# Subjective Evaluation of a Semi-Automatic Optical See-Through Head-Mounted Display Calibration Technique


Kenneth Moser
*Mississippi State University*

Yuta Itoh
*Technical University of Munich*

Kohei Oshima
*Nara Institute of Science & Technology*

J. Edward Swan II
*Mississippi State University*

Gudrun Klinker
*Technical University of Munich*

Christian Sandor
*Nara Institute of Science & Technology*



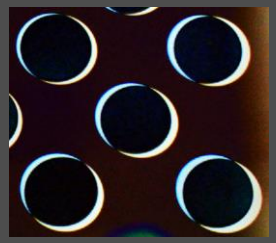


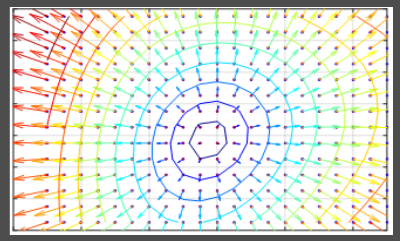

IEEE Virtual Reality 2015
Arles, France

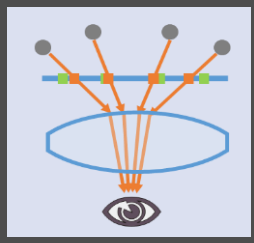
March 26, 2015

# Summary of our 3 talks

1. Better
   **HMD** model

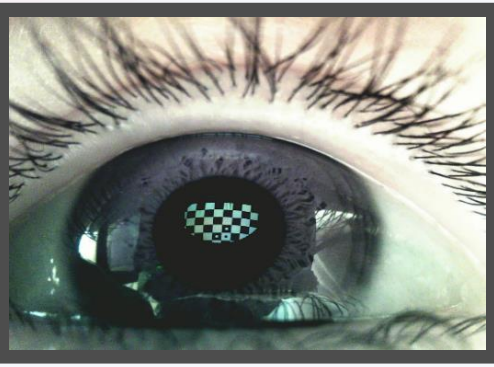


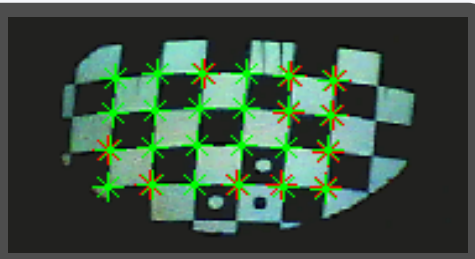




By Yuta

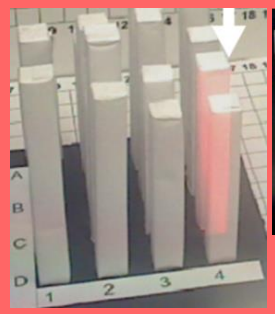
2. Better
   **Eye** localization

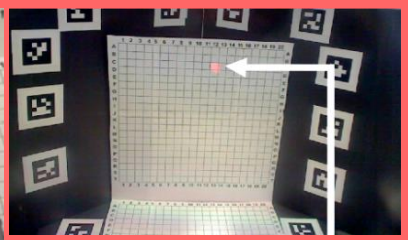




By Alex

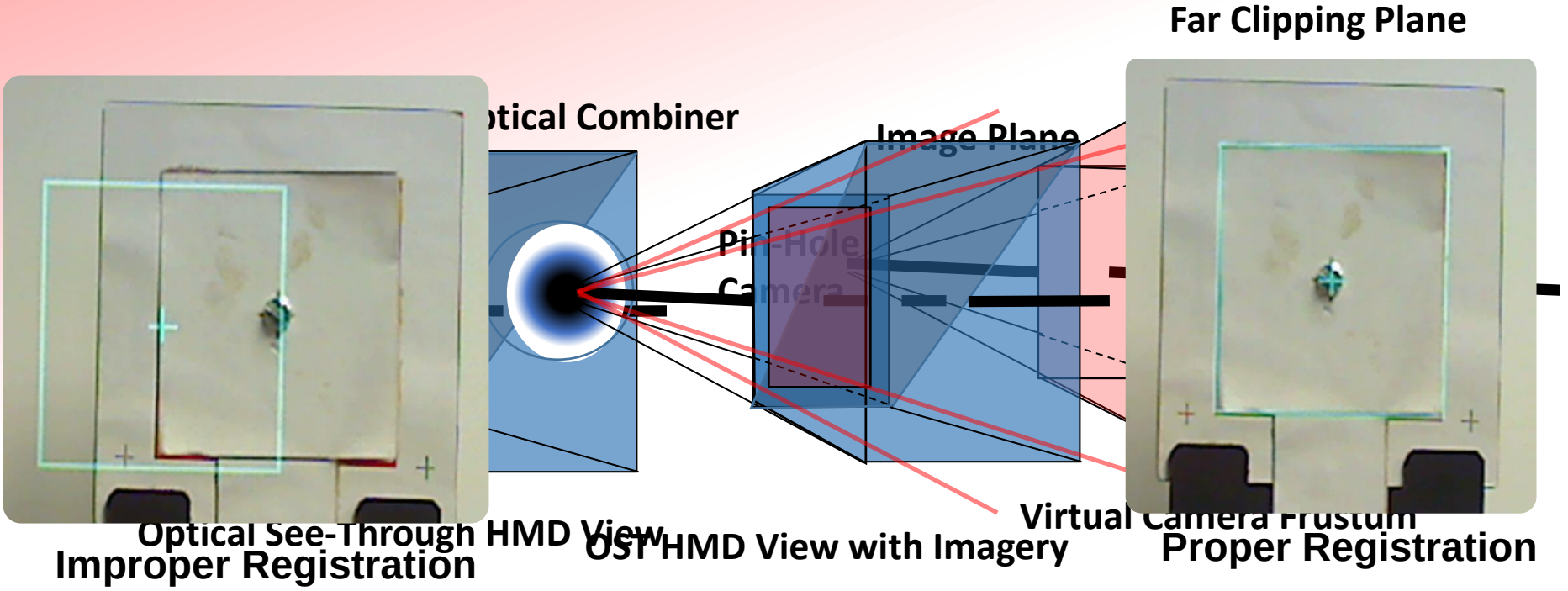
3. Formal
   **User** study



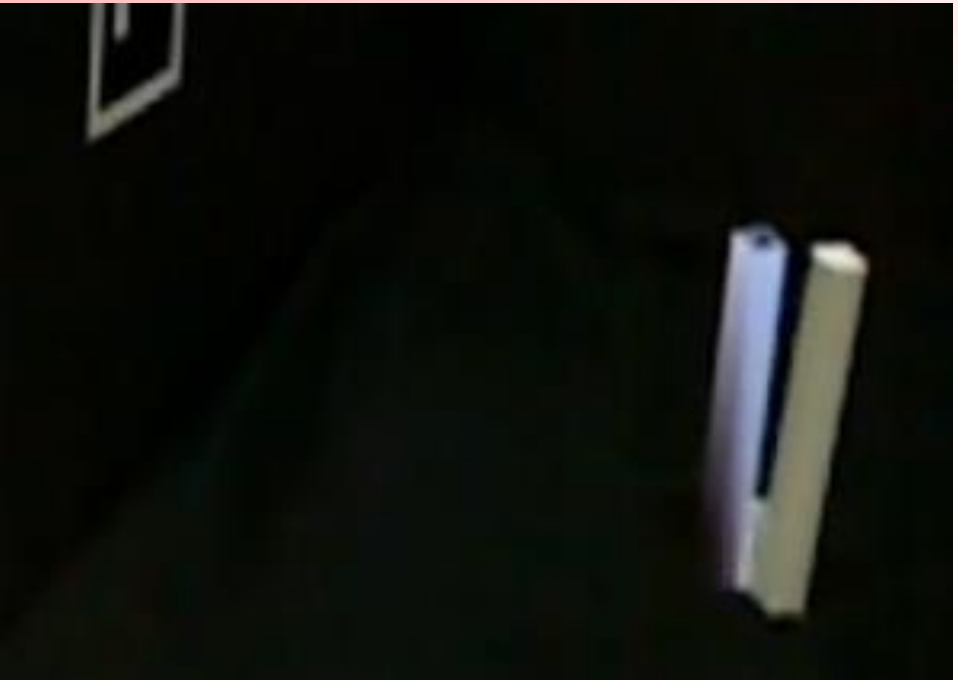
By Kenny

# OST-HMD Calibration



# OST-HMD Calibration

## Video Taken Through Camera Set Within The Display



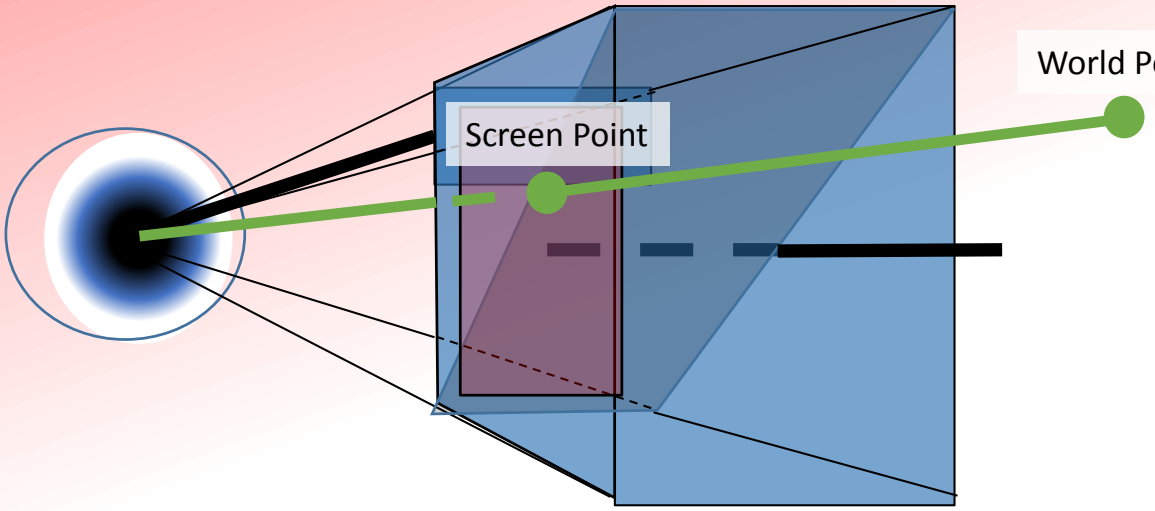
Less Accurate Registration

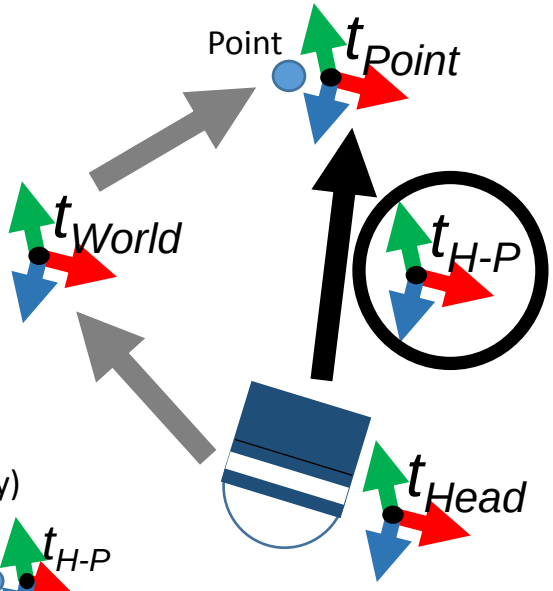
More Accurate Registration

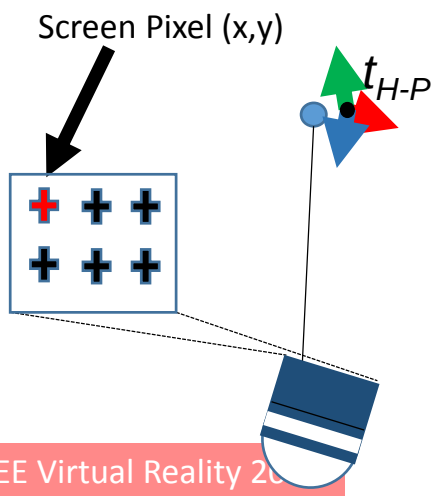
# Calibration Methods

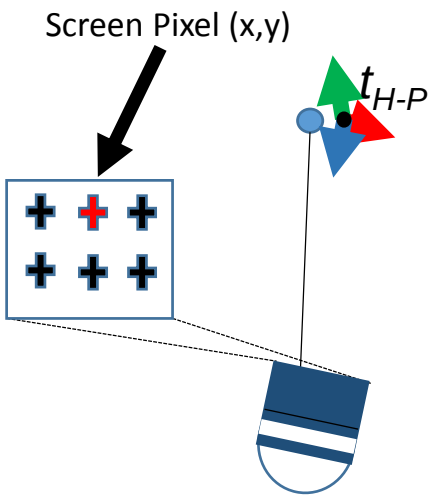
## Single Point Active Alignment Method (SPAAM)

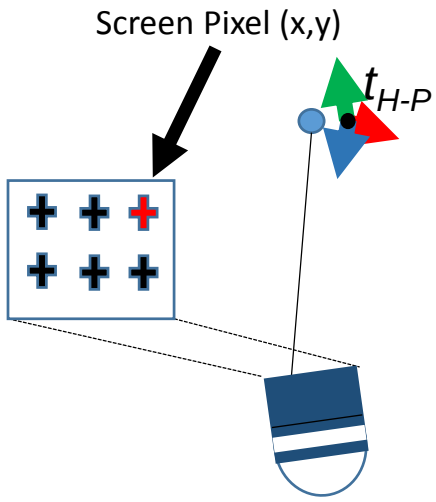
(Tuceryan & Navab, 2000)



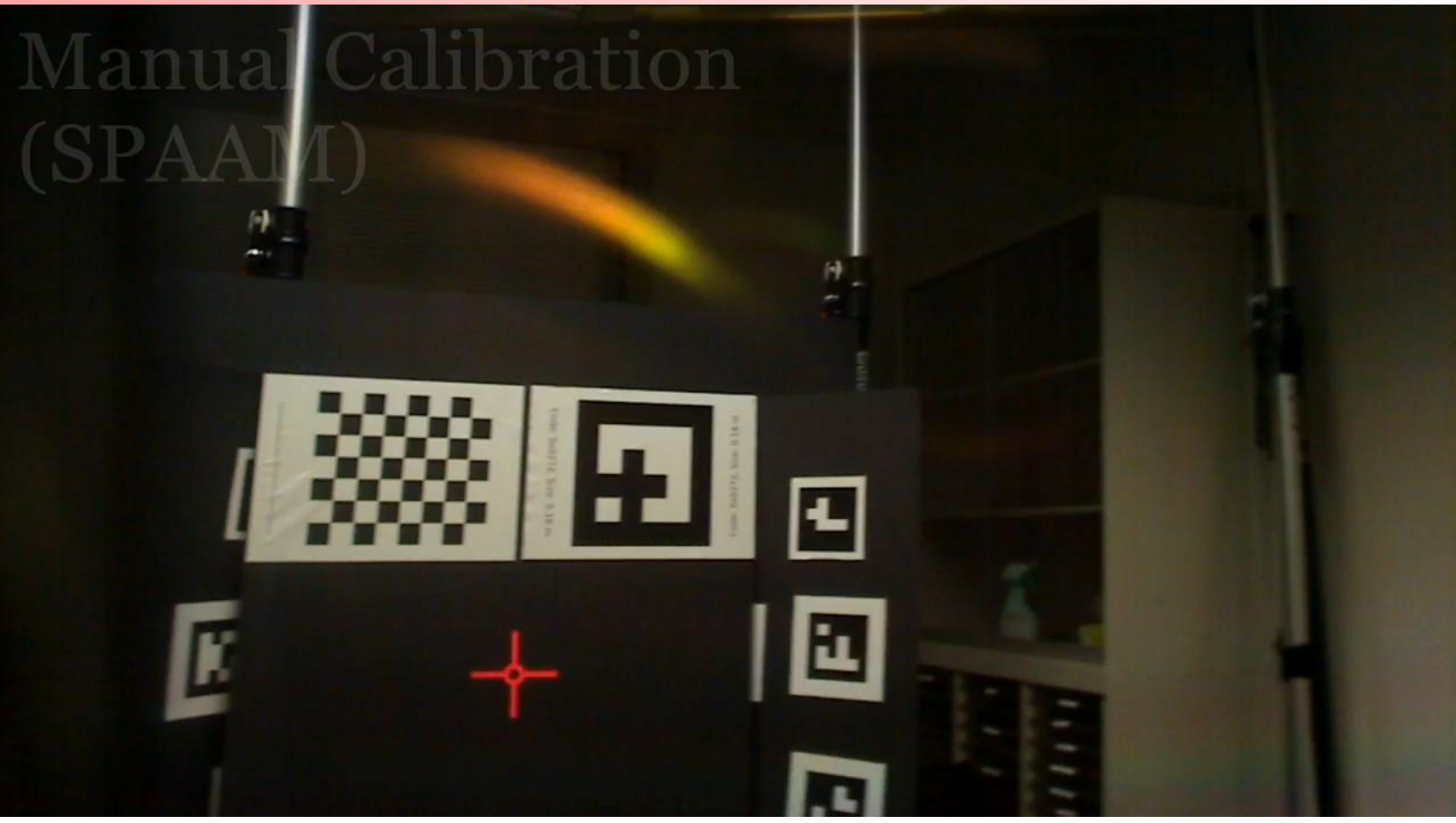



# Calibration Methods

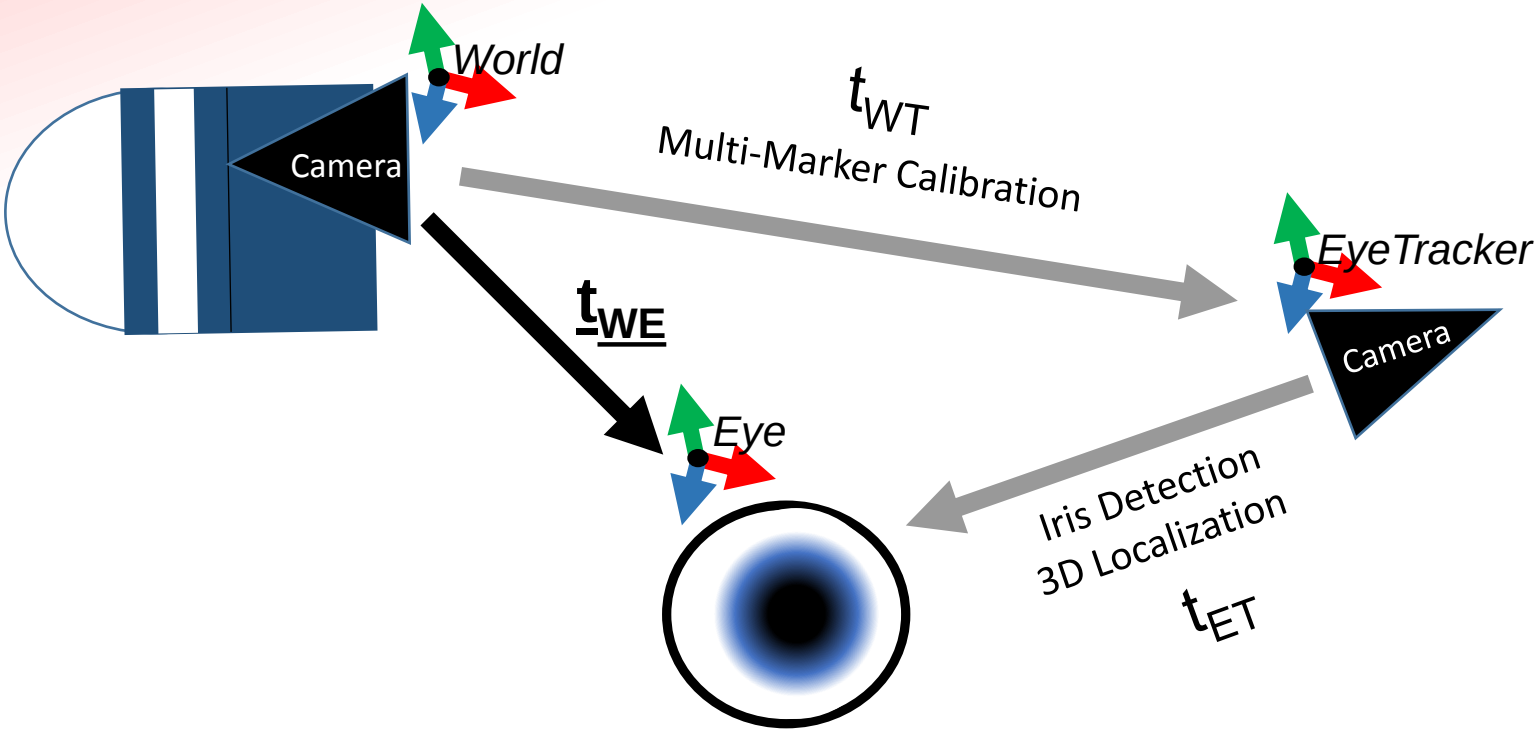


# Calibration Methods

Interaction Free Display Calibration (INDICA)

(Itoh & Klinker, 2014)

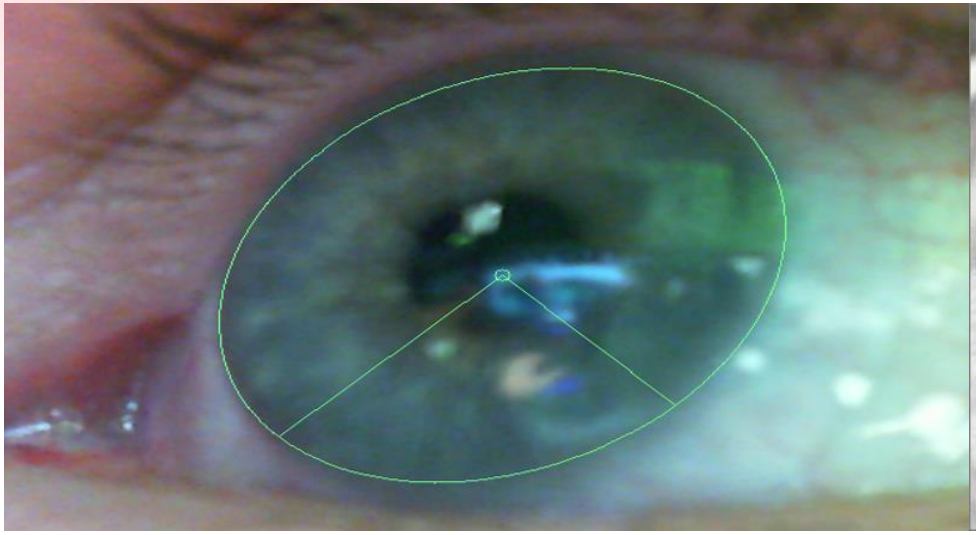
**Recycled INDICA**: Updates Calibration Matrix With Eye Location



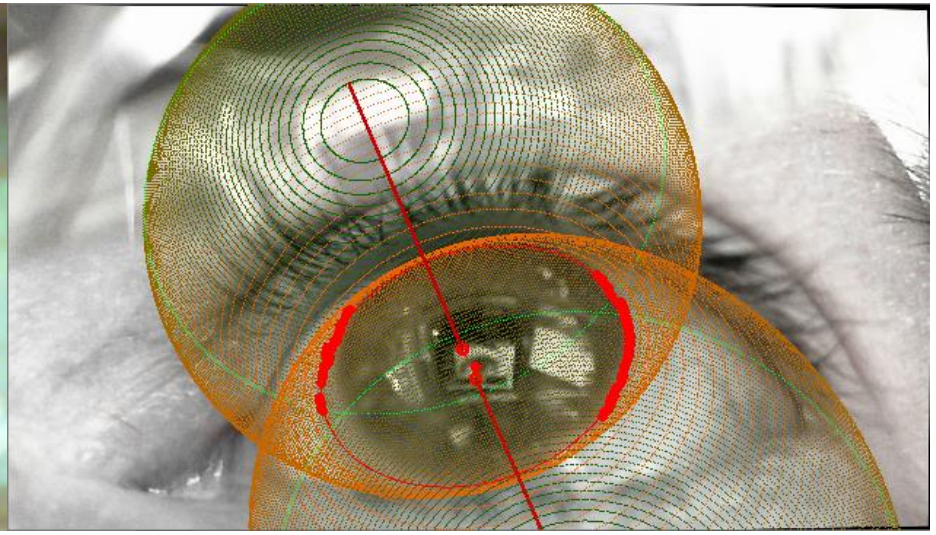
# Calibration Methods

Interaction Free Display Calibration (INDICA)
(Itoh & Klinker, 2014)

Eye Center Locations Determined Through Limbus Detection



Swirski, 2012



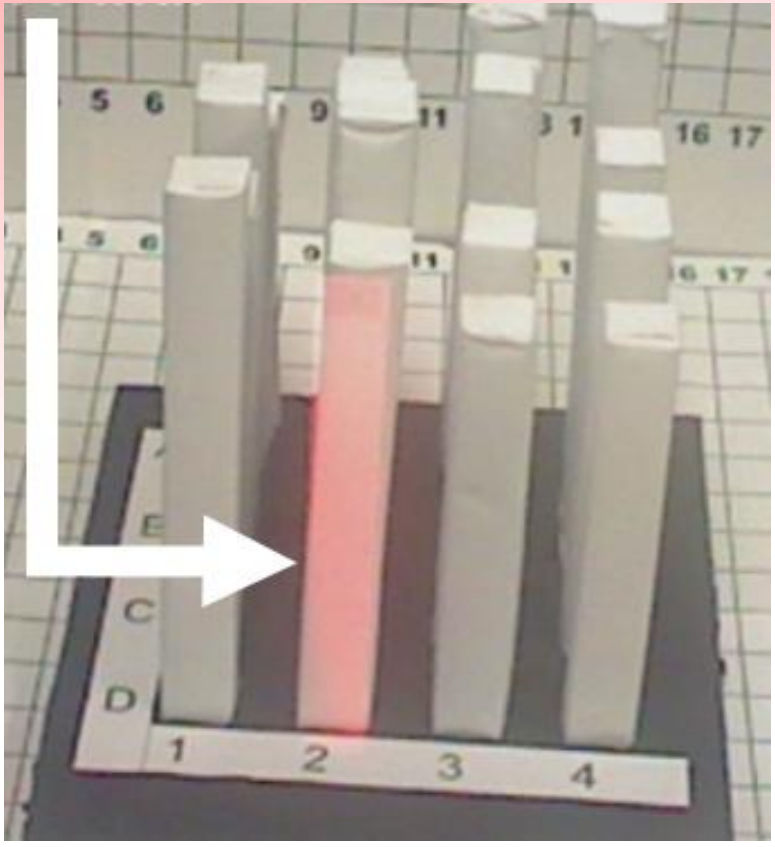
Nitschke, 2013

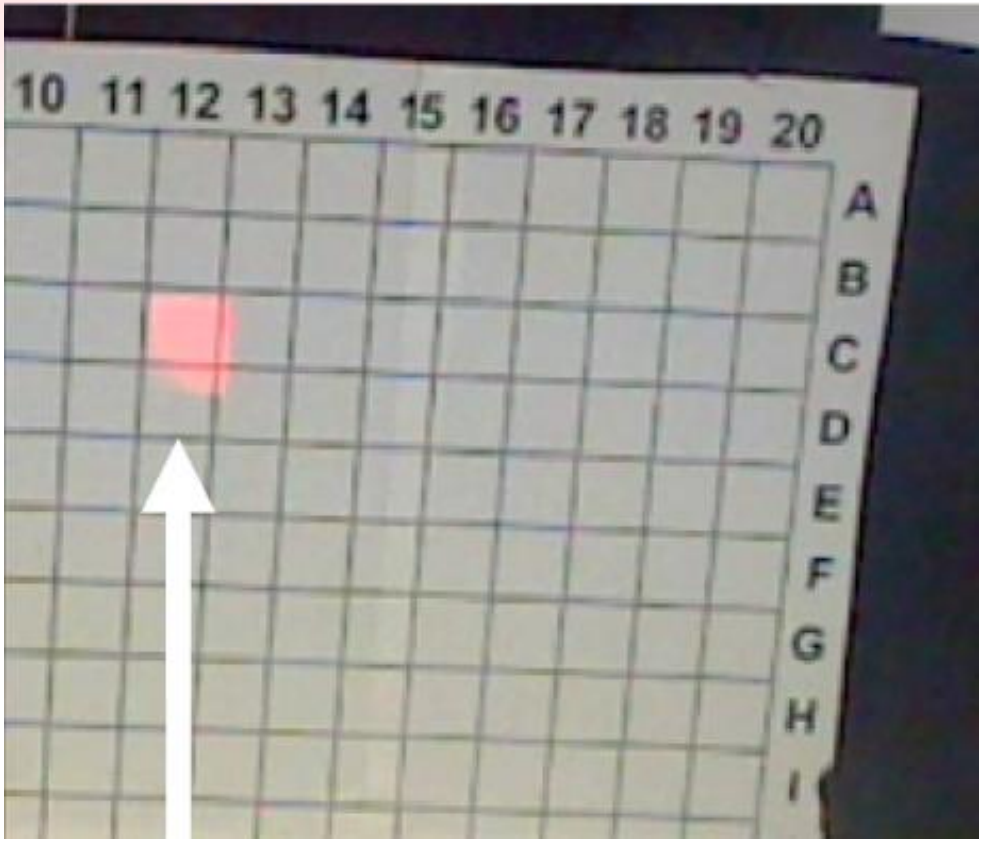
# Calibration Methods Evaluated

- SPAAM:
  - 20 screen - world alignments
  - taken over 1.5m – 3.0m distance to world point
- Degraded SPAAM:
  - reuse of SPAAM result
  - HMD removed and replaced
- Recycled INDICA:
  - Reuse intrinsic values from SPAAM calibration
  - Combine updated Eye Position for final result

# Subjective Evaluation Metrics

## perceived location error




*Virtual Objects in RED*

# Subjective Evaluation Metrics

## quality of registration with perceived location

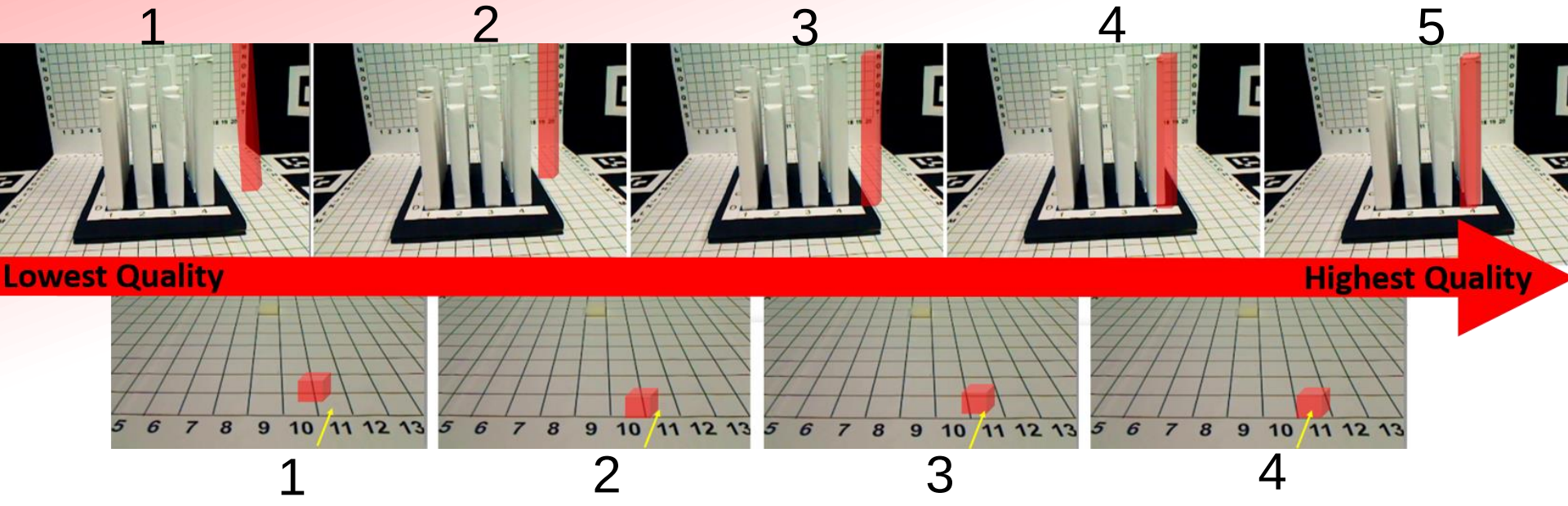
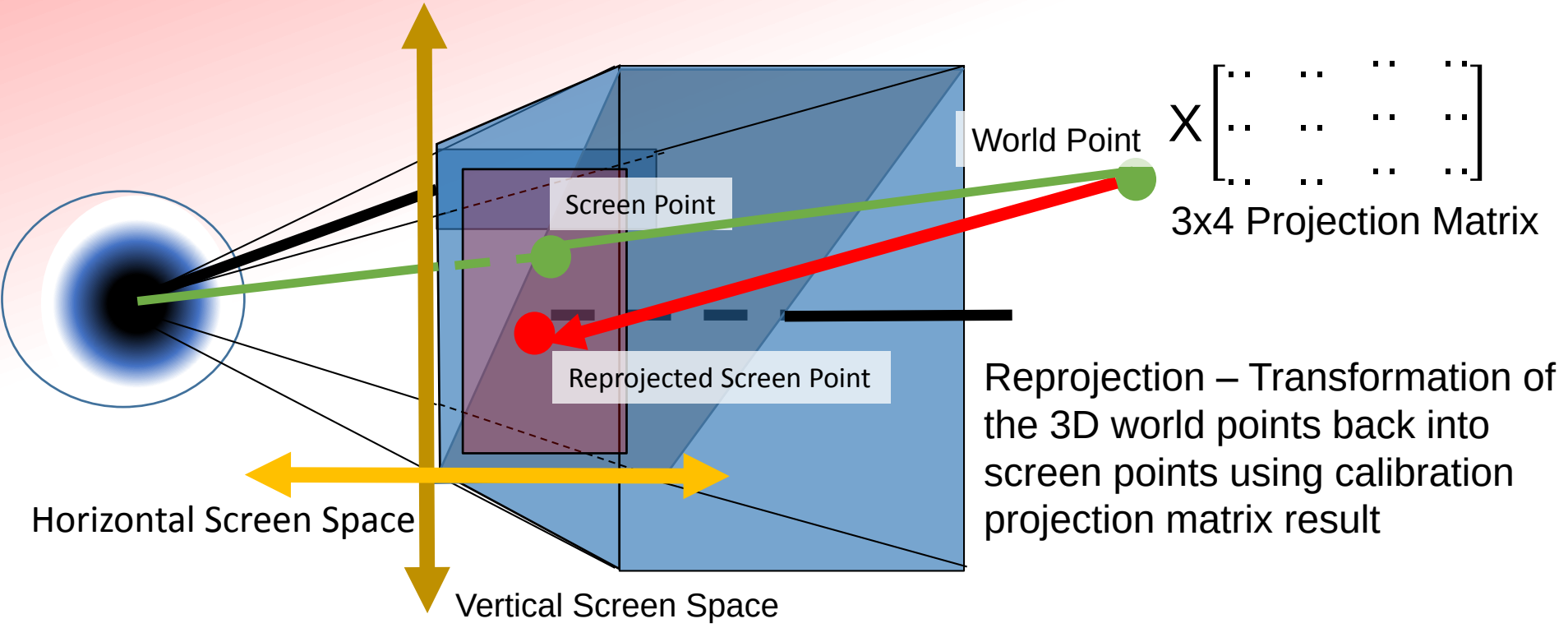



*diagram provided to subject before start of each trial set*

# Quantitative Evaluation Metrics

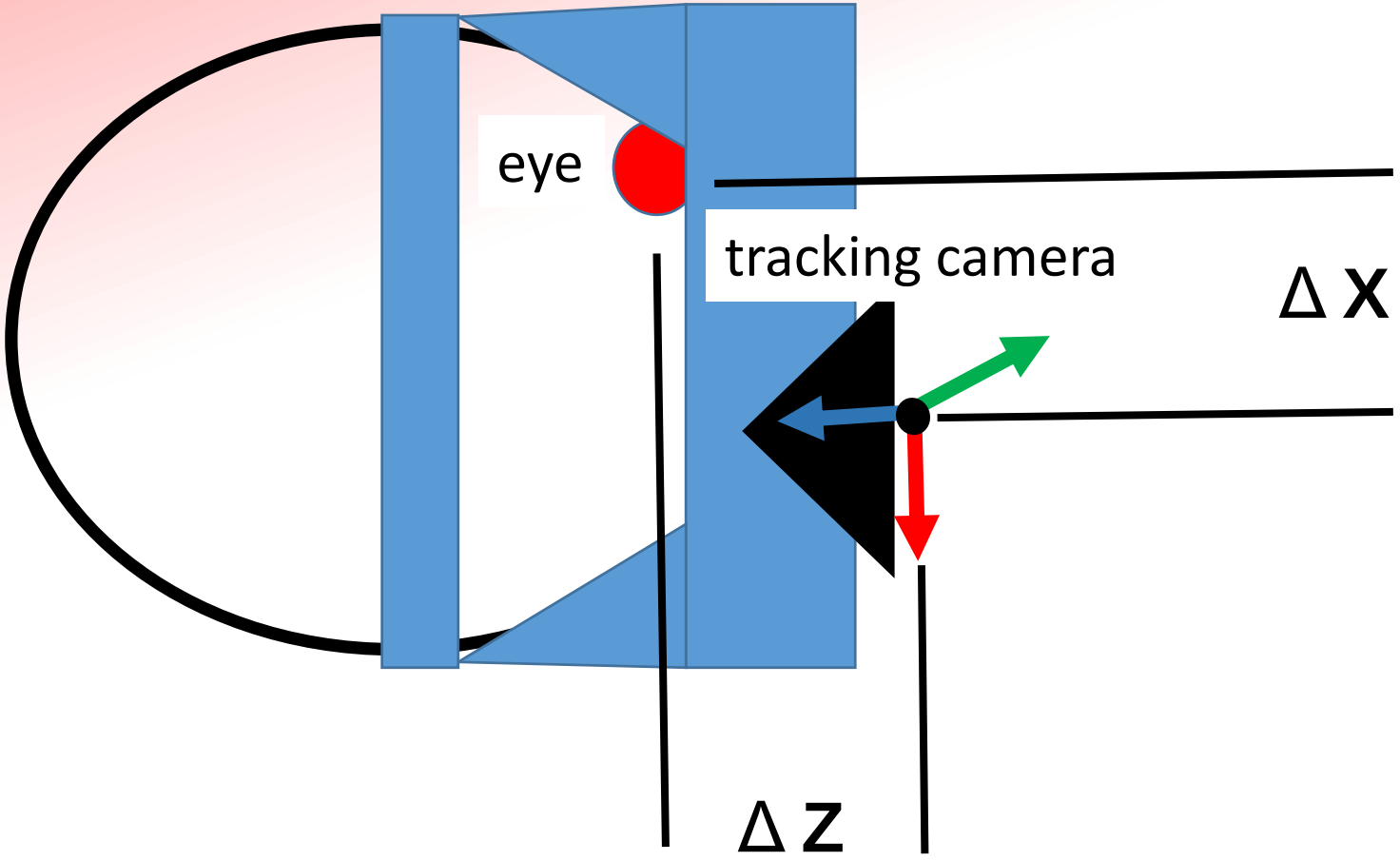
## variance of SPAAM alignment point reprojection




Reprojection – Transformation of the 3D world points back into screen points using calibration projection matrix result

# Quantitative Evaluation Metrics

## variance in eye location estimates



# System Hardware

**NVIS ST50**

1280 x 1024 per eye

HFOV $40^0$ / VFOV $32^0$

Left Eye (Monocular)

Right Eye Piece (Covered)

# System Hardware


**Logitech QuickCam Pro**
USB 2.0 Interface
Auto-Focus Disabled

World Tracking (head)
640 x 360 30fps

Eye Localization (left eye)
1280 x 720 (still images)

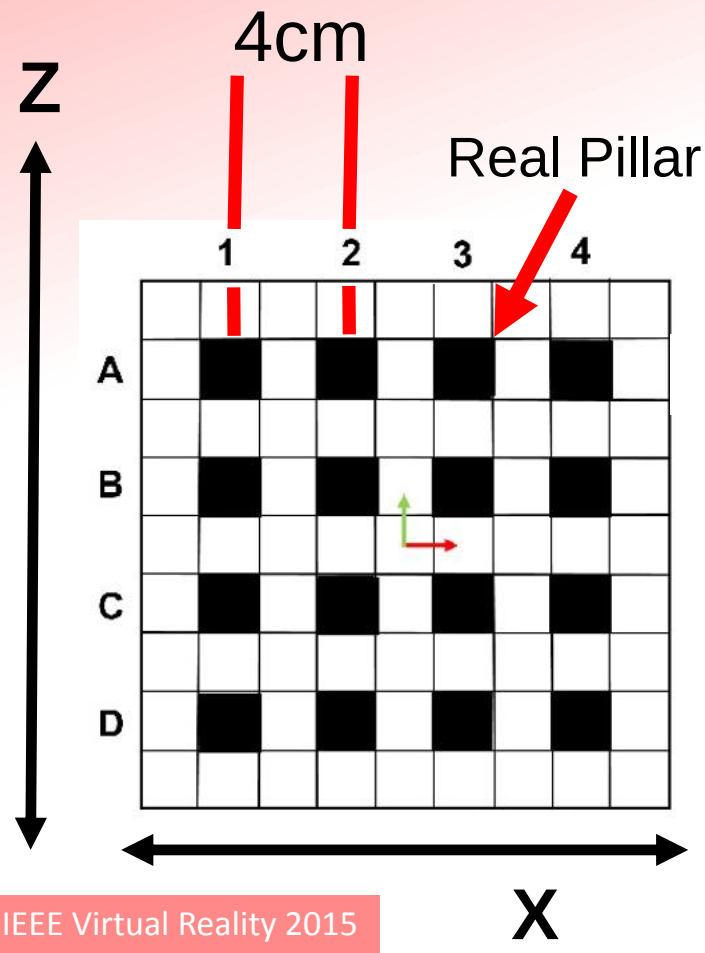
# Pillar Evaluation Tasks

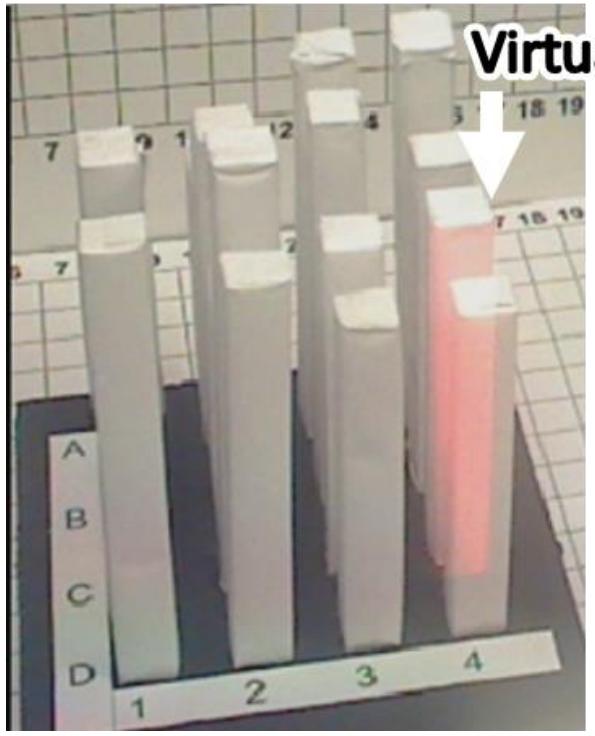
**verbally state location and registration quality**

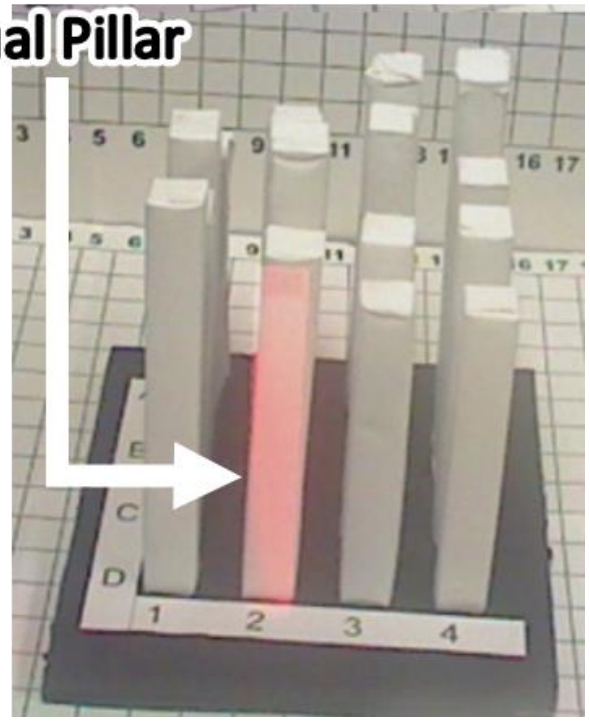
4 x 4 Grid of Pillars – 4cm spacing

Real Heights: 13.5 – 19.5cm

Virtual Height:15.5cm



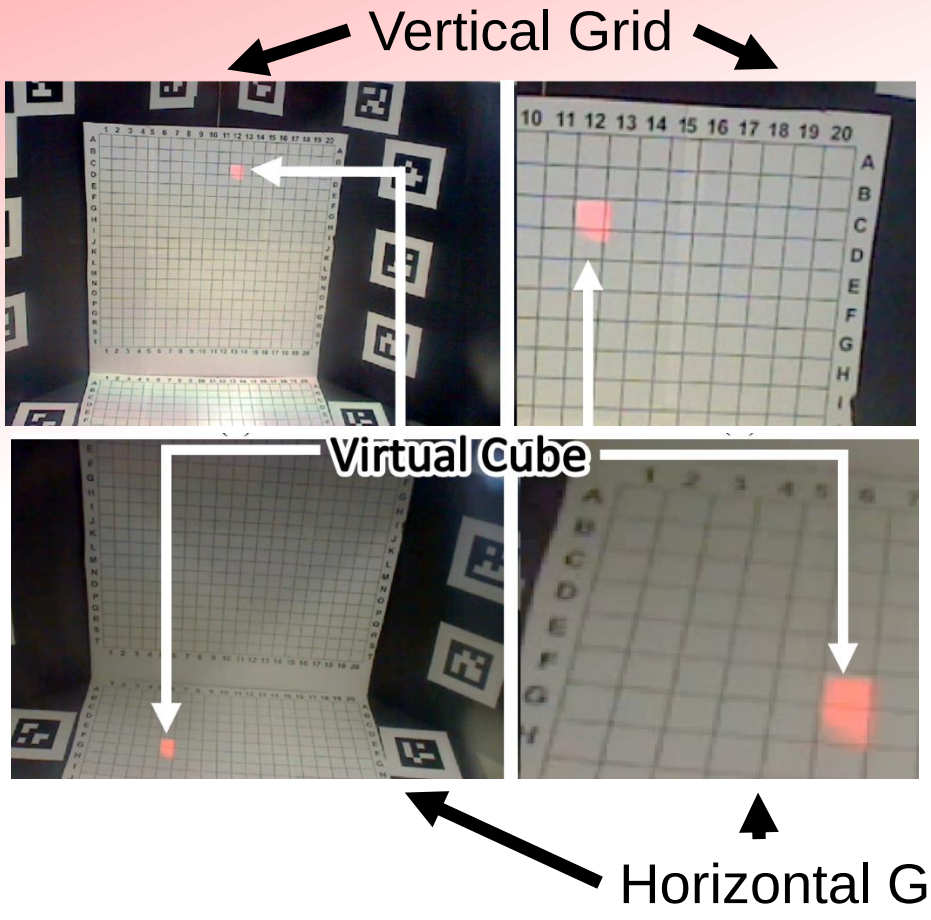


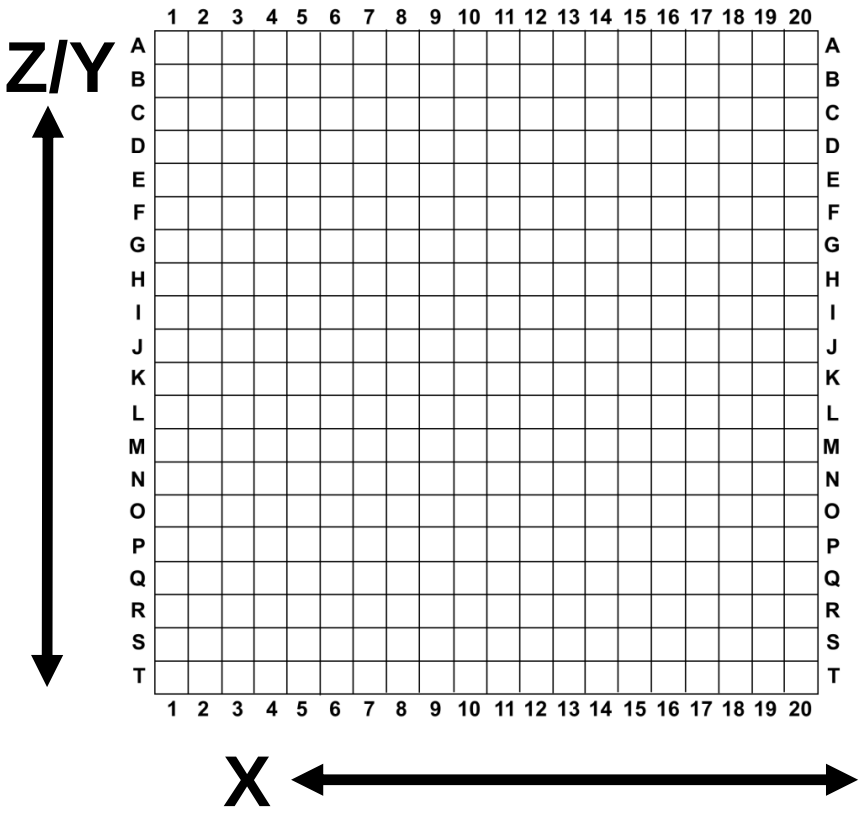
# Cube Evaluation Tasks

**verbally state location and registration quality**

Virtual Cube: 2cm x 2cm x 2cm

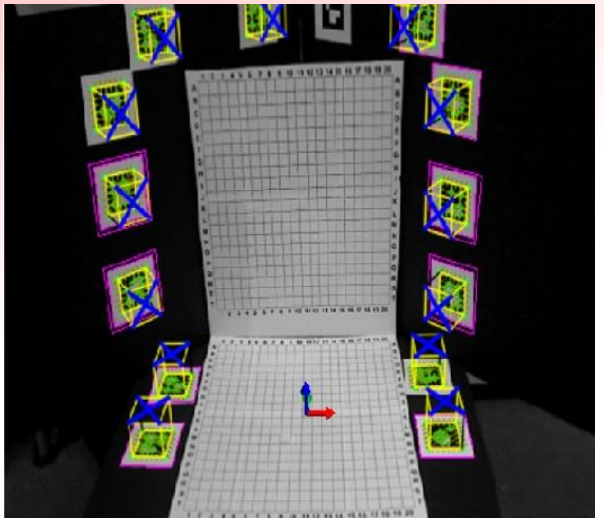
20 x 20 Grids of Squares



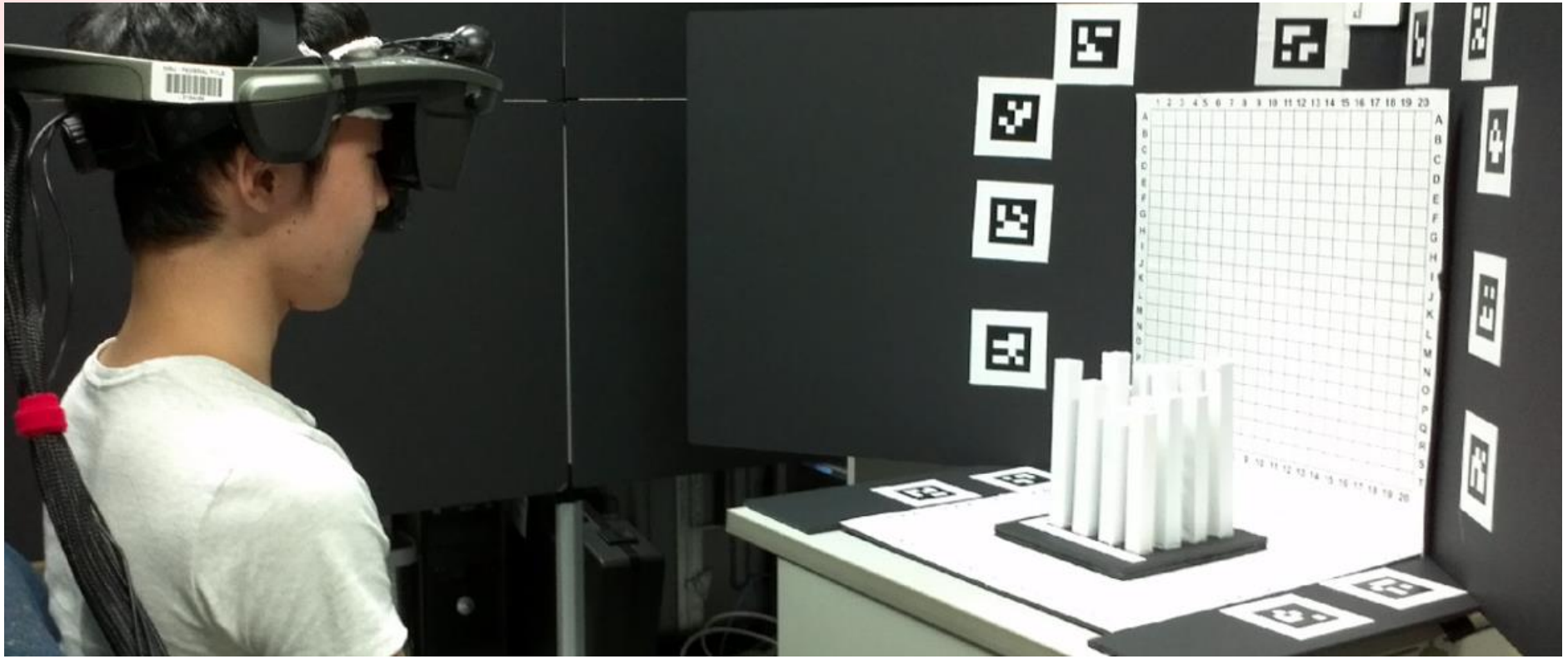



# System Hardware

Within-Subjects Design
13 Subjects (6 male / 7 female)
22 – 26 years of age
No prior experience with HMD's

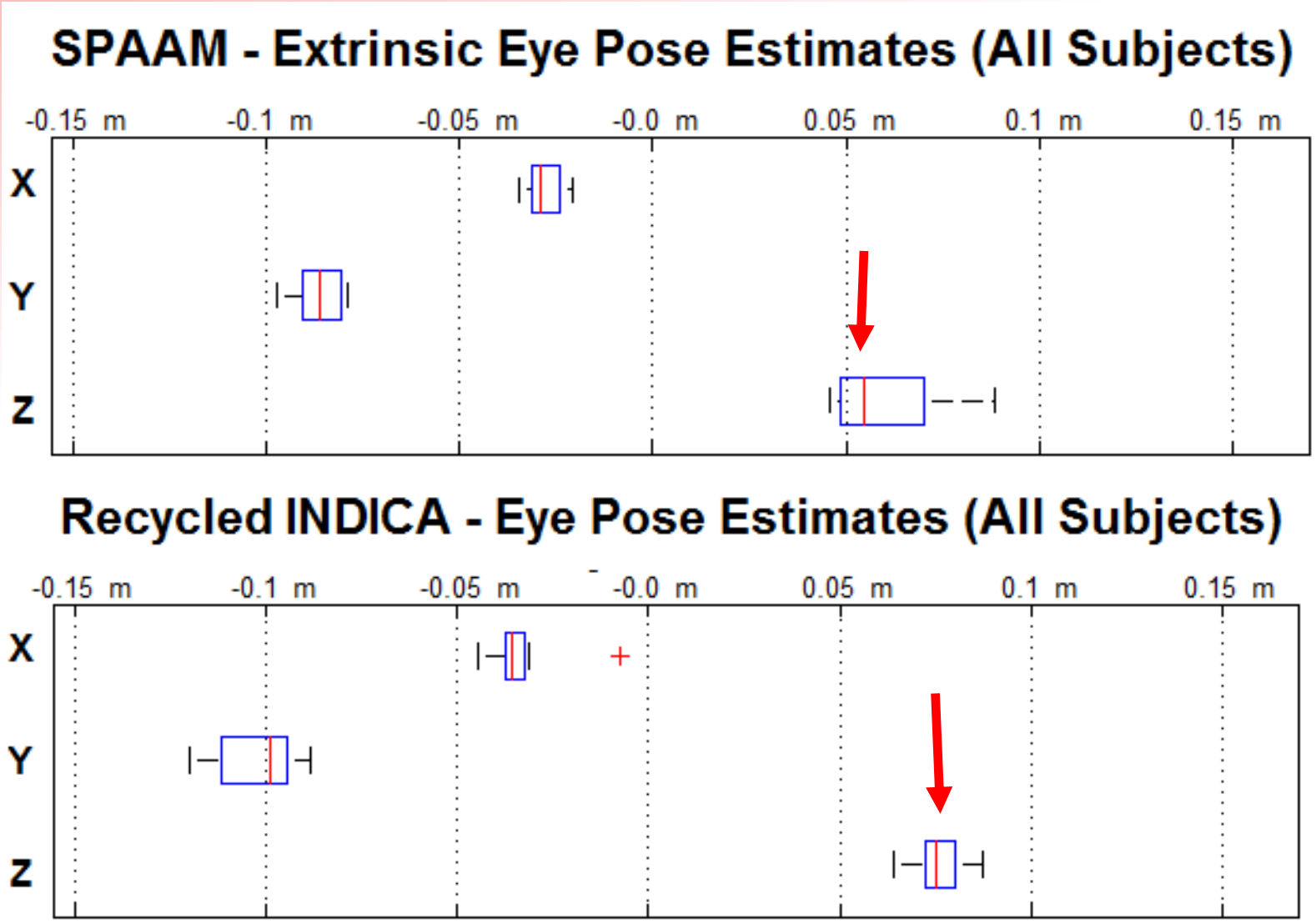


Tracking Performed by Ubitrack
Huber et al., 2007

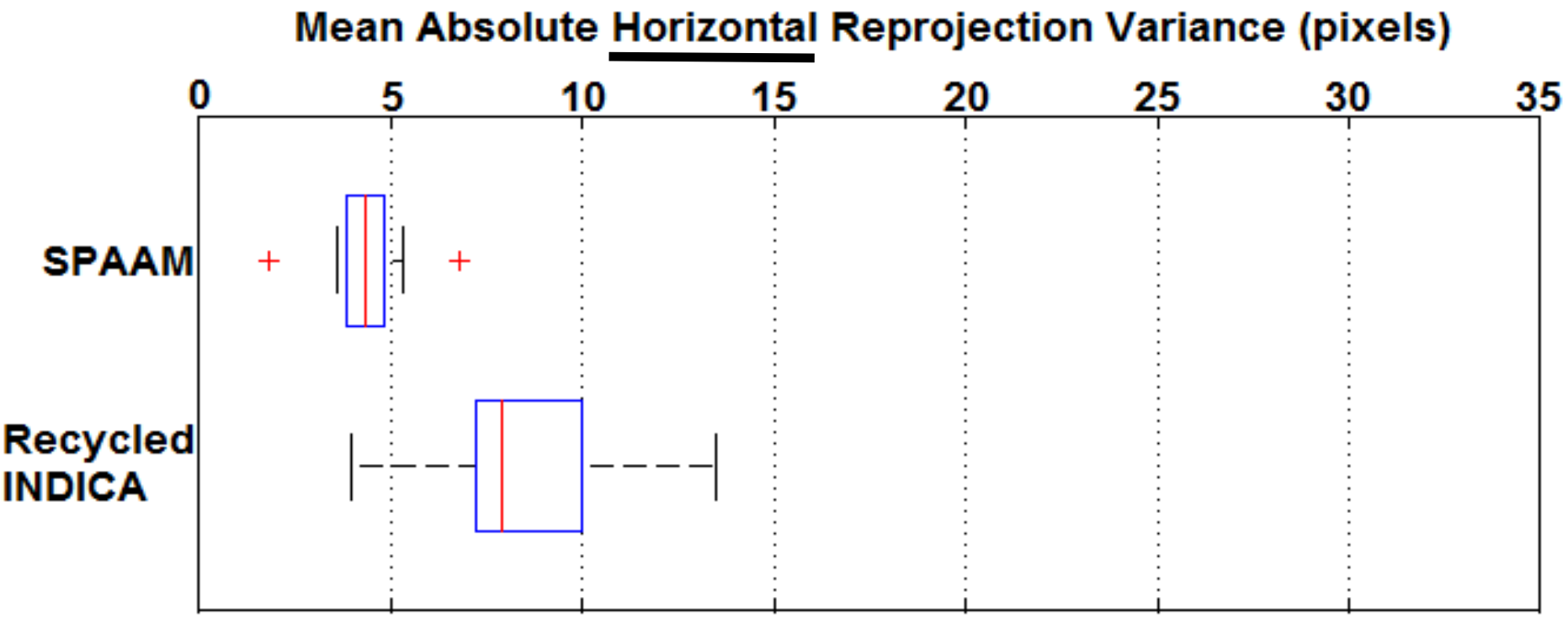


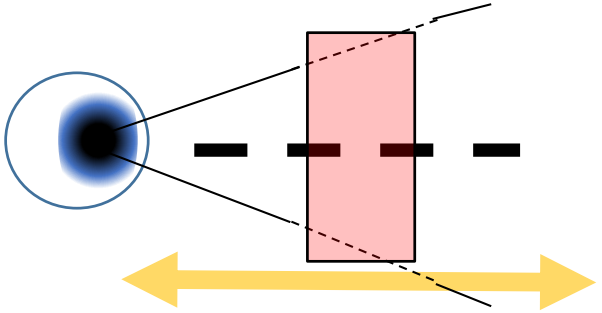
# Quantitative Result – Eye Location Estimation

SPAAM Extrinsic Parameters derived through QR Decomposition



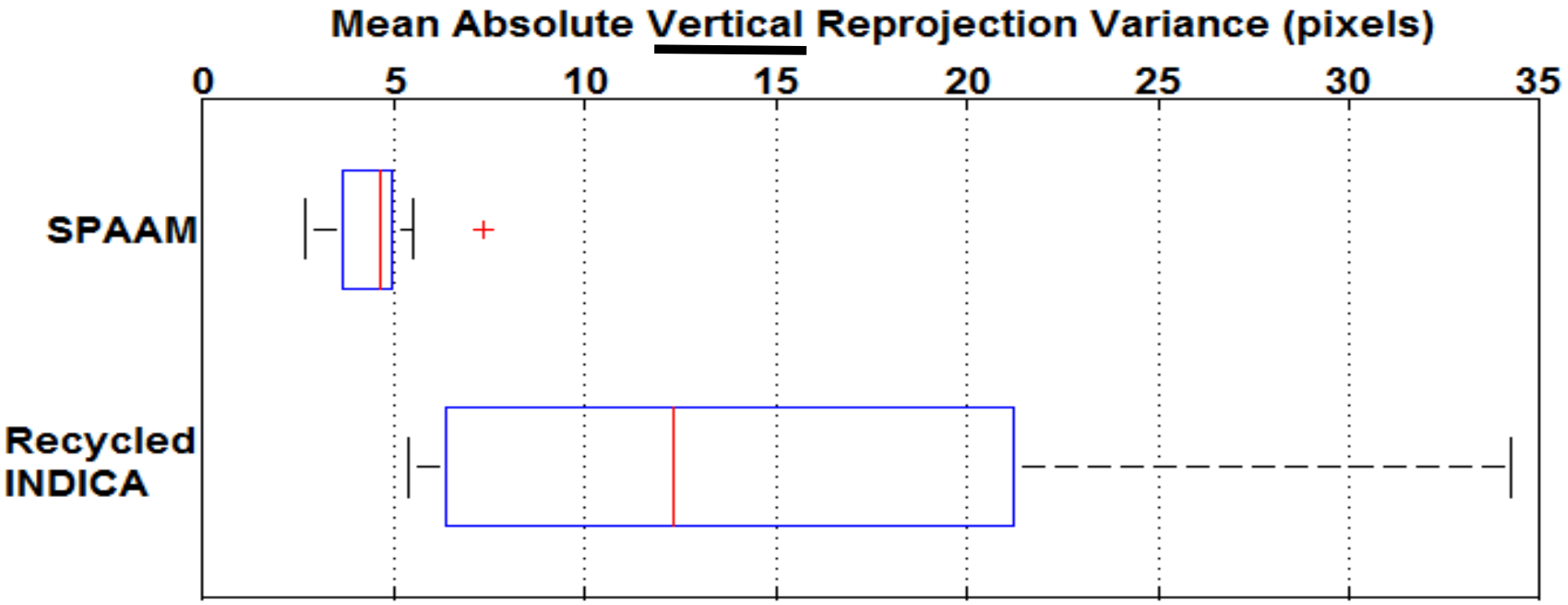
# Quantitative Result – Reprojection Variance

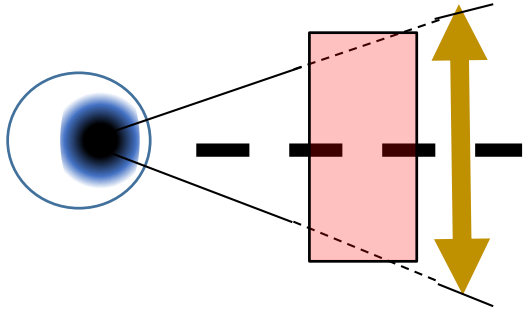


Horizontal Screen Space

# Quantitative Result – Reprojection Variance



Vertical Screen Space

# Subjective Result – Location Error

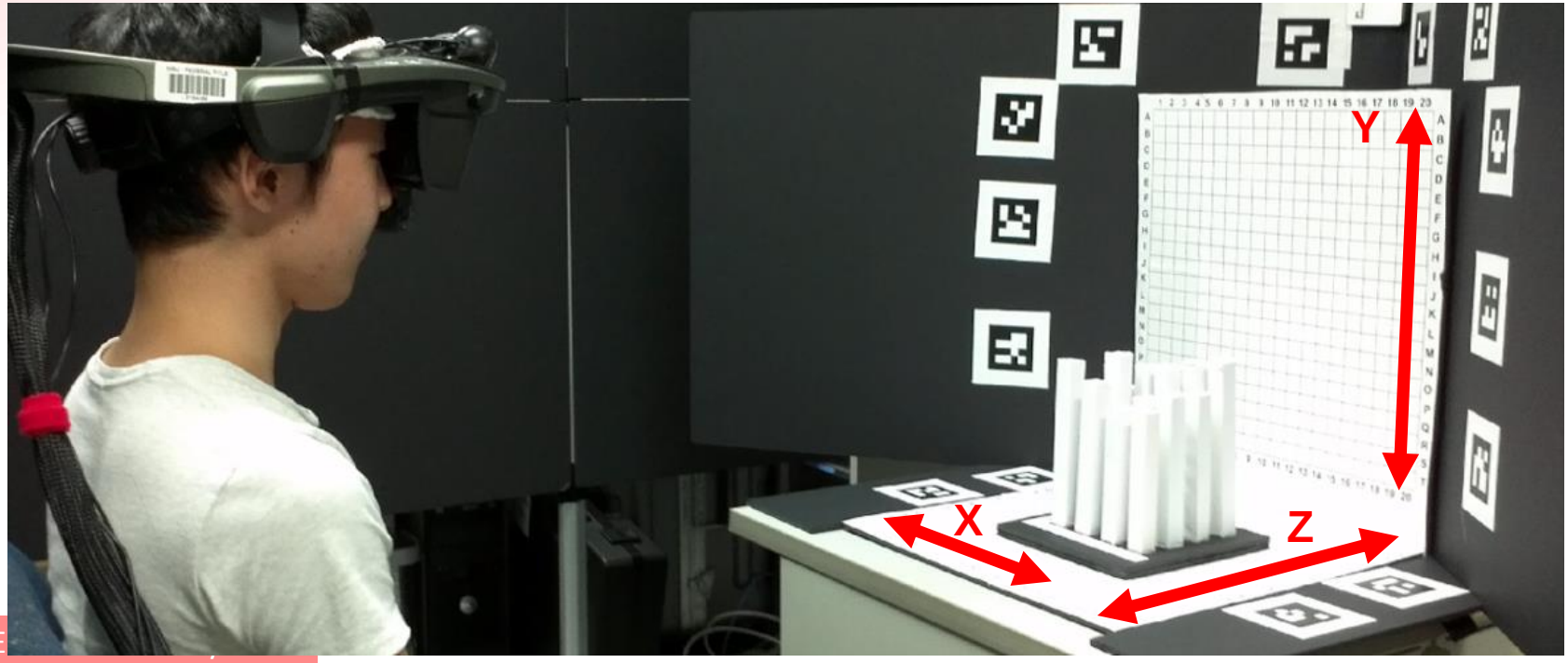
***Difference in cm between perceived and actual location***

**0:** No error (perfect registration)

**+Y:** Virtual perceived above
**- Y:** Virtual perceived below

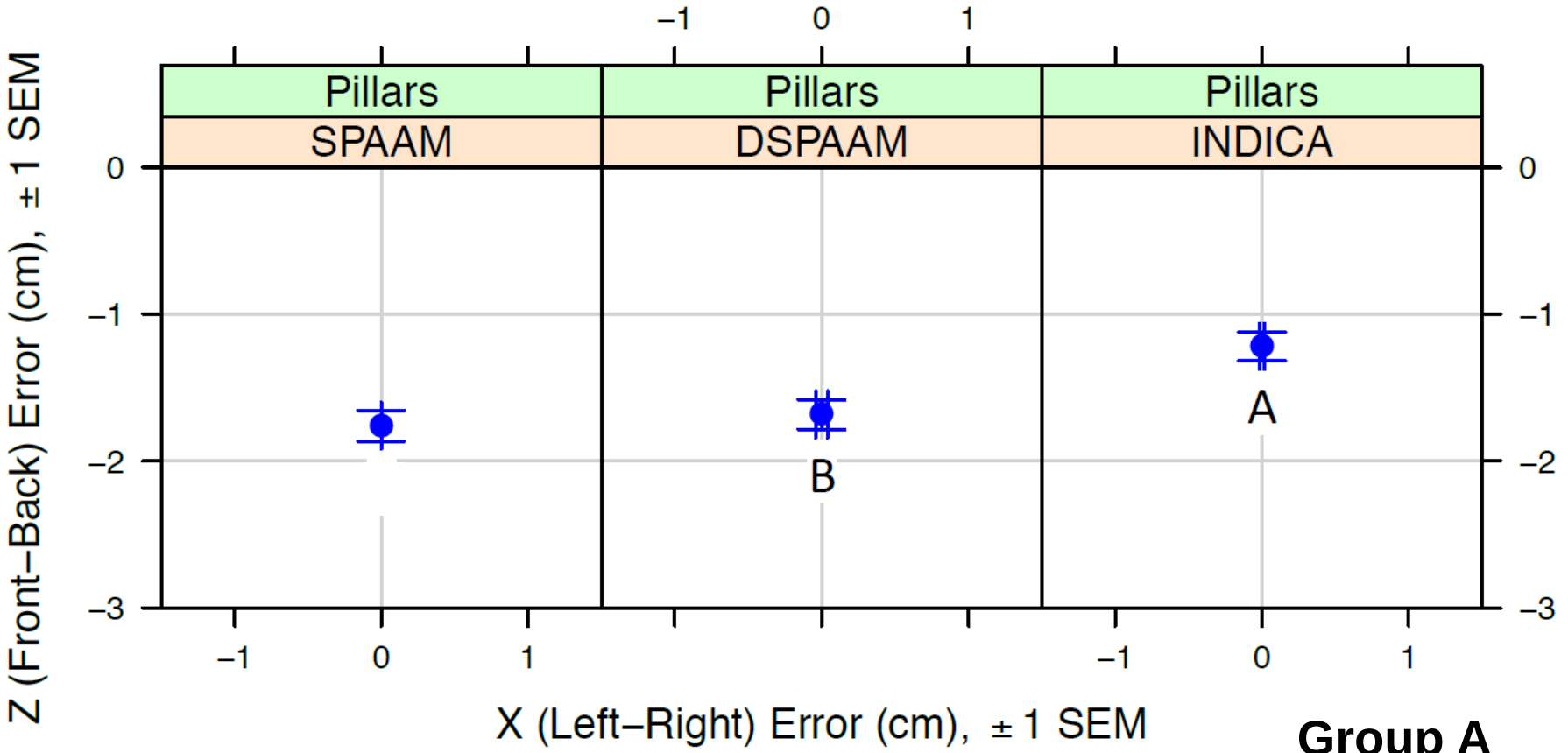
**+X:** Virtual perceived to the right
**- X:** Virtual perceived to the left

**+Z:** Virtual perceived further
**-Z:** Virtual perceived closer



# Subjective Result: Location Error – Pillars

## Difference Between Perceived & Actual Location




**Group B**

*No Significance**

**Group A**

*Significance in Z*

*F(2, 24) = 14.011*

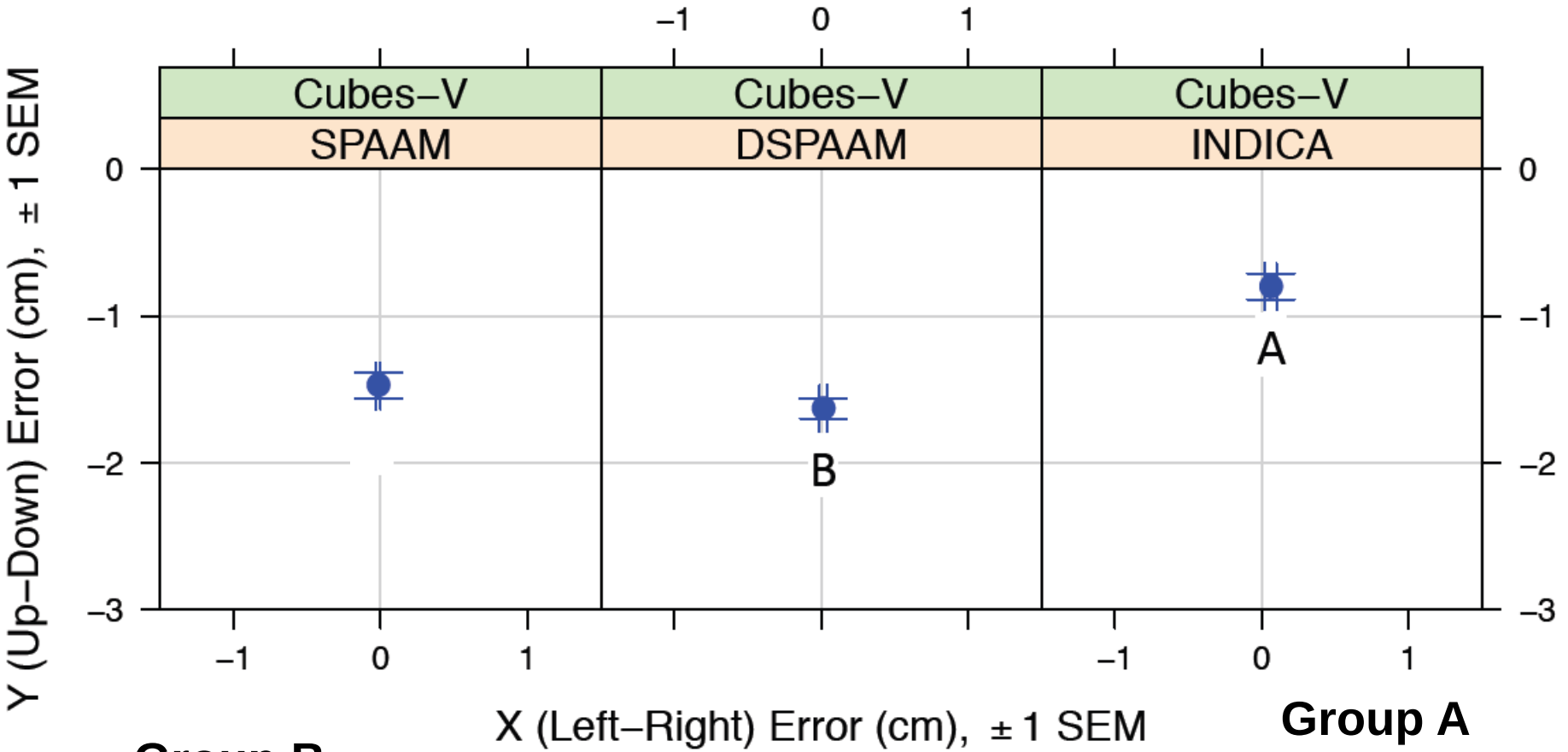
*p < .001*



* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

# Subjective Result: Location Error – Cubes Vertical Grid

## Difference Between Perceived & Actual Location




**Group B**

*No Significance**

**Group A**

*Significance in Y*

*F(2, 24) = 10.96*

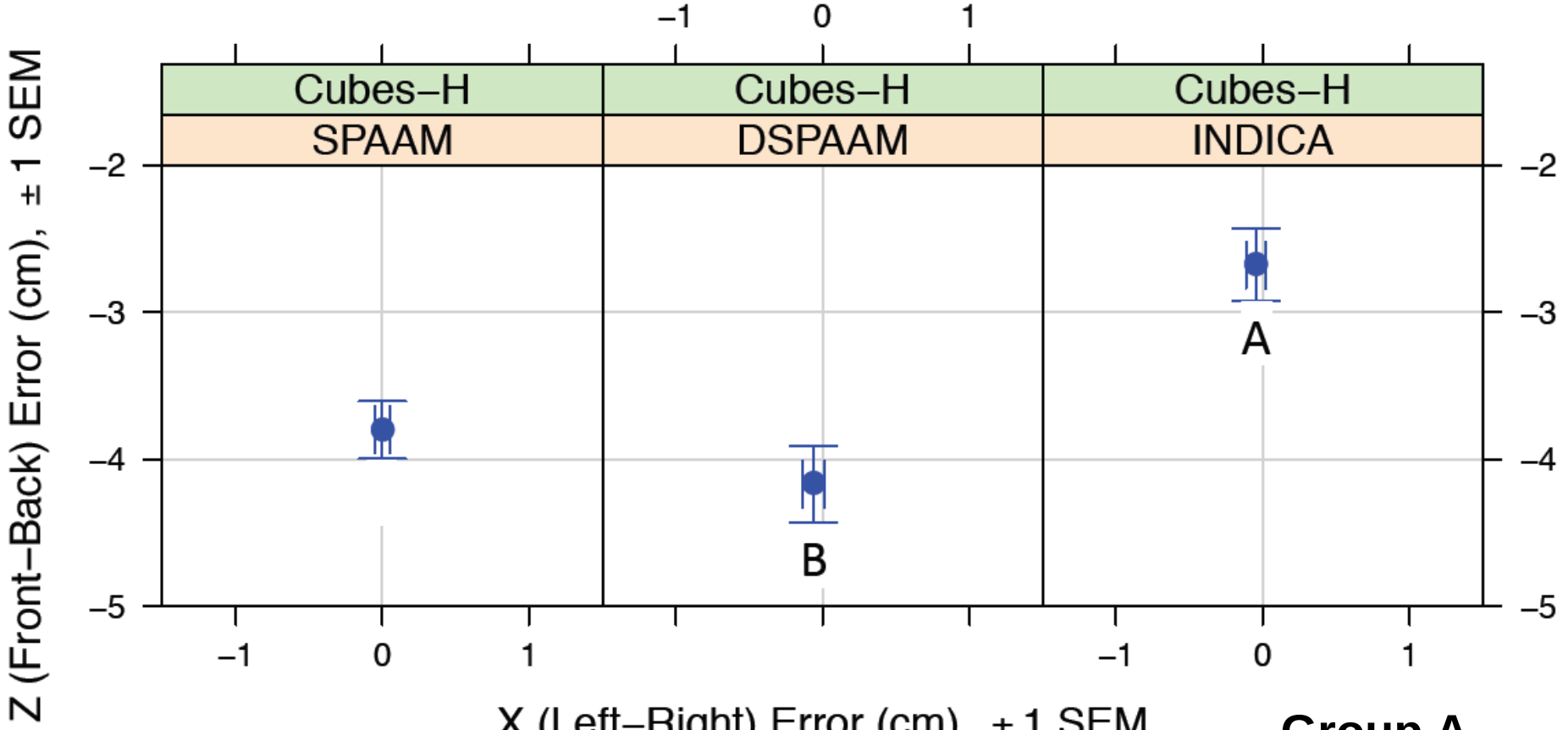
*p < .0016 ε = .75***

* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

**Mauchly's test indicated non-sphericity, p value adjusted by Huynh – Feldt $\varepsilon$

# Subjective Result: Location Error – Cubes Horizontal Grid

## Difference Between Perceived & Actual Location




**Group B**

*No Significance**

**Group A**

*Significance in Z*

*F(2, 24) = 7.37*

*p < .003*

* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

# Subjective Result: Location Error

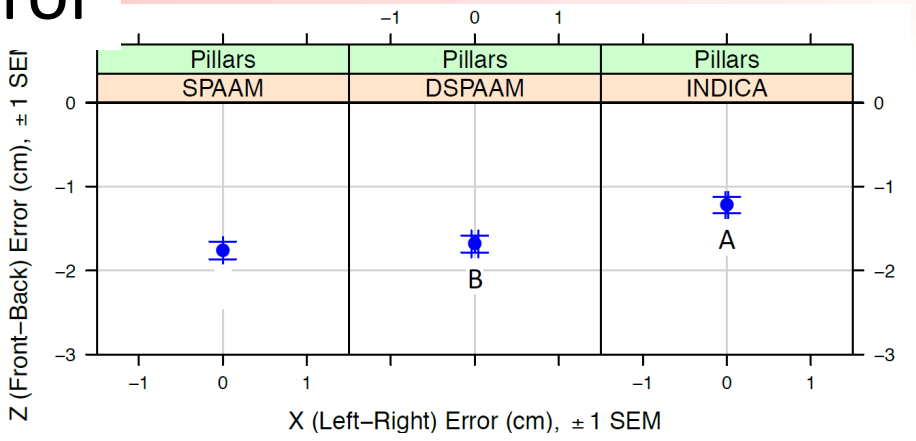
Quantitative Measures Not a Performance Prediction

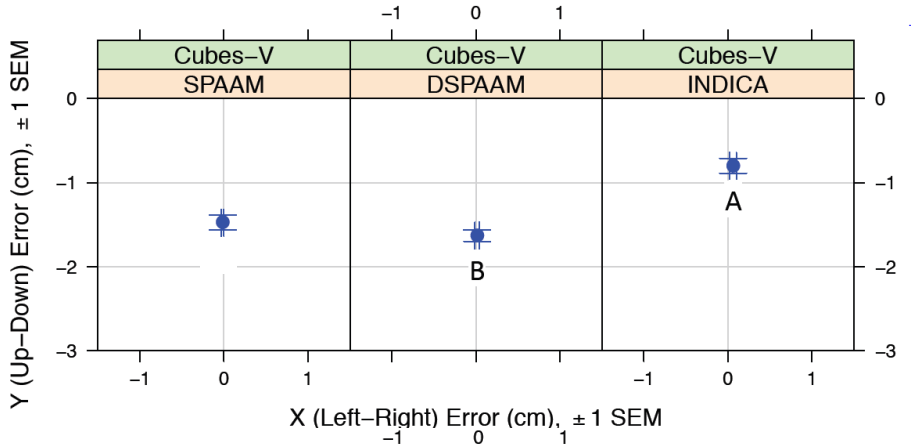
No Difference SPAAM/DSPAAM

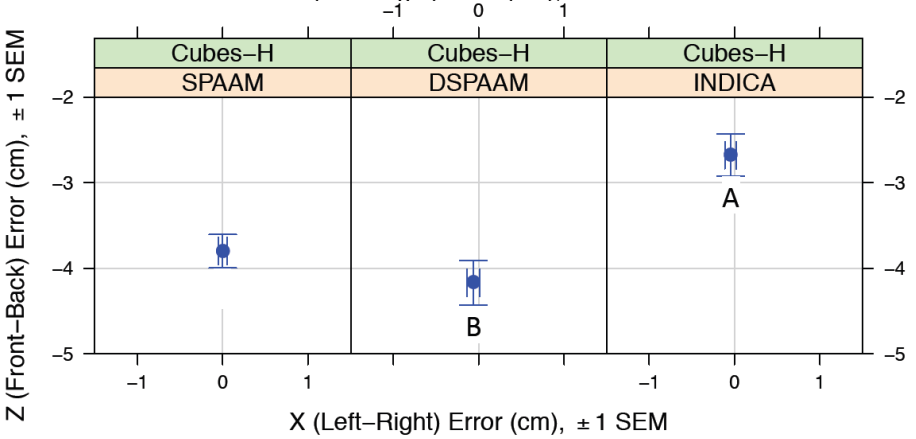
No Difference All Algorithms in X

Highest Overall Error in Z

INDICA Significantly Better Y/Z



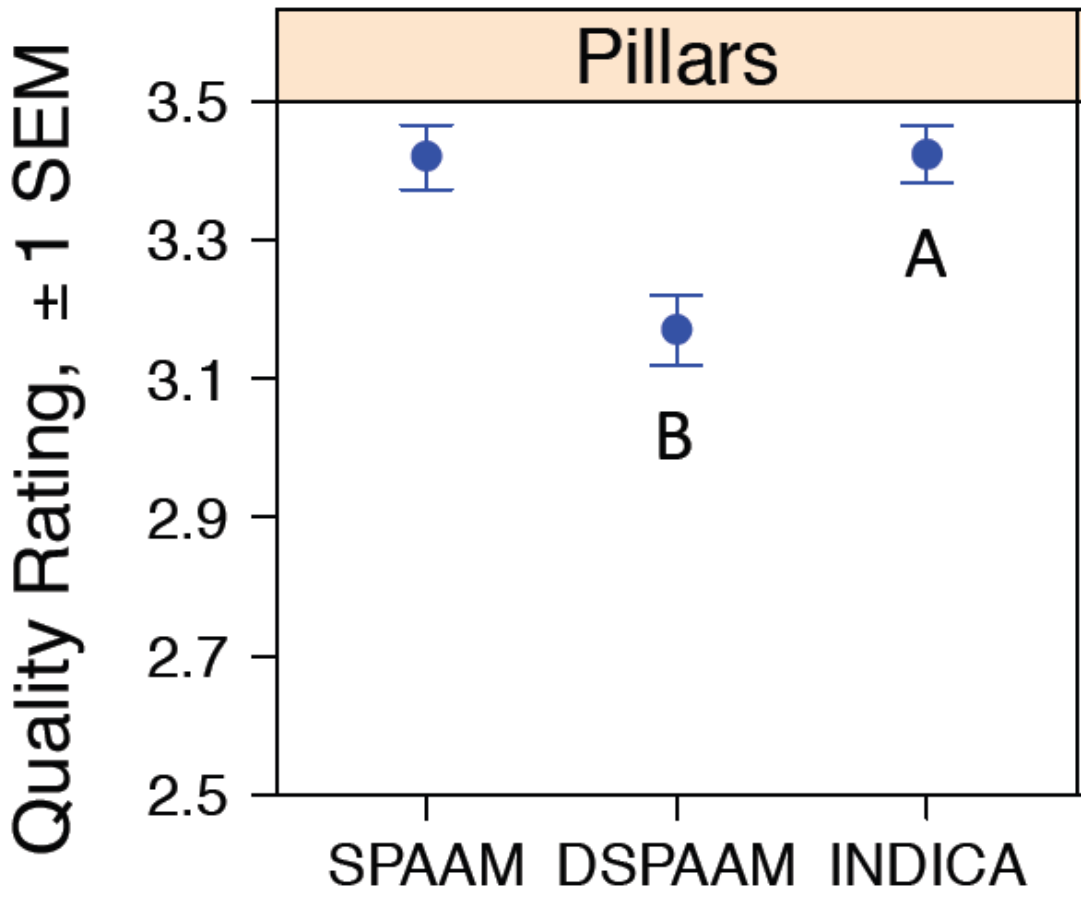



# Subjective Result: Registration Quality – Pillars

**Registration Quality with Chosen Location**

**Group A**
*No Significance**

**Group B**
*Significance in Quality*
*ANOVA: $F(2, 24) = 5.03$, $p < .015$*
*Friedman: $\chi^2(2) = 5.45$, $p < .066$*
*Kruskal-Wallis: $\chi^2(2) = 18.92$, $p < .001$*




* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

# Subjective Result: Registration Quality – Cubes Vertical

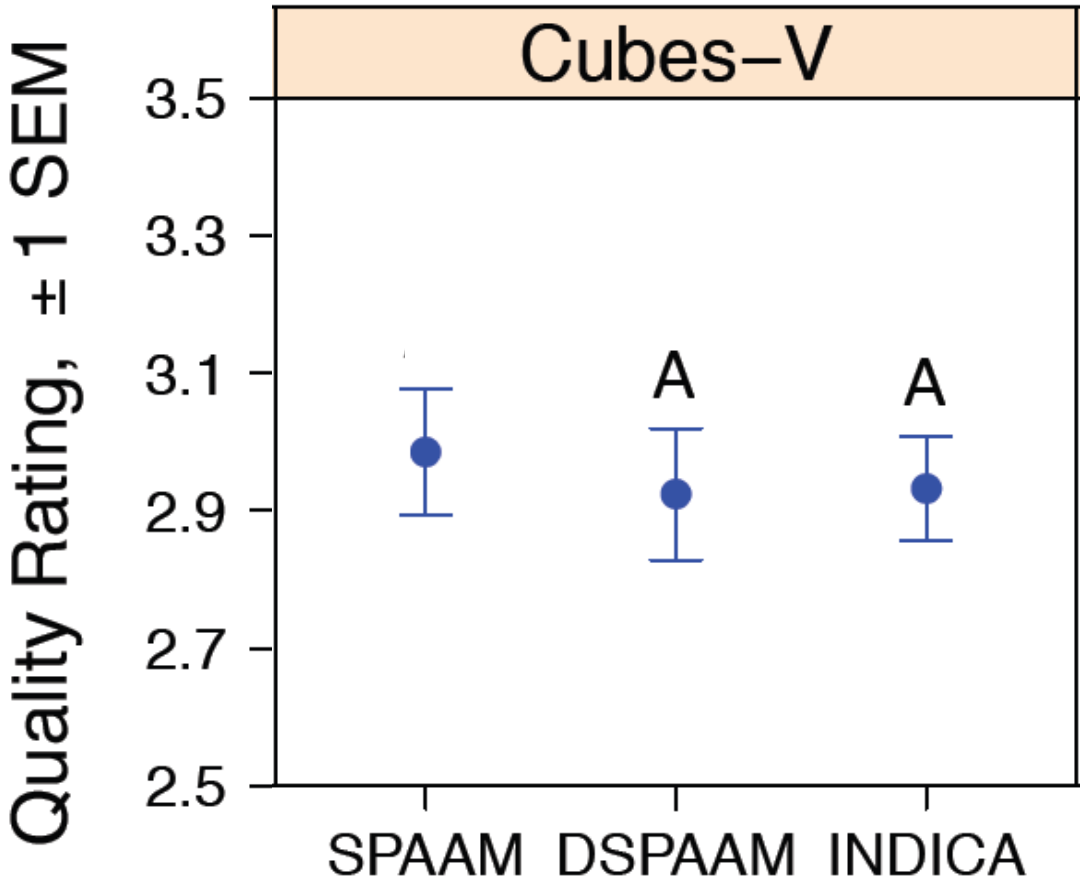
## Registration Quality with Chosen Location

**Group A**

*No Significance**

*Friedman: $\chi^2(2) = .15$*

*Kruskal-Wallis: $\chi^2(2) = .98$*




* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

# Subjective Result: Registration Quality – Cubes Horizontal

## Registration Quality with Chosen Location

**Group B**

*No Significance**

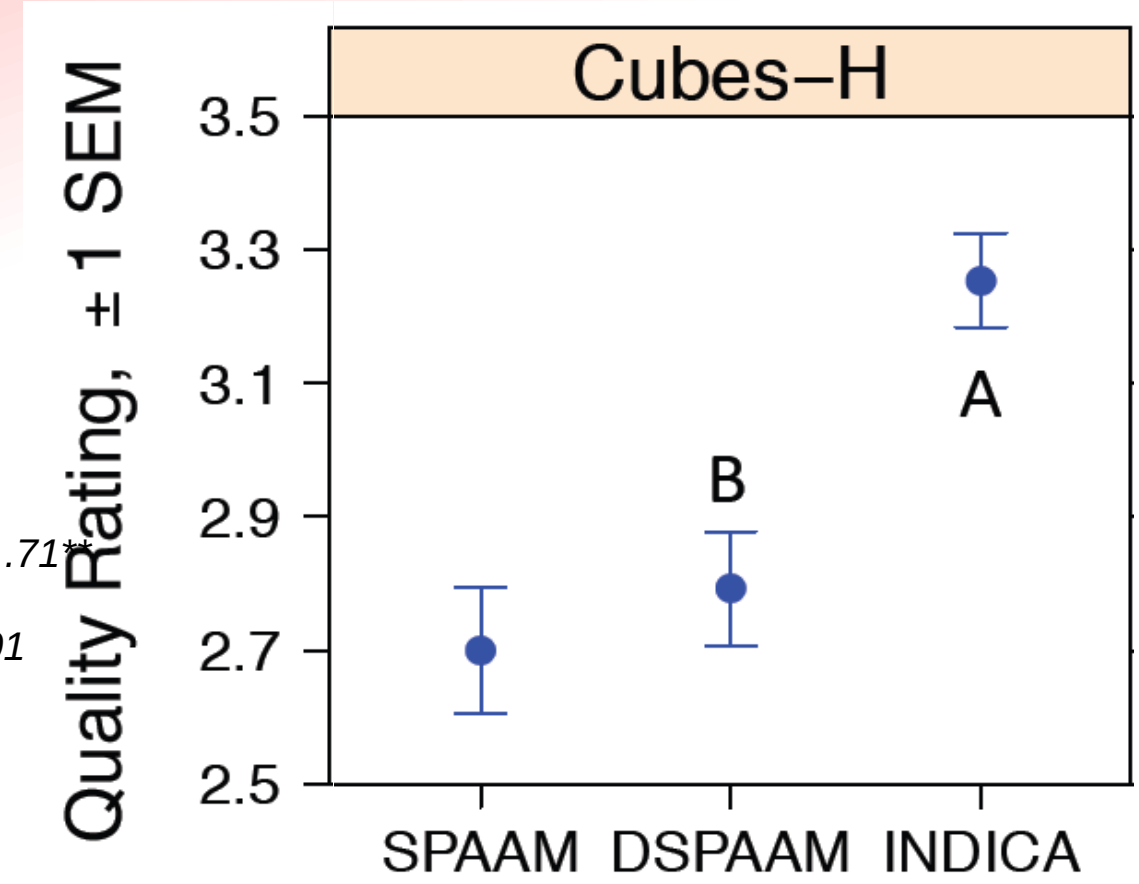
**Group A**

*Significance in Quality*

*ANOVA: $F(2, 24) = 6.65$, $p < .013$, $\varepsilon = .71$***

*Friedman: $\chi^2(2) = 13.06$, $p < 0.0015$*

*Kruskal-Wallis: $\chi^2(2) = 21.21$, $p < 0.001$*

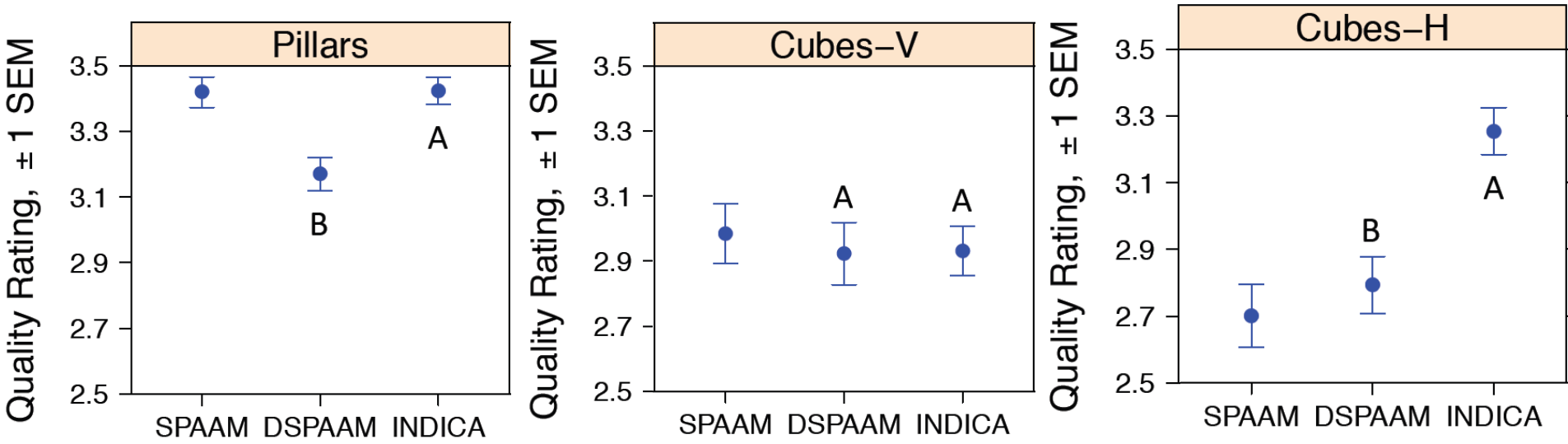



* $p \leq 0.05$ Ryan REGWQ post-hoc homogeneous subset test

**Mauchly's test indicated non-sphericity, p value adjusted by Huynh – Feldt $\varepsilon$

# Subjective Result: Registration Quality

## Registration Quality with Chosen Location




Quality Values Match Performance Measures

INDICA Quality is Equal or Better Than SPAAM

INDICA Quality Significantly Better in Z

# Results Discussion

***Why Apparent Disagreement in Quantitative & Subjective?***

~~Poor Eye Localization for INDICA?~~

Eye Location Values for INDICA Show Low Variance

Removal/Replacement of HMD Between Conditions

Reprojection Shows Closer to Actual Pixels Used in Alignment

INDICA Presumes a Simplistic HMD Model

***Why No Significant Performance Change in SPAAM/DSPAAM?***

HMD Specific Properties

- Fit on User's Head
- Exit Pupil Location

Resolution of Task Not High Enough to Find Significance

# Take Away

**SPAAM / Degraded SPAAM** – almost no difference

- Removal/Replacement little effect on accuracy
- Accuracy in X equal to INDICA
- Less favorable method (exit survey)

**INDICA** – Equal or Superior performance to SPAAM

- Significantly higher performance in Y/Z
- Significantly higher quality in Z
- **Recycled INDICA requires SPAAM intrinsics**
- Minimal requirement from user
- Less time to perform (user preferred)

# Future Work

***Evaluation of Full INDICA***

Remove induced error from SPAAM intrinsics (Full INDICA)
Utilize more robust eye localization (Alex's presentation)
Real time update of calibration (on-line)

***Binocular Task***

More relevant depth cue
Verification of SPAAM Z error

***Best VS Best***

Comparison of best possible calibrations SPAAM/INDICA
Removal of HMD distortion (Yuta's presentation)
Improvements to SPAAM to reduce impact of user error

# Special Thanks

Dr. Hirokazu Kato

Everyone at the Interactive Media Design Lab

All of the Anonymous Subjects

Dr. Goshiro Yamamoto

# Support


NSF Awards: IIA-1414772, IIS-1320909, IIS-1018413



NASA Mississippi Space Grant Consortium Fellowship


European Union Seventh Framework
PITN-GA-2012-316919-EDUSAFE

# Take Away

**SPAAM / Degraded SPAAM** – almost no difference

- Removal/Replacement little effect on accuracy
- Accuracy in X equal to INDICA
- Less favorable method (exit survey)

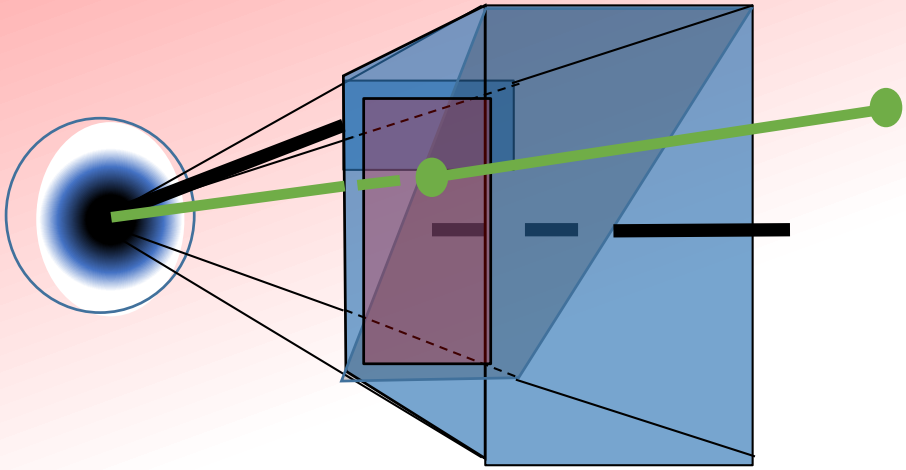
**INDICA** – Equal or Superior performance to SPAAM

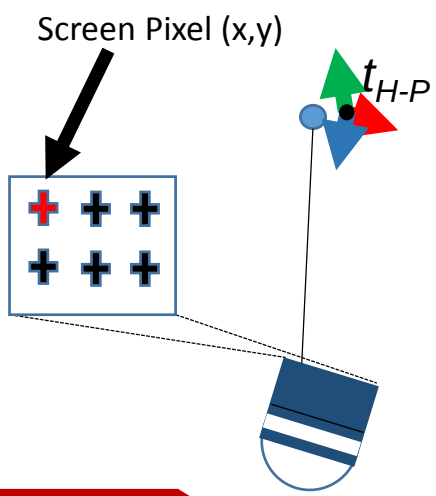
- Significantly higher performance in Y/Z
- Significantly higher quality in Z
- **Recycled INDICA requires SPAAM intrinsics**
- Minimal requirement from user
- Less time to perform (eye measures vs alignment)

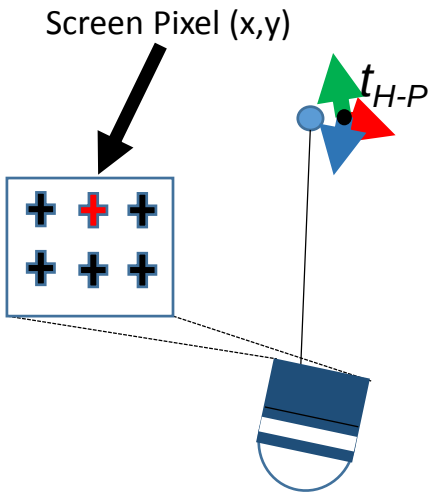
# Calibration Methods

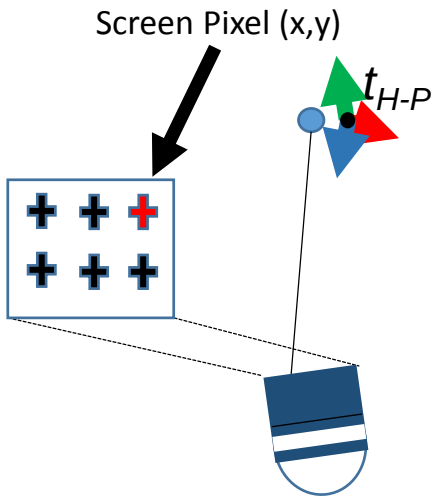
## Single Point Active Alignment Method (SPAAM)

(Tuceryan & Navab, 2000)





# Calibration Methods

## Interaction Free Display Calibration (INDICA)

(Itoh & klinker, 2014)

**Recycled Setup**

| | |
|---|---|
| **Intrinsic Calib. Params.** | **$K_E$** |
| **Translation World – Eye** | **$t_{WE}$** |
| **Translation World – Screen** | **$t_{WS}$** |

Full Setup

| | |
|---|---|
| Translation World – Screen | $t_{WS}$ |
| Pixel Scaling Factor | $\alpha_{(x,y)}$ |

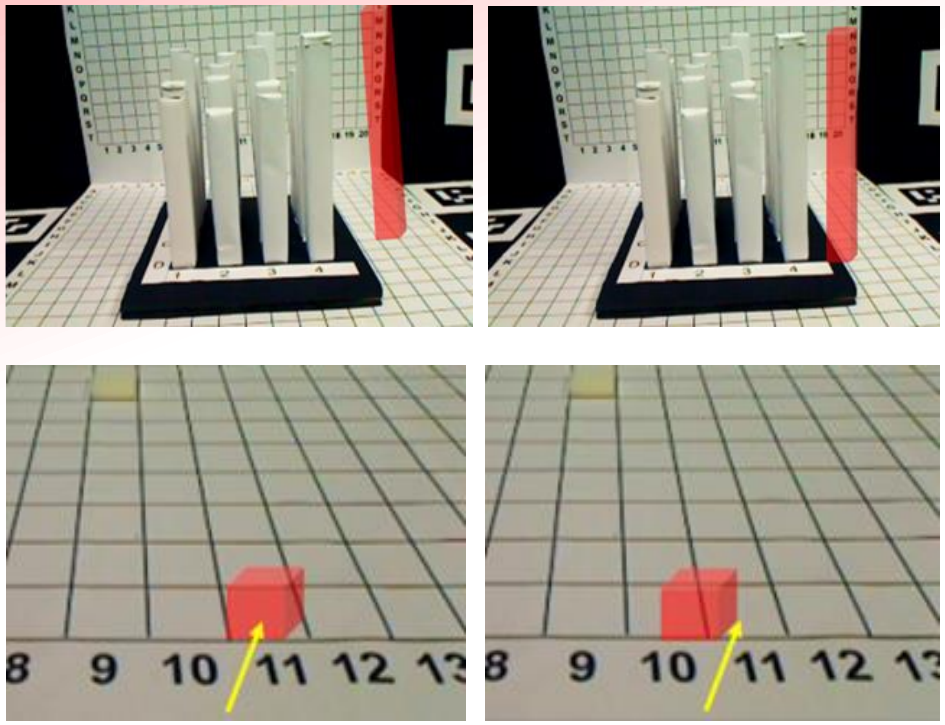
Required for Both

| | |
|---|---|
| Translation Eye -Tracker | $t_{ET}$ |
| Rotation World – Screen | $R_{WS}$ |
| Rotation World – Tracker | $R_{WT}$ |
| Translation World – Tracker | $t_{WT}$ |

# Evaluation Study

## *Registration Quality Comparison*

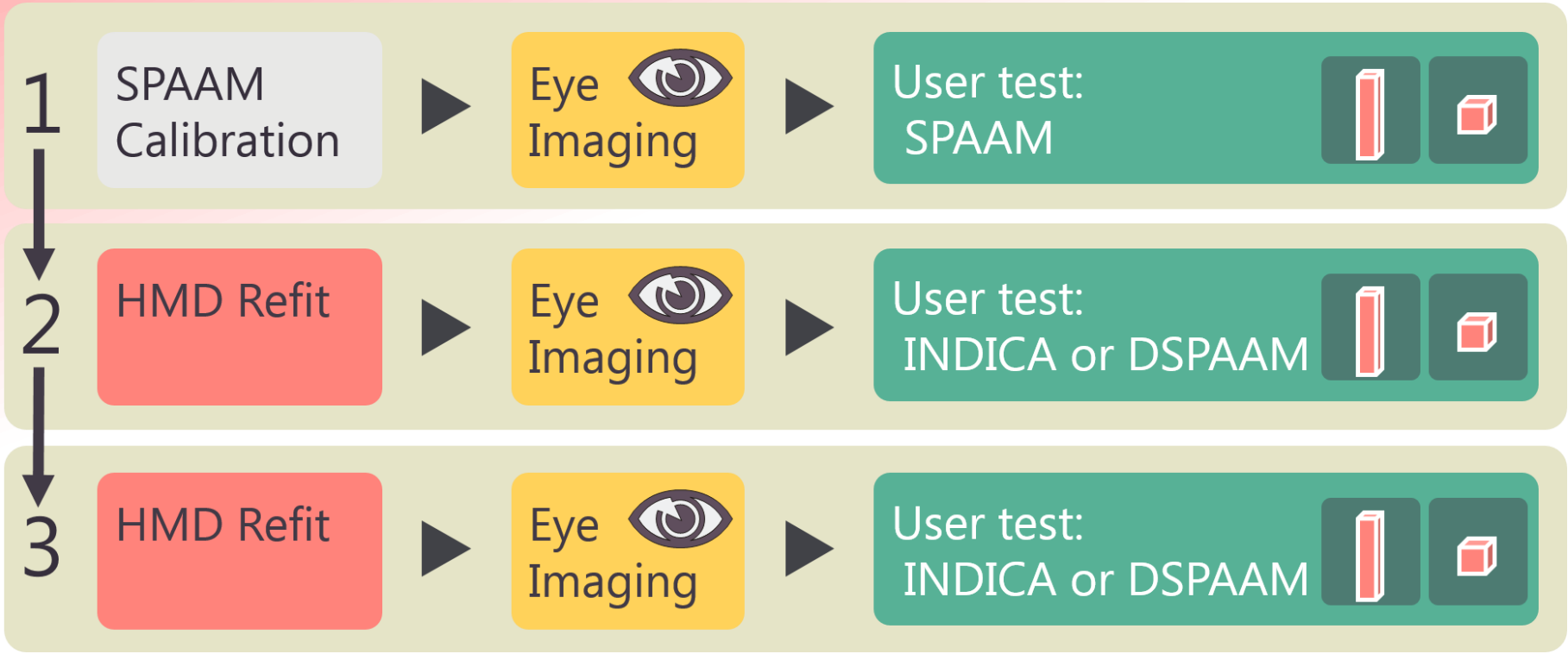
### *Perceived VS Intended Location*



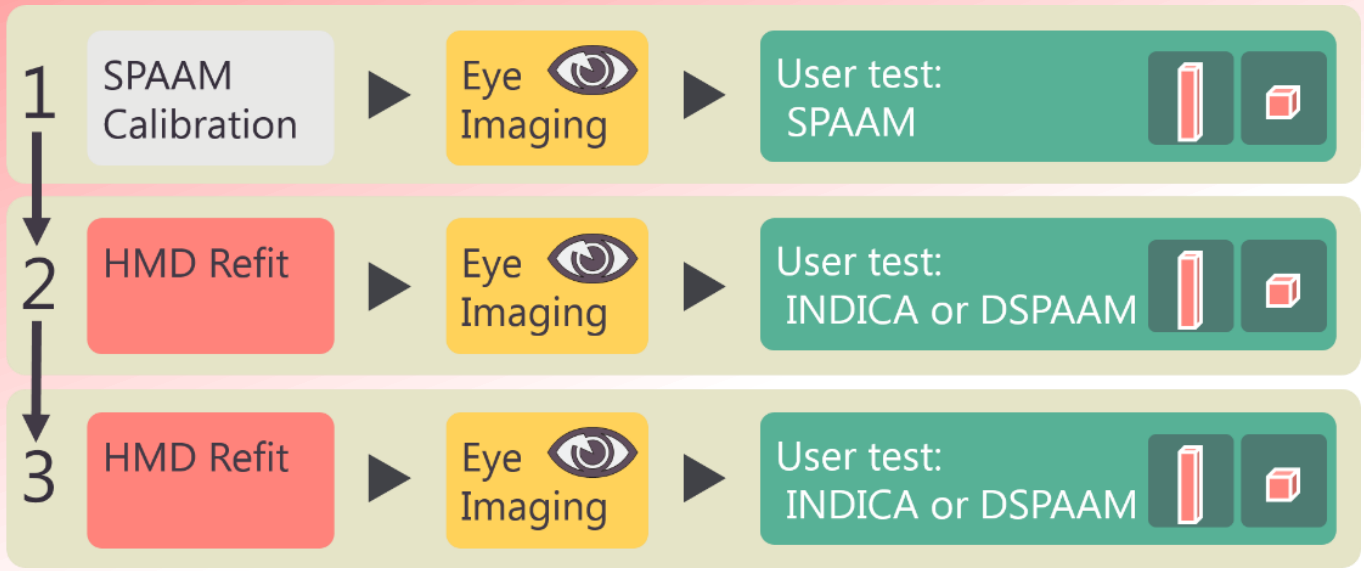

- SPAAM
- Degraded SPAAM:
  Reuse of SPAAM result after HMD replacement
- Recycled INDICA:
  Reuse intrinsic values from SPAAM calibration

# Evaluation Study

*Degraded SPAAM & Recycled INDICA rely on values from SPAAM calibration*



# Evaluation Study



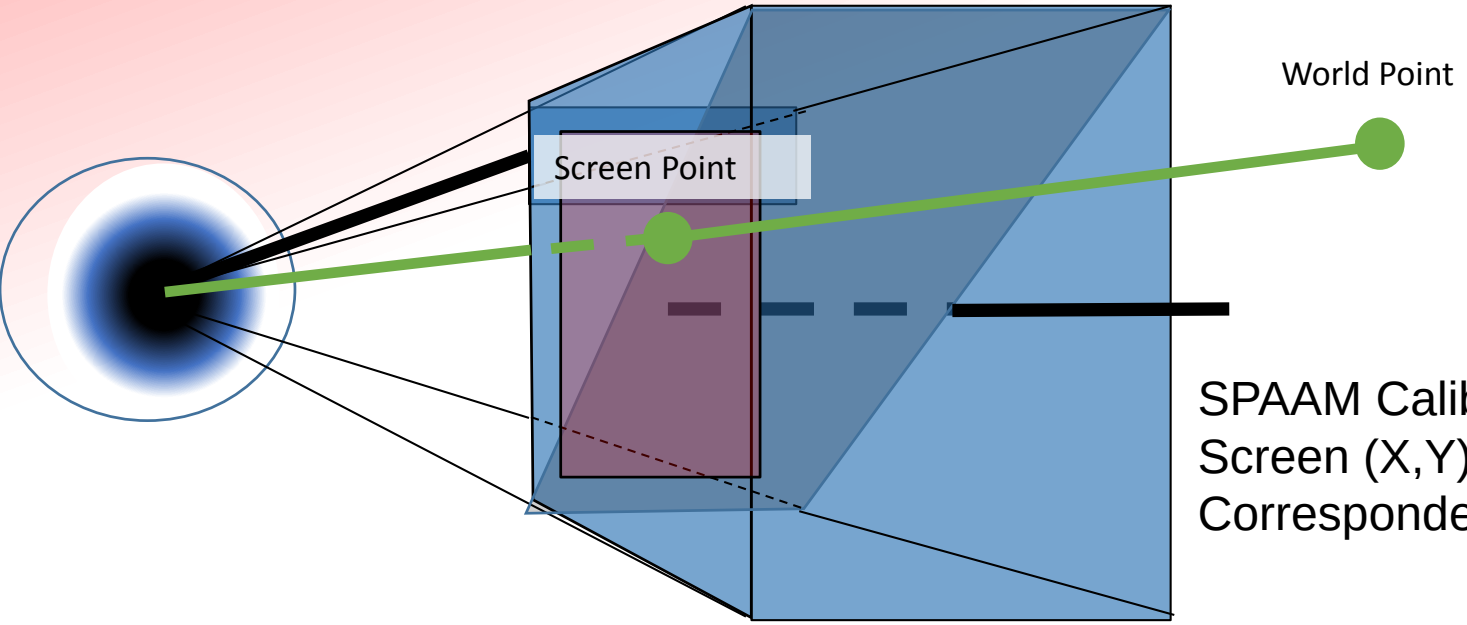

Within-Subjects Design
3 Alg. X 2 Tasks = 6 Conditions
16 Pillar Trials/20 Cube Trials per cond.

13 Subjects (6 male / 7 female)
22 – 26 years of age
No prior experience with HMD's

Order of D. SPAAM and R. INDICA, as well as Cube/Pillar task presentation, distributed such that no two subjects experienced the same sequence

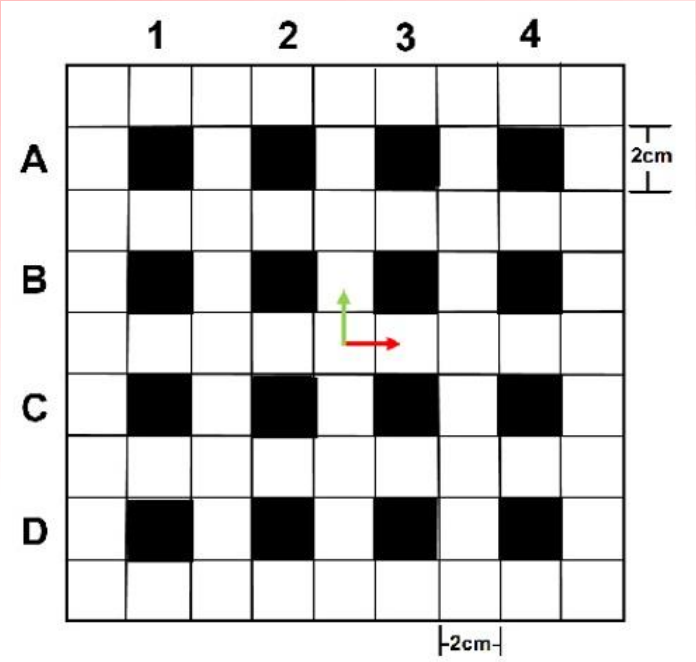
# Experimental Results & Discussion

## *Quantitative Measures – **Reprojection Estimates***



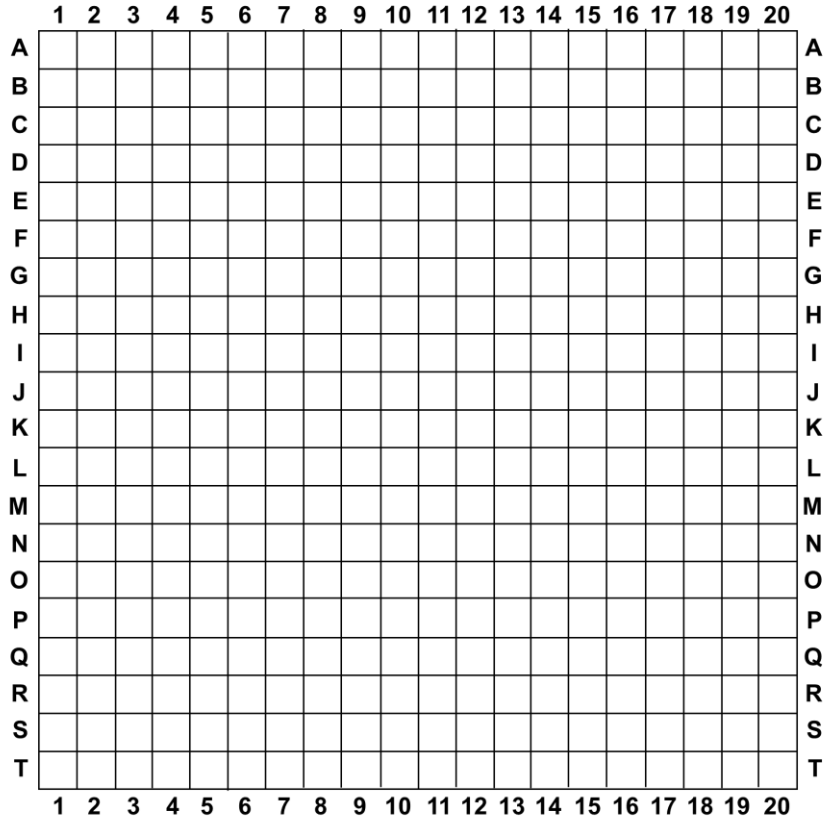

SPAAM Calibration Produces Screen (X,Y) and World (X, Y, Z) Correspondence Pairs

# Evaluation Tasks




Vertical Axis: A – Z (A-D)
Horizontal Axis: 1 – 20 (1-4

Discrete 2Axis Grids
Each Square 2cm x 2cm



# Experimental Results & Discussion

***Subjective Measures –Registration Quality***

***Verbal Response from Subject (1-4/5)***

Normalized both scales (Pillars/Cubes) into 1-4

Quality values are **not** Likert scale data – provided images create reference for quality range

Statistical Analysis on Quality Data:

- ANOVA
- Friedman
  - Less power compared to ANOVA
  - Reduces number of considered data points
- Kruskal-Wallis
  - More power compared to ANOVA
  - Does not consider within subject design

# Experimental Results & Discussion

***Performance Summary – Quantitative Measures***

SPAAM – Extrinsic/Reprojection Values Match Previous Findings

- Extrinsics show Higher variance along Z axis
- Low reprojection variance

*SPAAM result closely reproduces SPAAM alignments*

INDICA – Stable Eye Estimates / High Reprojection Variance.

- Manual Limbus Detection Required for best estimates
- Eye position in Z more consistent
- Higher reprojection variance

Reprojection not indication of result quality

INDICA reprojection shows actual pixel used (?)