Data mining for large scale image and video analysis

Pinar Duygulu

Large volumes of video

flickr

YAHOO!
SCREEN

- For YouTube alone
  - More than 1 billion unique user visits each month
  - Over 6 billion hours of video are watched each month
  - 100 hours of video are uploaded every minute

http://www.youtube.com/yt/press/statistics.html

Applications

- Analyzing video archives

First appearance of N. Sarkozy on TV

Sociology research: Influence of character smoking in movies

Education: How do I make a pizza?

- Surveillence

Where is my cat?

Predicting crowd behavior
Counting people

- Graphics

Motion capture and animation

Slide Credit: I.Laptev

Available Datasets

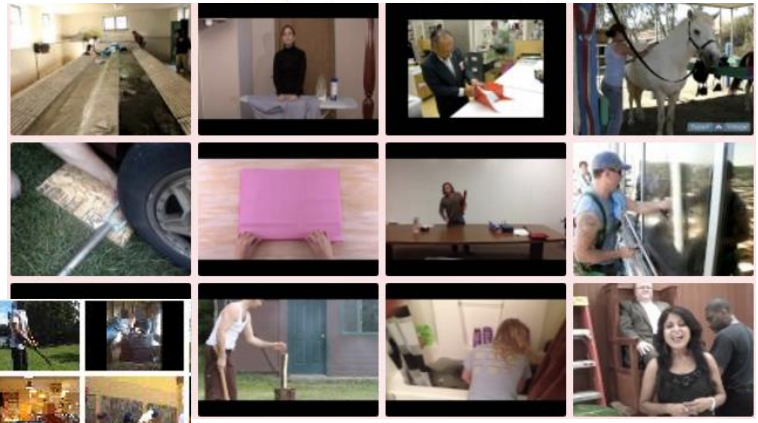
Dataset	#Classes
KTH	6
Weizmann	9
IXMAS	11
Hollywood	8
UCF Sports	9
Hollywood2	12
UCF YouTube	11
MSR	3
Olympic	16
UCF50	50
HMDB51	51

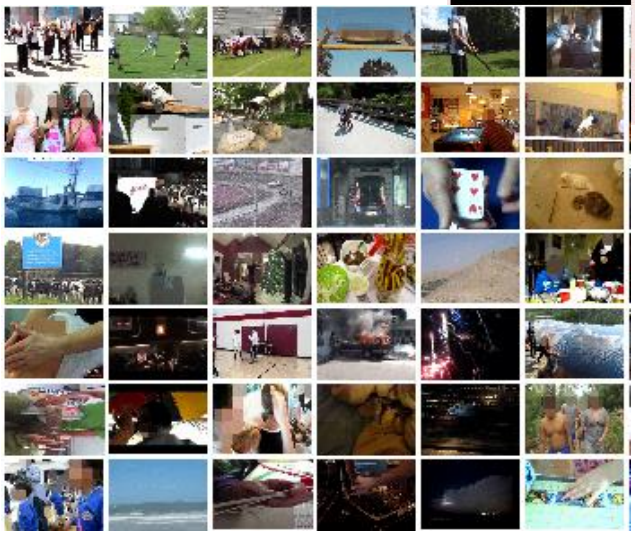
Yahoo! Recently released 100 million Flick data

http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/

Recent datasets

Veri Kümesi	#Sınıf	#Video
UCF-101	101	13320
ActivityNet	200	20000
FCVID	239	91223
Sports-1M	487	1M





Yahoo! Flickr
100M dataset
418.507 labeled video

Videos in the wild

- Unrestricted type of events with various activities

Harlem Shake : http://www.youtube.com/watch?v=4hpEnLtqUDg

Our attempts

- Videos as sequence of frames
  - Detect concepts in each frame
  - Utilize image search engines
- Discover important knowledge from videos itself
  - Discriminate parts
- Understand actions in videos
  - Simple but effective descriptors

Utilizing large volumes of weakly labeled images

Pinar Duygulu, November 2016, Ankara

Utilize image search results

Query : Ankara







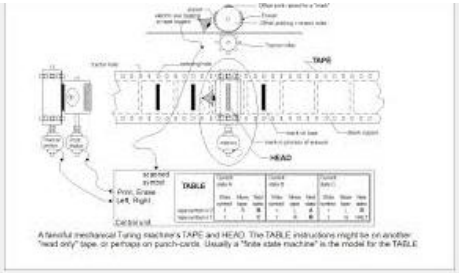





Single Dominant Category

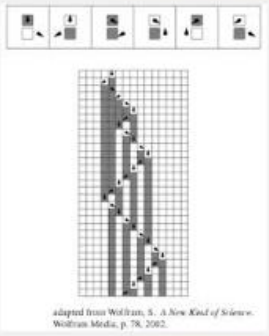
Query : Turing





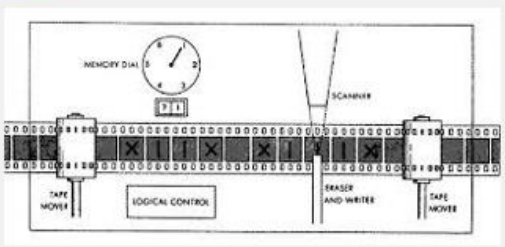




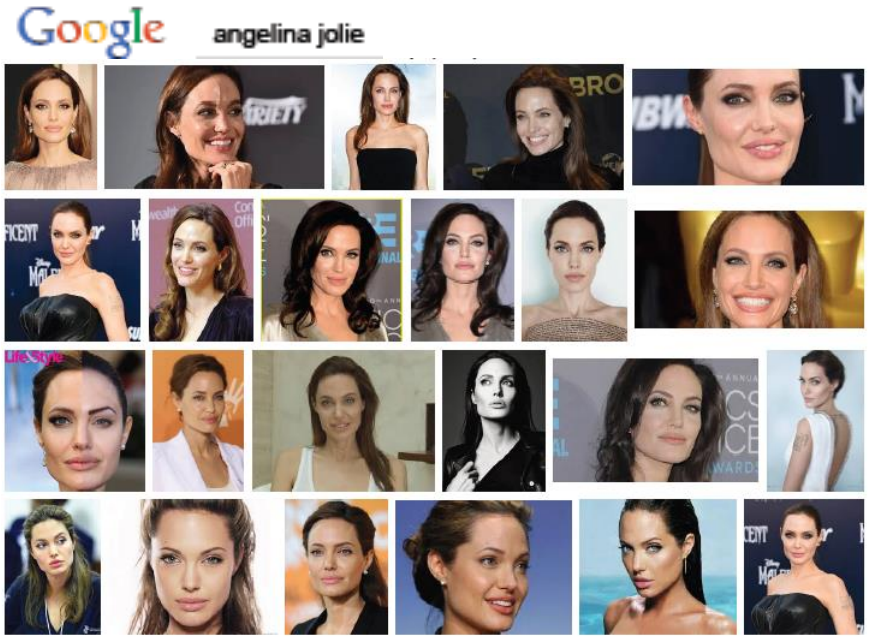









Google results



E.Golge, P.Duygulu. FAME: Face Association Through Model Evolution. In CVPR Workshops, 2015

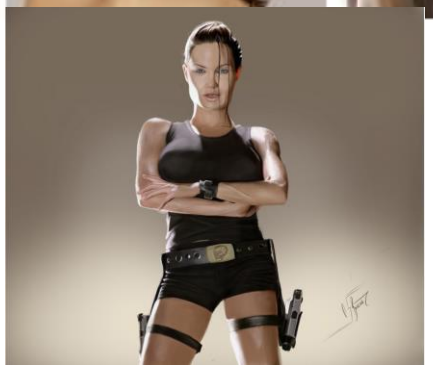
Pinar Duygulu, November 2016, Ankara









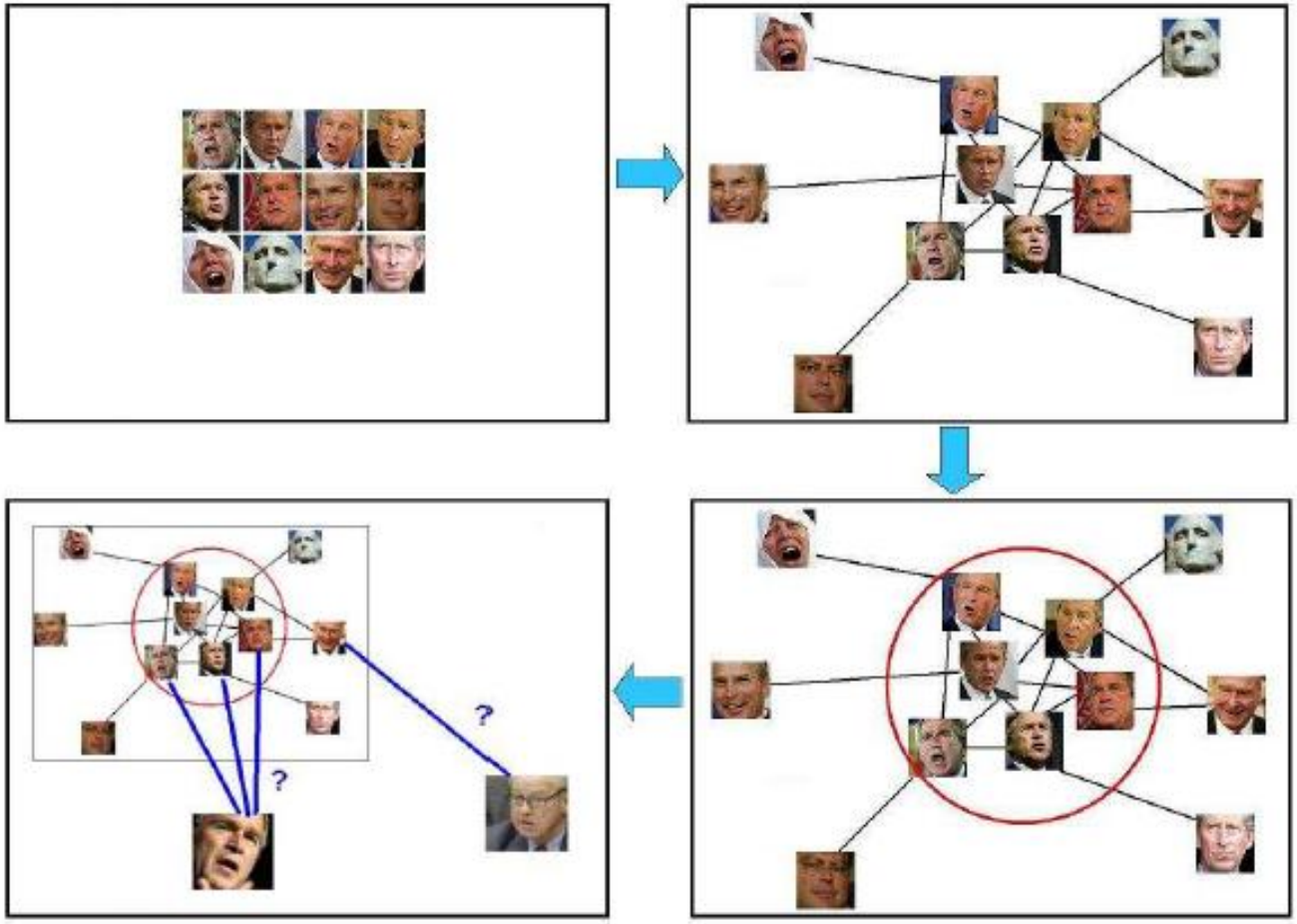



Pinar Duygulu, November 2016, Ankara

Pinar Duygulu, November 2016, Ankara

Pinar Duygulu, November 2016, Ankara

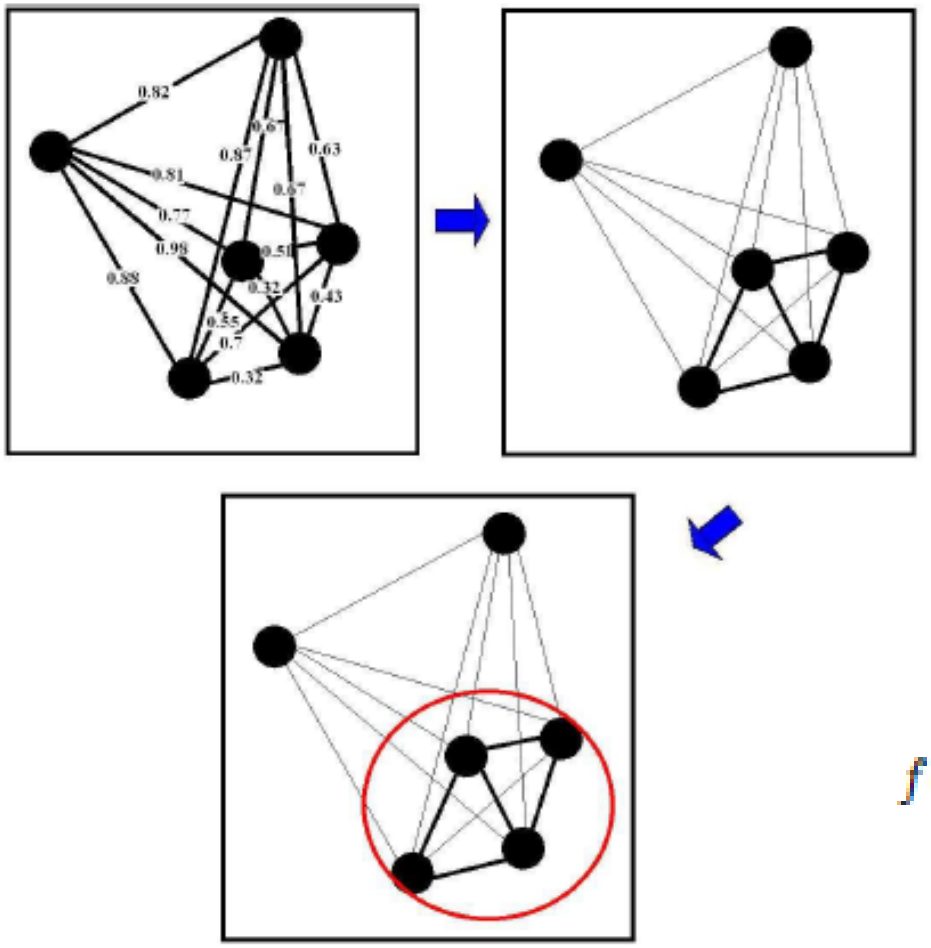
Naming faces



Among the faces associated with a name find the correct subset : The most similar subset of faces

Ozkan, D., Duygulu, P., ”Interesting Faces: A Graph Based Approach for Finding People in News”, Pattern Recognition, 2010
Ozkan, D., Duygulu, P., ”A Graph Based Approach for Naming Faces in News Photos”, CVPR, 2006
Ozkan, D., Duygulu, P., ”Finding People Frequently Appearing in News”, CIVR, 2006

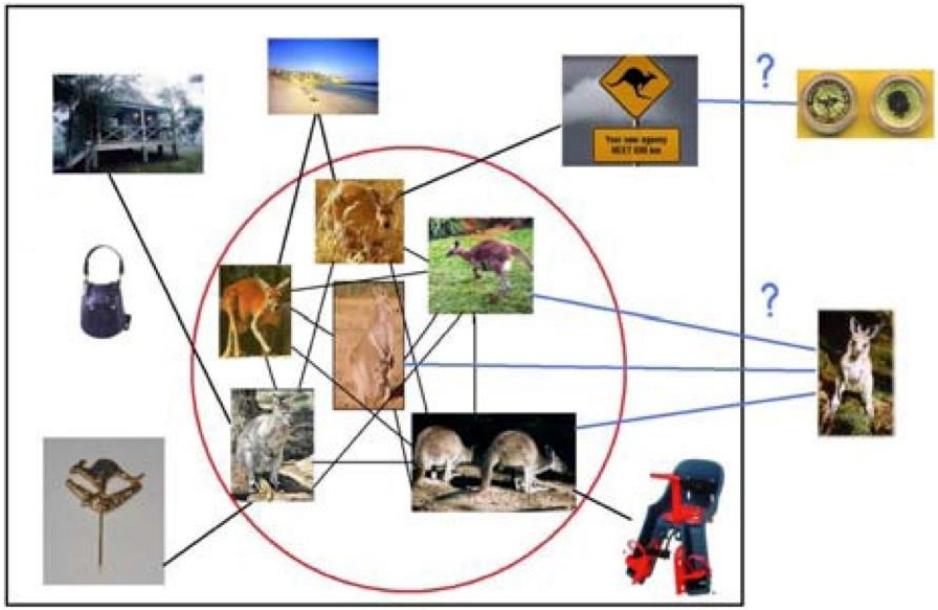
Finding Densest component



$$f(S) = \frac{|E(S)|}{|S|},$$

Node with the minimal degree is removed at each iteration (Charikar, 2000)

Image Re-ranking



Zitouni, H., Sevil, S. G., Ozkan, D., Duygulu, P., "Re-ranking of Image Search Results using a Graph Algorithm", ICPR 2008

Multiple meanings/variations











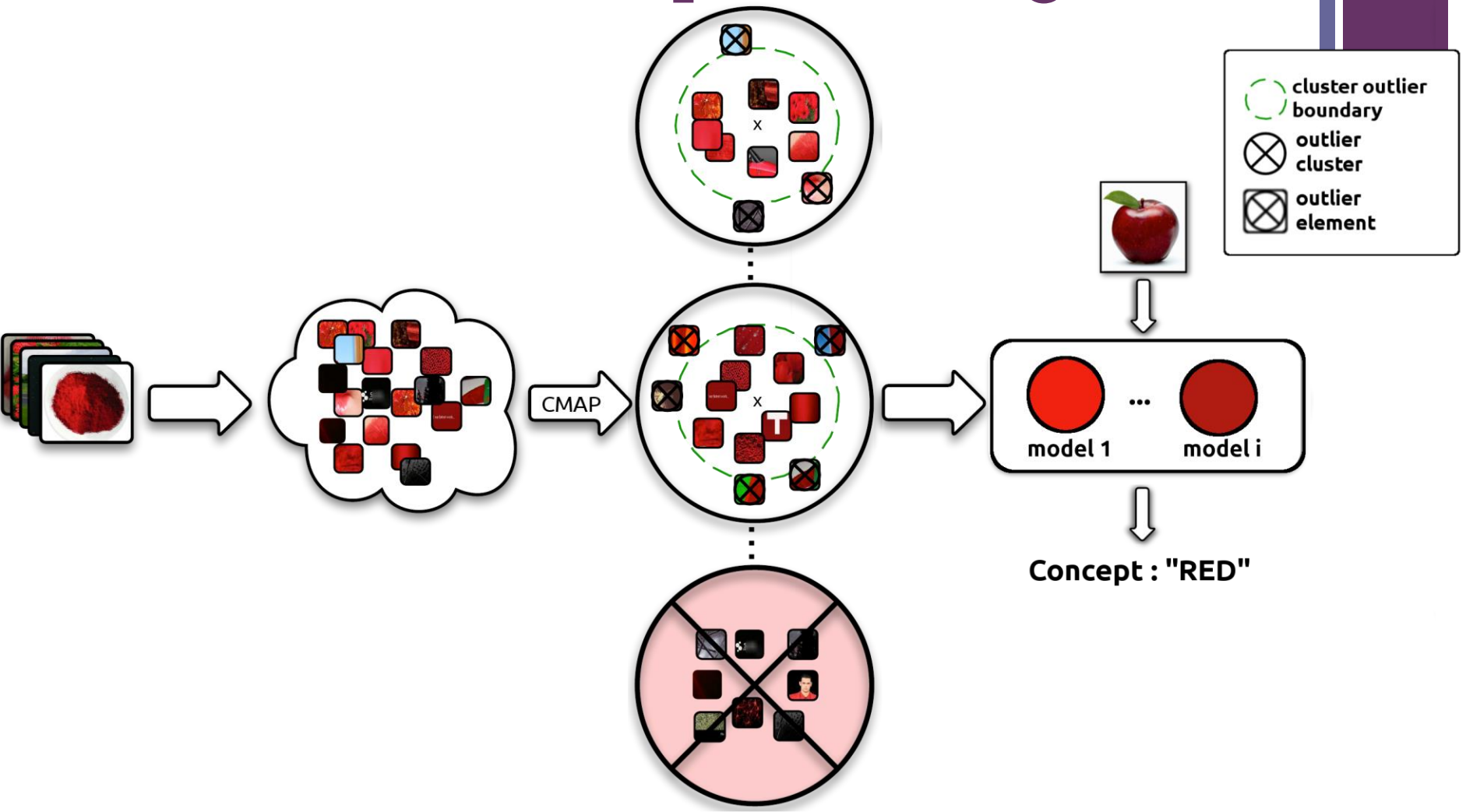







The concepts are observed in different forms requiring grouping and irrelevant elements to be eliminated.

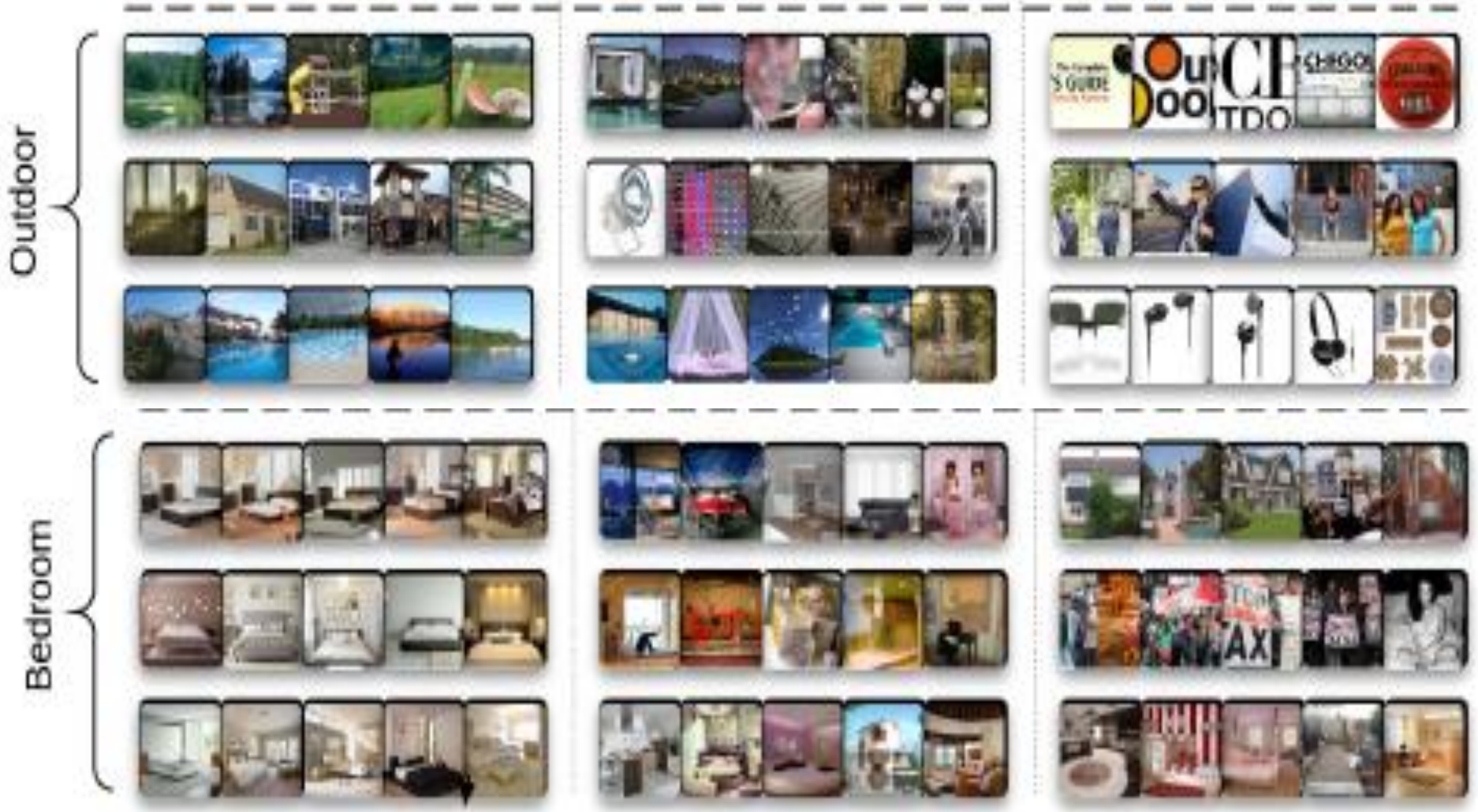
CMAP for Concept Learning



Golge, E., Duygulu, P., "Concept Maps: Mining Noisy Web Data for Concept Learning ", accepted to ECCV 2014

Color and Texture Attributes

Scene Concepts



Attribute and Scene Learning

Attribute learning for object recognition

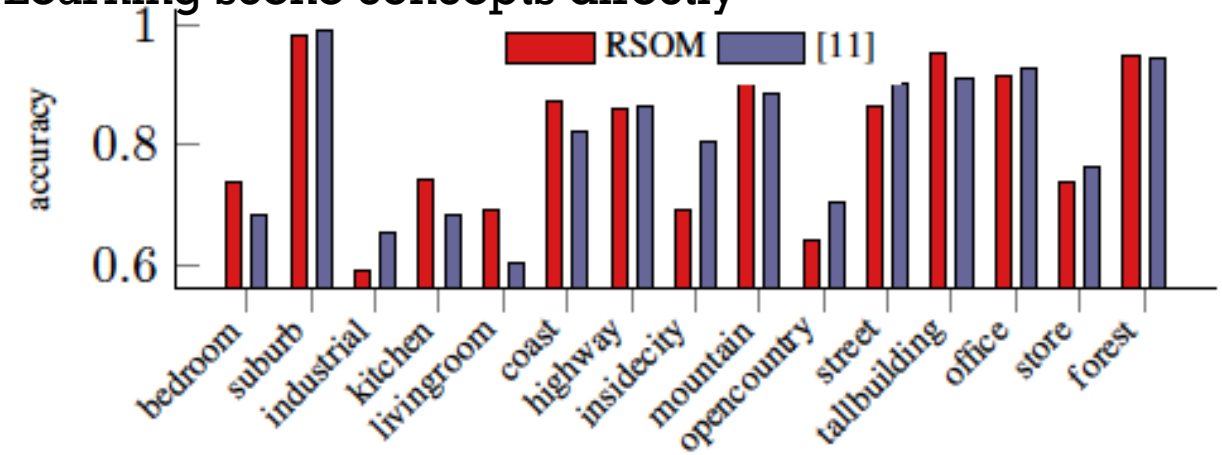
Method	RSOM-M	RSOM	PLSA-reg [22].
cars	0.97	0.92	0.93
shoes	1.0	0.97	0.99
dresses	1.0	1.0	0.99
pottery	0.98	0.92	0.94
overall	0.99	0.95	0.96

Attribute based scene recognition

Method	MIT-indoor [17]	Scene-15 [11]
RSOM-A	46.2%	82.7%
RSOM-S	-	80.7%
RSOM-S+HM	-	81.3%
Li *et al.* [12] VQ	47.6%	82.1%
Pandey *et al.* [16]	43.1%	-
Kwitt *et al.* [9]	44%	82.3%

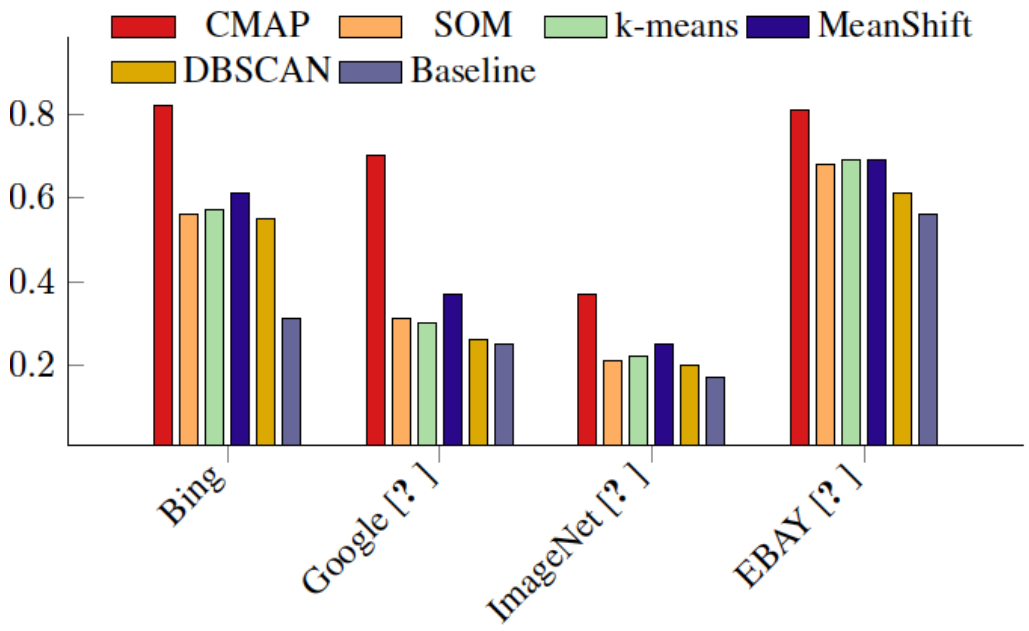
On ImageNet: 37.4% (RSOM), 36.8% (Russakovsky & Fei-Fei, 2012)

Learning scene concepts directly

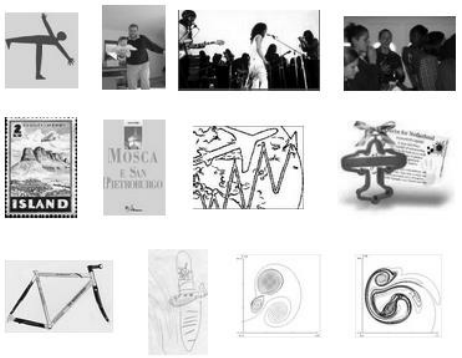
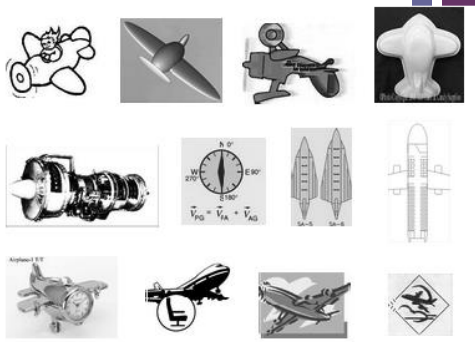


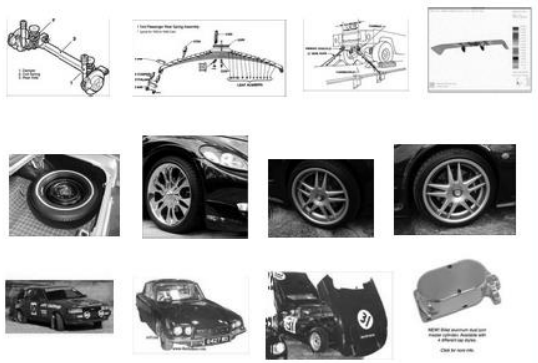
[17] Quattoni and Torralba,"Recognizing Indoor Scenes". 2009
[11] Lazebnik, Schmid, Ponce, "Beyond Bags of features: Spatial pyramid matching for recognizing natural scene categories", CVPR 2006
[22] Van de Weijer, Schmid, Verbeek, Larlus, "Learning Color Names for Real-world Applications", 2009

Comparison with other clustering methods



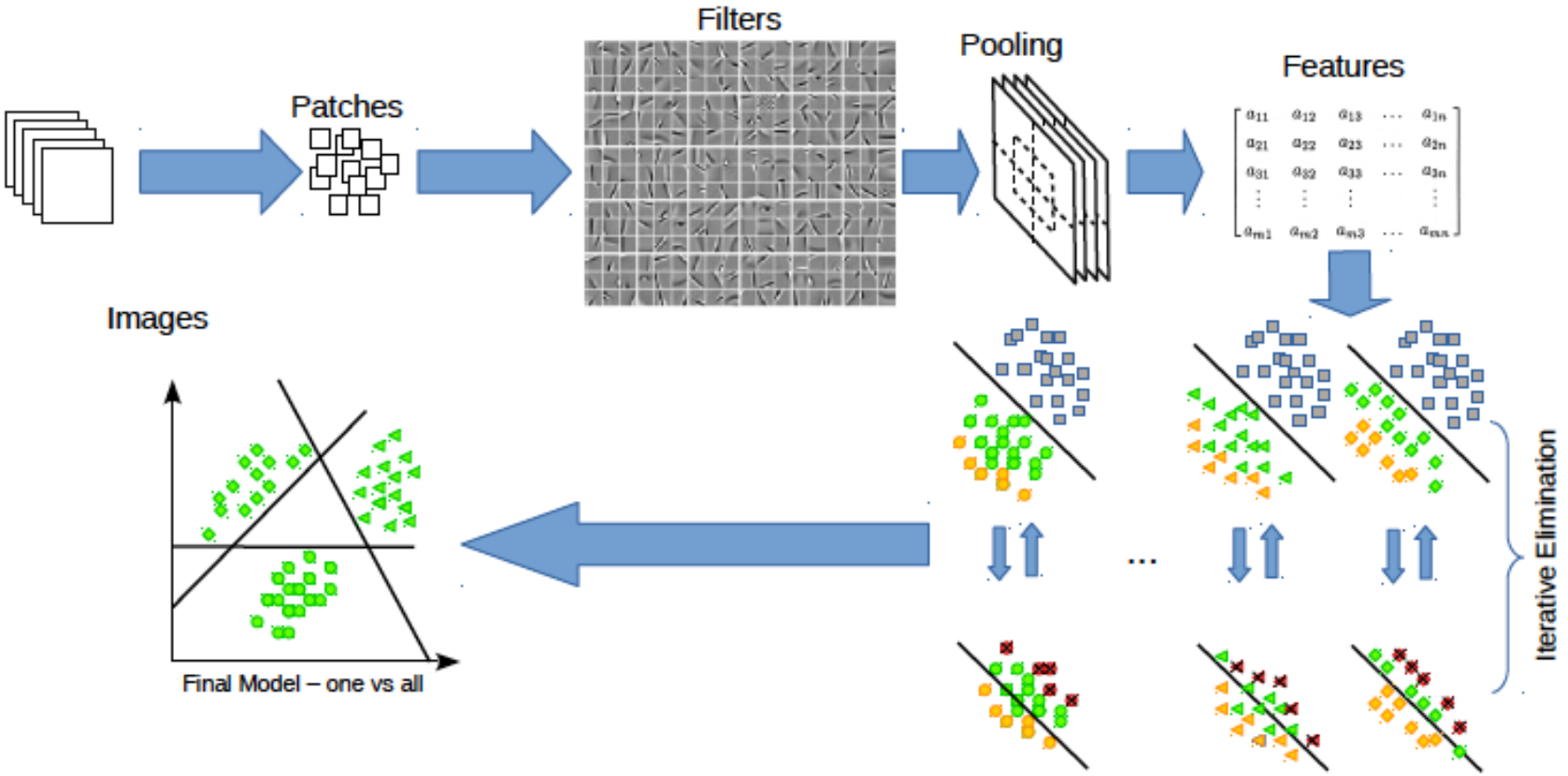
Objects

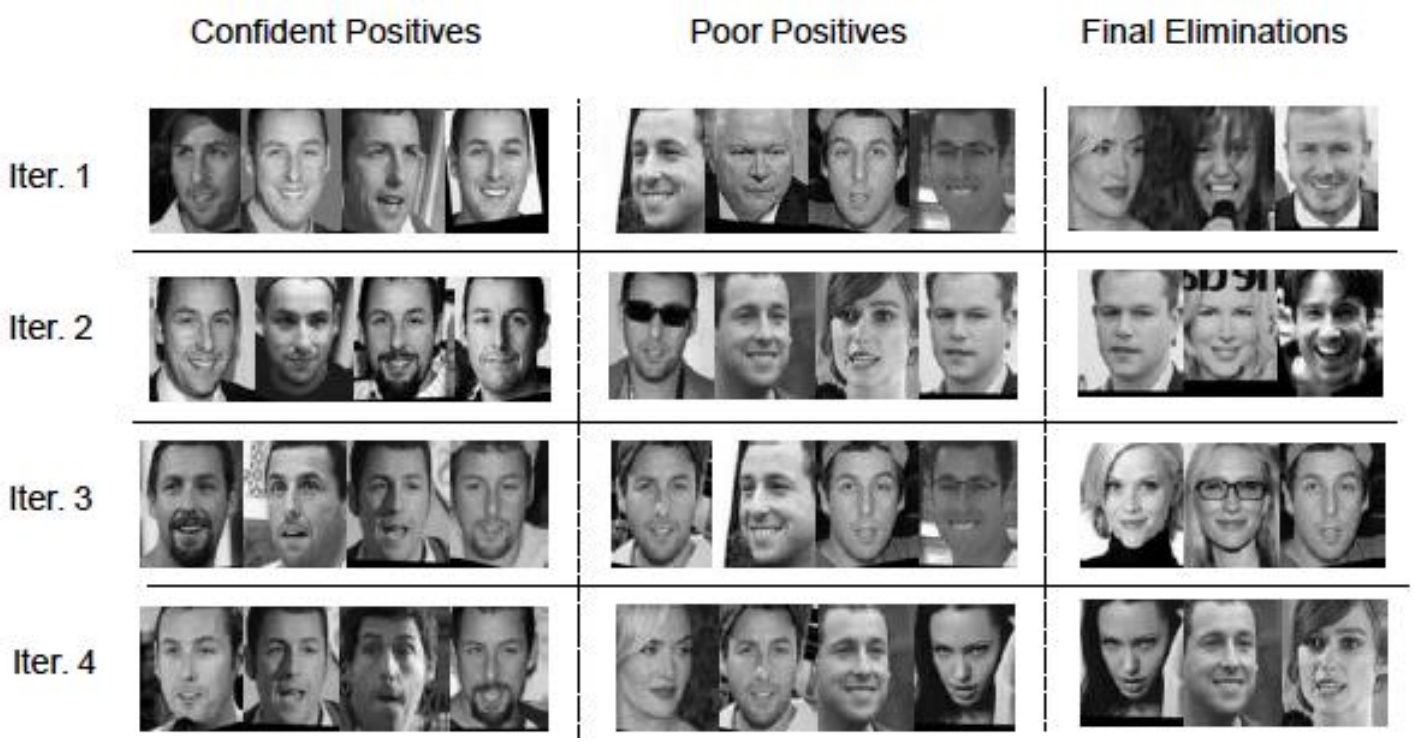







Faces

FAME: Face Association Through Model Evolution



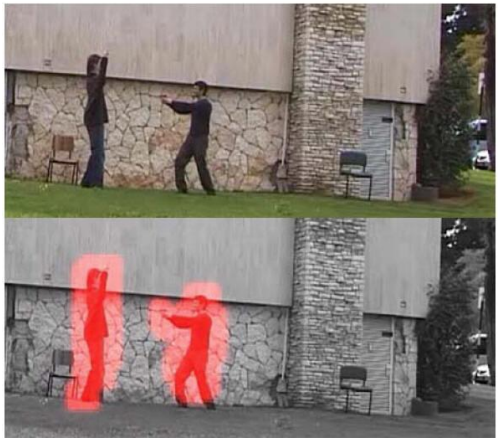


Method	Pinto *et al.* [34] (S)	Pinto et al.[34](M)	face.com [34]	Becker *et al.* [3]	*FAME*
Accuracy	75.6	87.1	82.1	85.9	90.75

Capture usualness in unusual videos

Pinar Duygulu, November 2016,
Ankara

Pinar Duygulu, November 2016, Ankara



Boiman and Irani ICCV 2005

Roshtkhari and Levine, CVPR 2013

Ito, Kitani, Bagnell, Hebert, 2012

Zhao, Fei-Fei, Xing, CVPR 2011

Usual versus unusual

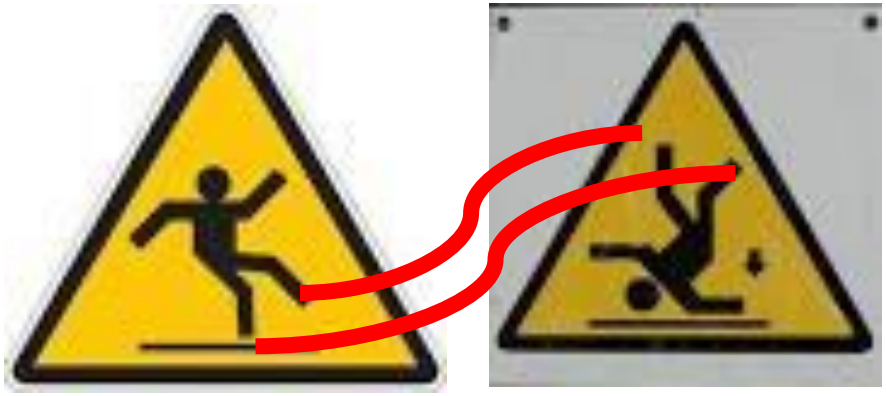





Pinar Duygulu, November 2016, Ankara

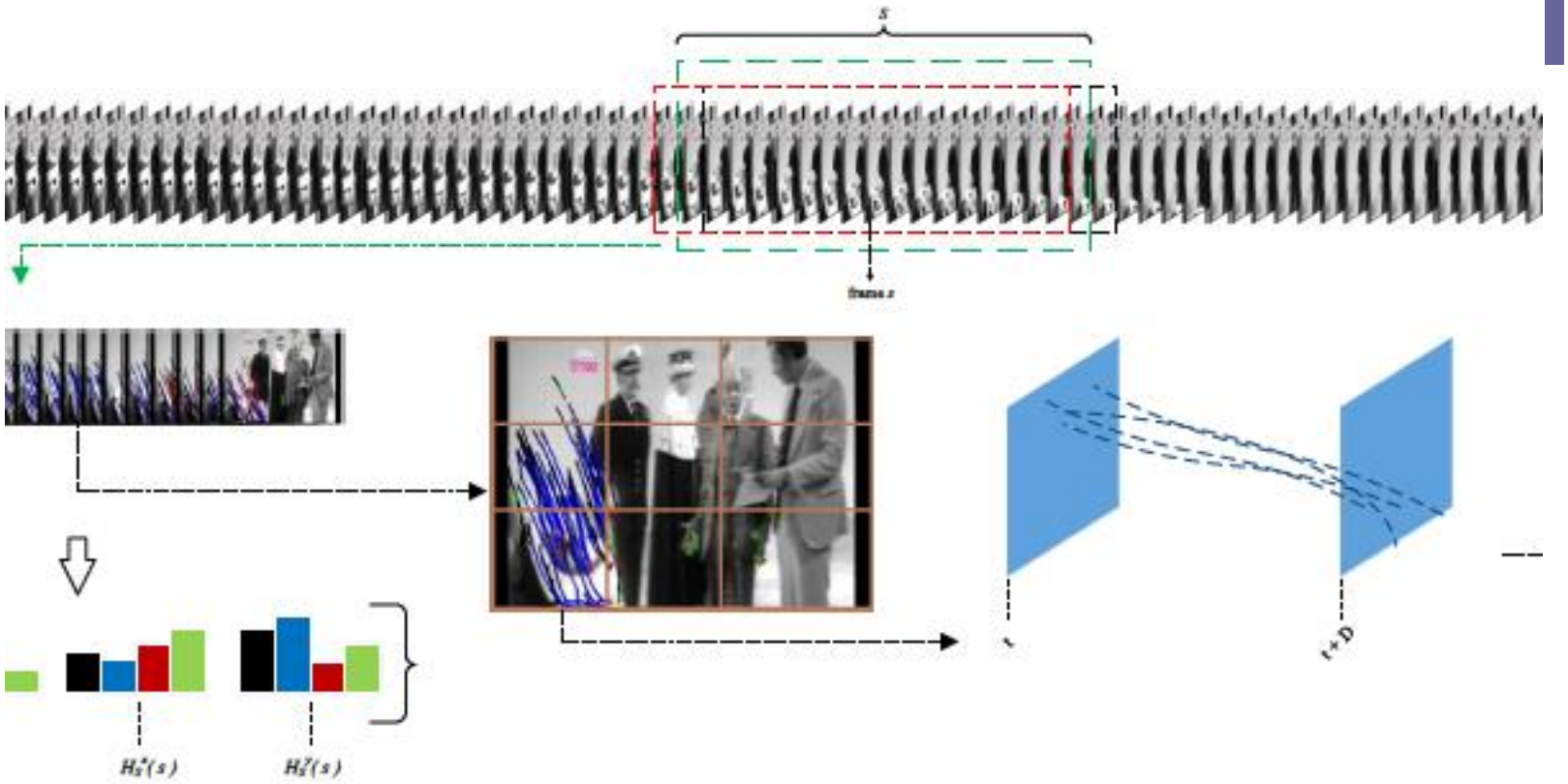
Usual versus unusual

Rapid motions

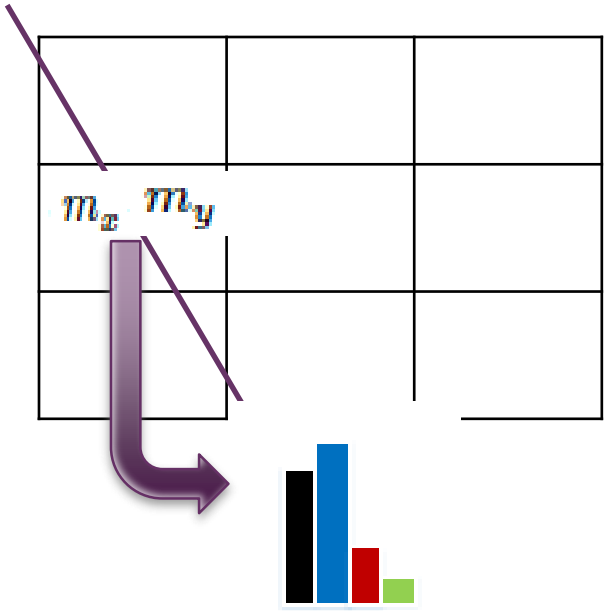


Fast speed
Large spatial extension

Trajectory Snippet Histograms



Representation



$$H_S^l = \sum_{t=s-(\|S\|/2)}^{s+(\|S\|/2)} H_S^l(t) \qquad H_S = (H_S^l, H_S^x, H_S^y)$$

$$H_S^l(t) = (H_S^l(t)_{[1,1]}, \ldots H_S^l(t)_{[1,N]}, \ldots H_S^l(t)_{[N,N]})$$

Velocity and spatial extension of the motion

$$T = (P_t, \ldots, P_{t+D-1}) \quad P_t = (x_t, y_t)$$

$$m_x = \frac{1}{D}\sum_t^{t+D-1} x_t, v_x = \frac{1}{D}\sum_t^{t+D-1} (x_t - m_x)^2$$

$$m_y = \frac{1}{D}\sum_t^{t+D-1} y_t, v_y = \frac{1}{D}\sum_t^{t+D-1} (y_t - m_y)^2,$$

$$l = \sum_t^{t+D-1} \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2}$$





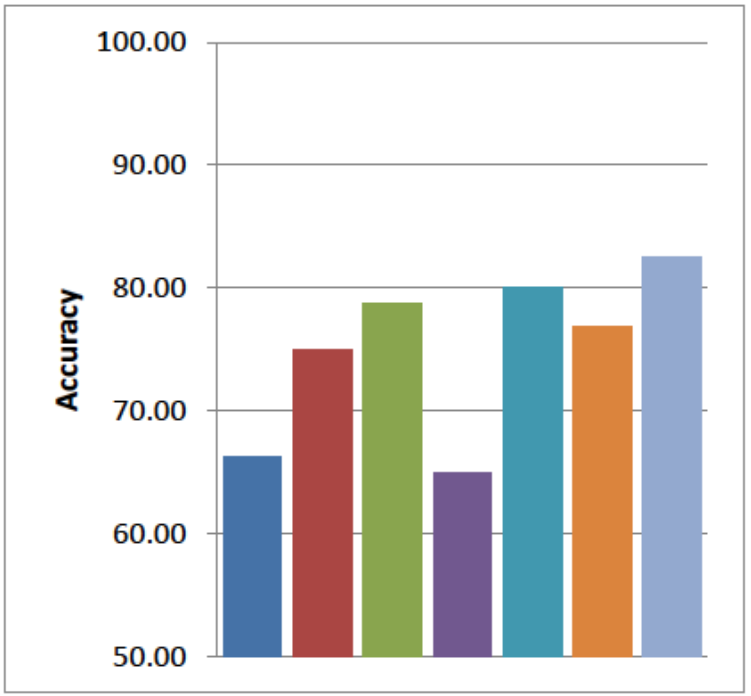




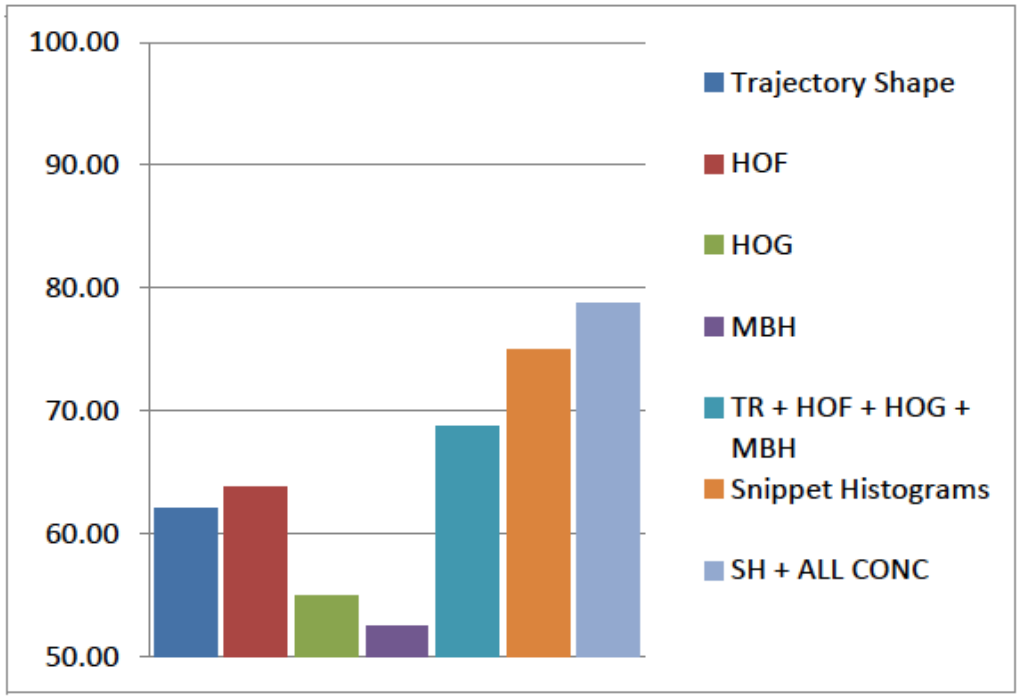
Pinar Duygulu, November 2016, Ankara

Pinar Duygulu, November 2016, Ankara

Classification



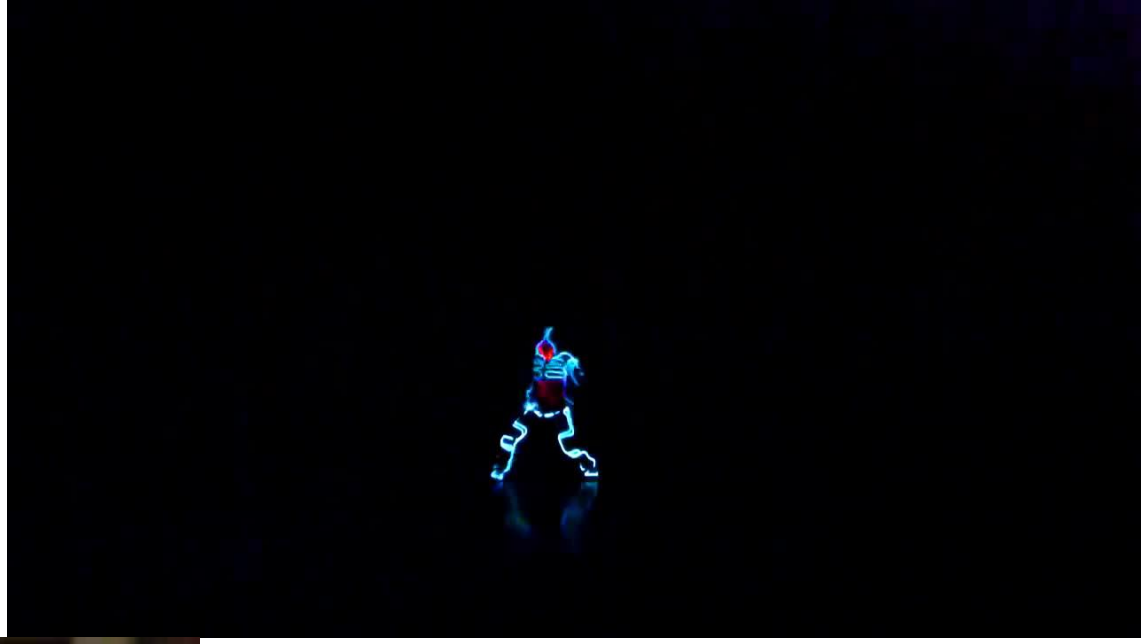
People Falling



Funny videos

Best SH: people falling 75%, funny videos76.25%
HOG3D : people falling 65%, funny videos 73.75%

Failure cases



Pinar Duygulu, November 2016, Ankara

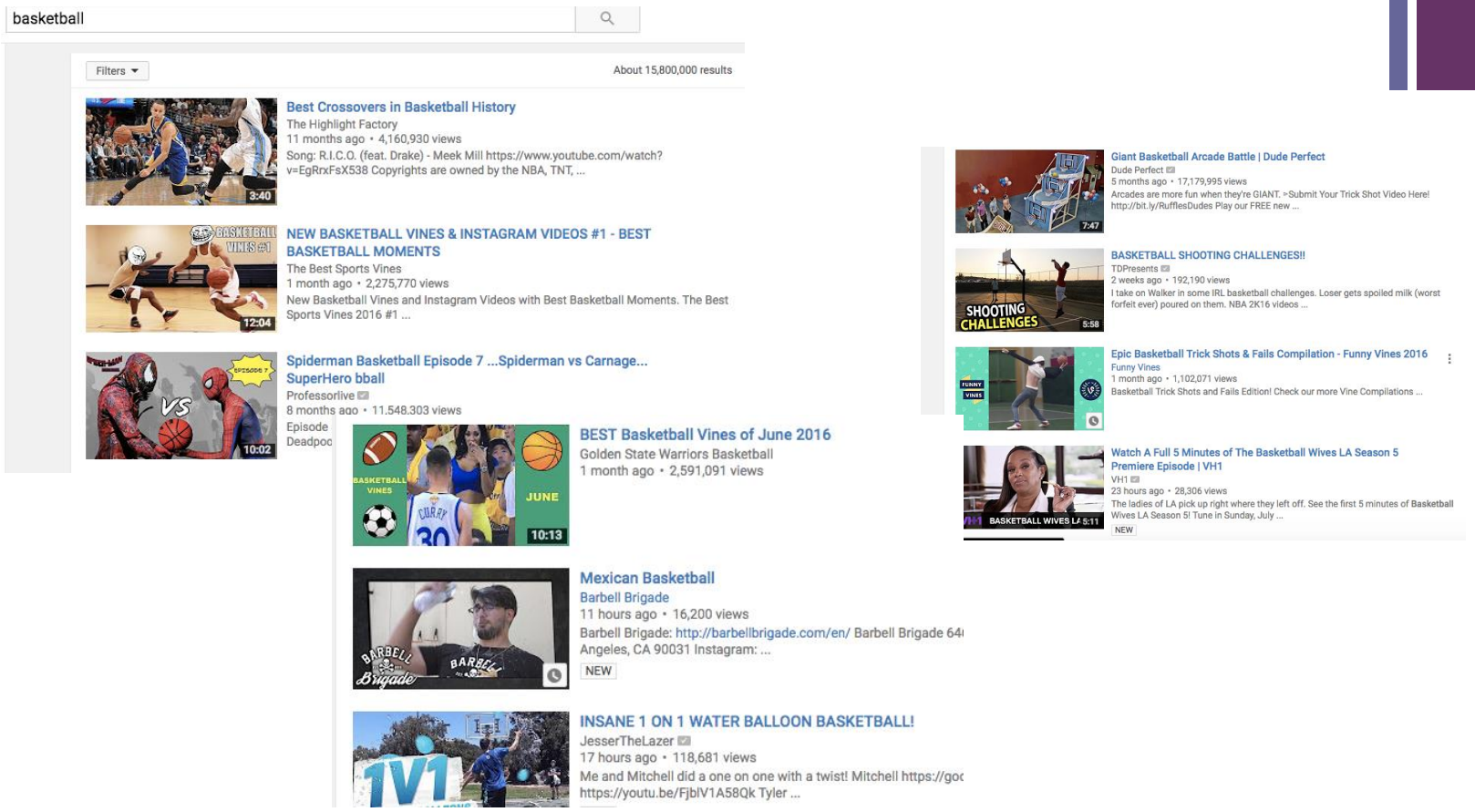
Pinar Duygulu, November 2016, Ankara

Limitations

Finding discriminative parts in videos

Pinar Duygulu, November 2016,
Ankara

Weakly labeled videos



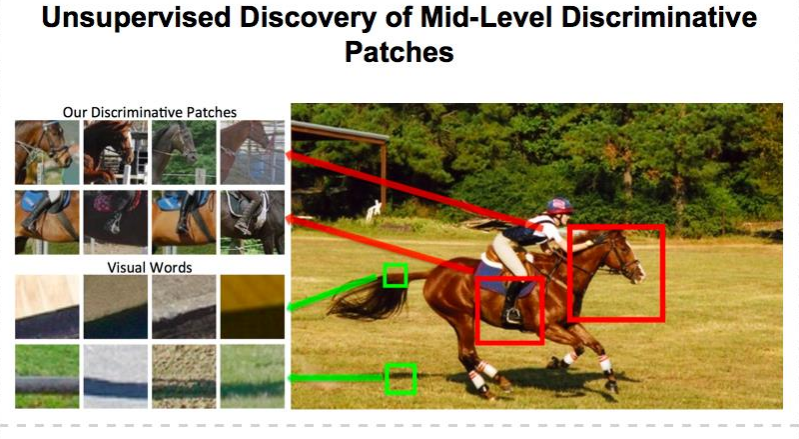
Pinar Duygulu, November 2016, Ankara

Birthday event

Mum üfleme

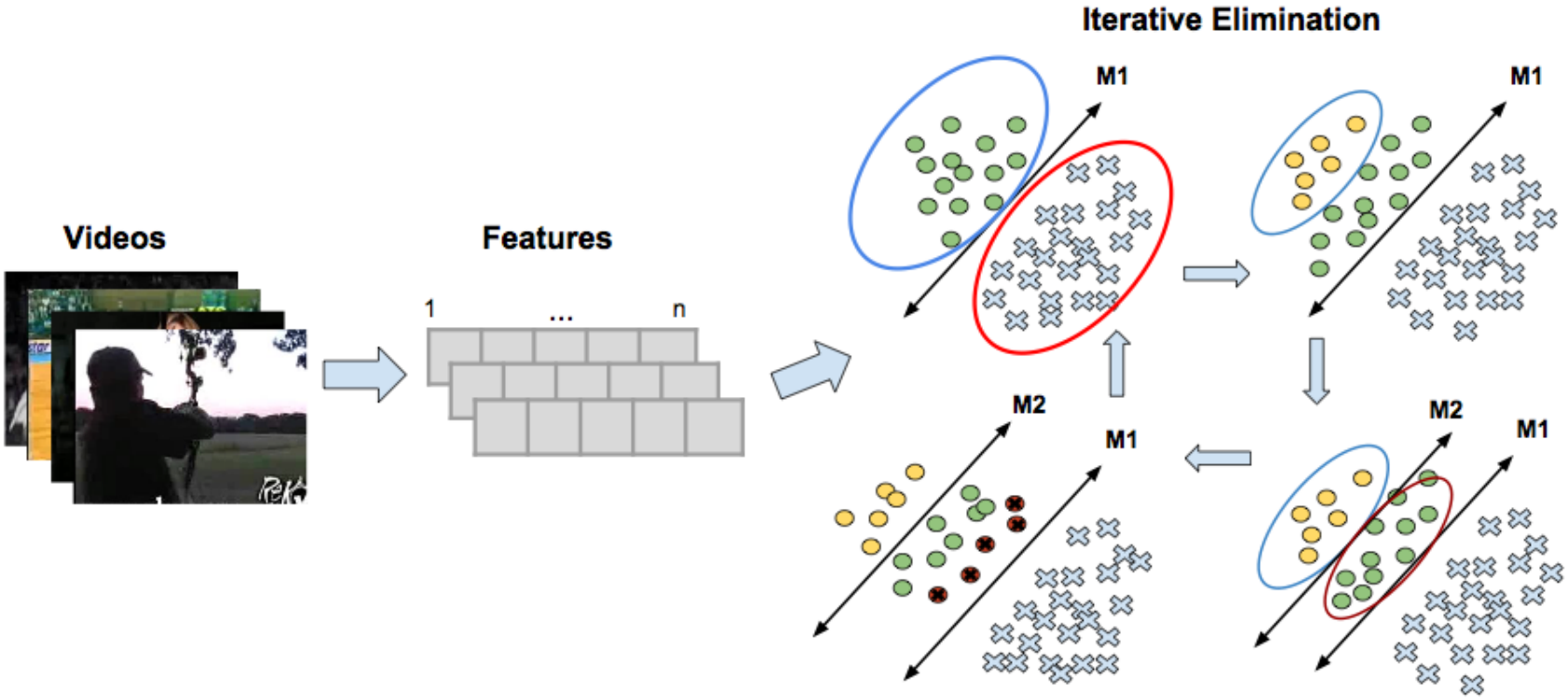
Finding discriminative parts

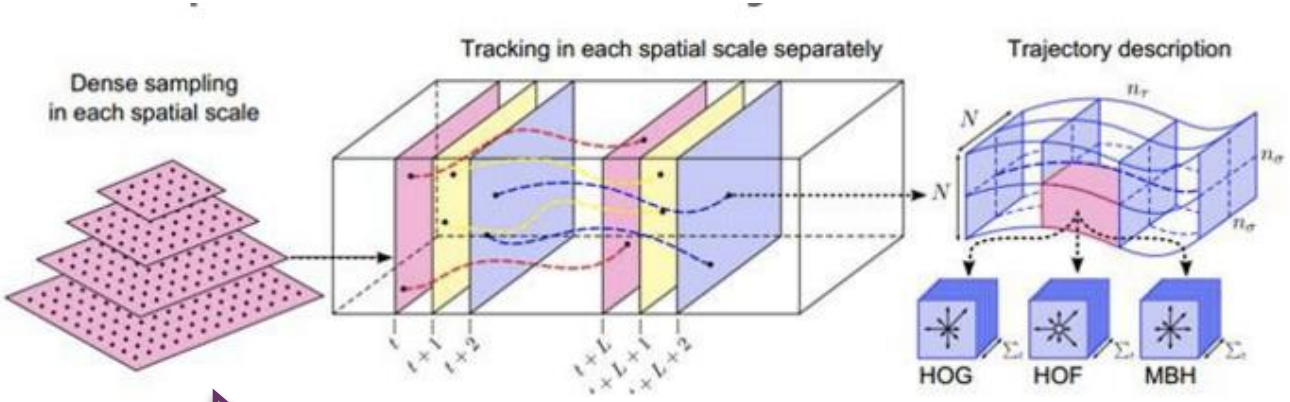
Singh ECCV 2012

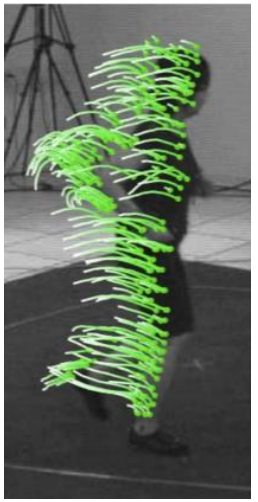


Jain CVPR 2013

AME



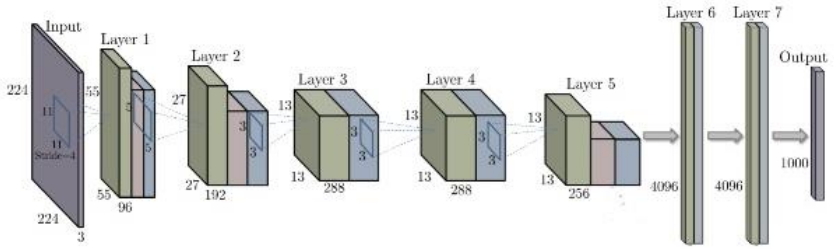




Input Video

CNN FC-6 Layer Features

Classification



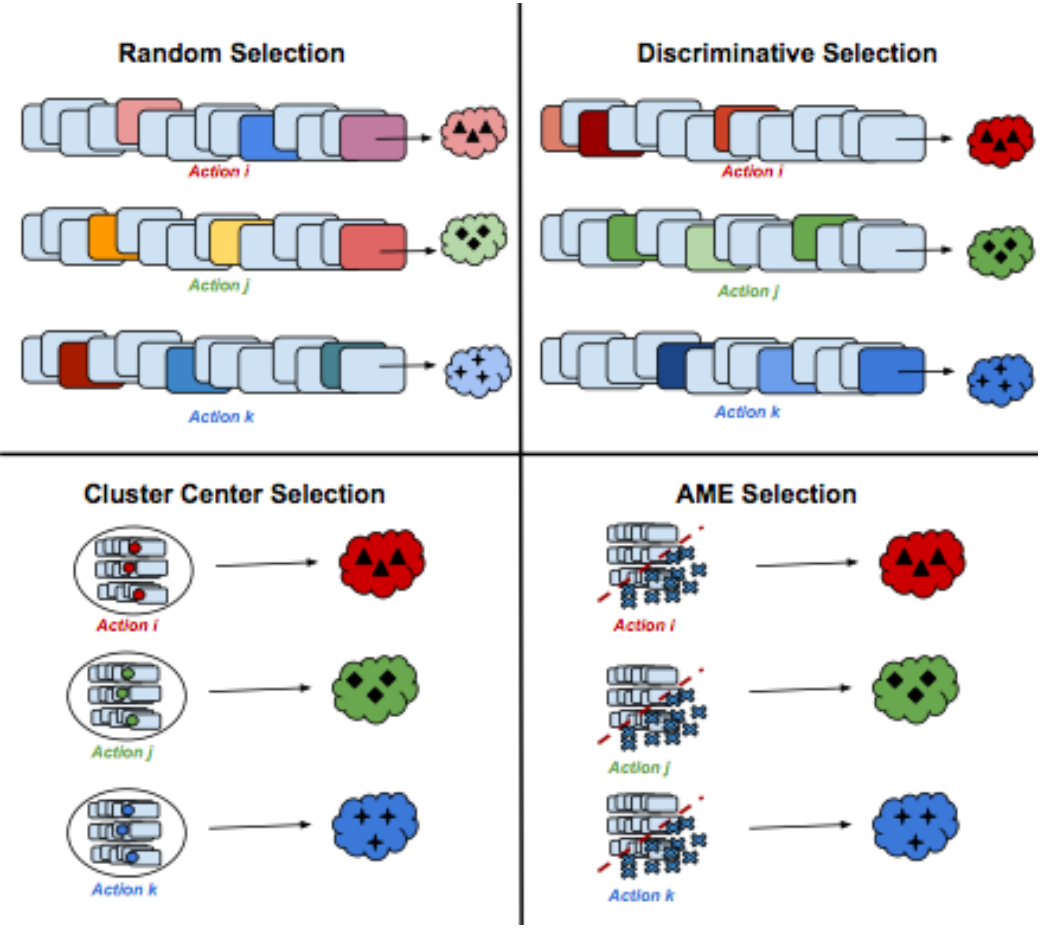
AlexNet [8,9,10]

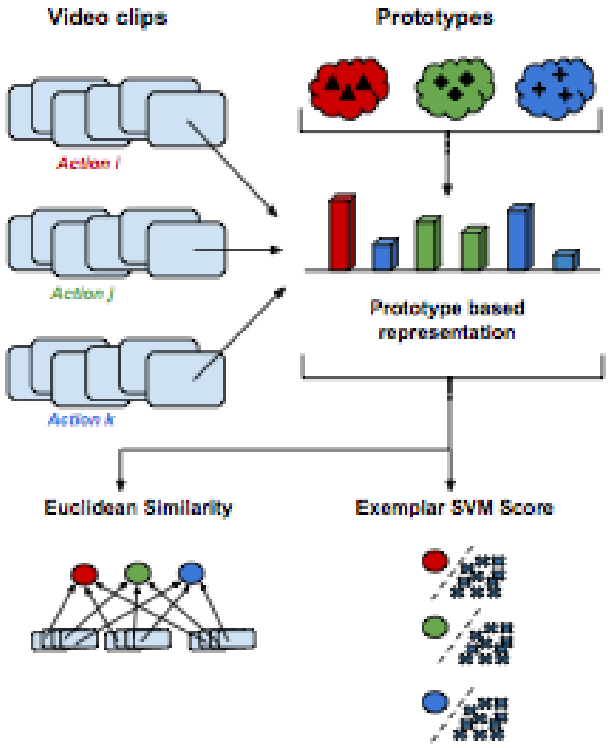
3D ConvNets [11,12]

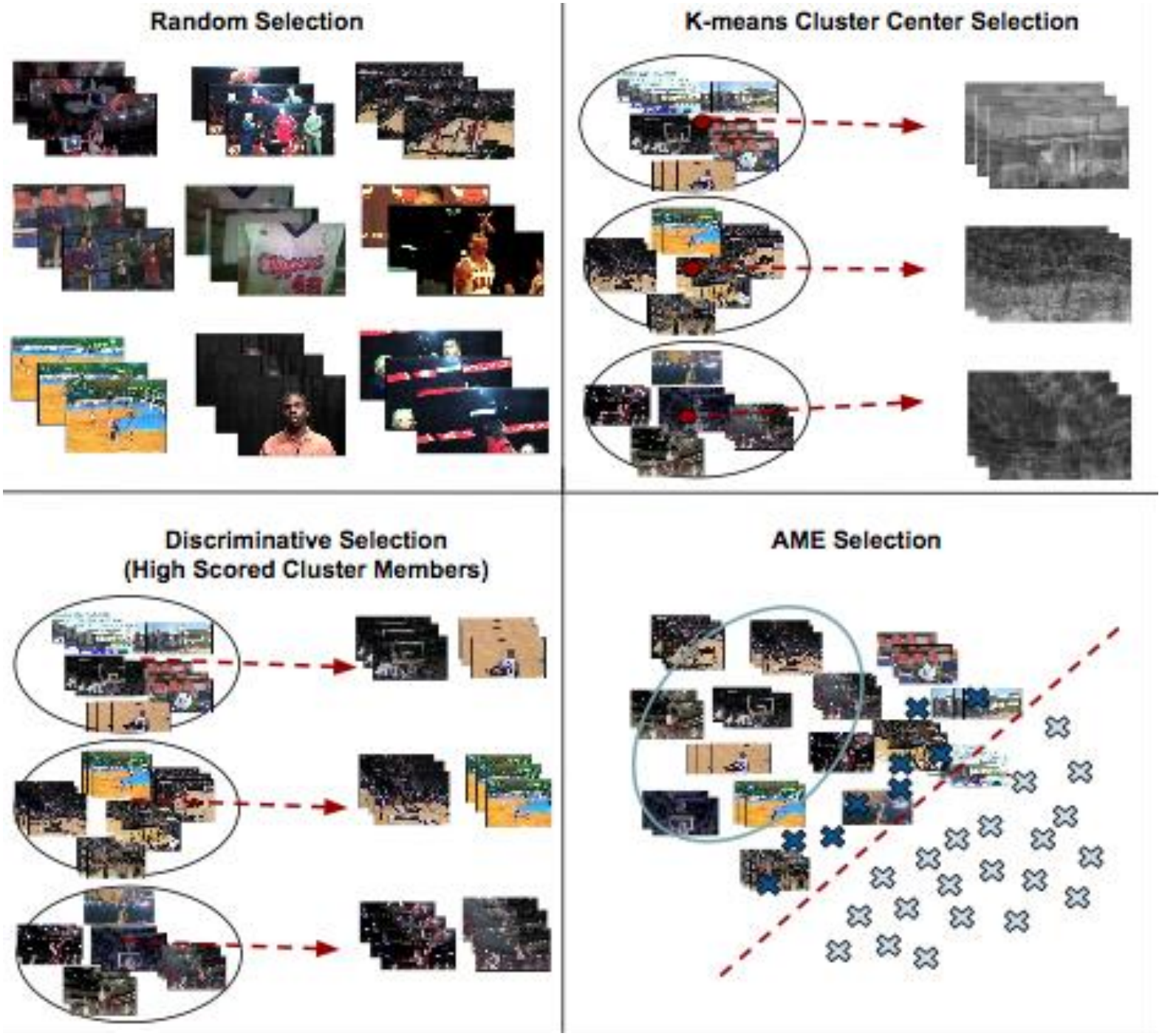
Example of successful eliminated instances by AME[1] for ActivityNet action classes.(1) "Archery" class. Baseline: 31.57%, AME: 44.73%. (2) "Checking tires" class. Baseline: 26.82%, AME: 41.46%. (3) "Platform diving" class. Baseline: 56.66%, AME: 73.33%

Example of unsuccessful eliminated instances for ActivityNet class "Windsurfing" with AME[1]. Baseline: 74.07%, AME[1]: 66.66%

Prototypes


Random Selection
Discriminative Selection
Cluster Center Selection
AME Selection
Action i
Action j
Action k


Video clips
Prototypes
Action i
Action j
Action k
Prototype based representation
Euclidean Similarity
Exemplar SVM Score


Random Selection
K-means Cluster Center Selection
Discriminative Selection
(High Scored Cluster Members)
AME Selection

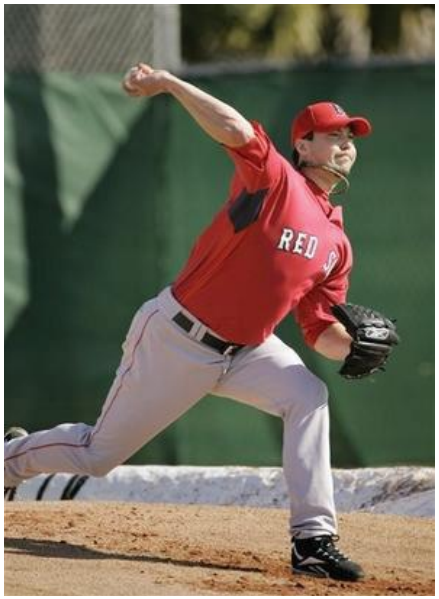
Human Activity Analysis

Pinar Duygulu, November 2016,
Ankara

What do these people do?

running

walking



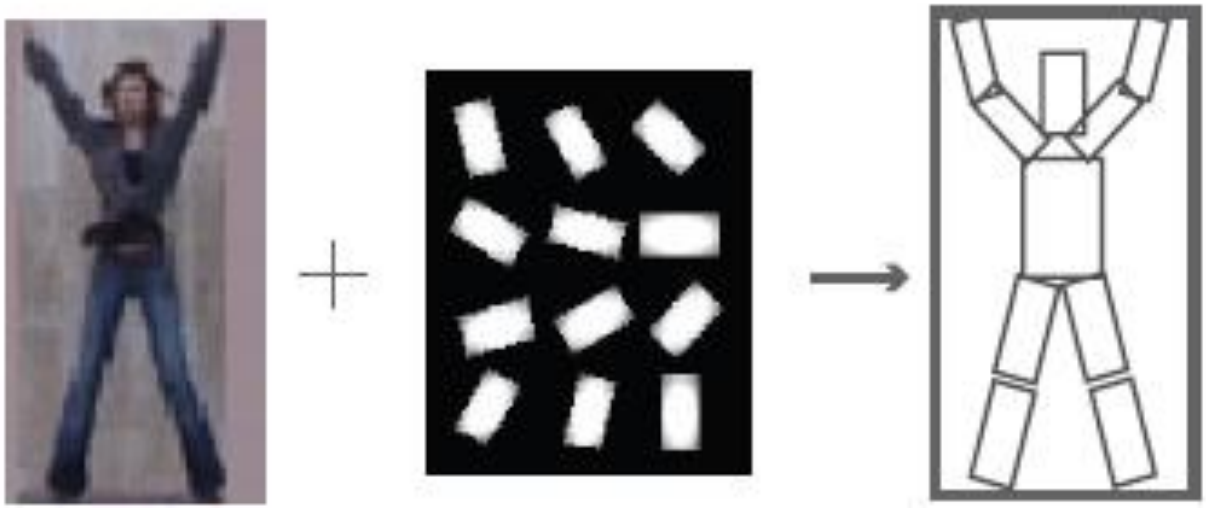
throwing

crouching

- Pose tells a lot about the actions.
- How can we describe the pose?

Pose as a Collection of Rectangles

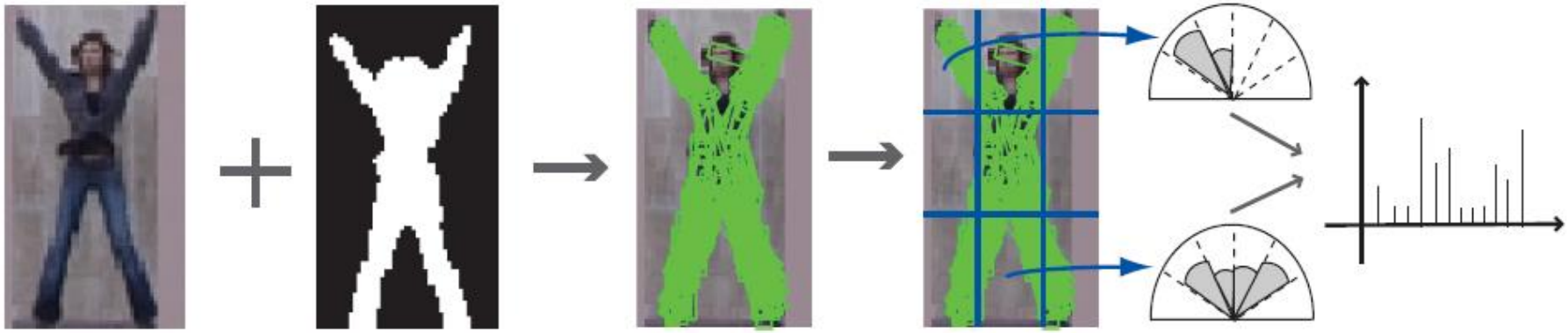
- Human body is composed of cylindrical parts.
- The projection of a cylinder on 2D is a rectangle.
- Body can be thought as a collection of rectangular regions
- We can represent the pose based on the orientation of these rectangles



Ikizler, N. Duygulu, P. ''Human Action Recognition Using Distribution of Oriented Rectanguar Patches'', Proc. 2nd Workshop on Human Motion: Understanding, Modeling, Capture and Animation, In conjunction with ICCV2007

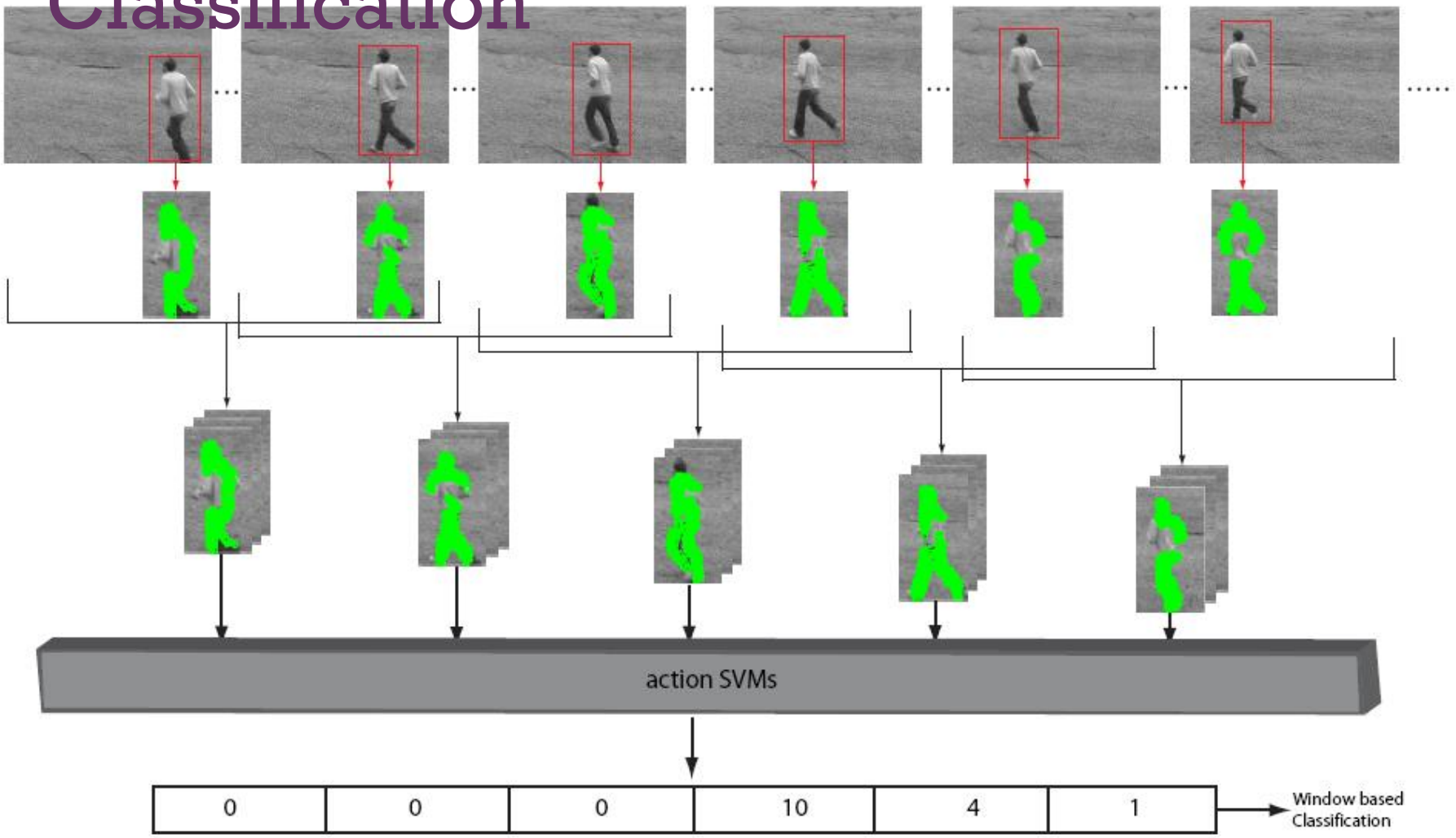
Ikizler, N. ve Duygulu P., ''Histogram of Oriented Rectangles: A New Pose descriptor for Human Action Recognition'', Image and Vision Computing, volume 27, Issue 10, pages 1515-1526, September 2009

Histogram of Oriented Rectangles (HOR)



- Rectangular regions are extracted over silhouettes using convolution of a zero-padded rectangular 2D Gaussian on different orientations and scales
  - 12 angles 15° apart

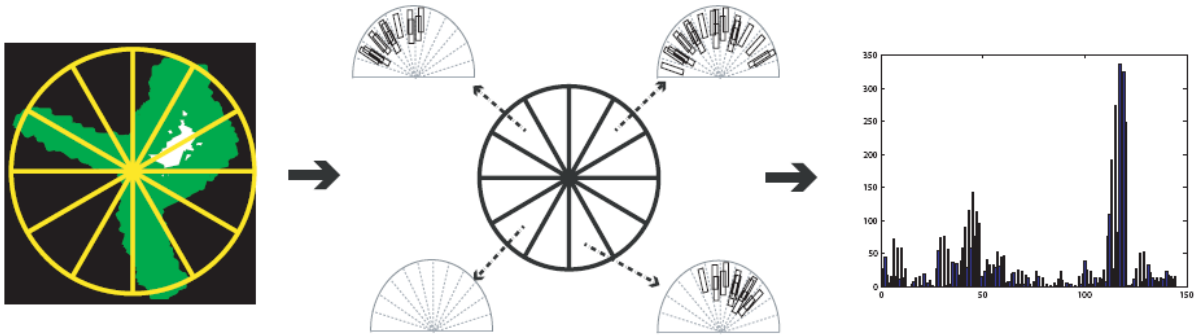
Classification



- Use snippets of frames and form histogram of oriented rectangles over a window (HORW)

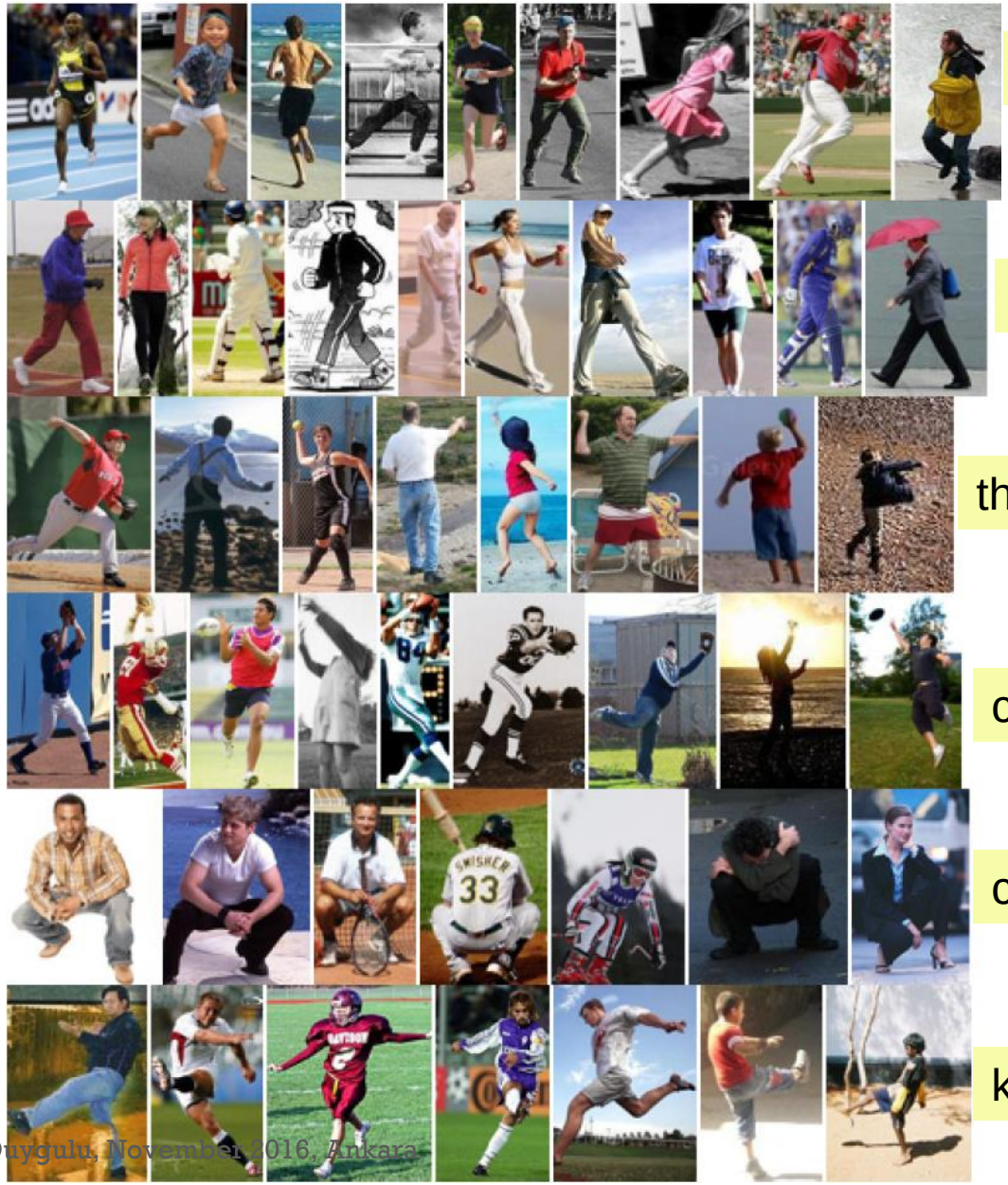
Action Recognition in Still Images

- Pose estimation by Ramanan's method
- Form Circular HORs (CHORs)
- Classification based on LDA+SVM



Ikizler, N., Cinbis, R. G., Pehlivan, S., Duygulu, P., "Recognizing actions from still images",Proc. 19th International Conference on Pattern Recognition (ICPR 2008)

Still Image Results



running

walking

throwing

catching

crouching

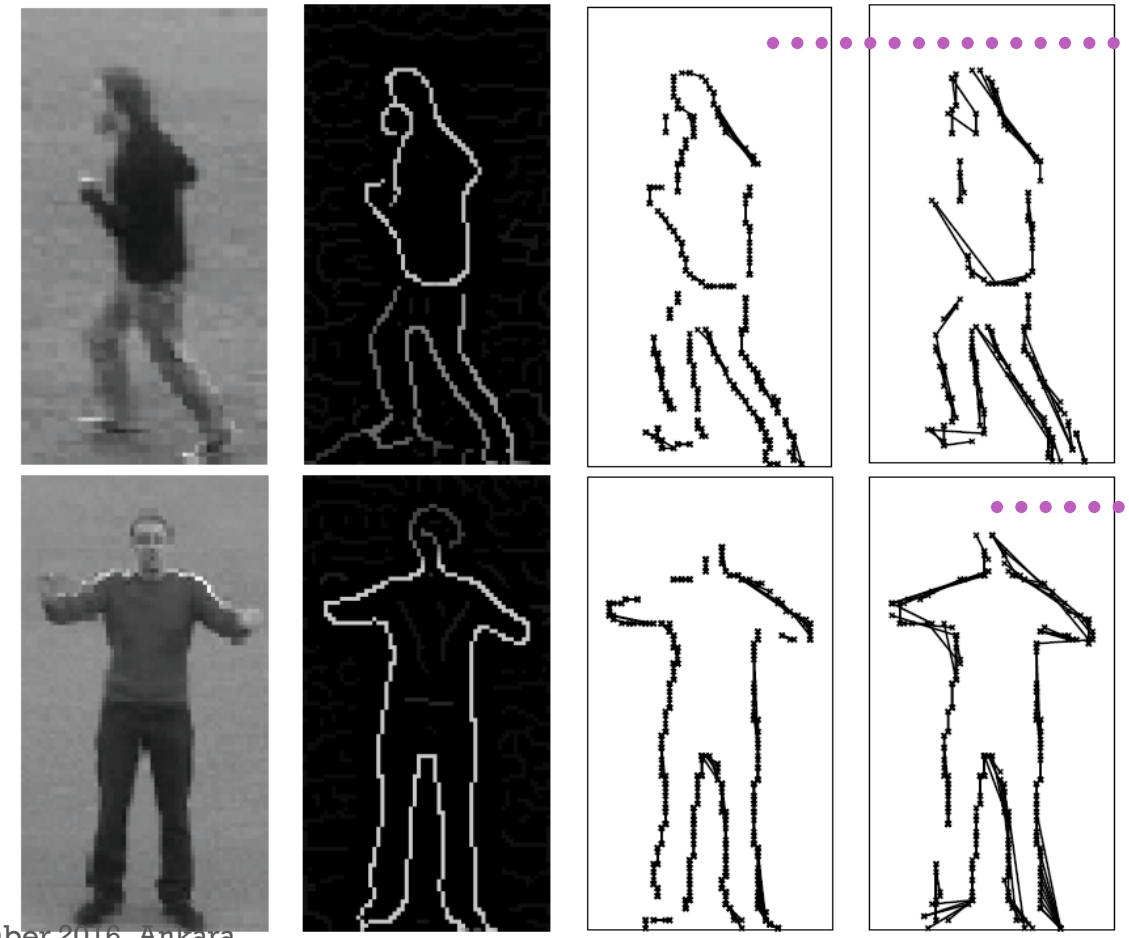
kicking

ActionWeb dataset - 467 images collected from the web

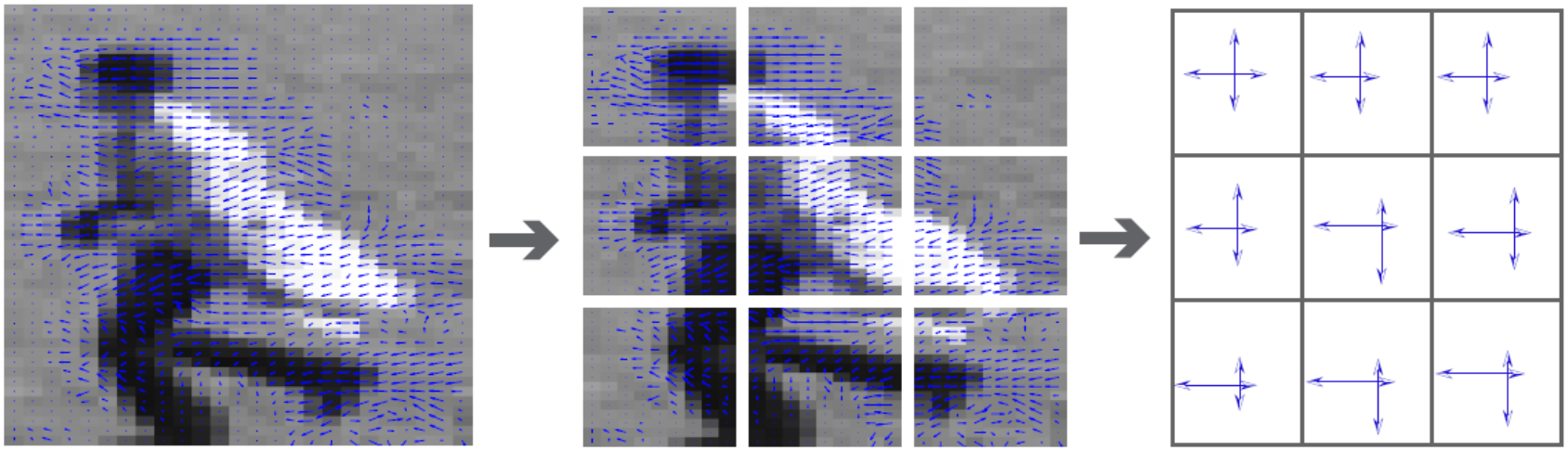
Correctly classified action images

Boundary-fitted Lines

- In the absence of silhouettes, we can use lines fitted to the boundaries (Pb) (Martin PAMI2004) of human figures

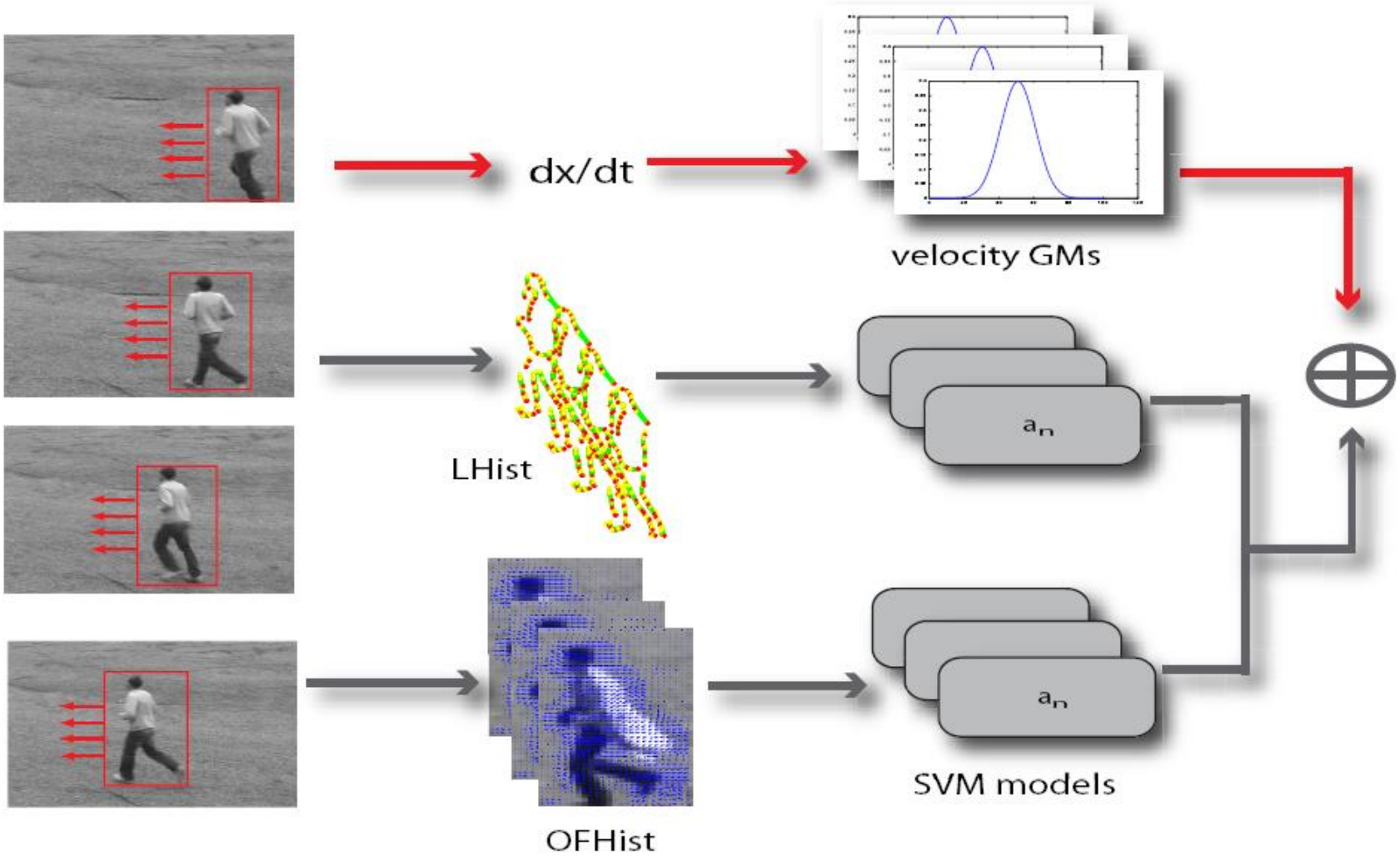


..and Optical Flow



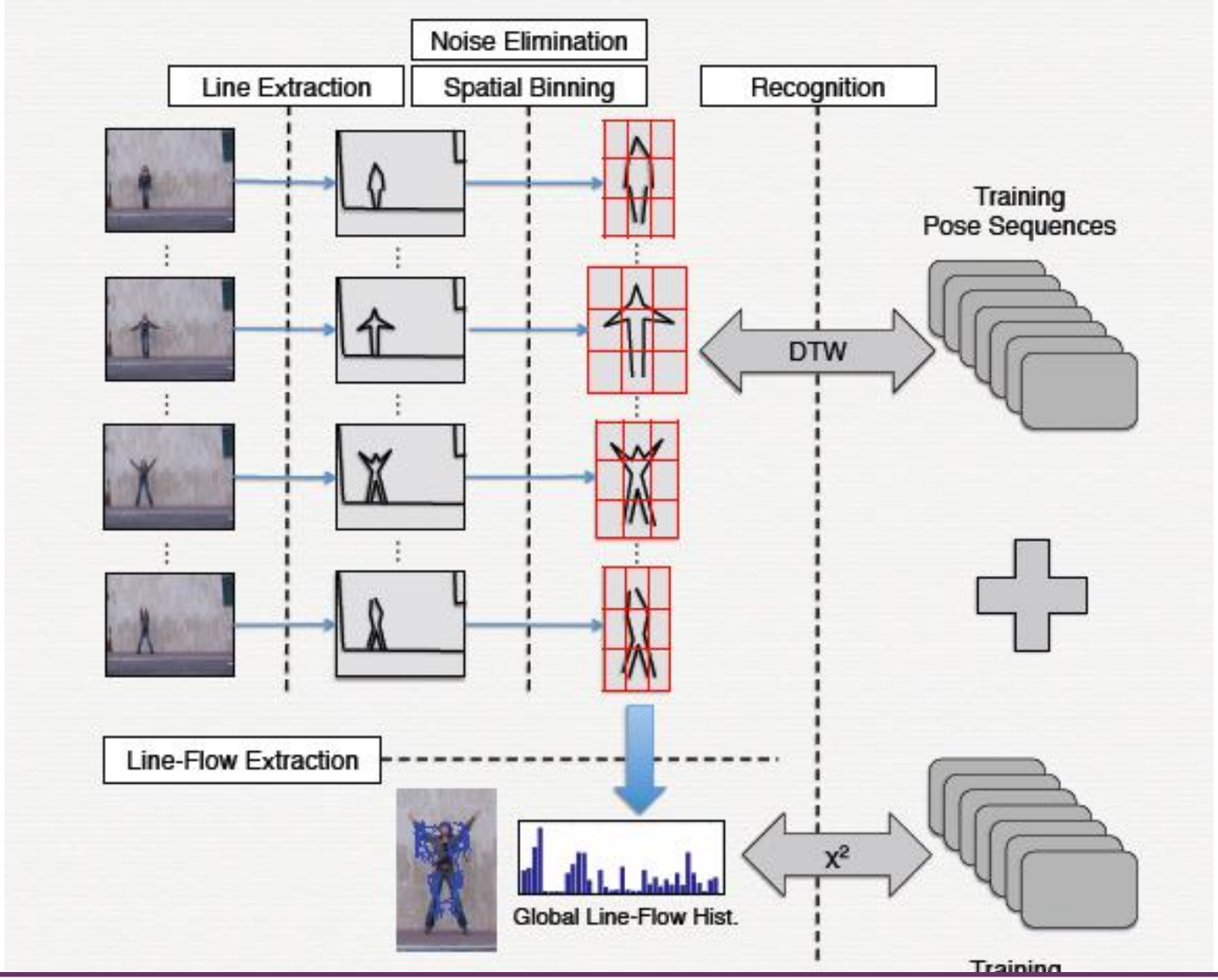
- Dense block-based optical flow calculation
  - L_1 block distance
  - 5x5 template size with a window size of 3

Recognition with LHist and OFHist



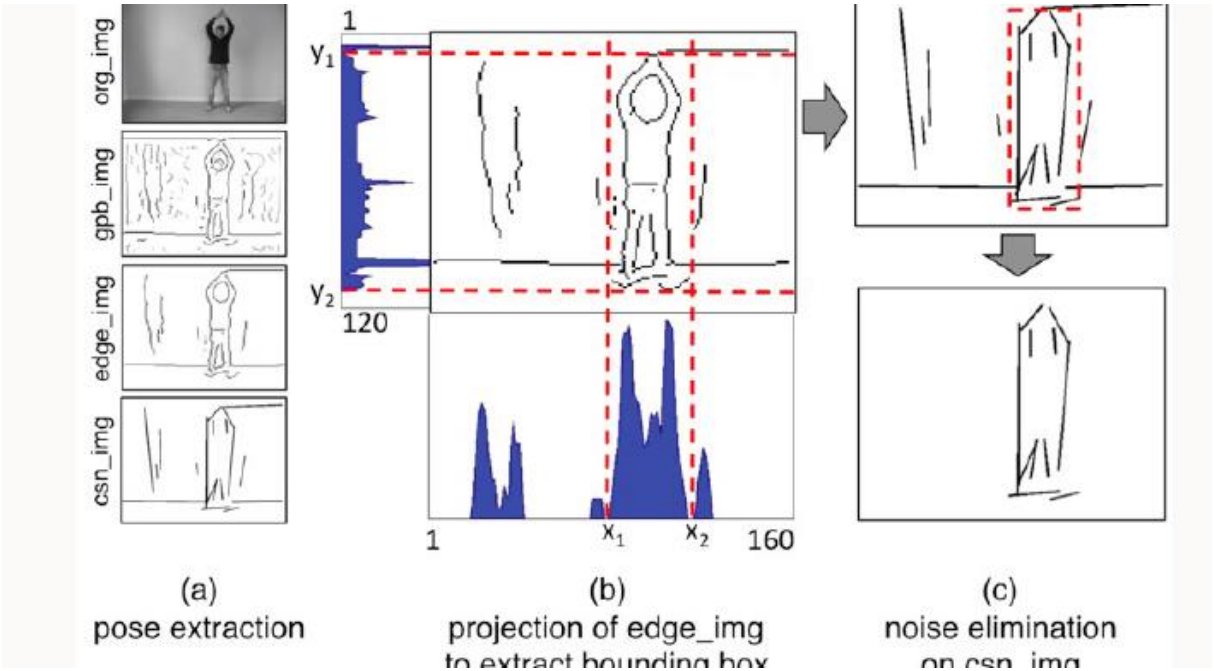
Ikizler, N., Cinbis, R. G., Duygulu, P., "Human action recognition with line and flow histograms", Proc. 19th International Conference on Pattern Recognition (ICPR 2008),

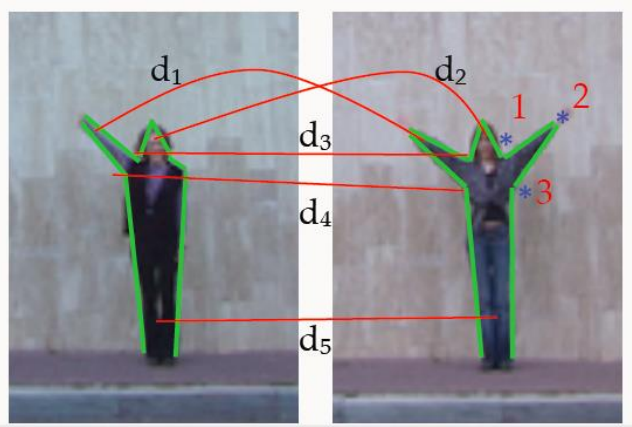
Pose as line segments

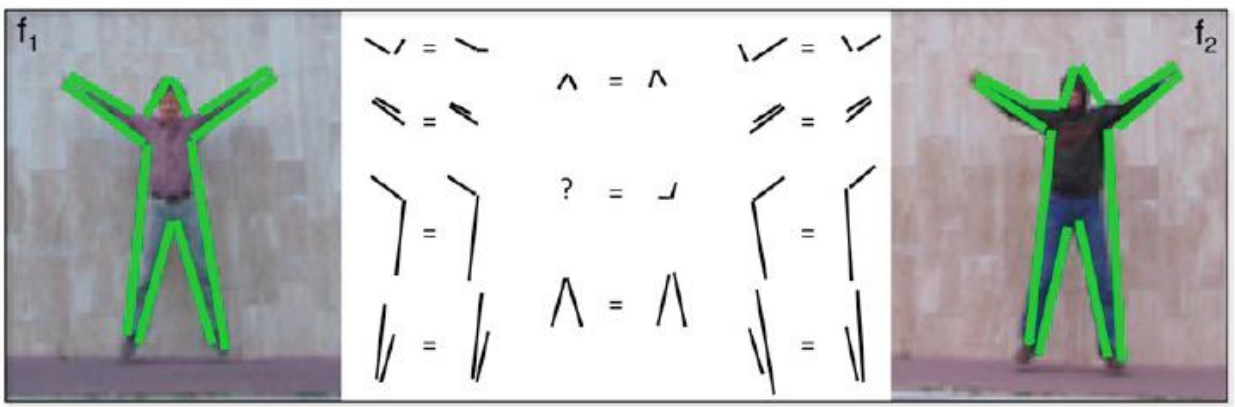


Baysal, S., Duygulu, P., "A Line Based Pose Representation For Human Action Recognition", Signal Processing: Image Communication, Volume 28, Issue 5, Pages 458-471, May 2013

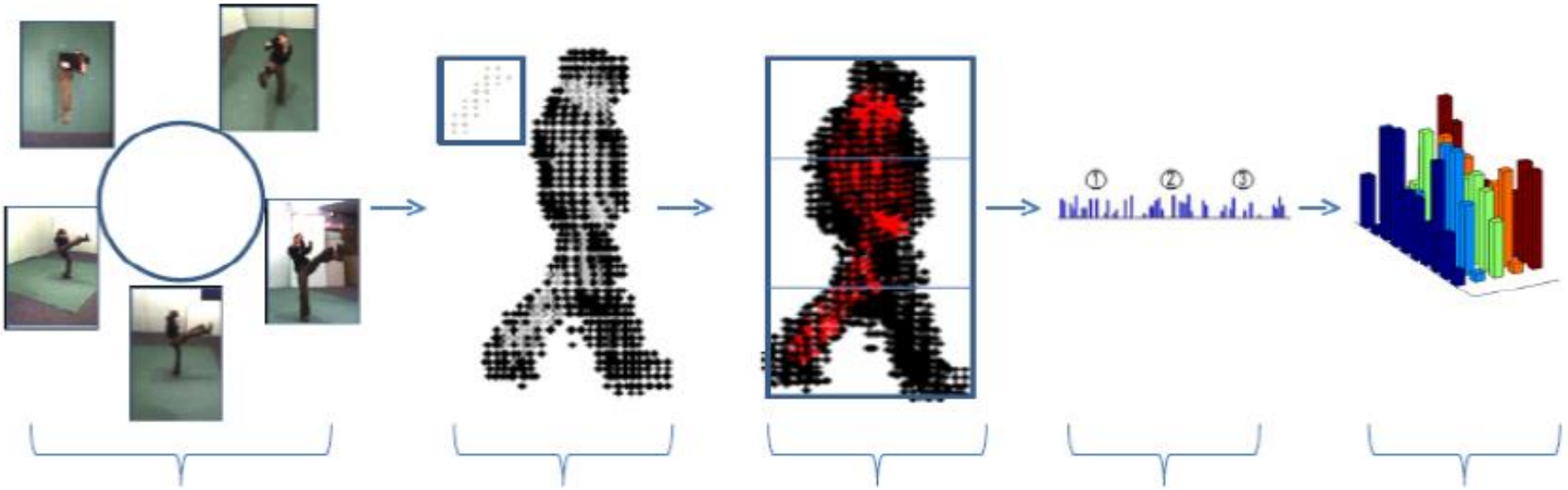
Line pairs


org_img
gpb_img
edge_img
csn_img
1
y1
y2
120
1
x1
x2
160
(a)
pose extraction
(b)
projection of edge_img
(c)
noise elimination

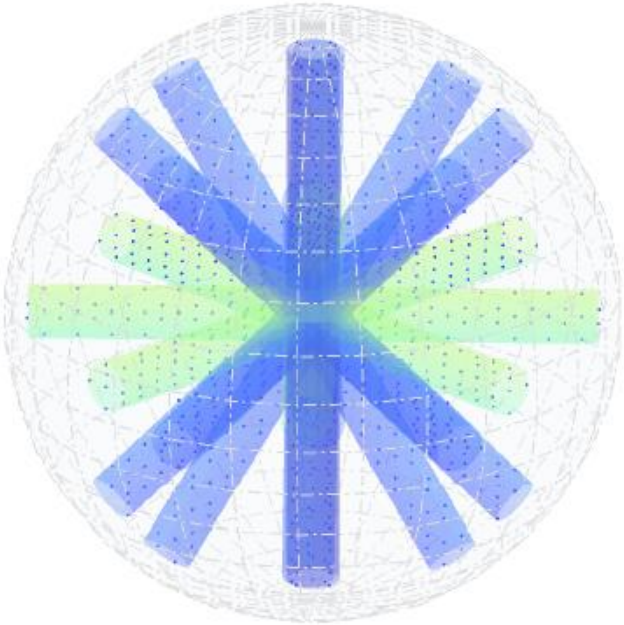

d1
d2
d3
d4
d5
1
2
3

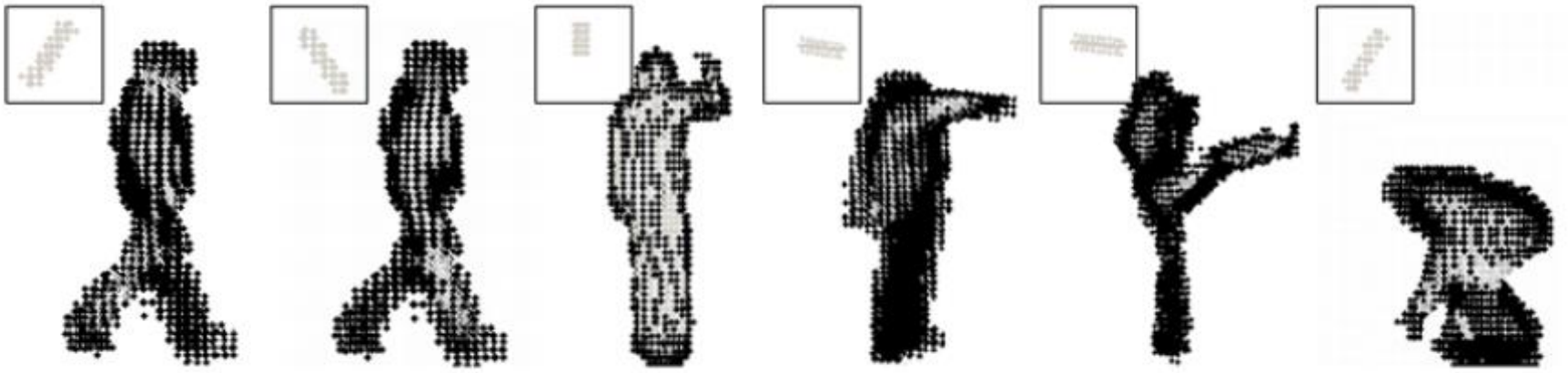

f1
f2

Multiple camera views



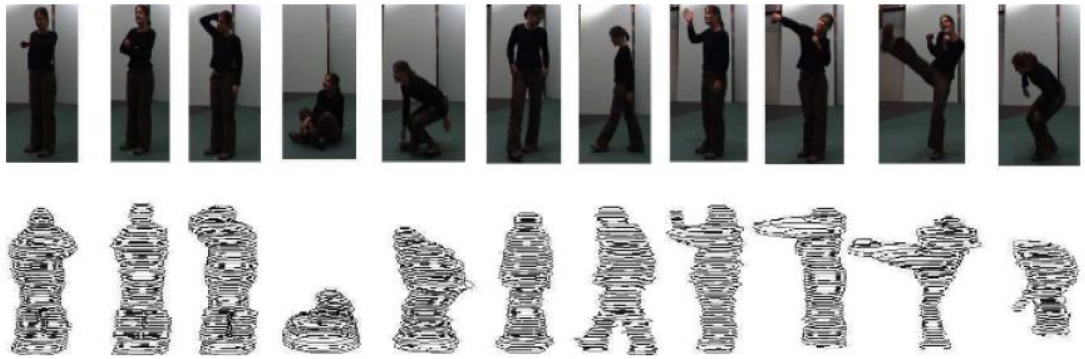
Oriented cylinders

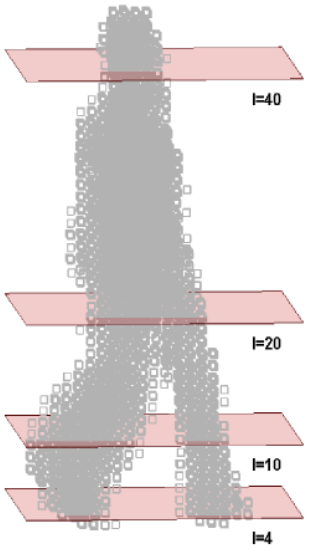


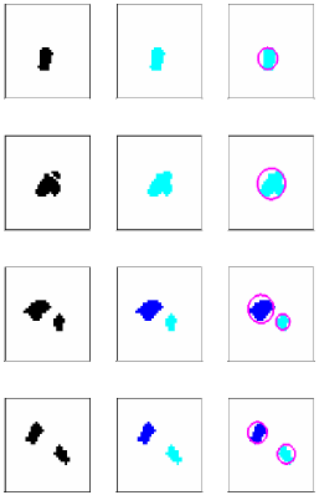


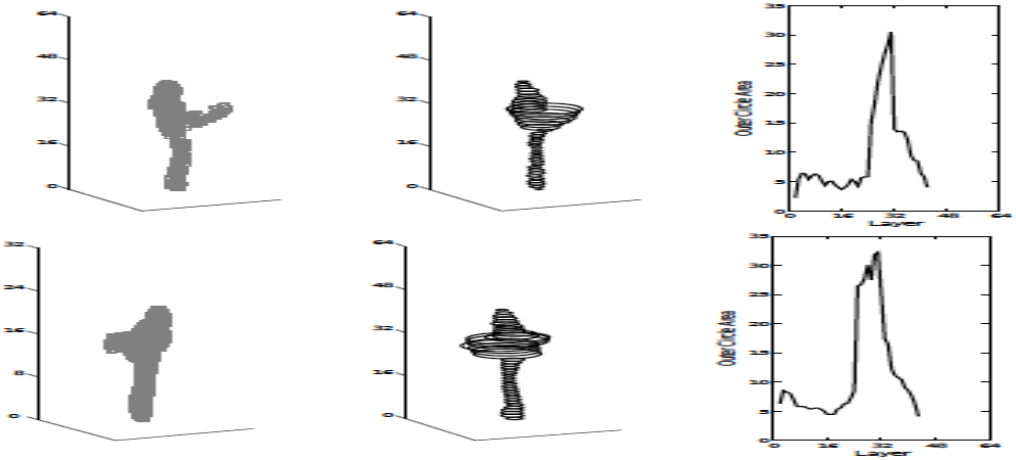
Pehlivan, S., . Duygulu, P. ”3D Human Pose Search using Oriented Cylinders”, IEEE Workshop on Search in 3D and Video (S3DV), in conjunction with ICCV 2009

Projections as circles







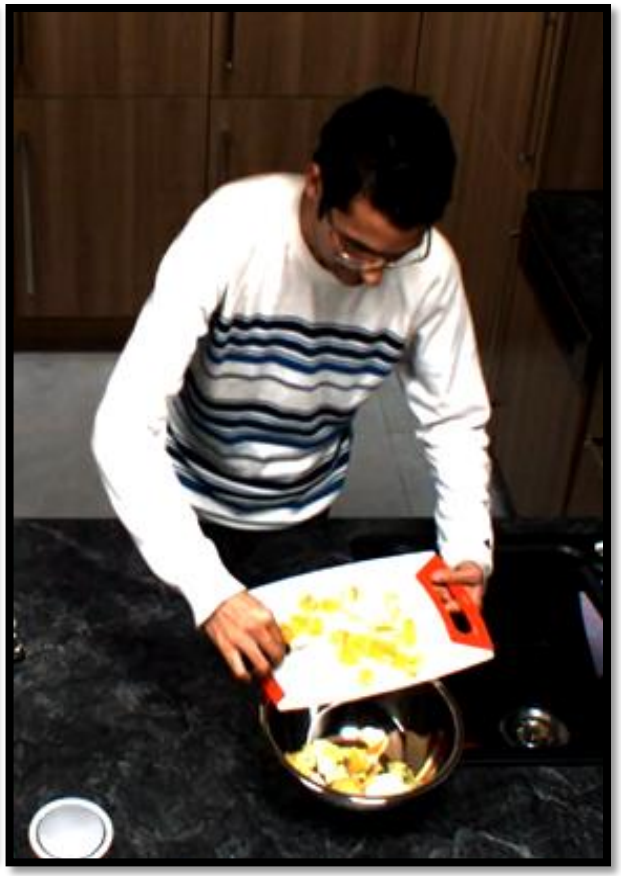


Pehlivan, S., Duygulu, P., ”A new pose-based representation for recognizing actions from multiple cameras”, Computer Vision and Image Understanding, volume 115, number 2, pages 140-151, February 2011

Assistive systems for Patient and Elderly care

Pinar Duygulu, November 2016,
Ankara

Cooking Activities: High Intra-class Variance



Iscen, A., Armagan, A., Duygulu, P., ''Knives are picked before slices are cut: Recognition through Activity Analysis'', Workshop on Cooking and Eating Activities, in conjunction with ACM Multimedia 2013.

Low Inter-class Variance

Cut apart, cut ends. cut slices, cut stripes, cut dice

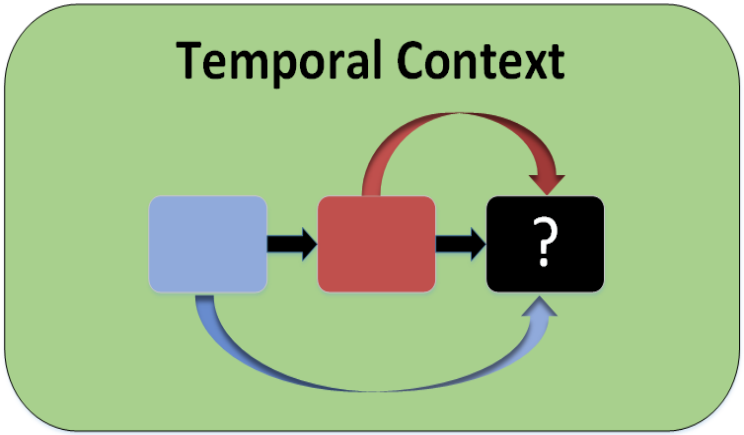


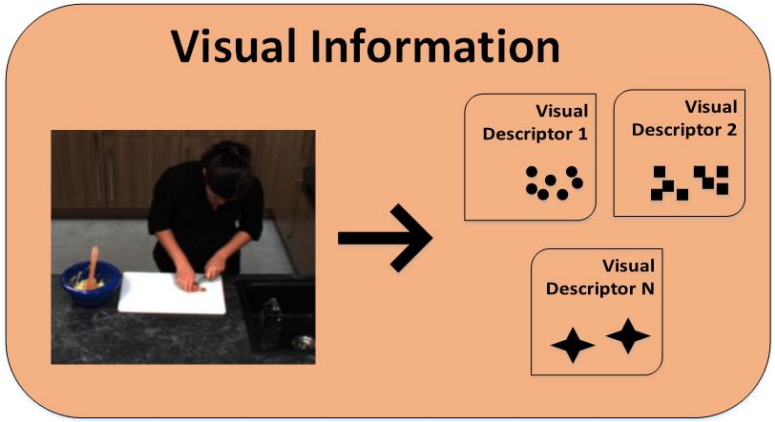



Solution

$$y = \underset{i}{argmax}\, P(c_i|x)$$

$$P(c_i|x) = T(c_i) \,.\, A(c_i, x)$$



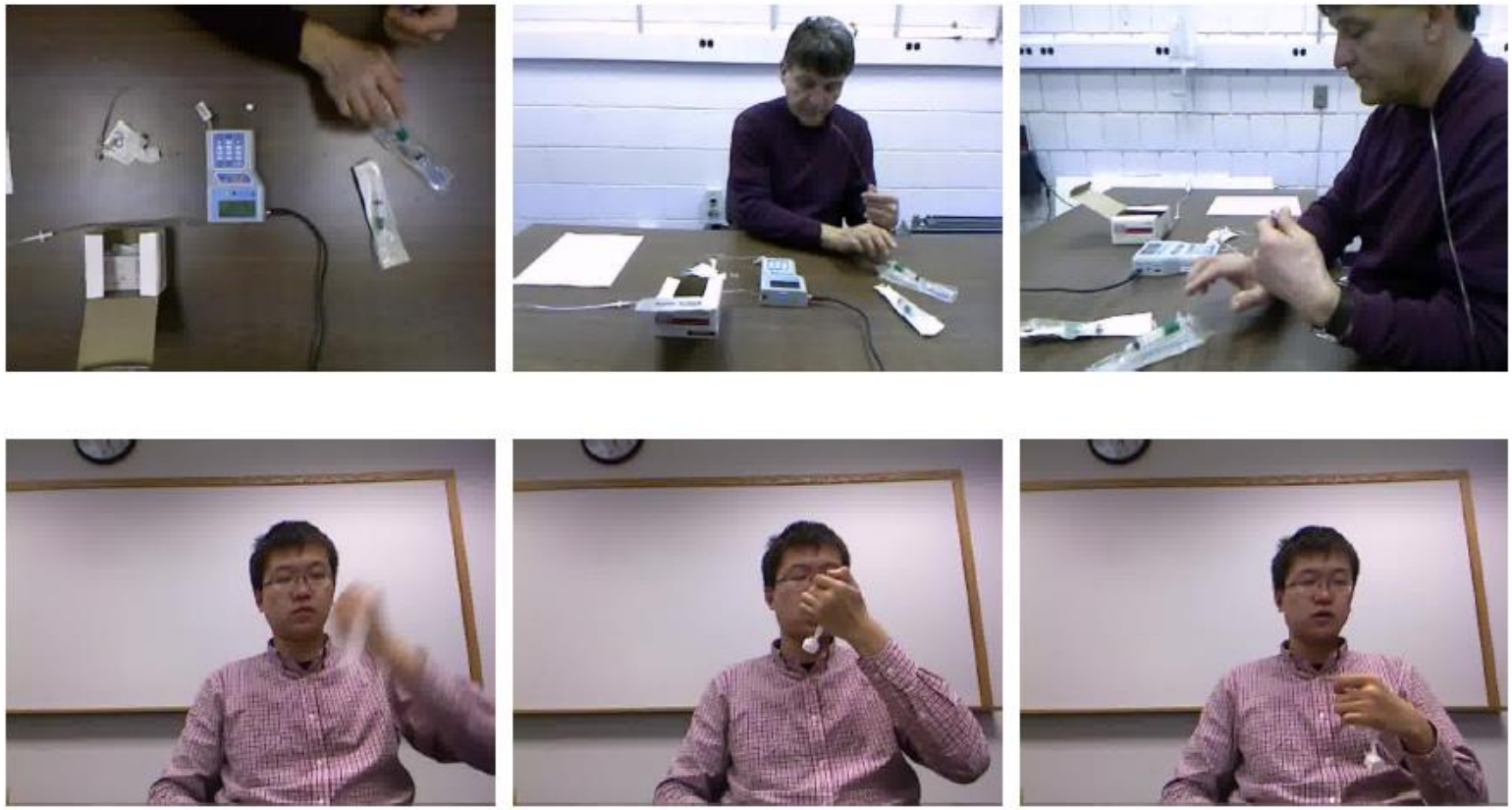


Put in Pan or Put in Bowl?

P(“put in bowl” | “cut dice”) >
P(“put in pan” | “cut dice”)

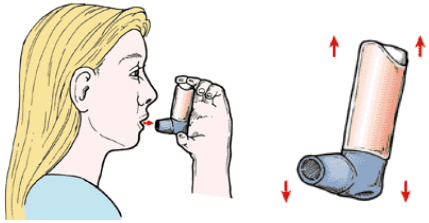
P(“put in pan” | “spread”) >
P(“put in bowl” | “spread”)

Medical Device Use

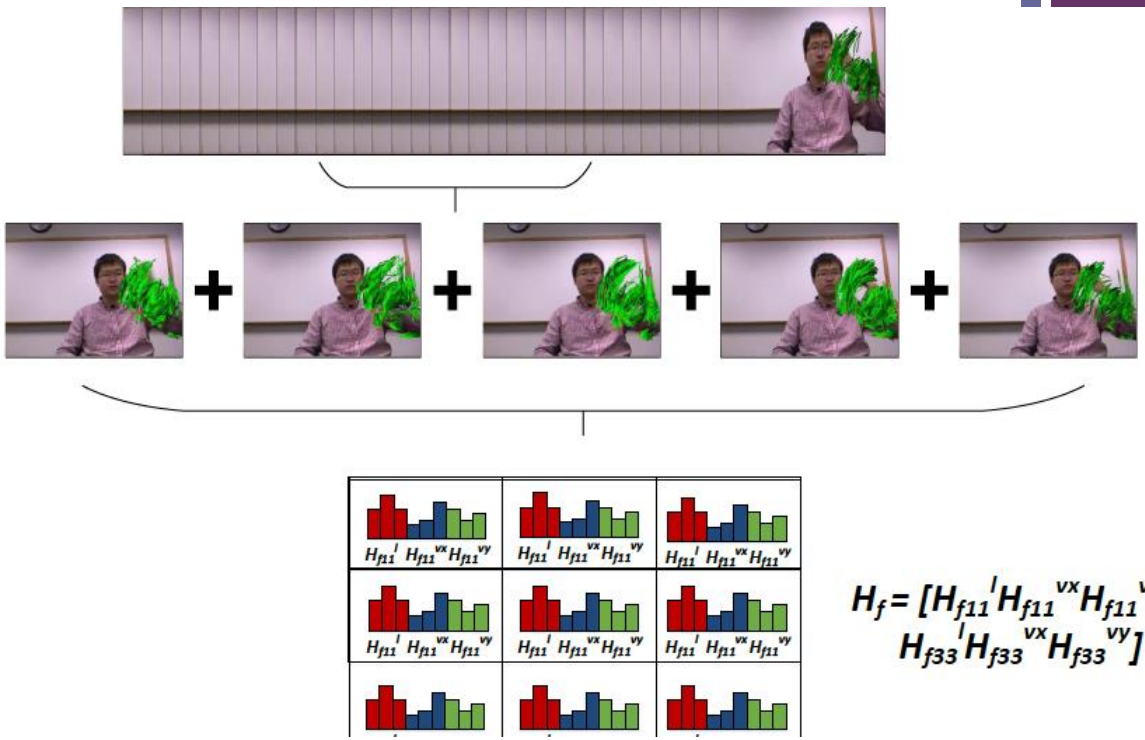


Iscen, A., Duygulu, P., ”Snippet Histograms for Assistive Technologies”, Workshop on Assistive computer Vision and Robotics, in conjunction with ECCV 2014.

Asthma Inhaler



- Shake the inhaler (for 5 second)
- Breathe out
- Put the inhaler about 2 inches in front of your mouth
- Breathe in and push down the button at the same time
- Hold your breath for 10 seconds
- Breathe out slowly



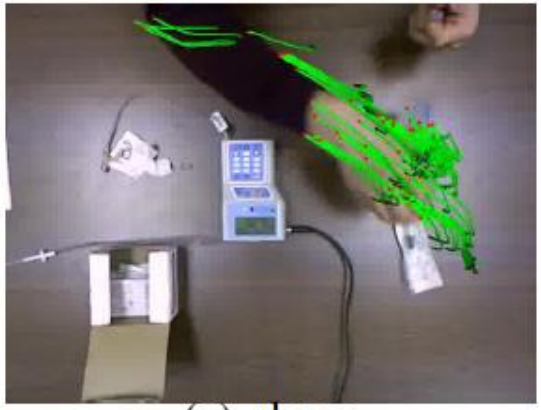
$$H_f = [H_{f11}{}^{l} H_{f11}{}^{vx} H_{f11}{}^{vy} \ldots H_{f33}{}^{l} H_{f33}{}^{vx} H_{f33}{}^{vy}]$$

	Trajectory	HOG	HOF	MBH	Snippet Hist
Recall	95.31	50.00	**100.00**	87.50	98.44
Precision	91.04	22.70	91.43	71.79	**100.00**
F-score	93.13	31.22	95.52	78.87	**99.21**

Infusion Pump

(a) front

(b) side



(c) above

Actions	Trajectory	HOG	HOF	MBH	Snippet Hist	ROI-BoW
Turn the pump on/off	91.52	91.52	90.83	92.39	**97.23**	89.40
Press buttons	79.93	80.28	80.10	79.76	83.91	**88.33**
Uncap tube end/arm port	84.26	85.64	83.56	85.47	**91.35**	65.41
Cap tube end/arm port	84.26	83.91	83.91	84.26	**89.45**	44.55
Clean tube end/arm port	70.24	73.18	77.51	74.05	75.78	**92.02**
Flush using syringe	88.75	88.24	88.06	87.20	92.56	**94.80**
Connect/disconnect	90.14	90.31	88.24	90.14	**92.73**	53.35
Average	84.16	84.73	84.60	84.75	**89.00**	75.41

Contributors

- Ahmet Iscen
- Eren Golge
- Anil Armagan
- Sermetcan Baysal
- Fadime Sener
- Hilal Zitouni
- Sare Gul Sevil
- Selen Pehlivan
- Gokberk Cinbis
- Derya Ozkan
- Nazli Ikizler


orsm.net