UCF-MANUS Arm

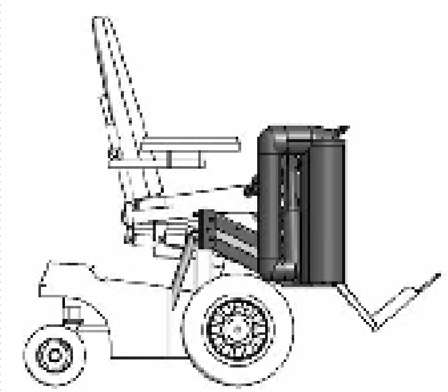
Stereo Vision-based Control of Assistive Robotic Manipulator in Unstructured Environments: Design, Implementation, and User Testing for Patients post SCI

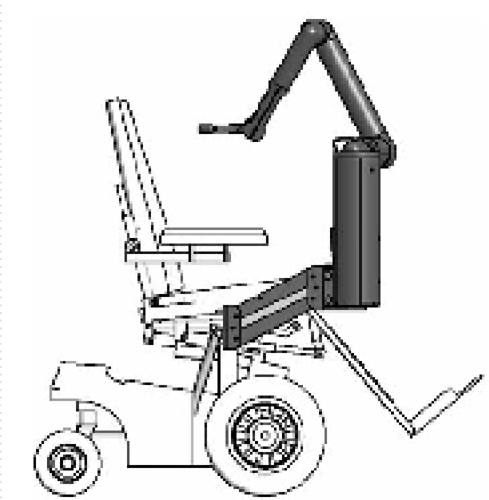
PI: Aman Behal, PhD
abehal@mail.ucf.edu

Assistive Robotics Lab.
University of Central Florida, Orlando, FL

Contents

- Introduction
- Problem Statement
- Approach
  - Gross-to-fine motion segmentation
  - Multi-sensory smart grasping
- Results
  - Setup – H/W and S/W
  - Demonstration Video
  - User Testing
- Concluding Remarks



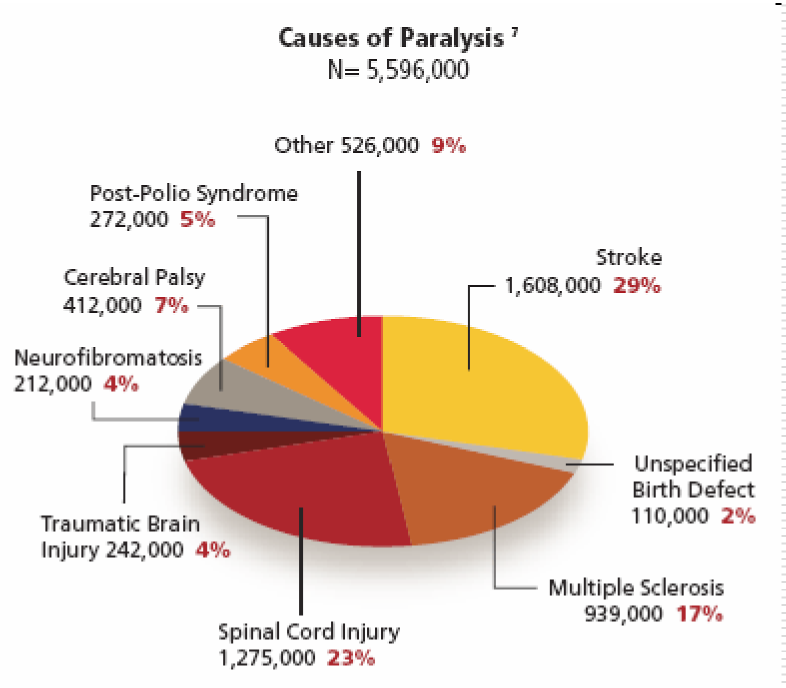


Introduction
- Statistics

- Approximately 0.4% of the U.S. population or some 1.3 million people reported being paralyzed due to a spinal cord injury (SCI). Every year, there is approximately an additional 11,000 spinal cord injuries.

- Many of these individuals are confined to power wheelchairs, have moderate to minimal function in their upper extremities, and require some amount of attendant care.
  - Average yearly expenses can range from $230k to $780k in the first year.
  - Estimated lifetime costs can range from $680k to over $3 million for a 25 year old.

- Two-thirds of mobility device users have limitations in one or more Instrumental Activities of Daily Living.
  - grocery shopping, using the telephone, meal preparation, light housework, etc.



[REF] Christopher and Dana Reeve Foundation, http://www.christopherreeve.org, 2009.

Introduction
- Robotic Caregiver

- Service Robot
  - Functions of service robots are generally related to the ordinary human life like repair, transfer, cleaning, health care, and so forth.

Human Caregiver

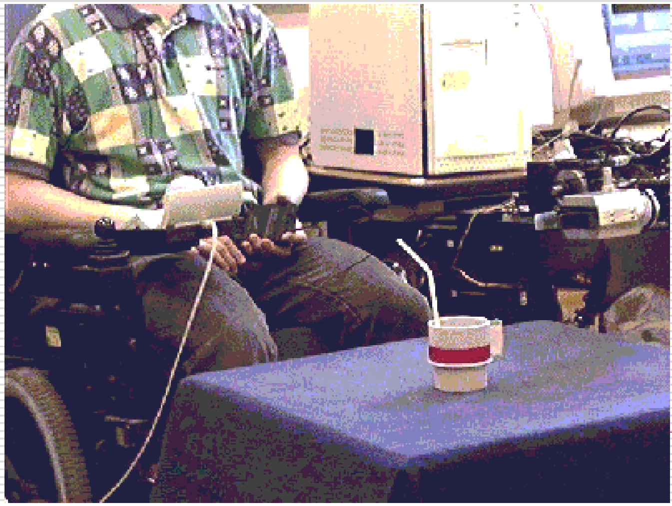
Robotic Caregiver: KARES II

Introduction
- Example of Wheelchair-based Robot

RWTH-ARM

Weston Robot

Raptor

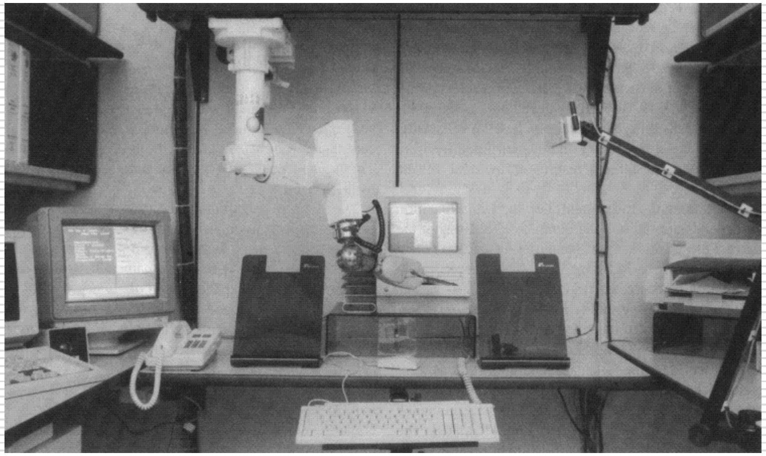


KARES I

FIRE

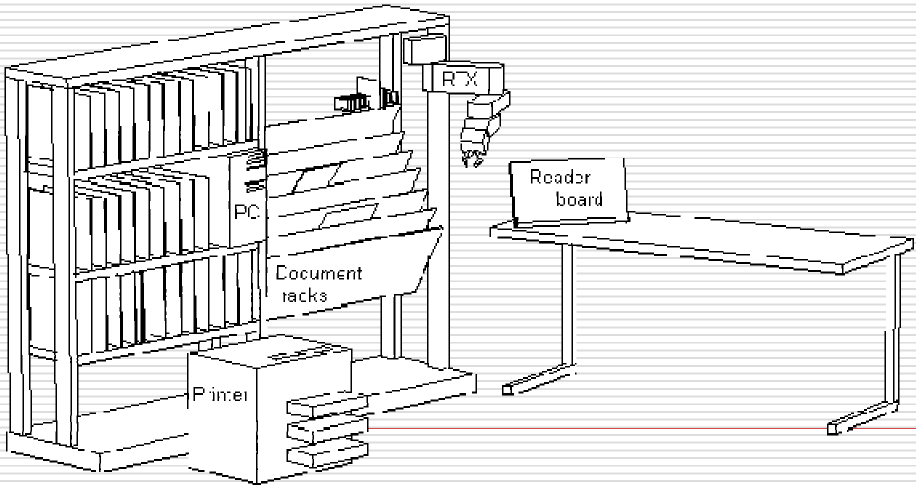
FRIEND

Introduction
- Example of Workstation-type Robot



Desktop Vocational Robot

ISAC



RAID workstation

AFMASTER

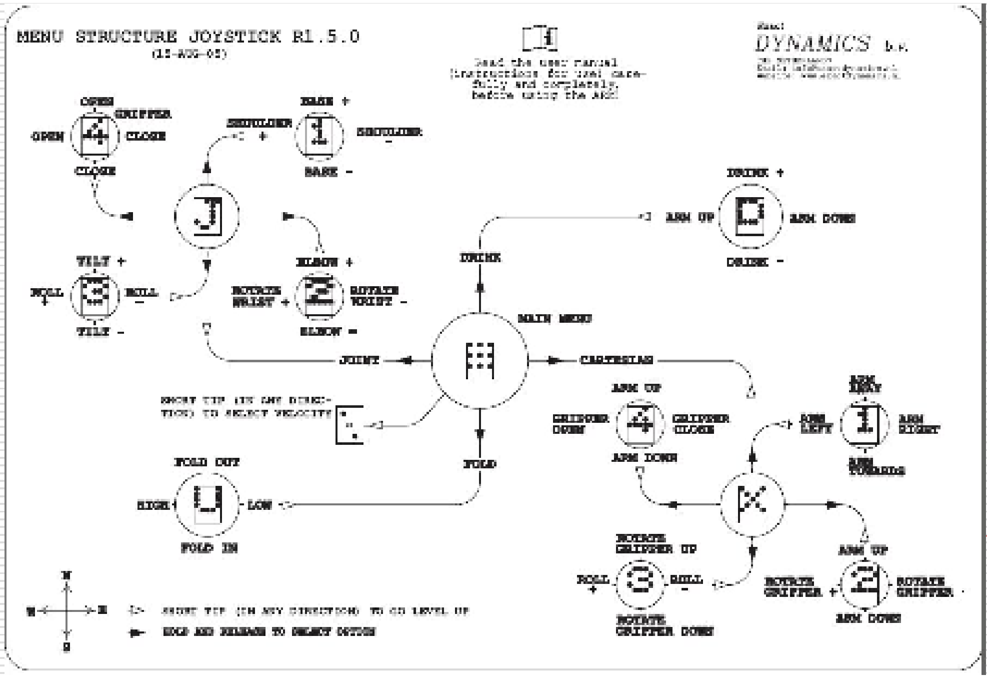
Introduction
- MANUS Assistive Robotic Manipulator (ARM)

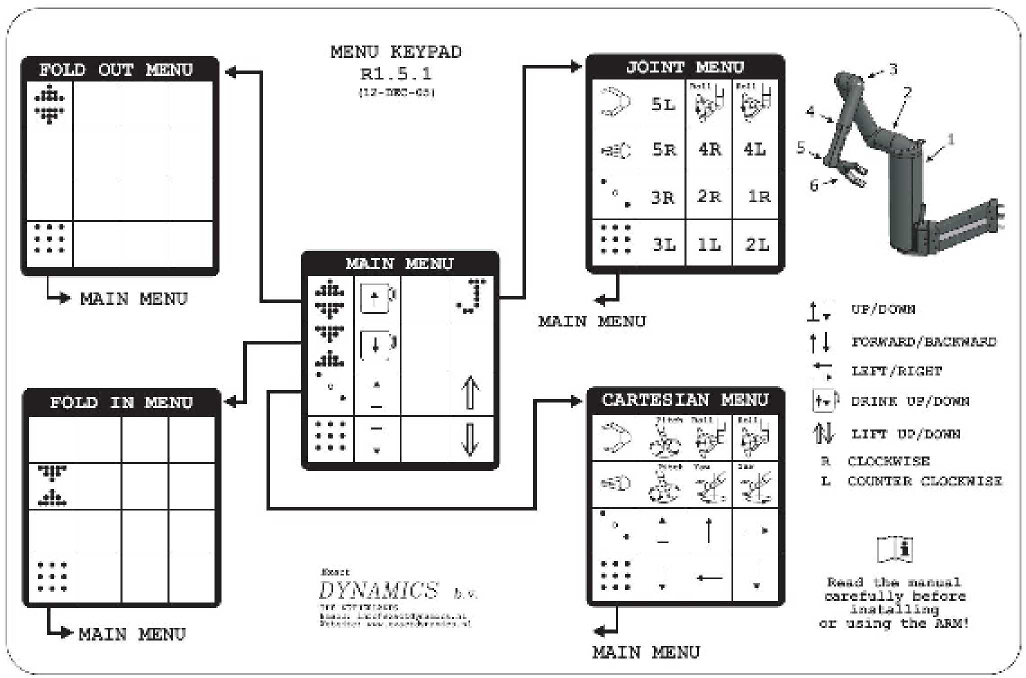
- Wheelchair Mounted Robotic Arms (WMRAs) such as MANUS and RAPTOR are capable of working in a variable workspace and an unstructured environment.
- They are capable of picking up miscellaneous objects from floor or shelves as well as carrying objects – tasks that have been identified by users as “high priority” in a collection of pre- and post-development surveys for a multitude of robotic assist devices.

Manus

Introduction
- MANUS users are concerned…

- The downside to the flexibility and versatility afforded by the MANUS is that it comes with the need for masterful control of a large number of degrees of freedom.
  - For many users (i.e., TBI), the cognitive load is excessive.
  - For other users, the entire process of shifting between layers of menus may be too tedious and frustrating.





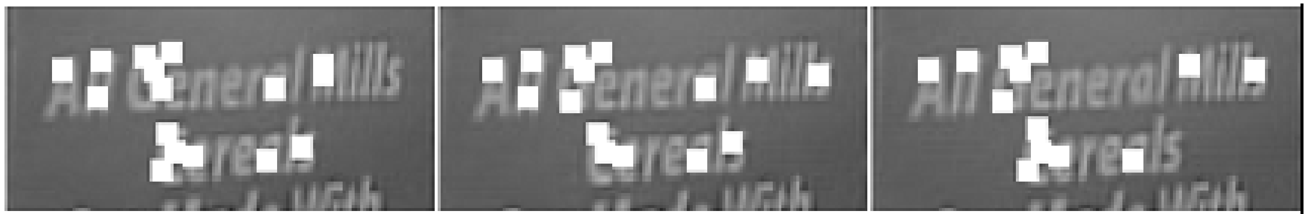
Introduction
- Related Work

- Vision based control and a variety of interfaces to simplify the operation of the MANUS
  - New Jersey Institute of Technology - an infrared sensory box, a stylus with joints mimicking the robot arm, and a computer mouse
  - TNO Science & Industry and the Delft University of Technology - the WMRA has been augmented with cameras, force torque sensors, and infrared distance sensors. It is operated by a wheelchair joystick and a switch in “pilot mode” to share autonomy between the robot arm and the user.
  - INRIA - a “one click” computer vision approach

Problem Statement
- Goal

- Design a motion control strategy for end-to-end automated object grasping
  - Joint angle feedback from the robot
  - Live video streams from an end-effector mounted stereo head

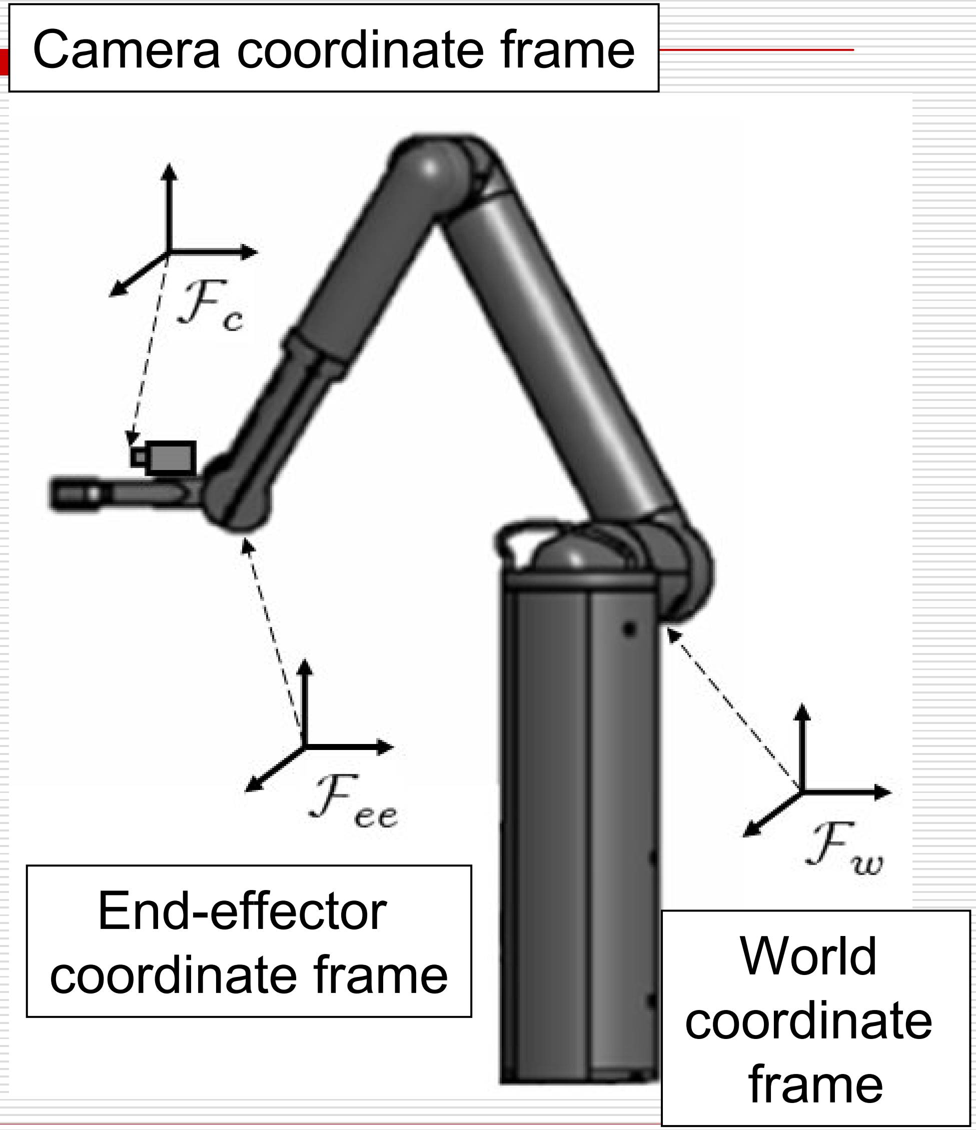
- We deal with everyday (ADL) objects in natural environments (i.e., variable illumination, background, etc.) that may be occluded by other objects in the vicinity or by virtue of their pose with respect to the end-effector.
  - Also, we work with natural features which may or may not be found/tracked in successive frames in the live view.



Problem Statement
- Nomenclature

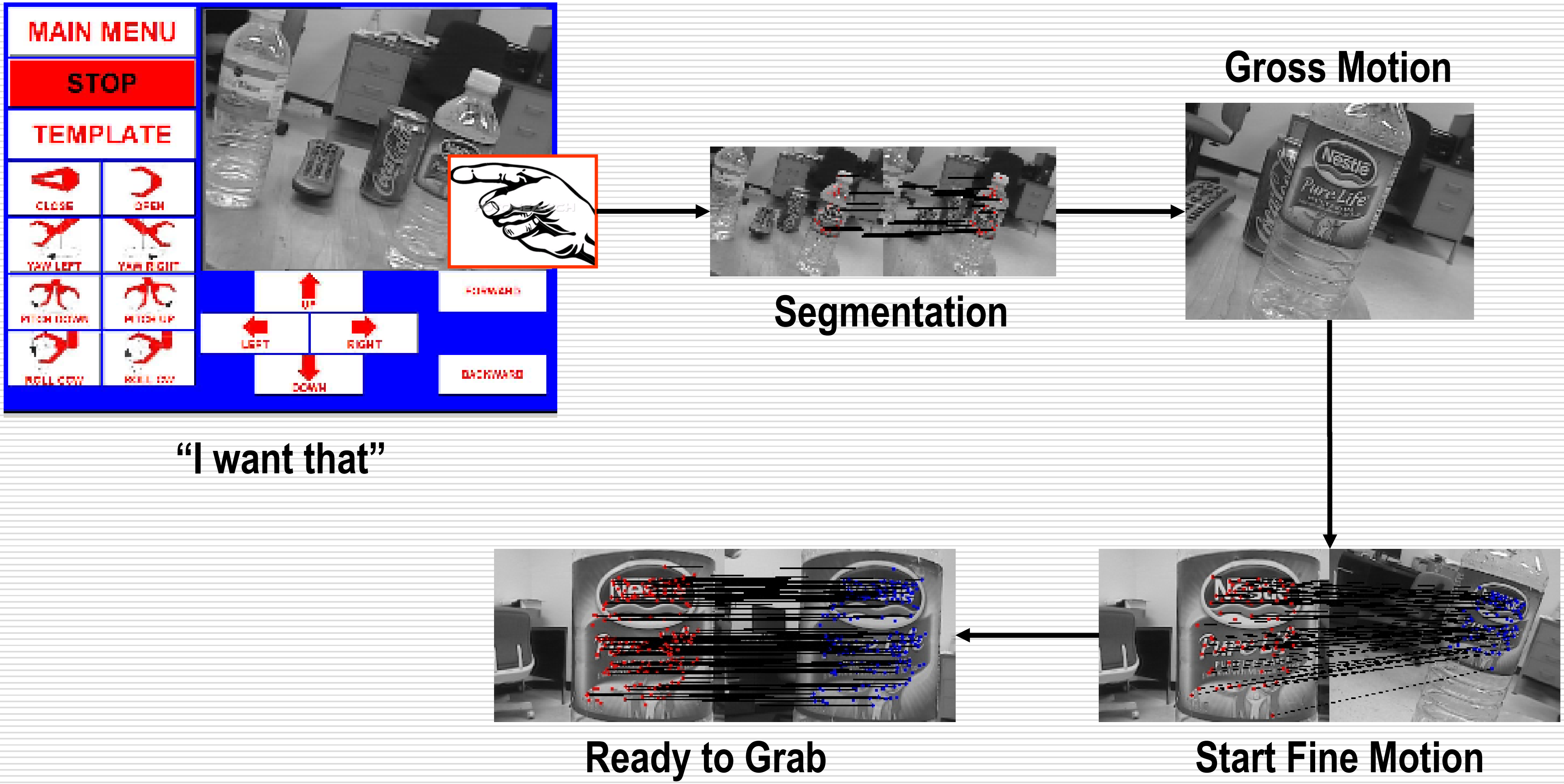
- $\mathbf{x}_z$: 3×1 position vector ***x*** in a coordinate frame $\mathcal{F}_z$
- $[\mathbf{x}_z]_y$ denotes its y^{th} component
- $[\mathbf{R}_z]_{xy}$: x,yth element of $\mathbf{R}_z$
- Rotation Transform Matrix $\mathbf{R}_{a2b}$
  - From the coordinate frame a to the coordinate frame b

$$\mathbf{R}_{c2w} = \mathbf{R}_{ee2w} \cdot \mathbf{R}_{c2ee}$$

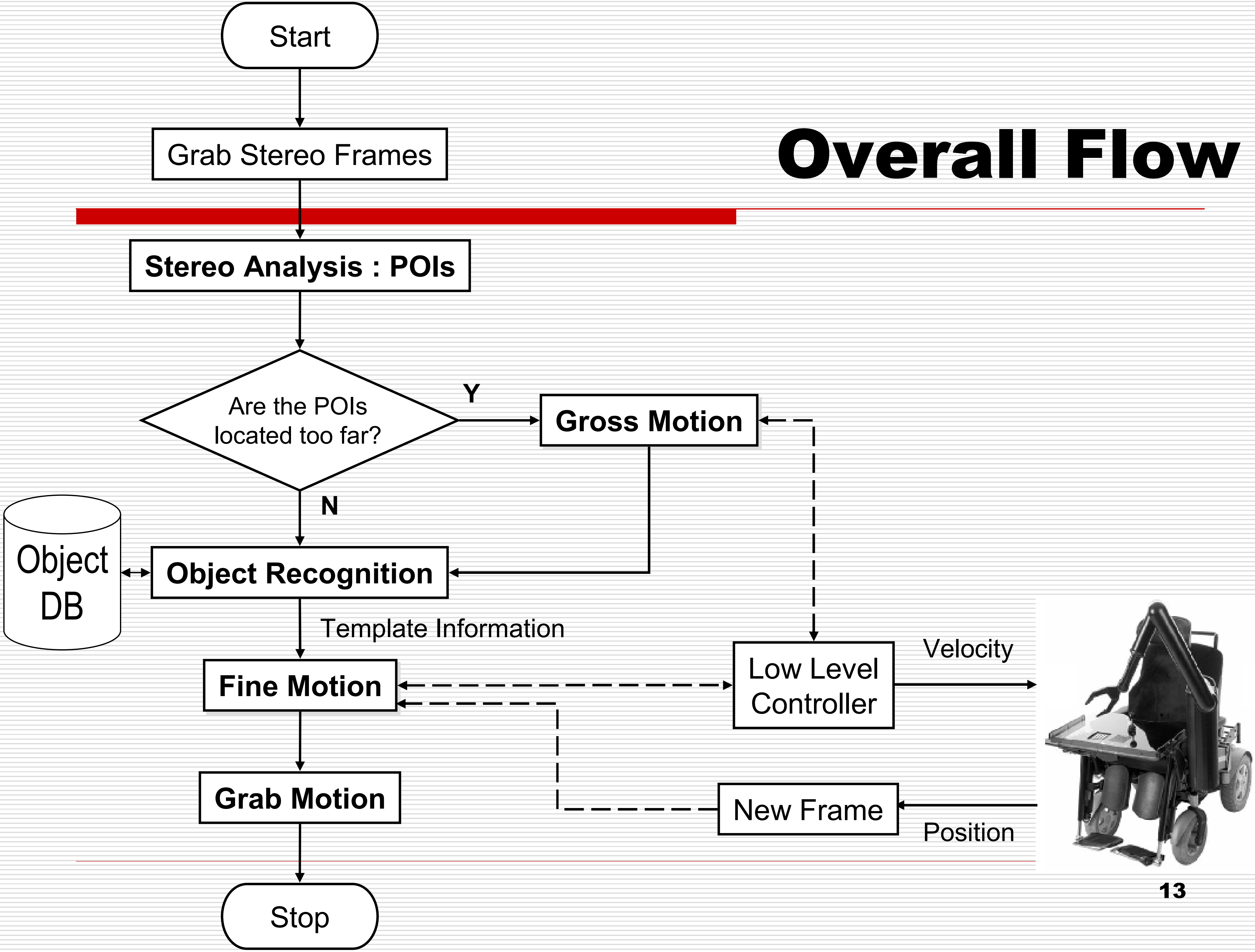


[REF] Homepage of Manus, http://www.exactdynamics.nl/, Exact Dynamics, Netherlands, 2009.

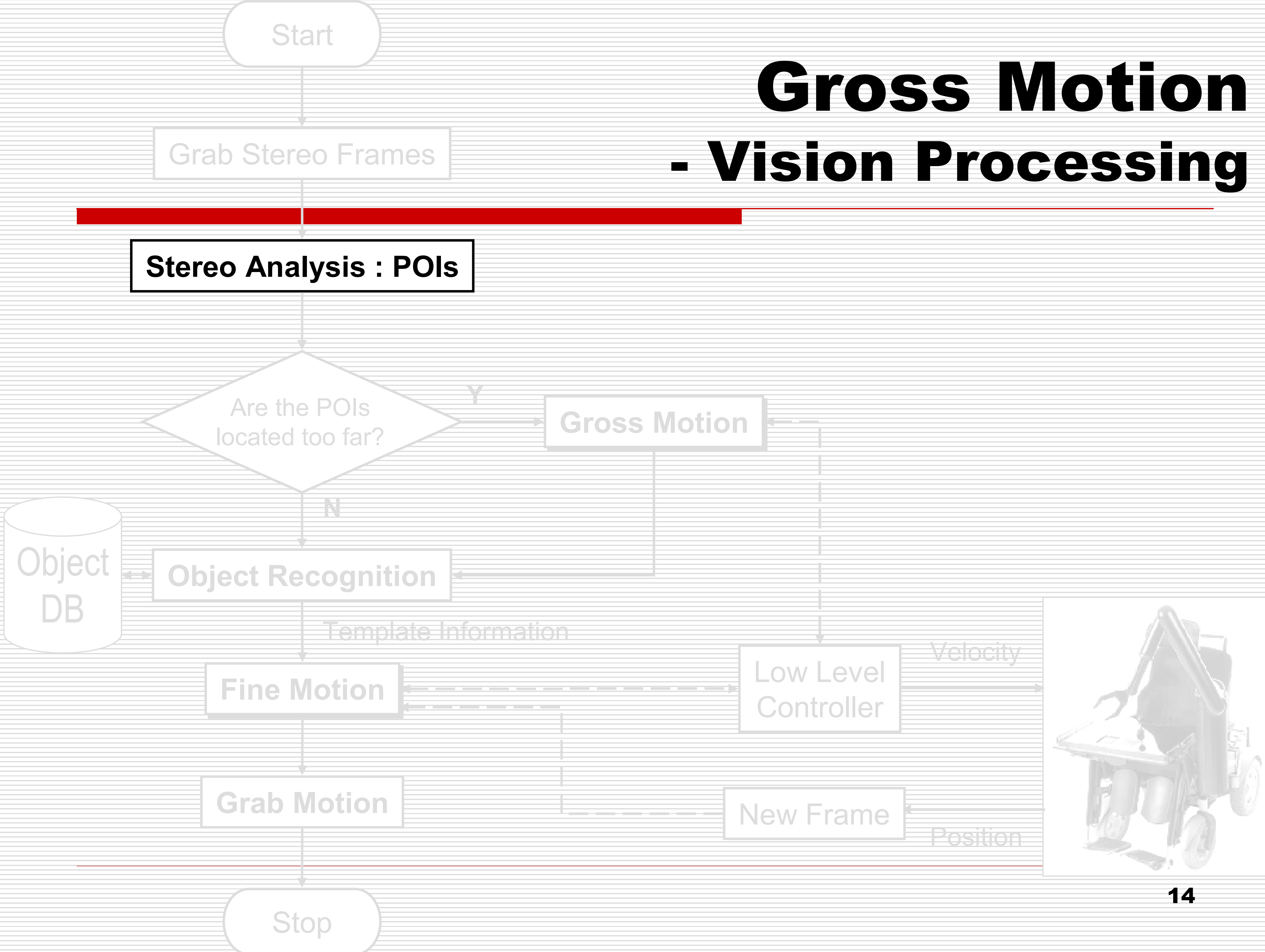
Approach
- Overview



Overall Flow



Gross Motion
- Vision Processing



Approach
- Gross Motion (I)

- Computing 3D information using SIFT feature descriptors
  - Least-squared optimization of error

$$\mathbf{e} = \left\| \bar{\mathbf{m}}_c^R - \hat{\mathbf{m}}_c^R \right\| \qquad \hat{\mathbf{m}}_c^R = \mathbf{R}_c^s \cdot \bar{\mathbf{m}}_c^L + \mathbf{T}_c^s.$$

  - 3D depth information

$$\mathbf{z}_c = \left(\mathbf{J}^T \cdot \mathbf{J}\right)^{-1} \left(\mathbf{J}^T \cdot \mathbf{R}_c^s \cdot \mathbf{m}_c^L\right)$$

$$\mathbf{z}_c = \left[\; z_c^R / z_c^L \quad 1/z_c^L \;\right] \qquad \mathbf{J} = \left[\; \mathbf{m}_c^R \quad -\mathbf{T}_c^s \;\right]$$

Approach
- Gross Motion (II)

- Initially, RANSAC is applied to cull outlier points.
- Constraints
  1. Depth ratio is used to eliminate points on the stool.
  2. World cubic formed by user's selected point is used to eliminate points on the closest bottle and coke can.
  3. Statistics on the residual 3D point cloud is used to eliminate points on the remote.

Initial View

Final View

Approach
- Gross Motion (III)

- Finding a surface normal vector on the target object
  - Laid down or upright

$$\mathbf{N}_w = \{\ \mathbf{n}_w^1, \quad \ldots, \quad \mathbf{n}_w^N\ \}$$

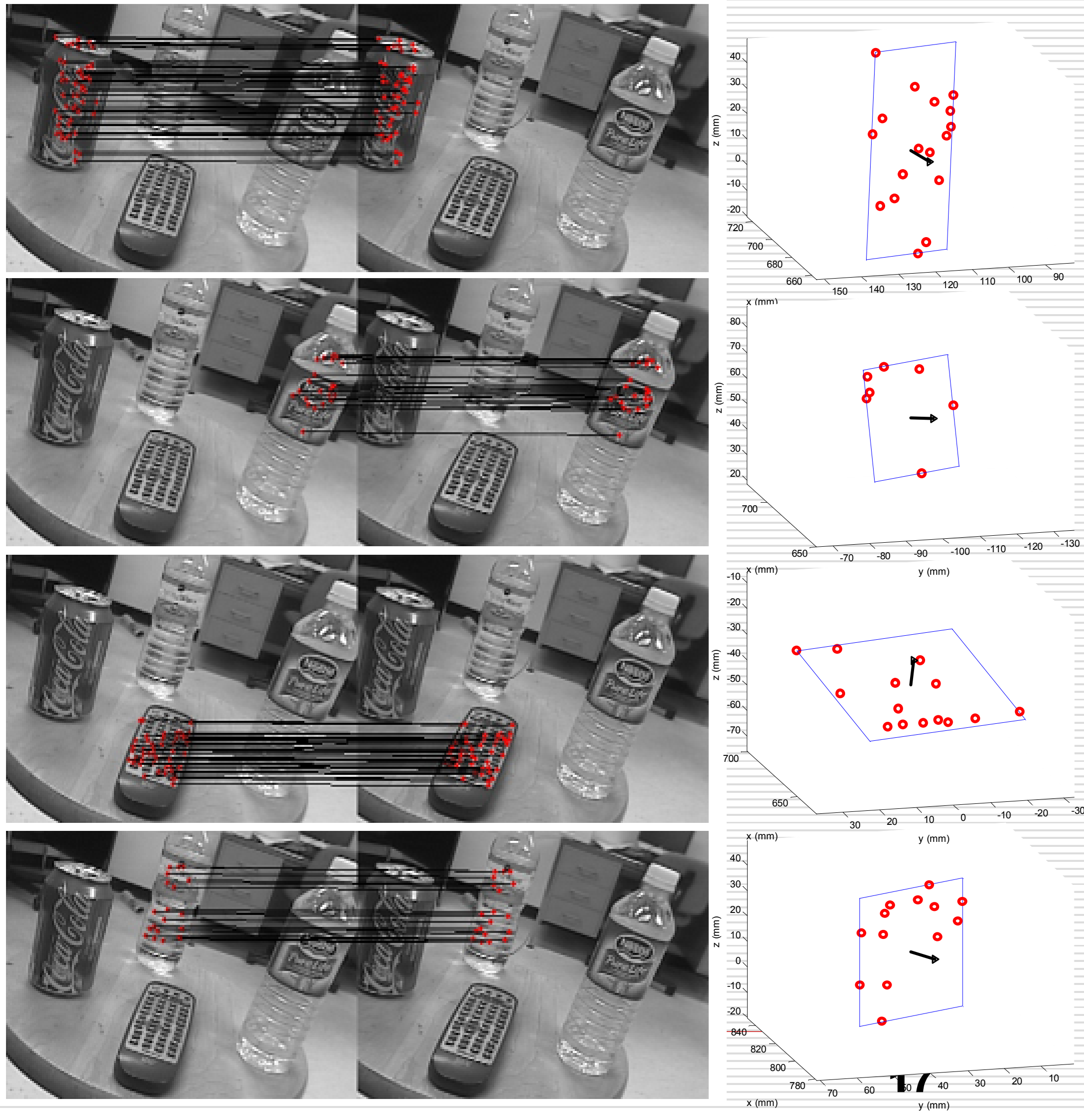
  - Voting by Optimization

$$\boxed{\arg\max_{\mathbf{j}} J(\mathbf{X}_w, \mathbf{n}_w^j)}$$

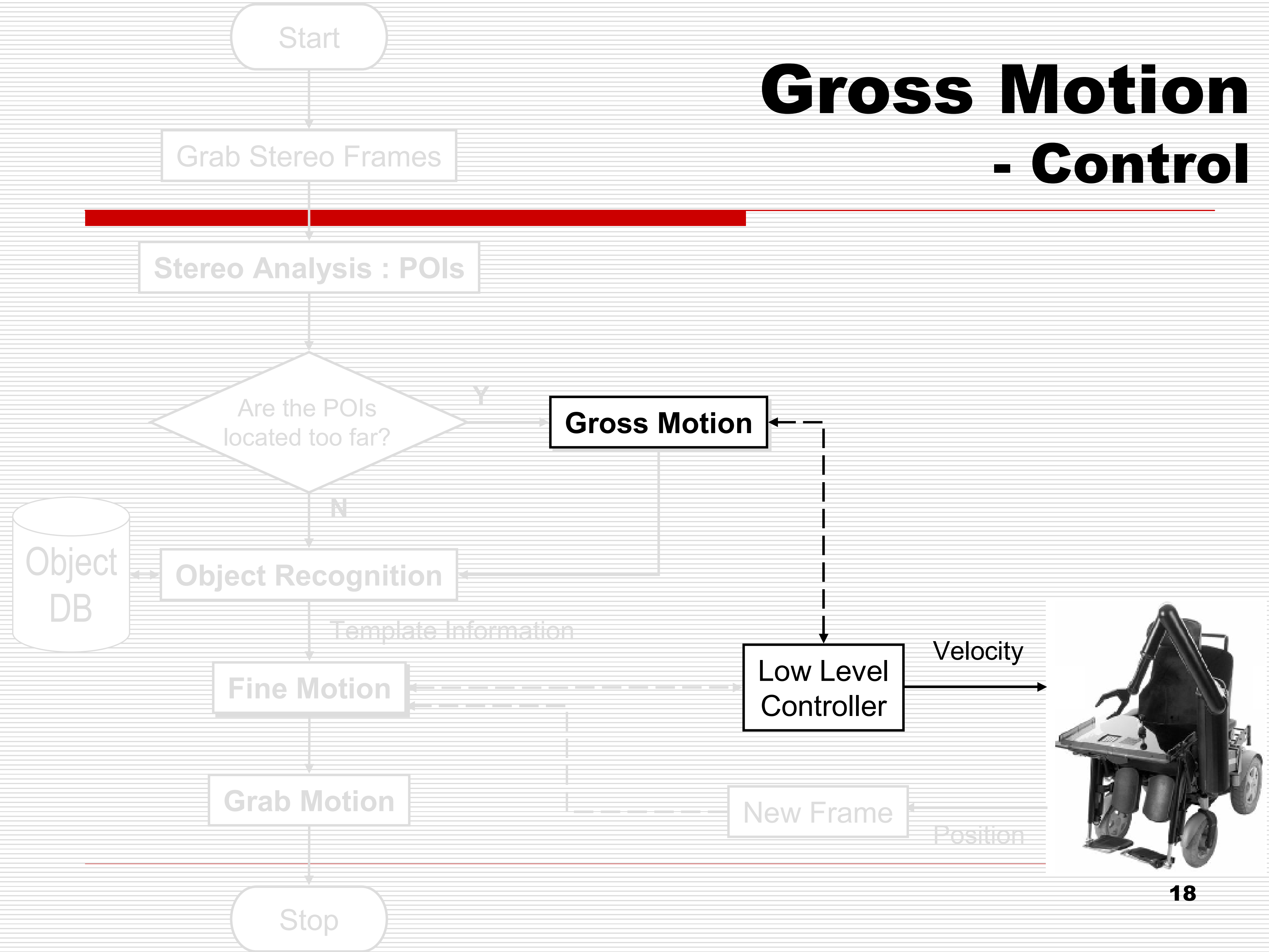
$$\mathbf{X}_w \triangleq \left[\mathbf{x}_w^1, \cdots, \mathbf{x}_w^{N_B}\right]$$

$$J(\mathbf{X}_w, \mathbf{n}_w^i) \triangleq \sum_{j=1}^{N_B} \sum_{k=1}^{N_B} f\left(\mathbf{W} \cdot \left(\mathbf{x}_w^k - \mathbf{x}_w^j\right)^T \mathbf{n}_w^i, \delta\right)$$

$$f(\alpha, \delta) = \begin{cases} 1 \text{ for } \alpha < \delta \\ 0 \text{ for } \alpha \geq \delta \end{cases}$$



Gross Motion
- Control



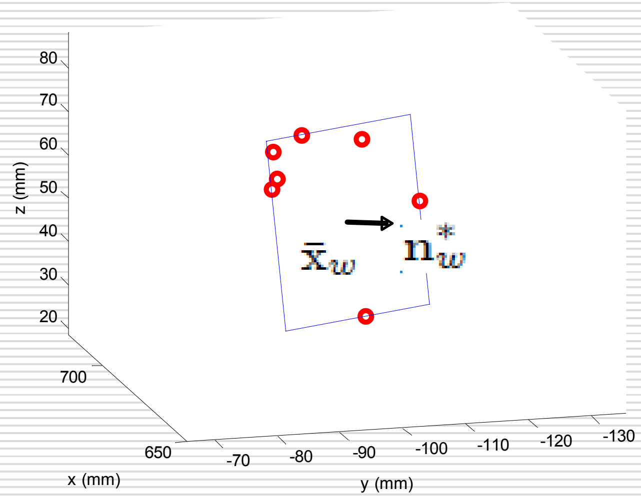
Approach
- Gross Motion (IV)

- Motion Control
  - Given the computed 3-D target position and the surface normal, one can compute the desired setpoints for position and orientation of the robot end-effector.

$$
\begin{aligned}
\mathbf{p}_w^* &= \mathbf{p}_w^t + \mathbf{p}_w^{ee} + {}^w\mathbf{R}_{ee} \cdot \mathbf{d}_{ee}^c \\
[\theta_w^*]_1 &= \operatorname{atan2}([\mathbf{n}_w^*]_2 , [\mathbf{n}_w^*]_1) \\
[\theta_w^*]_2 &= \operatorname{atan2}\left([\mathbf{n}_w^*]_3 , \sqrt{([\mathbf{n}_w^*]_1)^2 + ([\mathbf{n}_w^*]_2)^2}\right)
\end{aligned}
$$

$$
\mathbf{p}_w^t = {}^w\mathbf{R}_c \cdot \left(\mathbf{e}_c^t \cdot [\bar{\mathbf{x}}_w]_3 - \mathbf{n}_c^* \cdot d_c^o \right)
$$

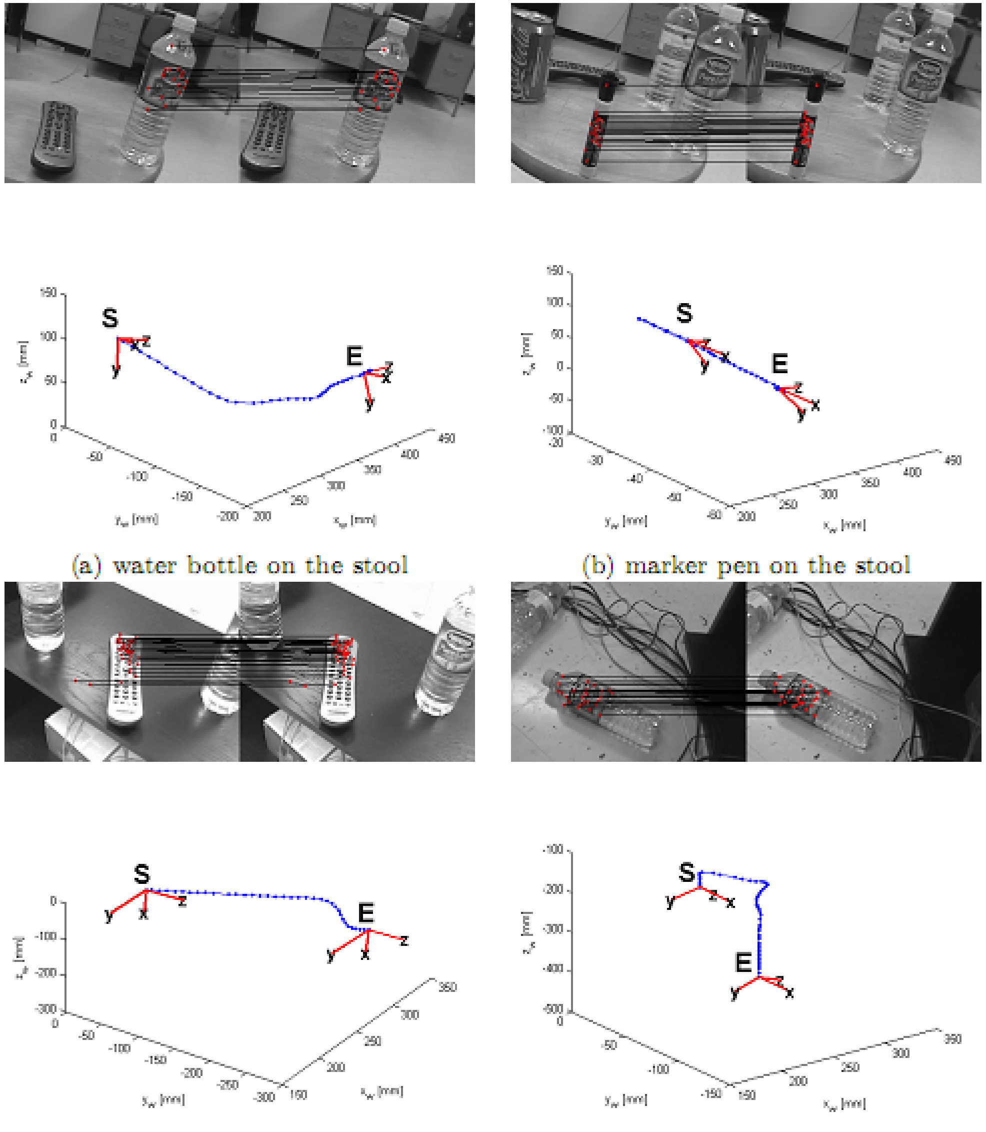
$$
\mathbf{e}_c^t \triangleq \left[([\mathbf{m}_c^s]_1 - [\mathbf{m}_c^o]_1) \; ([\mathbf{m}_c^s]_2 - [\mathbf{m}_c^o]_2) \; 1 \right]^T
$$



Approach
- Gross Motion (V)

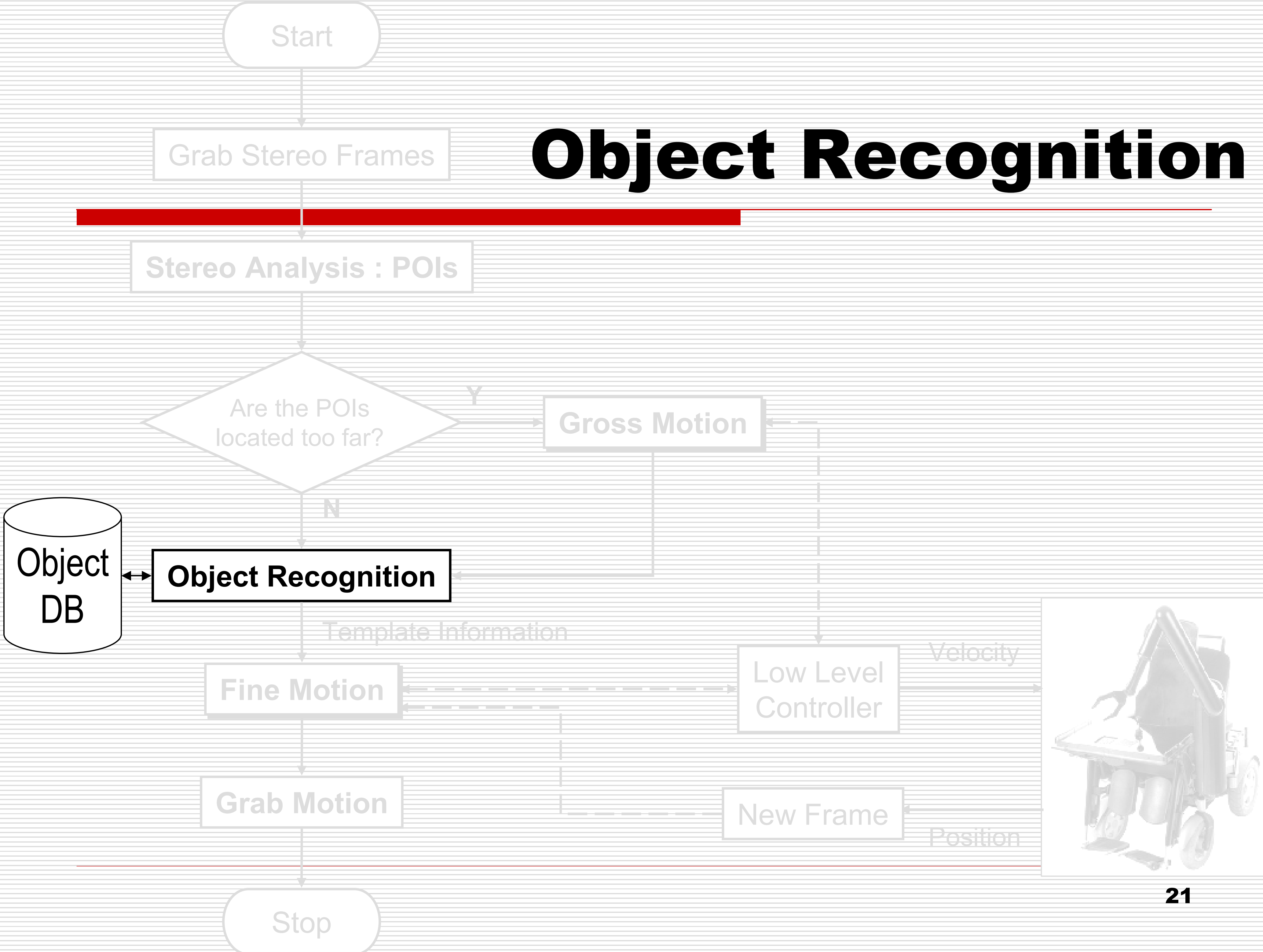
- Motion Control
  - P-control is employed to generate translational and rotational velocity commands.

$$\begin{bmatrix} \mathbf{v}_c \\ \omega_c \end{bmatrix} = \begin{bmatrix} \mathbf{K}_p & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times1} & \mathbf{K}_\theta \end{bmatrix} \begin{bmatrix} \mathbf{p}_w - \mathbf{p}_w^* \\ \theta_w - \theta_w^* \end{bmatrix}$$



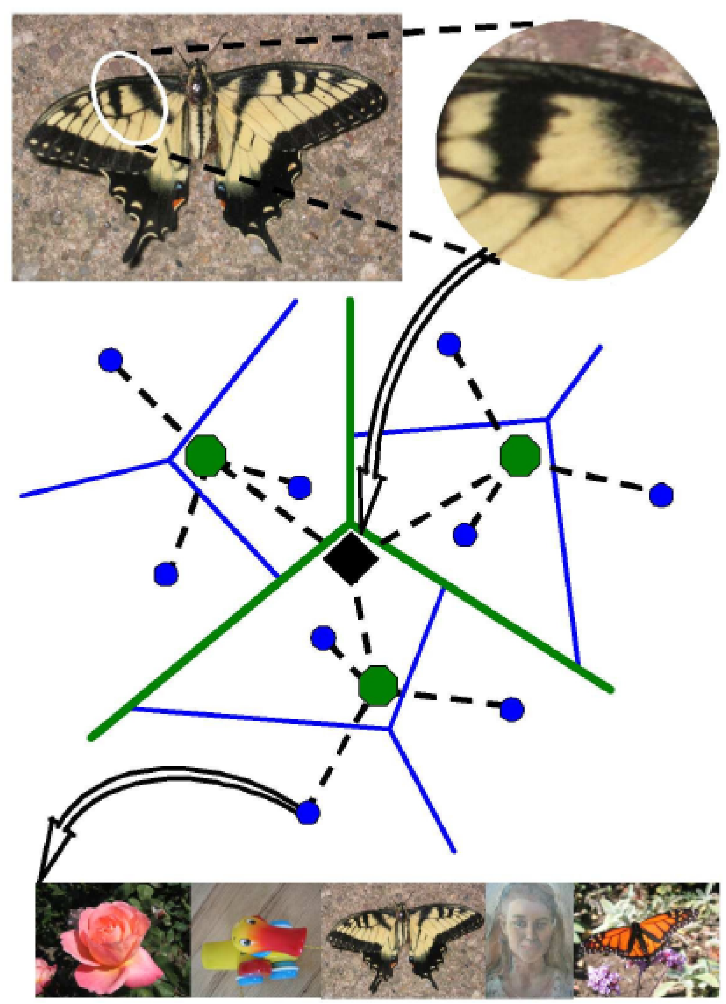
(a) water bottle on the stool (b) marker pen on the stool (c) remote on the low table (d) water bottle on the floor

Object Recognition



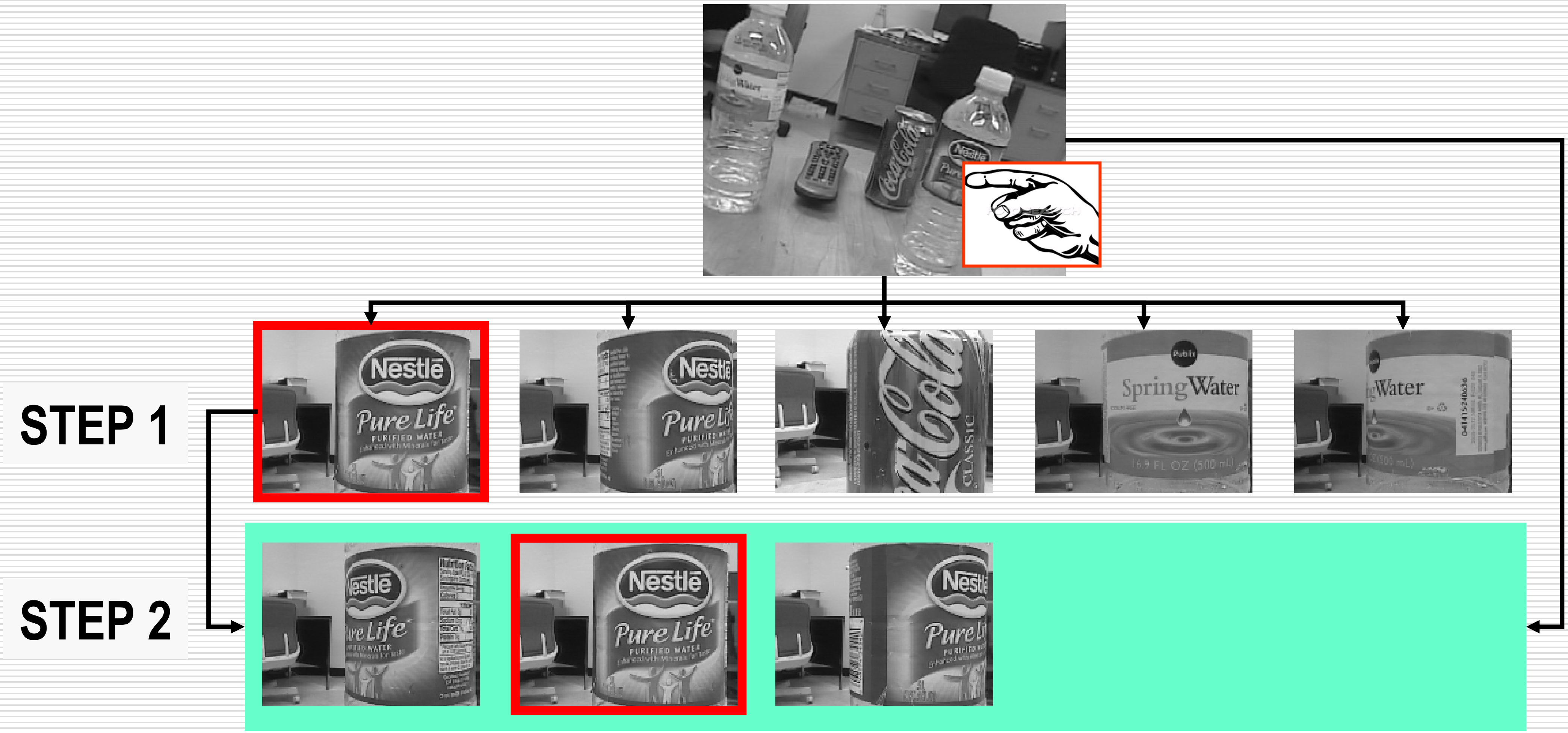
Approach
- Object Recognition (I)

- Object Recognition using SIFT descriptors
  - Computationally intractable especially when the database grows extremely large
- Vocabulary tree that provides for scalable recognition (SRVT)
  - A multi-level decision tree and visual keywords as leaf nodes
  - Easily extendible and scalable to deal with lots of different natural scenes

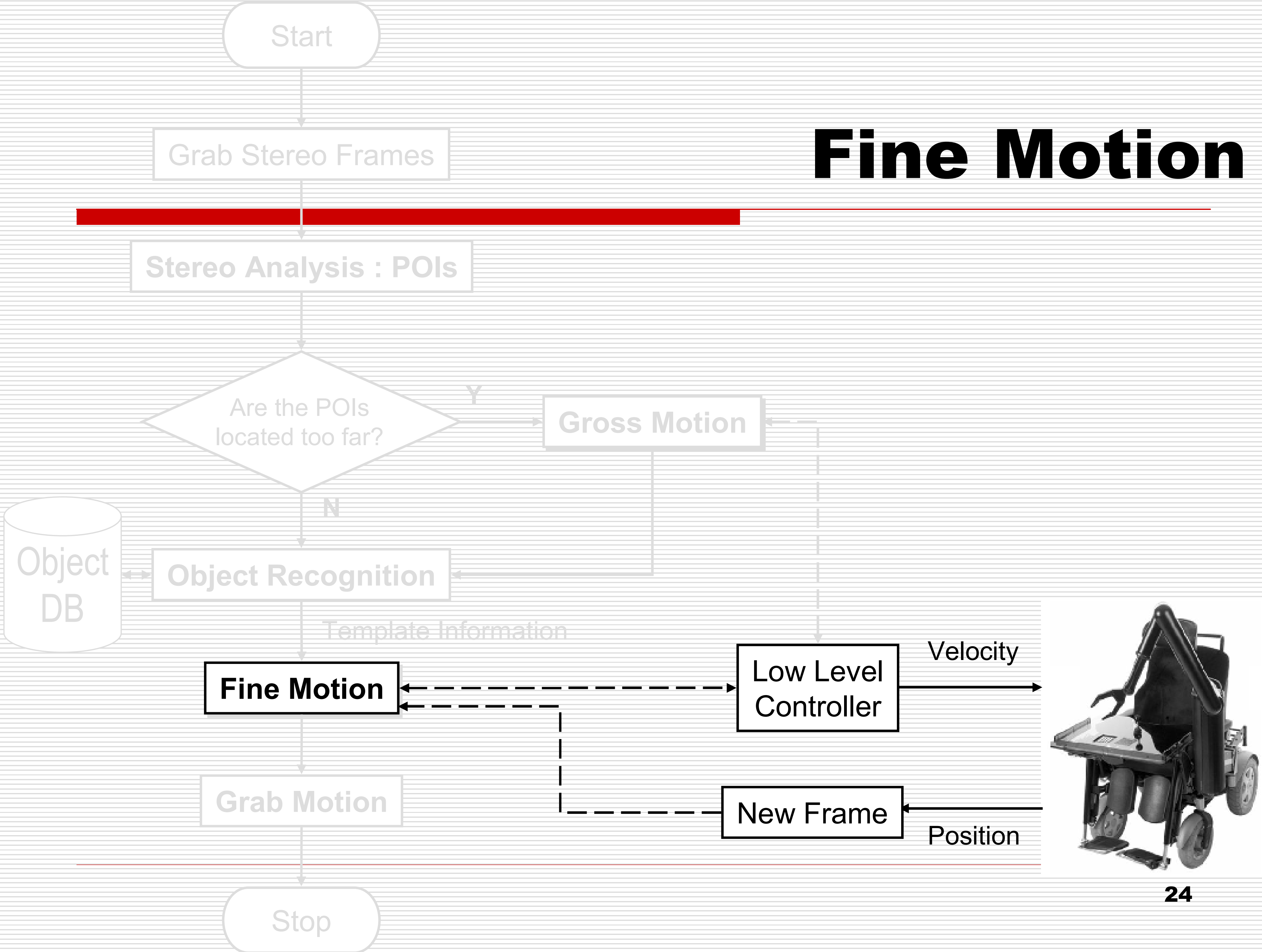


[REF] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree", CVPR 2006.

Approach
- Object Recognition (II)



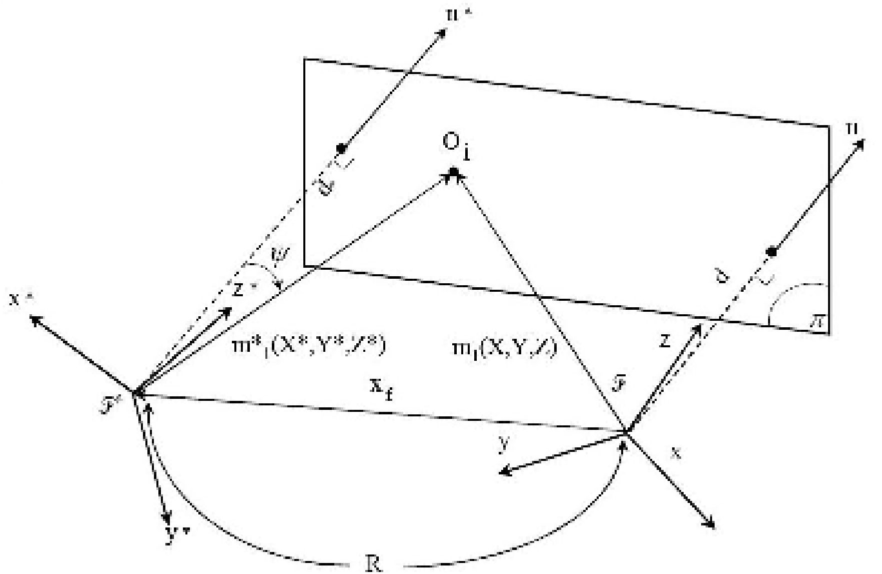
Fine Motion


Start
Grab Stereo Frames
Stereo Analysis : POIs
Are the POIs located too far?
Y
Gross Motion
N
Object DB
Object Recognition
Template Information
Fine Motion
Low Level Controller
Velocity
New Frame
Position
Grab Motion
Stop

Approach
- Fine Motion (I)

- A 2-1/2D (or hybrid) visual servoing scheme is adopted to generate translation and rotation motion to align current eye-in-hand view with the pertinent template database image.
  - Homography computation using matched local features
  - Euclidean decomposition of homography
    - Choice of one solution using auxiliary stereo frame
  - For translation motion, one of the local descriptors is used as an anchor point **m** to track in x-y plane of the camera coordinate frame.
  - For approaching motion, the depth ratio is used to define an error signal.

$$\log\left(\frac{Z_1}{Z_1^*}\right) = \log\left(\frac{(1+\mathbf{n}^T\mathbf{x}_h)\mathbf{n}^{*T}\mathbf{m}^*}{\mathbf{n}^T\mathbf{m}}\right)$$



Approach
- Fine Motion (II)

- Three basic control laws are developed as
  - 2D-IBVS (Image Based Visual Servoing)

$$\boxed{\mathbf{v}_w = \mathbf{K}_v \cdot {}^{w}\mathbf{R}_c \cdot \mathbf{e}_c}$$

  - HVS (Hybrid Visual Servoing)

$$\boxed{\mathbf{v}_w = \mathbf{K}_v \cdot {}^{w}\mathbf{R}_c \cdot \mathbf{e}_v}$$

$$\boxed{\omega_{ypr} = \mathbf{K}_\omega \cdot {}^{ypr}\mathbf{R}_w \cdot \mathbf{e}_w}$$

$$ {}^{ypr}\mathbf{R}_w = \begin{bmatrix} 0 & [{}^{w}\mathbf{R}_c]_{23} & [{}^{w}\mathbf{R}_c]_{13} \\ 0 & -[{}^{w}\mathbf{R}_c]_{13} & [{}^{w}\mathbf{R}_c]_{23} \\ 1 & 0 & [{}^{w}\mathbf{R}_c]_{33} \end{bmatrix}$$

  - HVS-T (Translation only)

$$\boxed{\mathbf{v}_w = \mathbf{K}_v \cdot {}^{w}\mathbf{R}_c \cdot \mathbf{e}_v}$$

Approach
- Fine Motion (III)

- Phase I – Centering the Target Object
  - 2D-IBVS

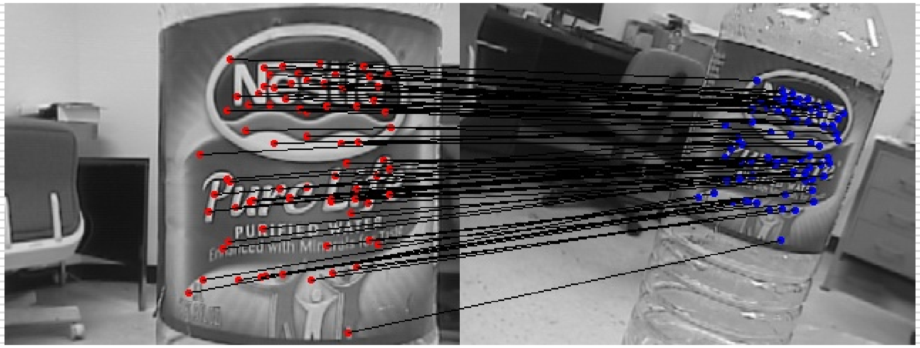
$$\mathbf{e}_c = \mathbf{p}_c^e - \mathbf{p}_c^o \qquad \mathbf{p}_c^o = \begin{bmatrix} \mathbf{m}_x^o & \mathbf{m}_y^o \end{bmatrix}^T \quad \mathbf{p}_c^e = \begin{bmatrix} \mathbf{m}_x^e & \mathbf{m}_y^e \end{bmatrix}^T$$

- Phase II – Alignment with Target Object's Template
  - HVS or HVS-T

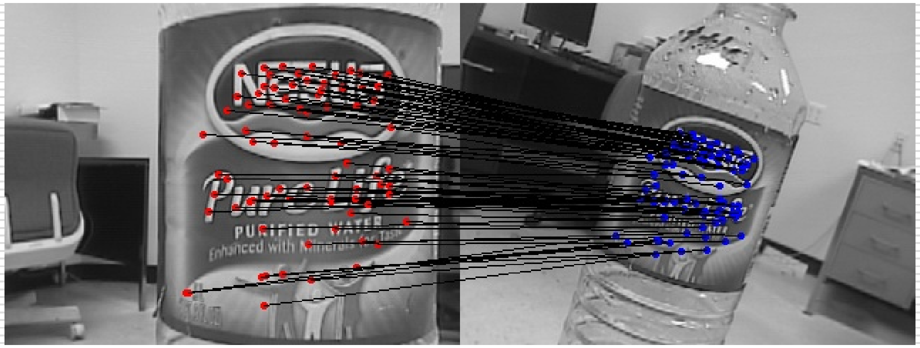
$$[\mathbf{e}_v]_{1,2} = \mathbf{p}_c^f - \mathbf{p}_c^*, \quad [\mathbf{e}_v]_3 = \log\left(\frac{Z_1}{Z_1^*}\right) \qquad \mathbf{p}_c^f = \begin{bmatrix} \mathbf{m}_x^f \; \mathbf{m}_y^f \; 1 \end{bmatrix}^T$$

$$\mathbf{e}_w = \theta_w^f - \theta_w^* \qquad \theta_w^* = \begin{bmatrix} \arctan([\mathbf{R}]_{21}, [\mathbf{R}]_{11}) \\ \arctan([\mathbf{R}]_{31}, \sqrt{([\mathbf{R}]_{32})^2 + ([\mathbf{R}]_{33})^2}) \\ -\arctan([\mathbf{R}]_{32}, [\mathbf{R}]_{33}) \end{bmatrix}$$

Approach
- Fine Motion (IV)



Initial Pose

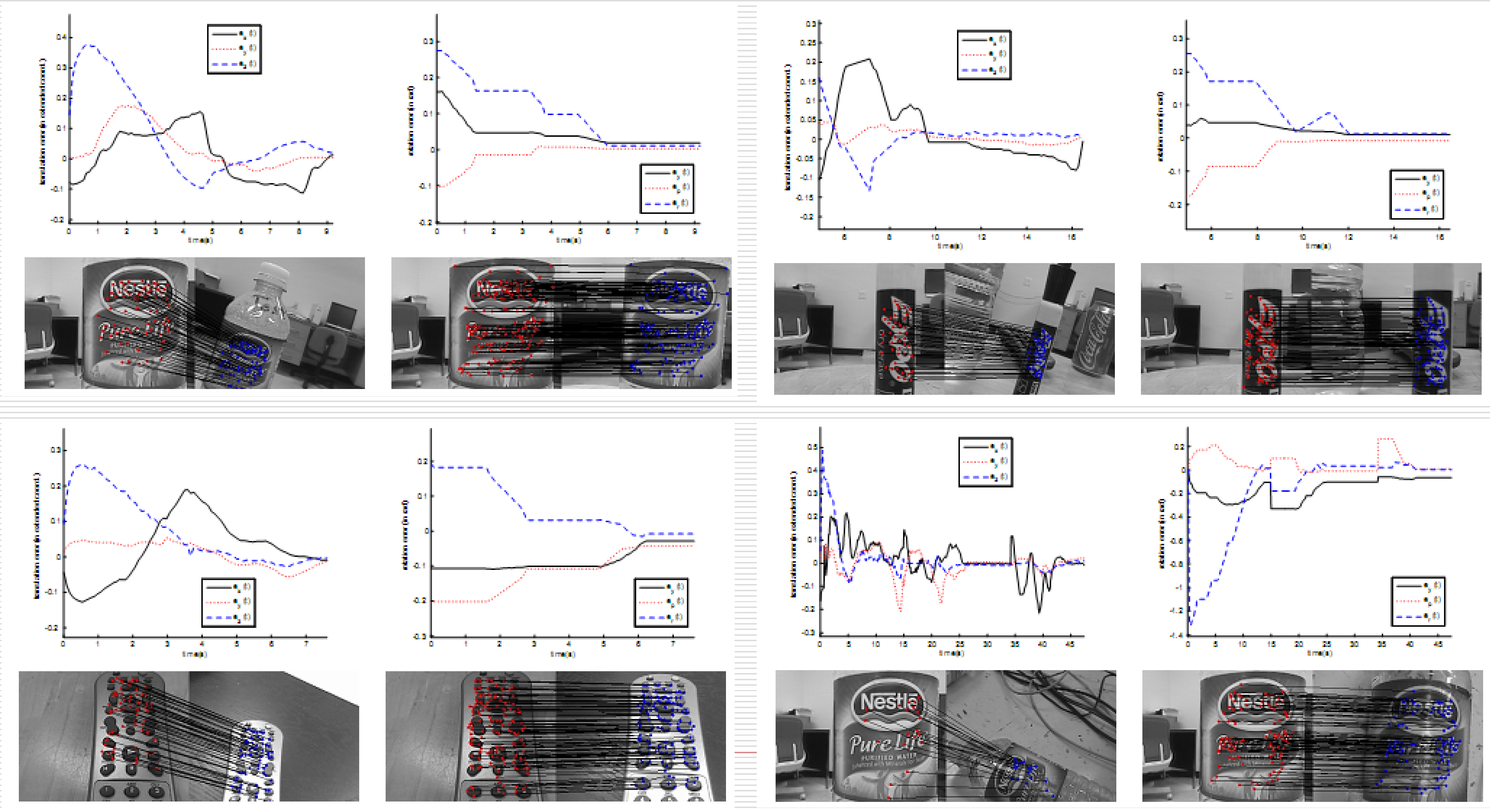


Phase I: Centering

Phase II: Alignment

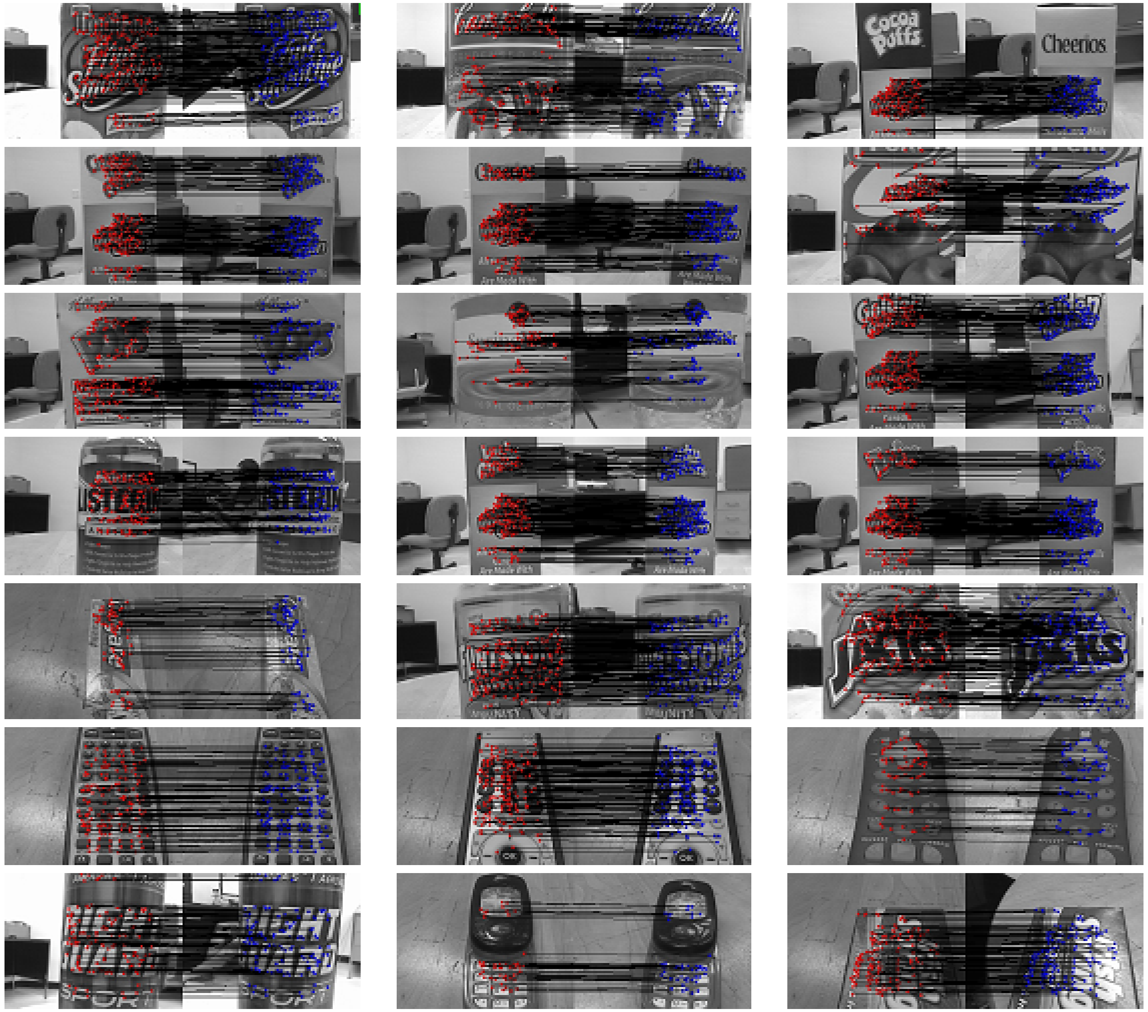
Final Pose

Approach
- Fine Motion (V)

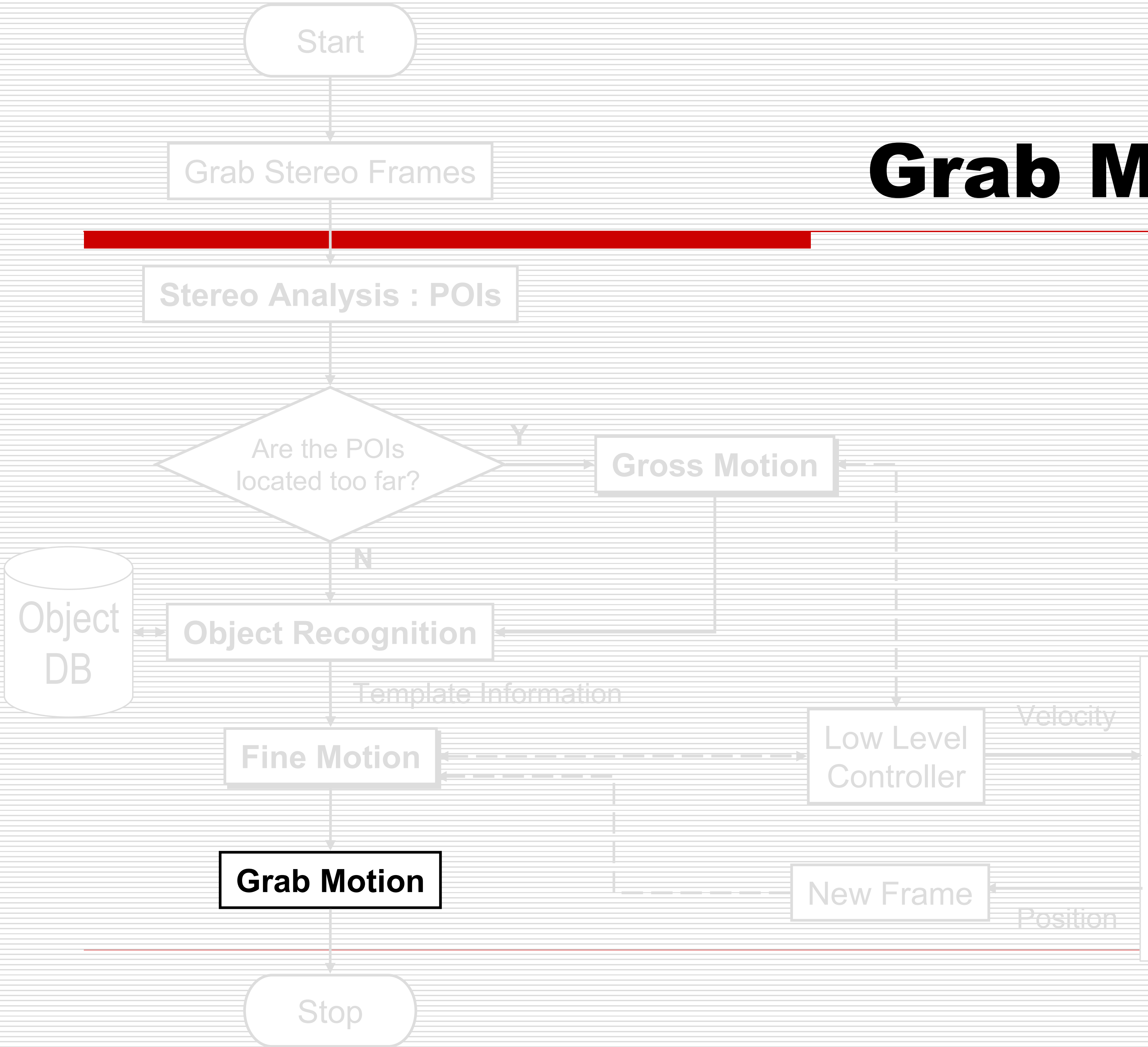


Approach
- Fine Motion (VI)

- More than 25 unique objects were tested.

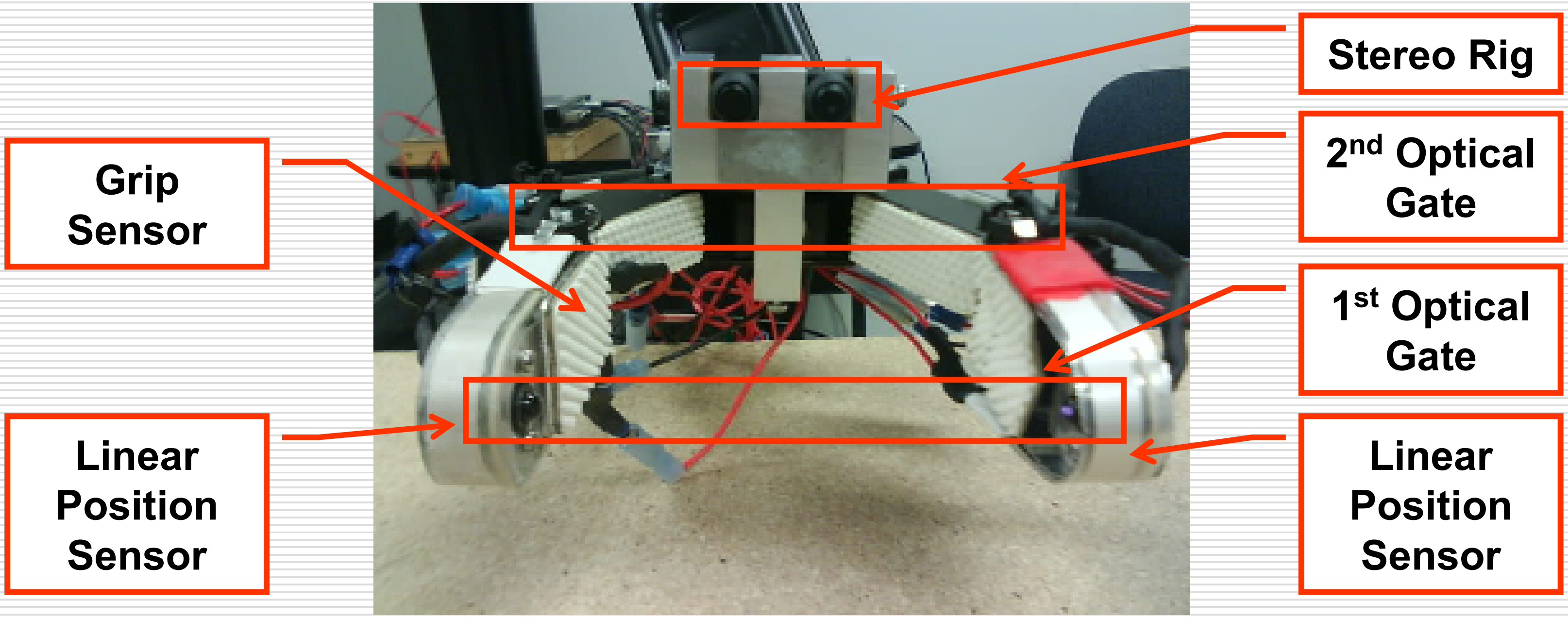


Grab Motion

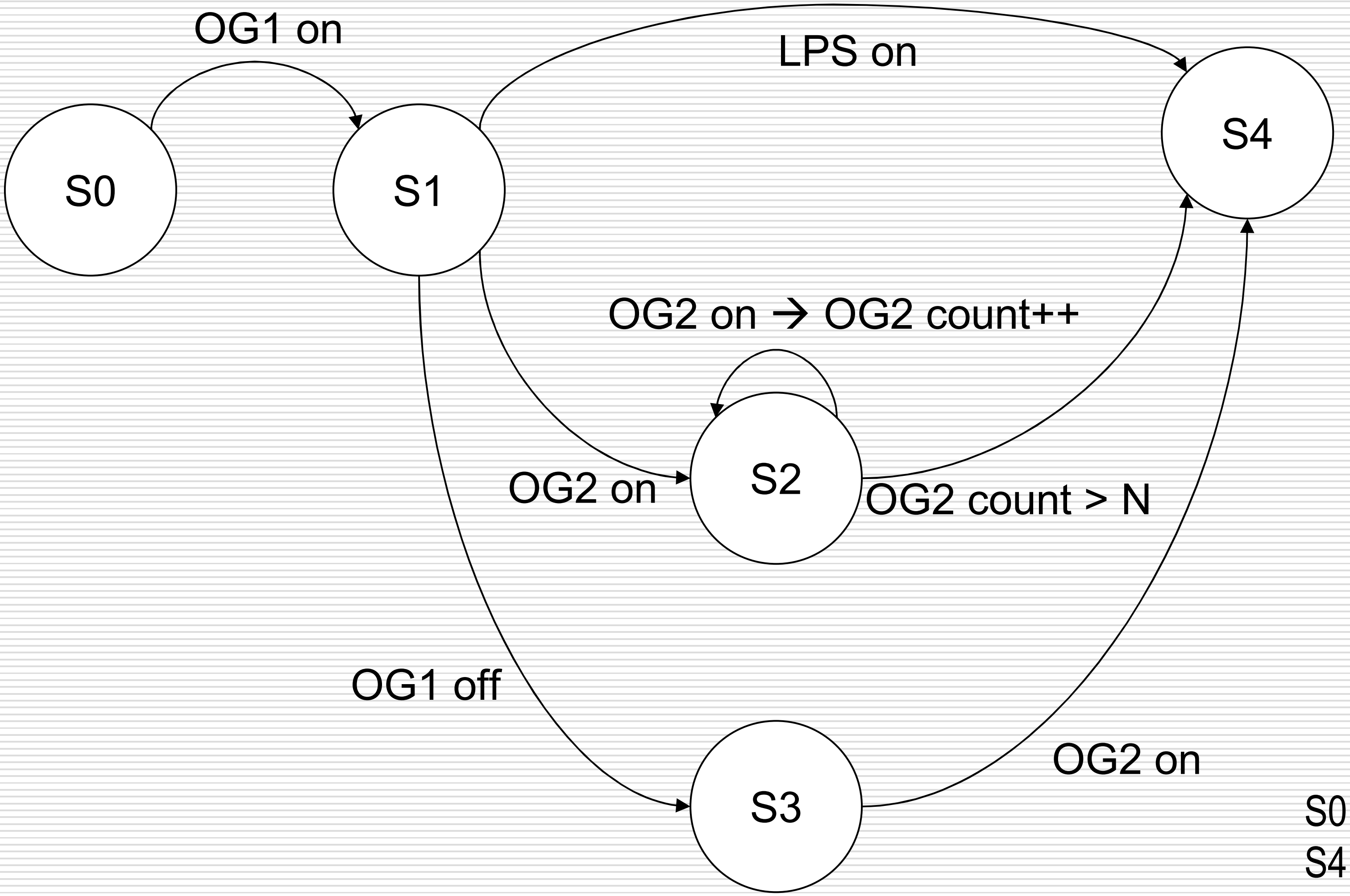


Start
Grab Stereo Frames
Stereo Analysis : POIs
Are the POIs located too far?
Y
Gross Motion
N
Object DB
Object Recognition
Template Information
Fine Motion
Low Level Controller
Velocity
Grab Motion
New Frame
Position
Stop

Approach
- Robust Grabbing H/W Configuration

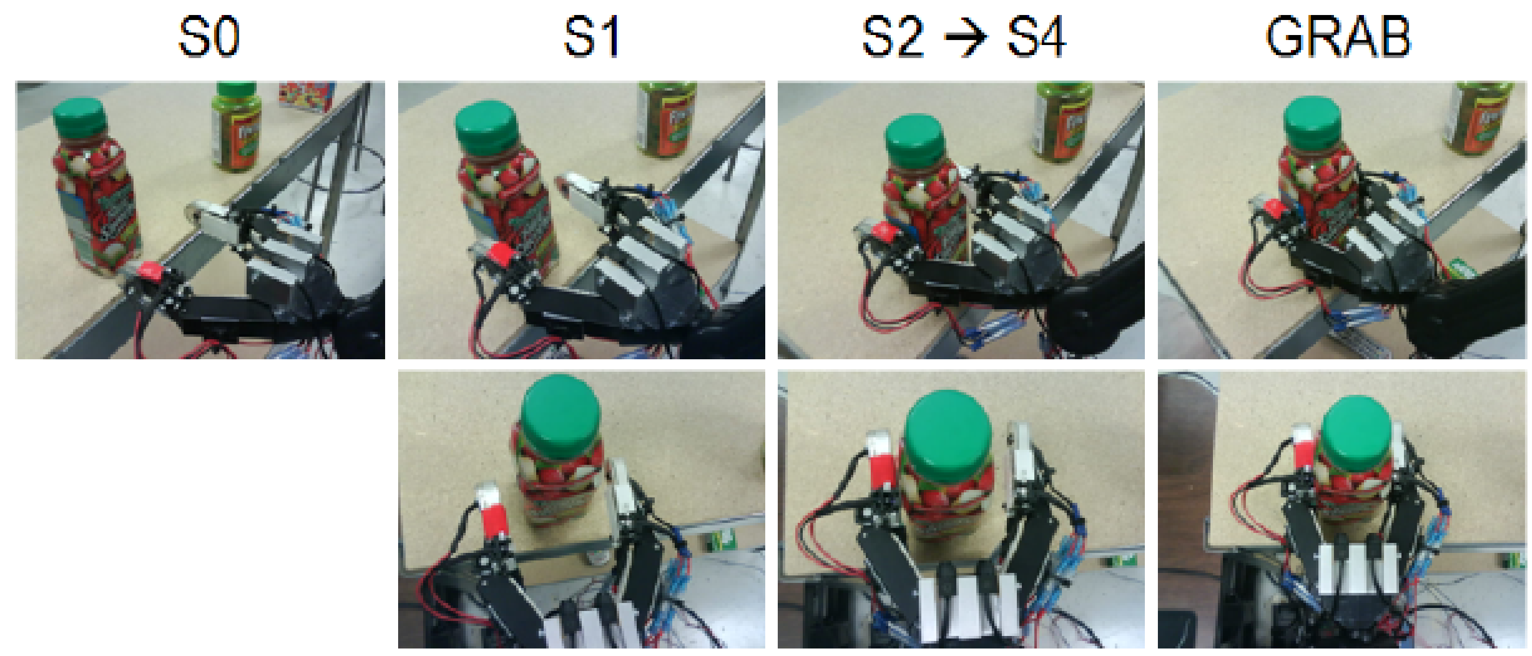


Approach
- Finite State Machine

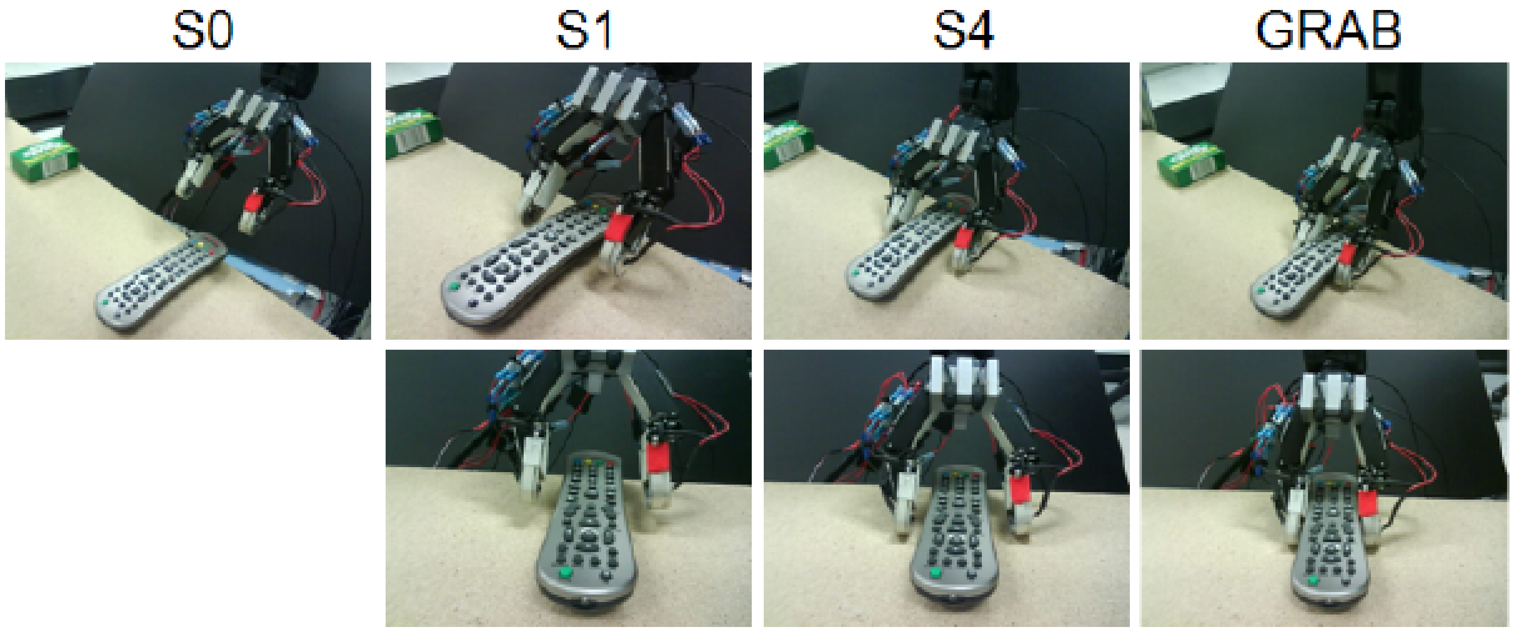


S0: Initial State
S4: Final State (Grab)

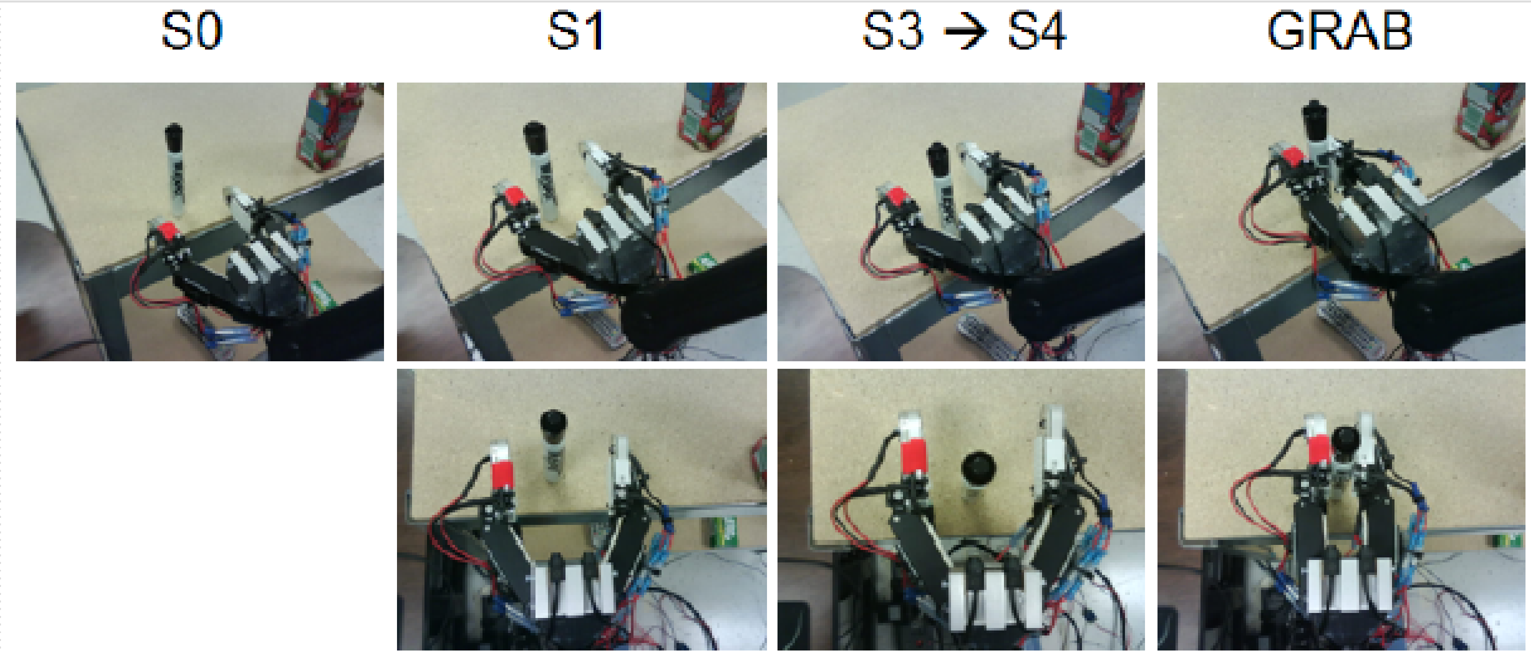
Approach
- Grabbing Example I: Upright Object



Approach
- Grabbing Example II: Laid-down Object



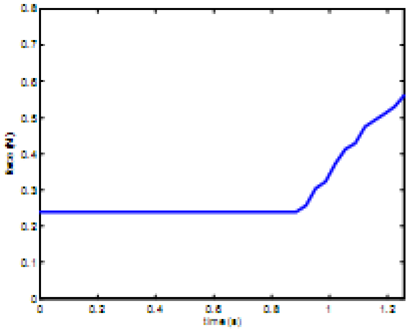
Approach
- Grabbing Example III: Upright & Thin Object

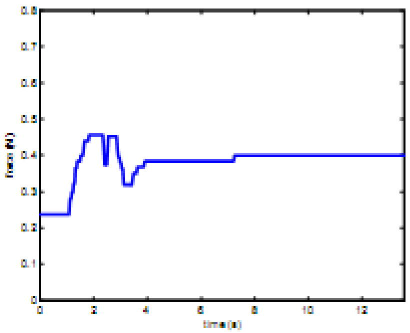


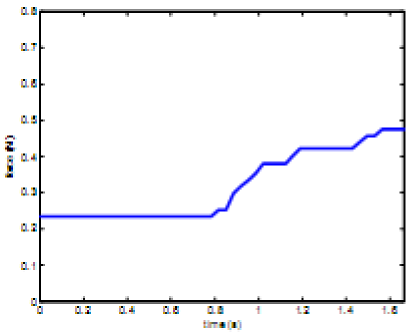
Approach
- Force-Adaptive Grasping

- Adaptive Grabbing vs. Non Adaptive Grabbing
  - AG & NAG works for Filled Coke Can
  - NAG fails for Empty Coke Can
  - AG works for Empty Coke Can
- AG - Finding a 'flatness'

$$g(t) = \int_{t-\Delta T}^{t} f'(\tau)d\tau$$

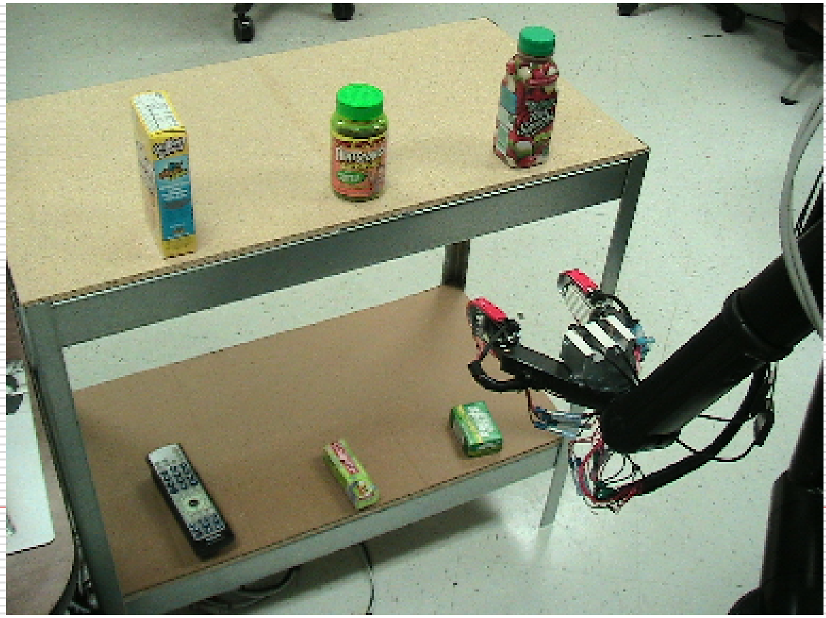


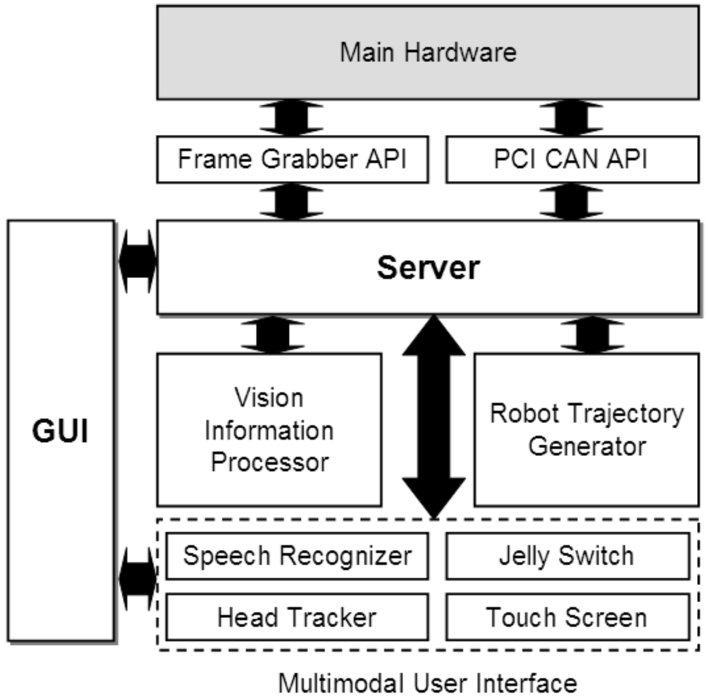




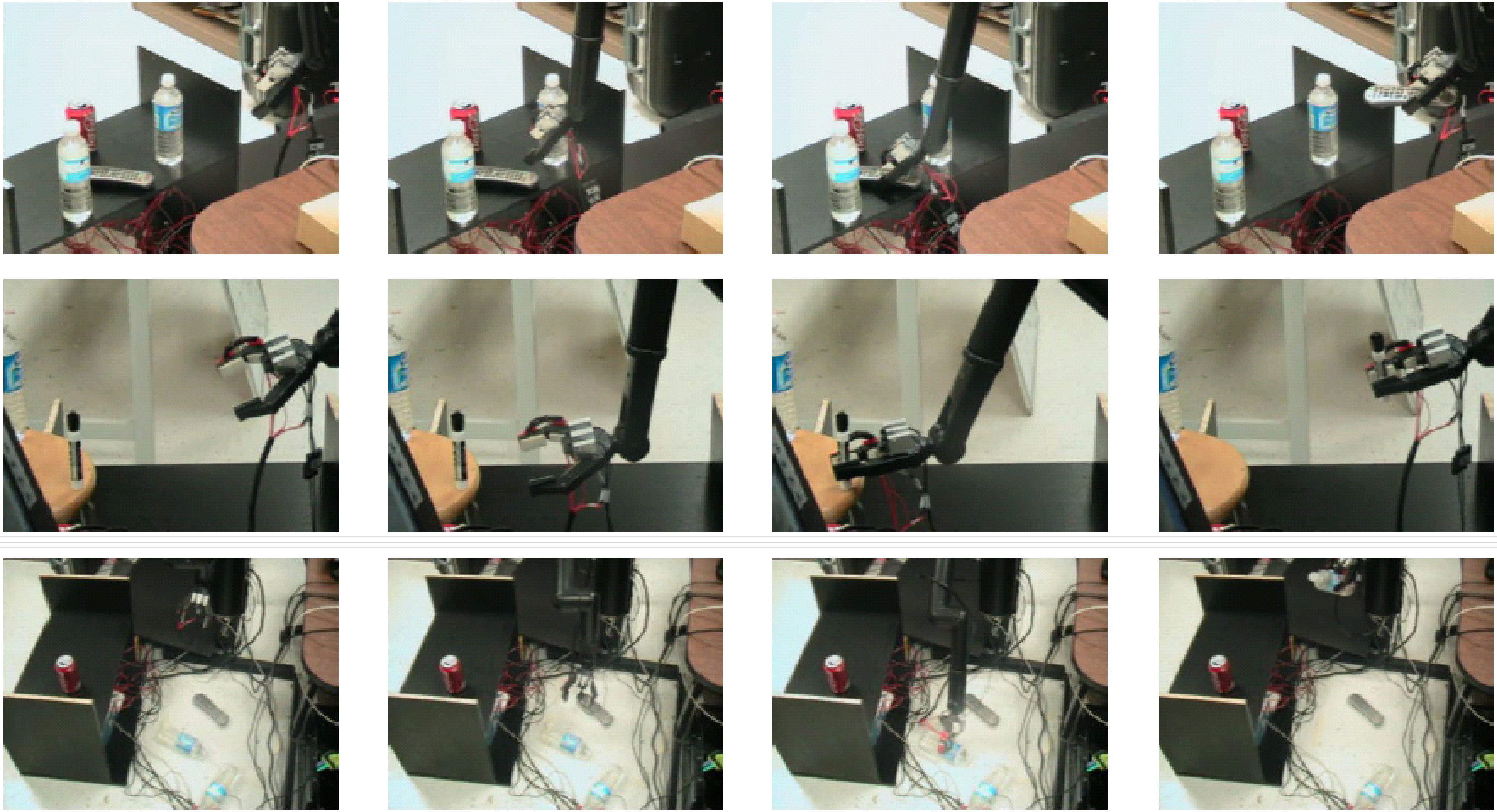
Results
- Setup

- Manus Robot + Multi-modal User Interfaces
- Sensors
  - Narrow baseline wide-angle stereo rig – Point Grey's Dragonfly 2
  - Smart Grabber – FSR, LPS, OG
- Software Architecture
  - Distributed TCP/IP Server-Client





Results
- Pick Up Multiple Objects



Results
- Object/Template Management

- Able to capture the object in the cluttered scene
- Add multiple view of the target object
- Web based template sharing and expansion

Results
- User Testing

- 2009.5.4-5.22 @ Orlando Health Rehabilitation Institute
- Ten individuals post SCI had participated in a three-week long study.
- Two H/W interfaces: Microphone & Trackball
- Two control methods: Auto vs. Cartesian (=Manual)
- Outcome measures
  - Time to task completion & Number of Clicks
  - Psychometrics & Baseline characteristics
    - MMSE, FIM, MVPT-R, PIADS, ASIA
- Analysis – Non-parametric
  - Wilcoxon Signed Rank Test
  - Spearman Correlation Coefficient

[REF] Orlando Health study examines technology developed by UCF, OHRI Website, June 16, 2009.

Results
- Lessons Learned

- The developed assist robotic device helps a lot to improve the user's satisfaction in performing the given pick-and-place tasks while great positive responses are observed through psychometric assessment and exit survey.

- There exists a disparity between the quantitative metrics and the psychometrics, For example, while their performance was better with Auto mode, they were more satisfied with the Cartesian mode.

- Baseline characteristics of subjects affect quantitative metrics. (MVPT-R)

- Segmented actions in Cartesian mode enable us to see the difference between Healthy Subjects (GroupH) and SCI Subjects (GroupS).

- Compared with GroupH, cautiousness and improper scene understating of GroupS makes a big difference in their task performance.
  - User's cautiousness degree is under development.

Concluding Remarks
- Summary

- For robust visual servoing tasks, gross-to-fine motion segmentation is used to control a WMRA in unstructured environments.
  - Gross motion brings the end effector close to and properly aligned with an object of interest to get a better resolution.
  - Fine motion moves the end effector to the best pose for grabbing an object with high precision.
- Developed system was tested with more than 25 unique objects in different height/background setup.
- User testing emerged that hybrid human-robot interaction is the most desirable way in the future.

Acknowledgements

- Funded by NSF grant #IIS-0649736

- Special Thanks to
  - UCF-REU program
    - Joshua Treadway, Ryan Lovelett, Marcos Hernandez, Christopher Hamilton, Brandon Gilzean, Simon Echeverry, Rafael Guerra
  - UMASS Lowell
    - Holly Yanco, Katherine Tsui
  - Orlando Health
    - Heather Godfrey, Rodney Olson, David Portee, Amanda Middekke, Tara McNally, Greta Bucks, Robert Melia, Carlos Carrasco

- http://www.eecs.ucf.edu/~abehal/AssistiveRobotics/