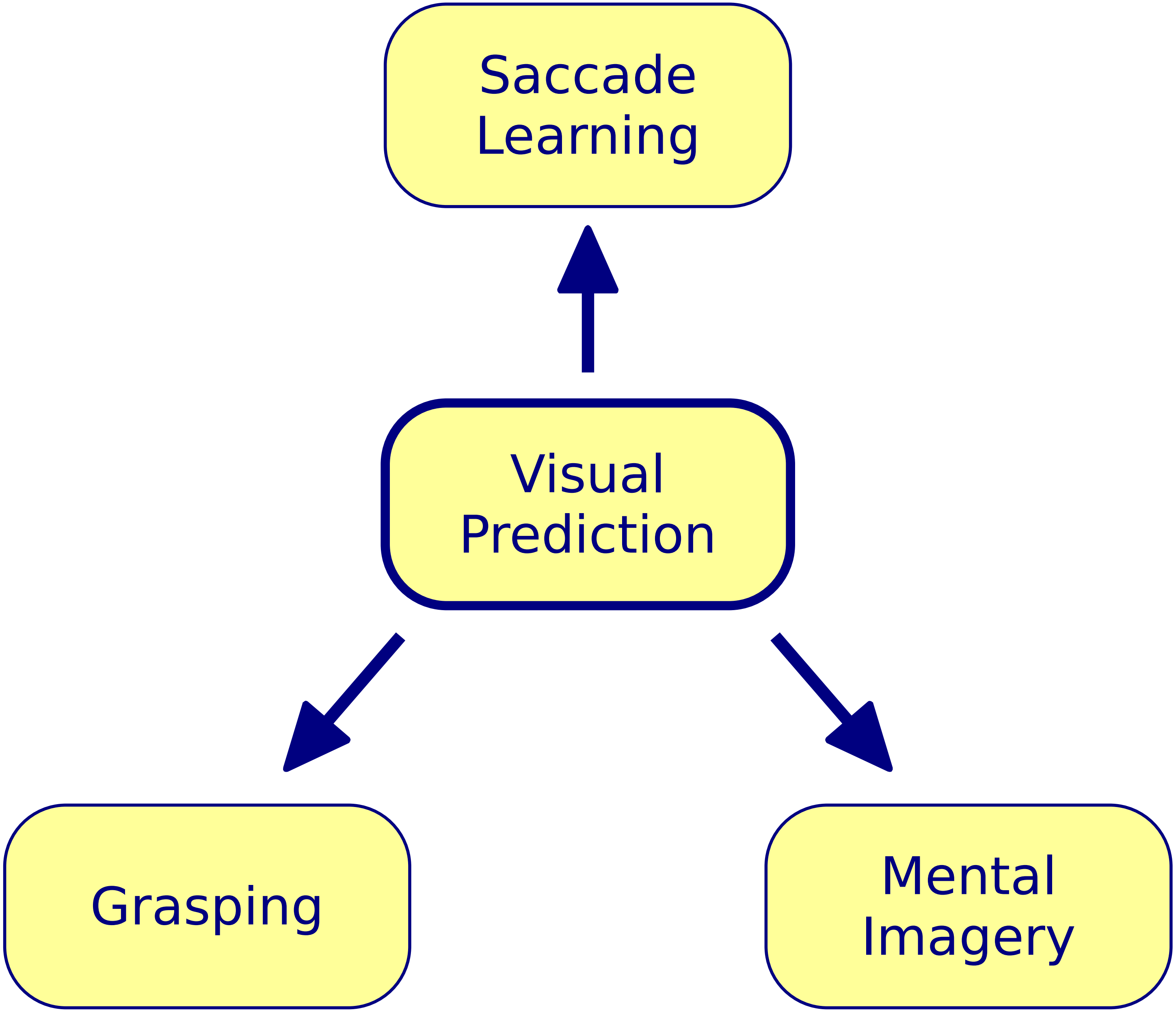
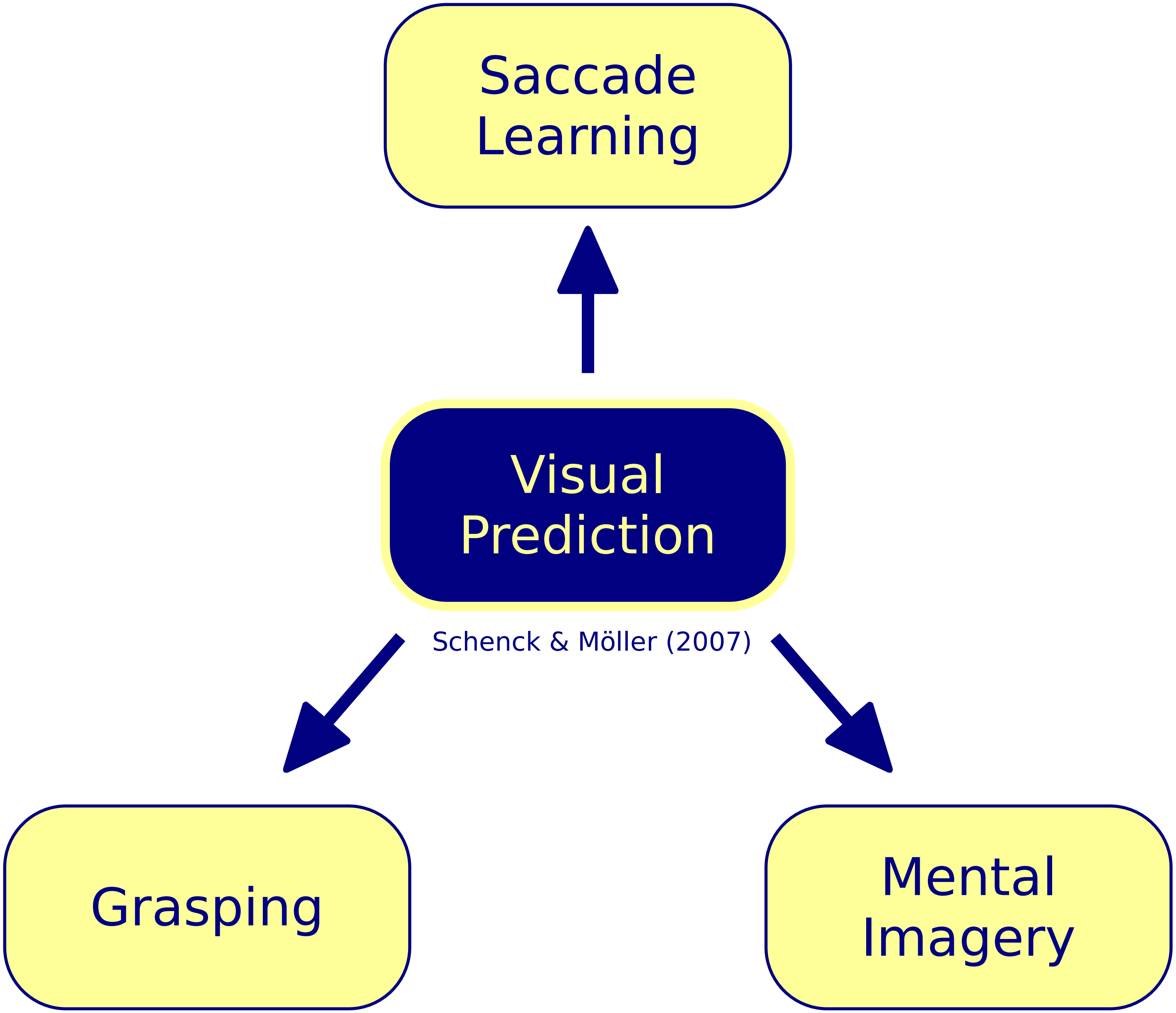
Application Scenarios for Visual Prediction

Wolfram Schenck, Ralf Möller

Computer Engineering Group
Faculty of Technology
Bielefeld University

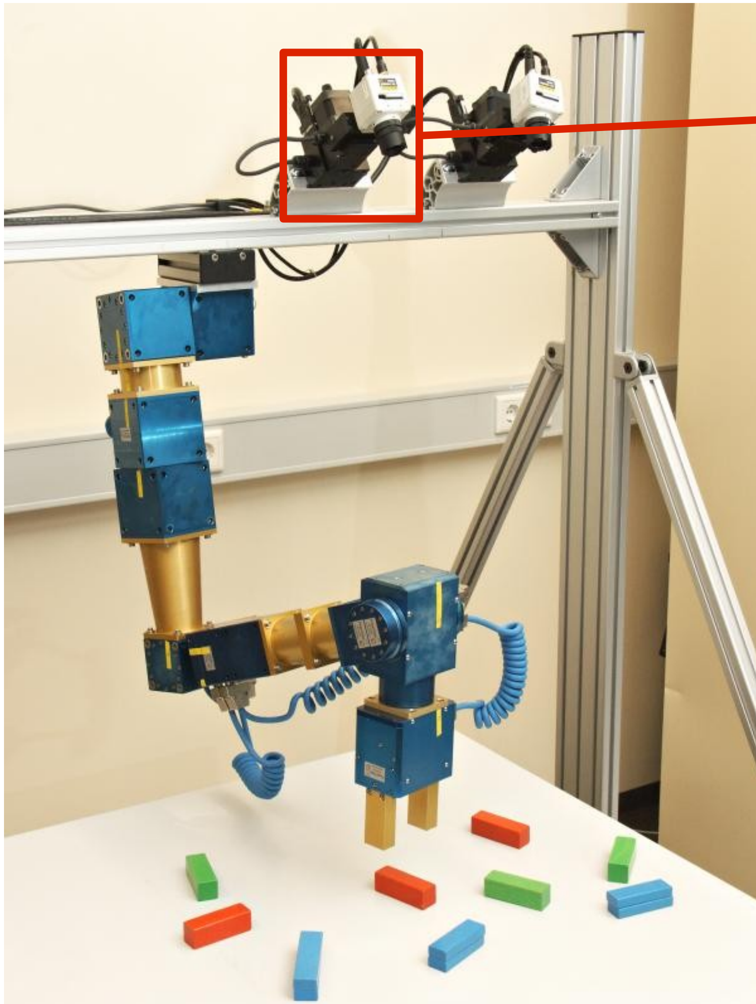
ABiALS 2011, Bielefeld, 22.02.2011


Saccade
Learning
Visual
Prediction
Grasping
Mental
Imagery

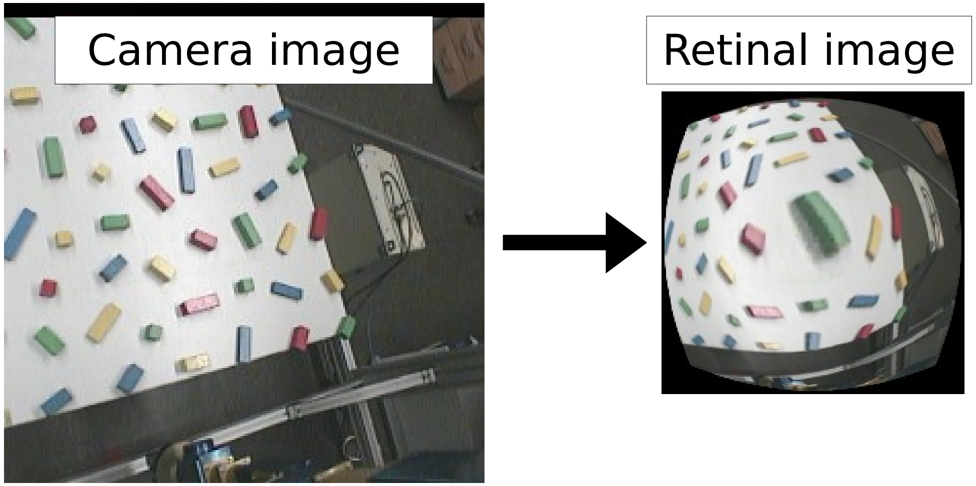


Saccade Learning
Visual Prediction
Schenck & Möller (2007)
Grasping
Mental Imagery

Setup

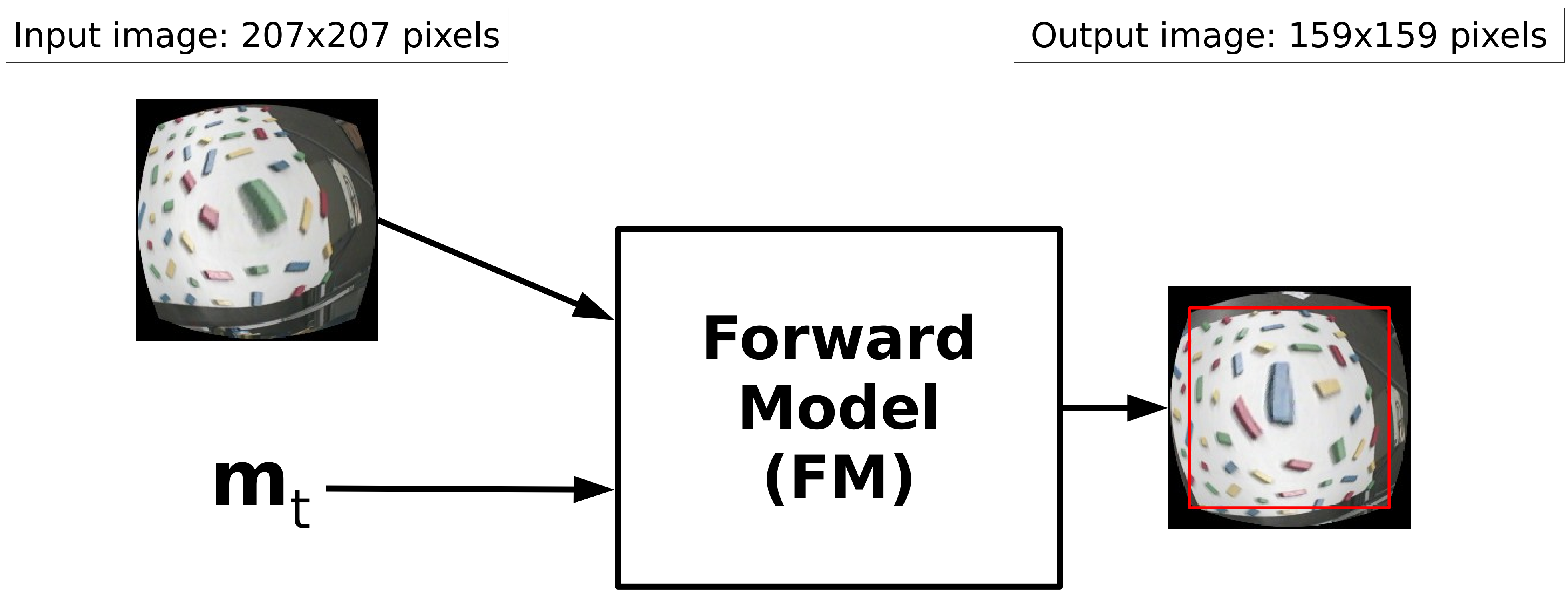


Two degrees of freedom:
- Δpan
- Δtilt



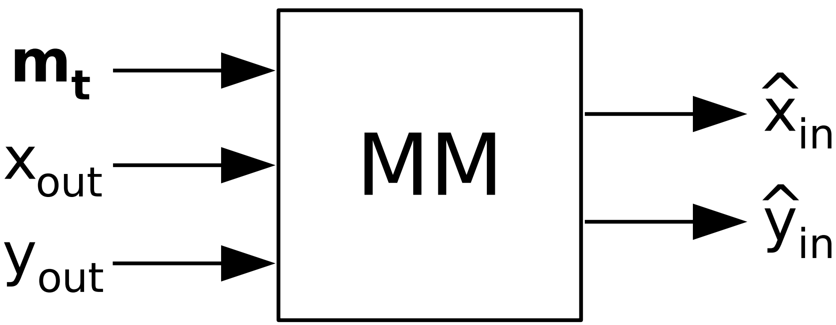
- Power function: $r_C = \lambda r_R^{\gamma} + (1-\lambda) r_R$
  - r_C: Normalized radius in camera image
  - r_R: Normalized radius in retinal image
  - $\lambda = 0.8$, $\gamma = 2.5$

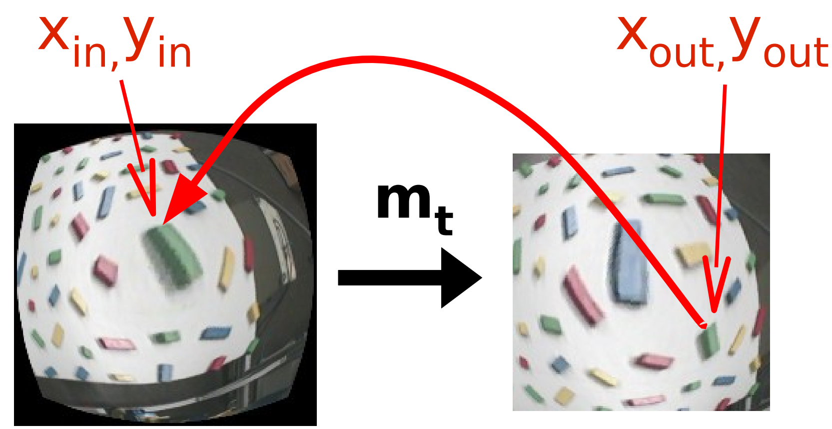
Visual Forward Model for Saccades



$\mathbf{m}_t = (\Delta pan_t, \Delta tilt_t)$

Mapping and Validator Model

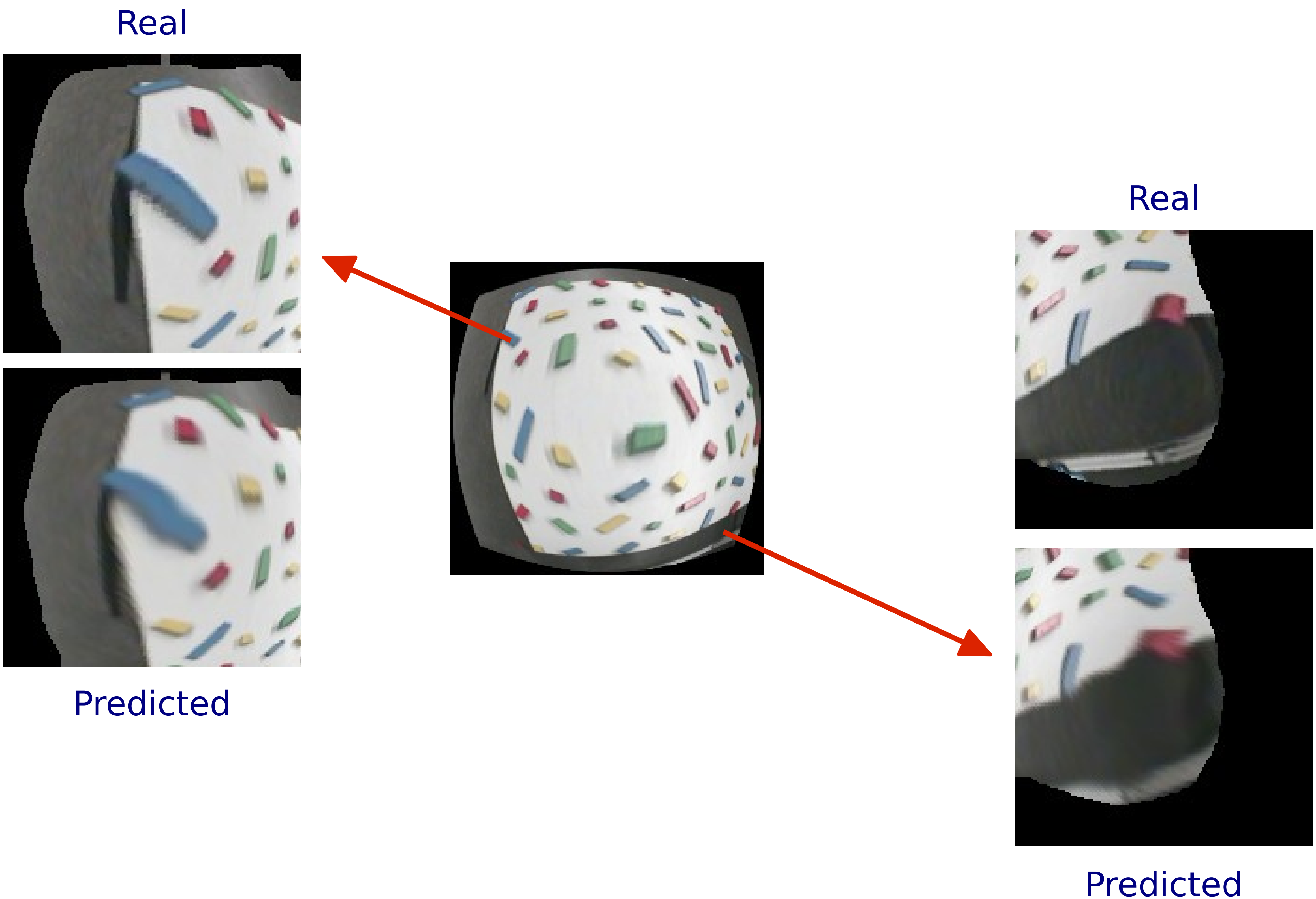


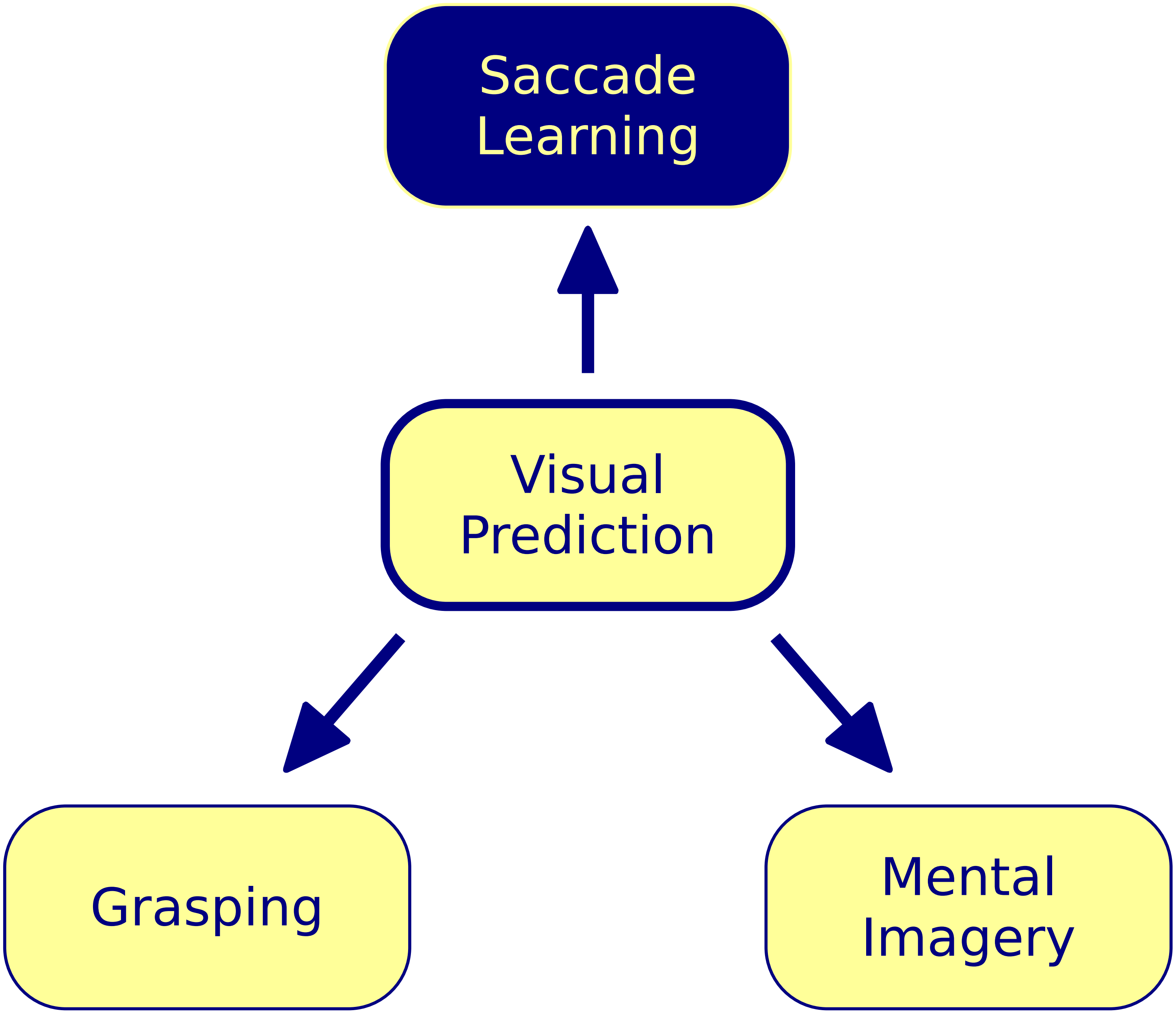


Predictive remapping: Visual prediction in the brain by shifting receptive fields (e.g., Duhamel et al., 1992; Umeno & Goldberg, 1997)

$\mathbf{m}_t = (\Delta pan_t, \Delta tilt_t)$

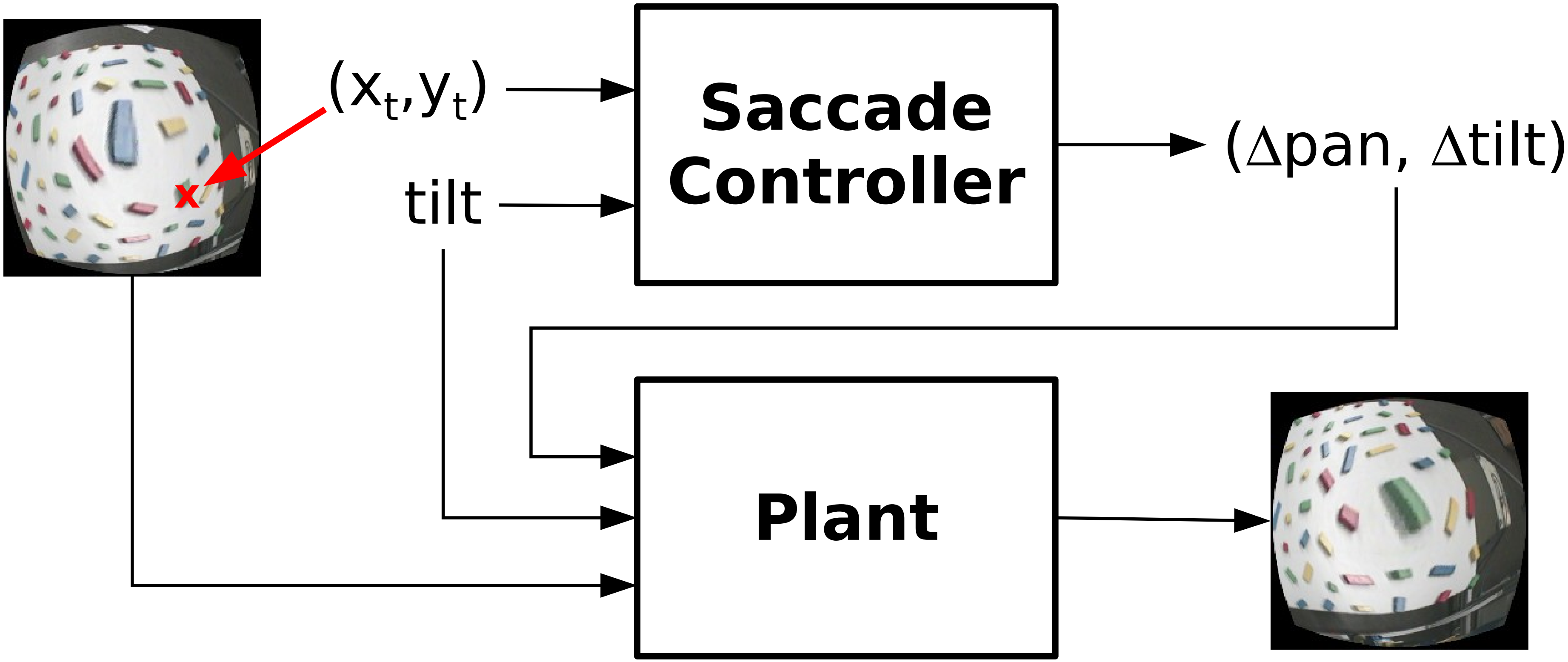
Prediction Examples




Saccade Learning
Visual Prediction
Grasping
Mental Imagery

Saccade Control Task

- **Task:** Schedule commands (Δpan, Δtilt) such that a target object is projected onto the center of the camera image

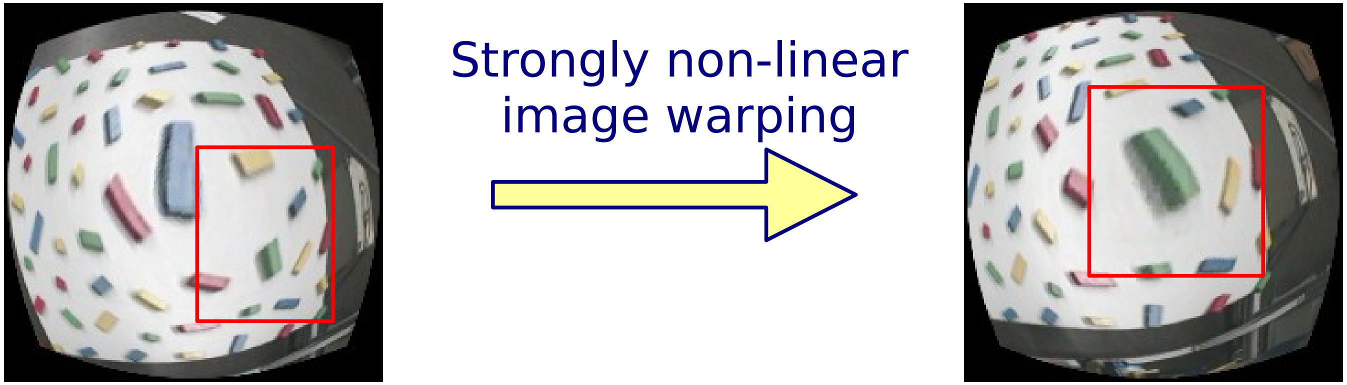


Problem of the Missing Teacher Signal

- **Goal:** Train an adaptive saccade controller
- **Required:** Learning examples associating sensory input and the correct motor output
- **Problem:** The correct motor output (and thus the motor error $\Delta\mathbf{m}$) is unknown...
- **Solution:** “Direct Inverse Modeling” (DIM) (e.g., Kuperstein, 1990)
- **But:** DIM requires $\mathbf{s}_t$ and $\mathbf{s}_{t+1}$

Target Re-Identification Problem

- Sensory state for the saccade learning task: target position $\mathbf{s}_t = (x_t, y_t)$
- To determine $\mathbf{s}_{t+1}$ after saccade excution:
  - Target re-identification necessary
  - In past studies with plain camera images: By correlation approach
- Problem: Correlation approach not applicable to retinal images



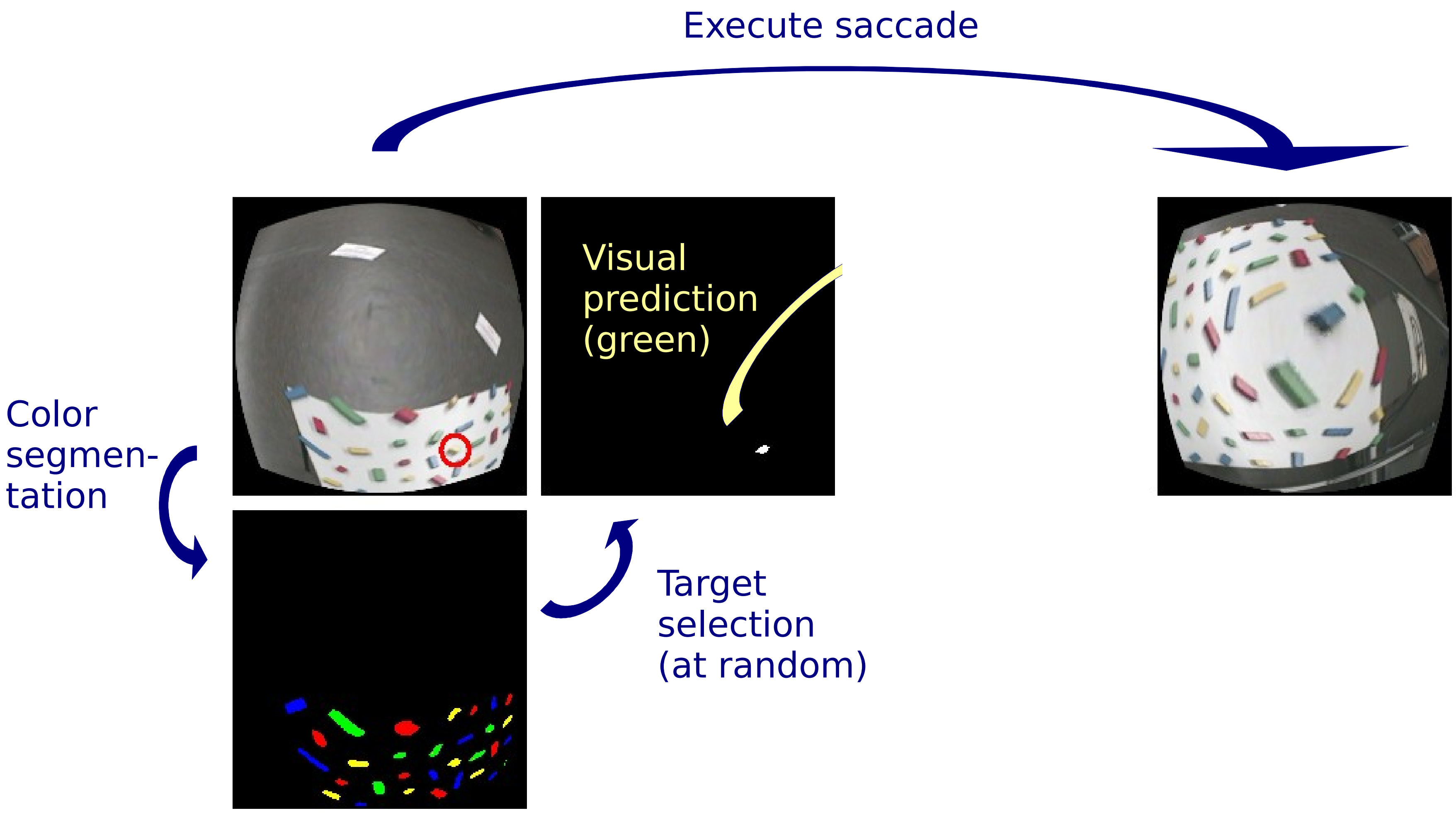
Visual Prediction for Target Re-Ident.

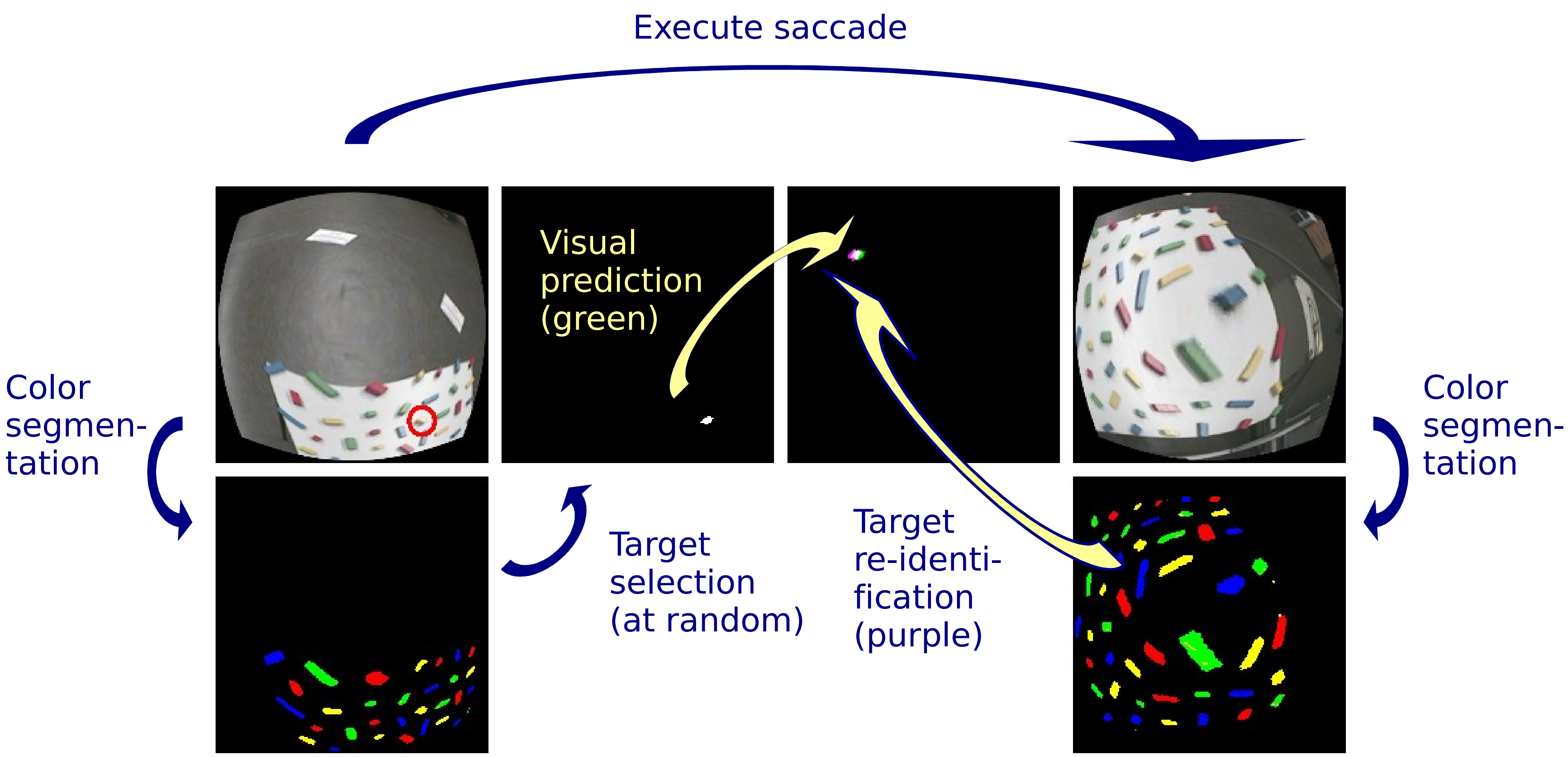
- **Solution:**
  - Predict location of target object after saccade by visual FM
  - Search for real target in close vicinity of the predicted location
    - Use other distinct features like color or shape for final re-identification

Basic Image Processing Steps

Color segmen-tation

Basic Image Processing Steps



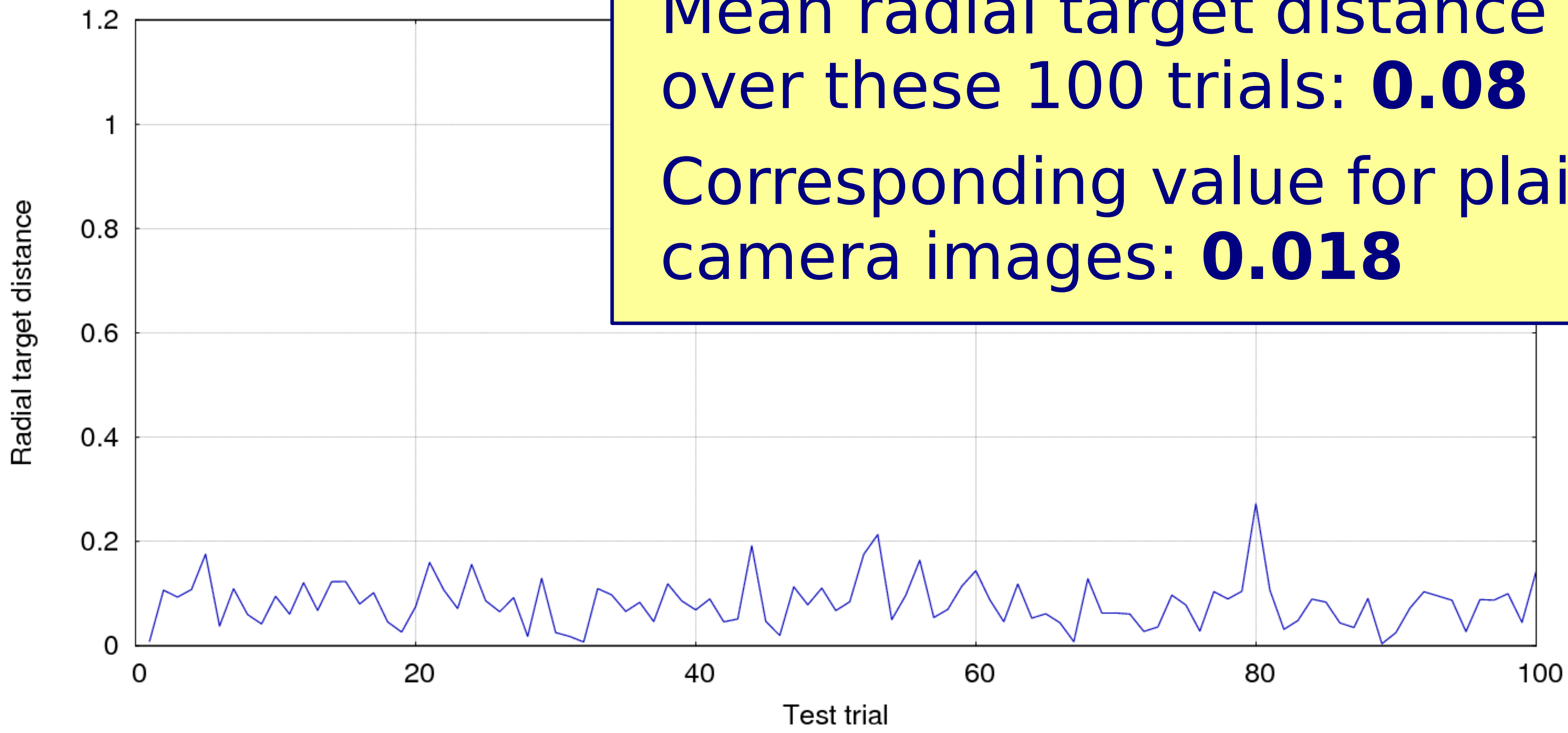

Execute saccade
Visual prediction (green)
Color segmen-tation
Color segmen-tation
Target selection (at random)
Target re-identi-fication (purple)

Experiments

- Controller network:
  - Multi-Layer Perceptron (MLP)
  - Weight adaptation by online back-propagation
  - Training set: 6000 randomly collected learning examples
- Indicator for saccade precision:
  Radial target distance
  - Distance between center of mass of target object and image center
    (normalized to range [0;1] along the horizontal/vertical direction)

Training Results

- Controller training over 3,000 epochs:
  $\Rightarrow$ Test error down from 0.09 to 0.0002
- Controller network performance over 100 trials:

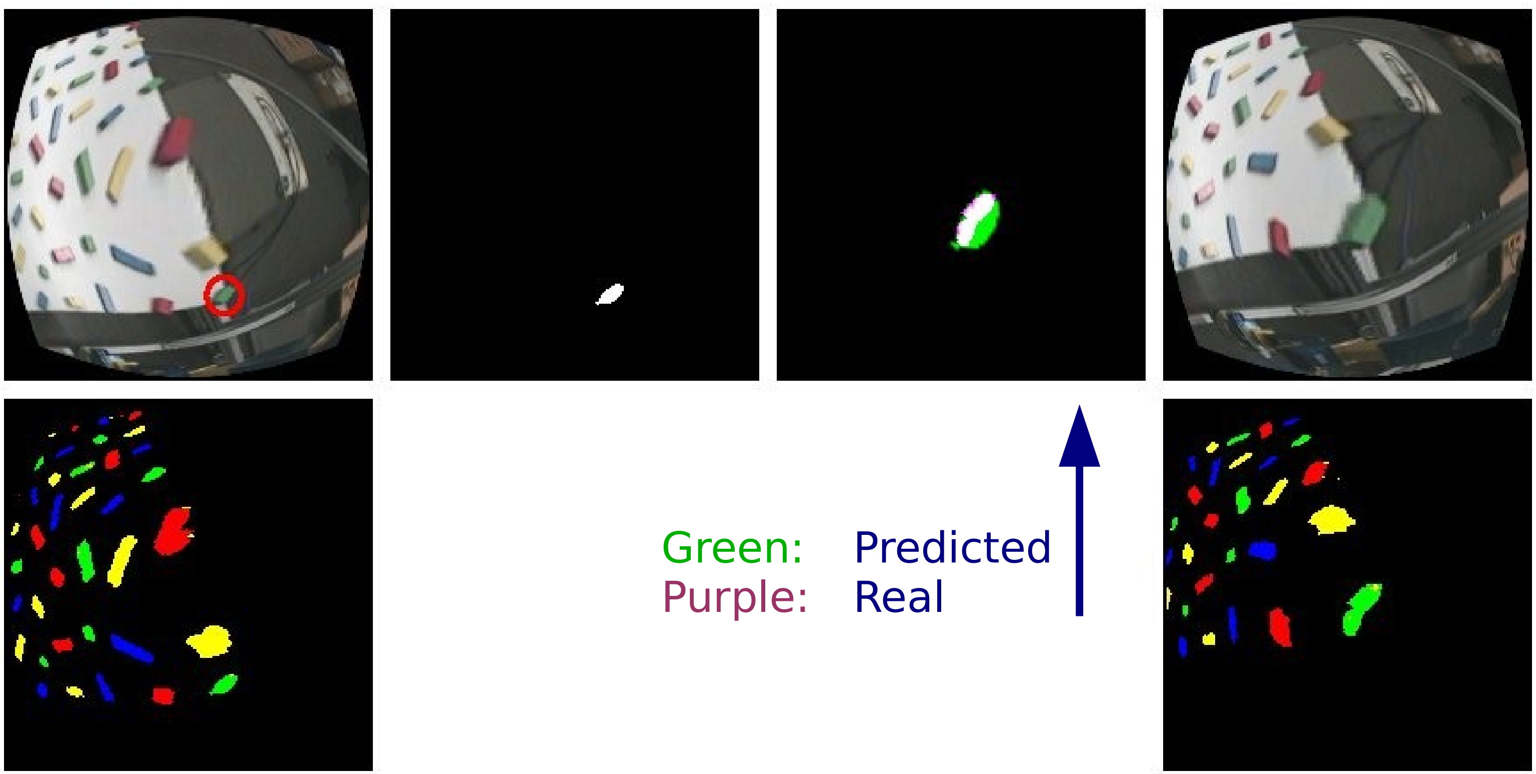


Mean radial target distance over these 100 trials: **0.08**

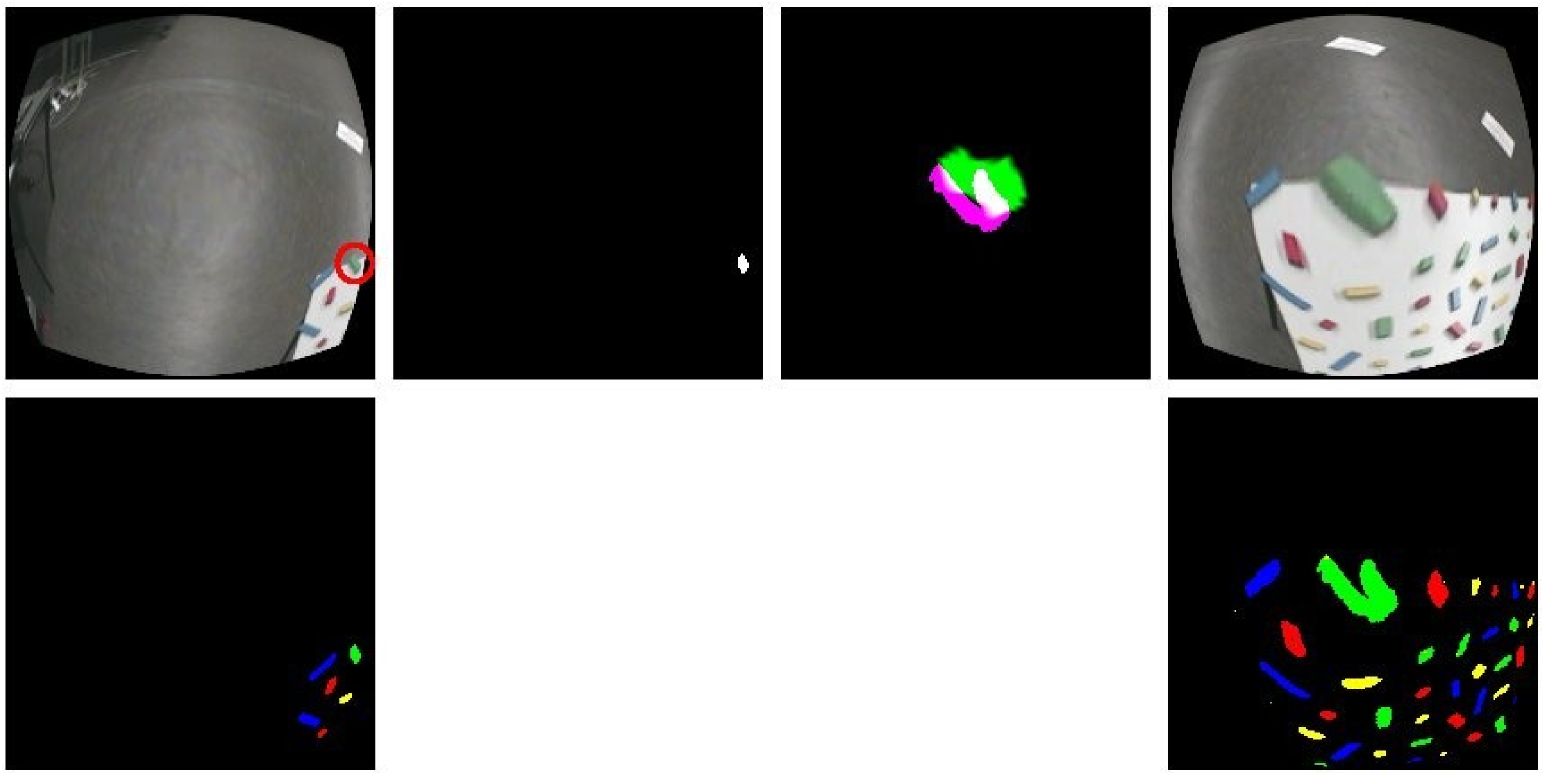
Corresponding value for plain camera images: **0.018**

Results for DIM: Examples

Pre-saccadic **Post-saccadic**



Pre-saccadic **Post-saccadic**

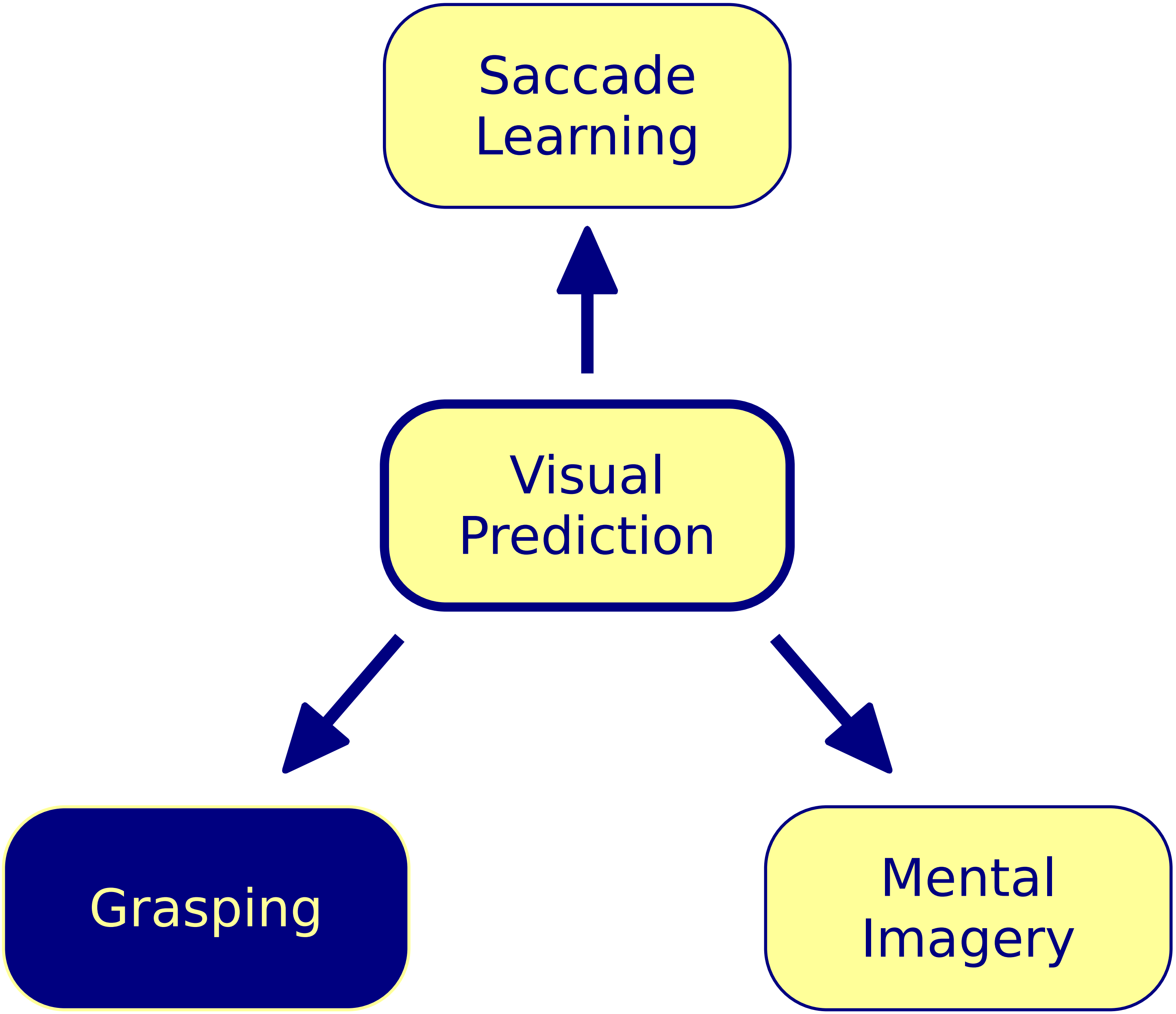


Intermediate Discussion

- Successful saccade learning on retinal images
- Problem of the missing teacher signal solved by direct inverse modeling
- Problem of target re-identification solved by visual prediction

Link to “spatial representations and dynamic interactions”:

Visual prediction enables the agent to shift freely between different eye-centered frames of reference (“different” with respect to the gaze direction)

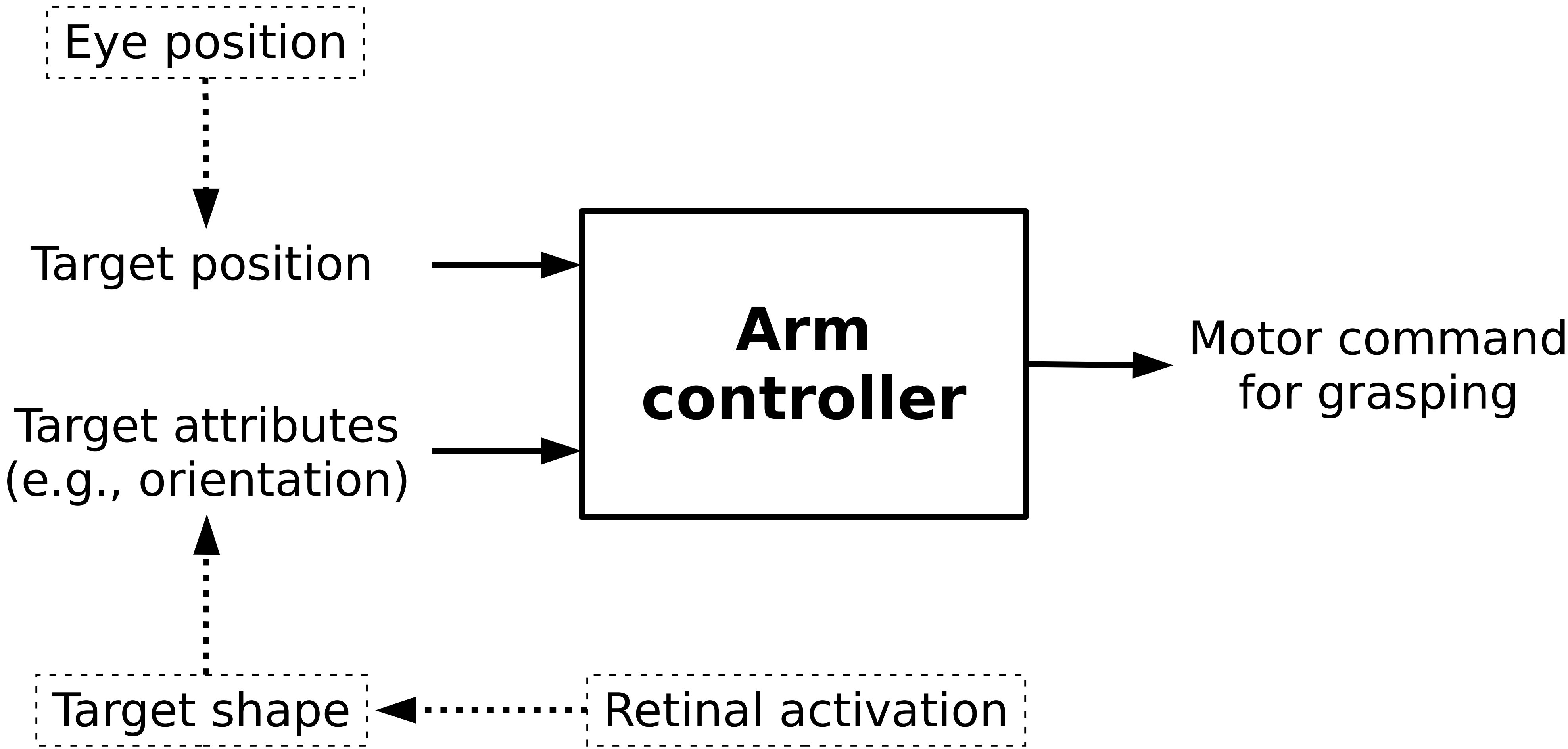


Schenck, Hoffmann, & Möller (2009)

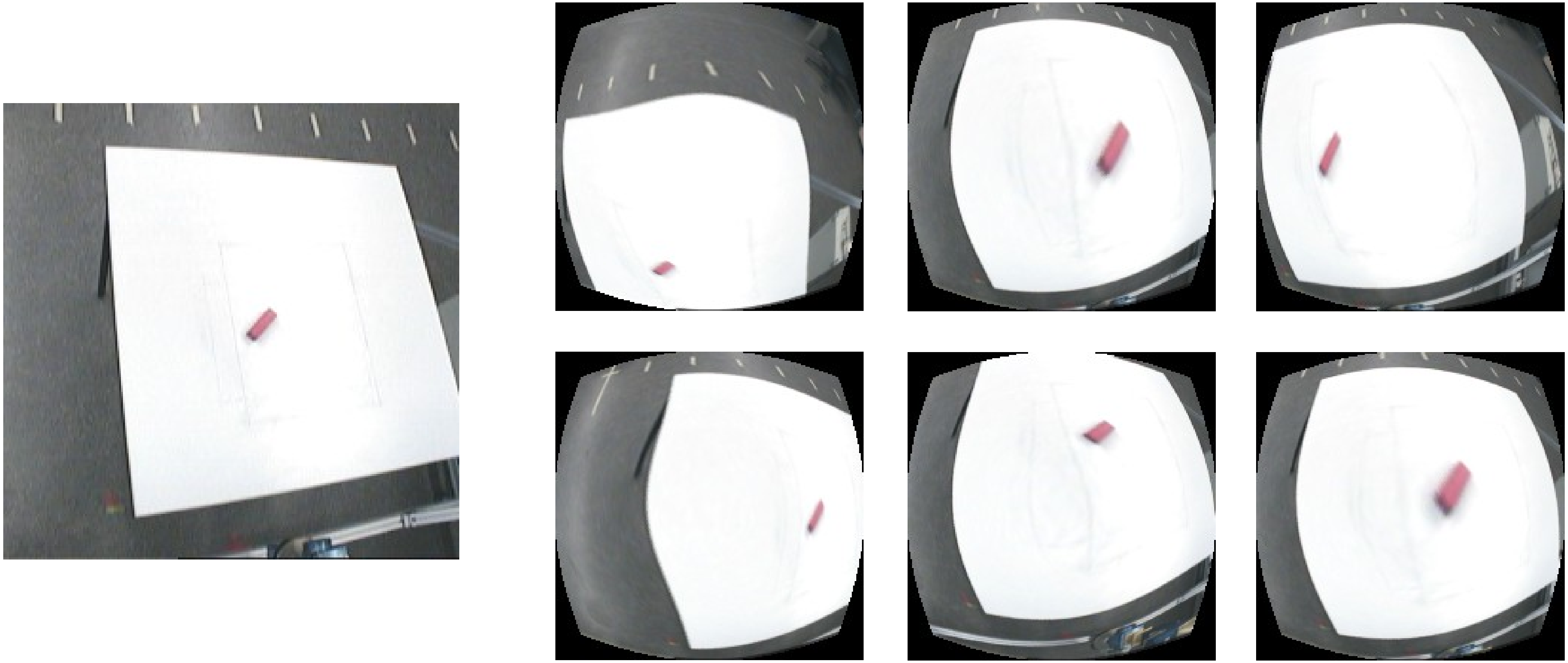
Grasping to Extrafoveal Targets

- Usually: Saccades precede arm movements for reaching and grasping
  - Target objects are projected onto the fovea
- Arm movements towards extrafoveal target objects are possible, but with less precision (e.g., Vercher et al., 1994)

Arm Control Scheme

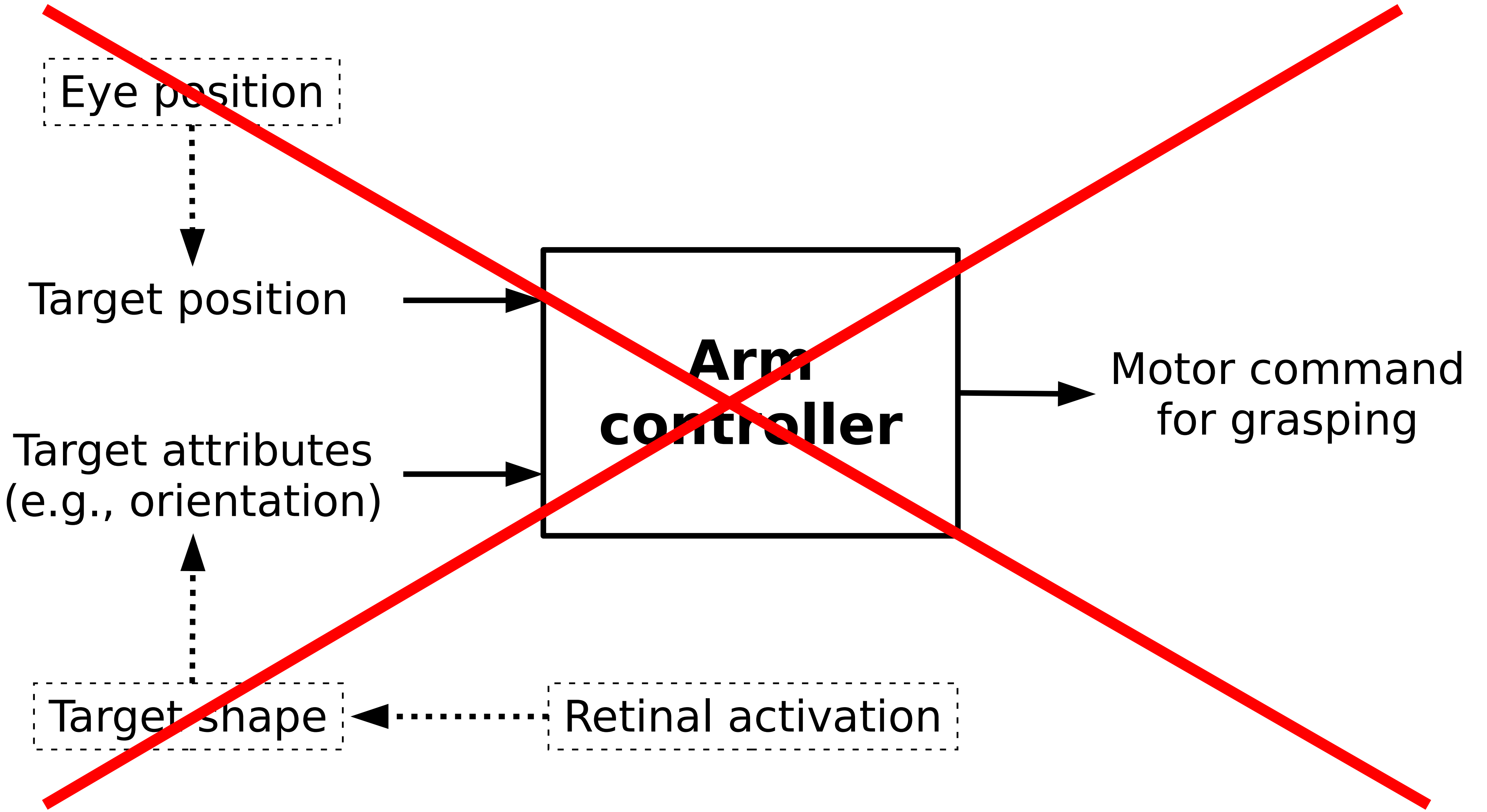


Retinal Variance



➔ Depending on retinal position, sensor activation differs considerably!

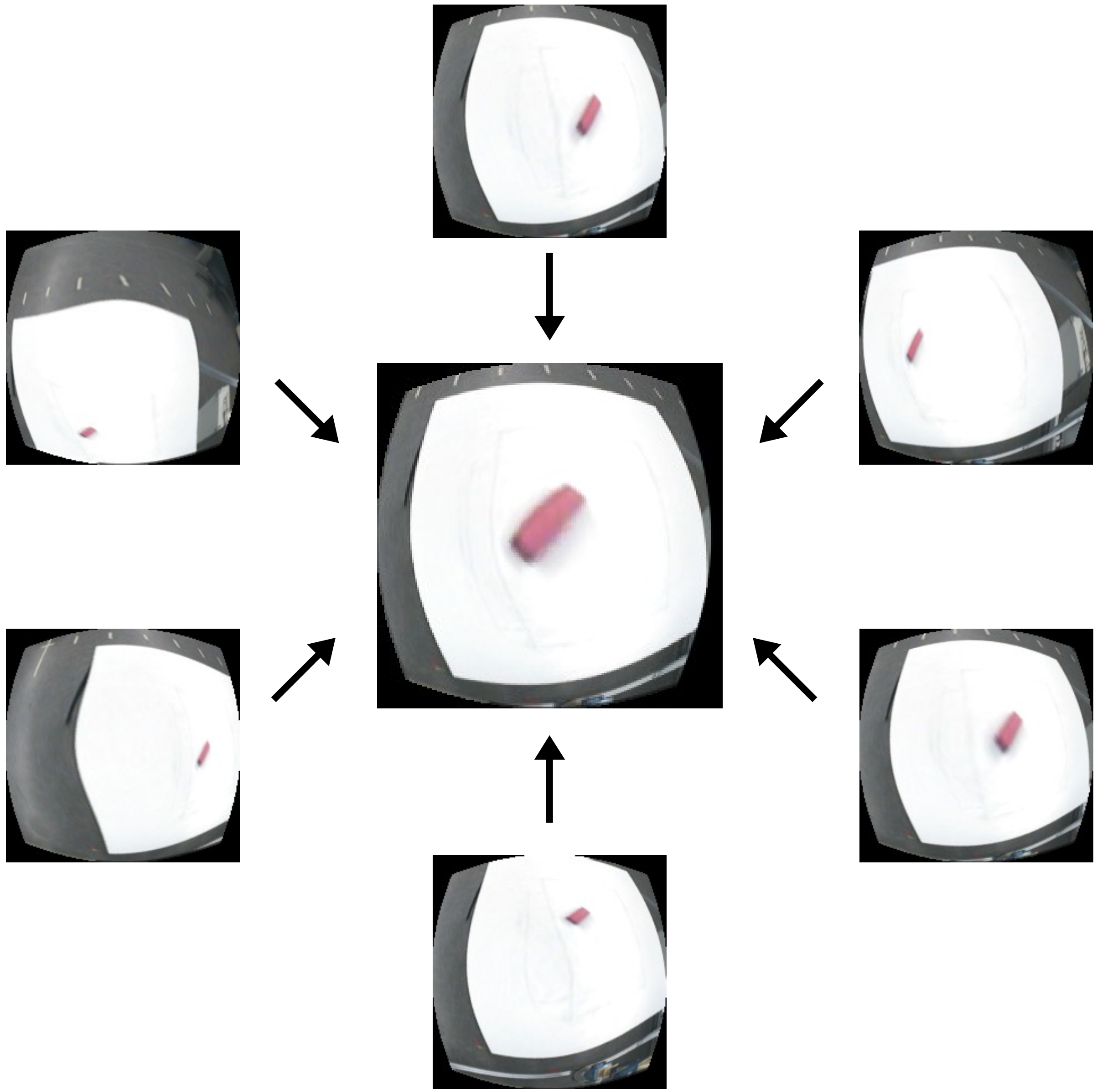
Arm Control Scheme



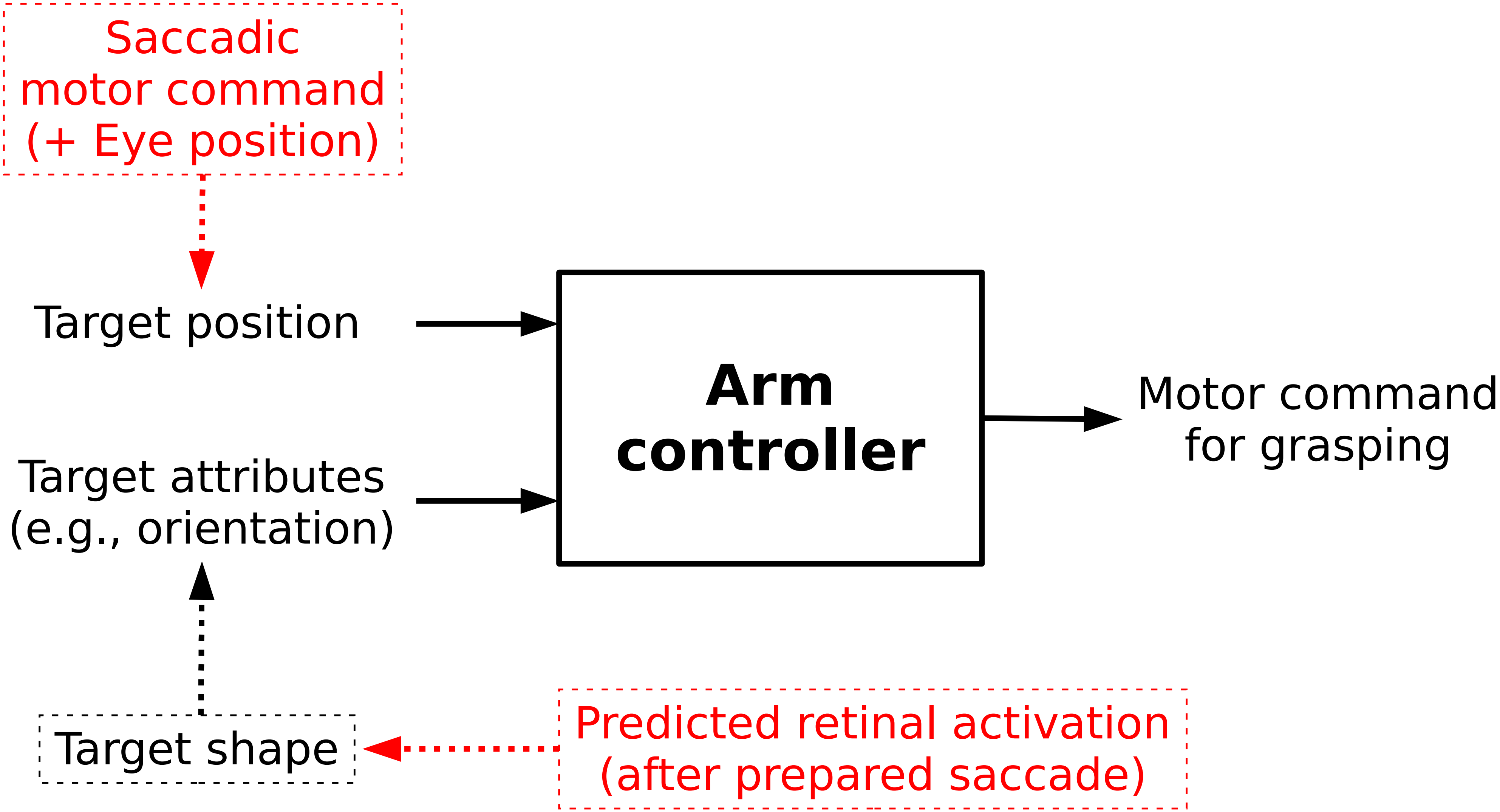
Premotor Theory

- Spatial attention: Consequence of *preparation* of goal-directed movements (Rizzolatti et al, 1994)
  - Attention shifts are accompanied by the preparation of eye movements
- Additional hypothesis:
  - The preparation of eye movements triggers a prediction of the retinal images after the movement

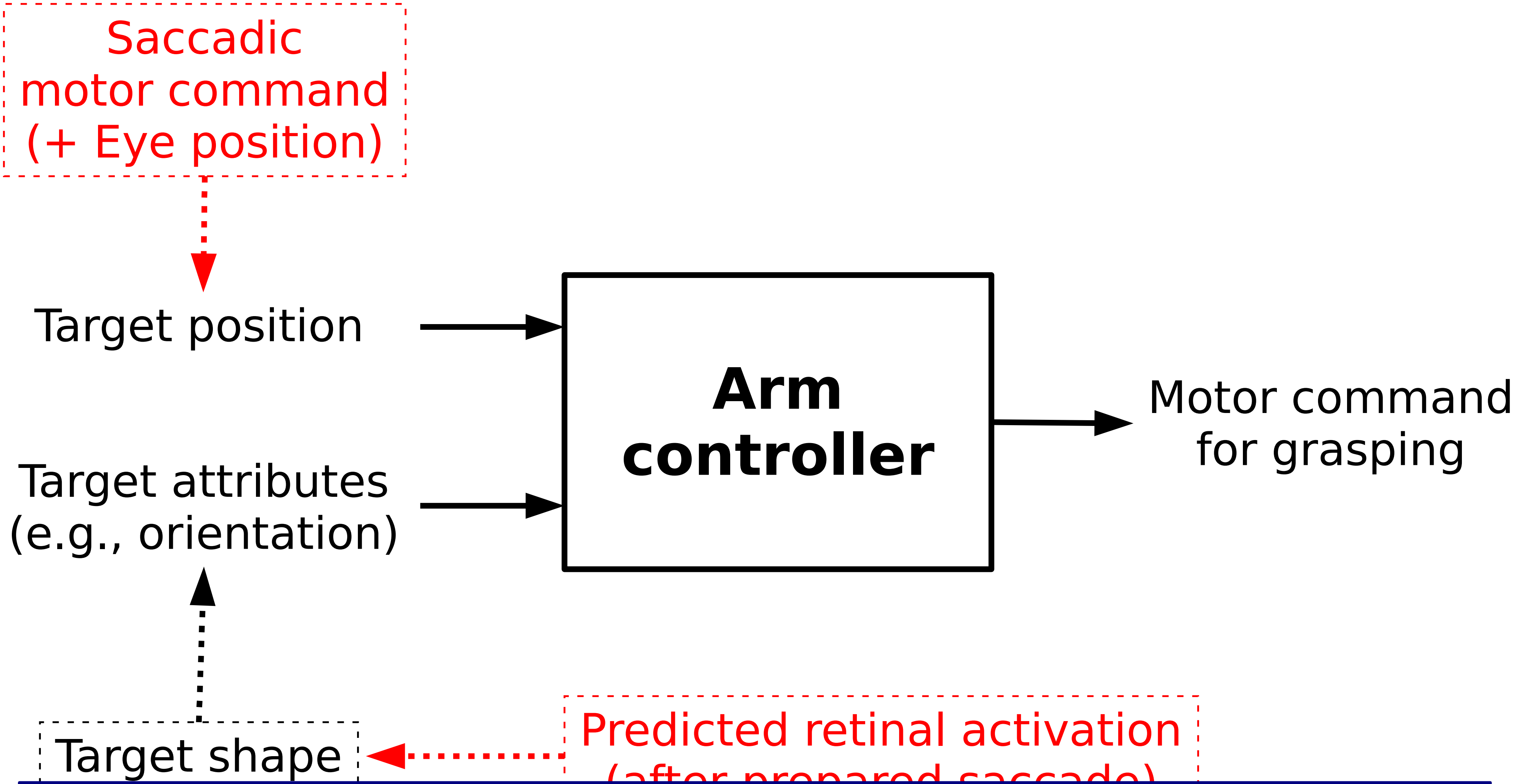
Retinal Invariance Through Visual Pred.



Revised Arm Control Scheme



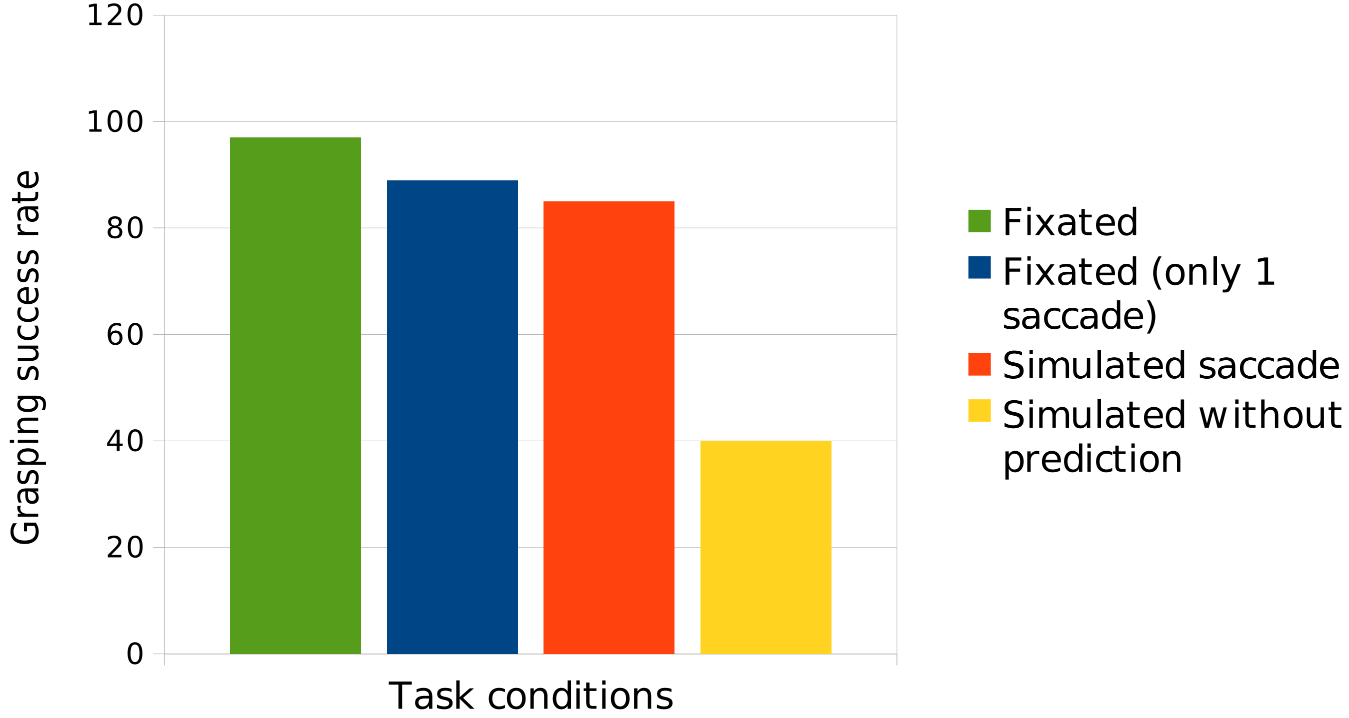
Revised Arm Control Scheme



➔ Single arm controller for both fixated and non-fixated target objects!

Experimental Results

- Grasping performance over 100 trials in four different experimental conditions

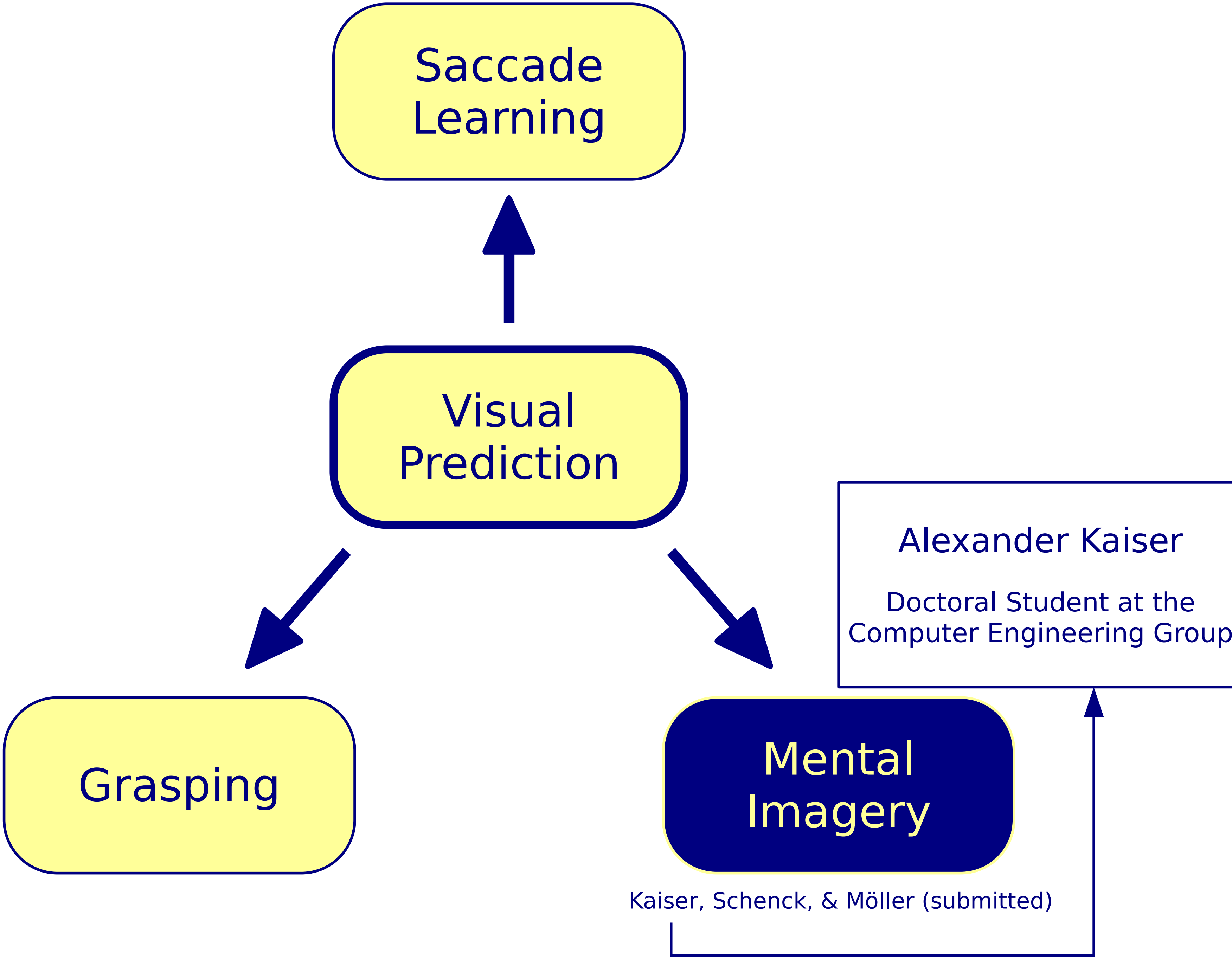


Intermediate Discussion

- Model successful in grasping to extrafoveal targets
  - Real-world robotic test
- Visual prediction might be an important component of visuomotor coordination in this task domain

Link to “spatial representations and dynamic interactions”:

Visual prediction (in combination with a saccade controller) enables (kind of) invariance of the retinal representation with respect to the gaze direction

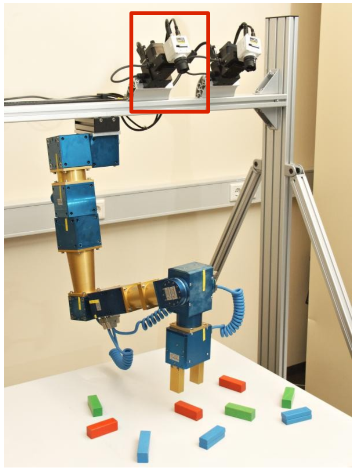

Saccade Learning
Visual Prediction
Grasping
Mental Imagery
Alexander Kaiser
Doctoral Student at the Computer Engineering Group
Kaiser, Schenck, & Möller (submitted)

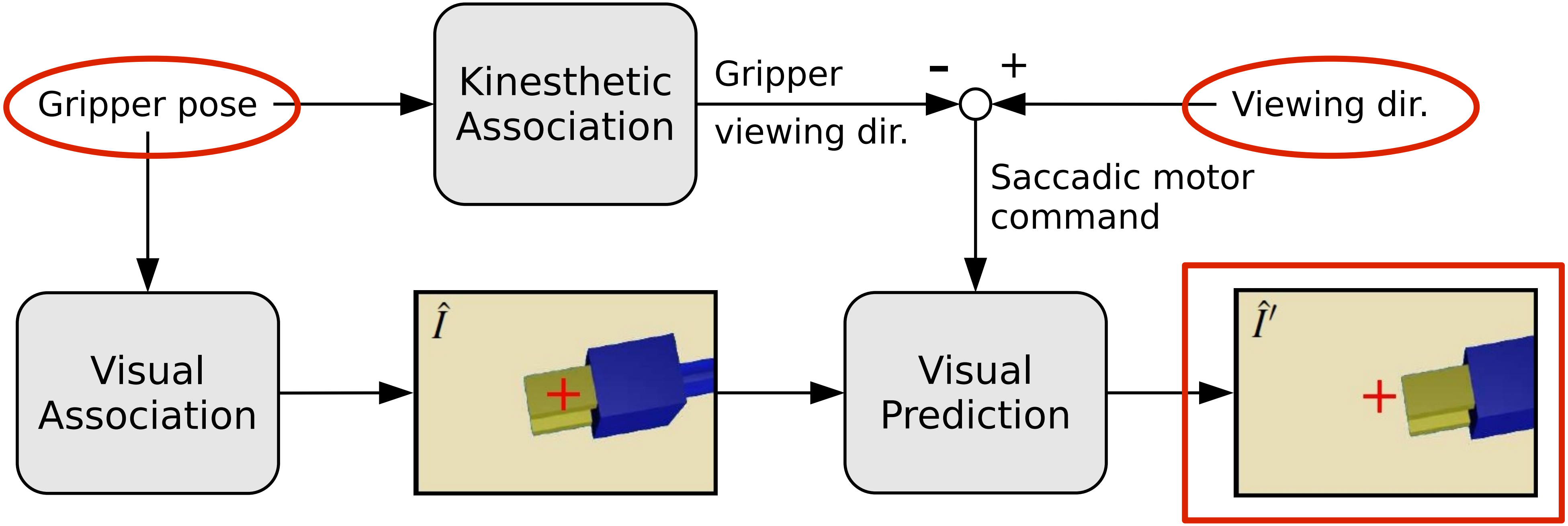
Mental Imagery for Motor Simulation

- Simulation theories: Simulation of action sequences and their accompanying sensory or system states as basis for perception and cognition (e.g., Hesslow, 2002; Möller, 1999; Ziemke at al., 2005)
- Covert (simulated) sensory and motor states: “Mental images”
- **Here:** Mental images of visual gripper states (to be used in forthcoming cognitive architectures)

Overall Model

- Input: Gripper pose, viewing direction
- Output: Image of the gripper
  (simulated "mental" image)





Sorry, further slides...

- … on this topic removed since this work has been recently submitted for publication (decision pending)

Intermediate Discussion

- Model successful in generating “mental images” of visual gripper states
- Holistic approach: The image is generated as a whole in its original “raw” format, not on the basis of single features
- Need for visual prediction results from the decomposition of the overall problem in a visual association and a prediction part

Link to “spatial representations and dynamic interactions”:

Visual prediction allows to generalize from a purely foveal representation (target fixated) to other gaze directions

Summary and Open Questions

Summary: Three Studies

- Visual prediction in the context of saccade learning
  - Purpose: Facilitates target re-identification
- Visual prediction in the context of grasping
  - Purpose: Transforms extrafoveal visual representations of target objects into foveal representations
- Visual prediction in the context of mental imagery
  - Purpose: Facilitates the generation of “mental images” by simplifying the visual association task

Open Questions

- Algorithms to learn visual prediction for motor tasks with many degrees of freedom?
- Visual prediction for movements where depth information is relevant?
- Multi-step visual prediction – stability, precision?

Thank you for your attention!