

Motivation

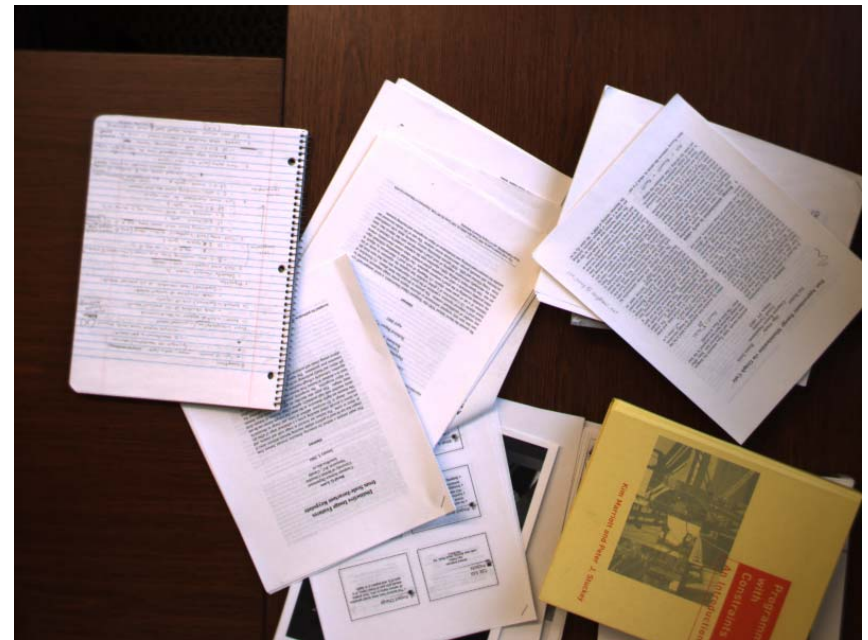


Where is my W-2 Form?

Video-based Tracking

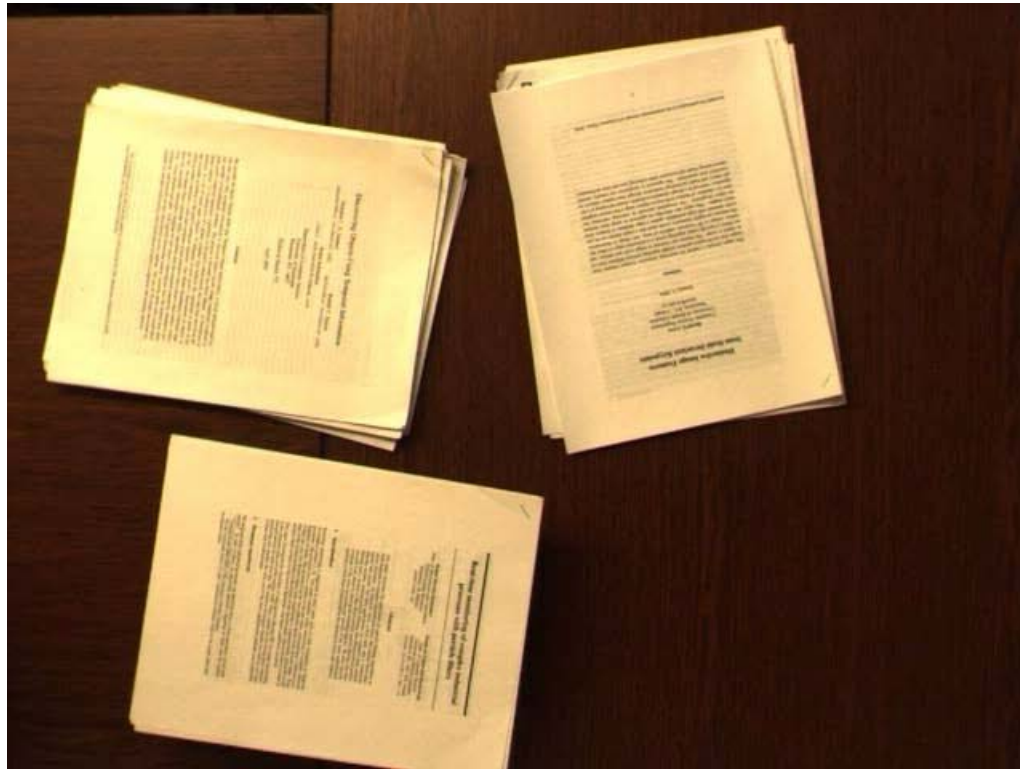


Overhead video camera



Camera view of the desk

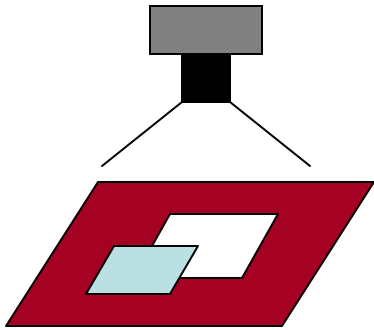
Example



40 minutes, 1024x768 @ 15 fps

System Overview

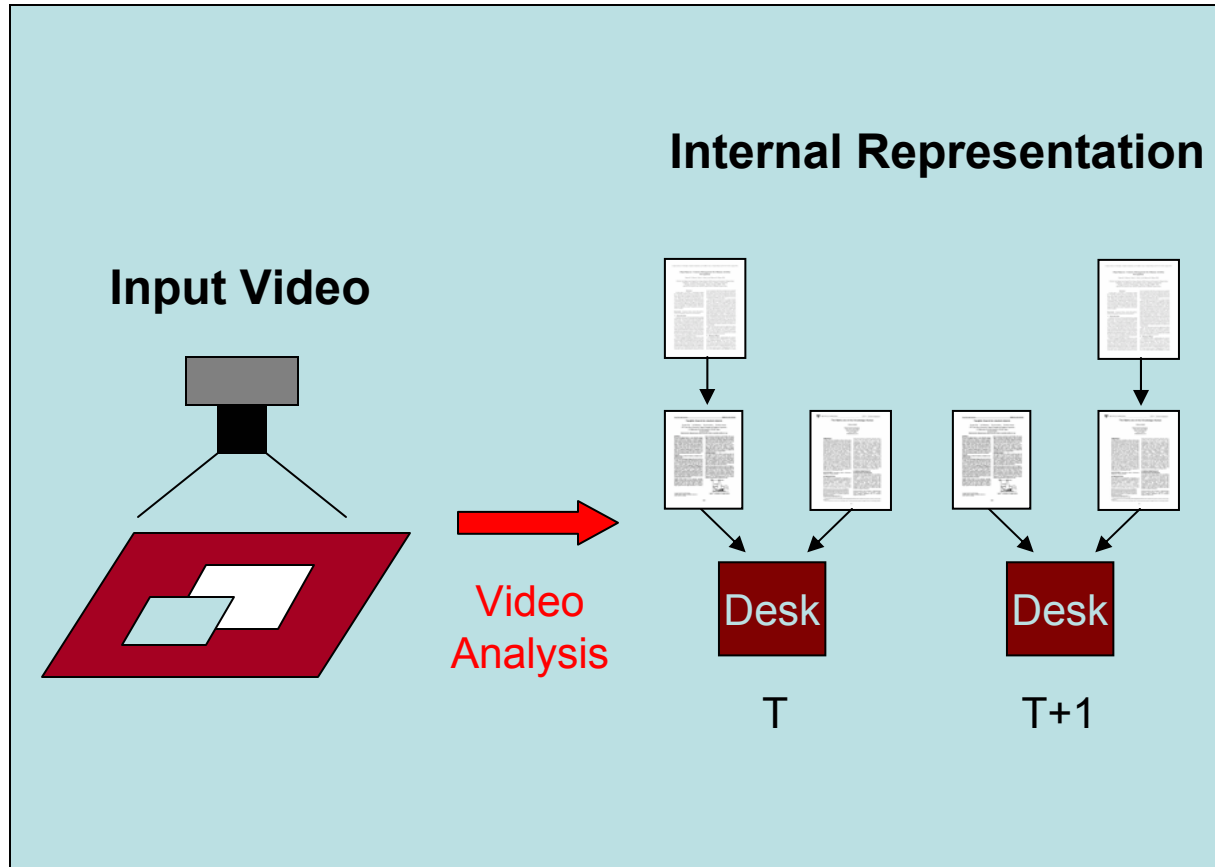
Input Video



User



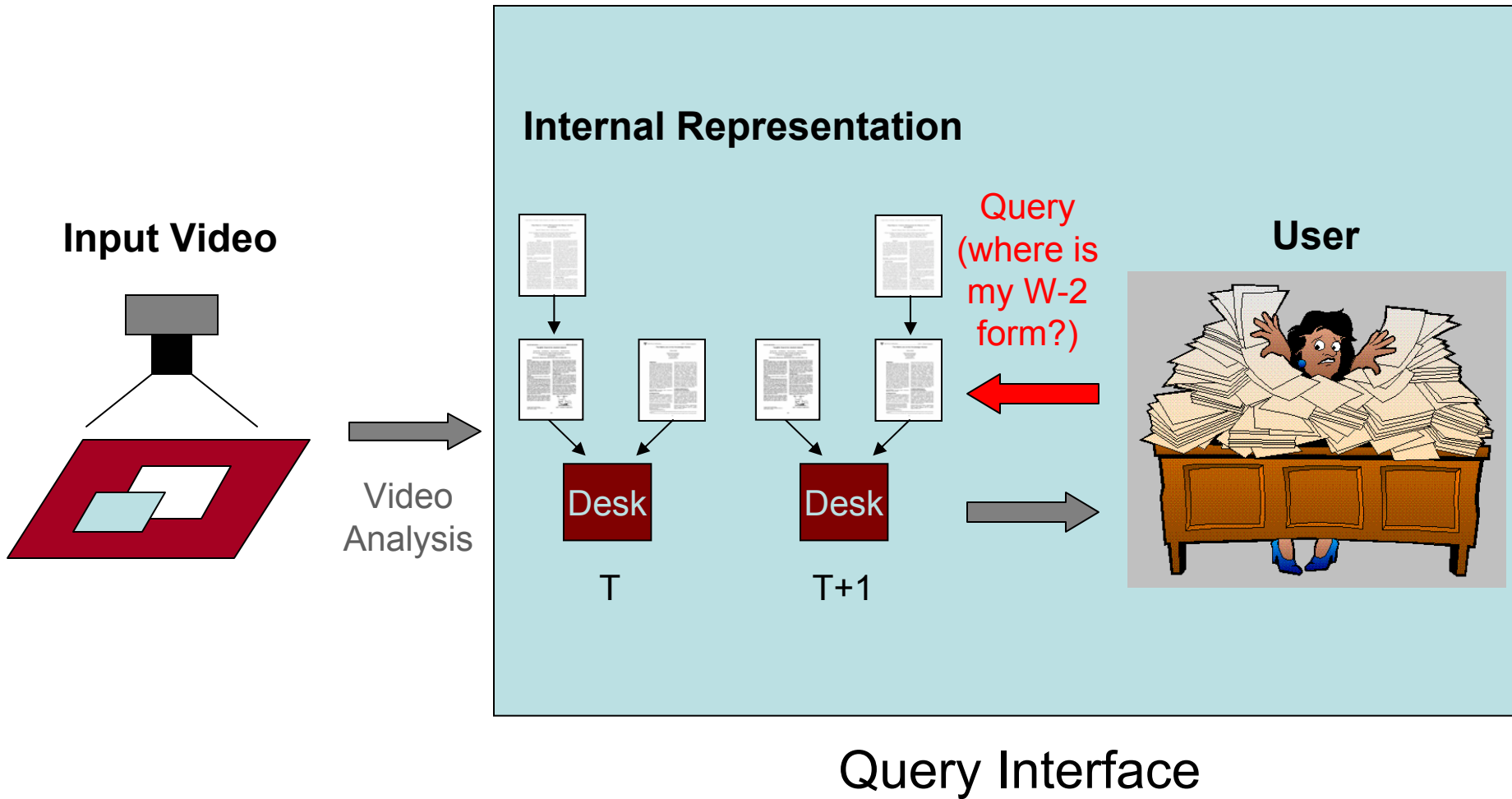
System Overview



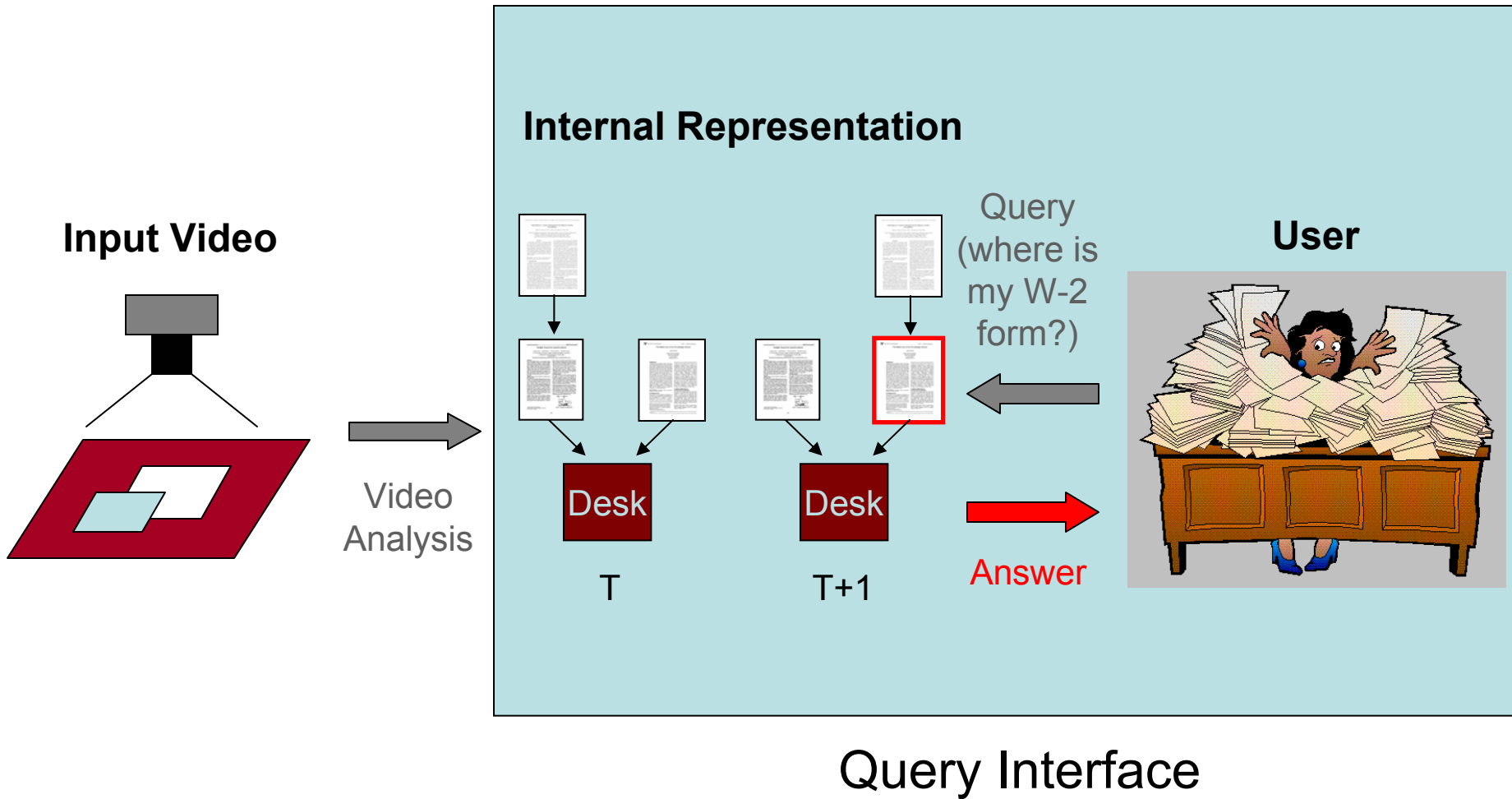
Vision Engine



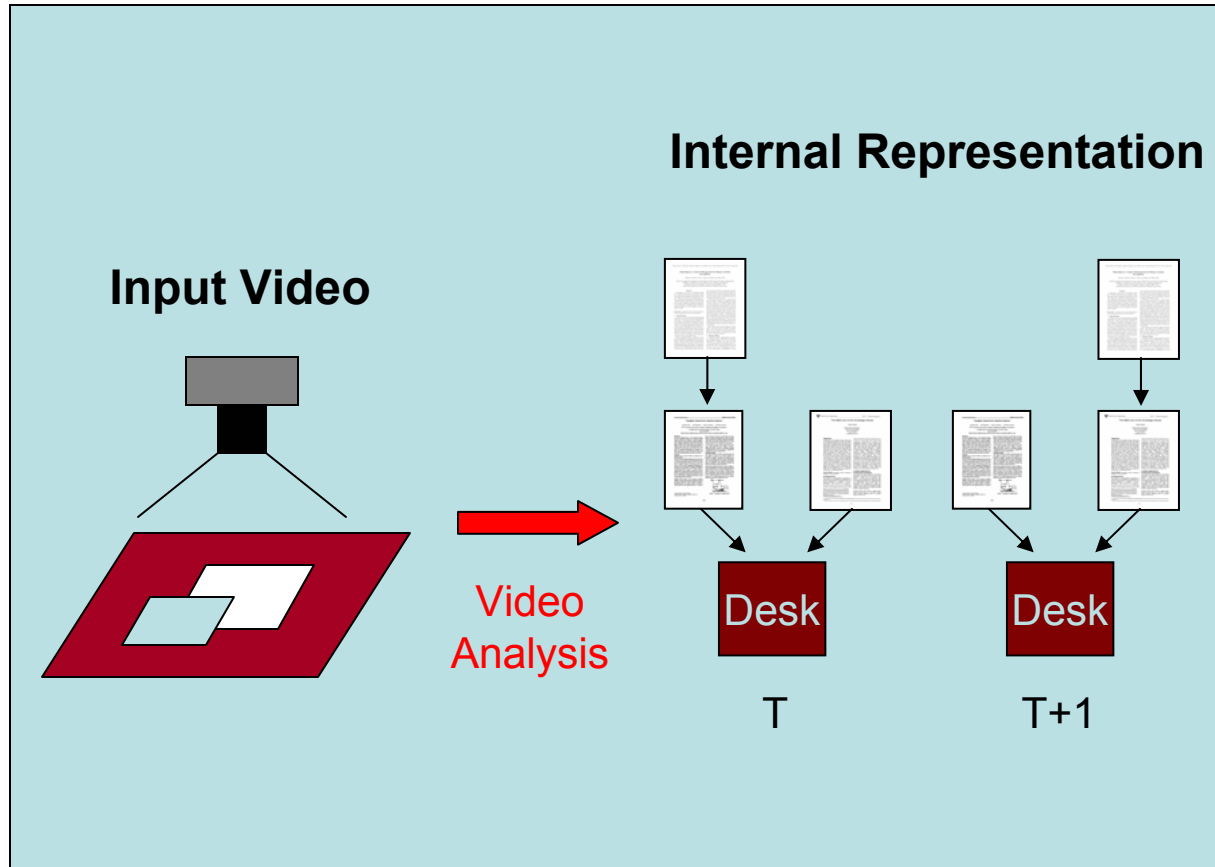
System Overview



System Overview



System Overview

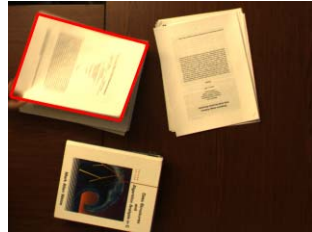
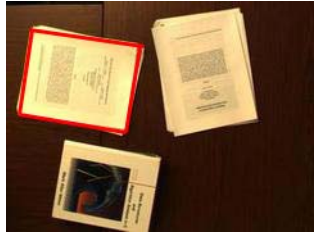


Vision Engine



User

Vision Problem



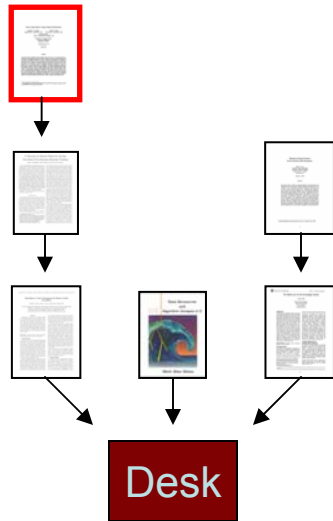
Vision Problem

Event



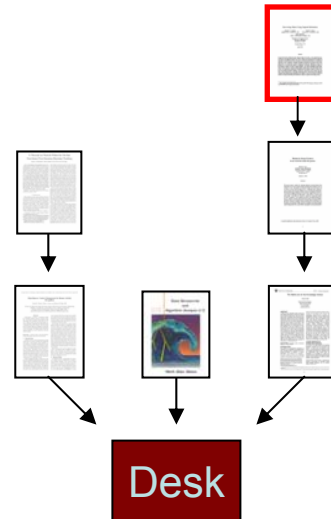
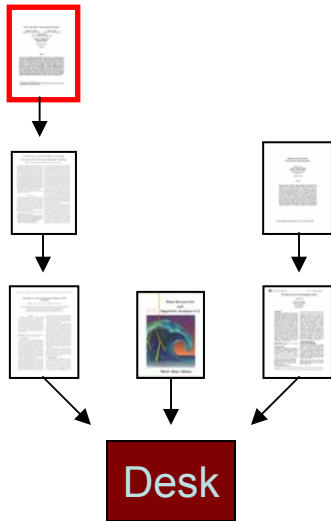
Vision Problem

Event



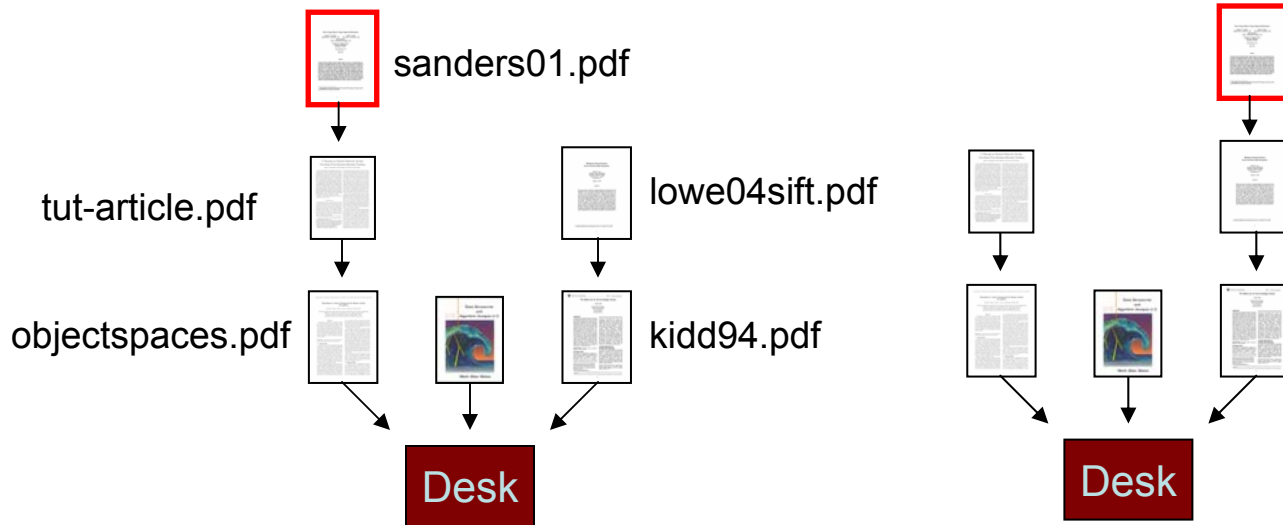
Vision Problem

Event



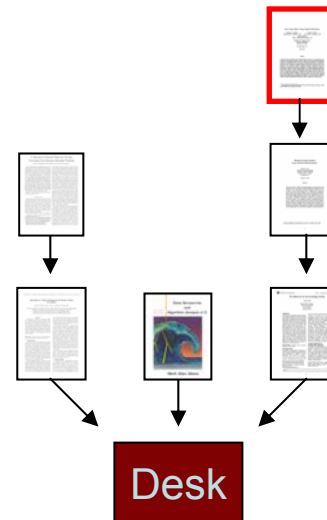
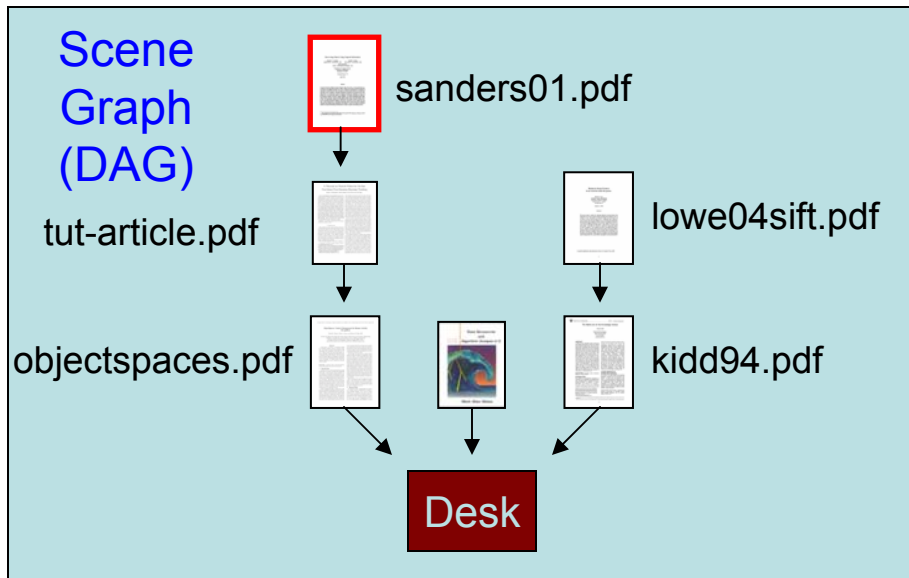
Vision Problem

Event



Vision Problem

Event



Assumptions

- Simplifying
 - Corresponding electronic copy exists

Assumptions

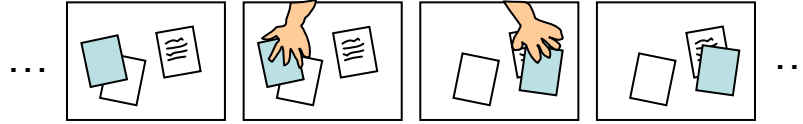
- Simplifying
 - Corresponding electronic copy exists
 - 3 event types: move/entry/exit
 - One document at a time
 - Only topmost document can move
 - No duplicate copies of same document

Assumptions

- Simplifying
 - Corresponding electronic copy exists
 - 3 event types: move/entry/exit
 - One document at a time
 - Only topmost document can move
 - No duplicate copies of same document
- Other
 - Desk need not be initially empty

Algorithm Overview

Input
Frames



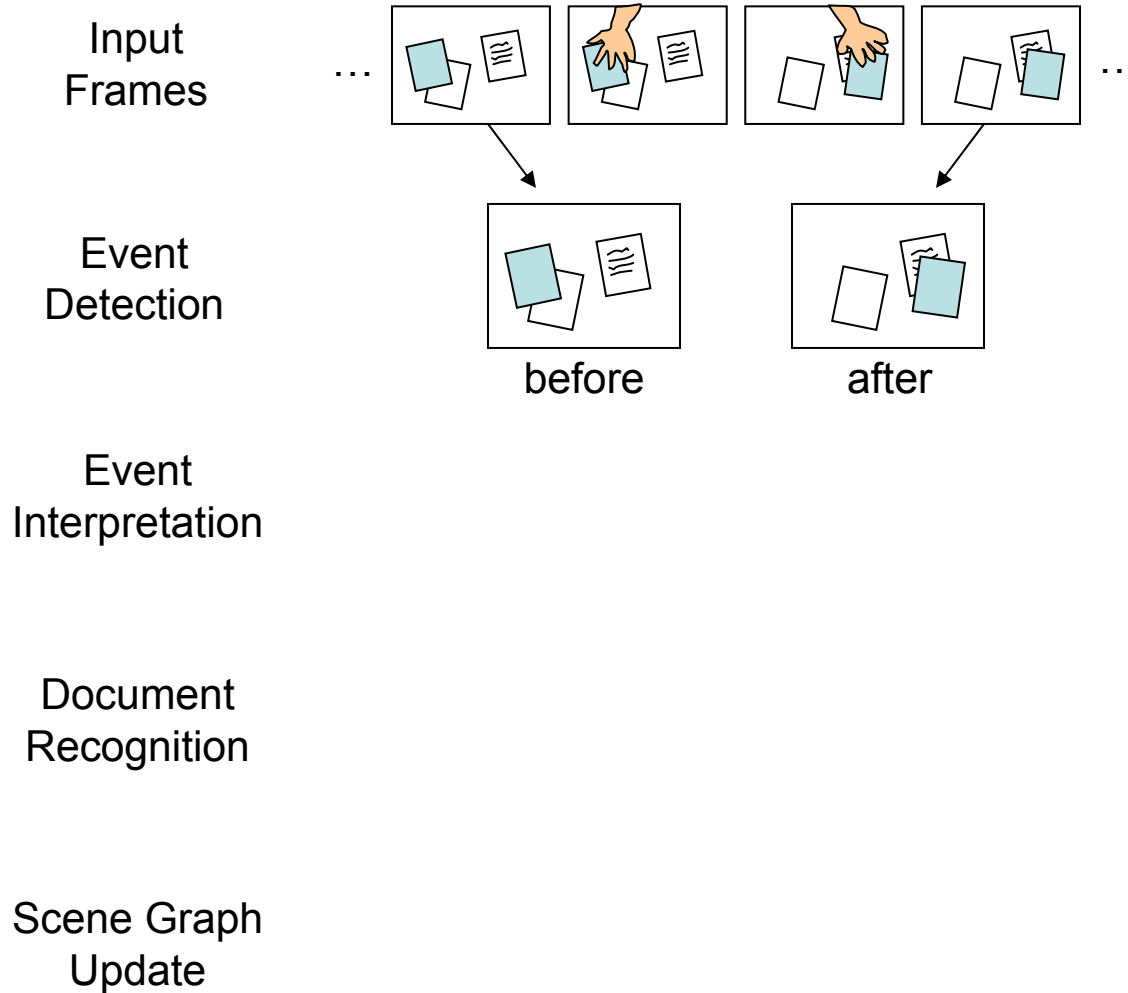
Event
Detection

Event
Interpretation

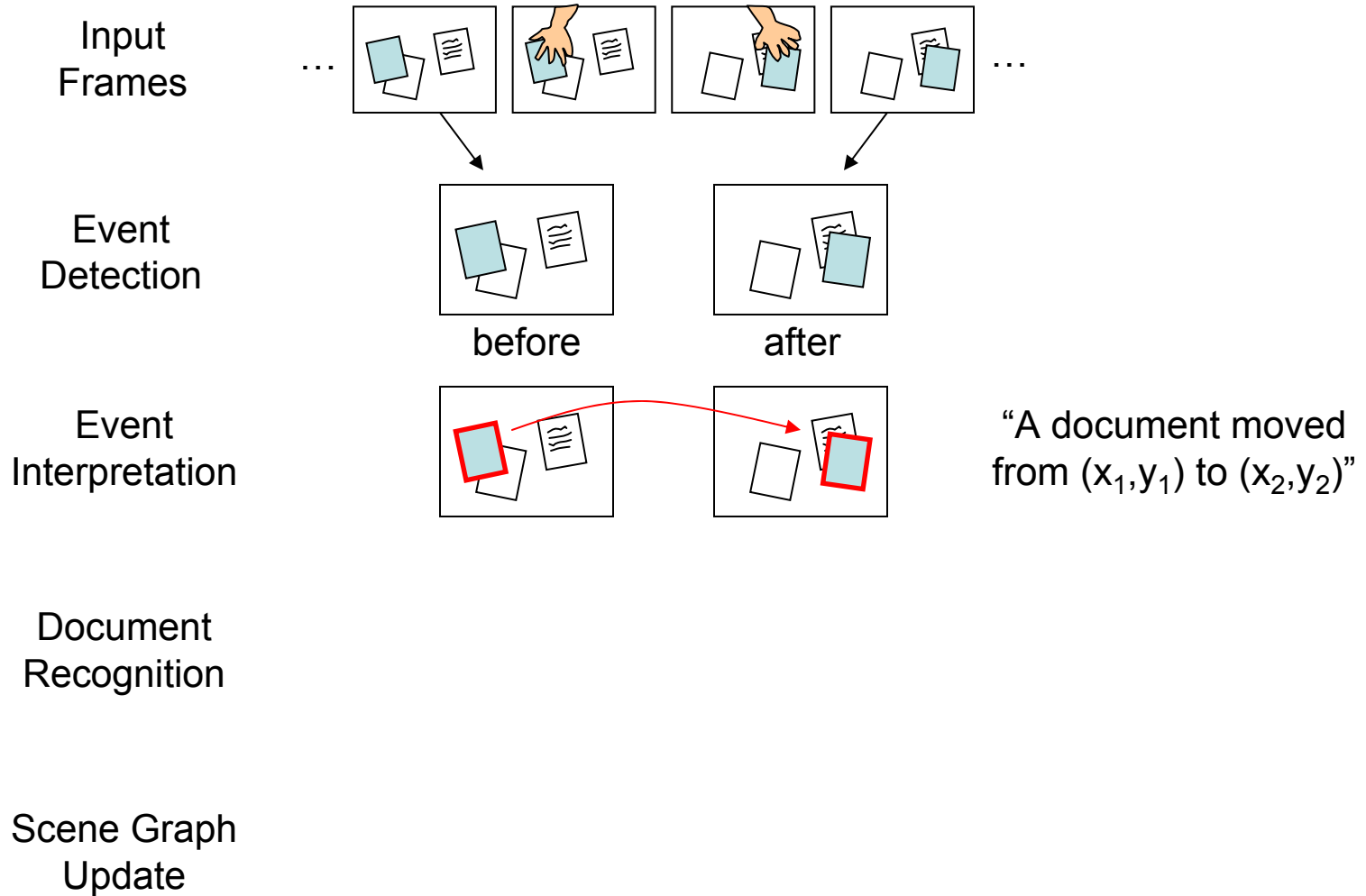
Document
Recognition

Scene Graph
Update

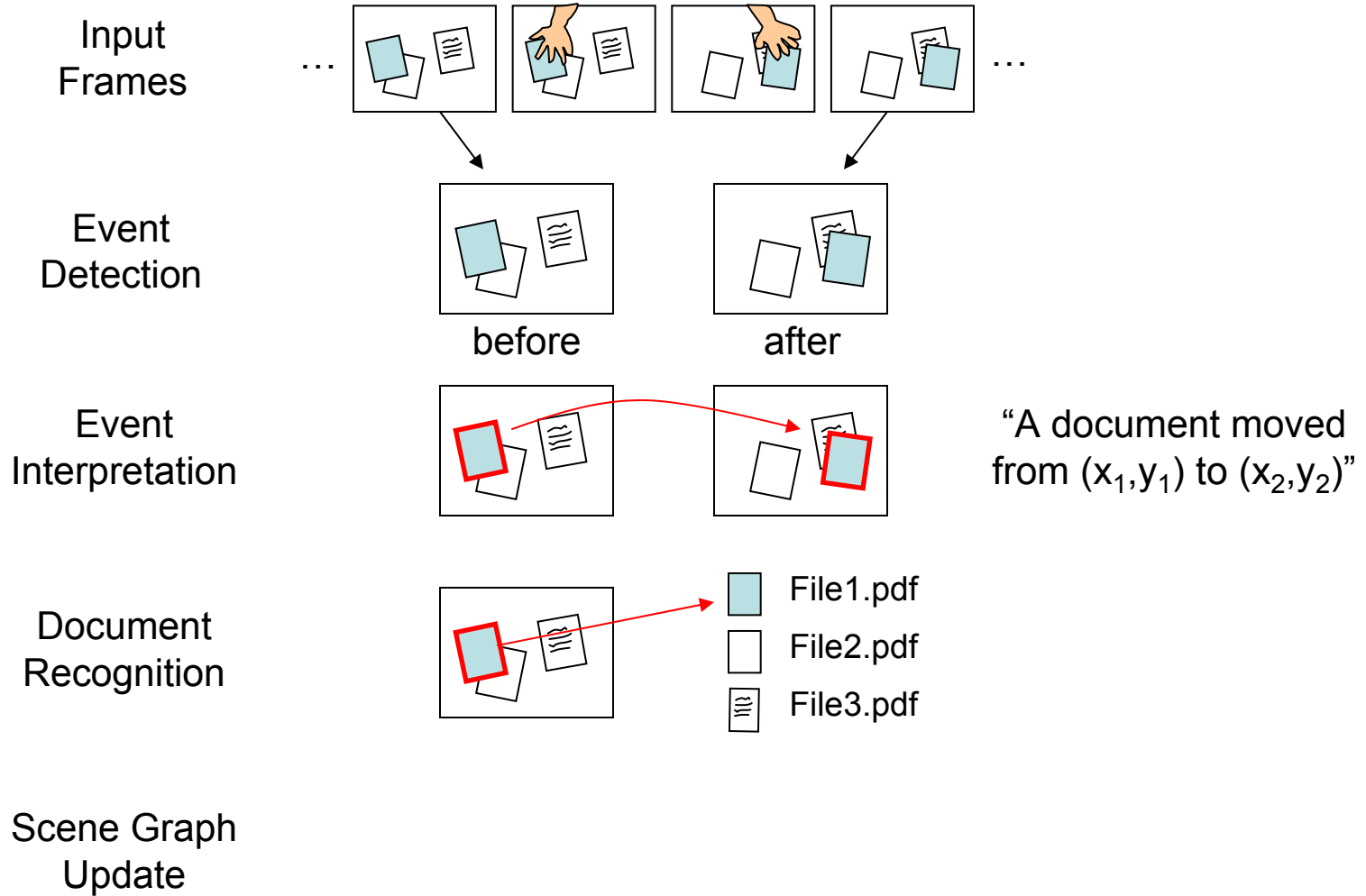
Algorithm Overview



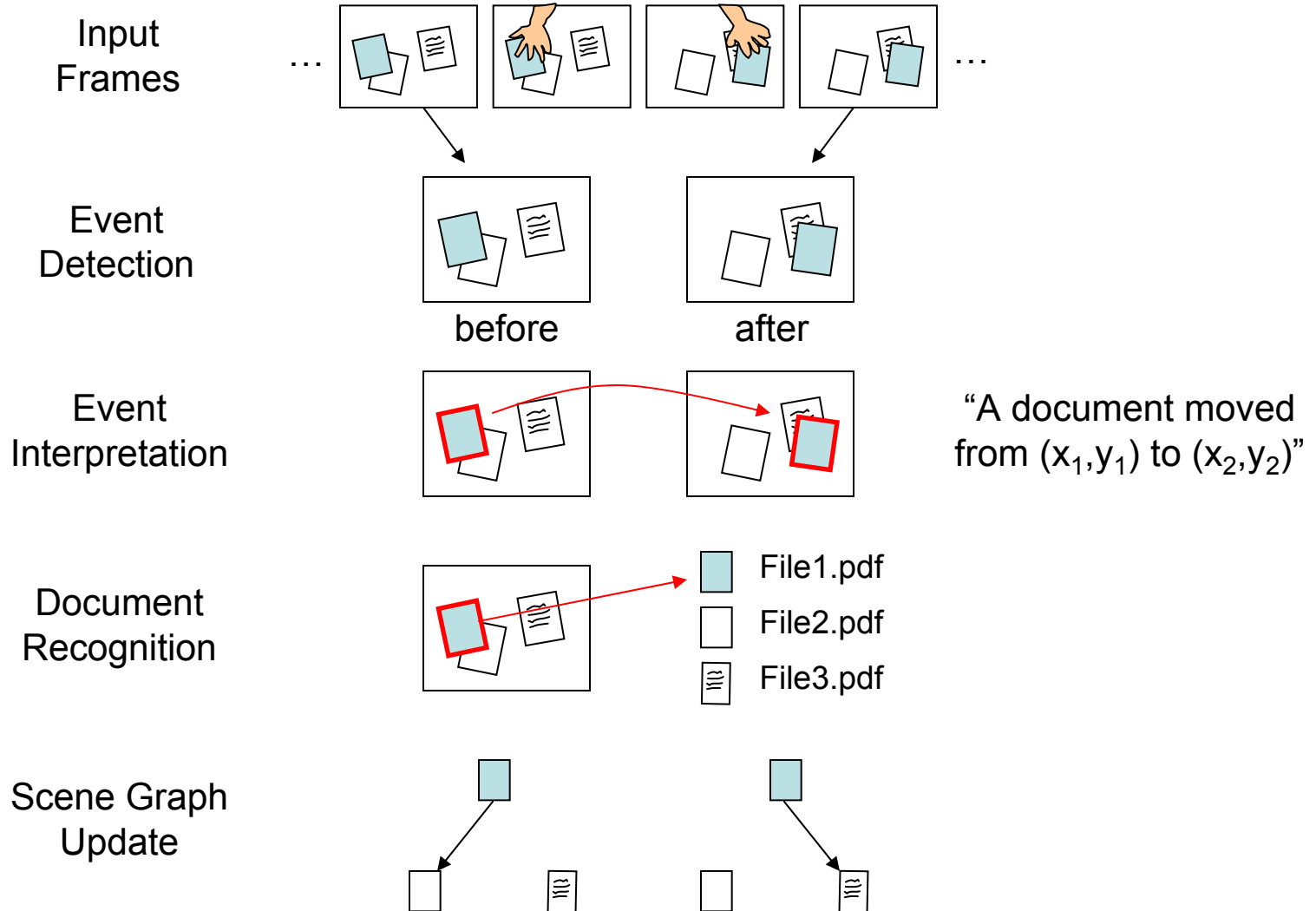
Algorithm Overview



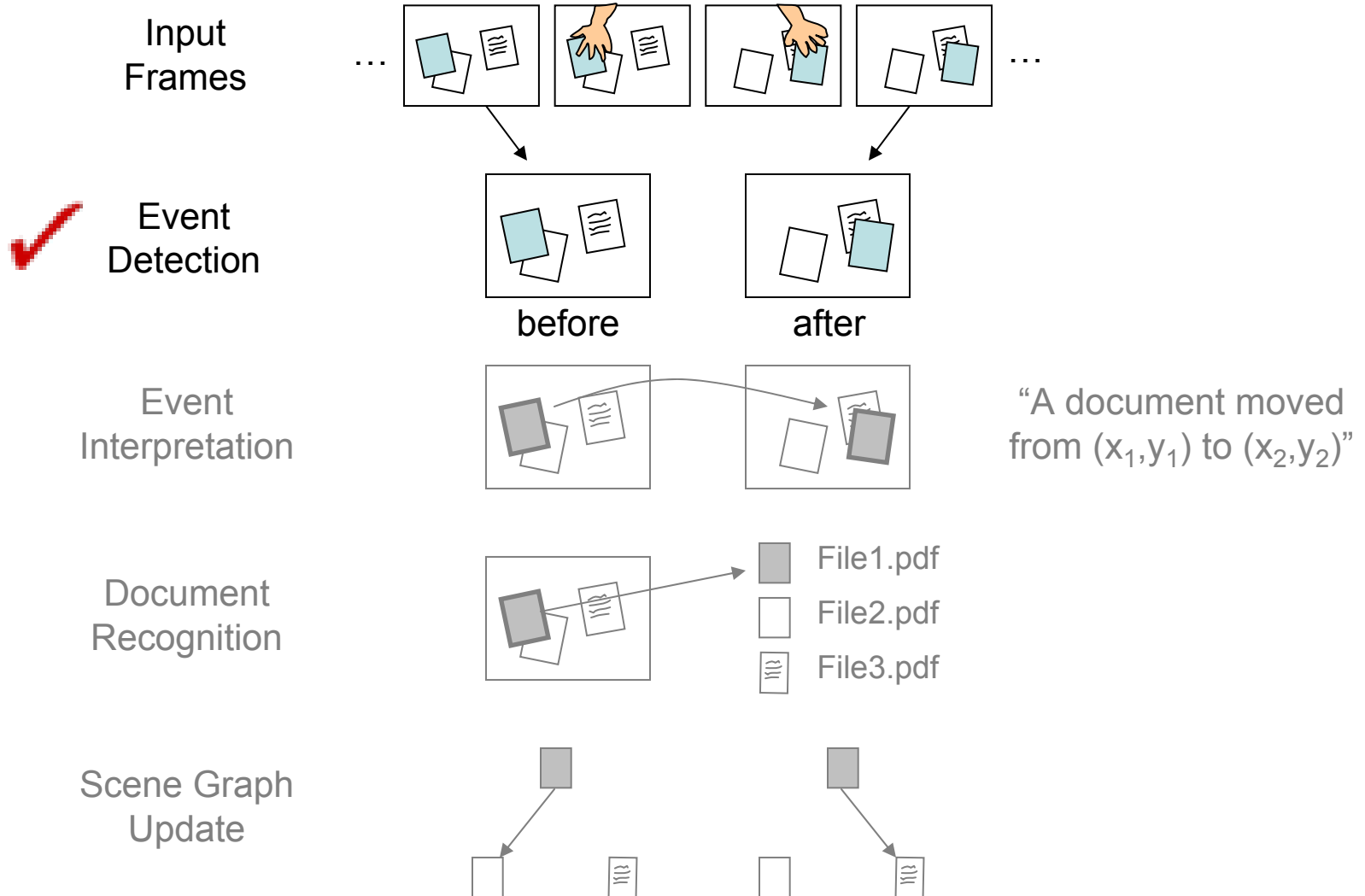
Algorithm Overview



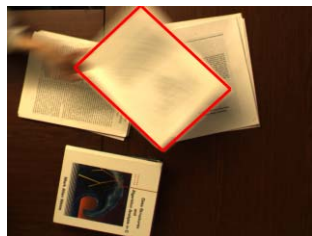
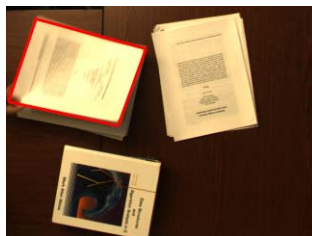
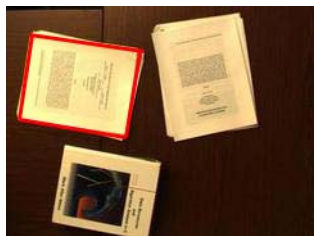
Algorithm Overview



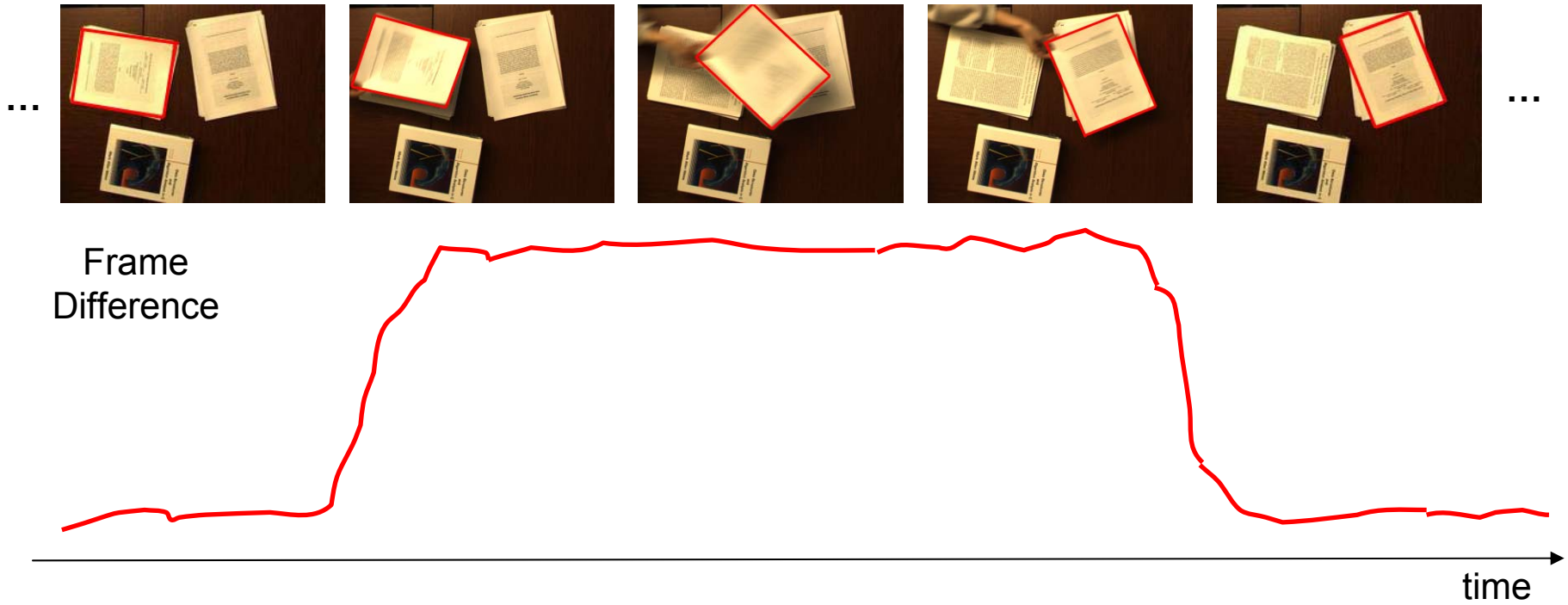
Algorithm Overview



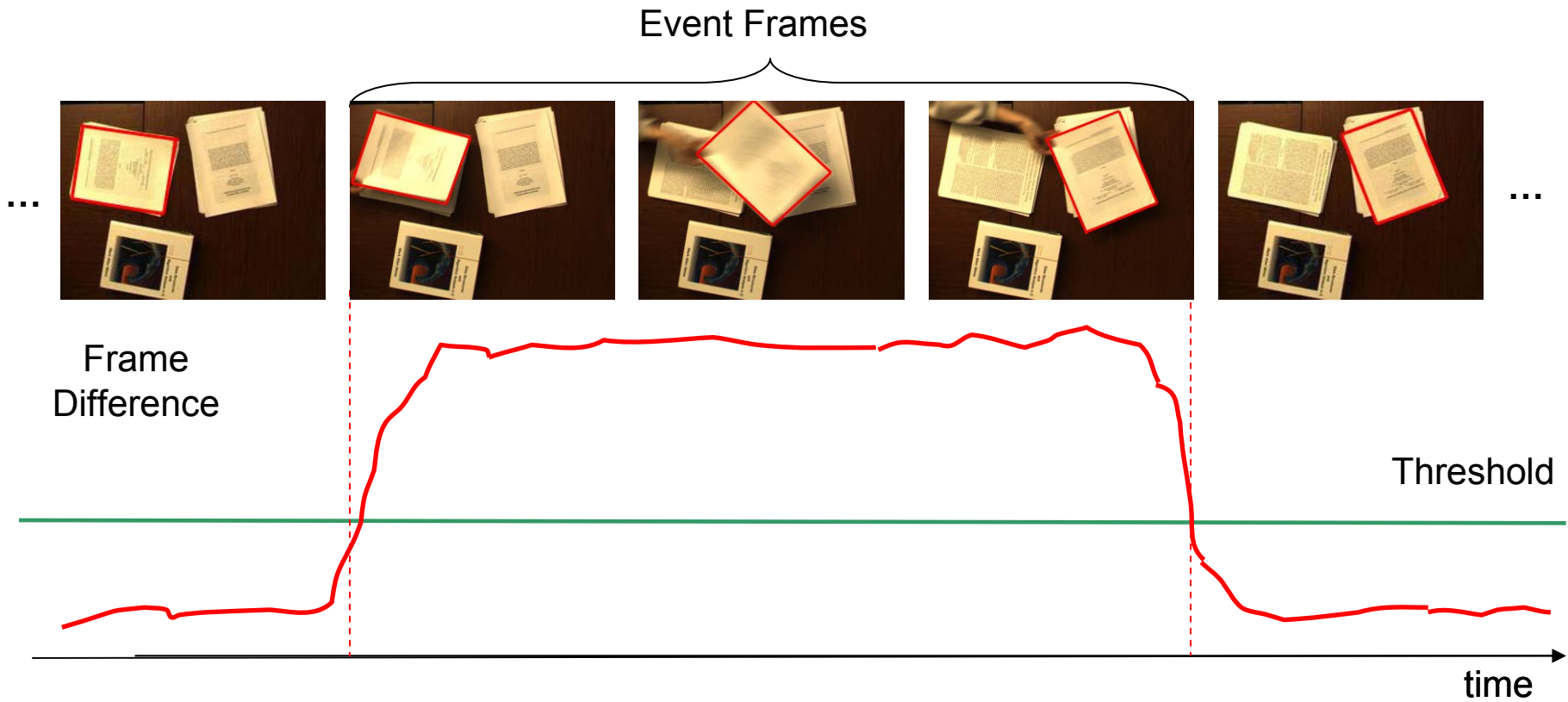
Event Detection



Event Detection

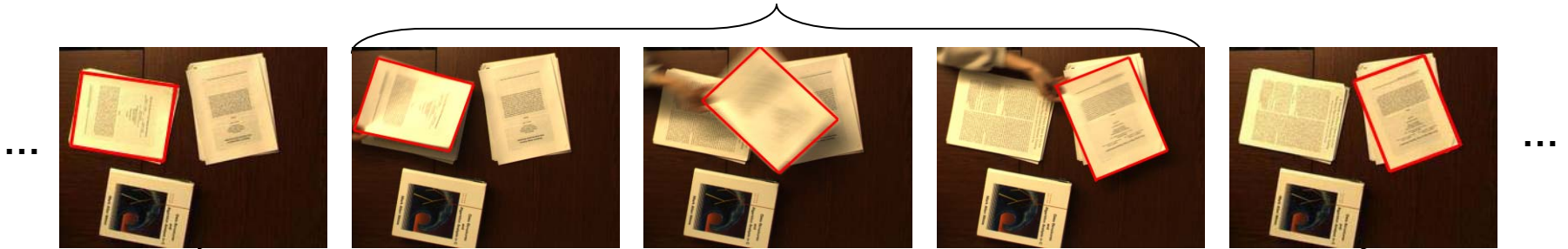


Event Detection



Event Detection

Event Frames



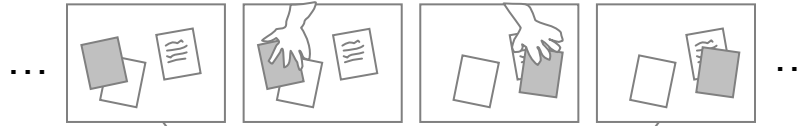
before



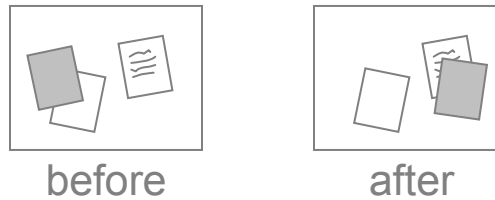
after

Algorithm Overview

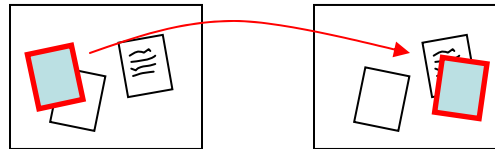
Input
Frames



Event
Detection

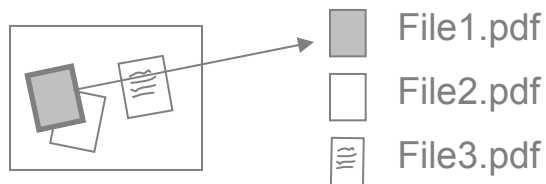


✓ Event
Interpretation



“A document moved
from (x_1, y_1) to (x_2, y_2) ”

Document
Recognition



Scene Graph
Update

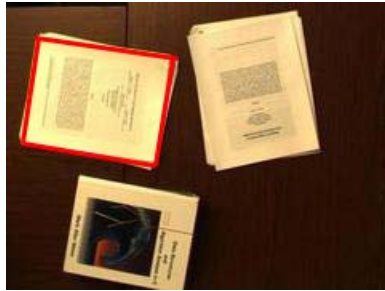


Event Interpretation

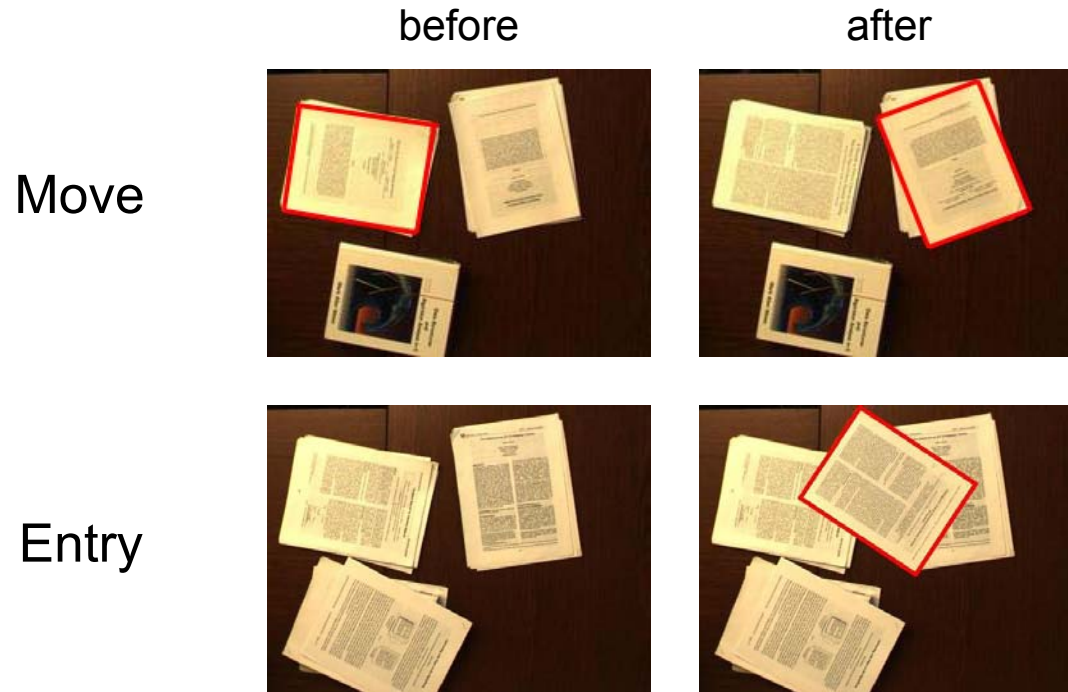
before

after

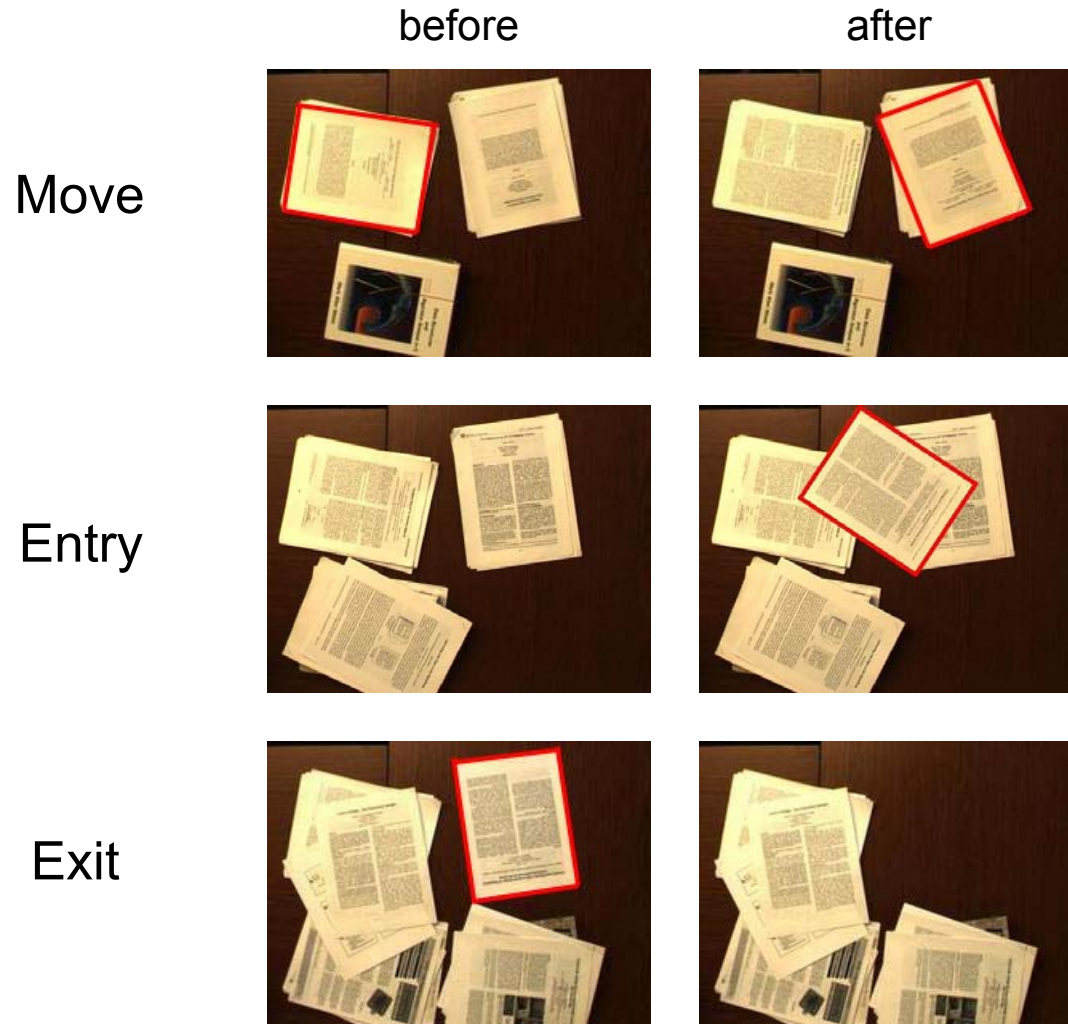
Move



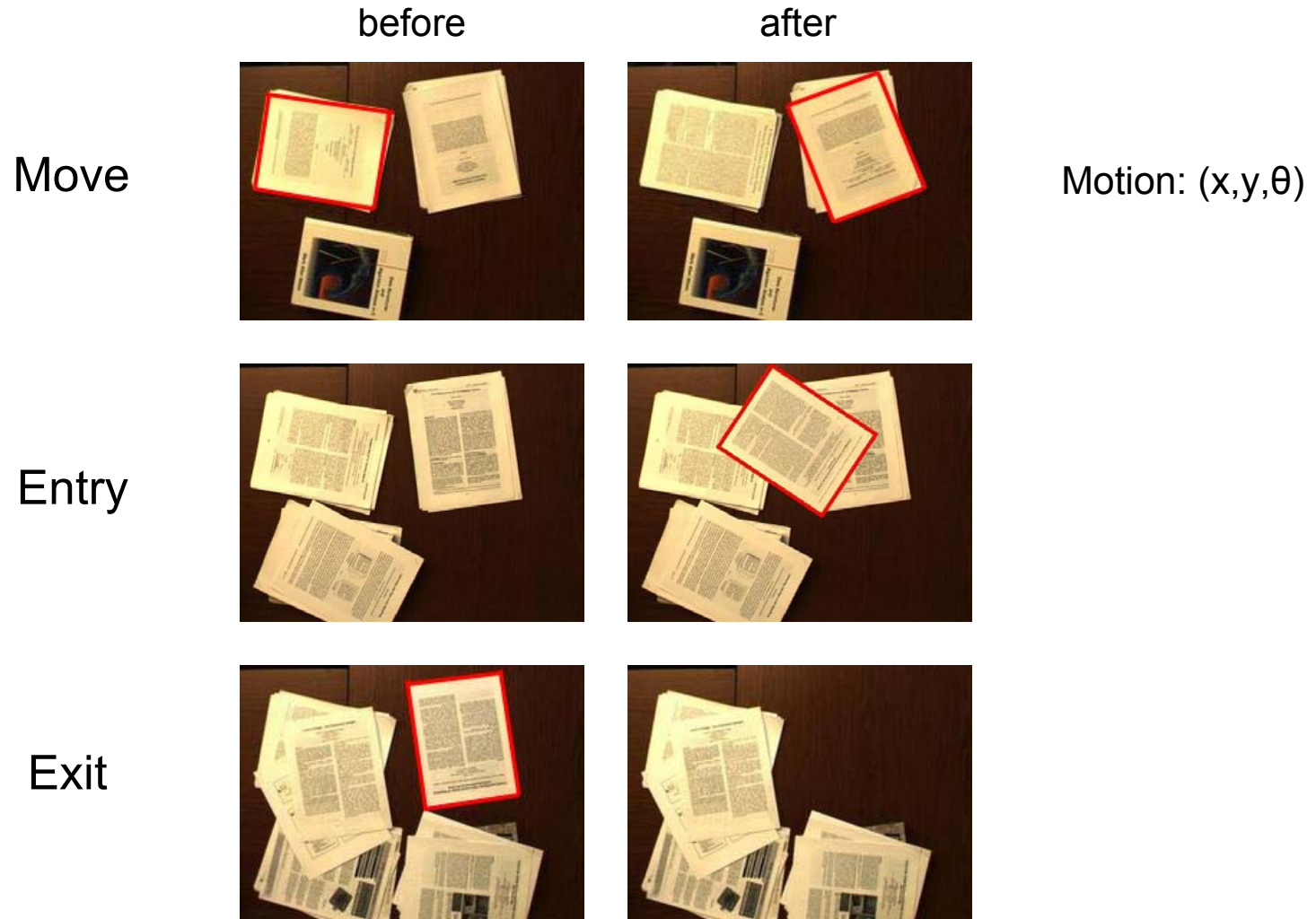
Event Interpretation



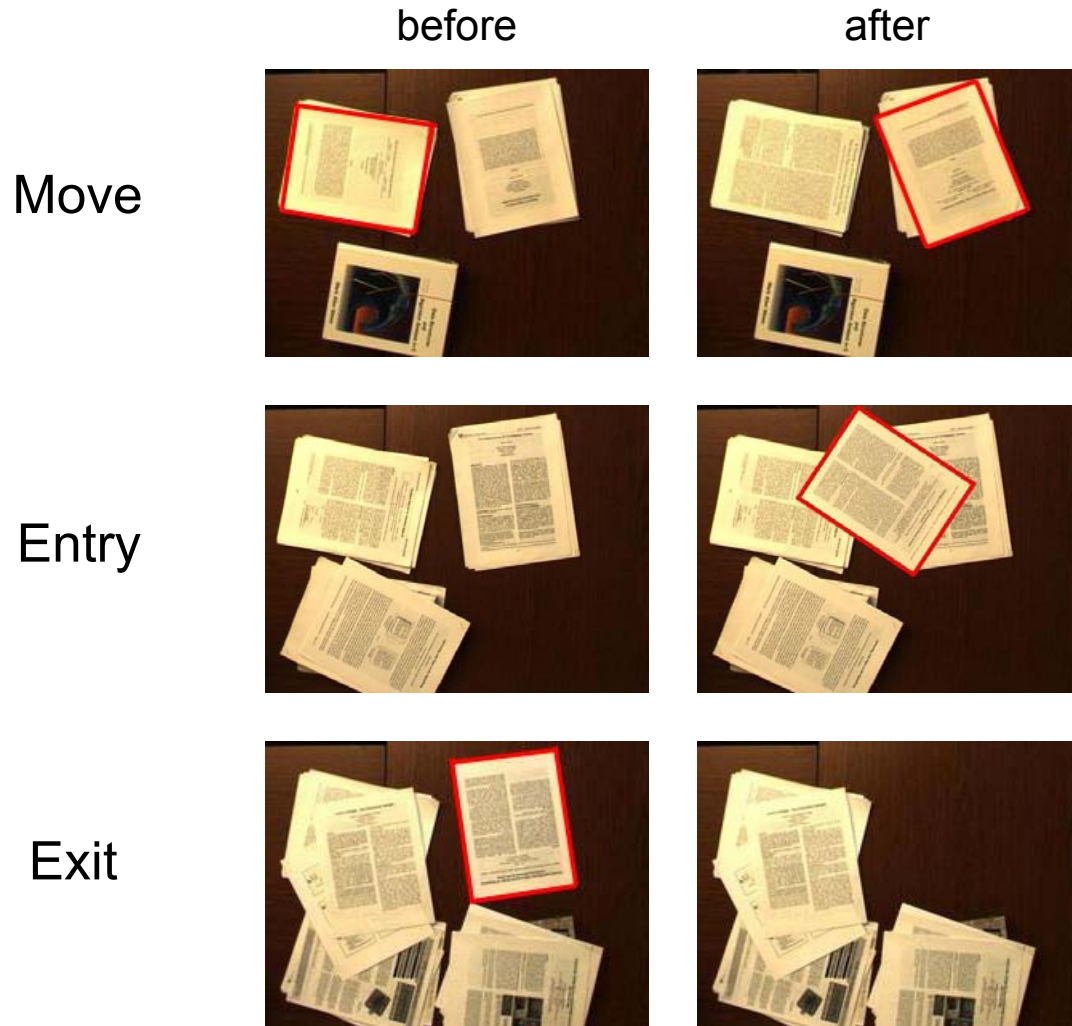
Event Interpretation



Event Interpretation

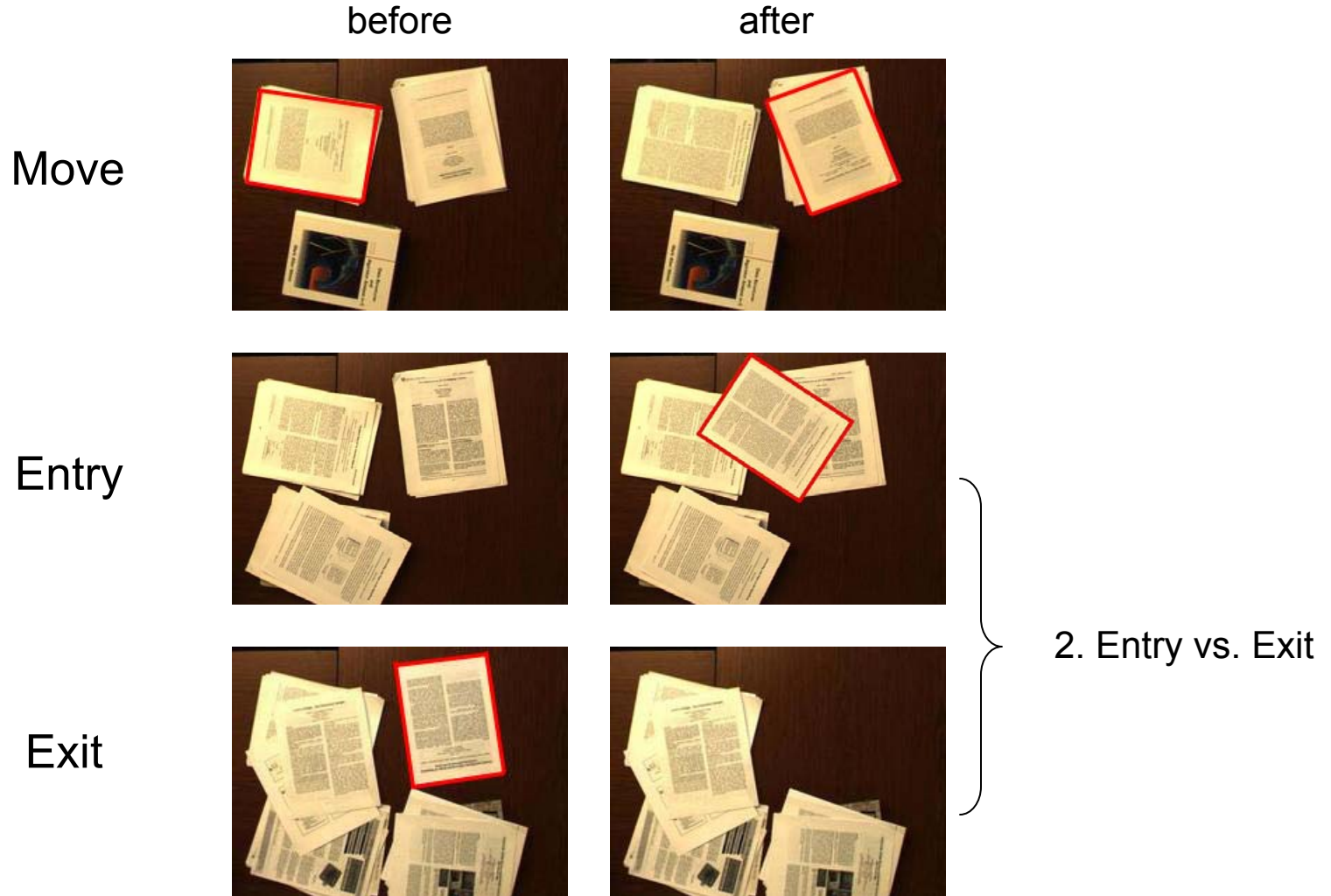


Event Interpretation



1. Move vs. Entry/Exit

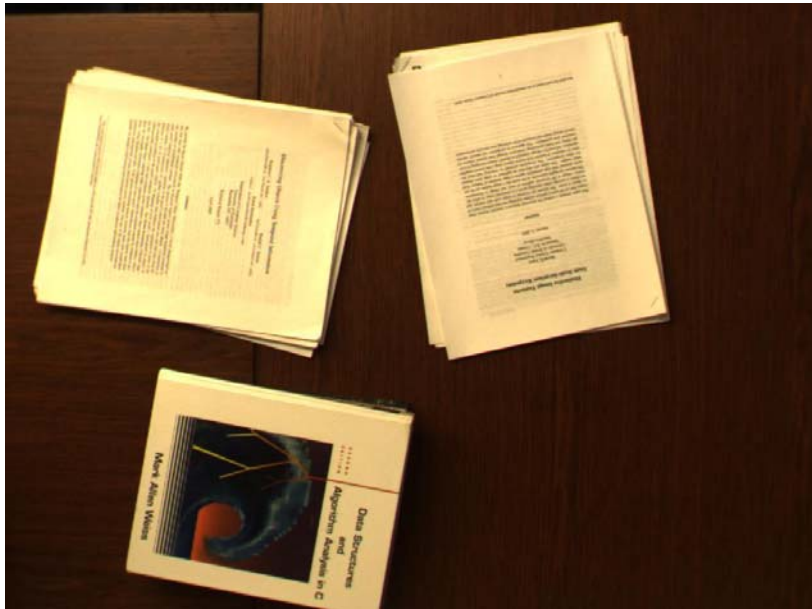
Event Interpretation



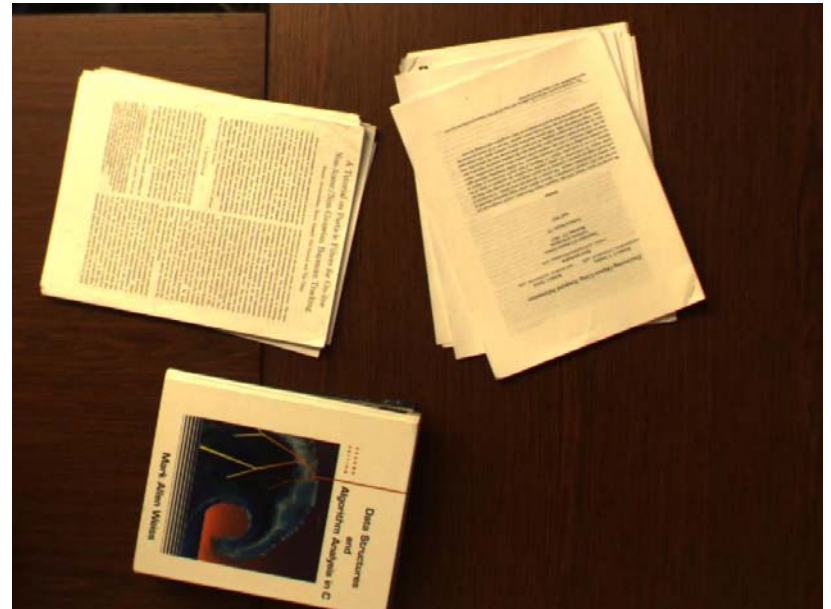
Event Interpretation

- Use SIFT [Lowe 99]
 - Scale Invariant Feature Transform
 - Distinctive feature descriptor
 - Reliable object recognition

Move vs. Entry/Exit

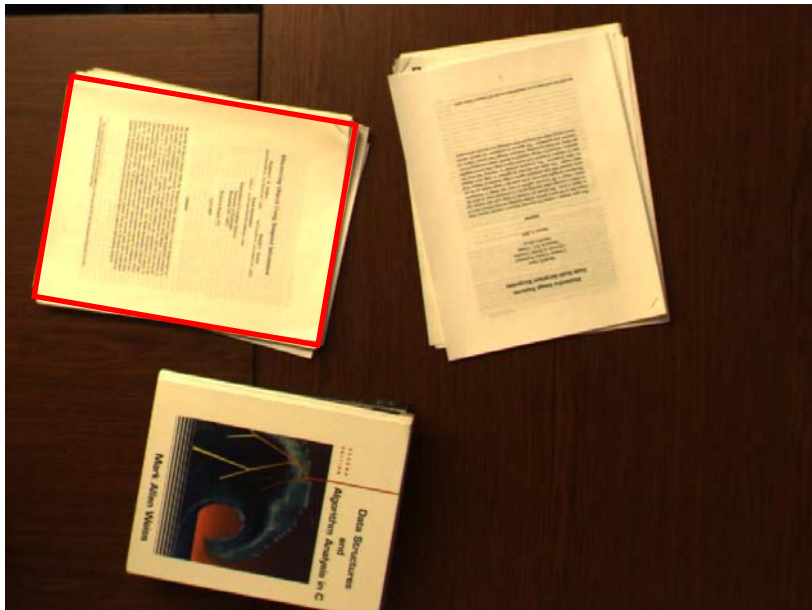


before

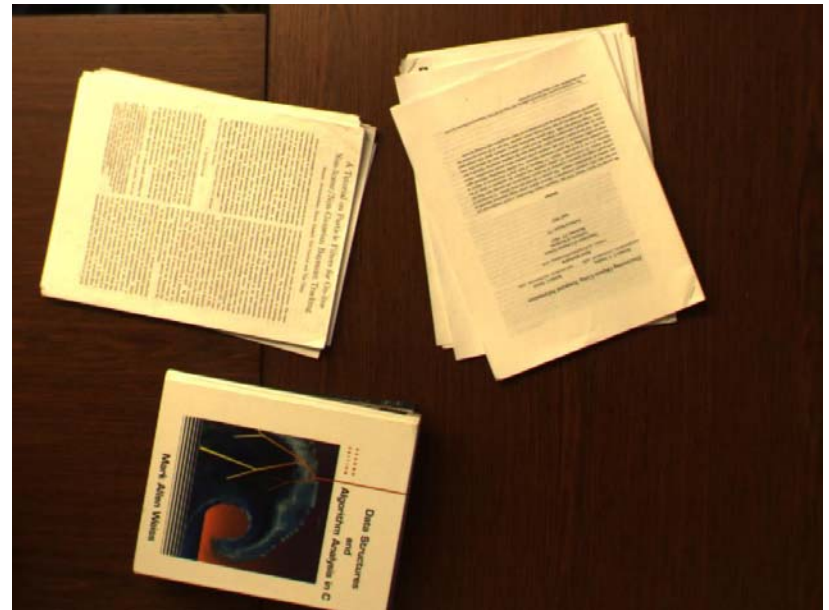


after

Move vs. Entry/Exit

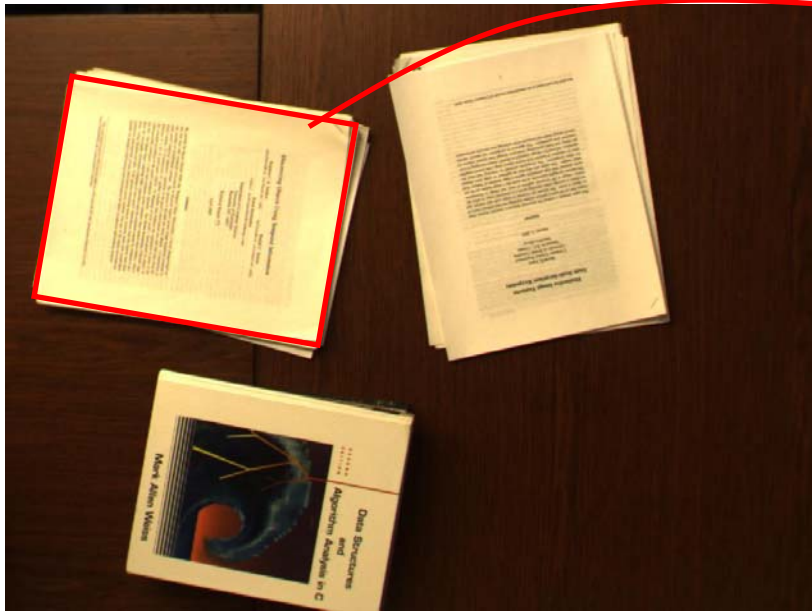


before

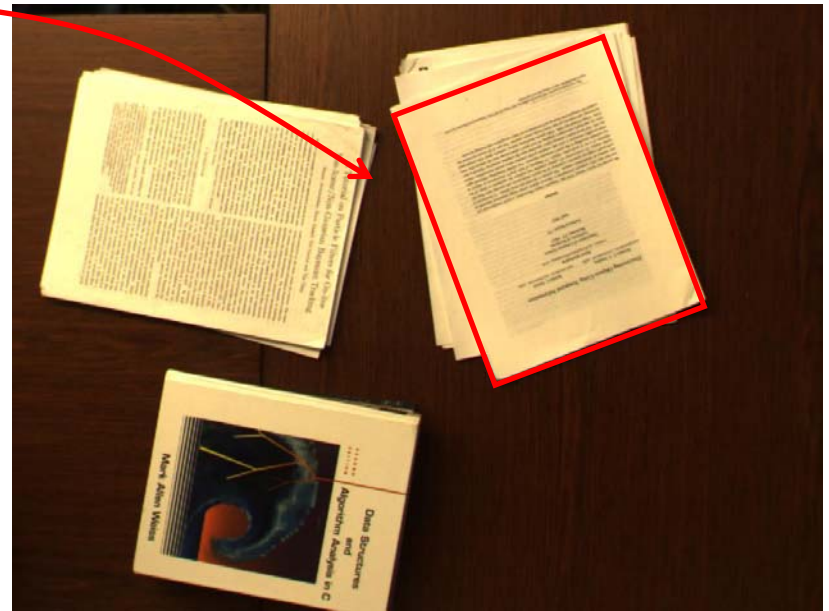


after

Move vs. Entry/Exit

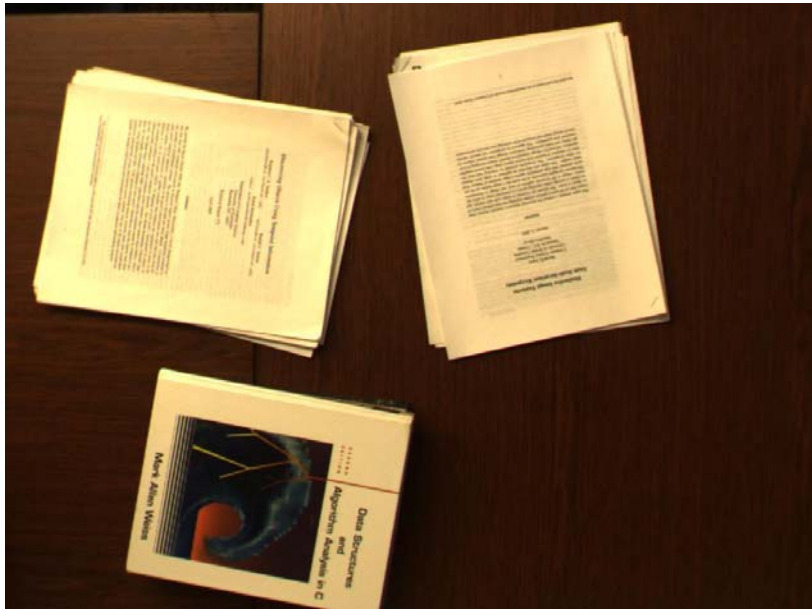


before

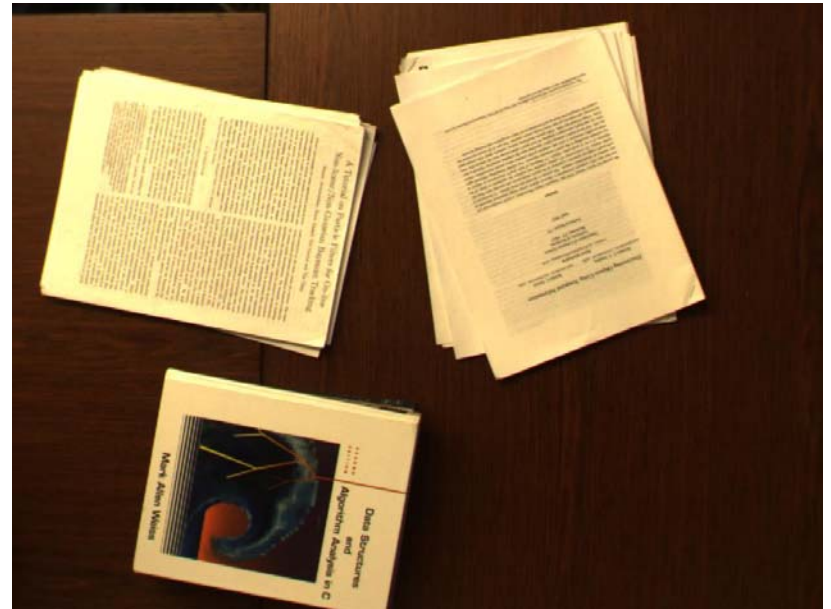


after

Move vs. Entry/Exit

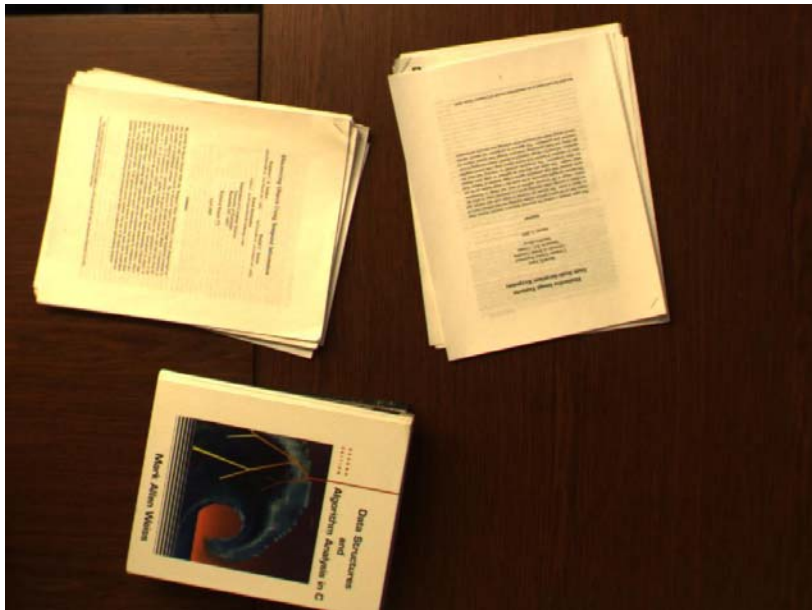


before

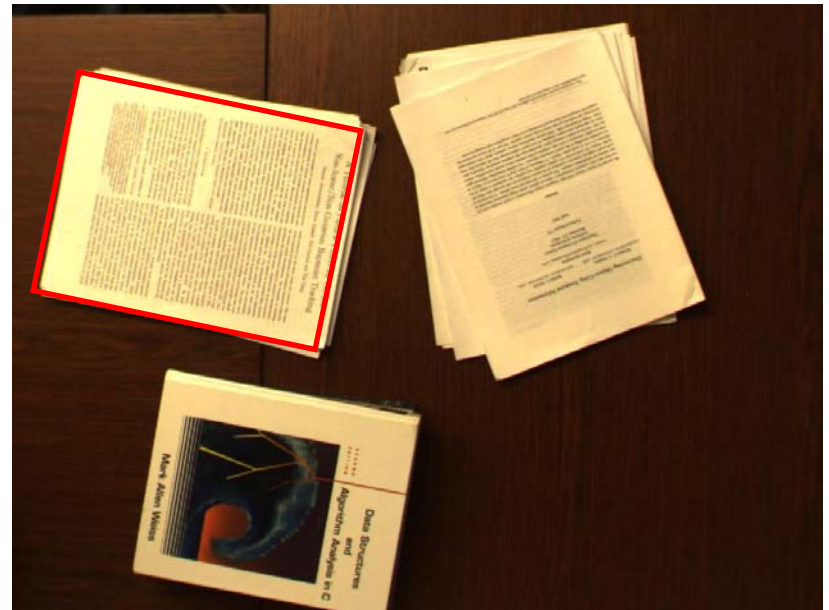


after

Move vs. Entry/Exit

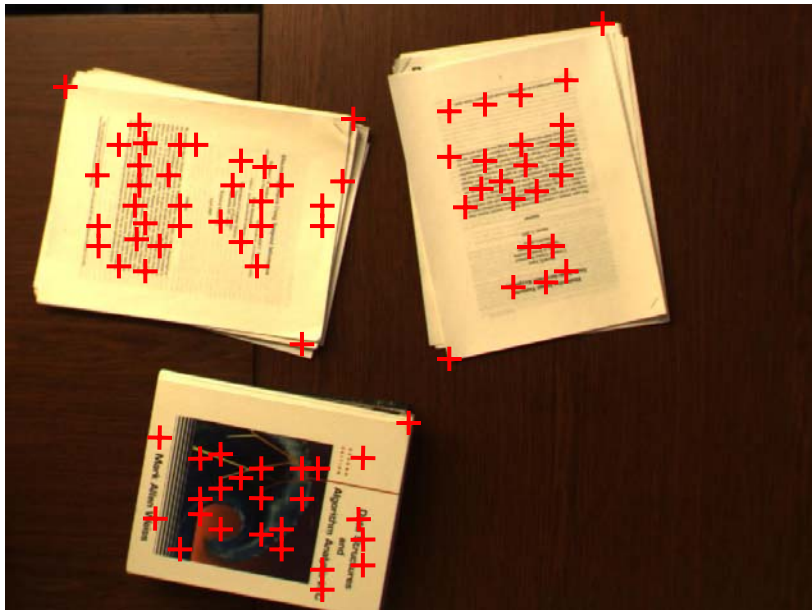


before

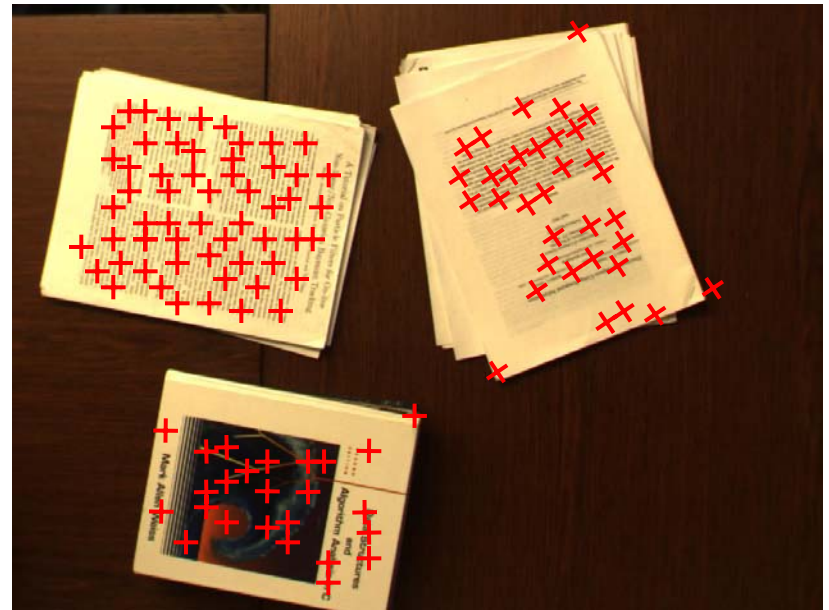


after

Move vs. Entry/Exit

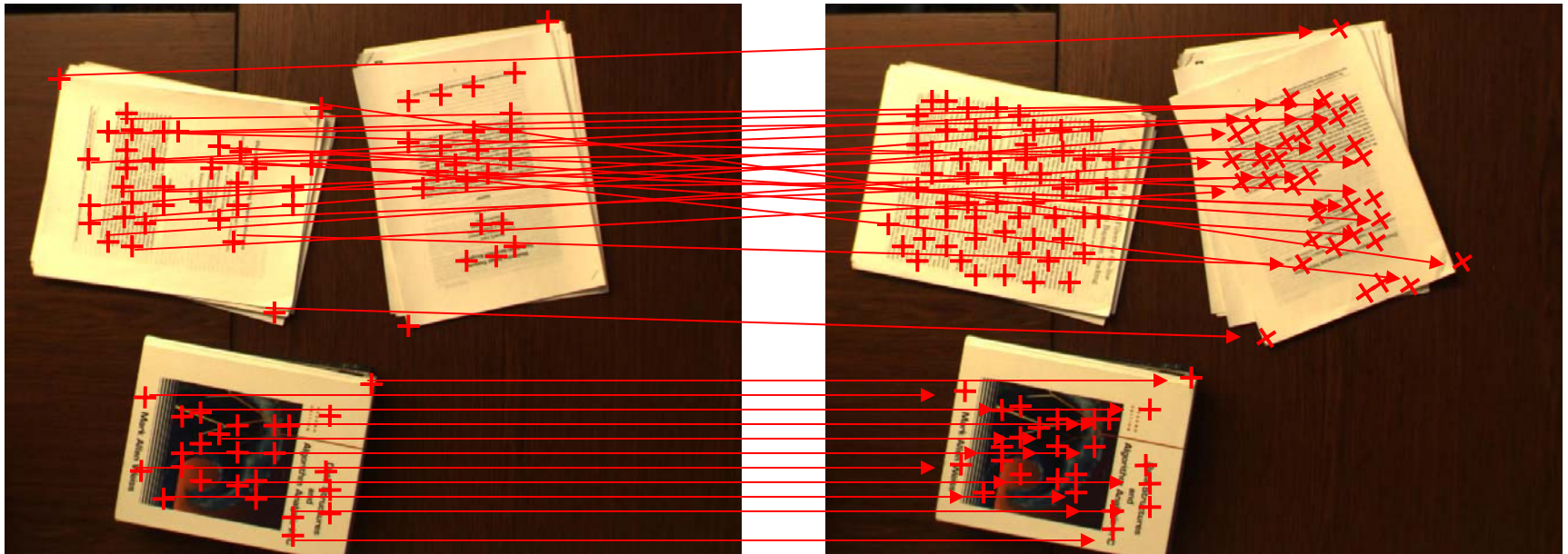


before



after

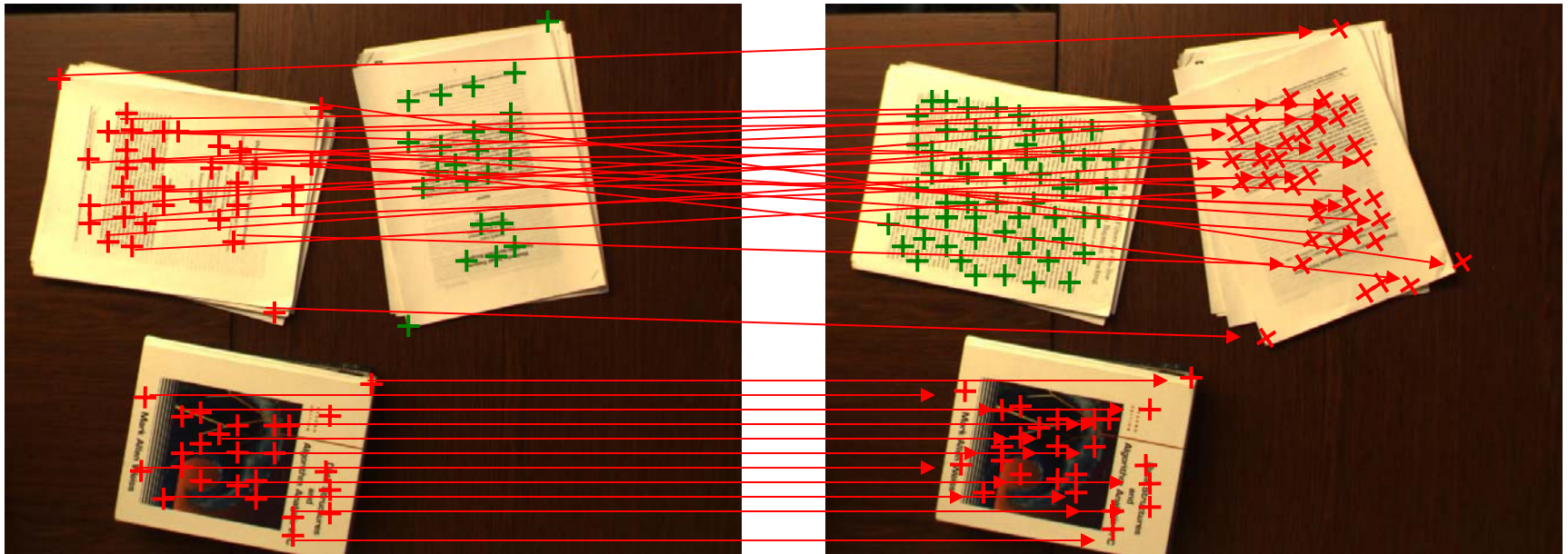
Move vs. Entry/Exit



before

after

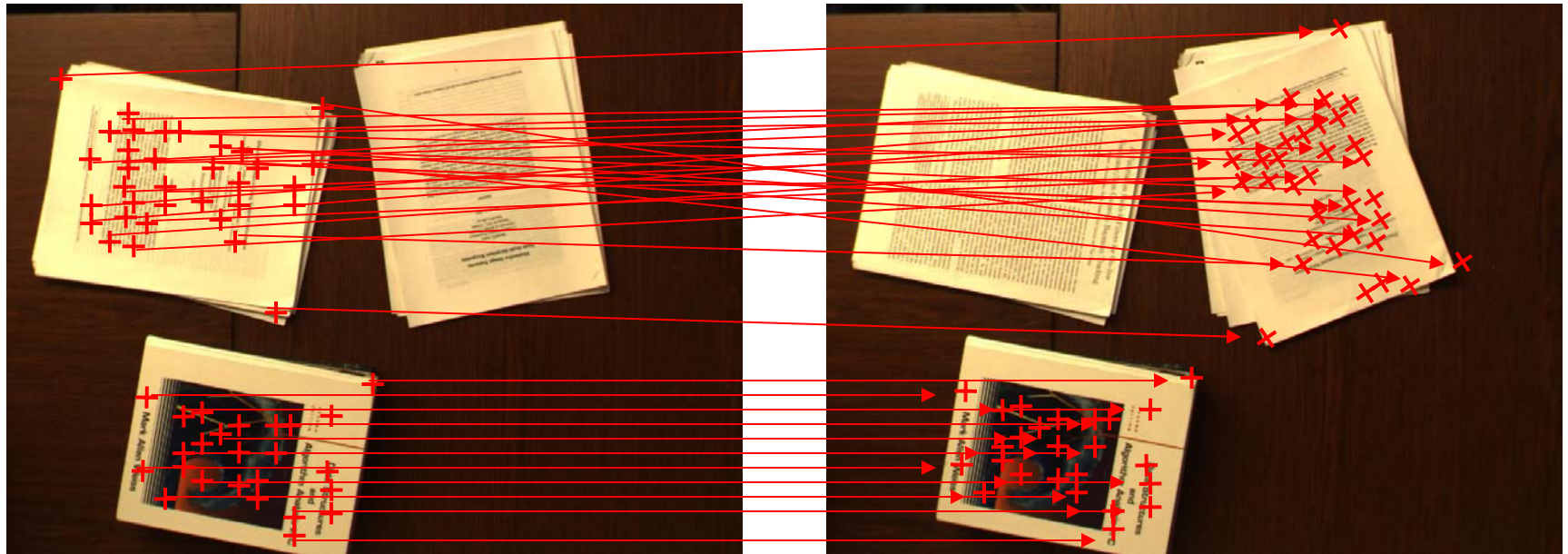
Move vs. Entry/Exit



before

after

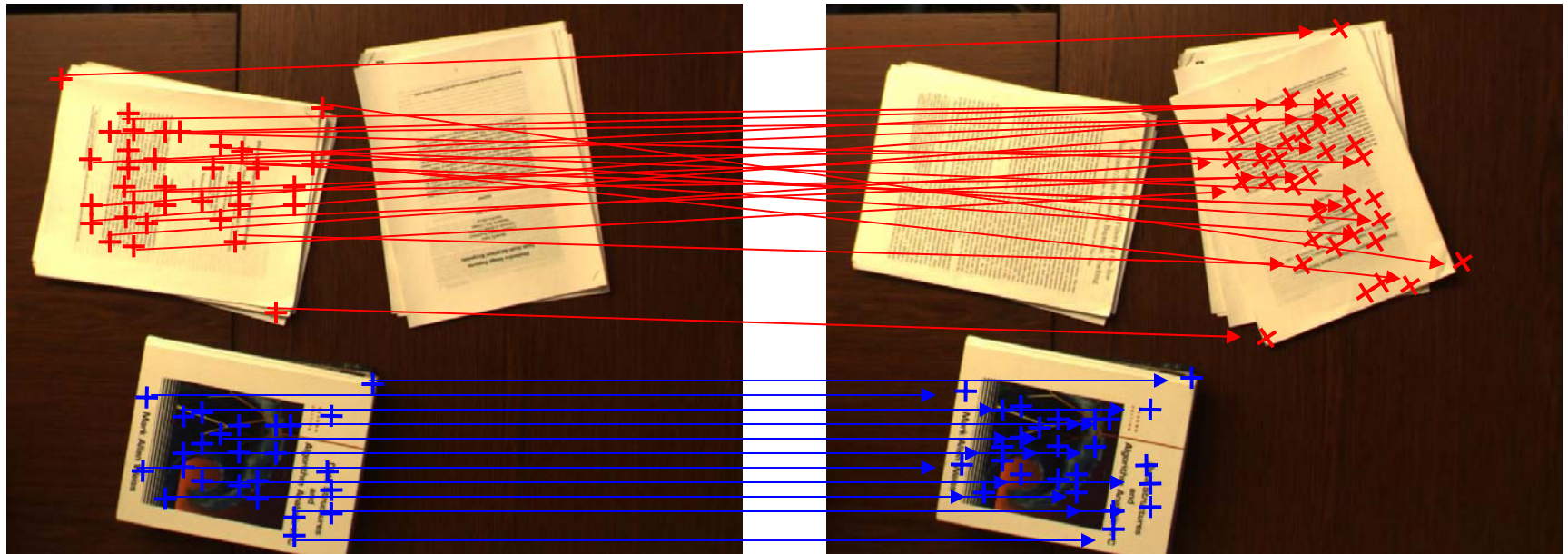
Move vs. Entry/Exit



before

after

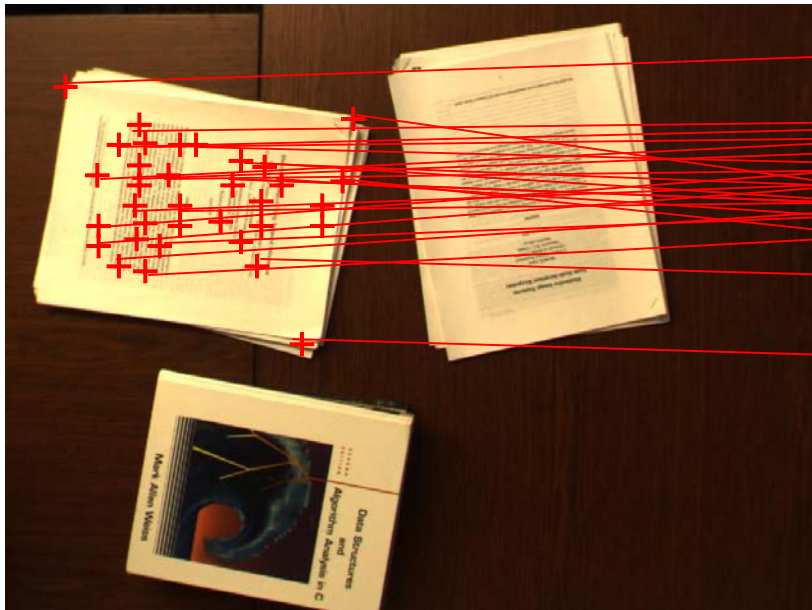
Move vs. Entry/Exit



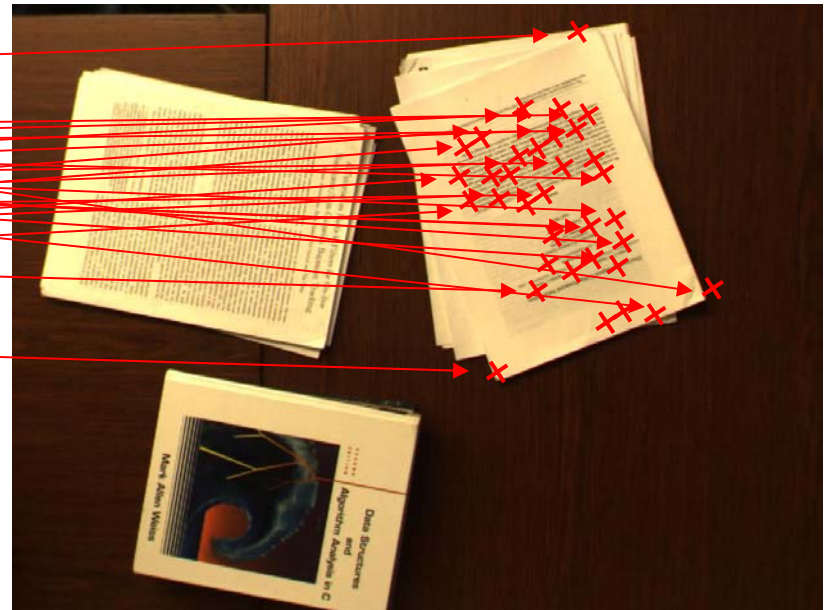
before

after

Move vs. Entry/Exit



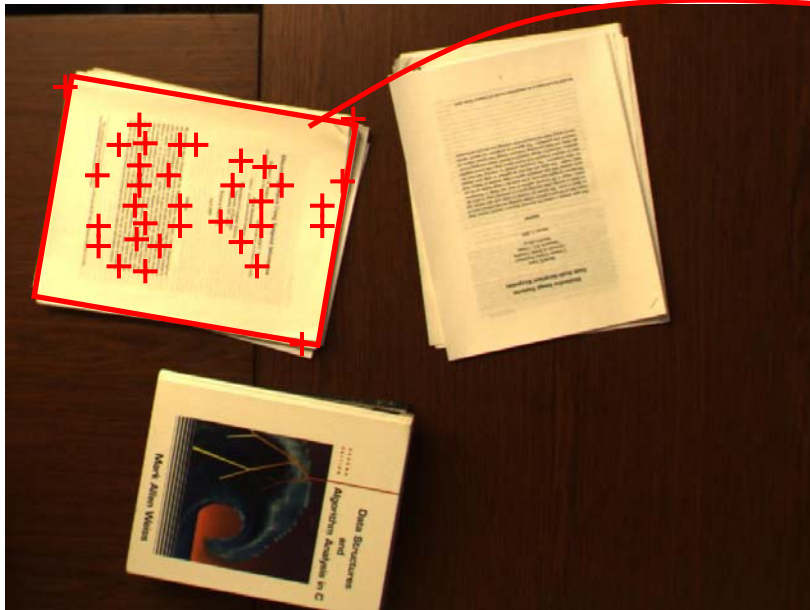
before



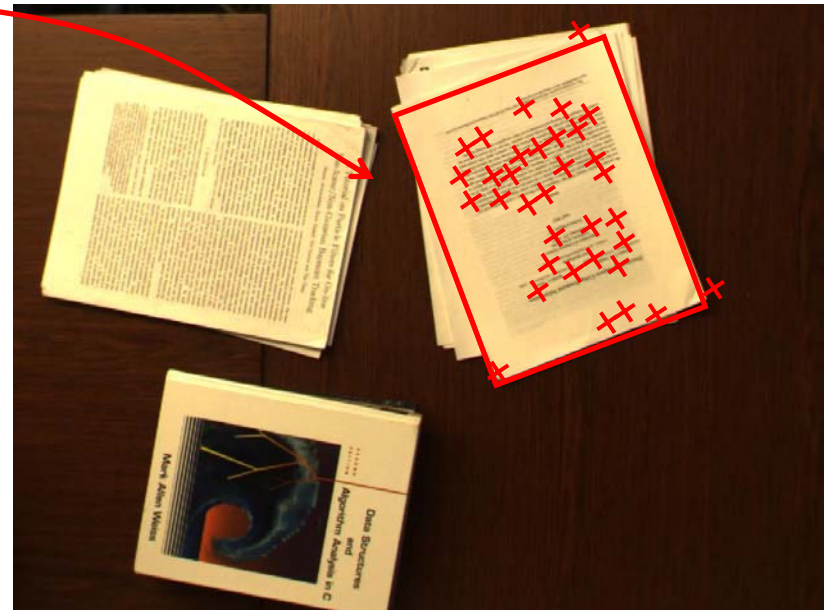
after

Move vs. Entry/Exit

Motion: (x,y,θ)



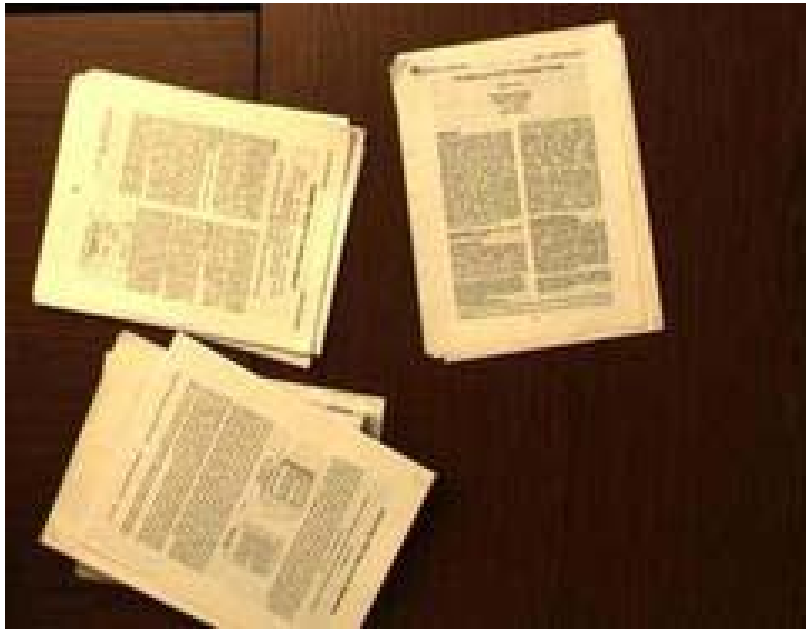
before



after

Entry vs. Exit

Example 1 (entry)



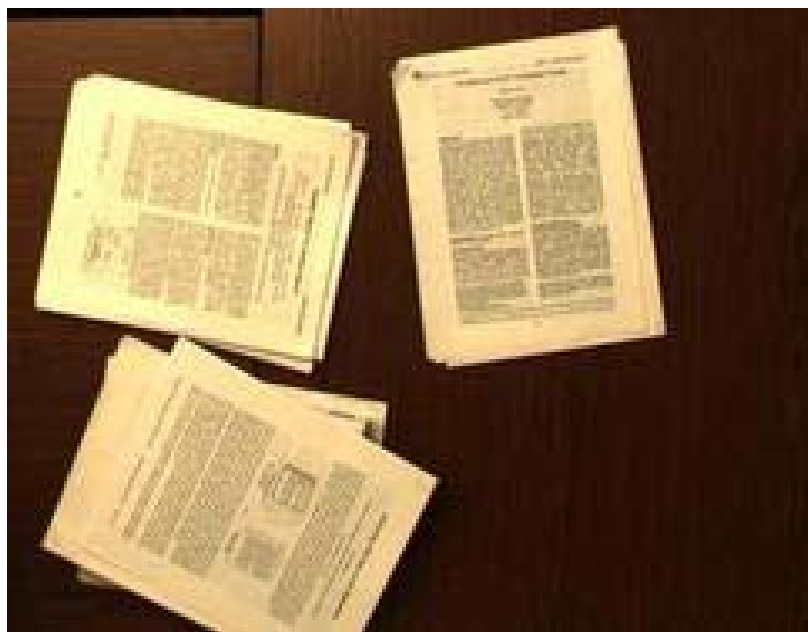
before



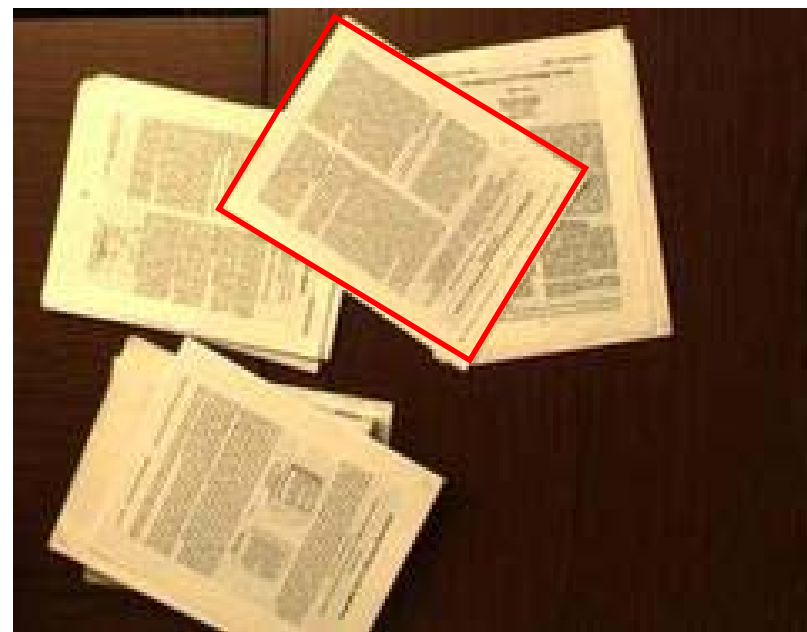
after

Entry vs. Exit

Example 1 (entry)



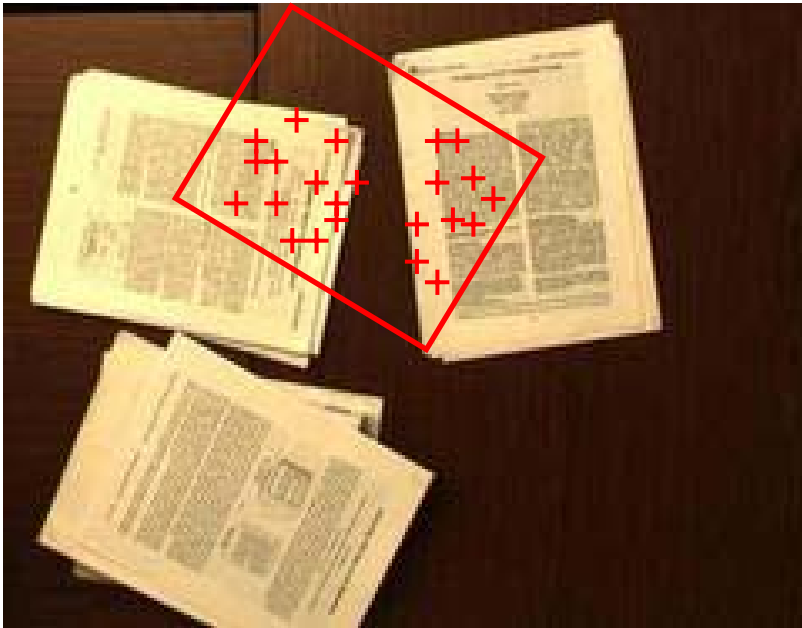
before



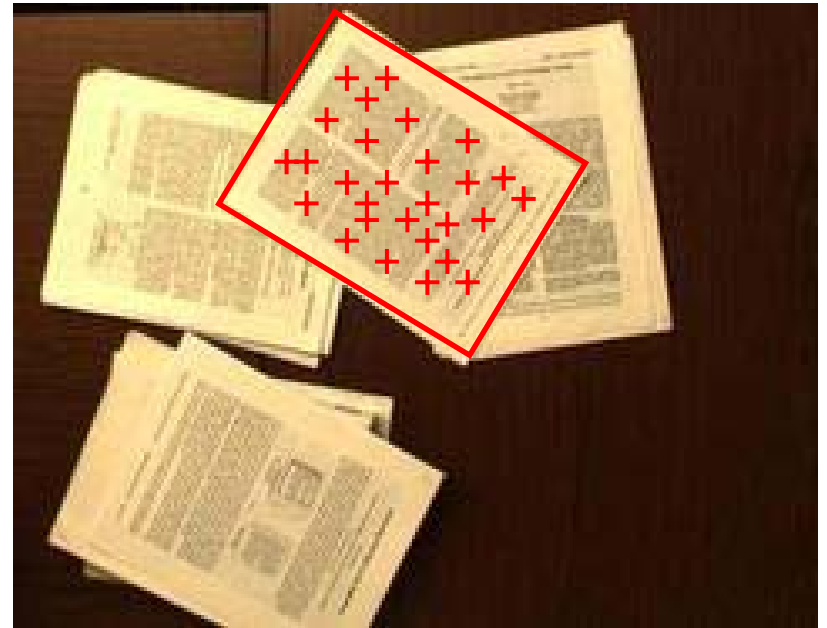
after

Entry vs. Exit

Example 1 (entry)

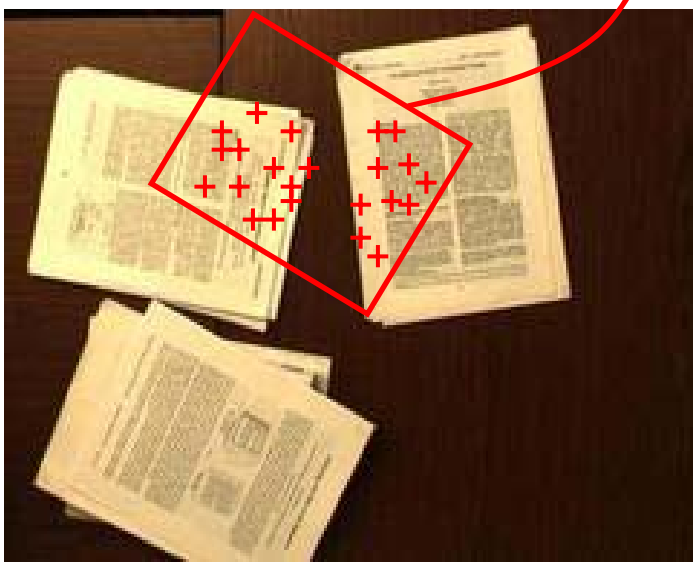
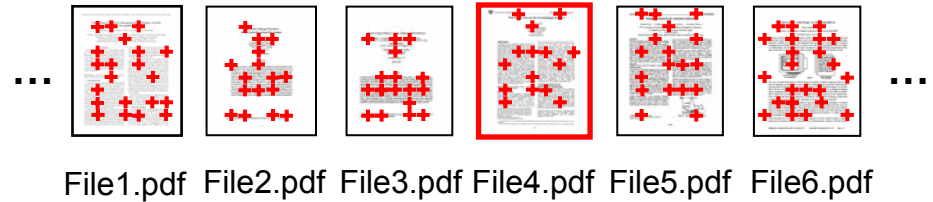


before

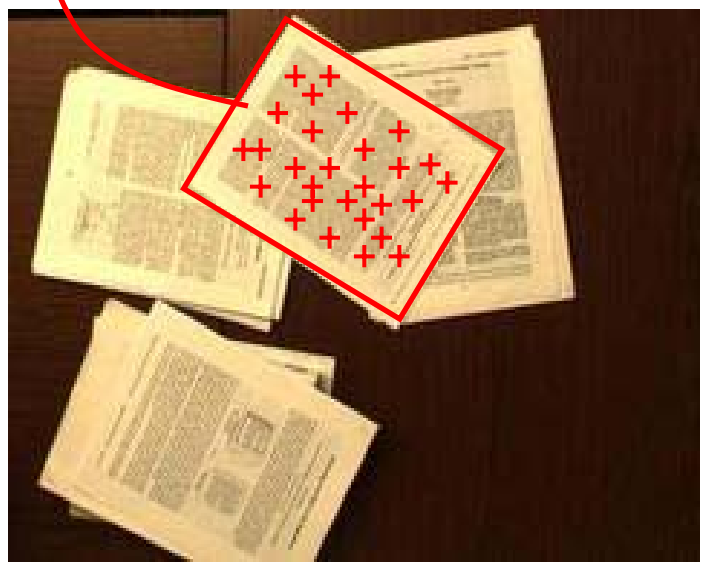


after

Entry vs. Exit



before



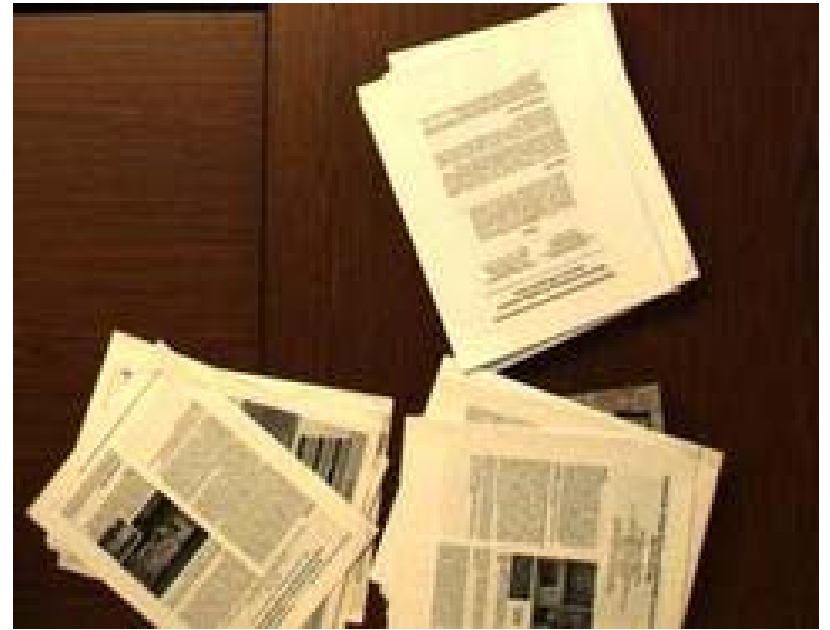
after

Entry vs. Exit

Example 2 (entry)

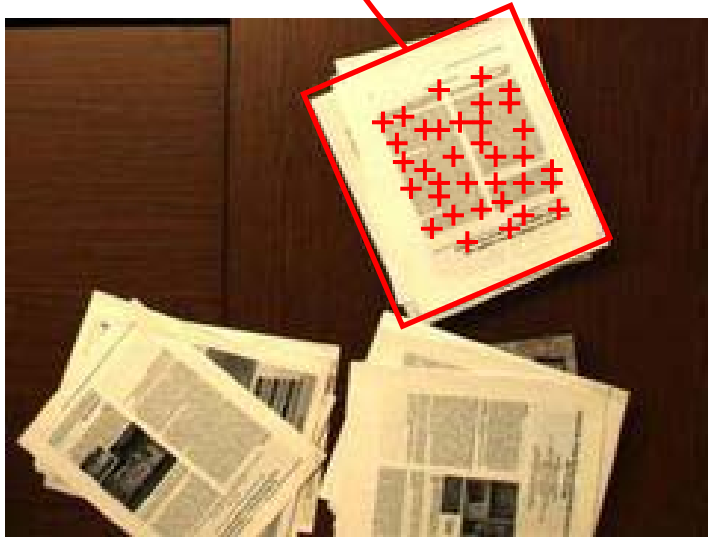
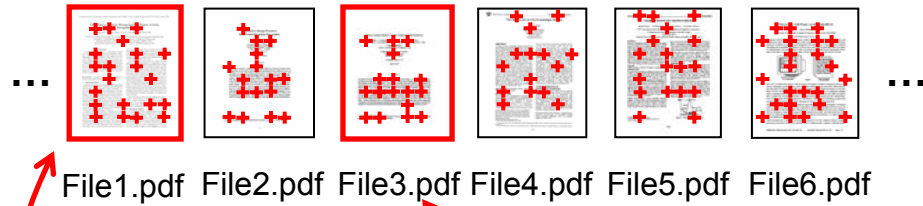


before

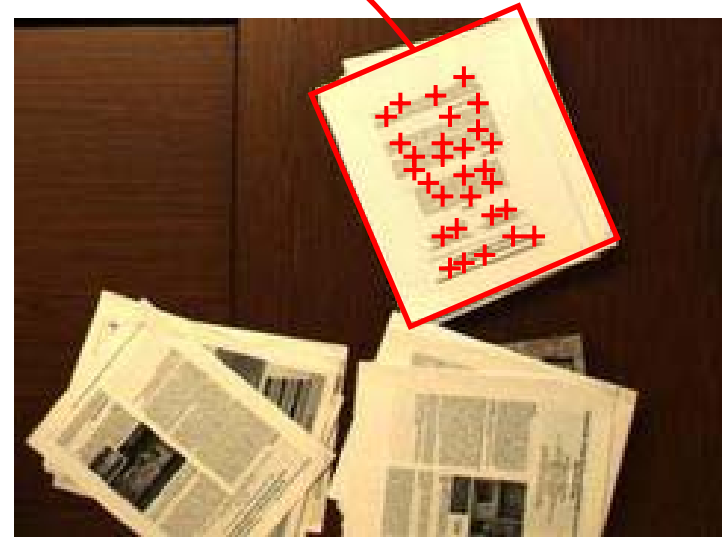


after

Entry vs. Exit

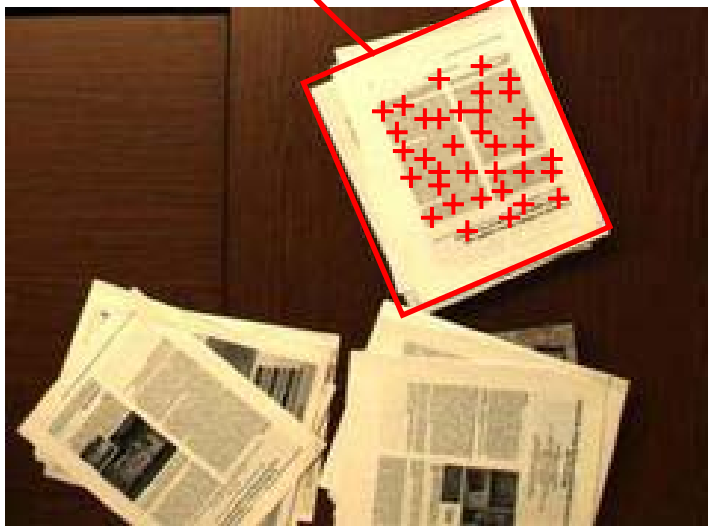
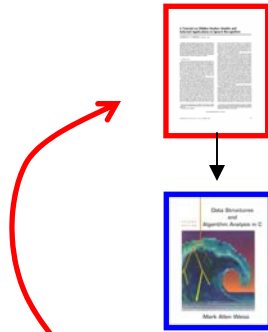


before

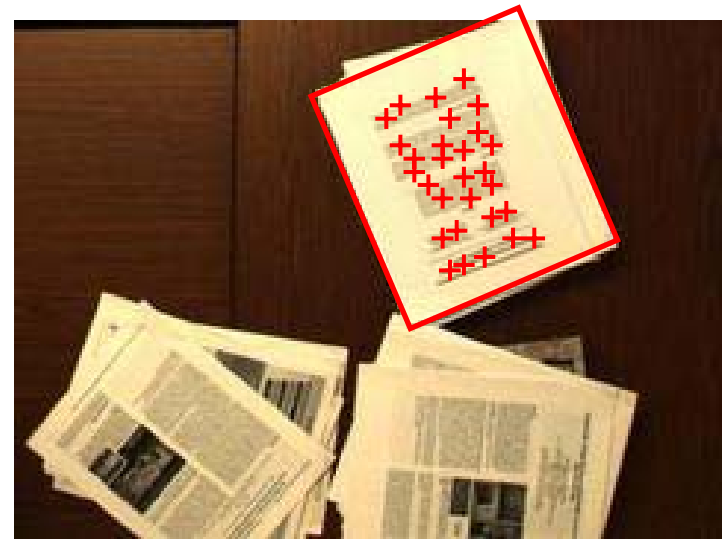


after

Entry vs. Exit

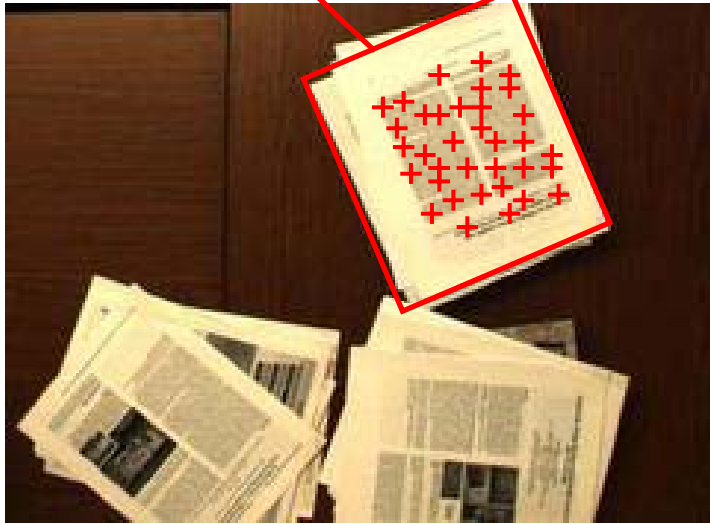
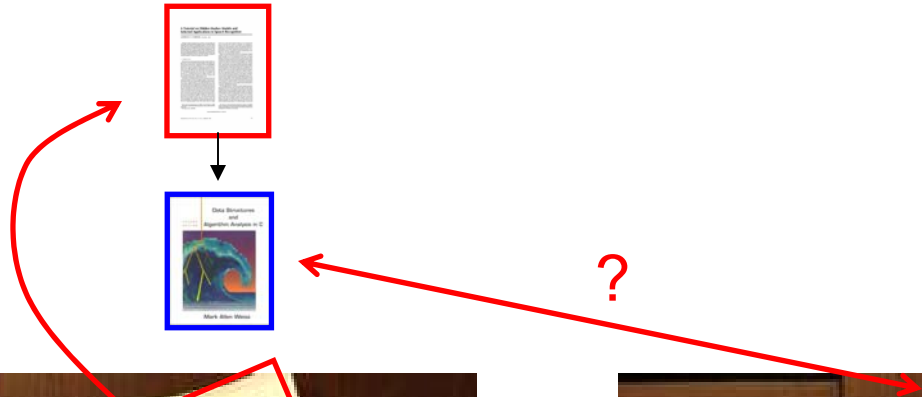


before

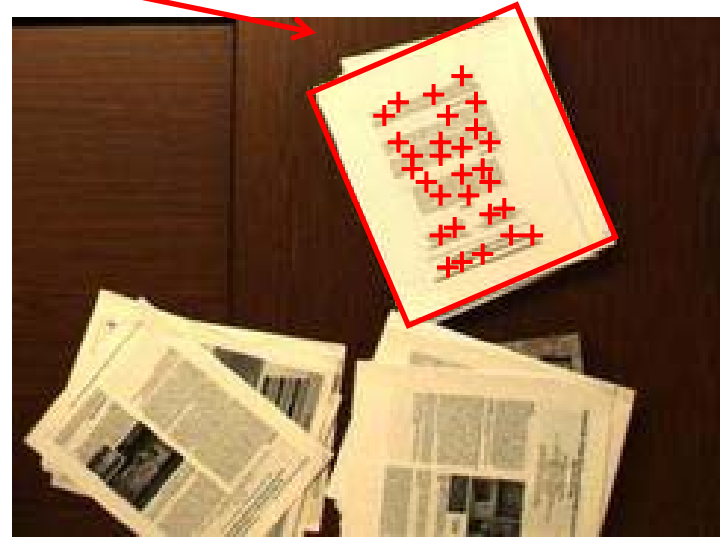


after

Entry vs. Exit

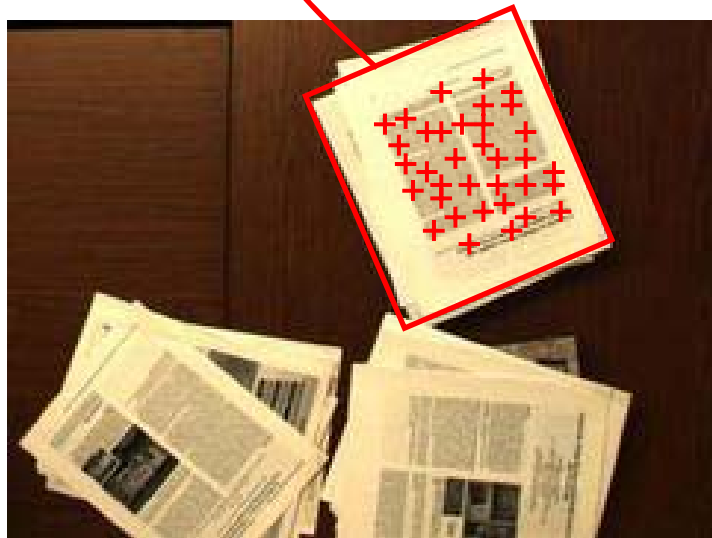
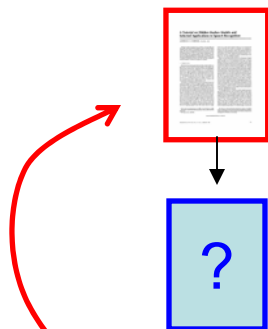


before

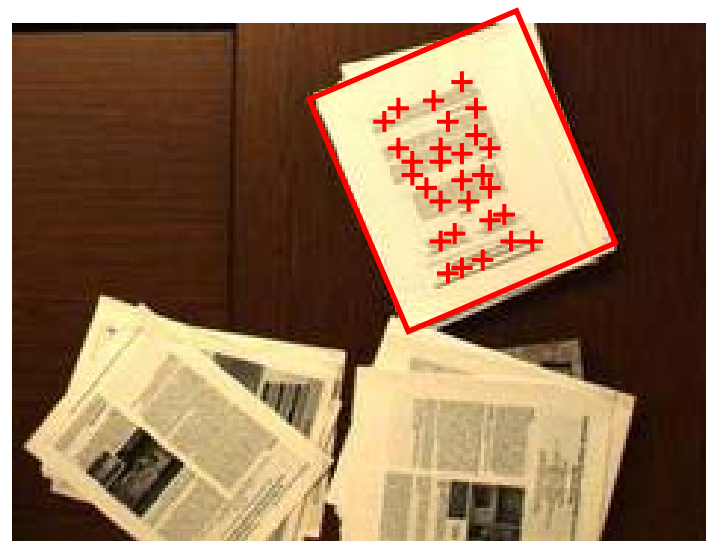


after

Entry vs. Exit

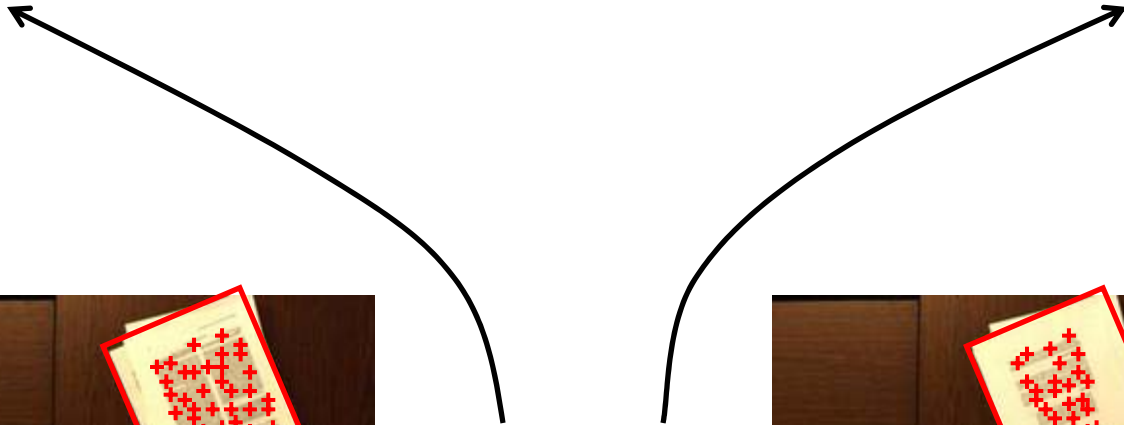


before



after

Entry vs. Exit



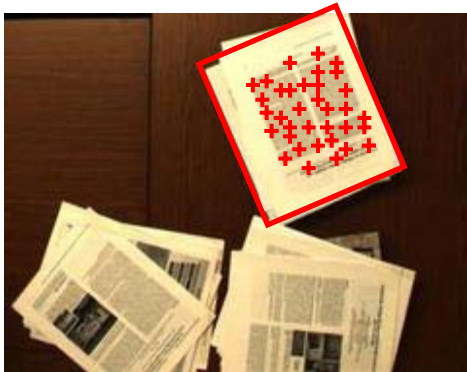
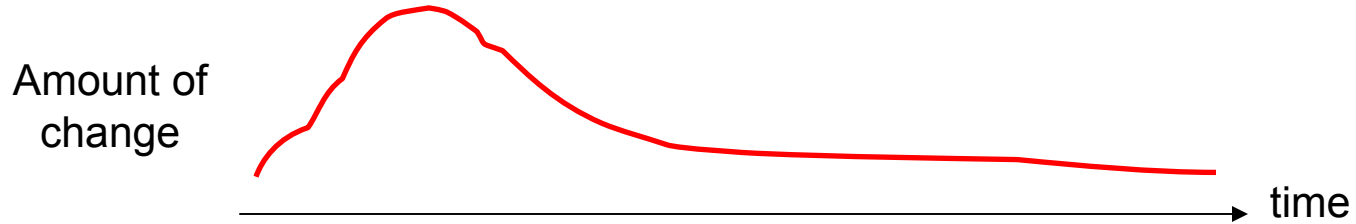
before

.....



after

Entry vs. Exit



before

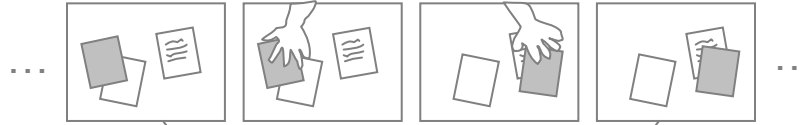
.....



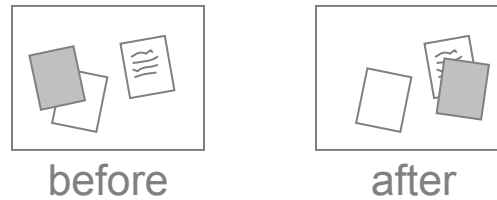
after

Algorithm Overview

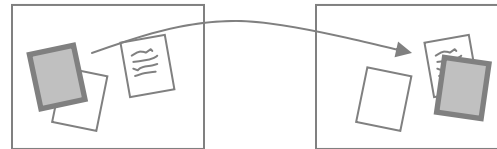
Input
Frames



Event
Detection

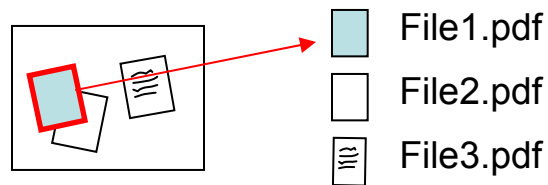


Event
Interpretation



“A document moved
from (x_1, y_1) to (x_2, y_2) ”

✓ Document
Recognition

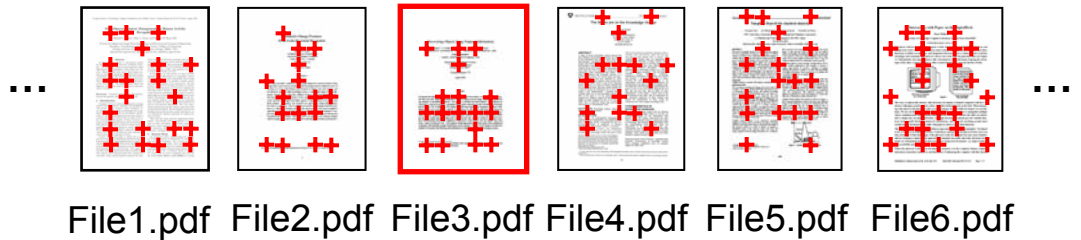
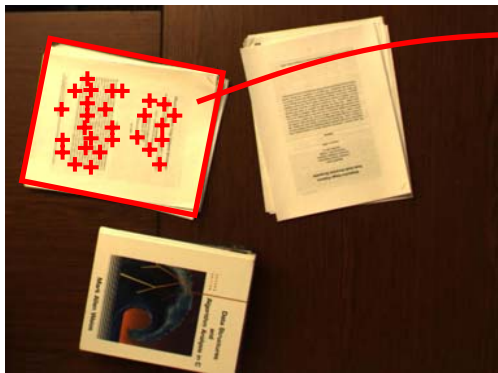


Scene Graph
Update



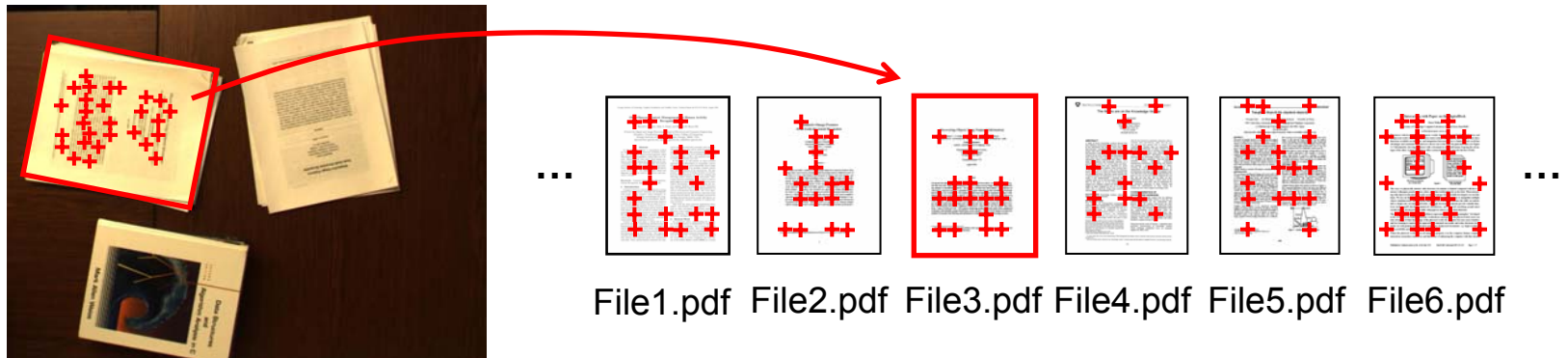
Document Recognition

- Match against PDF image database



Document Recognition

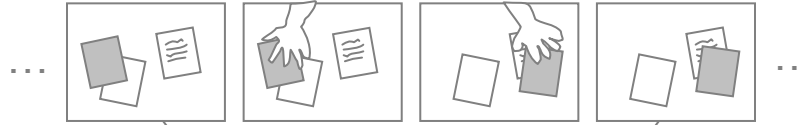
- Match against PDF image database



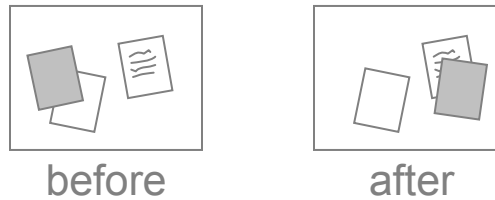
- Performance
 - Can differentiate between ~100 (or more) documents
 - ~200x300 pixels per document for reliable match

Algorithm Overview

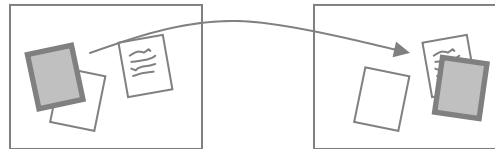
Input
Frames



Event
Detection

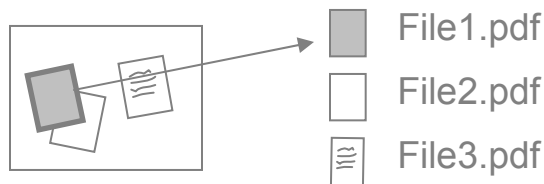


Event
Interpretation

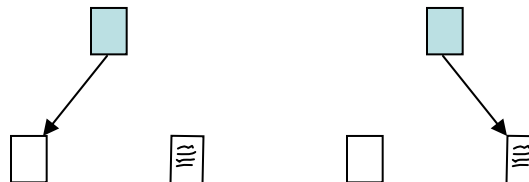


“A document moved
from (x_1, y_1) to (x_2, y_2) ”

Document
Recognition

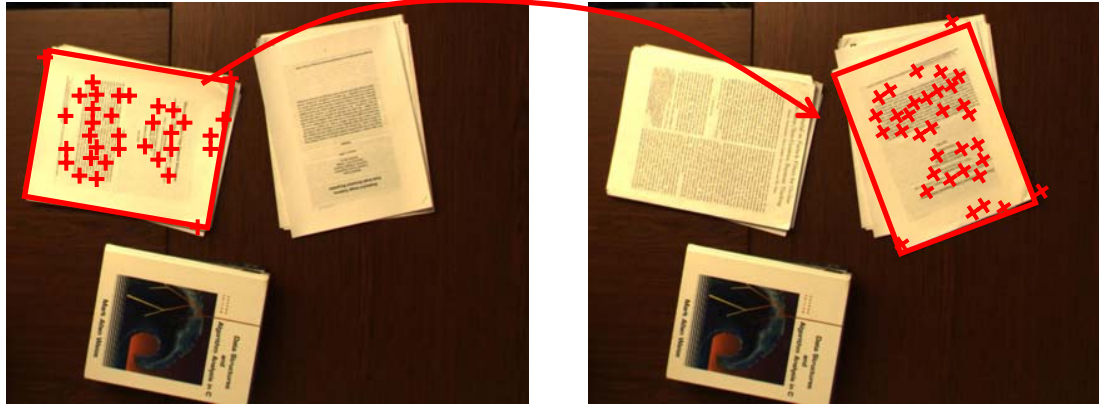


Scene Graph
Update



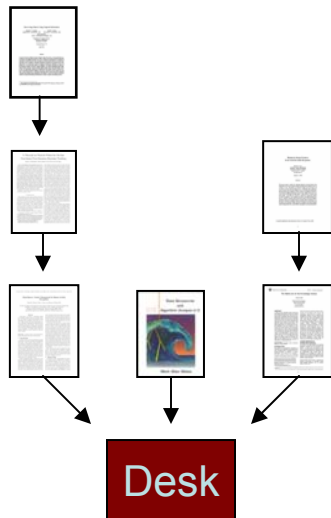
Scene Graph Update

Motion: (x,y,θ)



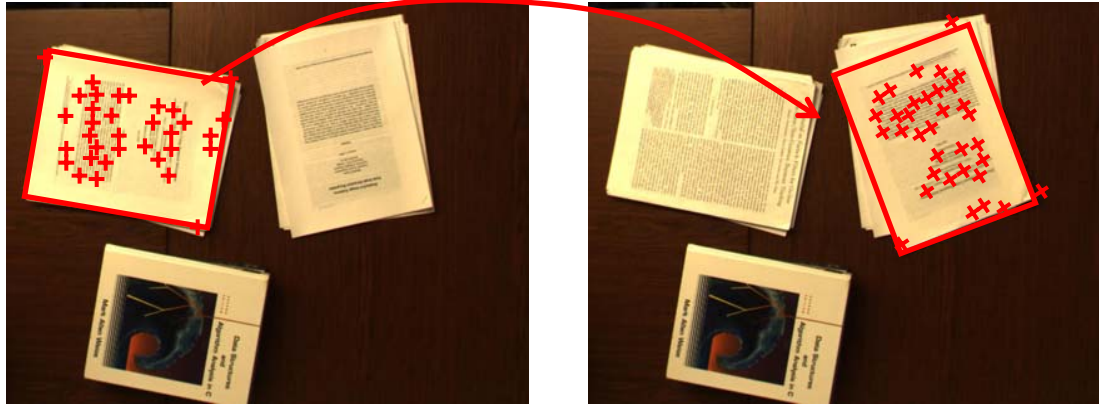
before

after



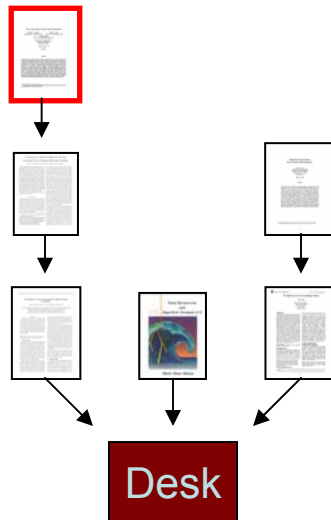
Scene Graph Update

Motion: (x,y,θ)



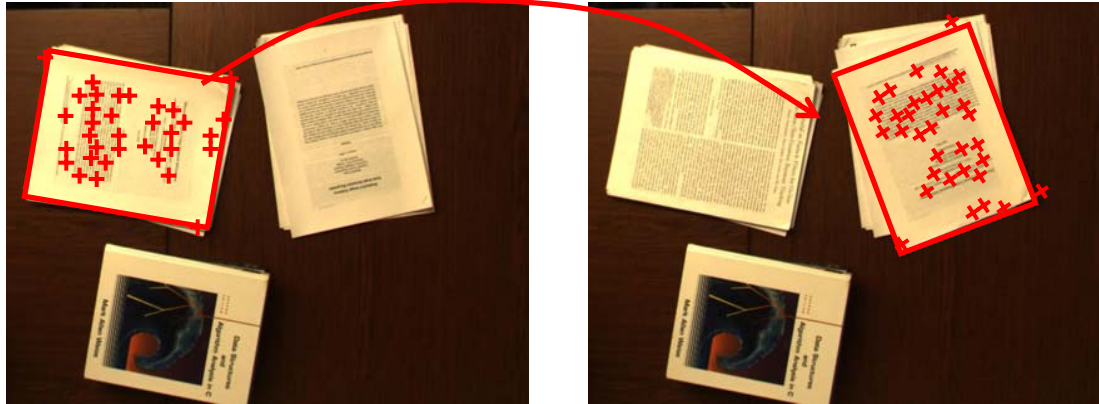
before

after



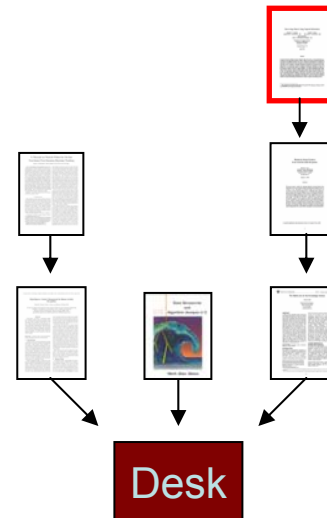
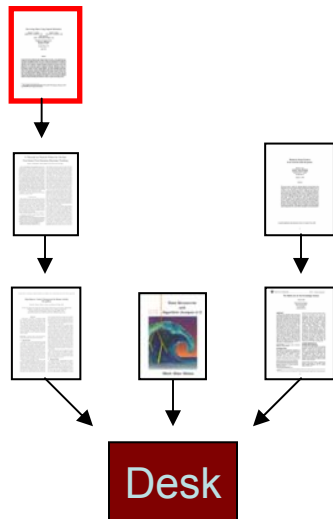
Scene Graph Update

Motion: (x, y, θ)



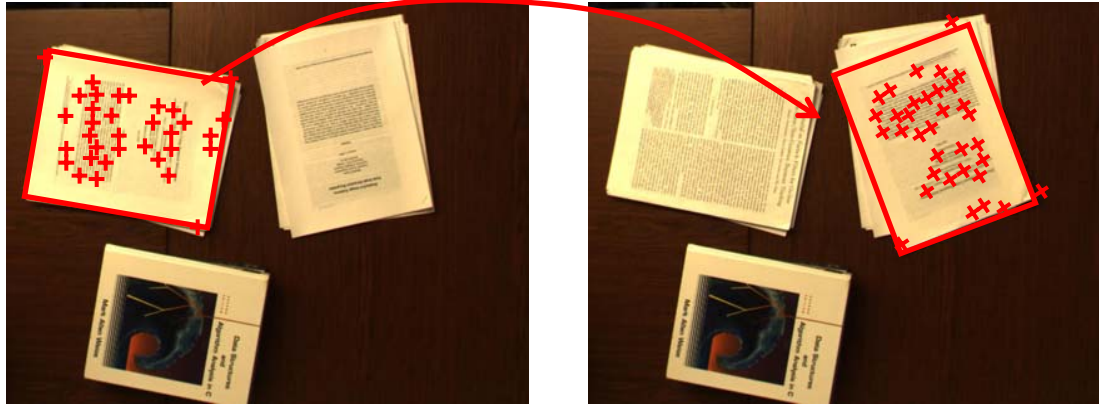
before

after



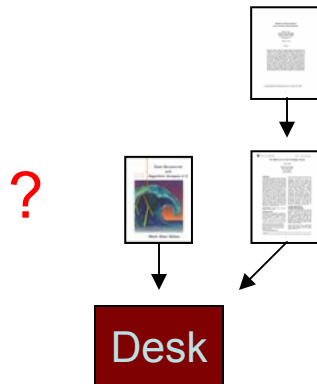
Scene Graph Update

Motion: (x, y, θ)



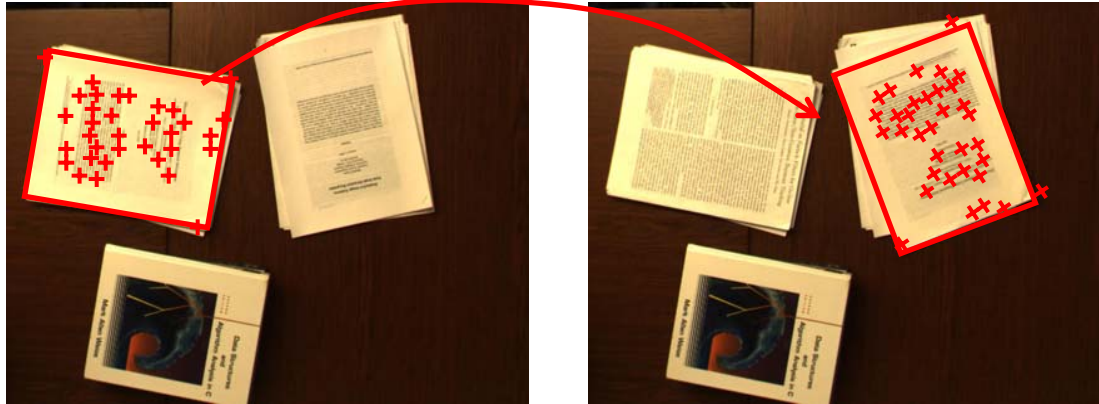
before

after



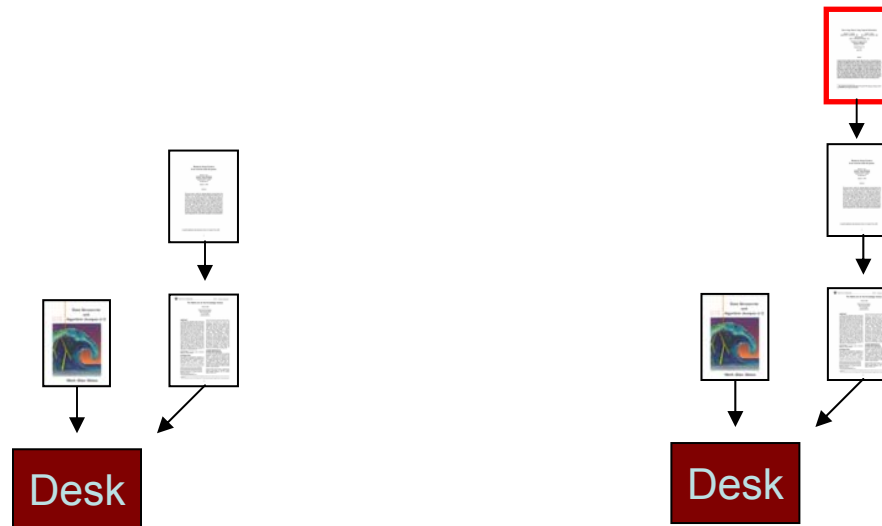
Scene Graph Update

Motion: (x, y, θ)

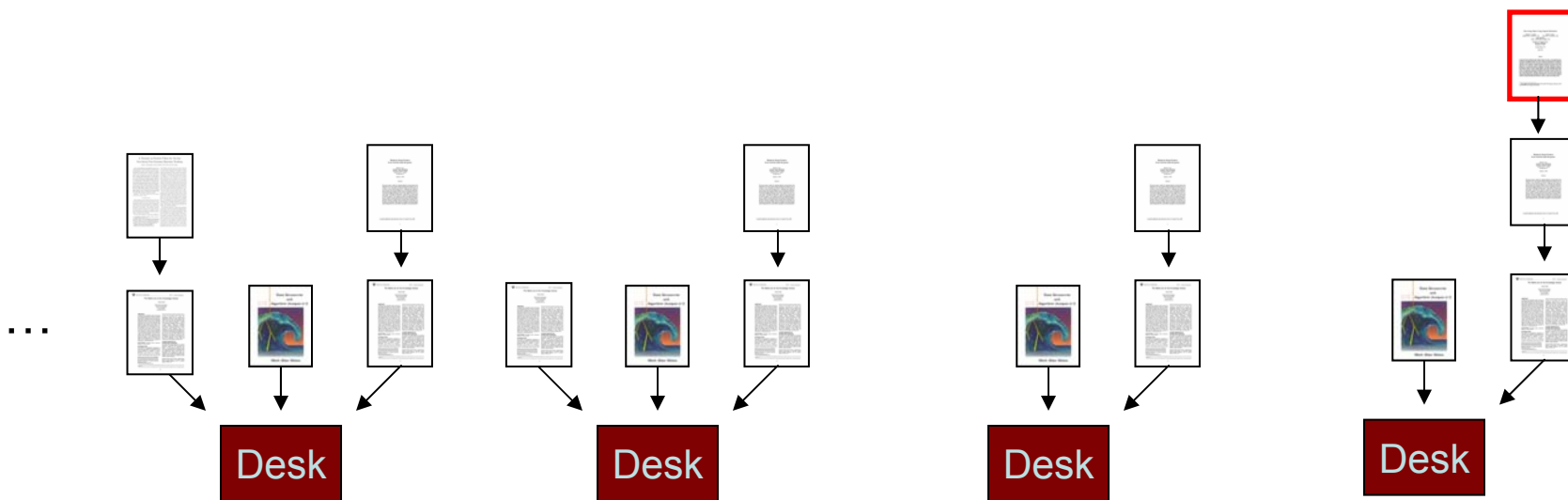


before

after



Scene Graph Update



Scene Graph Update

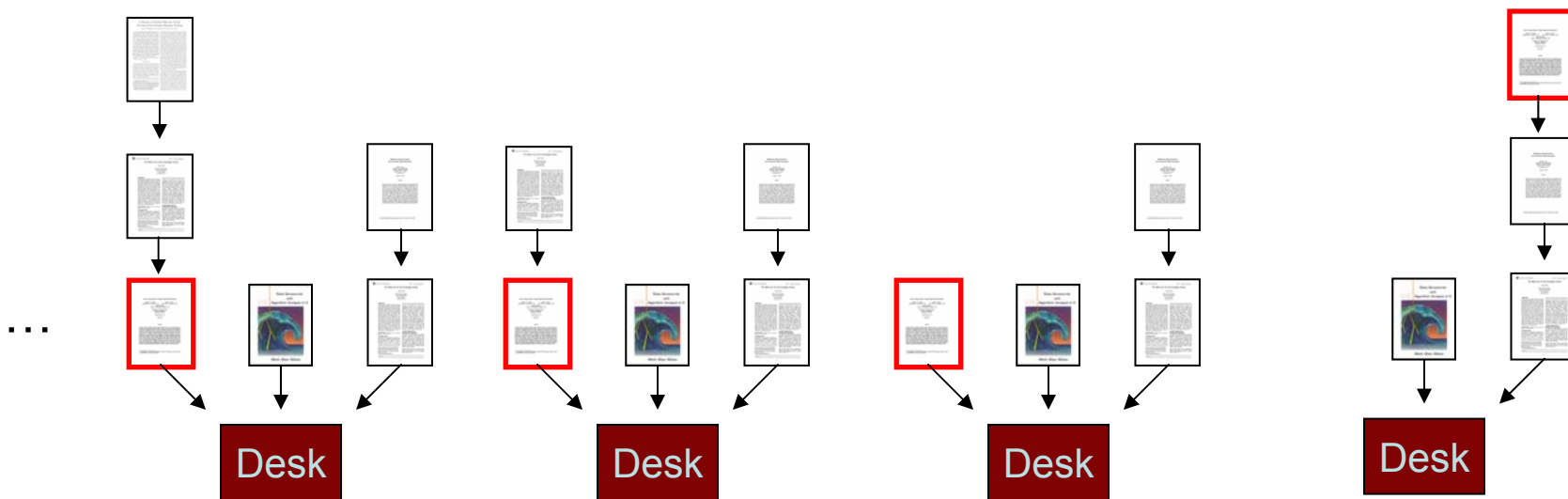
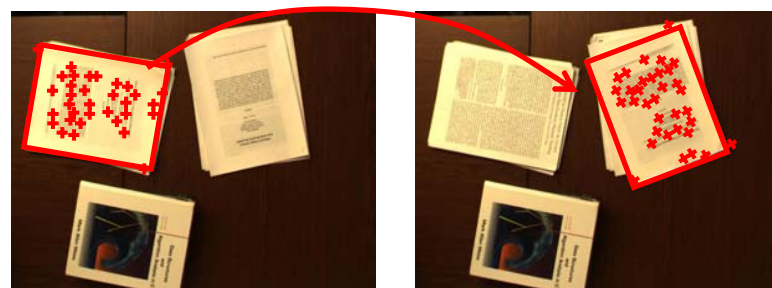
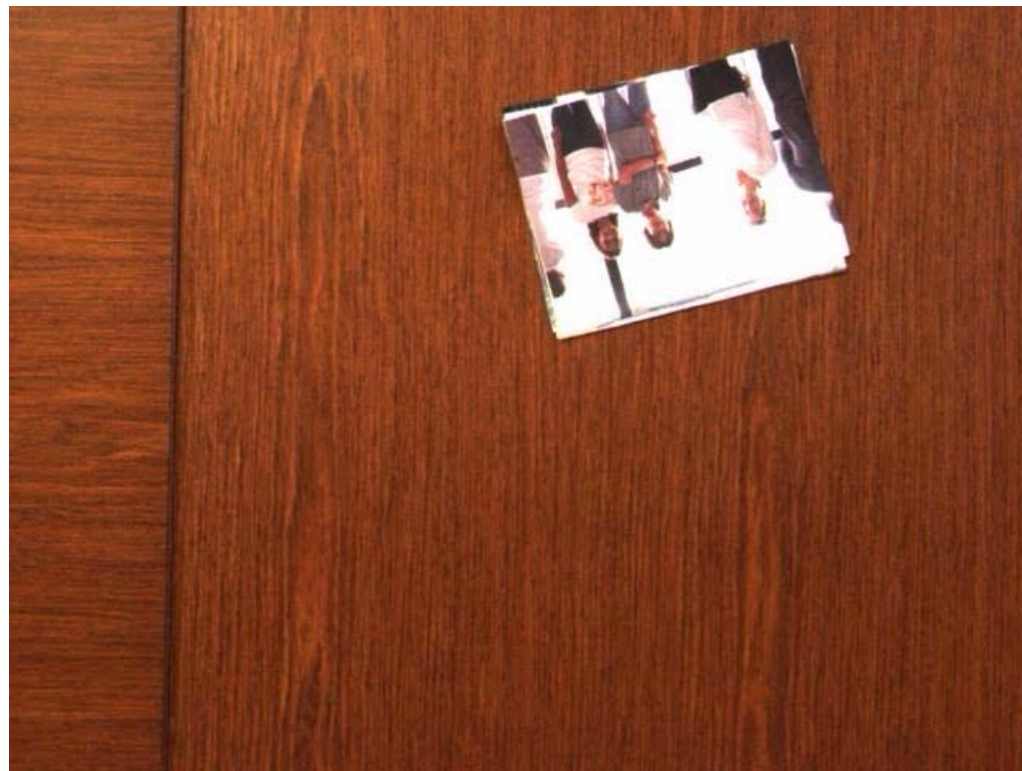


Photo Sorting Example



Current Directions

- Handle more realistic desktops
- Speed up processing time
- Other useful functionalities
 - Written annotation
 - Version management
 - Bookmark
 - Multi-user queries

For More Information

- Our publications

- Jiwon Kim, Steven M. Seitz, Maneesh Agrawala. The Office of the Past. *IEEE Workshop on Real-time Vision for HCI*, 2004.
- Jiwon Kim, Steven M. Seitz, Maneesh Agrawala. Video-based Document Tracking: Unifying Your Physical and Electronic Desktops. To appear in *Proceedings of UIST*, 2004.

- Other related work

- David G. Lowe. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 2004.
- Pierre Wellner. Interacting with paper on the DigitalDesk. *Communications of the ACM*, 36(7):86.97, 1993.
- D. Rus and P. deSantis. The self-organizing desk. In *Proceedings of International Joint Conference on Artificial Intelligence*, 1997.

For More Information

- Our publications

- Jiwon Kim, Steven M. Seitz, Maneesh Agrawala. The Office of the Past. *IEEE Workshop on Real-time Vision for HCI*, 2004.
- Jiwon Kim, Steven M. Seitz, Maneesh Agrawala. Video-based Document Tracking: Unifying Your Physical and Electronic Desktops. To appear in *Proceedings of UIST*, 2004.

- Other related work

- David G. Lowe. Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 1999.
- Pierre Wellner. Interacting with paper on the DigitalDesk. *Communications of the ACM*, 36(7):86.97, 1993.
- D. Rus and P. deSantis. The self-organizing desk. In *Proceedings of International Joint Conference on Artificial Intelligence*, 1997.