Fronz, and Nassir Navab. "Visual marker detection and decoding in AR systems: A comparative study." Proceedings of the 1st International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2002.

Goal

Assess the strengths and weaknesses of four marker tracking systems.

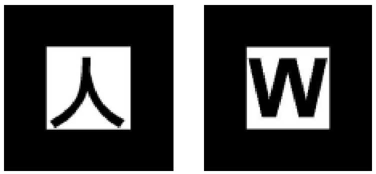

(a) ATK


(b) HOM

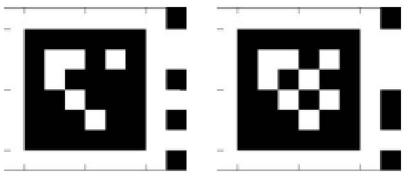

(c) IGD


(d) SCR

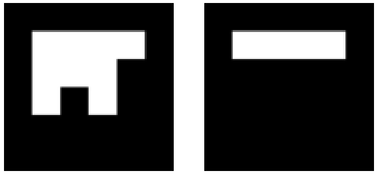
Marker Systems



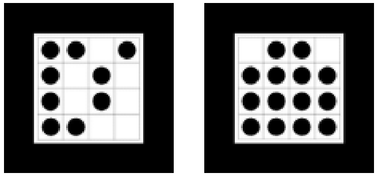
- ARToolKit (ATK)



- Hoffman marker system (HOM)



- Institut Graphische Datenverarbeitung (IGD)



- Siemens Corporate Research (SCR)

Assessment Criteria

- 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 ($ms/frame$).

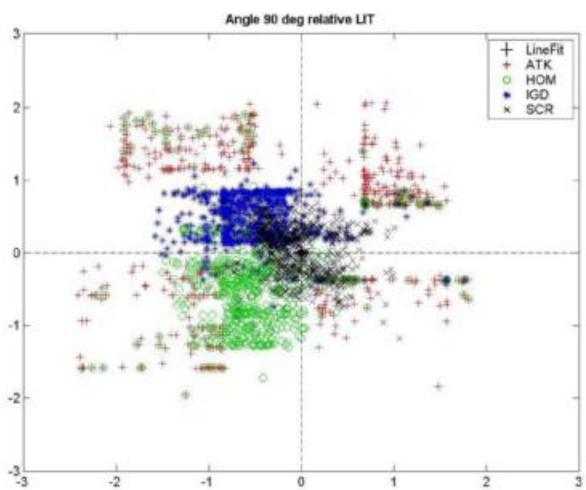
Size	ROM/MPF	Atk	Hom	Igd	Scr	ScrT
320 × 240	$68 \times 68/1$	4.1	5.1	6.2	11.6	3.5
	$61 \times 70/1$	4.1	4.9	6.4	11.9	3.1
	$188 \times 148/3$	7.1	10.3	—	14.9	—
	$257 \times 207/10$	23.9	35.5	—	21.9	—
640 × 480	$200 \times 200/1$	13.1	13.6	19.8	58.2	22.1
	$514 \times 414/10$	41.6	51.0	—	72.9	—
	$258 \times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

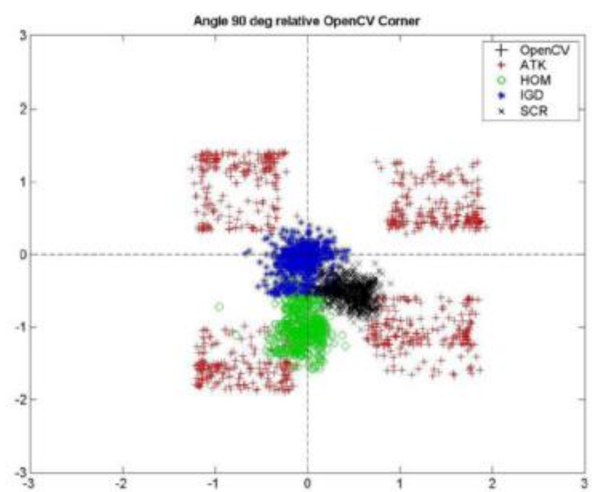
- ROM = region of markers (pixels)
  - “The smallest rectangular region that contains all the markers in the image.”
- MPF = markers per frame
- Technical difficulties with IGD for multiple markers
- ScrT is a special “tracking mode” for SCR. Only works for single marker.

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)





Accuracy

Table 2: '*Errors*' related to the LIT points (Average Distance)/(Standard Deviation) (in pixels).

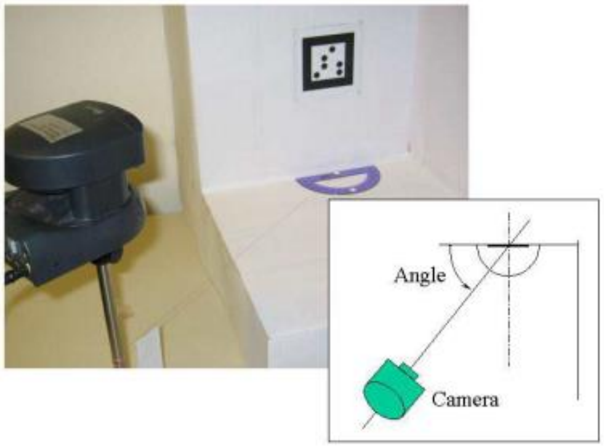
angle	Atk	Hom	Igd	Scr
90°	1.43/0.57	0.98/0.48	0.84/0.29	0.46/0.26
75°	1.43/0.51	0.84/0.43	0.85/0.31	0.58/0.34
60°	1.27/0.43	0.88/0.35	0.84/0.33	0.61/0.34
45°	1.57/0.42	0.92/0.46	0.99/0.44	0.63/0.40
30°	1.28/0.39	0.92/0.37	0.94/0.38	0.63/0.32
Avg.	1.40/0.46	0.91/0.42	0.89/0.35	0.58/0.33

Table 3: '*Error*' related to the OCV ~~LIT~~ points (Average Distance)/(Standard Deviation) (in pixels).

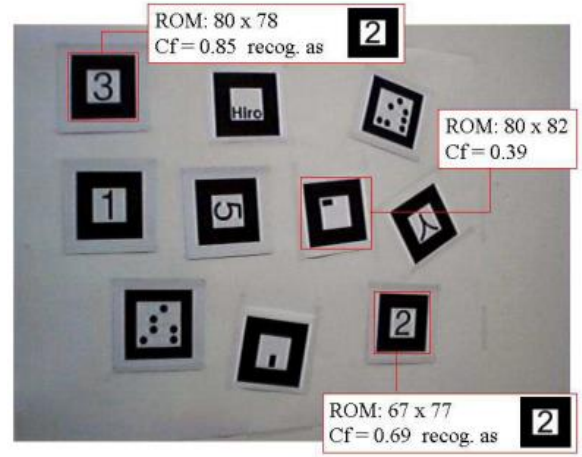
angle	Atk	Hom	Igd	Scr
90°	1.55/0.32	1.22/0.14	0.17/0.10	0.59/0.14
75°	1.44/0.07	1.12/0.05	0.22/0.11	0.59/0.09
60°	1.42/0.12	1.17/0.05	0.37/0.13	0.78/0.24
45°	1.50/0.21	1.05/0.13	0.39/0.11	0.53/0.12
30°	1.23/0.18	1.16/0.06	0.44/0.24	0.71/0.17
Avg.	1.43/0.18	1.14/0.09	0.32/0.14	0.64/0.15

- 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

(a) Large ROM (b) Small ROM

Small Region of Marker

(1) Perfect (5) Worst

Poor Focus

Reliability - Projective Distortion

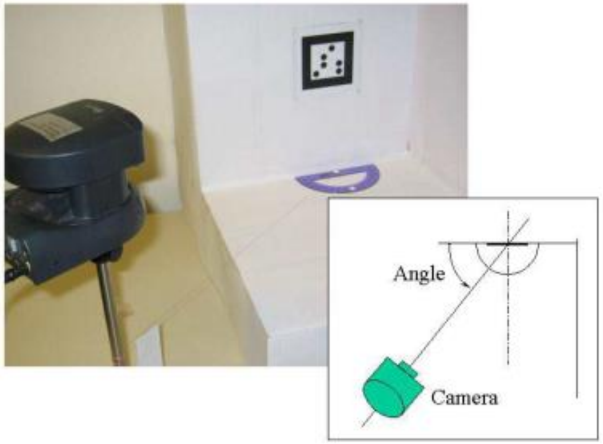


Table 4: Marker recognition rate under prospective distortion (%).

angle	Atk	Hom	Igd	Scr
90°	100	100	100	100
75°	100	100	100	100
60°	100	100	100	100
45°	100	100	100	98
30°	100	100	100	95
15°	71/(cf≥0.50) 16/(cf≥0.67) 8/(cf≥0.75)	100	0	7

- 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

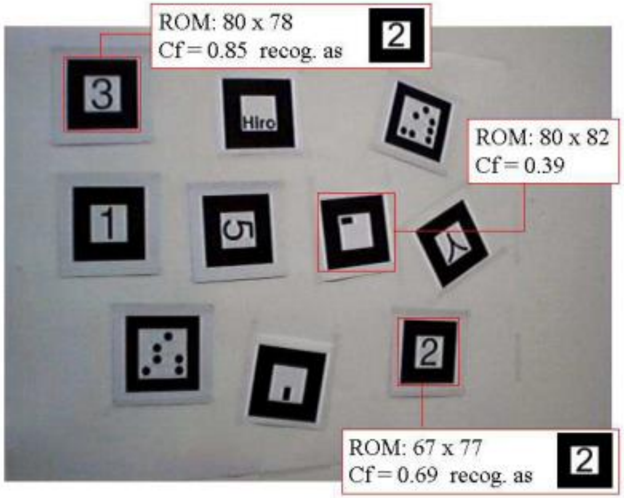


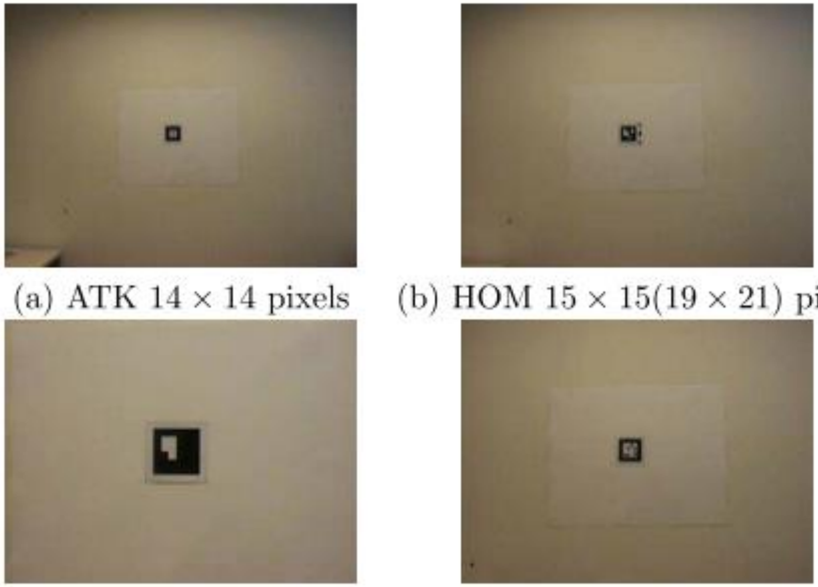
Table 5: Marker recognition rate with multiple markers (%).

Size	ROM/MPF	Atk/cf	Hom	Scr
640 $\times$ 480	$(514 \times 414)/10$	90/(cf≥0.50) 59/(cf≥0.67) 46/(cf≥0.75)	100	81
	$(258 \times 218)/10$	83/(cf≥0.50) 38/(cf≥0.57) 29/(cf≥0.75)	100	72
320 $\times$ 240	$(257 \times 207)/10$	83/(cf≥0.50) 39/(cf≥0.67) 14/(cf≥0.75)	100	52
	$(188 \times 148)/3$	100/(cf≥0.50) 86/(cf≥0.67) 58/(cf≥0.75)	100	93

- 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.

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.



(a) ATK 14×14 pixels (b) HOM $15 \times 15 (19 \times 21)$ pixels (c) IGD 44×44 pixels (d) SCR 21×21 pixels

Figure 18: Marker recognition with small region of interest (image size 320×240 pixels).

Reliability – Poor Focus

(1) Perfect

(2) Good

(3) Bad

(4) Worse

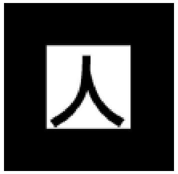
Table 6: Recognition rate with poorly focused videos (%).

Focus	Atk	Hom	Igd	Scr
Perfect	100 (cf=0.79)	100	100	100
Good	100 (cf=0.81)	100	100	100
Bad	100 (cf=0.63)	100	28	97
Worse	100 (cf=0.56)	0	12	0
Worst	100 (cf=0.73)	0	0	0

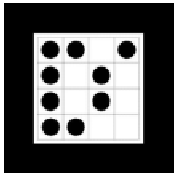
- ATK's confidence values were very questionable.
- Showed higher confidence for the most unfocused images.
- HOM's confidence metric was very reliable.

Usability

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



(ATK)



(SCR)

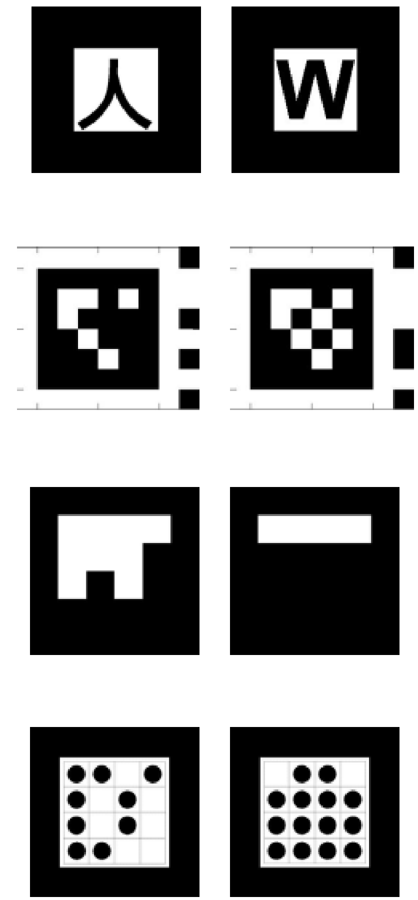
- A downside to ATK:
  - Custom markers require extra pattern registration.
  - The other marker systems use systematic grid patterns.
  - They can generate thousands of distinct markers with no extra steps.
  - ATK does not scale well to applications requiring hundreds of markers.

- 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?