Artificial and robotic vision

Eugenio Culurciello

Lecture 9: vision systems

e-Lab

visual intelligence

“being able to track targets of interests while they are on the scene, keep targets in memory if they disappear, recognize them when they re-appear”

Targets: objects or parts of the scene

Scene: environment perceived by the visual system

Memory: a list of targets and their identifying features

keep a list of targets in memory that can be used to infer higher order statistics and behavior of the targets in the scene: the foundations of intelligence, behavior, interaction with environment

Computer Vision and some history...

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

e-Lab

looking at computer vision with a bio-inspired eye

What did we learn?

problem

Pieces in isolation:
- segmentation
- stereo, optical flow
- recognition
- tracking

...

Eugenio Culurciello

looking at computer vision with a bio-inspired eye

Pieces in isolation: **cannot do that!**
- the system is **entangled**
- **parts** help the **whole**

- it is **inefficient** to do the same operation in multiple algorithms/system!!!

e-Lab

unified visual model

solution

Why a unified model?

- to replicate human vision in one system, or:
- to tackle general problems, new problems with one system
- train only one system, not a bag of tricks
- to re-use the same processing elements for multiple modules:
    - --> use gabor filters for categorization also segmentation, etc
    - --> use contrast normalization as input to multiple modules
    - --> use 1st/2nd layer filters as features for multiple modules

problem

biggest problem in vision is:

object has to be found in background (bg)

you cannot remove the bg
you cannot segment the object
you cannot just learn object features w/out bg

so how do we know there is an object there?

Segmentation

- segmentation of images: best techniques are graph based - Pedro F. Felzenszwalb

http://www.cs.brown.edu/~pff/bp/

- Also fast color segmentations are good:

Segmentation

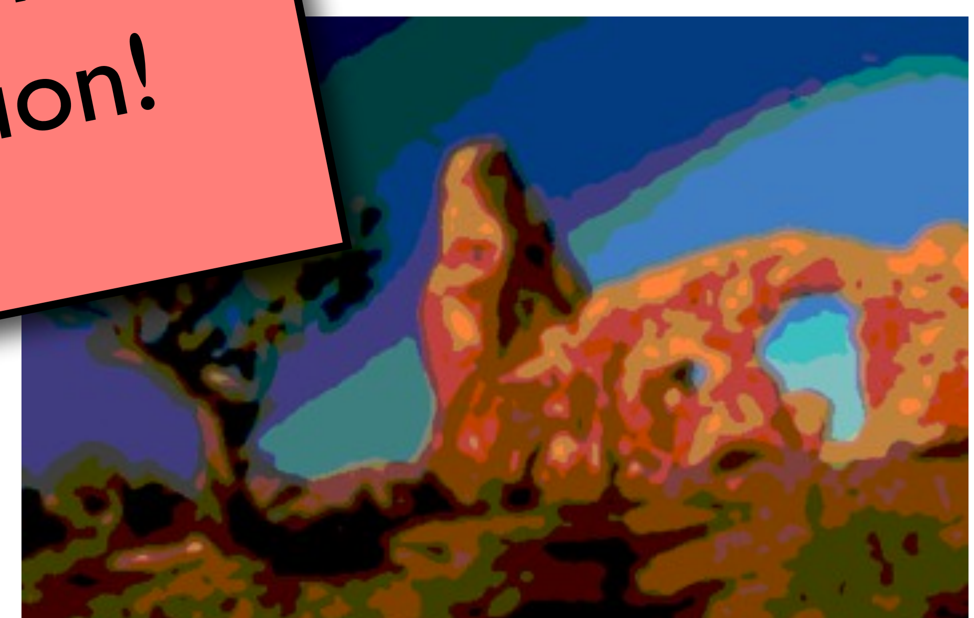
problem

- segmentation of images: b t techniques are graph based - P zwalb

htt du/~pff/bp/

not a unified system: another method used in isolation that performs redundant computation!

- Also fast co re good:



Eugenio Culurciello

proto-objects and segmentation

How do we know something is an object?

- we group some “features”
- areas that look similar
  in color
  in texture
- areas that move together
- prior knowledge of “objectness”

proto-objects and segmentation

problem

How do we know something is an object?

- we group some “features”
- areas that look similar
in color
in texture
- areas that move together
- prior knowledge of “objectness”

We still do not have good models to do all this at once and in real-time

e-Lab

tracking offers some insights:

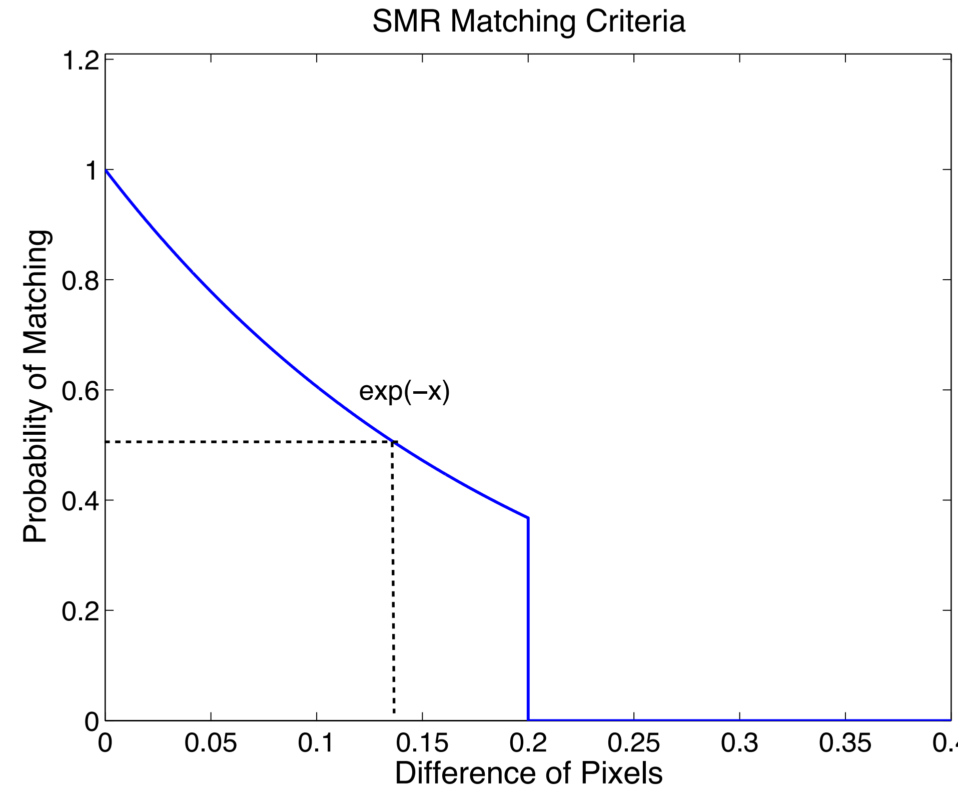
solution

because... you need to know "objectness" in order to track

but neuroscience model of tracking only mentions we track in feature space, not which and how

Blaser, E., Pylyshyn, Z., Holcombe, A., et al. (2000). Tracking an object through feature space. Nature, 408(6809):196–198.

SMR tracking



For each pixel, find the absolute difference of pixel values. Are they matching?

Yes

Maybe
p = %50

No

For each pixel, accumulate the probabilities: **SMR**

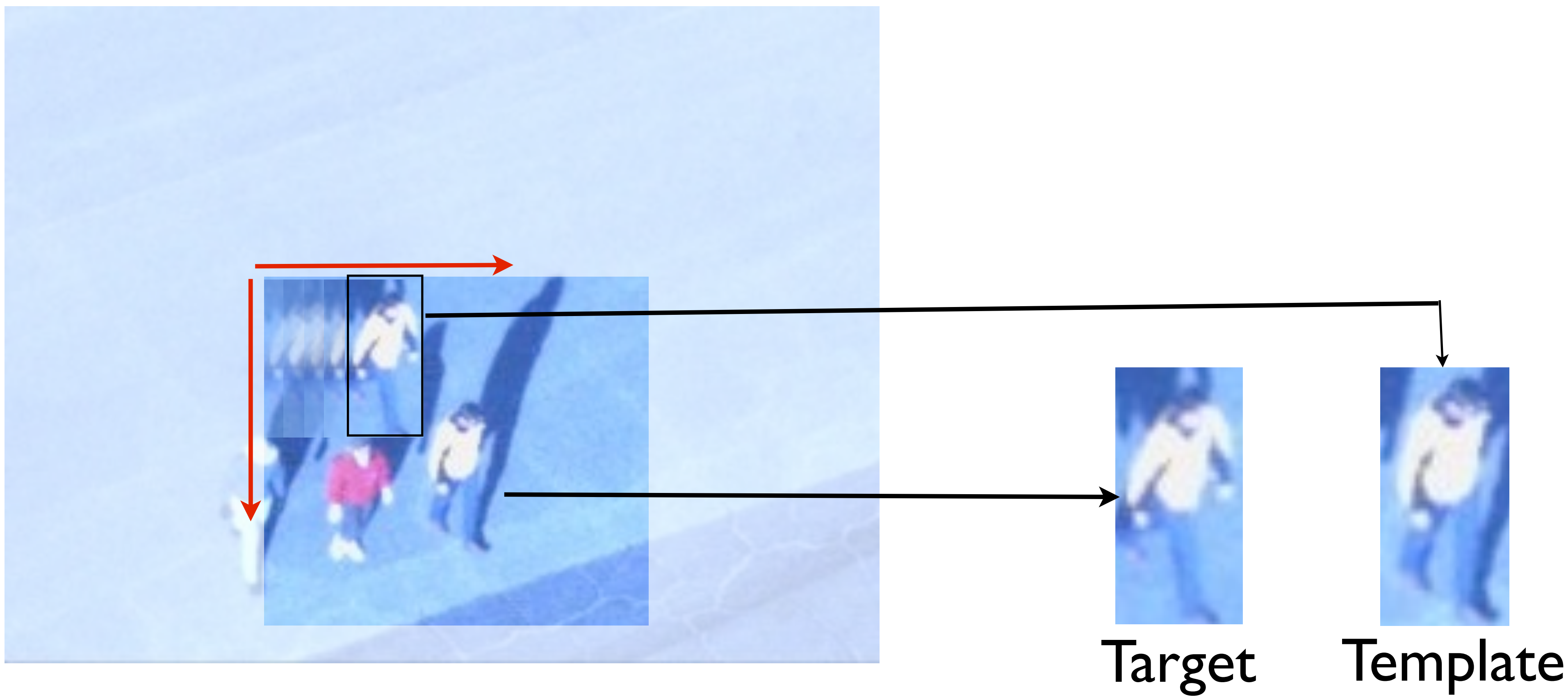
Outliers: Pixels dramatically changed

Should not effect the matching similarity.

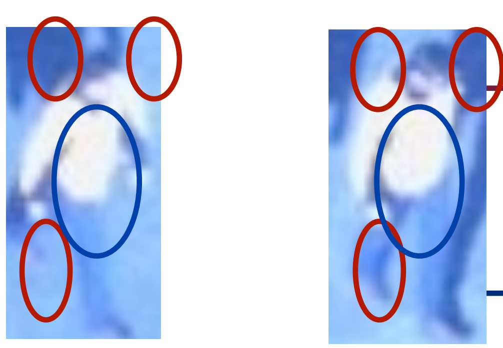
Biggest Similarity Matching Ratio is the one with the biggest number of pixels that matches with the template. → SMR Tracker's Result

e-Lab

SMR tracking



Same object for us, very different for computer. Why?



Different
Shadow, right leg, arm

Same
Upper body, left leg, head

Computer looks as a whole.

We look at the majority.

SMR tracking

problem

SMR only tracks some features but has no notion of "objectness"

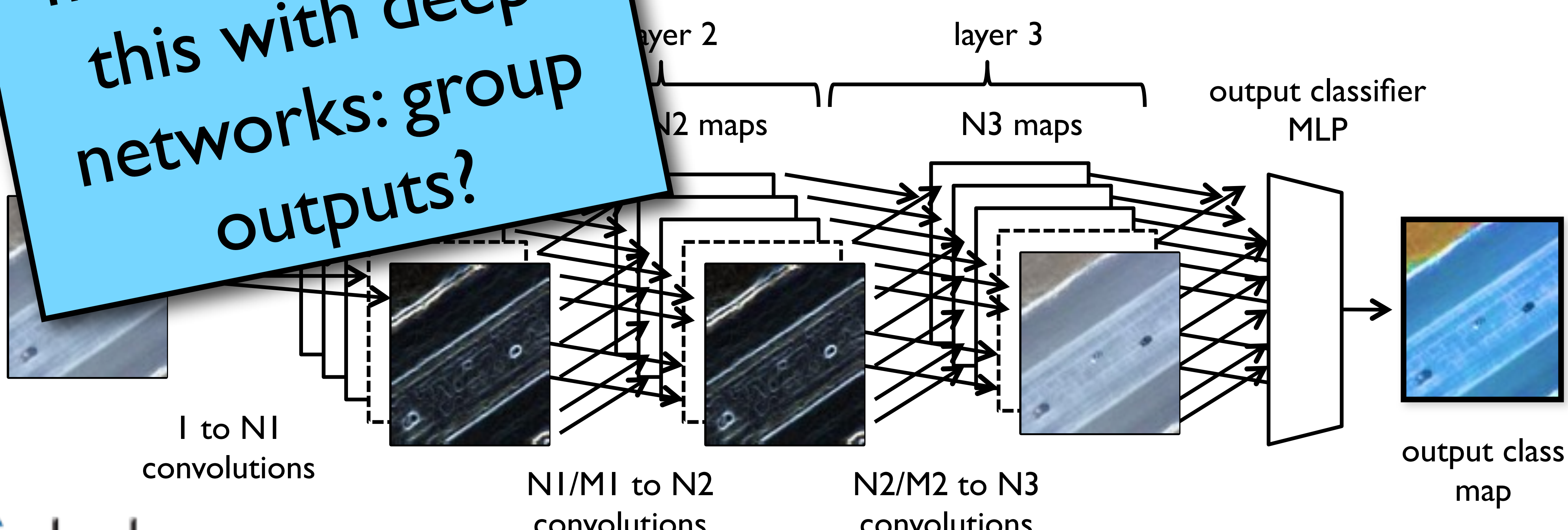
not a unified system: another method used in isolation that performs redundant computation!

proto-objects and segmentation

solution

We still do not have good models to do all this at once and in real-time

find a way to do this with deep networks: group outputs?



e-Lab

Eugenio Culurciello

Deep Networks as trackers

solution

[Learning Convolutional Feature Hierachies for Visual Recognition](#) ,
K. Kavukcuoglu, P. Sermanet, Y. Boureau, K. Gregor, M. Mathieu and Y. LeCun,
Advances in Neural Information Processing Systems [9 pages]

e-Lab

Deep Networks for segmentation

solution

We still do not have good models to do all this at once and in real-time

Srini Turaga: nice work in this area!

Jain V, Turaga SC, Seung HS (pdf)
Machines that learn to segment images: a crucial technology of connectomics.*Current Opinion in Neurobiology*, 2010.

Turaga SC, Briggman KL, Helmstaedter M, Denk W, Seung HS (pdf)
Maximin learning of image segmentation. *NIPS*, 2009.

e-Lab

optical flow

segmentation supports region motion estimates

http://people.csail.mit.edu/celiu/OpticalFlow/

e-Lab

optical flow

segmentation supports region motion estimates

not a unified system: another method used in isolation that performs redundant computation!

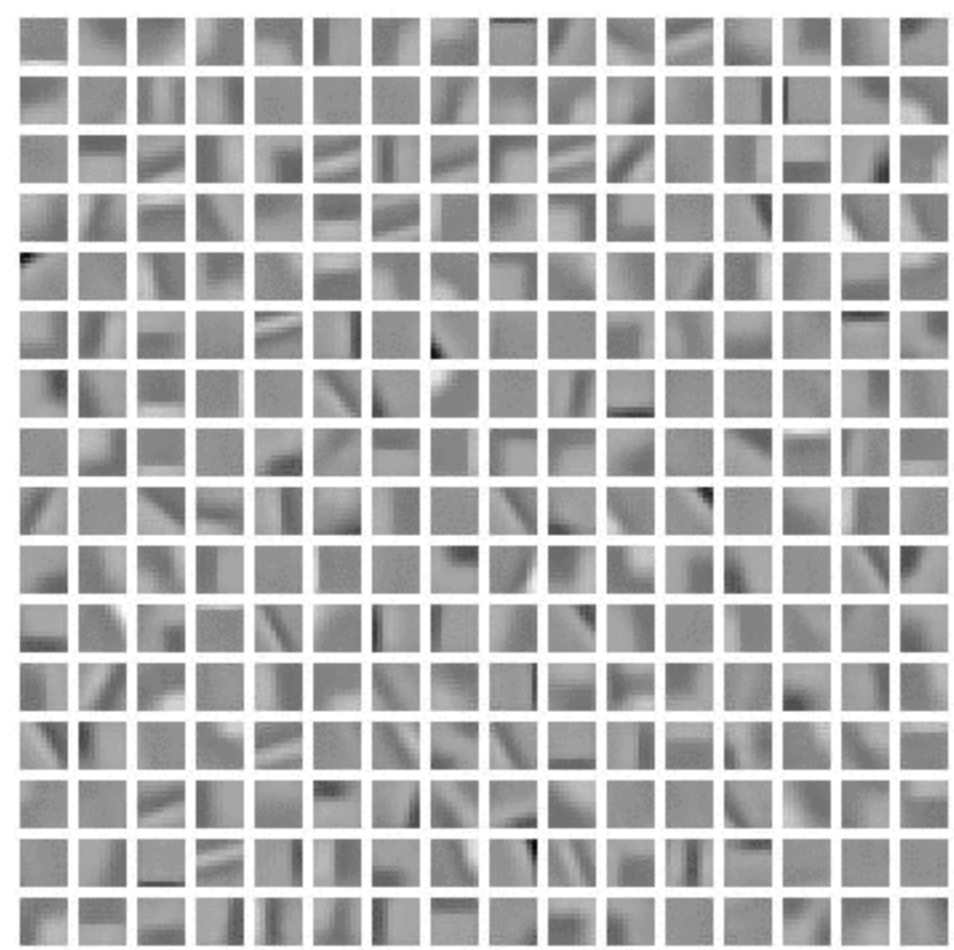
http:/ … /celiu/OpticalFlow/

e-Lab

clustering learning: motion filters

same patch location for multiple frames

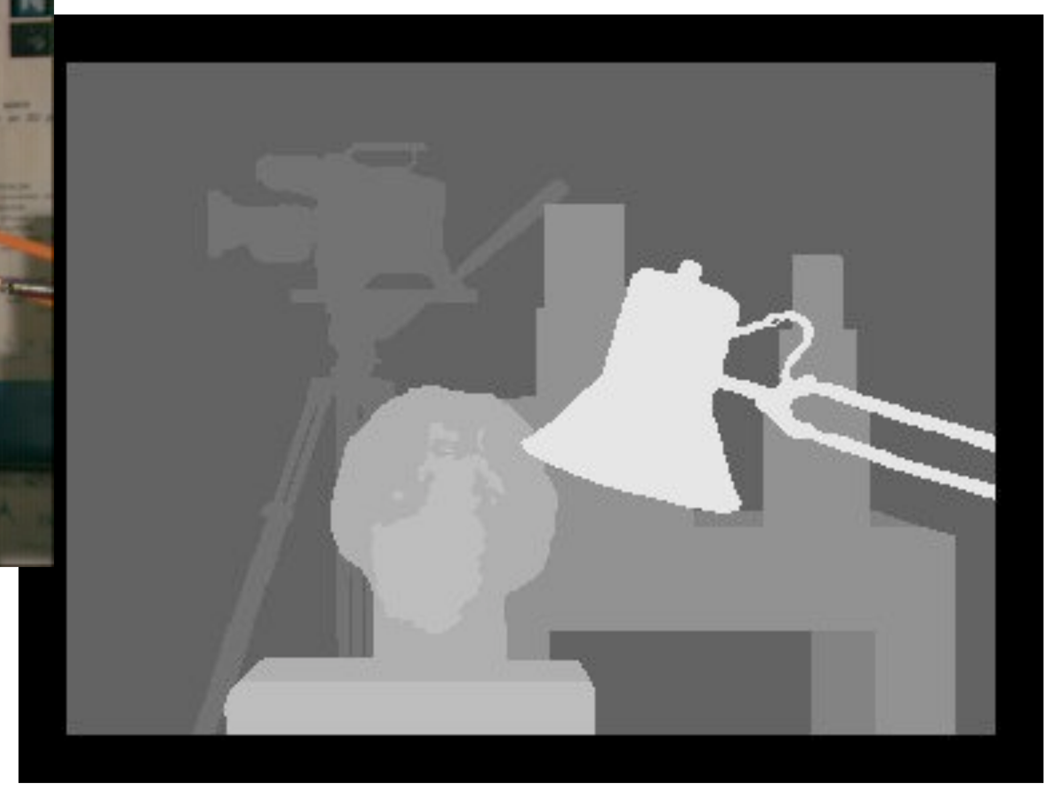
run k-means on group of patches

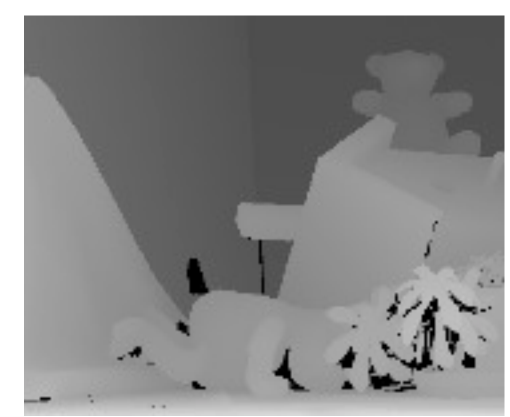


Eugenio Culurciello

stereo and 3D

segmentation supports stereo correspondence





http://vision.middlebury.edu/stereo/

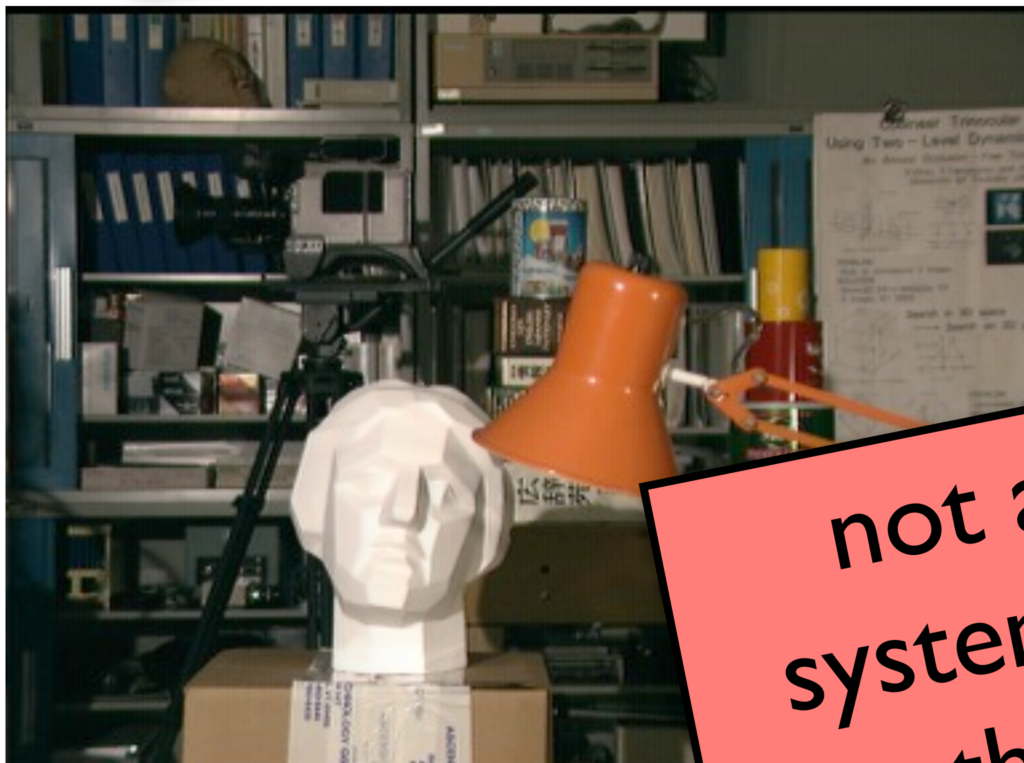
e-Lab

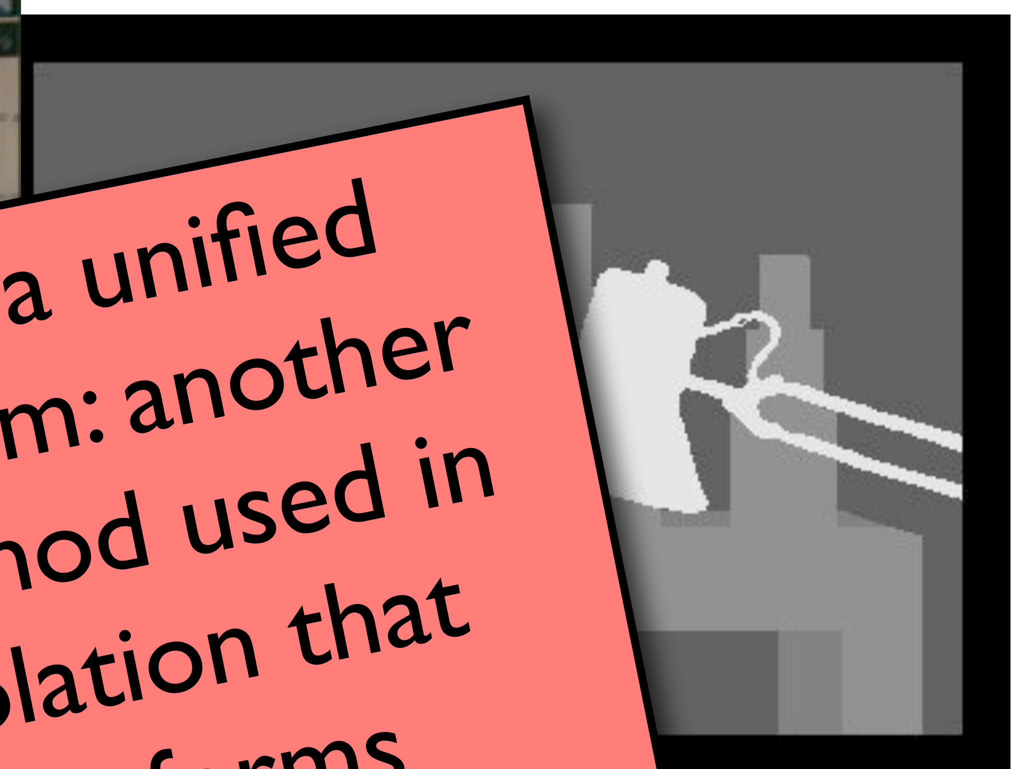
http://www.cs.brown.edu/~pff/bp/

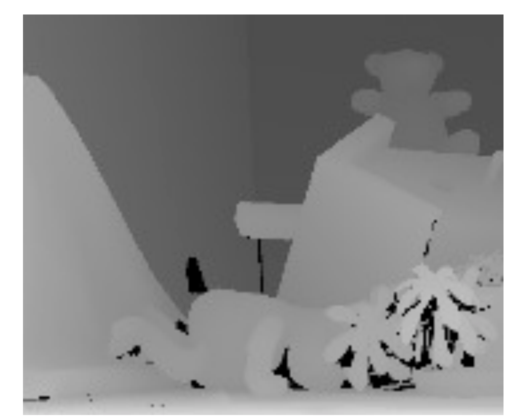
stereo and 3D

problem

s... supports stereo correspondence







not a unified system: another method used in isolation that performs redundant computation!

http://visi... tereo/

http://w... ff/bp/

clustering learning: stereo frames input

solution

same patch location for multiple frames

run k-means
on group of
patches

as was done for
motion features!

deep networks for stereo/ disparity

solution

[DARPA LAGR Program: Learning Applied to Long-Range Vision using a Collision-Free Navigation Platform]

P. Sermanet, R. Hadsell, M. Scoffier, M. Grimes, J. Ben, A. Erkan, C. Crudele, U. Muller, Y. LeCun, in video competitions of Association for the Advancement of Artificial Intelligence (AAAI) and Learning and Adaptive Behavior in Robotic Systems (LAB-RS)

deep networks for vision systems!

solution

compute with
ONE network
NOT IN isolation:

- segmentation
- stereo, optical flow
- recognition
- tracking

...