

Video-Based Document Tracking: Unifying Your Physical and Electronic Desktops

Jiwon Kim Steven M. Seitz Maneesh Agrawala
University of Washington Microsoft Research

Motivation



Unifying physical and electronic desktops

Video camera



Record video of paper
on physical desktop

Unifying physical and electronic desktops

Video camera



Record video of paper
on physical desktop
– Tracking

Unifying physical and electronic desktops

Video camera



Record video of paper on physical desktop

- Tracking
- Recognition

Unifying physical and electronic desktops

Video camera



Record video of paper on physical desktop

- Tracking
- Recognition
- Linking

Applications

Video camera



Desktop

Applications

Video camera



- Find lost document

Applications

Video camera



- Find lost document
- Browse remote desk

Applications

Video camera



- Find lost document
- Browse remote desk
- Find electronic version

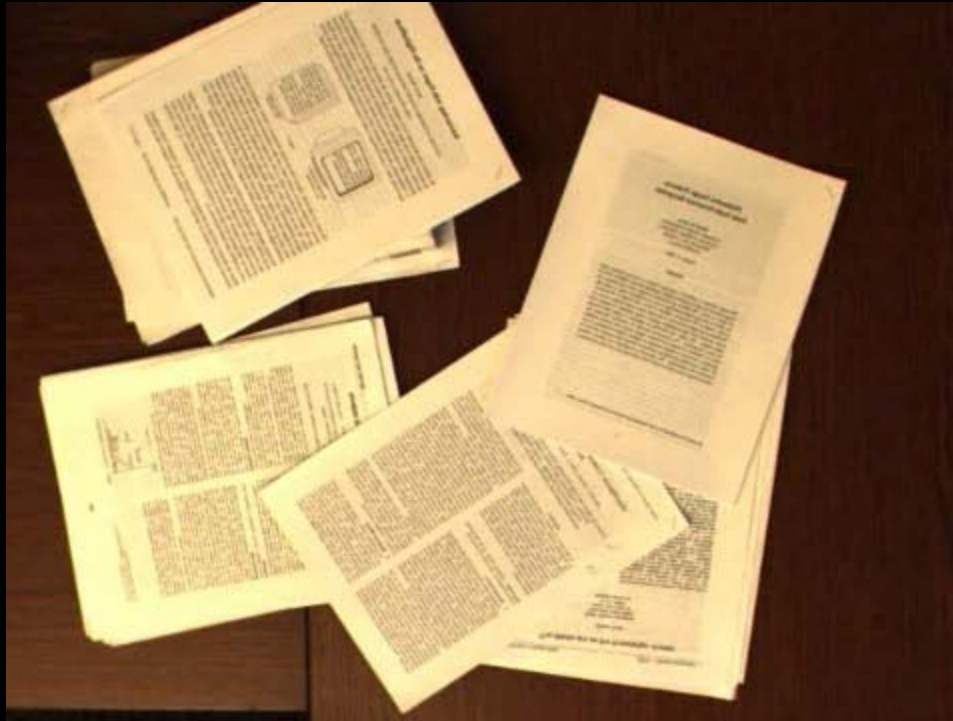
Applications

Video camera



- Find lost document
- Browse remote desk
- Find electronic version
- History-based queries

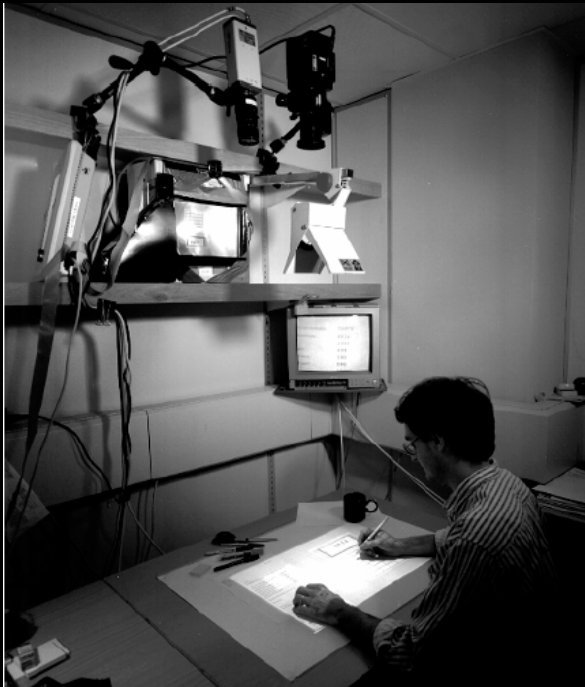
Example Input Video



Demo – Remote Desktop

Related Work

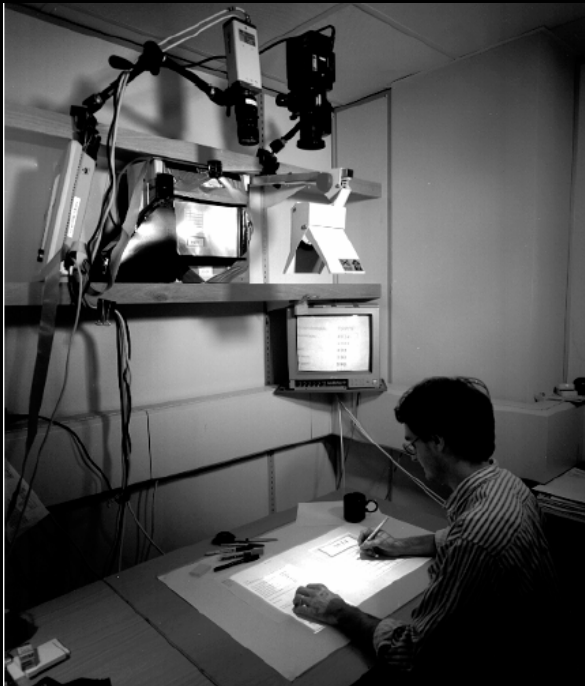
- Interactive desktops



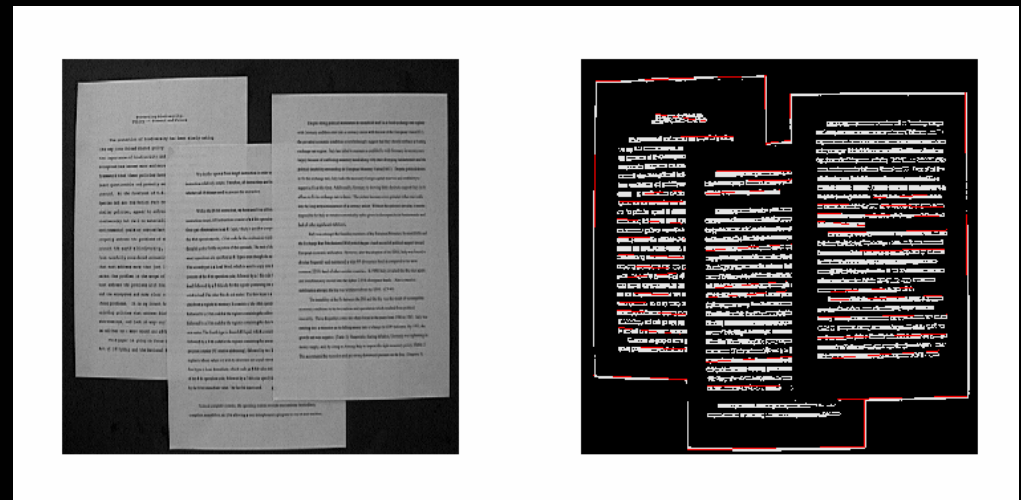
DigitalDesk [Wellner 93]

Related Work

- Interactive desktops



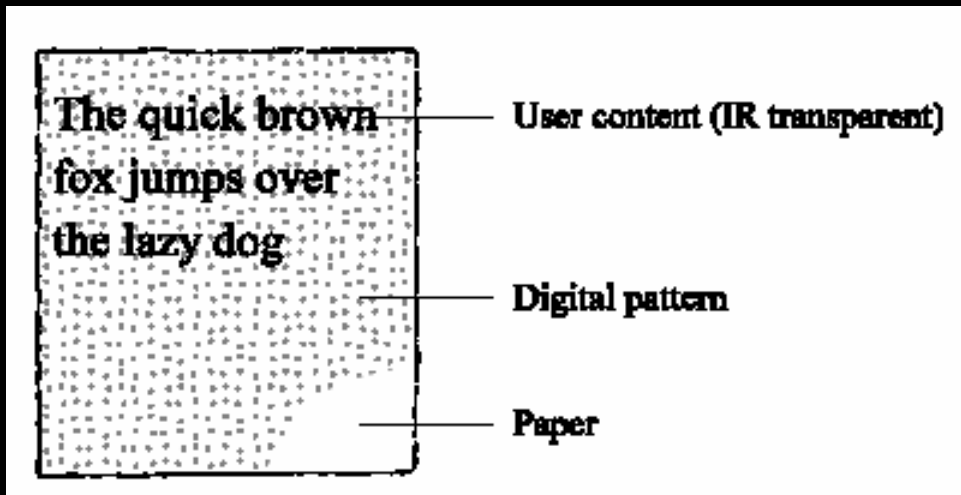
DigitalDesk [Wellner 93]



Self-Organizing Desk
[Rus et al. 97]

Related Work

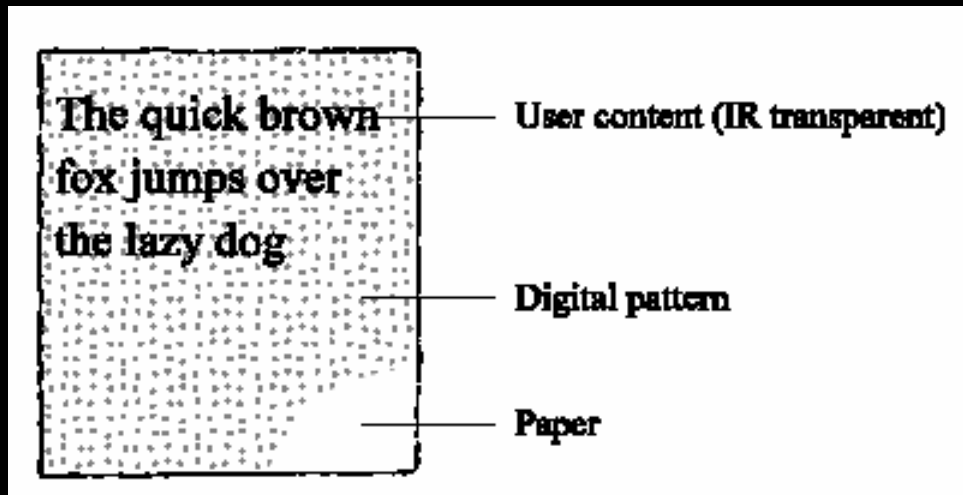
- Interactive desktops
- **Augmented paper**



PADD [Guimbretière 03]

Related Work

- Interactive desktops
- **Augmented paper**



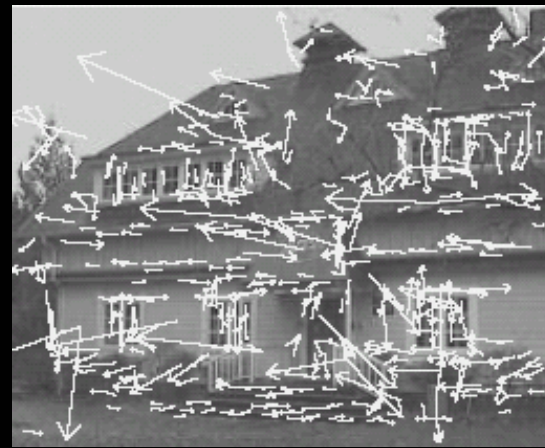
PADD [Guimbretière 03]



**CyberCode
[Rekimoto et al. 00]**

Related Work

- Interactive desktops
- Alternative media
- Object tracking & recognition



SIFT [Lowe 04]

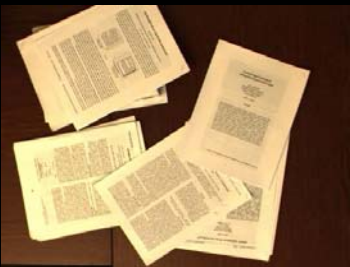
System Overview



System Overview



Video of desk

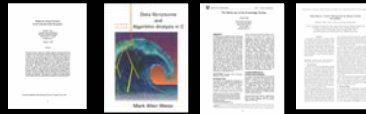


System Overview



Video of desk

Images from PDF



System Overview

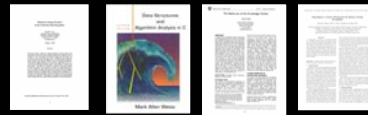


Video of desk

Images from PDF



Track & recognize

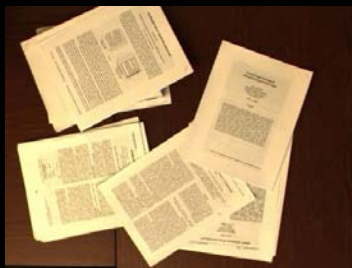


System Overview

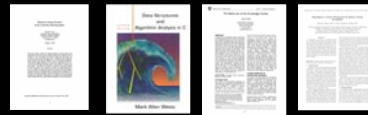


Video of desk

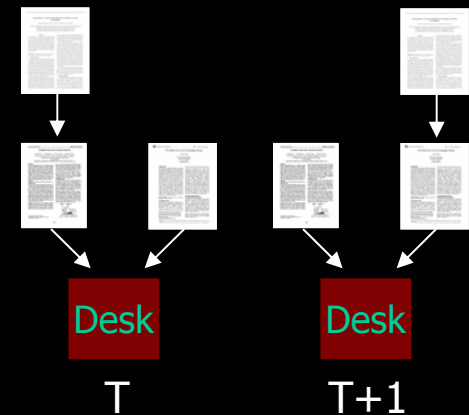
Images from PDF



Track & recognize



Internal representation

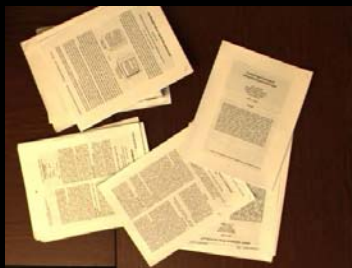


System Overview

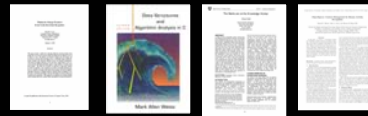


Video of desk

Images from PDF

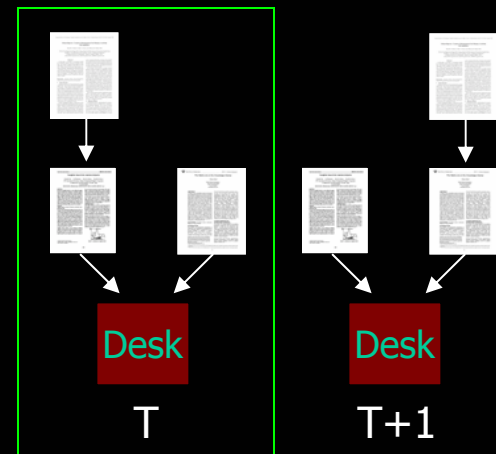


Track & recognize



Scene Graph

Internal representation



System Overview

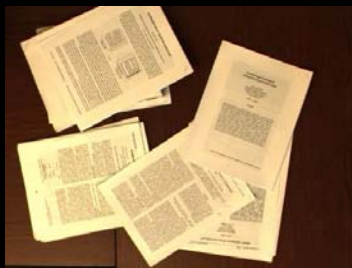


Where is my W-2?

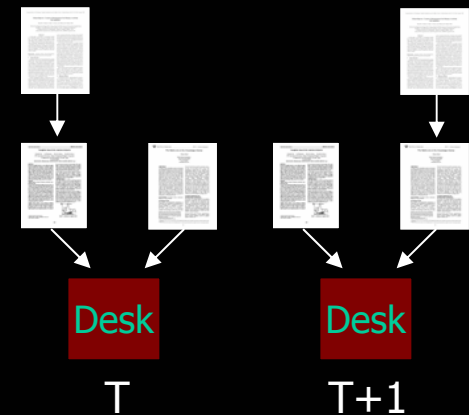
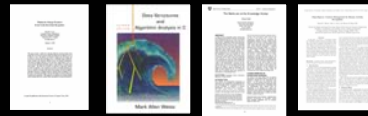
Video of desk

Images from PDF

Internal representation



Track & recognize



System Overview



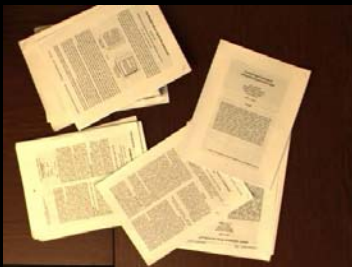
Where is my W-2?

Answer

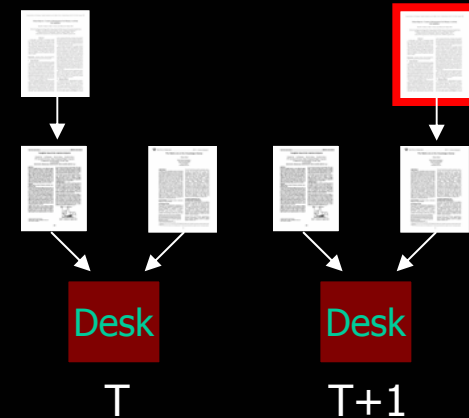
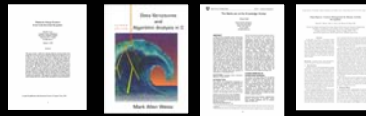
Video of desk

Images from PDF

Internal representation



Track & recognize



System Overview

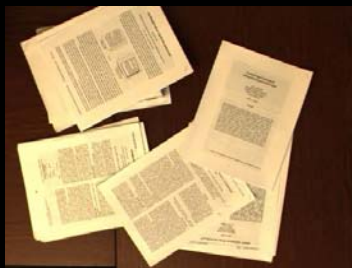


Where is my W-2?

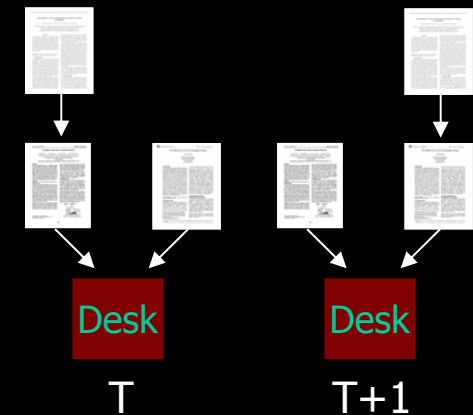
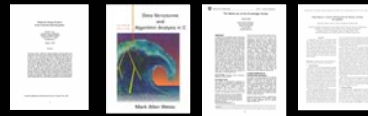
Video of desk

Images from PDF

Internal representation



Track & recognize



Tracking & Recognition



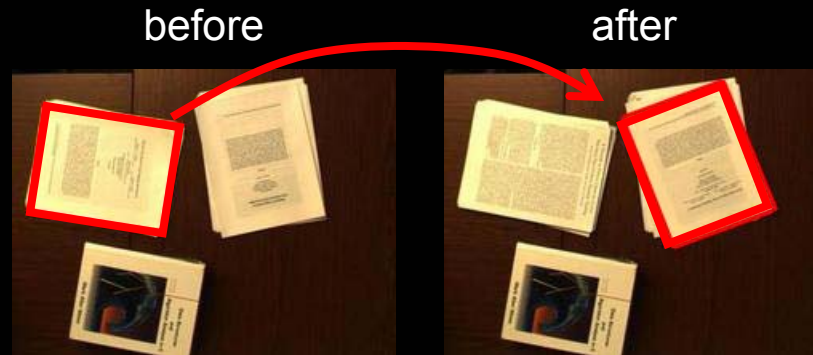
Tracking & Recognition

Event

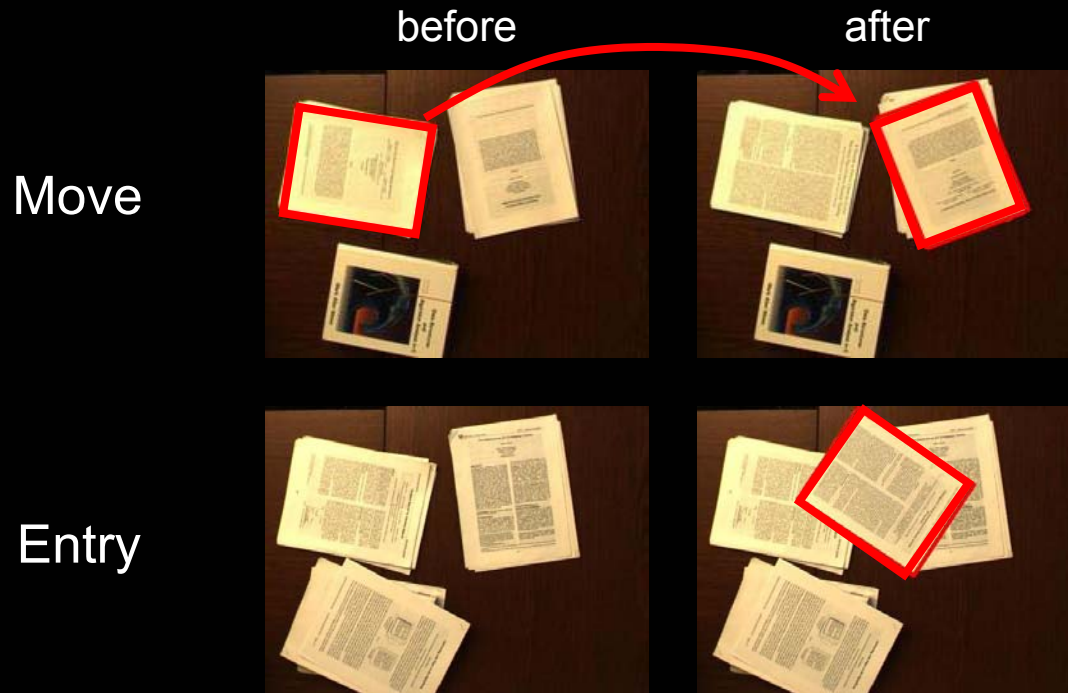


Event Types

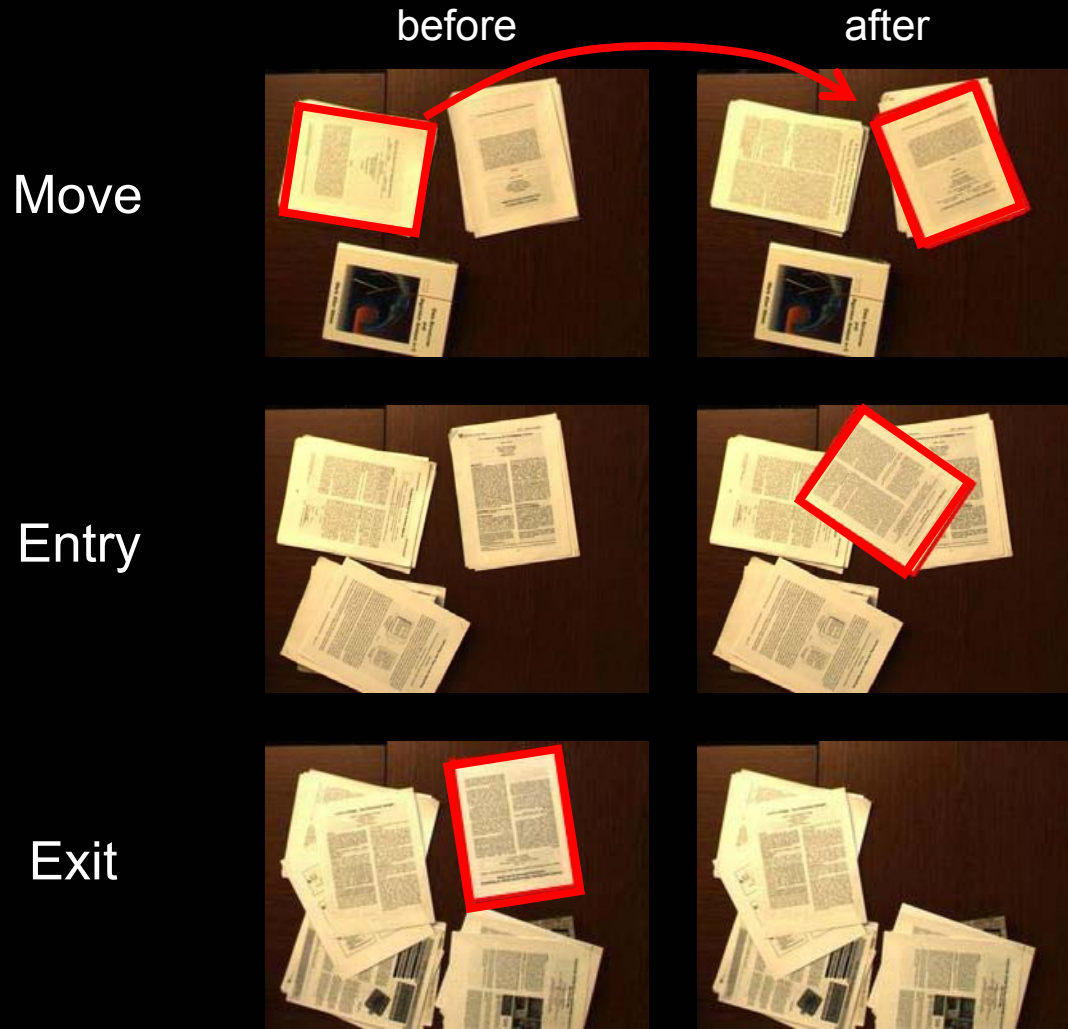
Move



Event Types

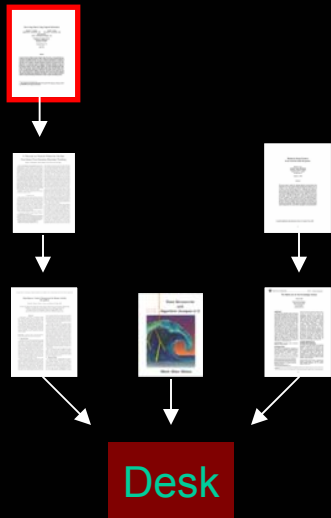


Event Types



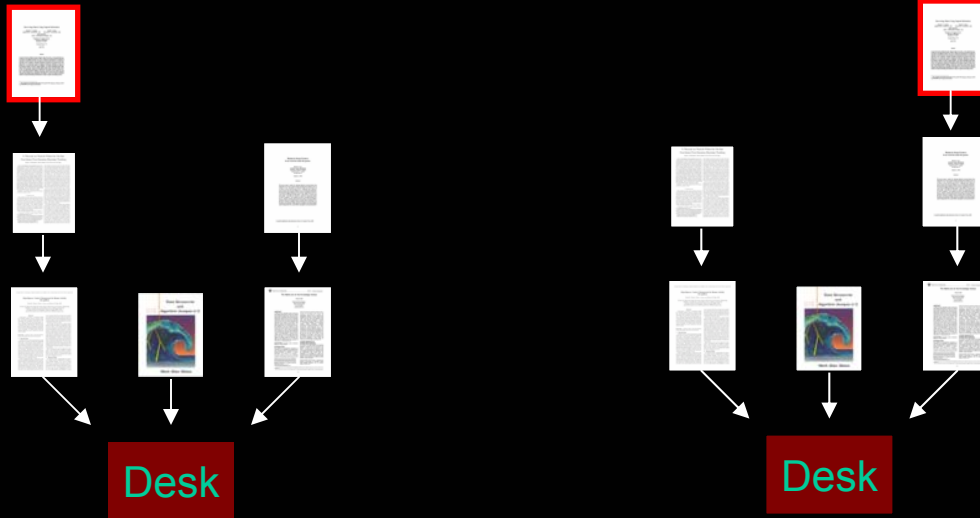
Tracking & Recognition

Event



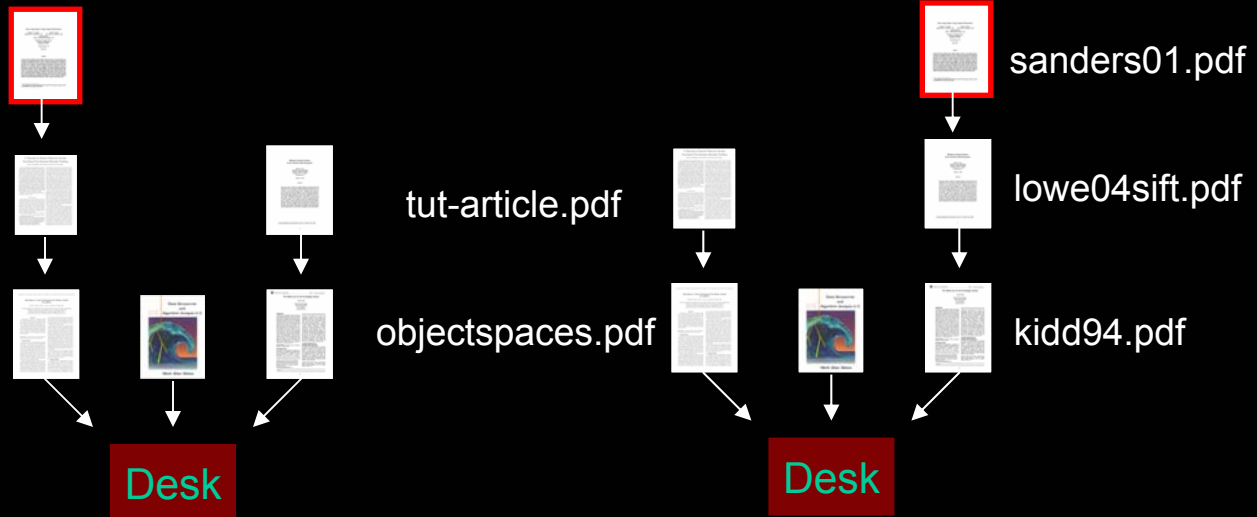
Tracking & Recognition

Event



Tracking & Recognition

Event



Assumptions

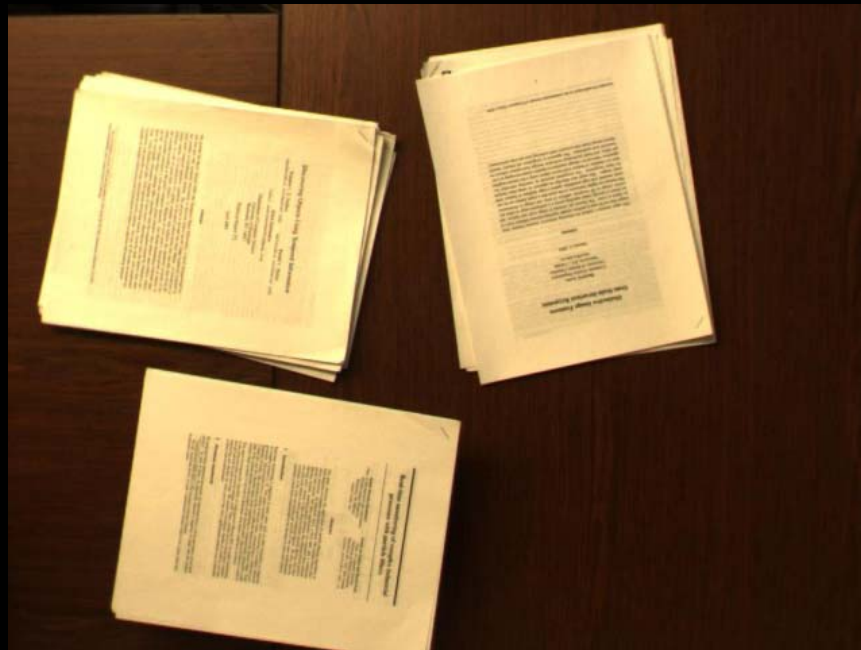
- Document
 - Corresponding electronic copy exists
 - No duplicates of same document

Assumptions

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

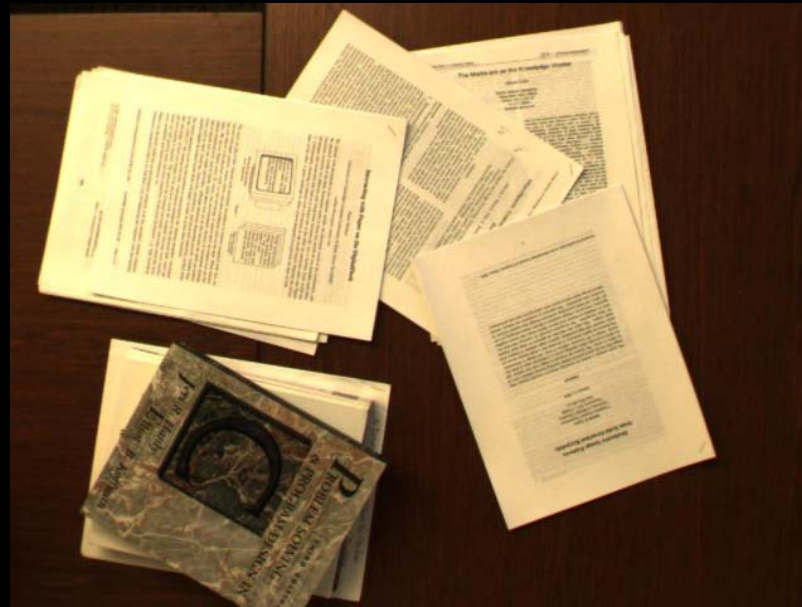
Non-Assumptions

- Desk need not be initially empty



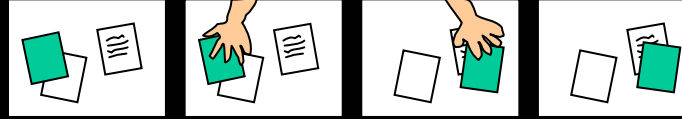
Non-Assumptions

- Desk need not be initially empty
- Stacks may overlap

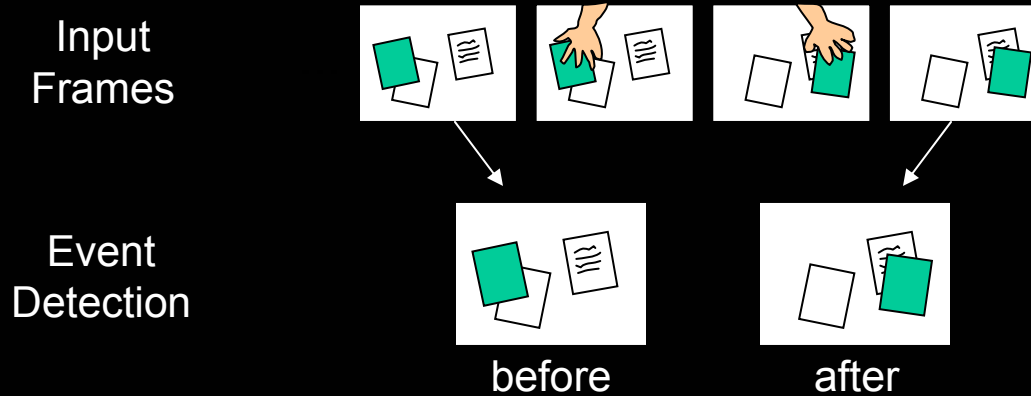


Algorithm Overview

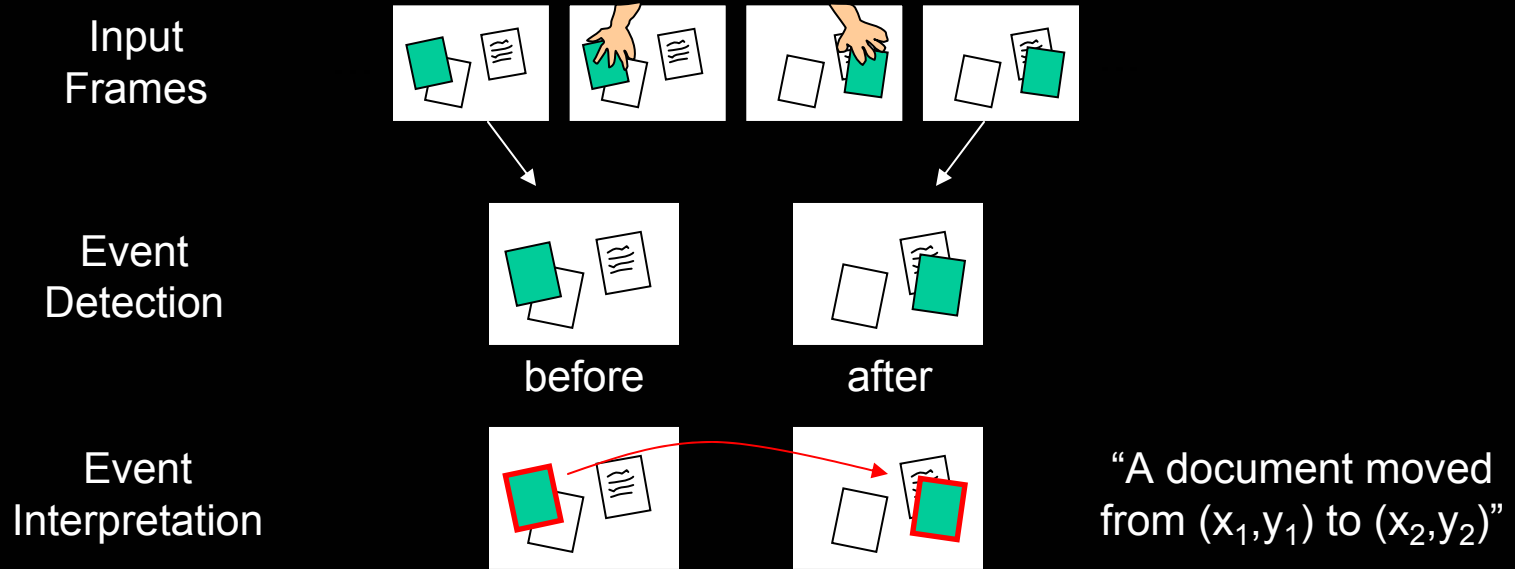
Input
Frames



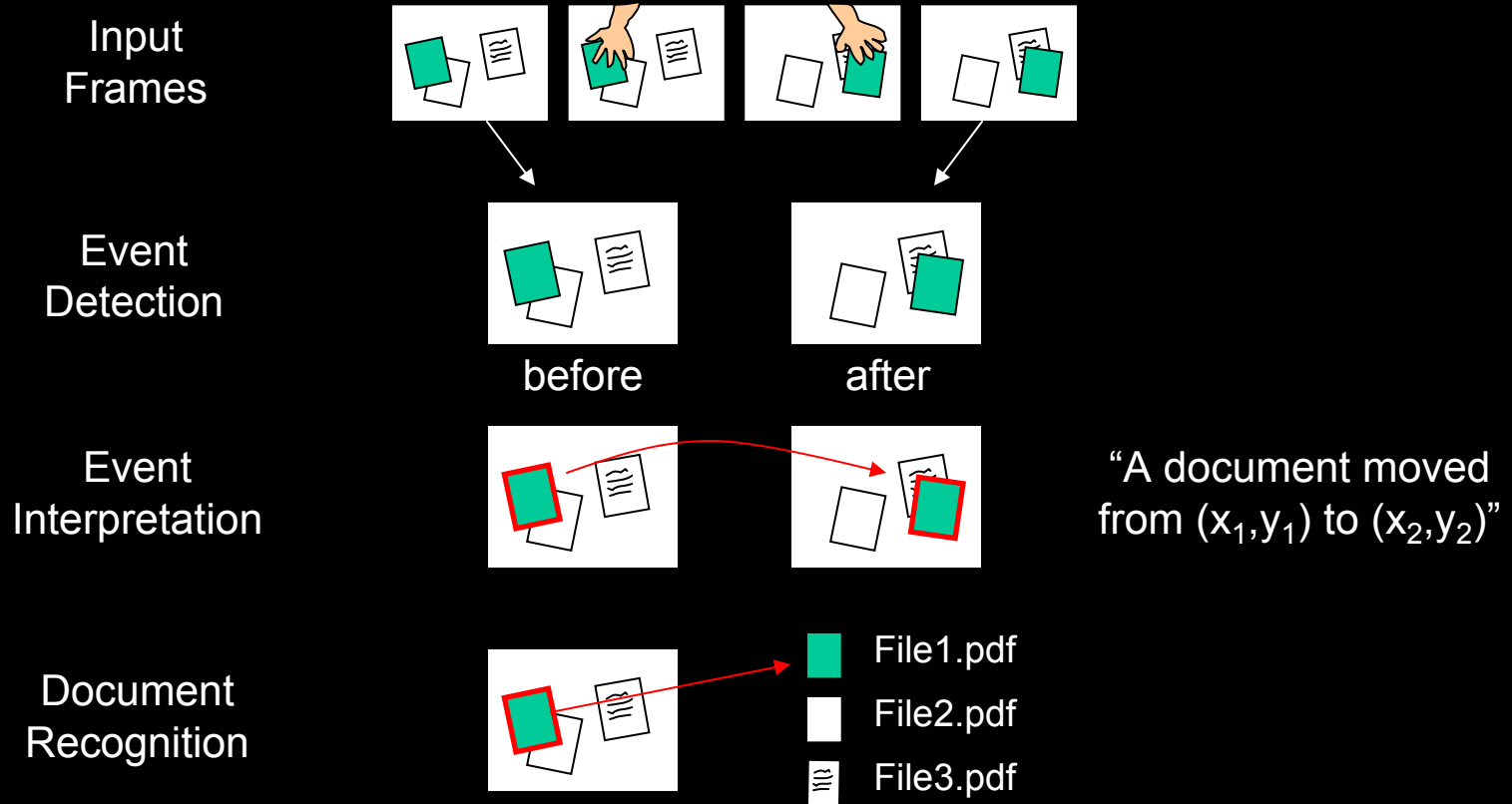
Algorithm Overview



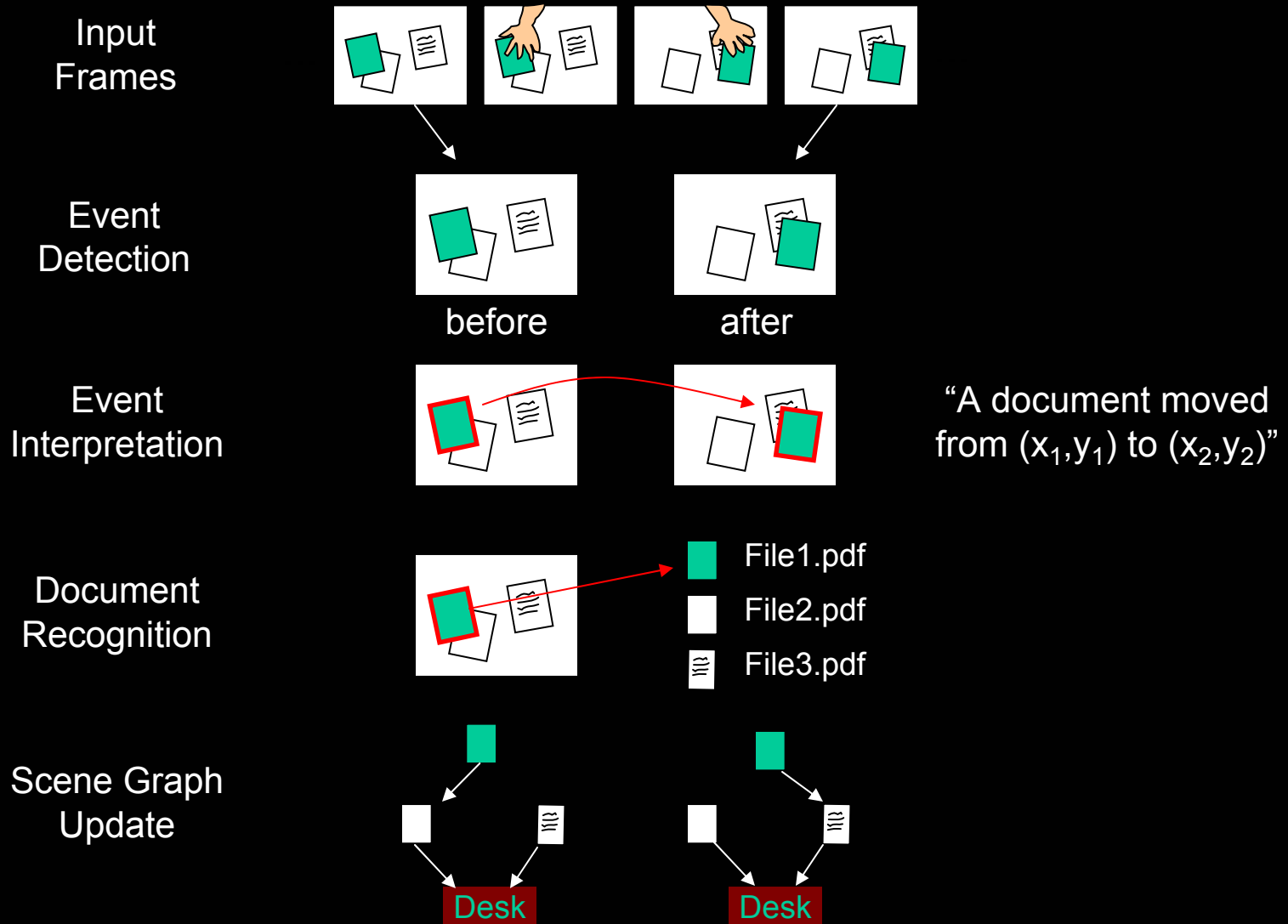
Algorithm Overview



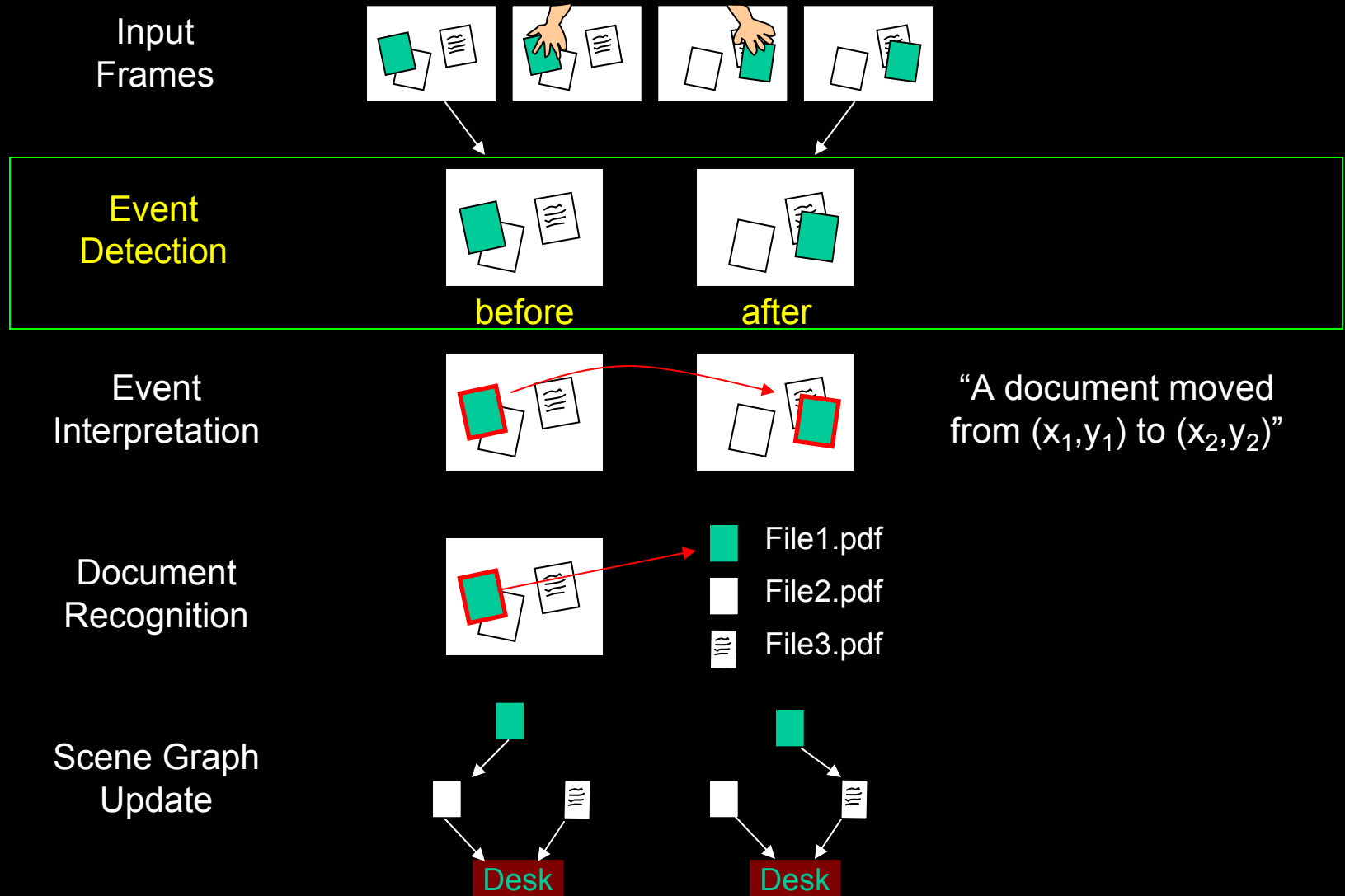
Algorithm Overview



Algorithm Overview



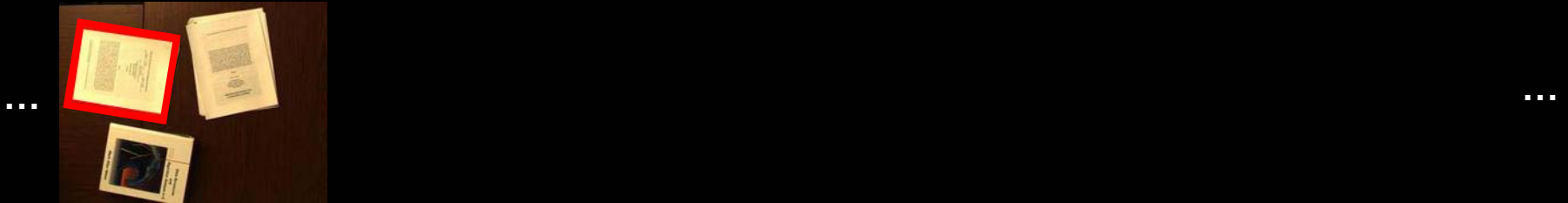
Algorithm Overview



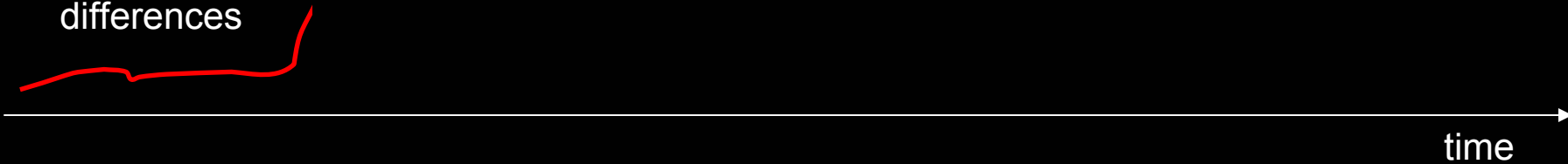
Event Detection



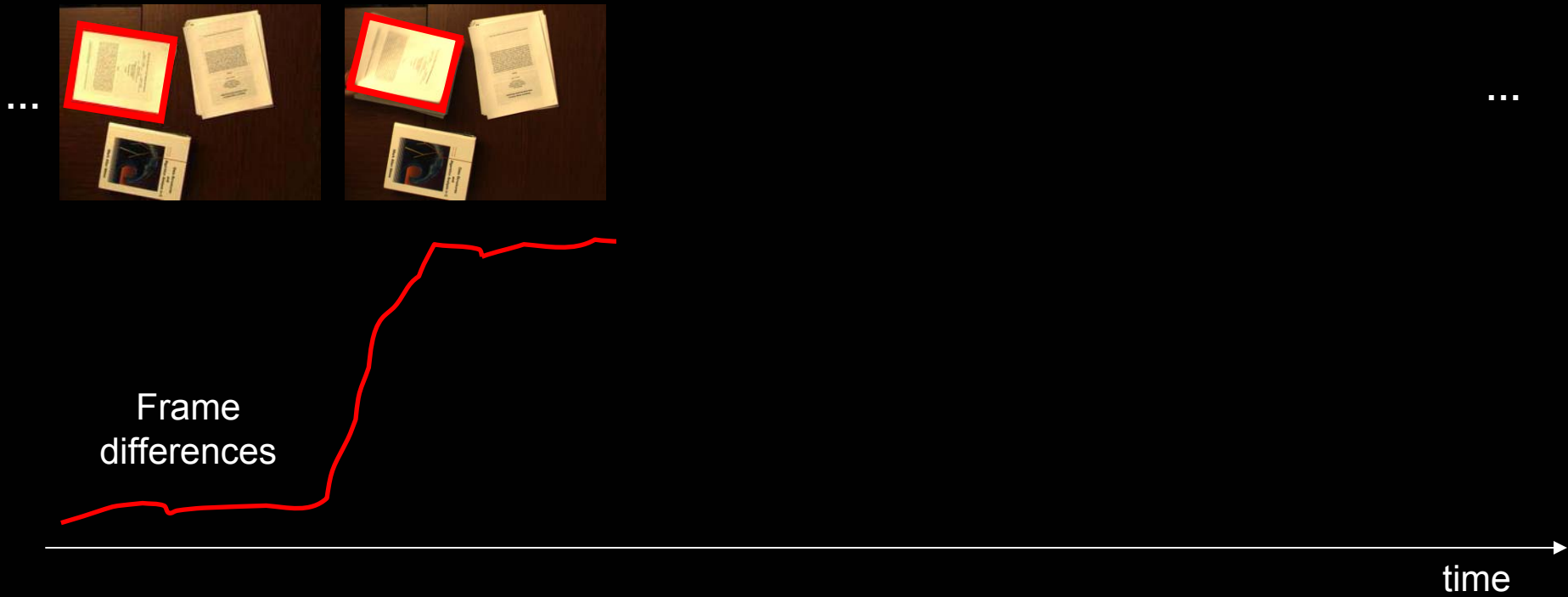
Event Detection



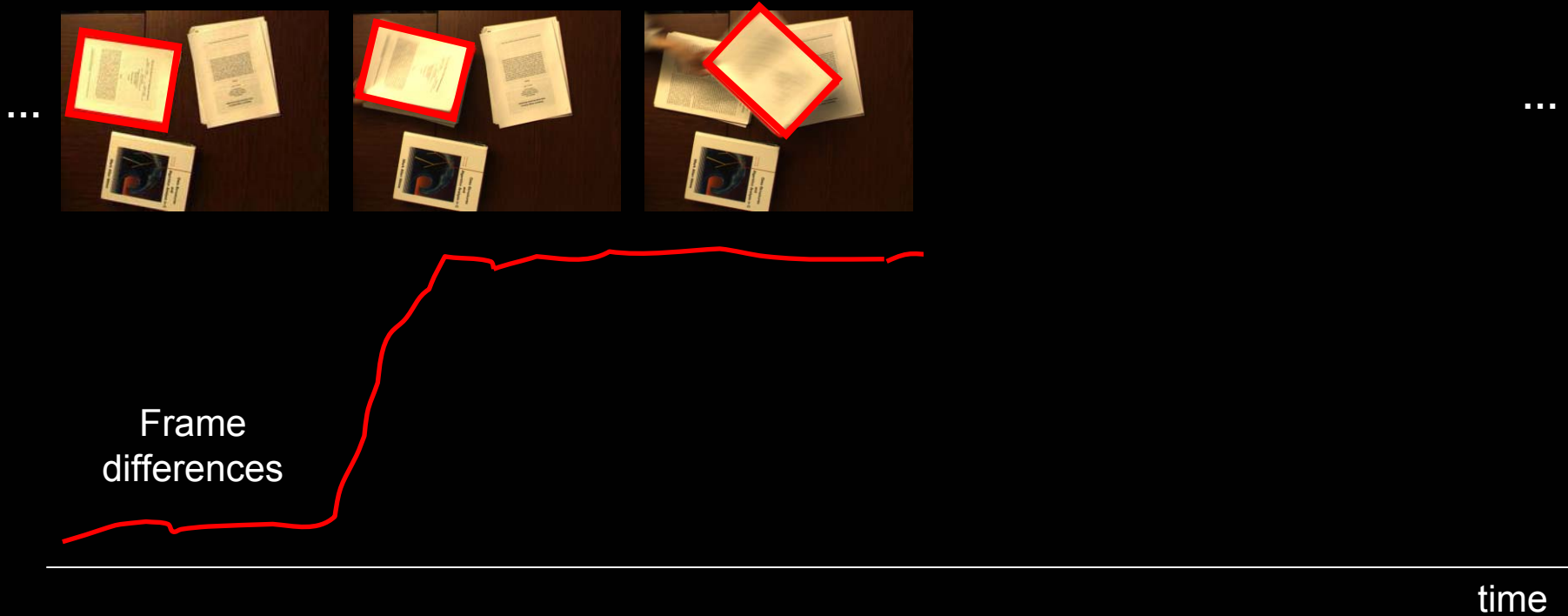
Frame
differences



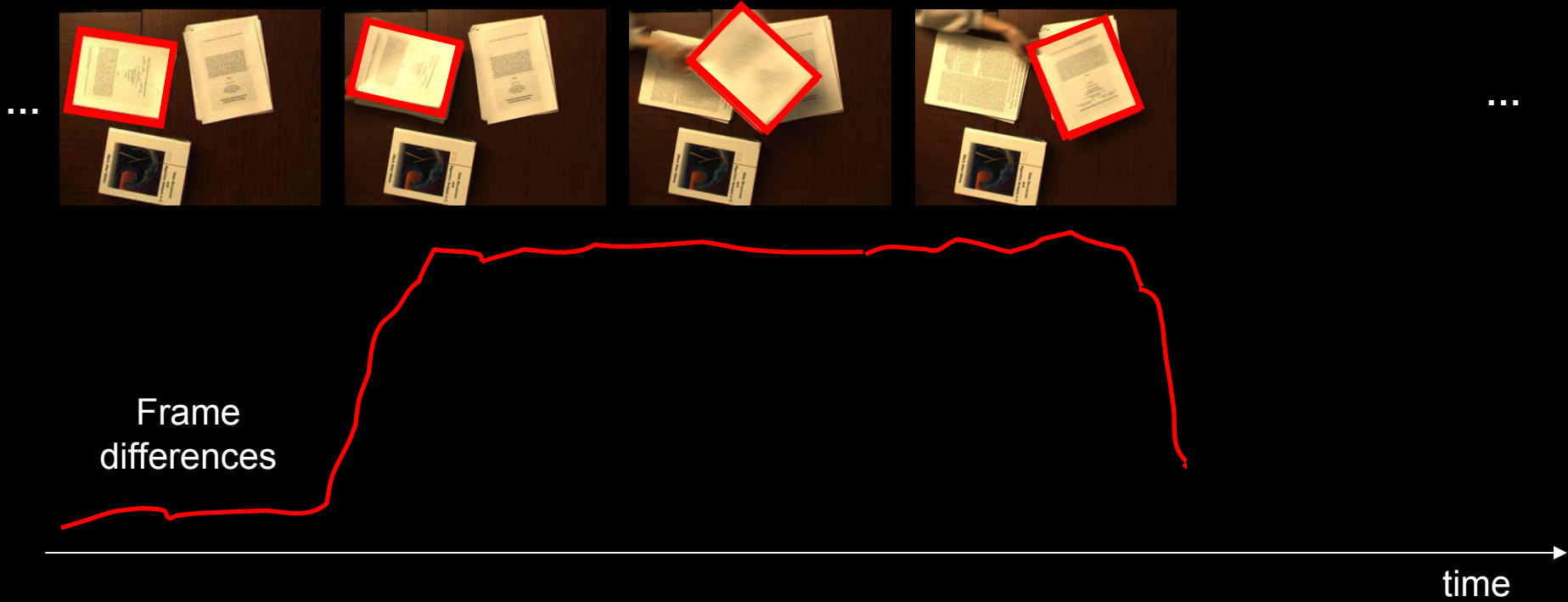
Event Detection



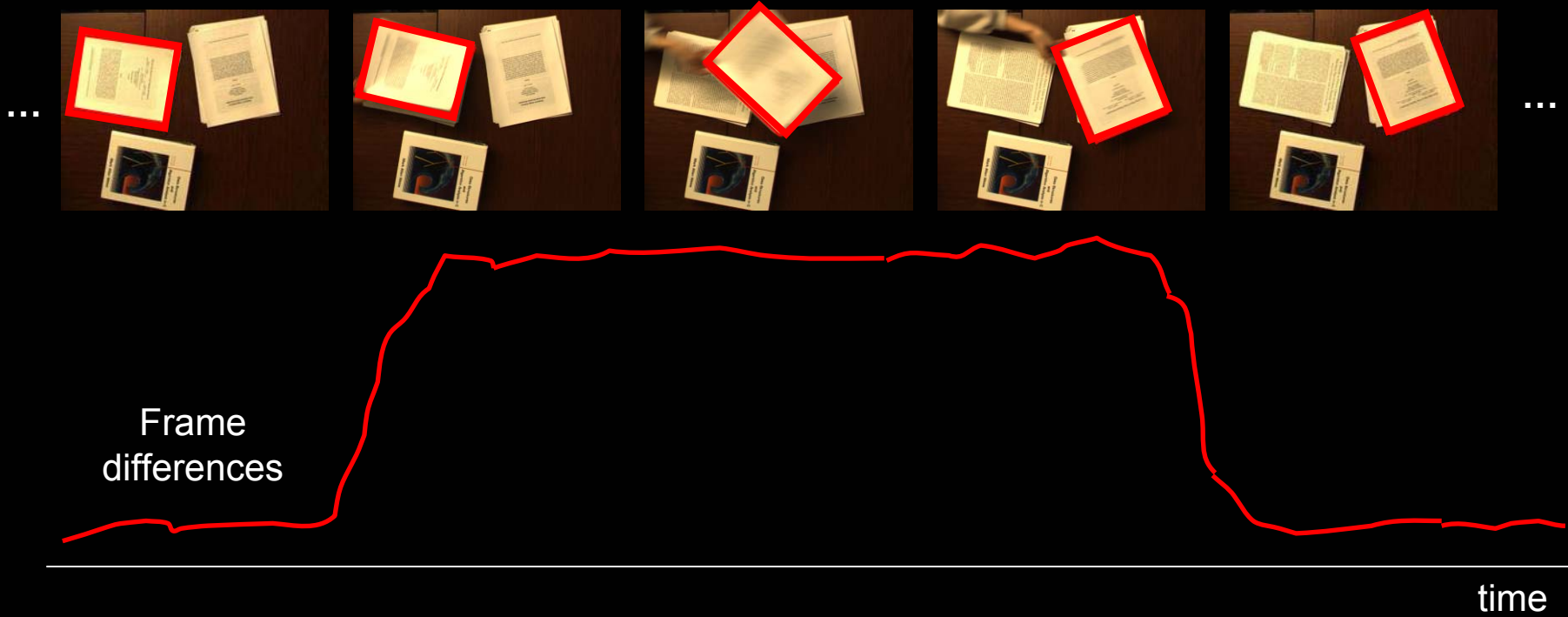
Event Detection



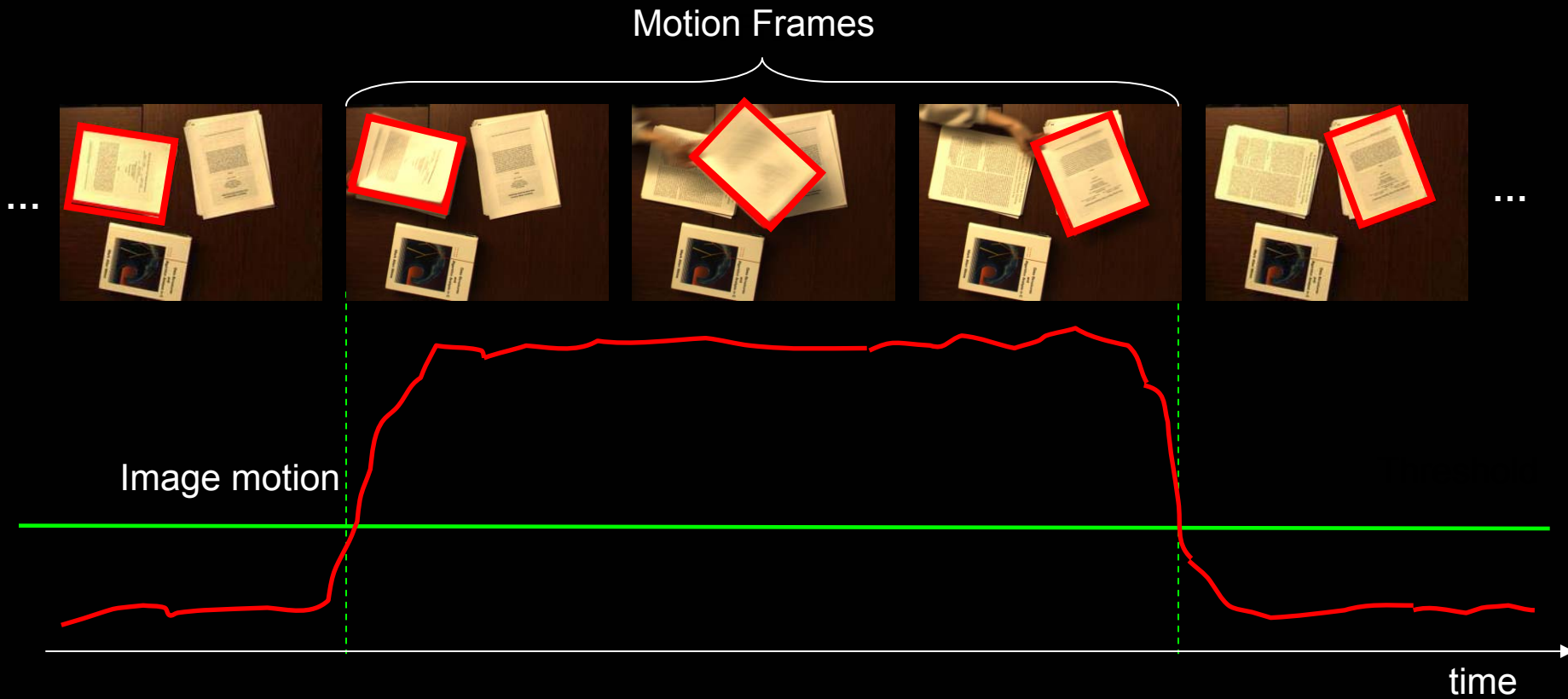
Event Detection



Event Detection

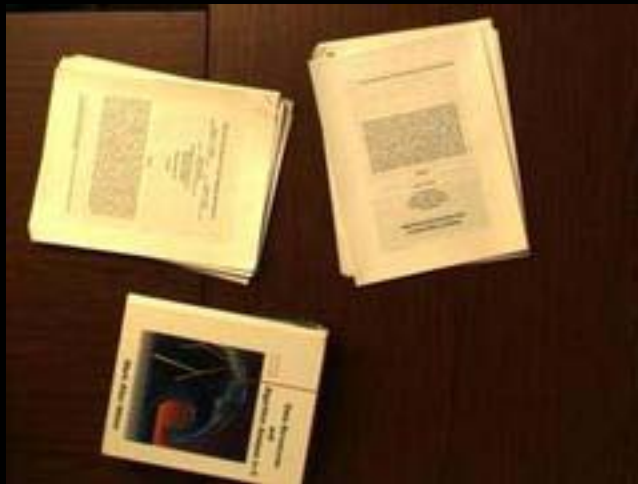
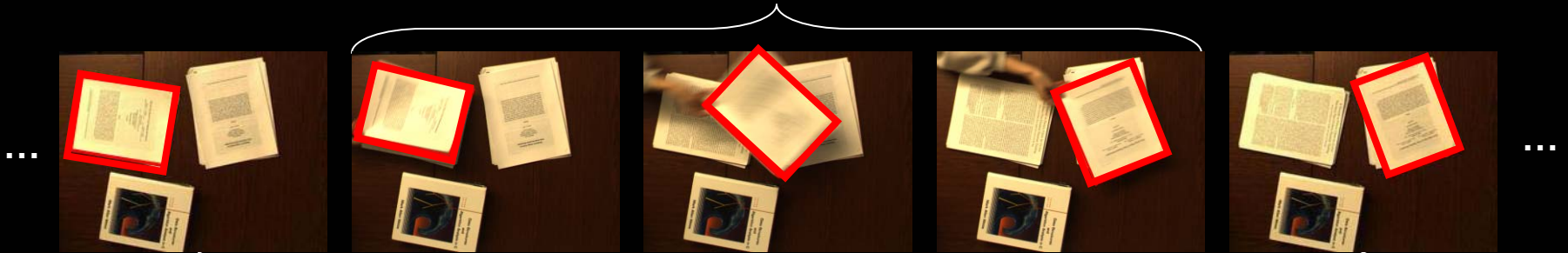


Event Detection



Event Detection

Motion Frames

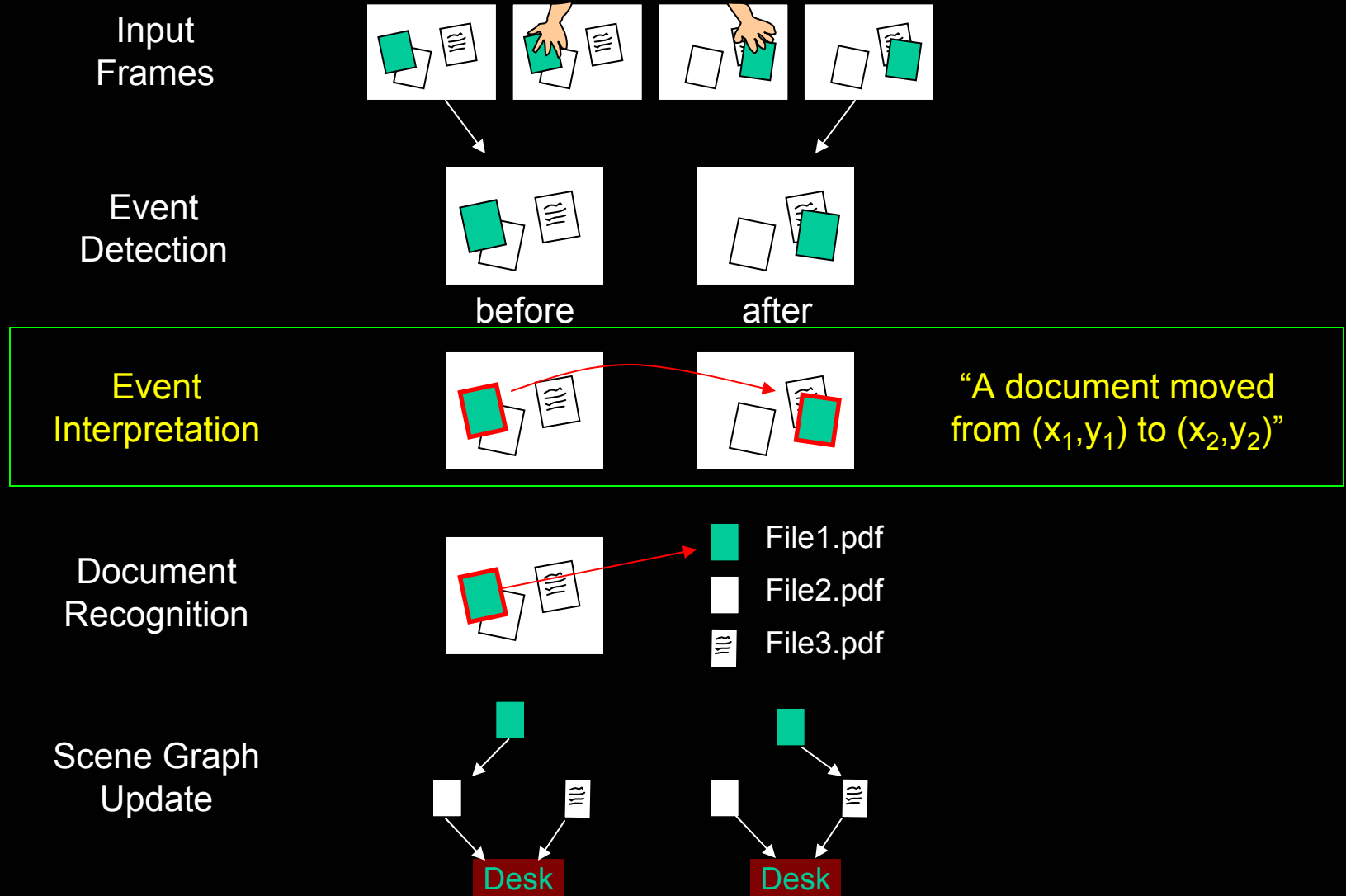


before

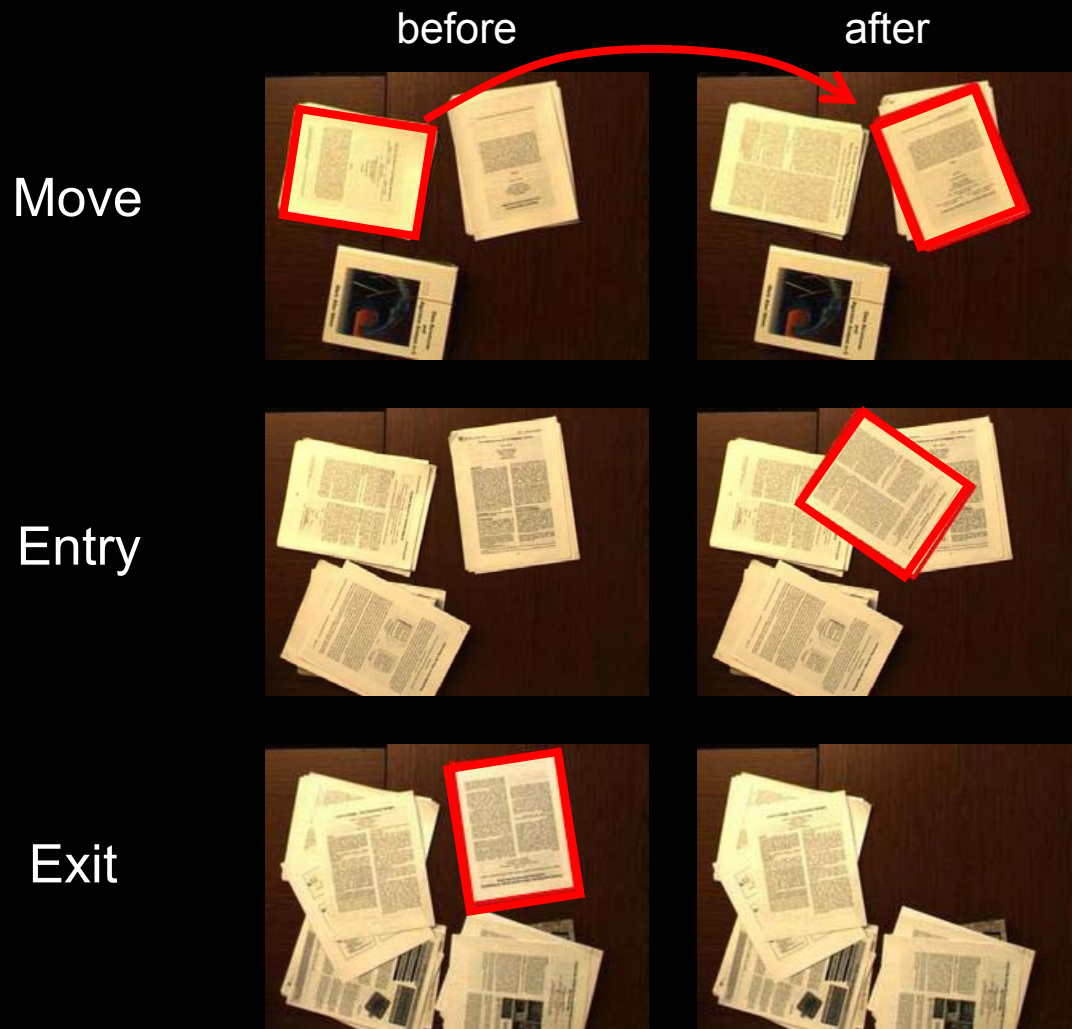


after

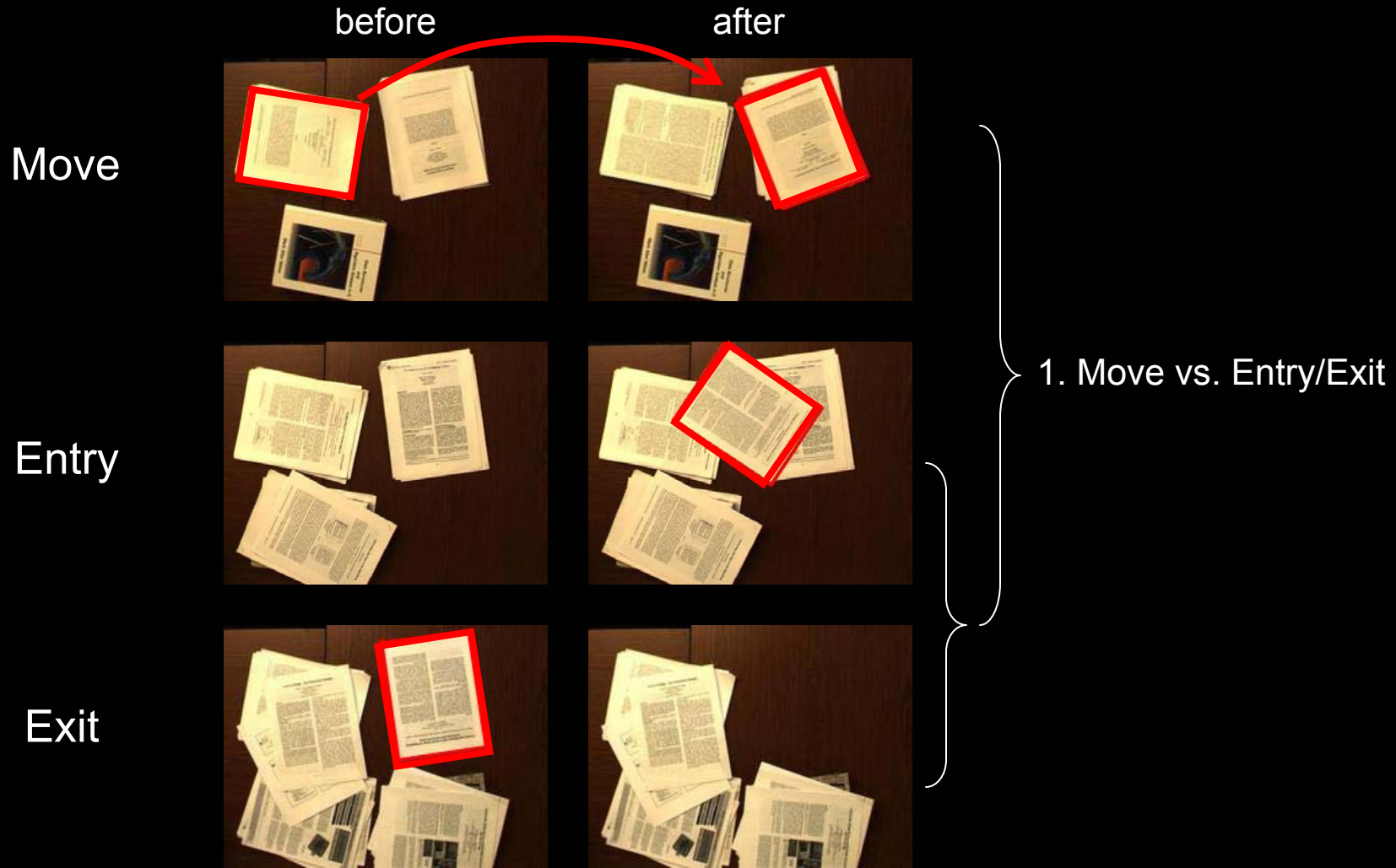
Algorithm Overview



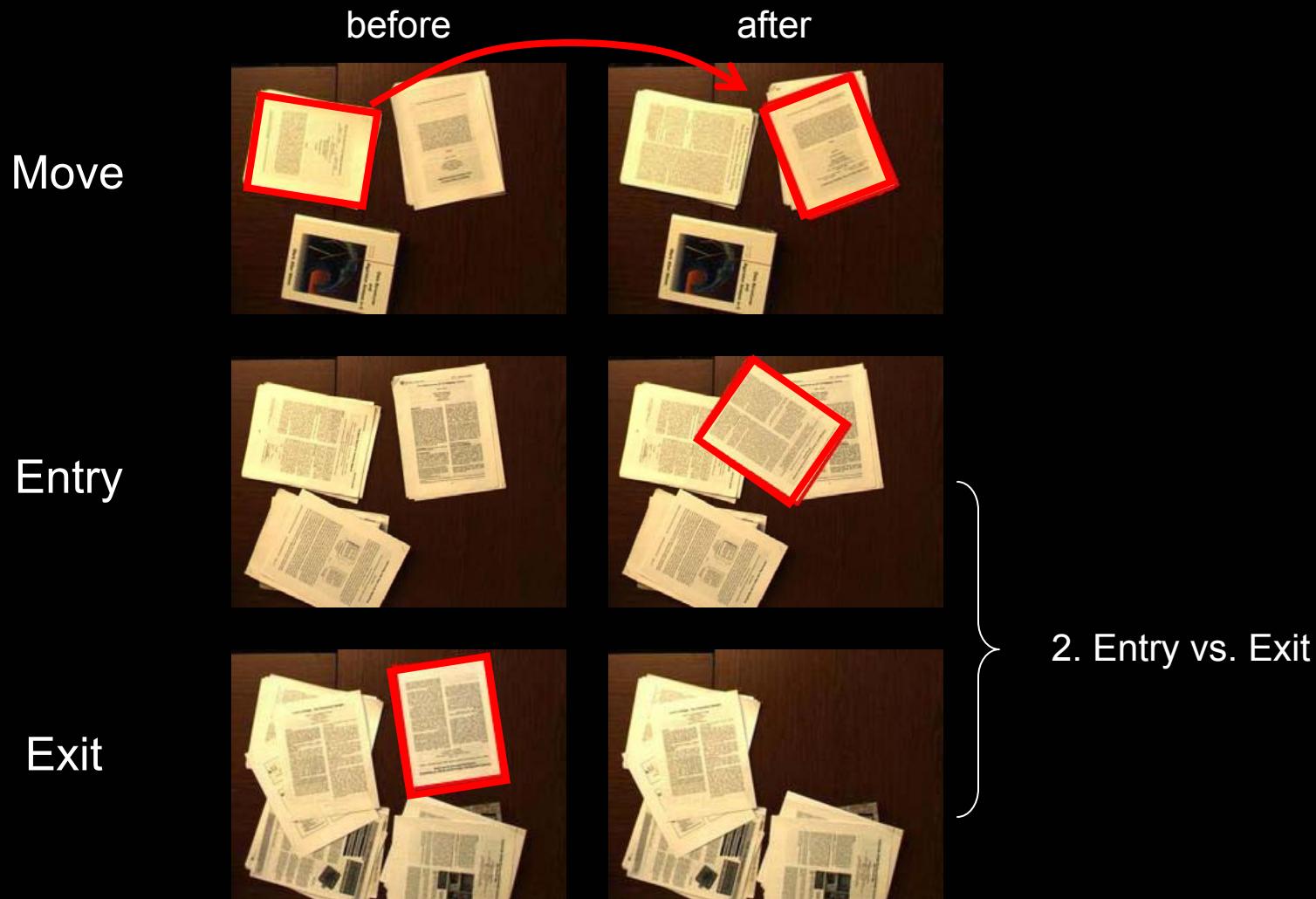
Event Interpretation



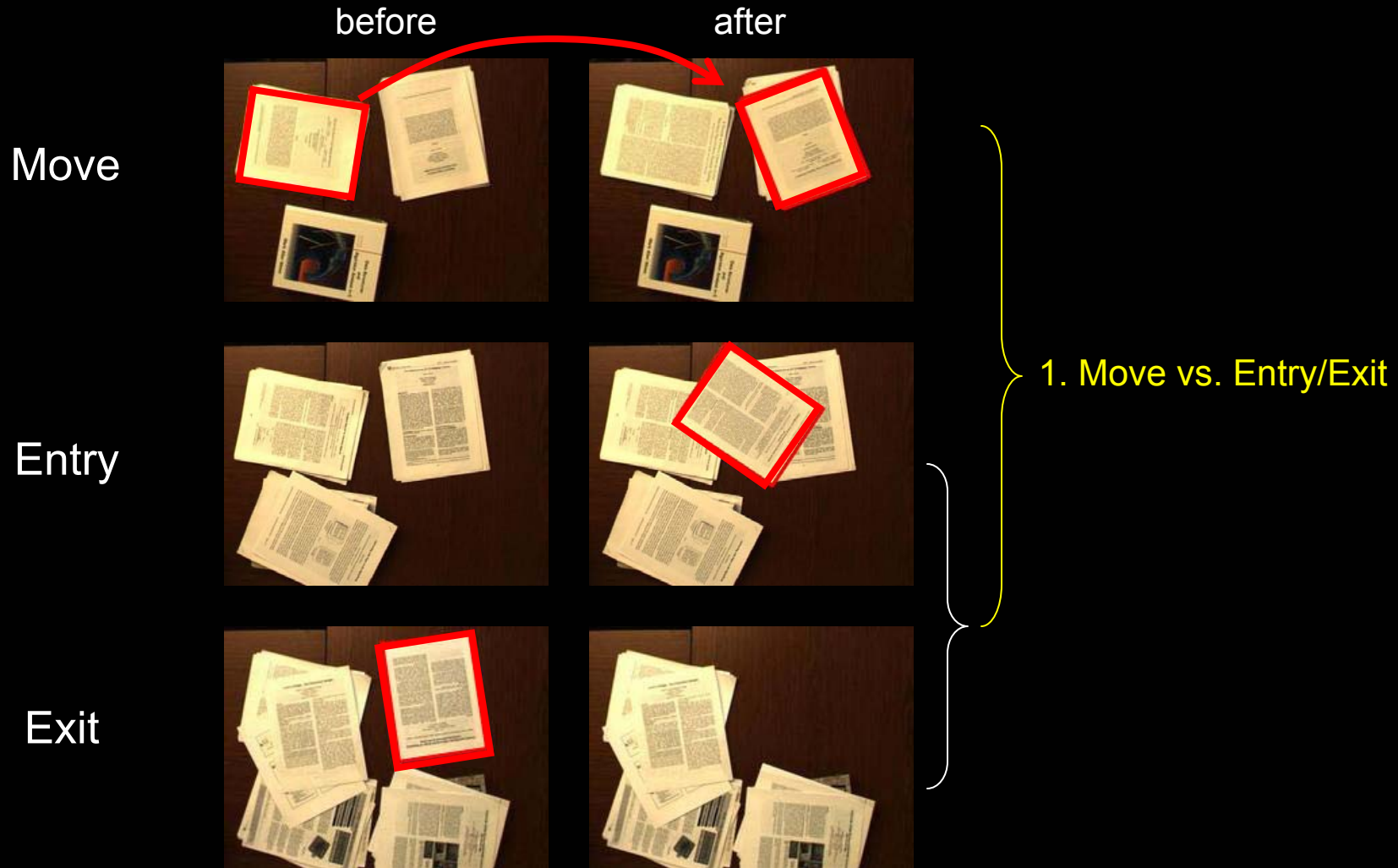
Event Interpretation



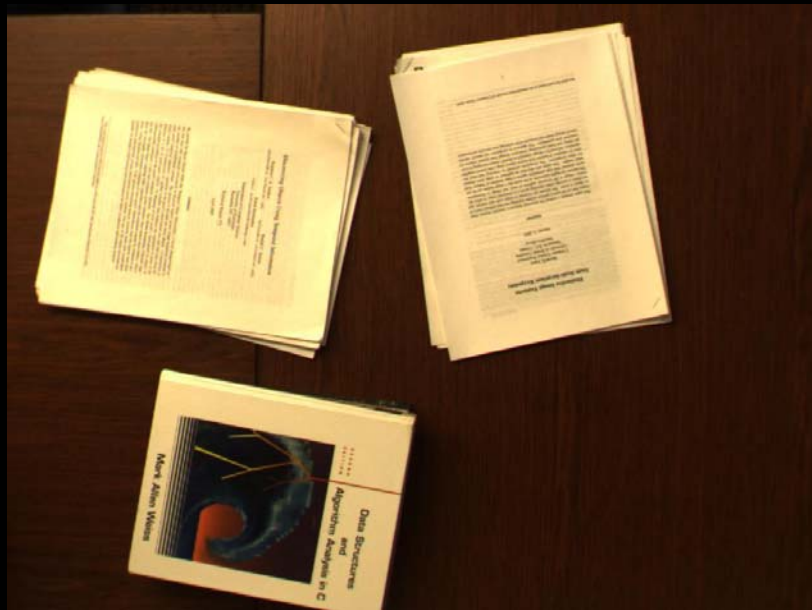
Event Interpretation



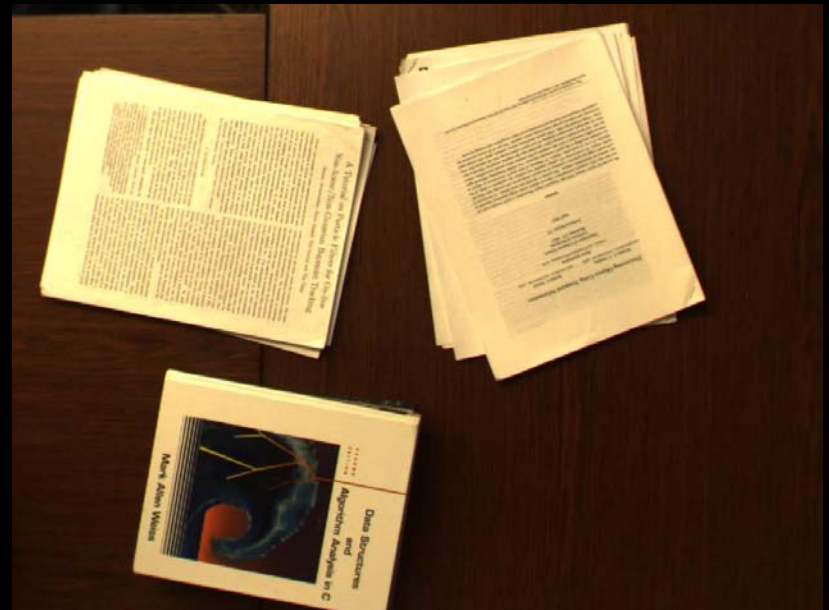
Event Interpretation



Move vs. Entry/Exit

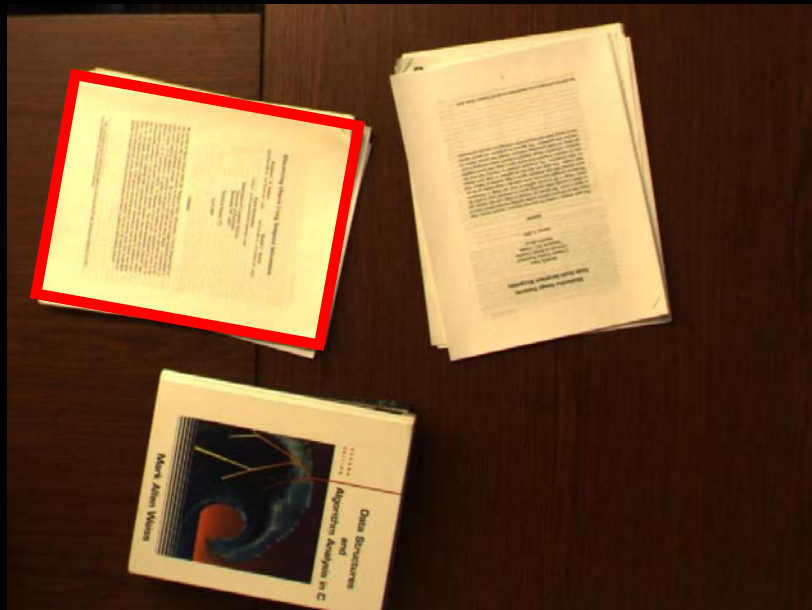


before

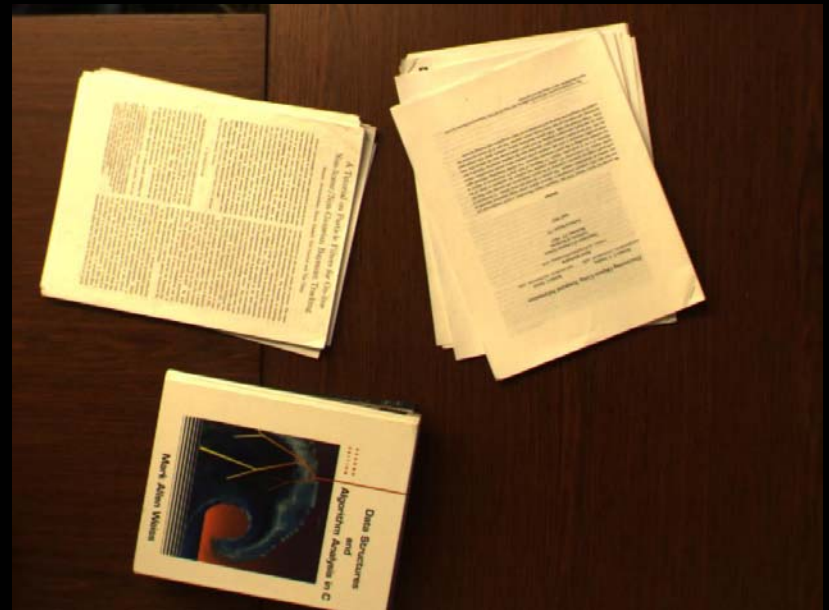


after

Move vs. Entry/Exit

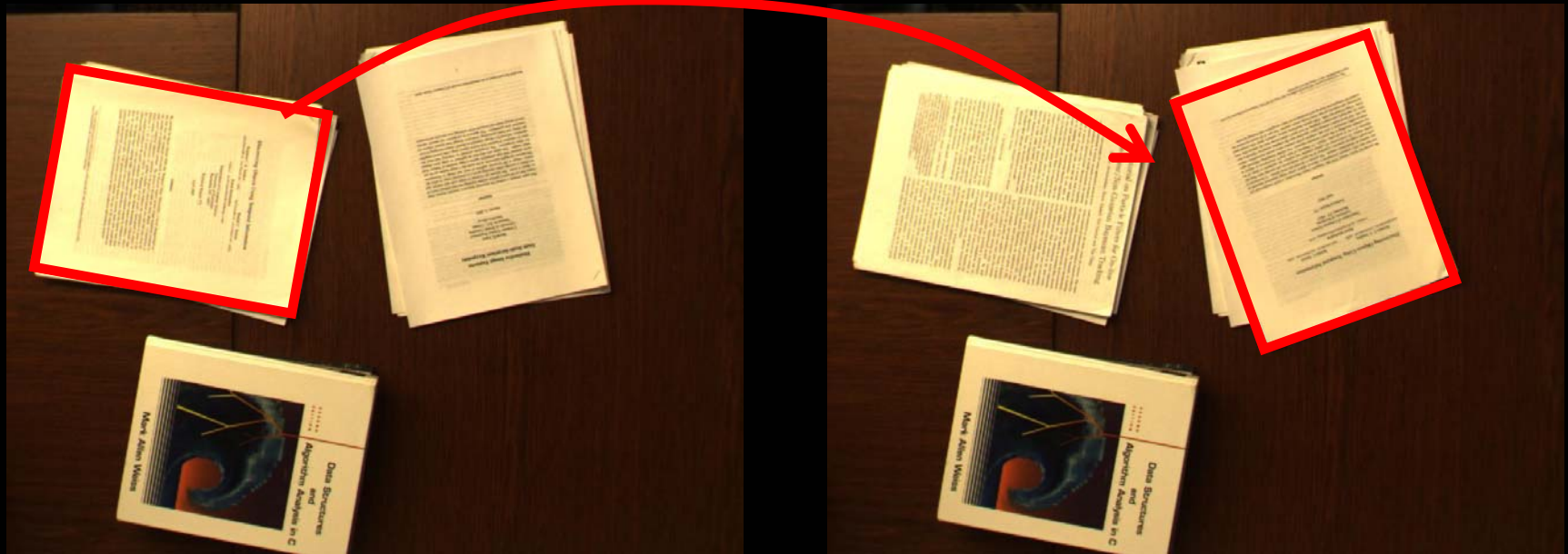


before



after

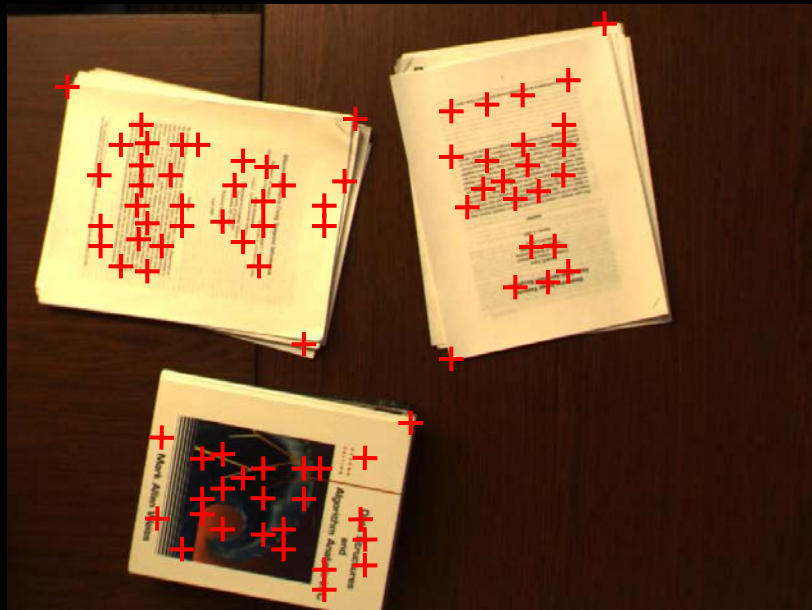
Move vs. Entry/Exit



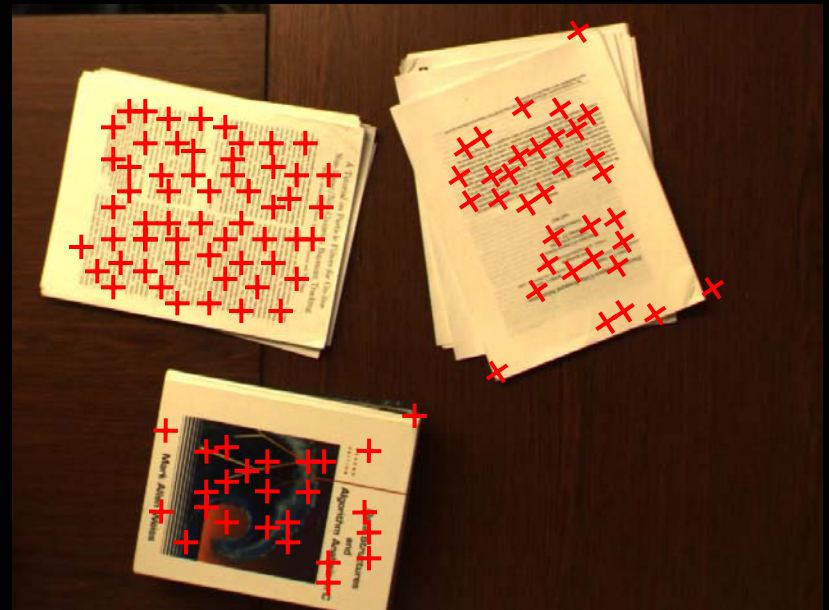
before

after

Move vs. Entry/Exit



before

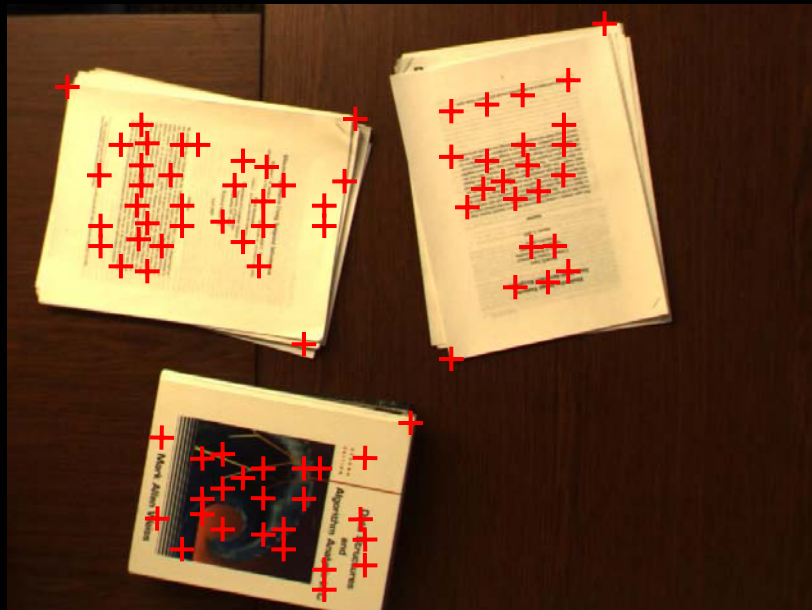


after

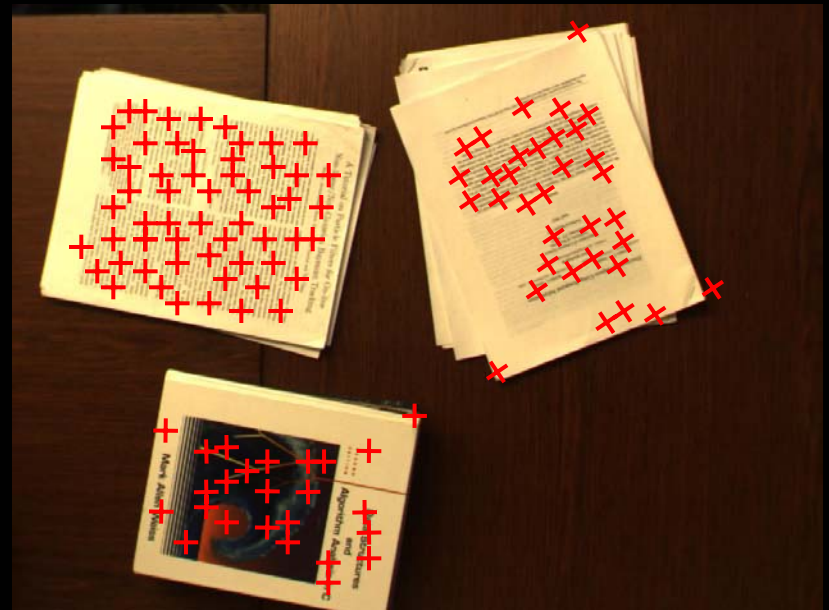
Event Interpretation

- Use SIFT [Lowe 04]
 - Rotation- and scale-invariant
 - Highly distinctive (128-bit vector)

Move vs. Entry/Exit

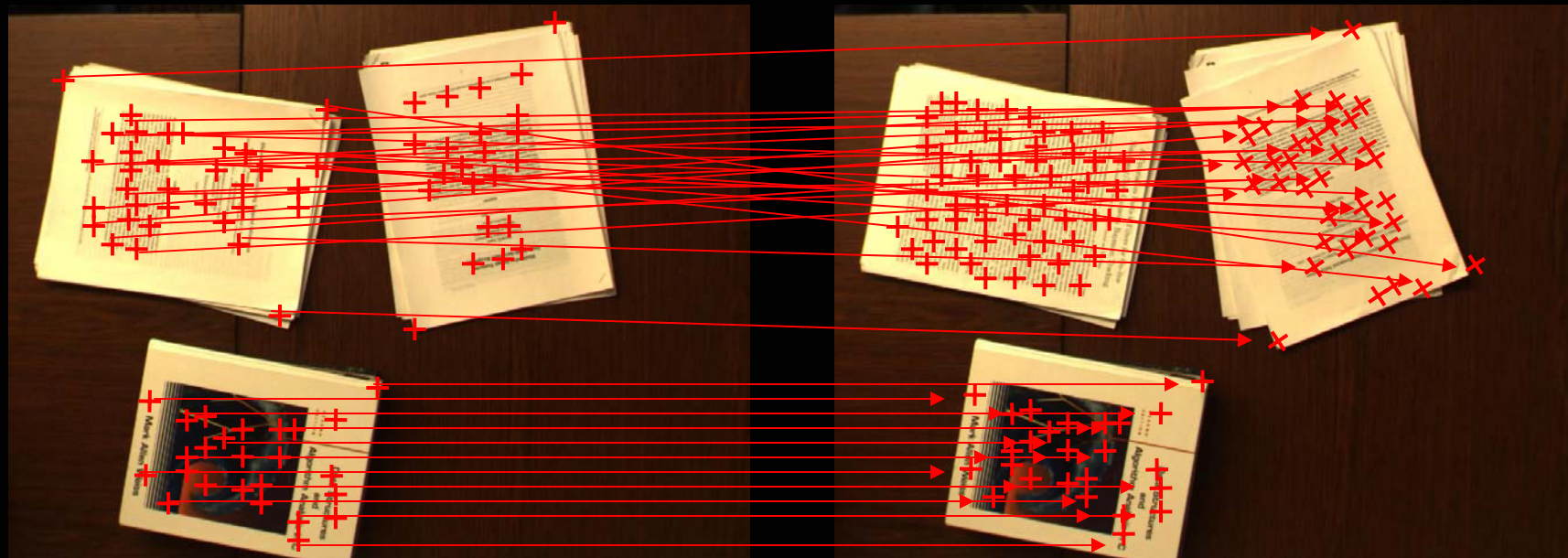


before



after

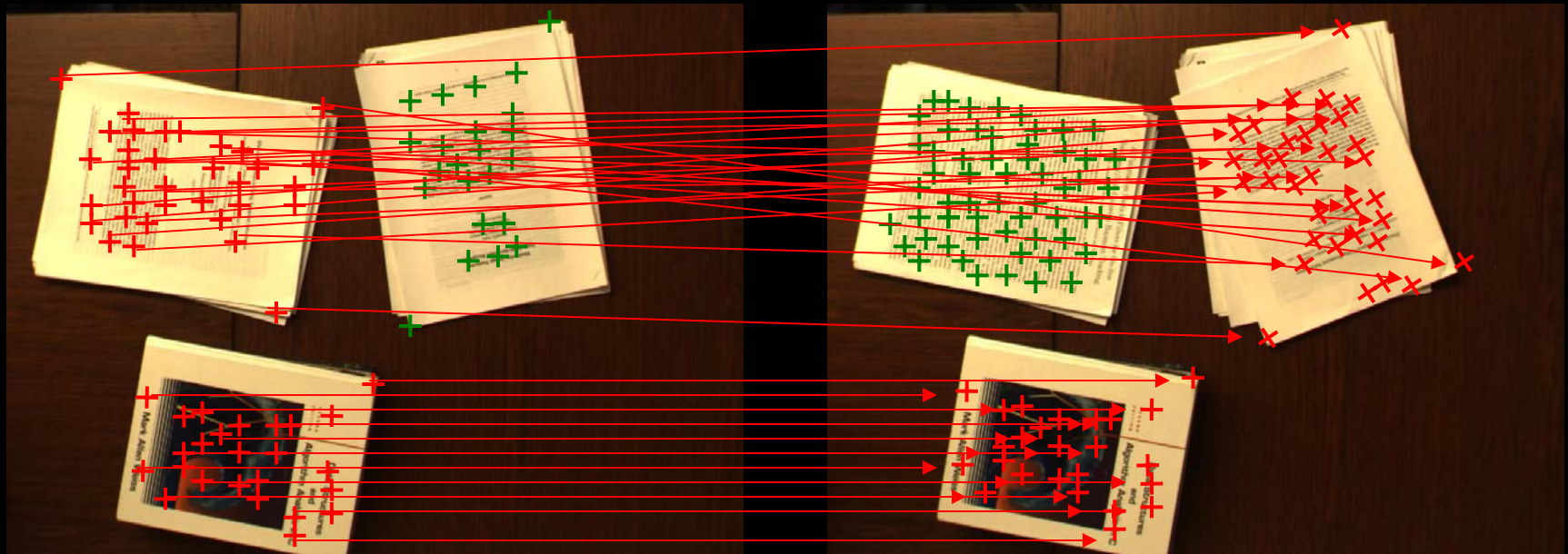
Move vs. Entry/Exit



before

after

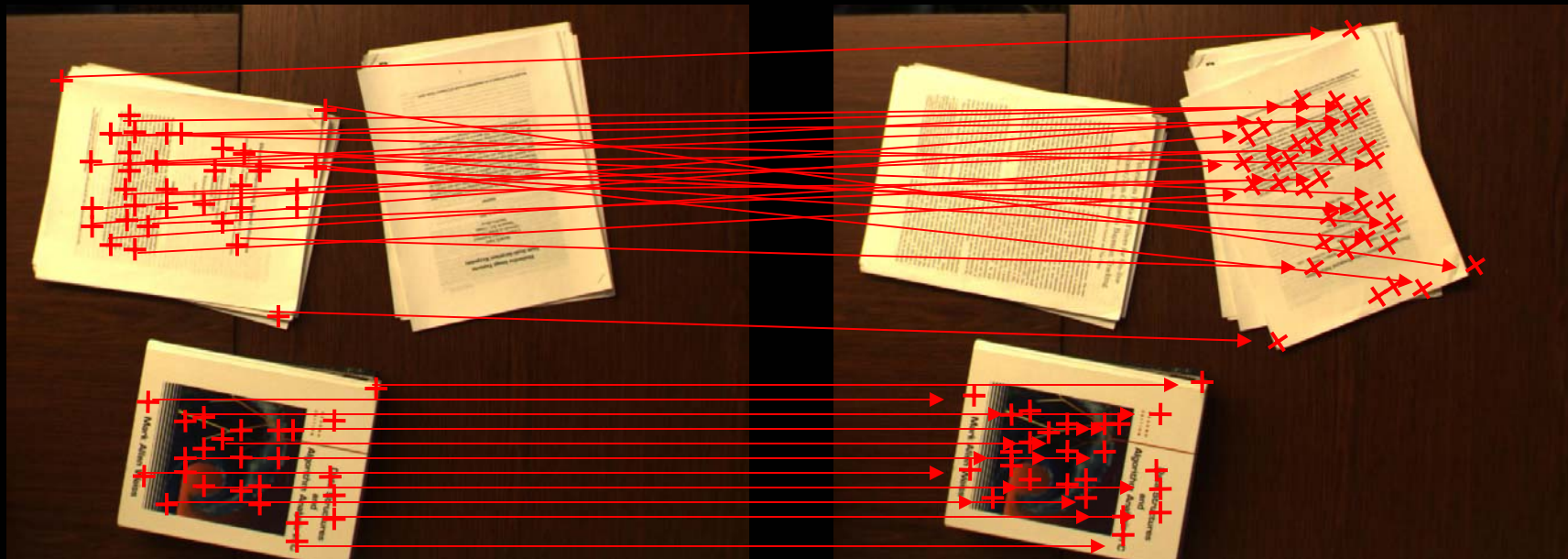
Move vs. Entry/Exit



before

after

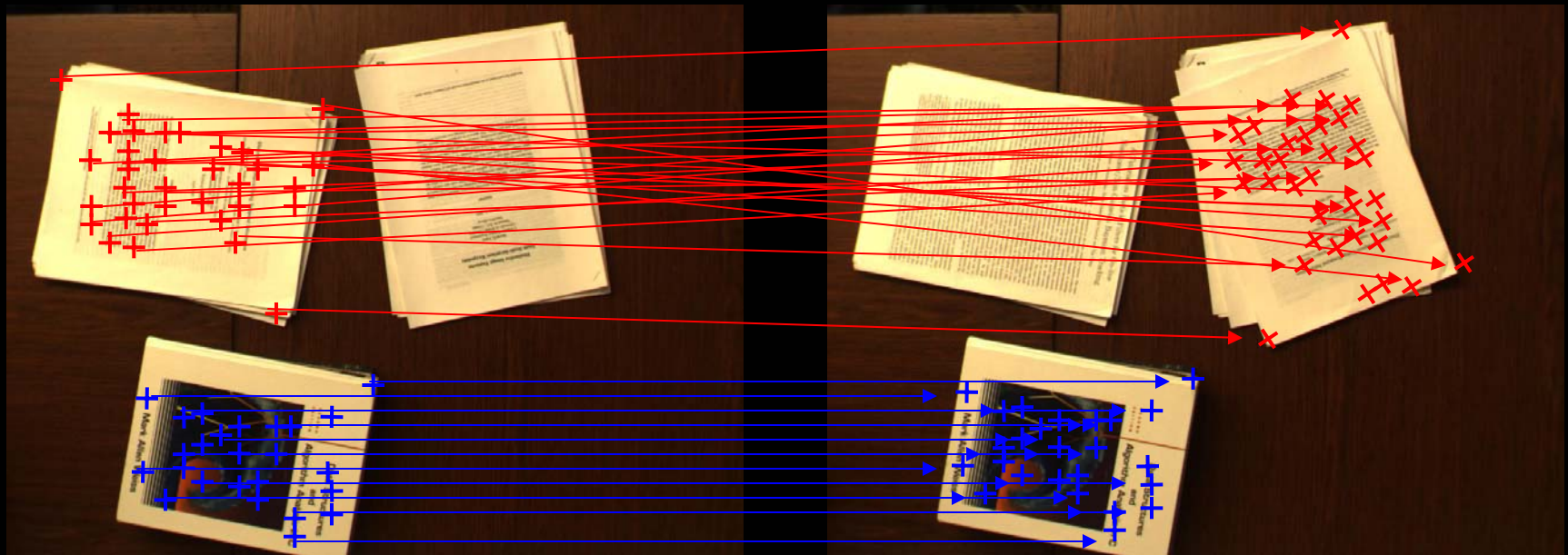
Move vs. Entry/Exit



before

after

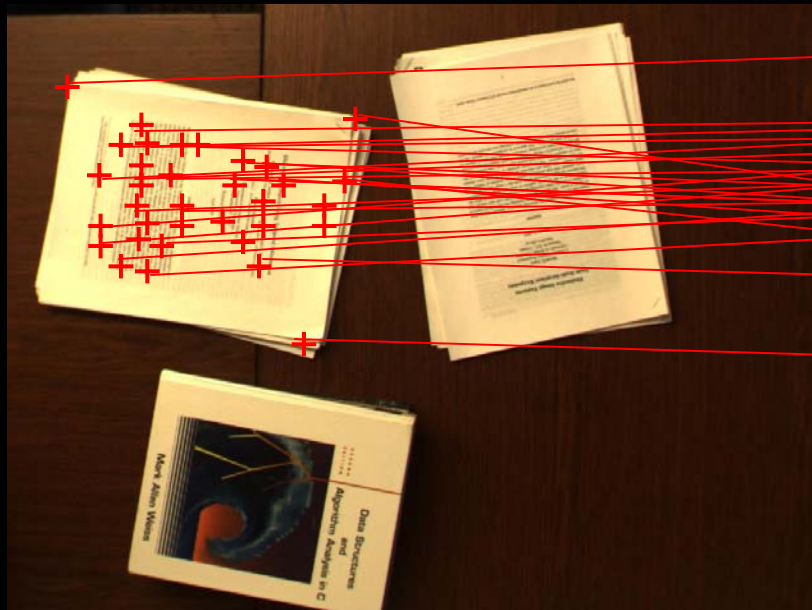
Move vs. Entry/Exit



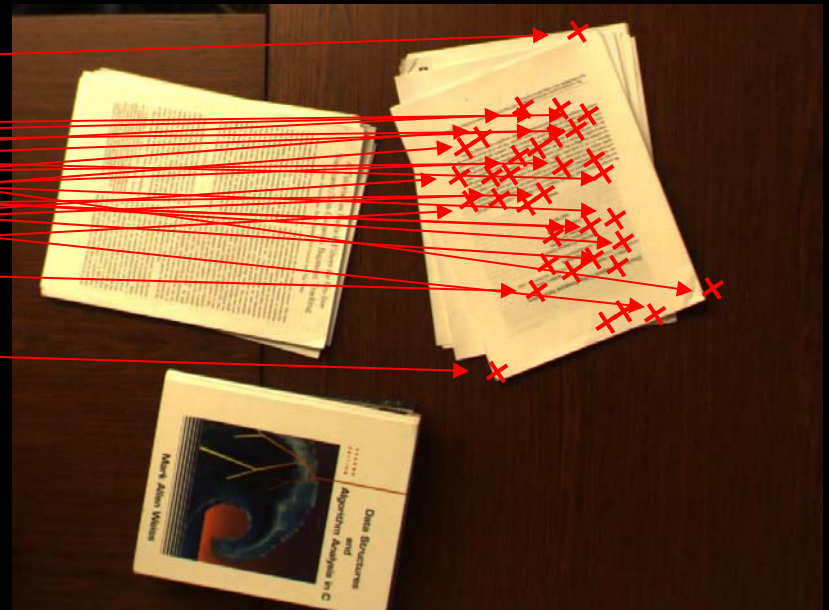
before

after

Move vs. Entry/Exit



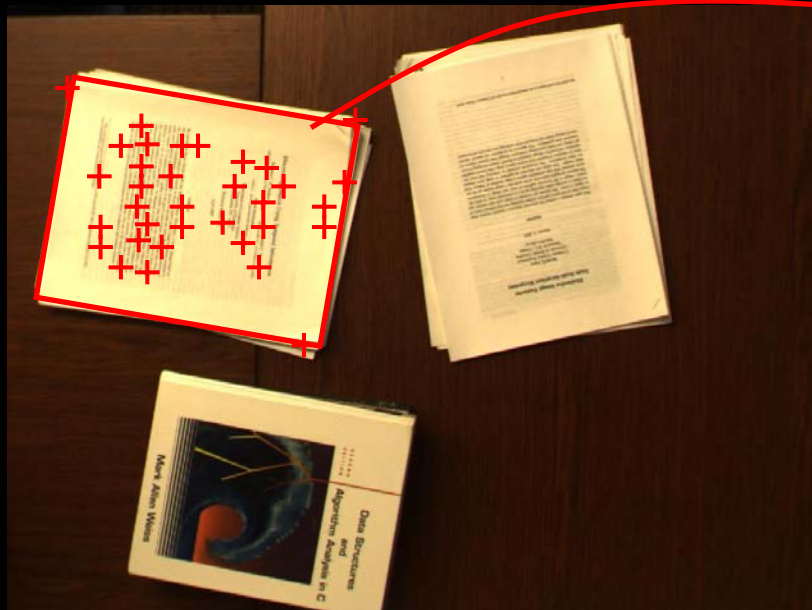
before



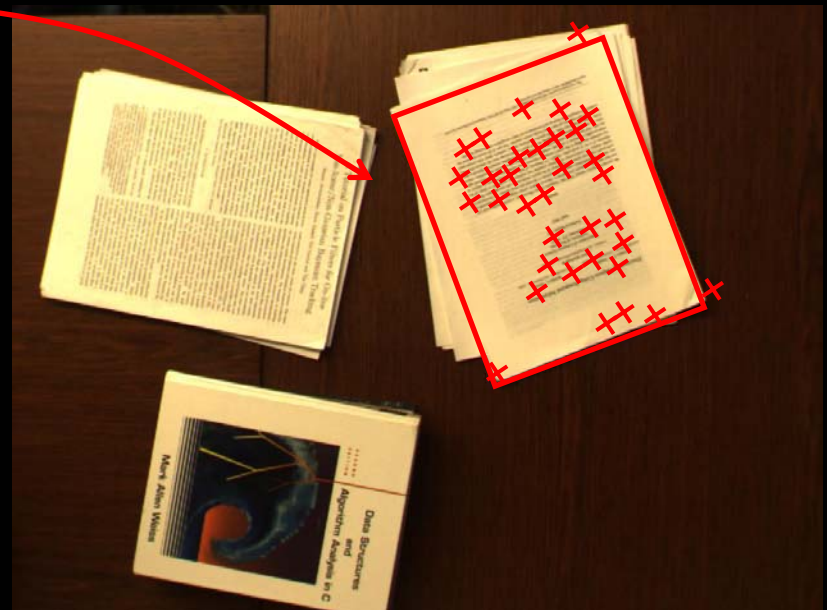
after

Move vs. Entry/Exit

Motion: (x,y,θ)



before



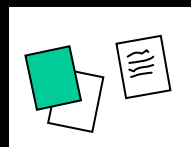
after

Algorithm Overview

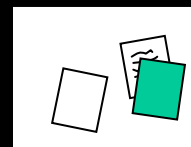
Input
Frames



Event
Detection

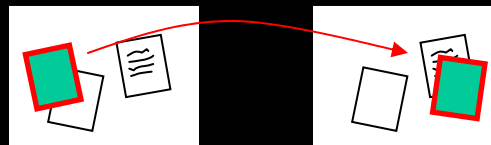


before



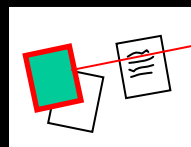
after

Event
Interpretation

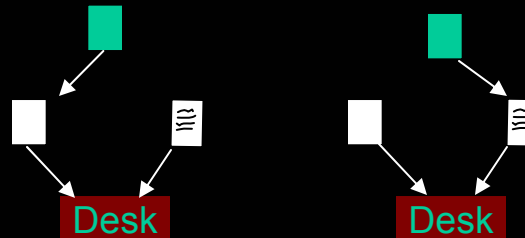


“A document moved
from (x_1, y_1) to (x_2, y_2) ”

Document
Recognition

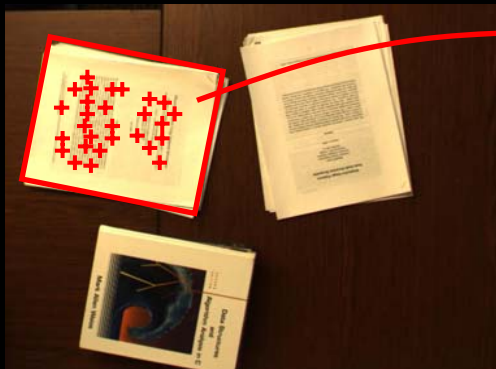


Scene Graph
Update



Document Recognition

- Match against PDF image database



File1.pdf File2.pdf File3.pdf File4.pdf File5.pdf File6.pdf

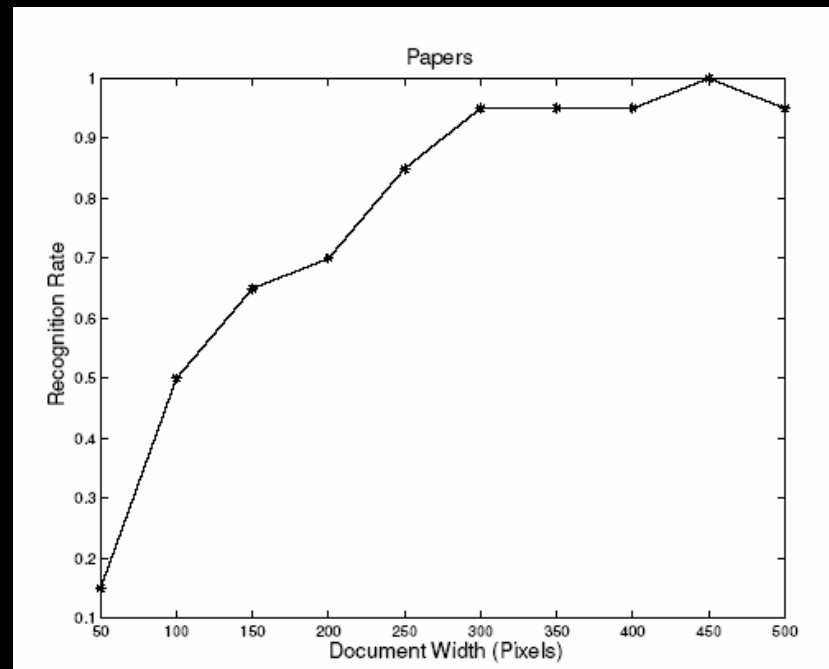
Document Recognition

- Performance analysis
 - Tested 20 pages against database of 162 pages

Document Recognition

- Performance analysis
 - Tested 20 pages against database of 162 pages
 - $\sim 200 \times 300$ pixels per document for reliable match

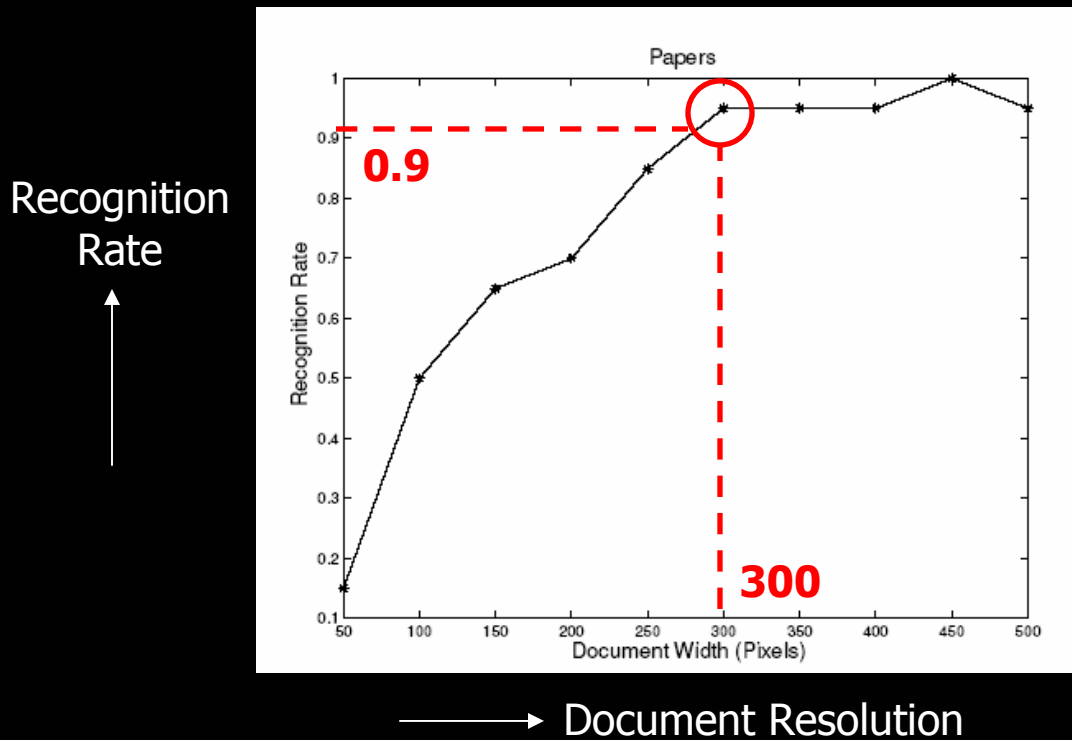
Recognition Rate



Document Resolution

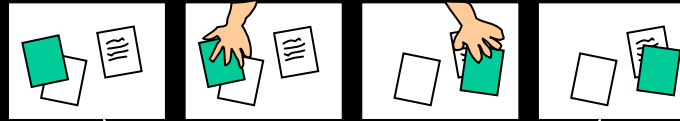
Document Recognition

- Performance analysis
 - Tested 20 pages against database of 162 pages
 - $\sim 200 \times 300$ pixels per document for reliable match



Algorithm Overview

Input
Frames



Event
Detection

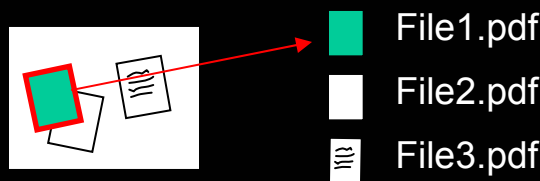


Event
Interpretation

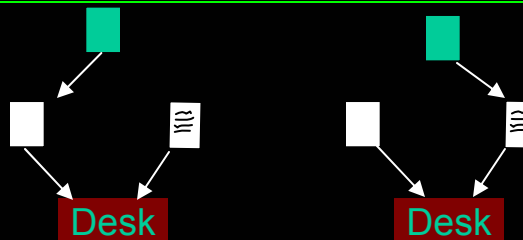


“A document moved
from (x_1, y_1) to (x_2, y_2) ”

Document
Recognition

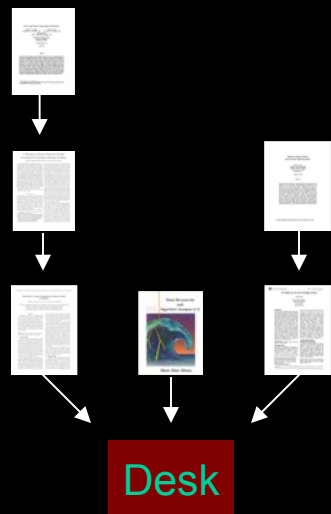
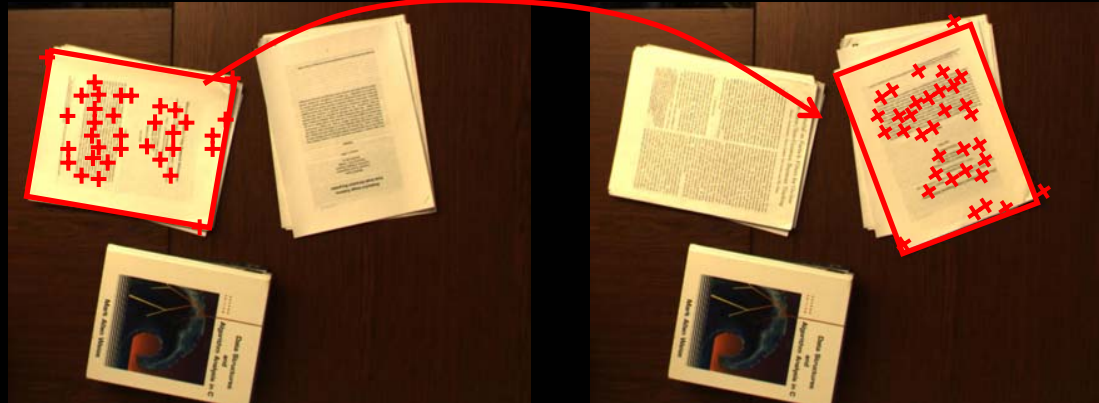


Scene Graph
Update



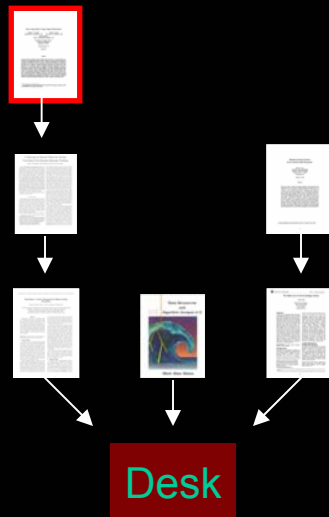
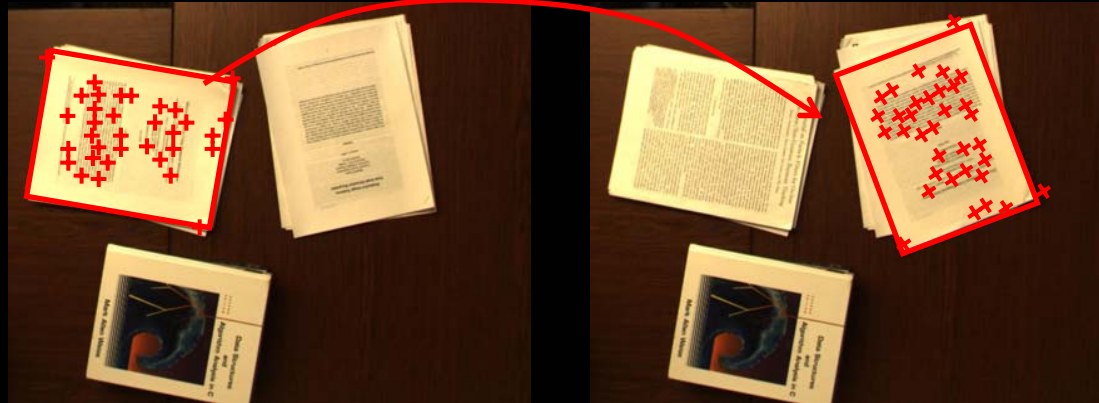
Scene Graph Update

Motion: (x,y,θ)



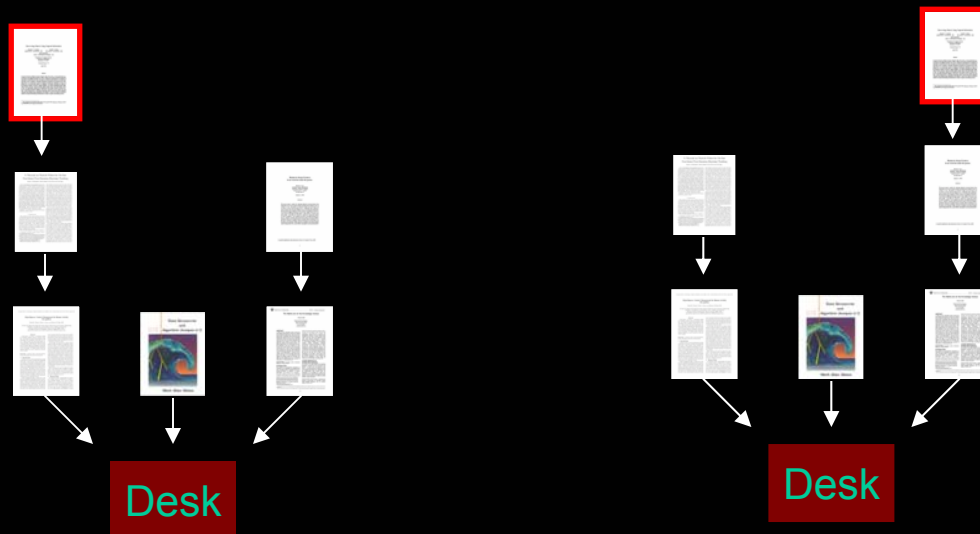
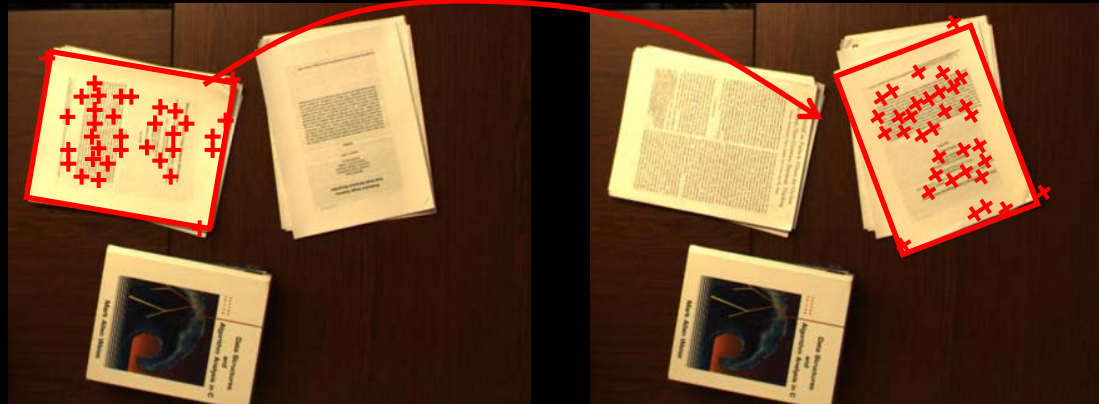
Scene Graph Update

Motion: (x,y,θ)



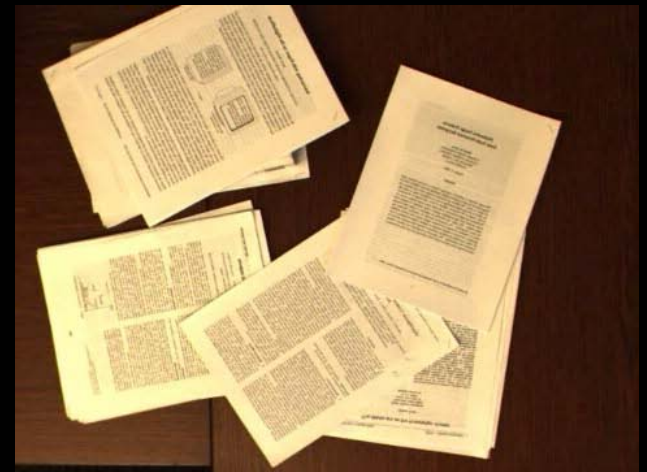
Scene Graph Update

Motion: (x, y, θ)



Results

- Input video
 - ~40 minutes
 - 1024x768 @ 15 fps
 - 22 documents, 49 events
- Running time
 - Video processed offline
 - No optimization
 - A few hours for entire video



Demo – Paper tracking

Photo Sorting Example

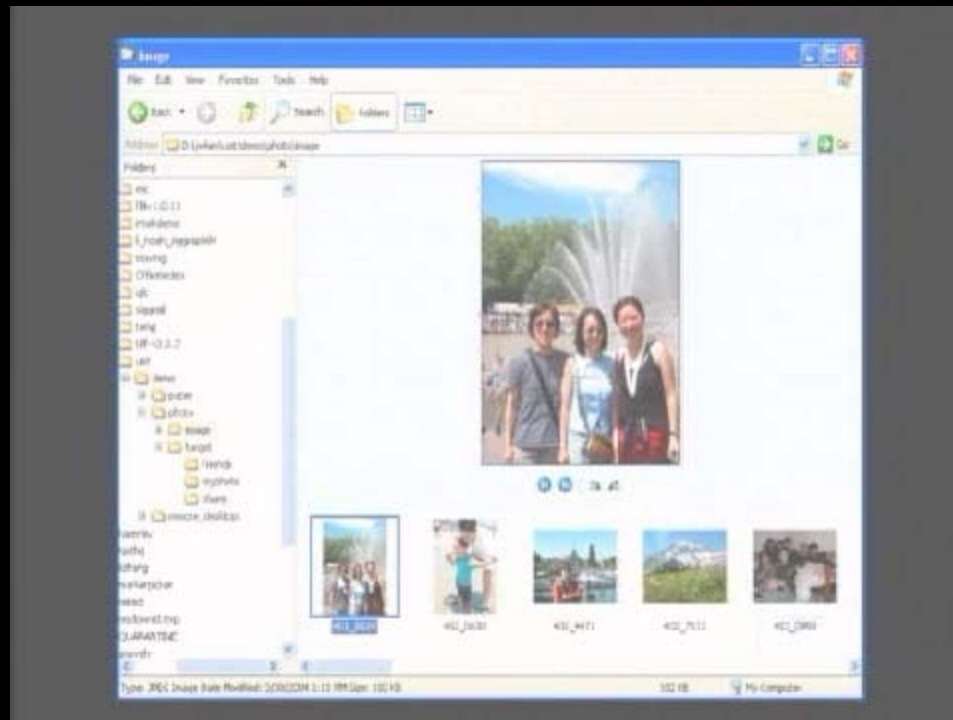


Photo Sorting Example



Demo – Photo Sorting

Future Work

- Enhance realism
- More applications

Future Work

- Enhance realism
 - Handle more realistic desktops



Moving a stack of documents



Documents with no electronic versions

Future Work

- Enhance realism
 - Handle more realistic desktops
 - Real-time performance

Future Work

- More applications
 - Support other document tasks
 - E.g., attach reminder, cluster documents

Future Work

- More applications
 - Support other document tasks
 - E.g., attach reminder, cluster documents
 - Beyond documents

Future Work

- More applications
 - Support other document tasks
 - E.g., attach reminder, cluster documents
 - Beyond documents



Future Work

- More applications
 - Support other document tasks
 - E.g., attach reminder, cluster documents
 - Beyond documents



Future Work

- More applications
 - Support other document tasks
 - E.g., attach reminder, cluster documents
 - Beyond documents



Acknowledgments

- NSF
- Intel Corp.
- Li Zhang