Büyük ölçekli video analizi

Pinar Duygulu

Çok fazla sayıda video

flickr

YAHOO!
SCREEN

- Sadece YouTube'da
  - Her ay 1 milyardan fazla erişim
  - 6 milyar saatten fazla izlenme
  - Her dakika yüklenen yüzlerce saat video

https://youtu.be/IPq_LanED60

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

Uygulamalar

- 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

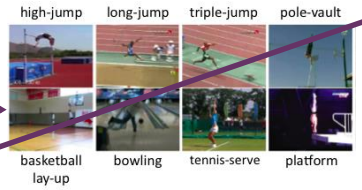
- Graphics

Motion capture and animation

Alıntı: I.Laptev

Araştırma için kullanılan veri kümeleri

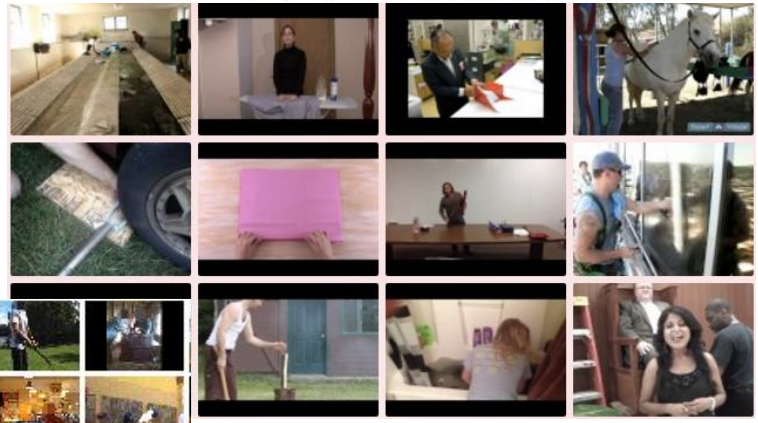
Veri kümesi	Sınıf sayısı
KTH	6
Weizmann	9
IXMAS	11
Hollywood	8
UCF Sports	9
Hollywood2	12
UCF YouTube	11
MSR	3
Olympic	16
UCF50	50
HMDB51	51

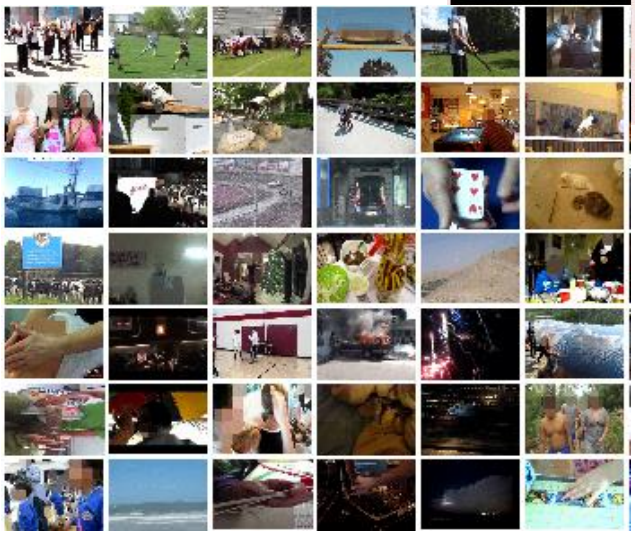


http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/

Daha güncel veri kümeleri

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





Yahoo! Flickr
100M veri kümesi
418.507 etiketli video

Sınırsız ve kısıtlamasız doğal(!) videolar

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

Hareketleri nasıl tanırız?

Zamansal değişim

Poz

Kullanılan nesneler

İçinde bulunulan ortam

Öğrenme

Etiketlenmiş veri

Etiketler

Öznitelik çıkarımı

Öğrenme

Model

Deneme

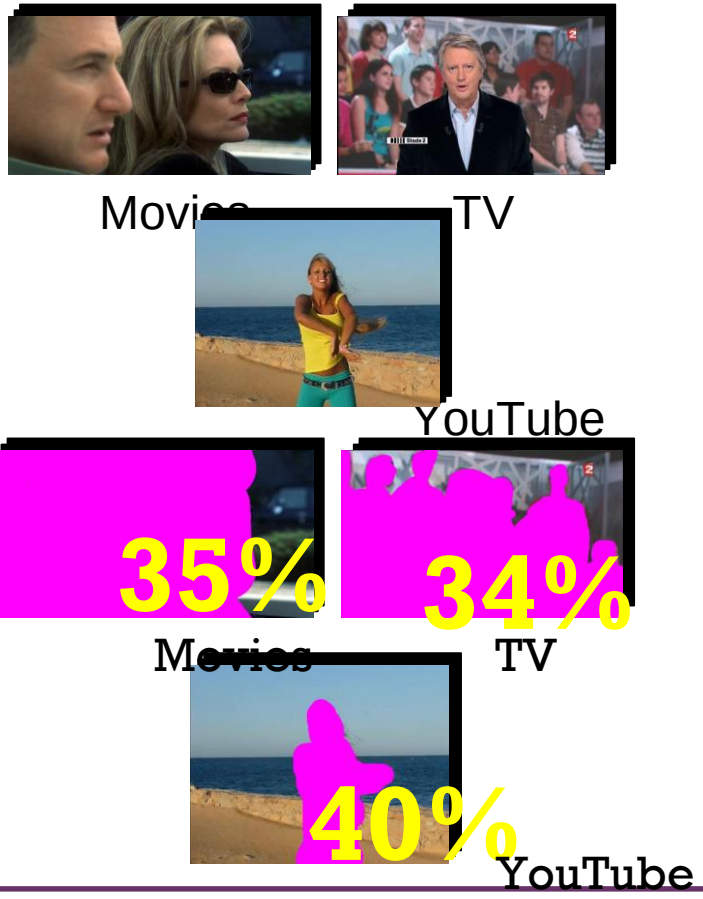
Yeni veri

Öznitelik çıkarımı

Model

Tahmin

Tek resimden poz bilgisi ile hareket tanıma



Silhouette description [Sullivan & Carlsson, 2002]

Histogram of gradients (HOG) [Dalal & Triggs 2005]

Human body part layout [Felzenszwalb & Huttenlocher, 2000]

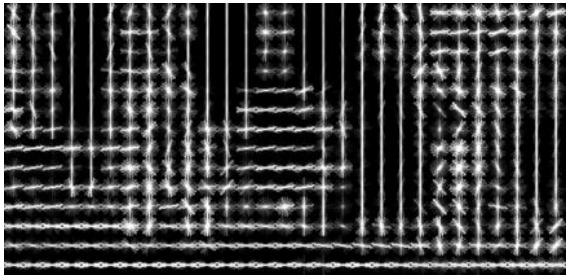
Alıntı: C. Schmid

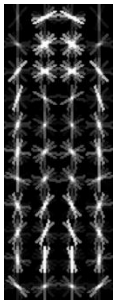
Histogram of Oriented Gradients Deformable Part based models

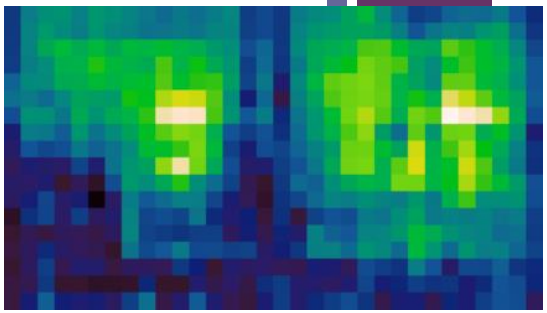
HOG feature map

Template

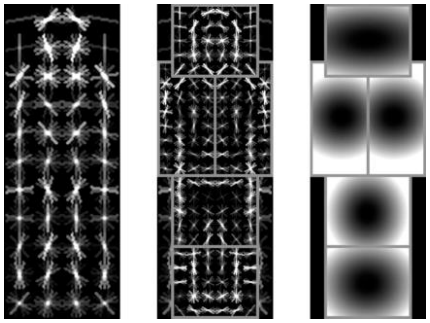
Detector response map

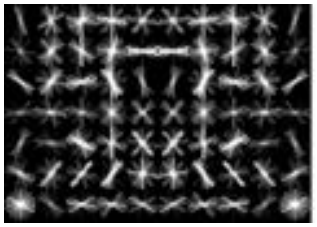


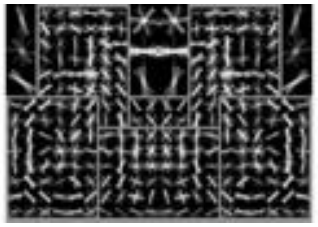


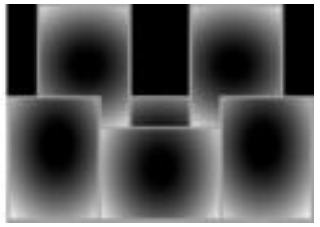


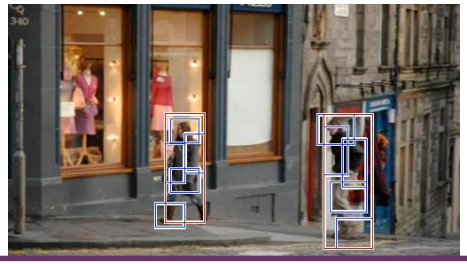
N. Dalal and B. Triggs, Histograms of Oriented Gradients for Human Detection, CVPR 2005

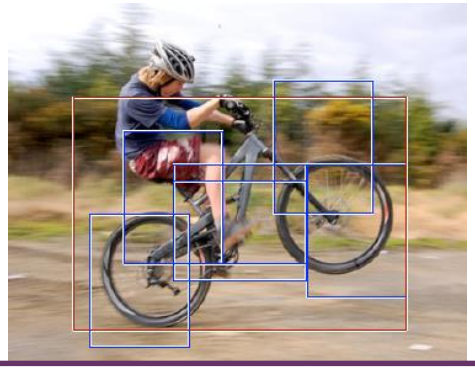








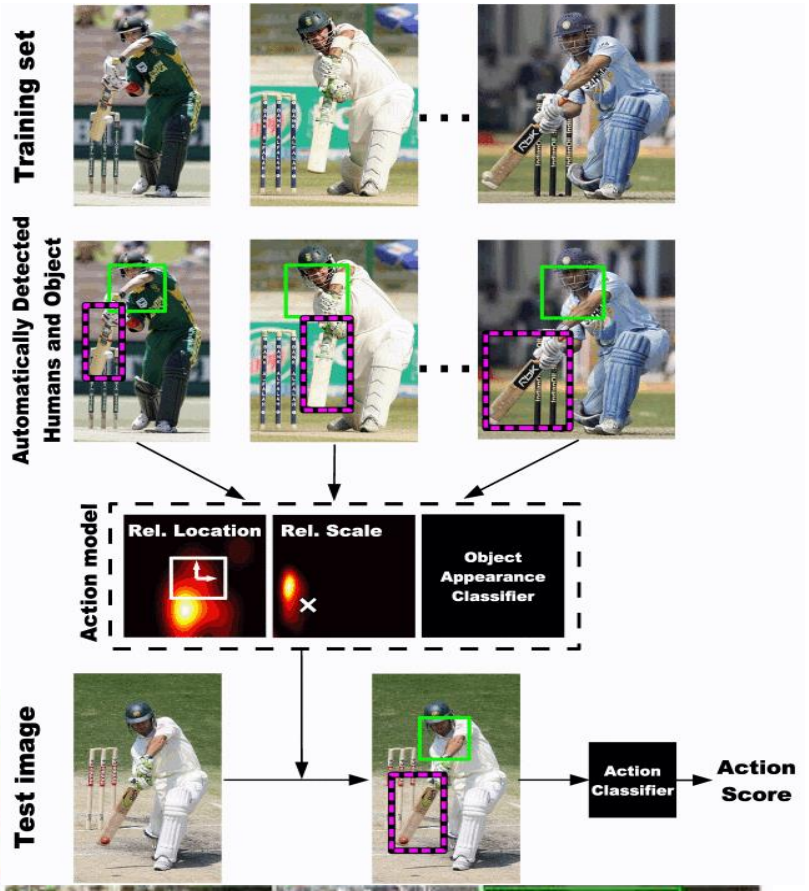


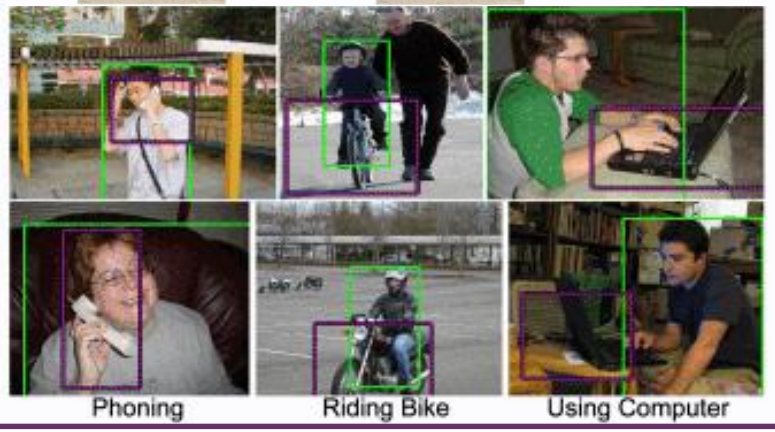


P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part Based Models, PAMI 32(9), 2010

Alıntı: Svetlana Lazebnik

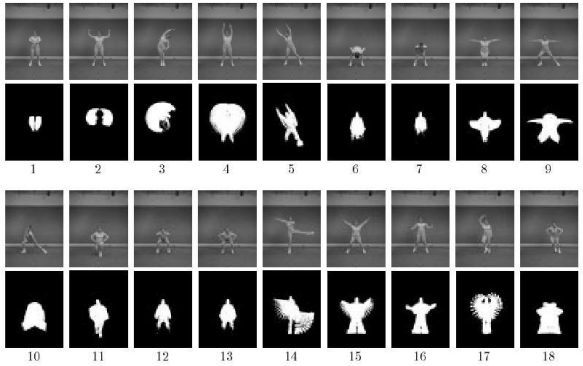
Kullanılan nesne ile hareketi tanıma



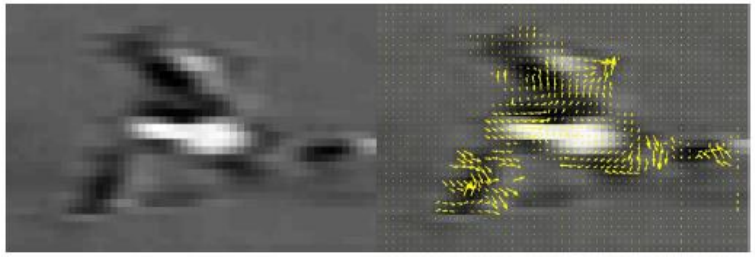


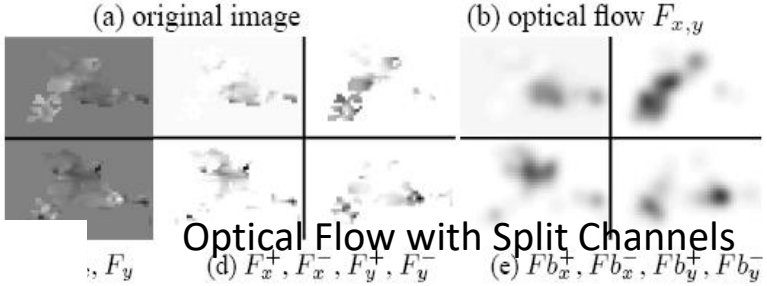
Prest, Schmid & Ferrari , Weakly-supervised learning of interactions between humans and objects, 2011

Hareket tanıma için zamansal değişim

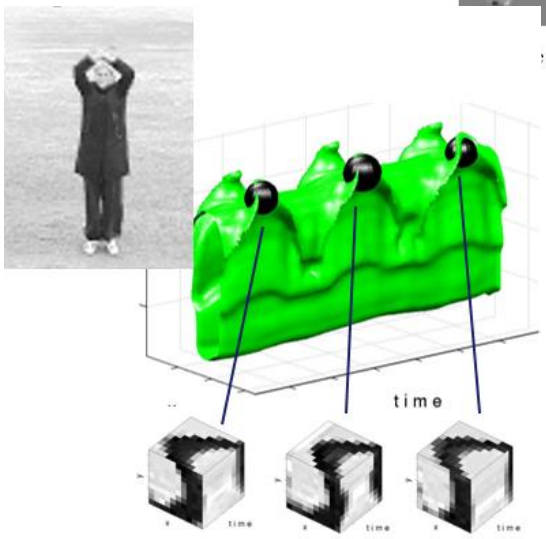


Motion Energy and Motion History





Optical Flow with Split Channels

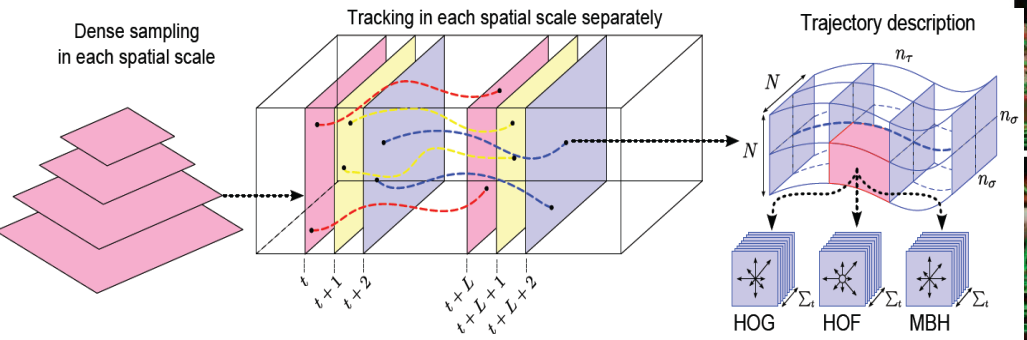


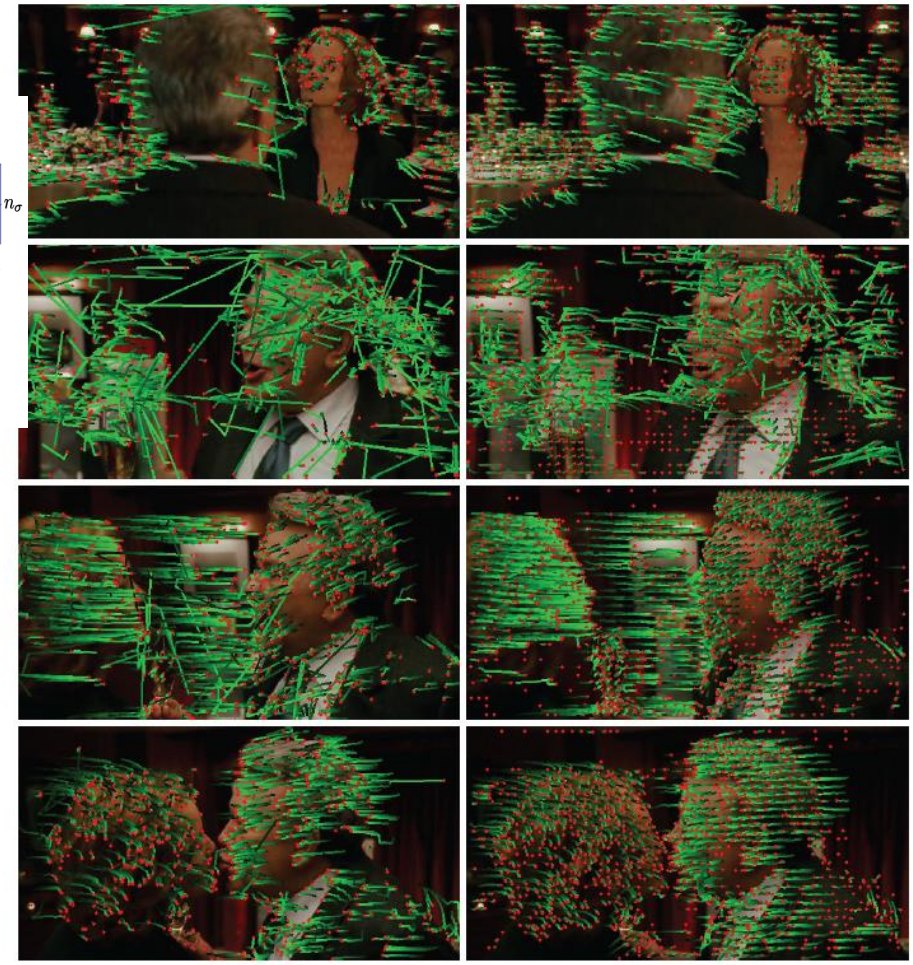
Space Time Interest Points

Space-Time Volumes

A. F. Bobick, J. W. Davis, “The Recognition of Human Movement Using Temporal Templates”, PAMI 2001
A. A. Efros, A. Berg, G. Mori, J. Malik, Recognizing Action at a Distance, ICCV 2003
I. Laptev, “On Space-Time Interest Points”, IJCV, 2005

Yoğun İz takibi

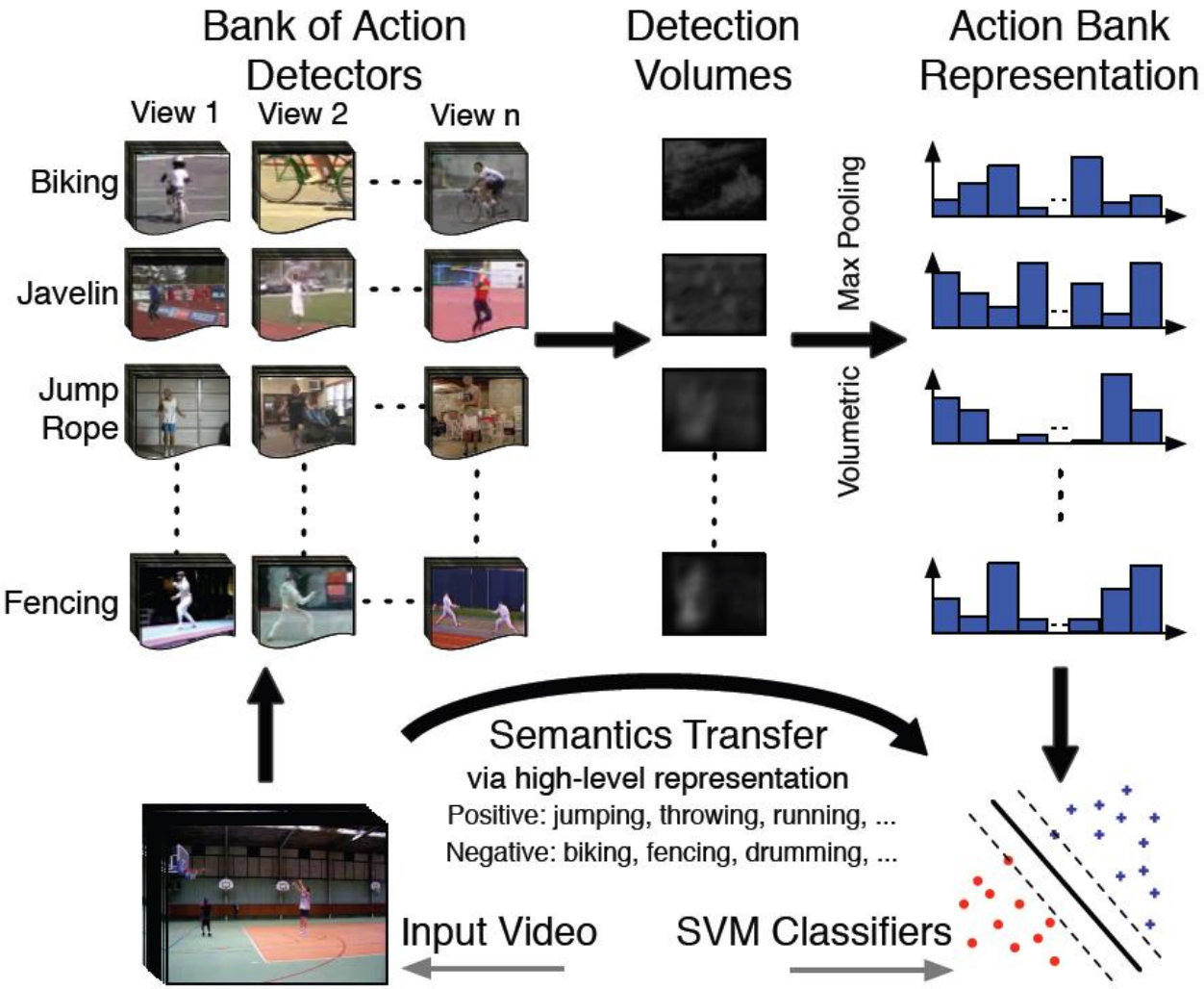




KLT Dense trajectories

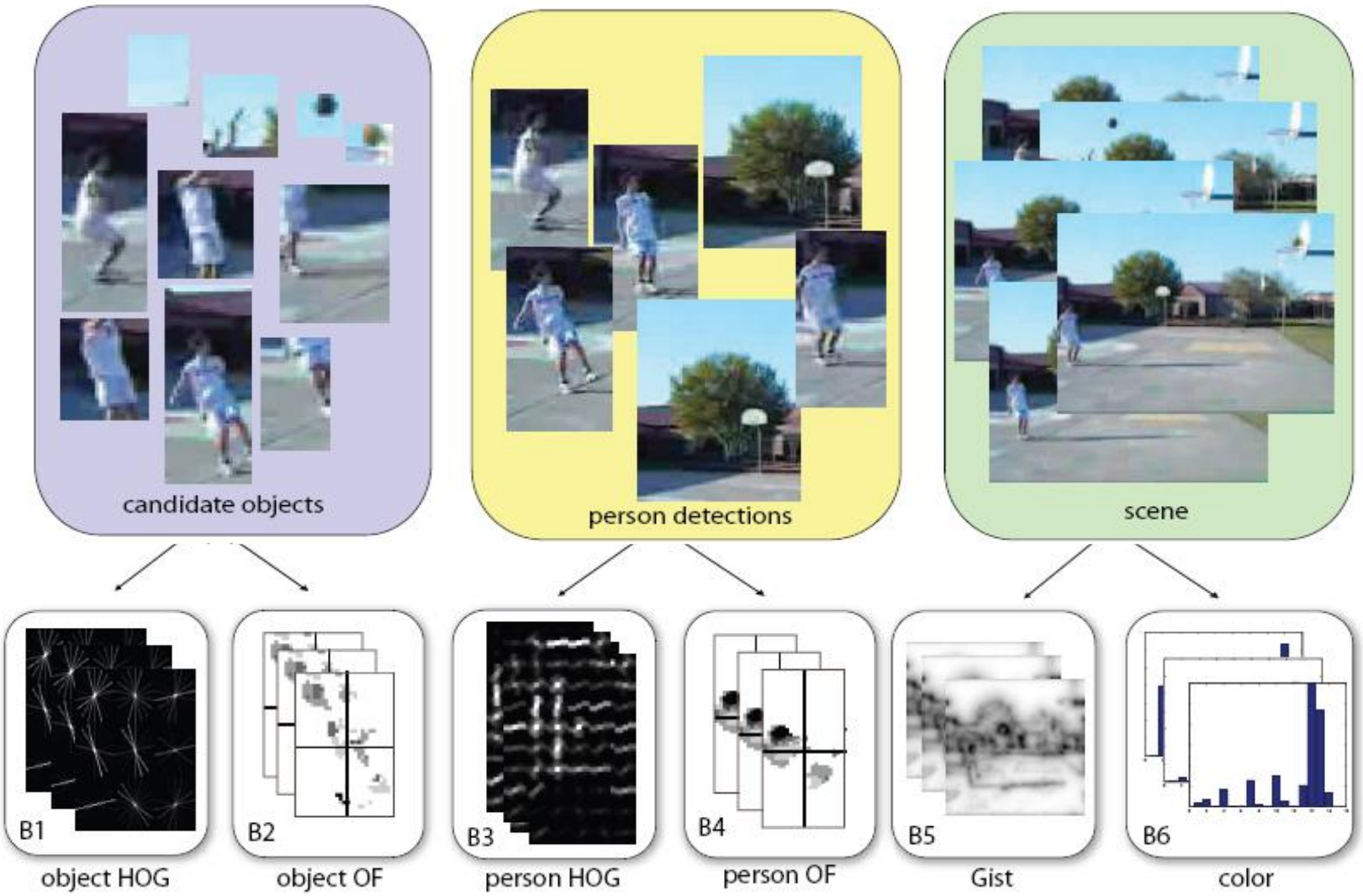
H. Wang, A. Klaser, C. Schmid, C.-L. Liu. Action Recognition by Dense Trajectories. CVPR, 2011.

Hareket Bankası (Action bank)



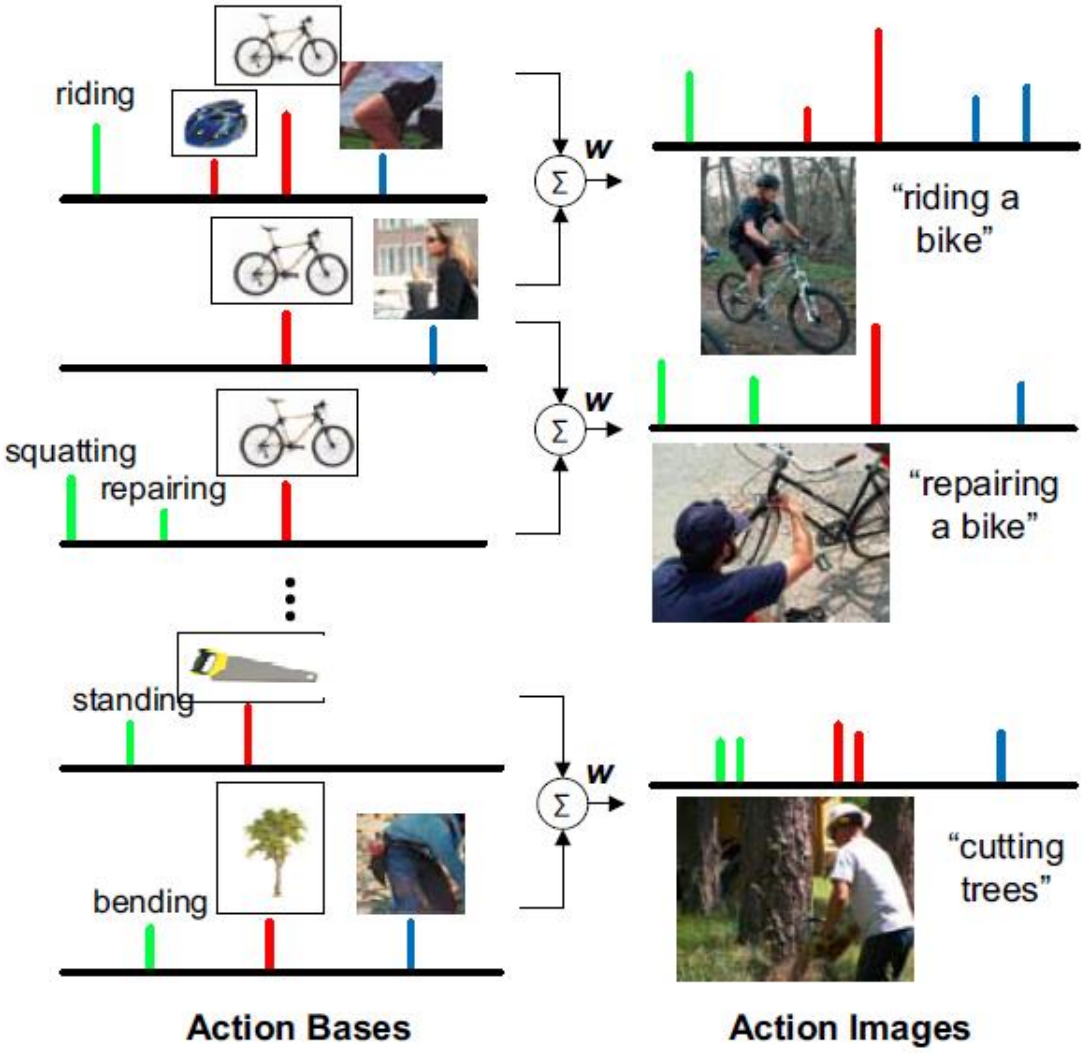
S. Sadanand, J. J. Corso. Action Bank: A High-Level Representation of Activity in Video, CVPR, 2012.

Sahne, nesne ve hareket birlikte



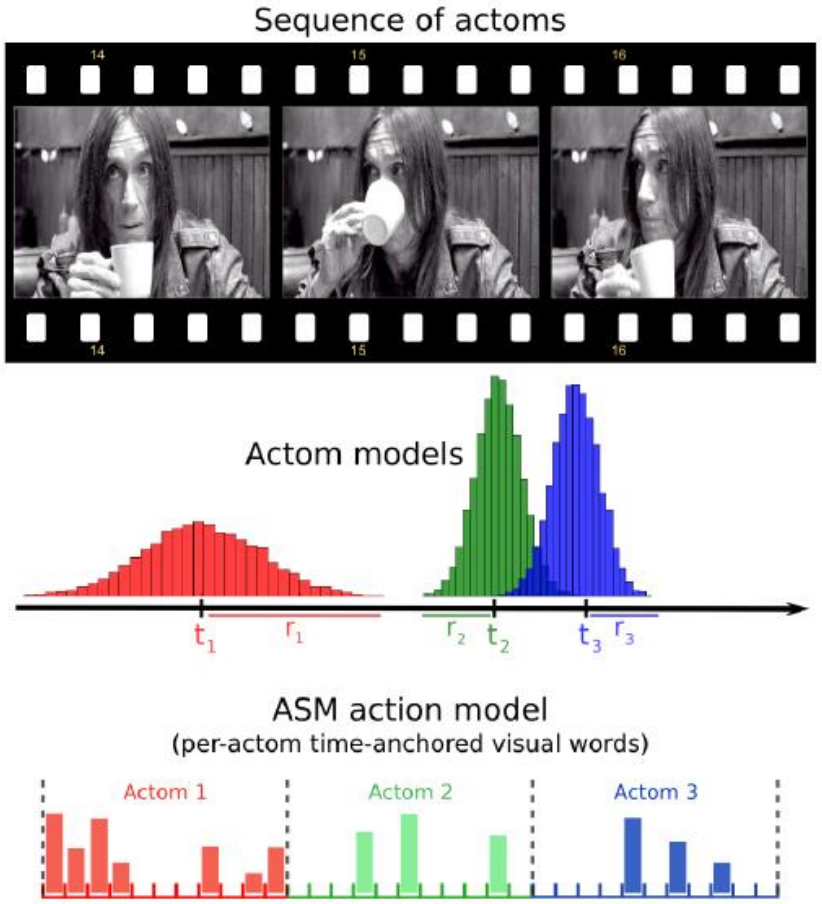
N. Ikizler-Cinbis, S. Sclaroff, Object, Scene and Actions: Combining Multiple Features for Human Action Recognition, ECCV, 2010.

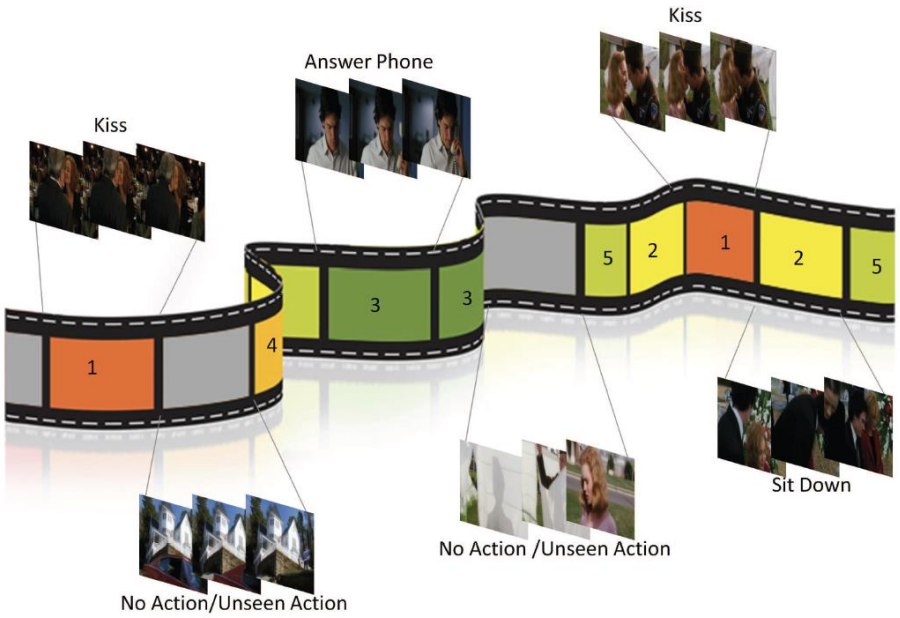
Hareket tanımlayıcı nitelikler



B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas and L. Fei-Fei, Human action recognition by learning bases of action attributes and parts, ICCV 2011

Bileşik hareketler - Hareket dizileri





A. Gaidon, Z. Harchaoui, and C. Schmid. Actom sequence models for efficient action detection. CVPR, 2011.

M. Hoai, Z. Lan, and F. Torre. Joint segmentation and classification of human actions in video. CVPR, 2011.

Hareket mi Olay mı

- Hareket: içme, oturma

Coffee & Cigarettes dataset

Hollywood dataset

- Olay: sandviç yapma

Alıntı: C. Schmid

TRECVID-Multimedia Event Detection

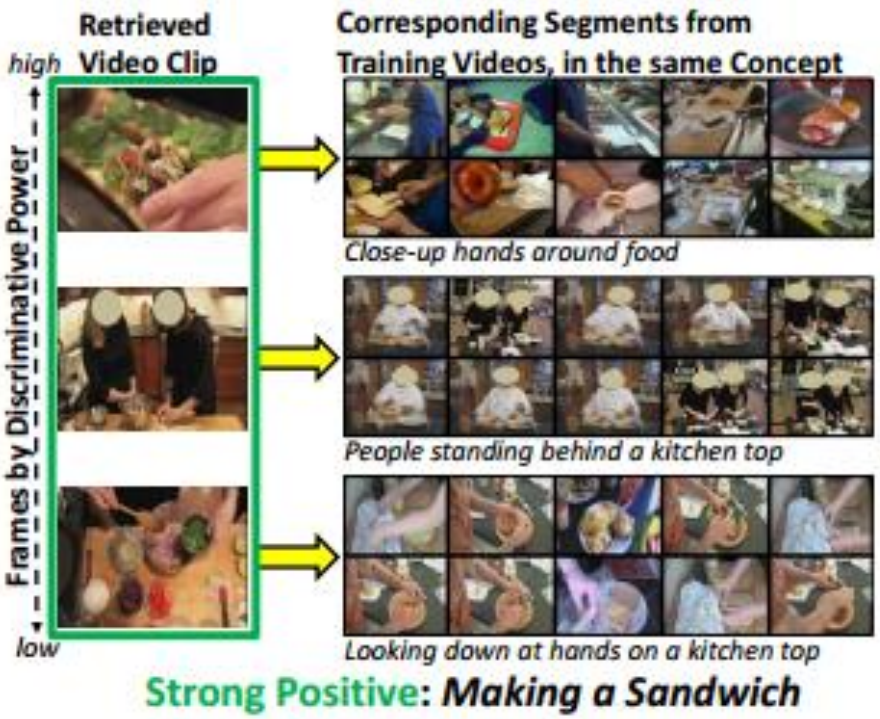
Attempting a board trick

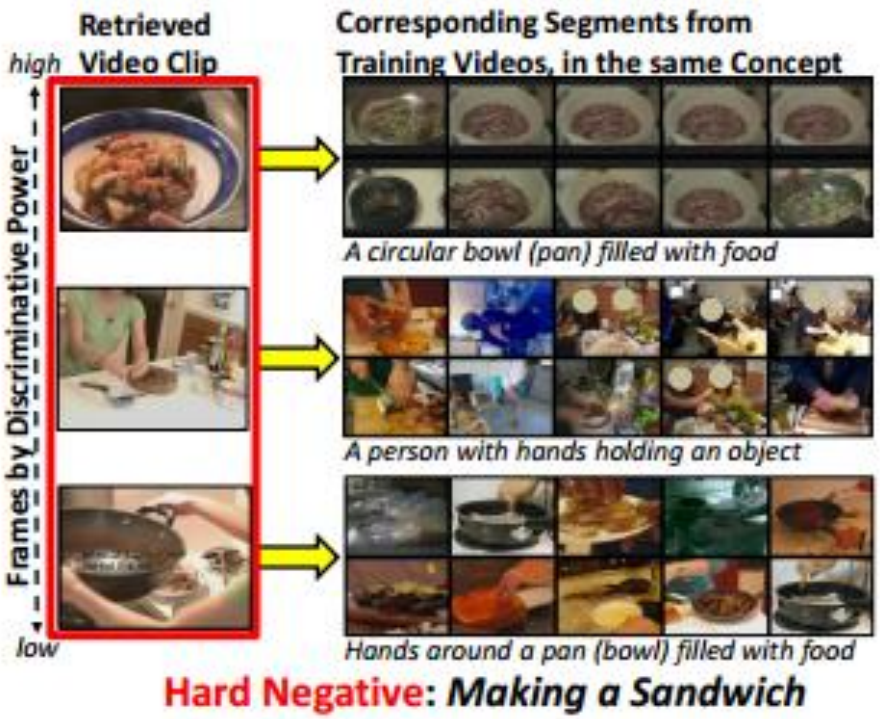
Feeding an animal

Wedding ceremony

Getting a vehicle unstuck

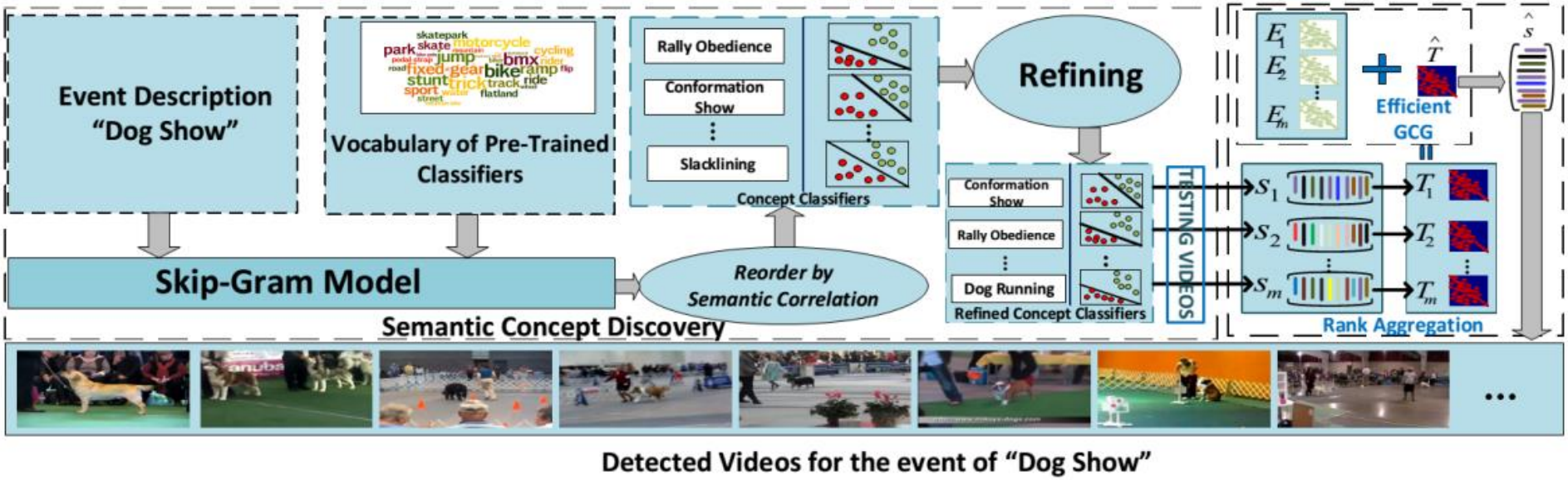
Alıntı: C. Schmid





S. Oh, S. McCloskey, I. Kim, A. Vahdat, K. Cannons, H. Hajimirsadeghi , G. Mori, A. Perera, M. Pandey, J. Corso, Multimedia Event Detection with Multimodal Feature Fusion and Temporal Concept Localization, Machine Vision and Applications, 2014

Örnek olmadan olay tanıma (zero-shot event detection)



Detected Videos for the event of "Dog Show"

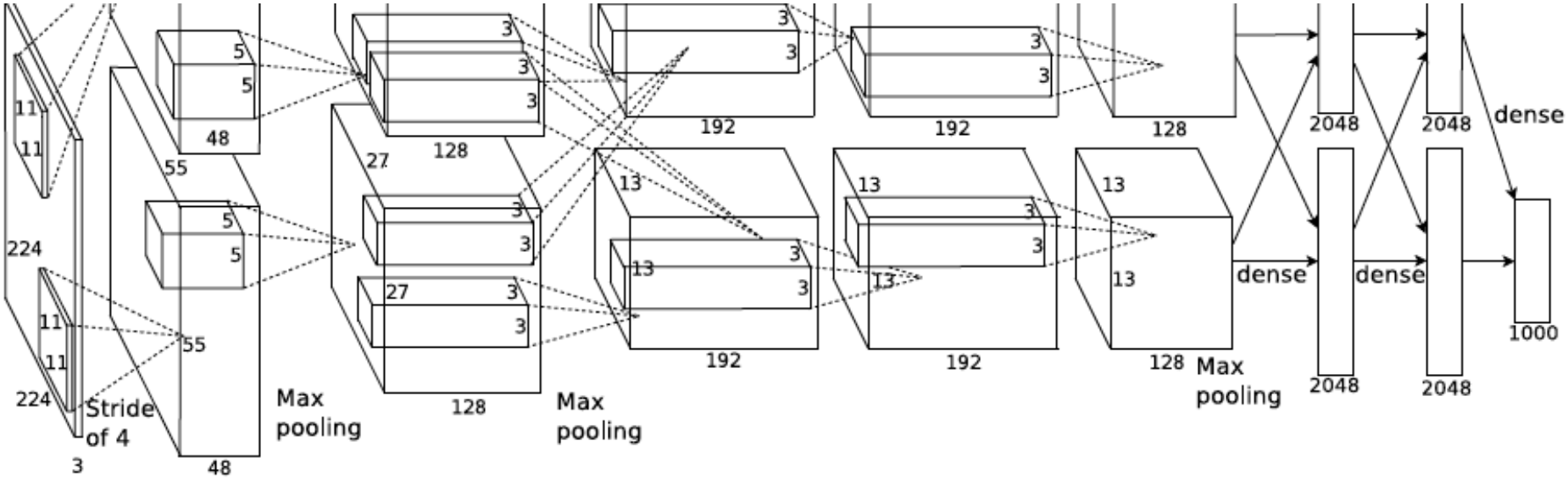
X. Chang, Y. Yang , A. Hauptmann , E. Xing, Y-L. Yu, Semantic Concept Discovery for Large-Scale Zero-Shot Event Detection , IJCAI 2015

Imge/ Video

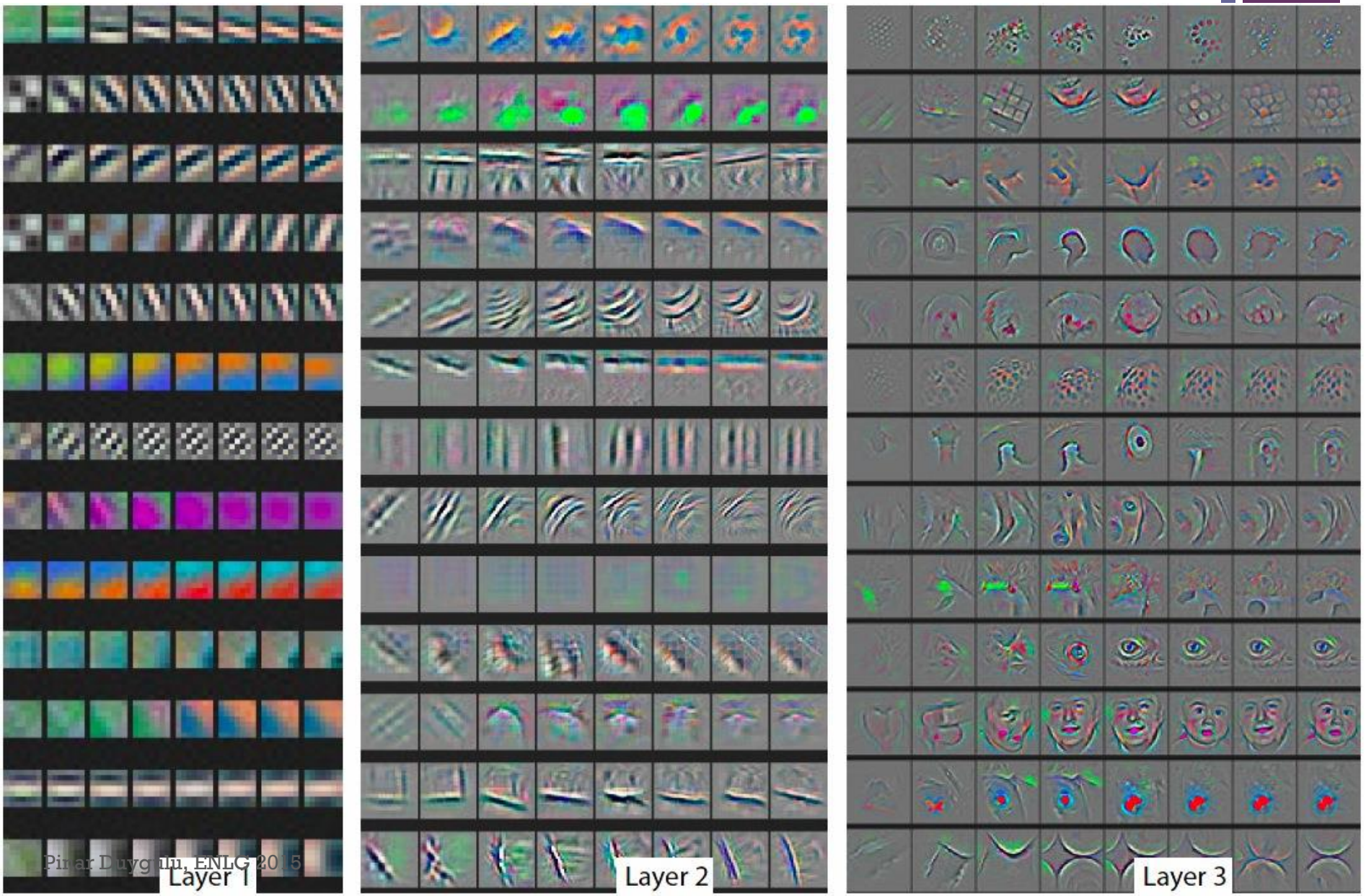
Özel olarak geliştirilmiş öznitelik

Sınıflandırıcı

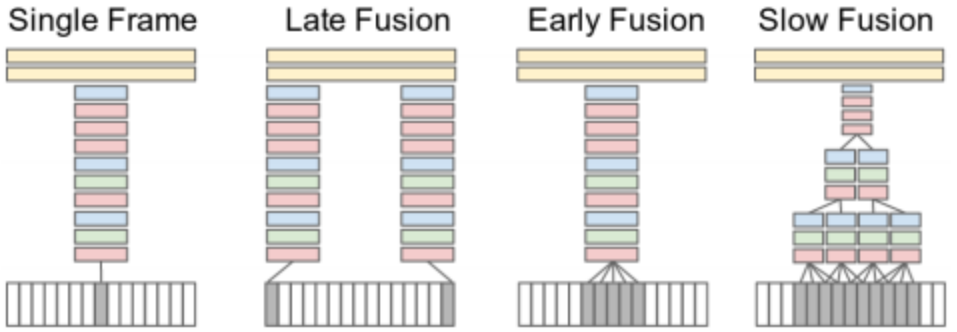
Etiket



A. Krizhevsky, I. Sutskever, and G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012


Layer 1
Layer 2
Layer 3

Video için CNN



footbag
single frame predictions:
crossfit
weight pulling
triathlon
disc dog
powerbocking
motion-aware predictions:
footbag
freestyle football
freestyle bmx
unicycle
decathlon

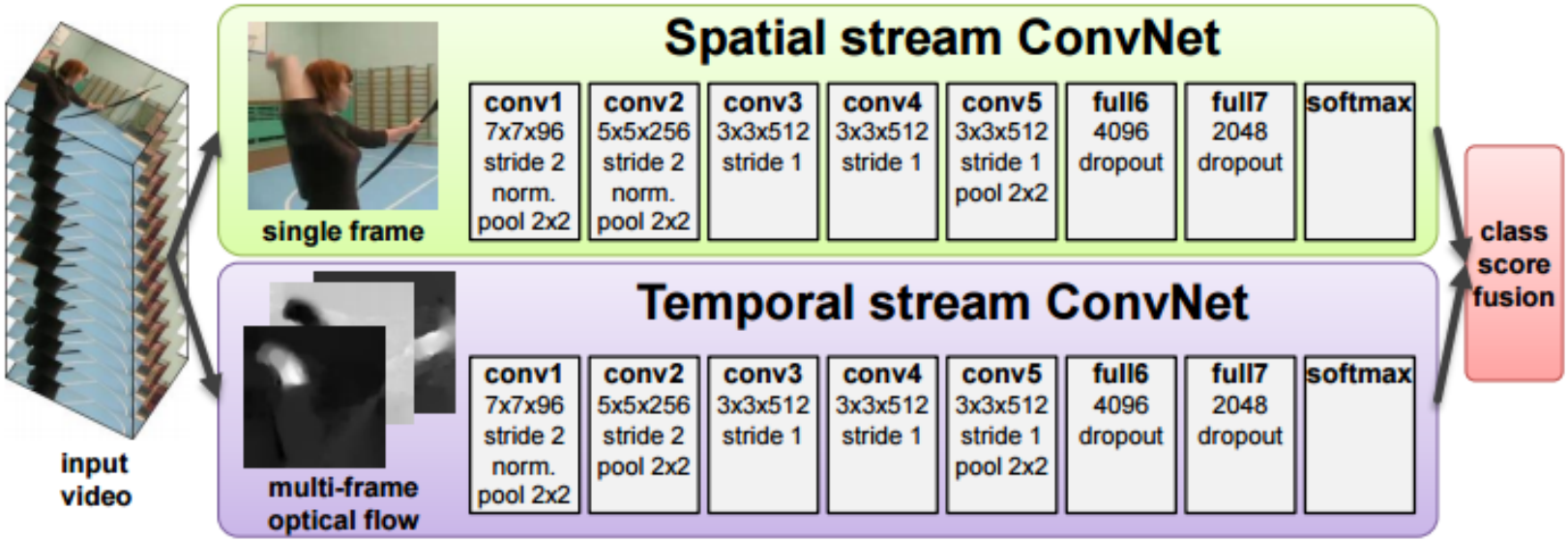
juggling club
single frame predictions:
acrobatics
wing tsun
freestyle slalom skating
trapeze
unicycle
motion-aware predictions:
juggling club
kalaripayattu
baton twirling
acrobatics
color guard (flag spinning)

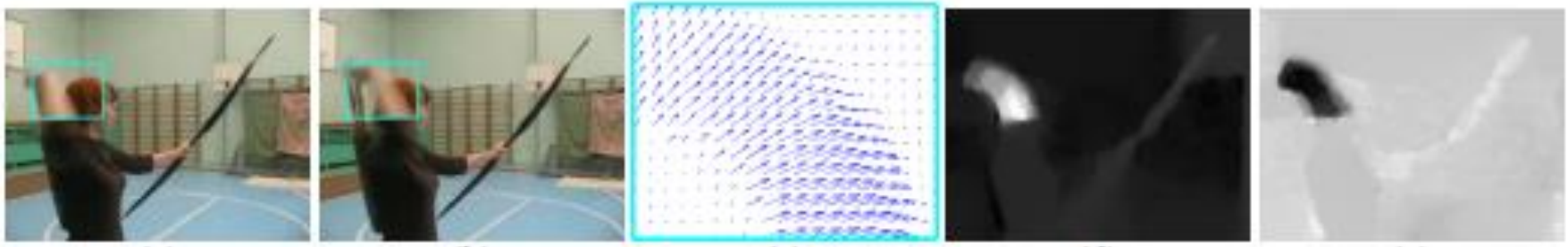
slacklining
single frame predictions:
rope climbing
beach tennis
rings (gymnastics)
inline speed skating
modern pentathlon
motion-aware predictions:
slacklining
rope climbing
beach handball
footvolley
streetball

short track motor racing
single frame predictions:
short track motor racing
touring car racing
drifting (motorsport)
motorcycle racing
time attack
motion-aware predictions:
dirt track racing
drifting (motorsport)
stock car racing
rallycross
auto racing

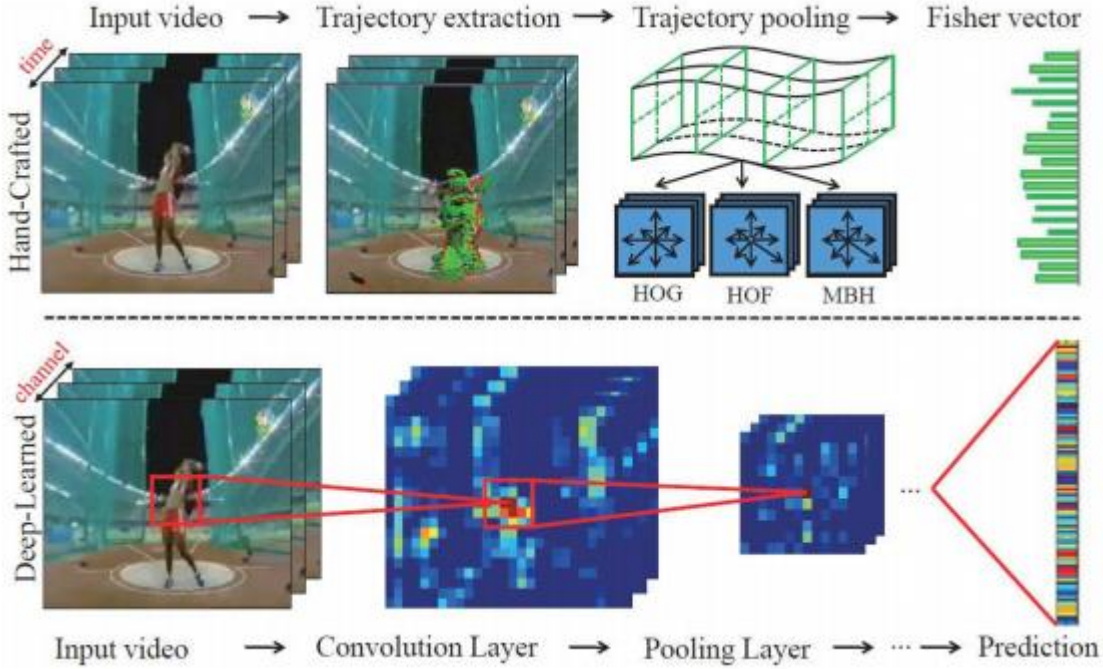
A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale Video Classification with Convolutional Neural Networks , CVPR 2014

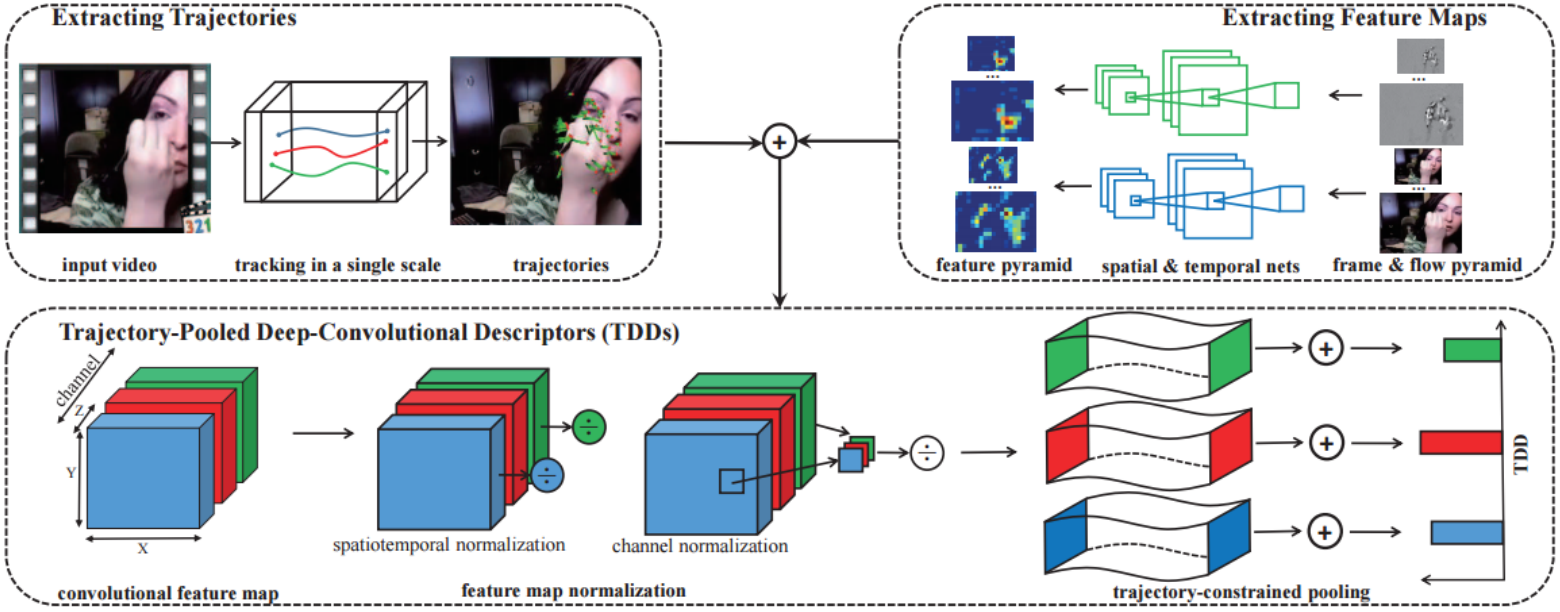
Two stream ConvNets





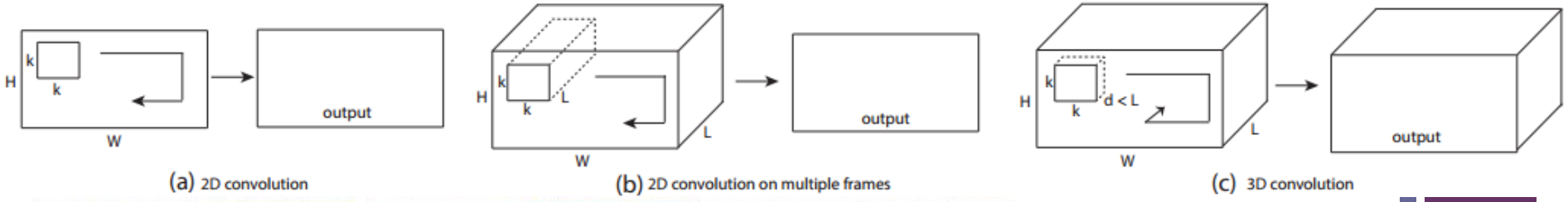
K. Simonyan, A. Zisserman, Two-Stream Convolutional Networks for Action Recognition in Videos , 2014





Limin Wang, Yu Qiao, Xiaoou Tang, Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors, CVPR 2015

3D ConvNets



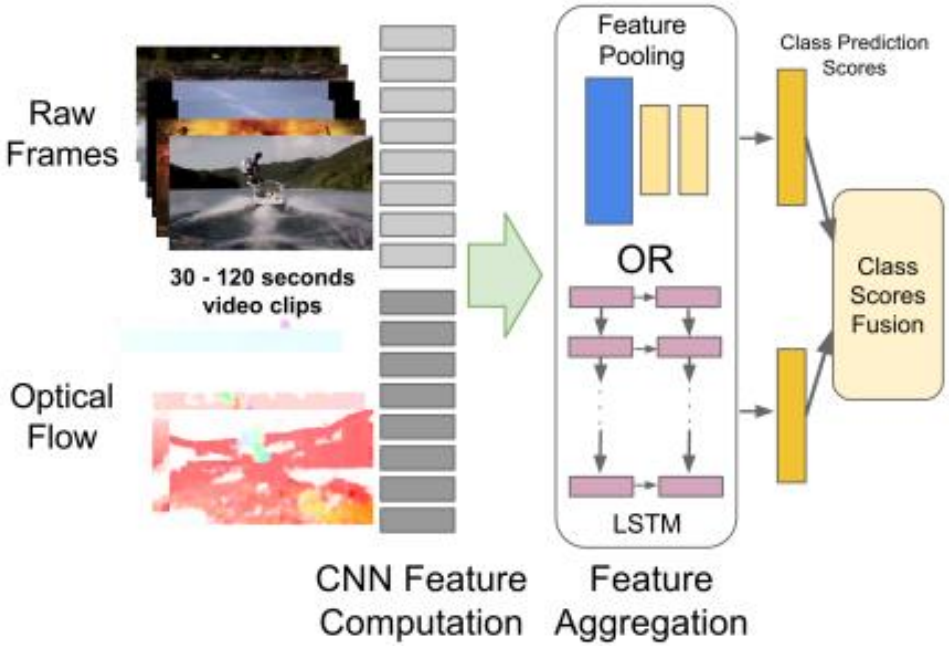
Moving edges and blobs

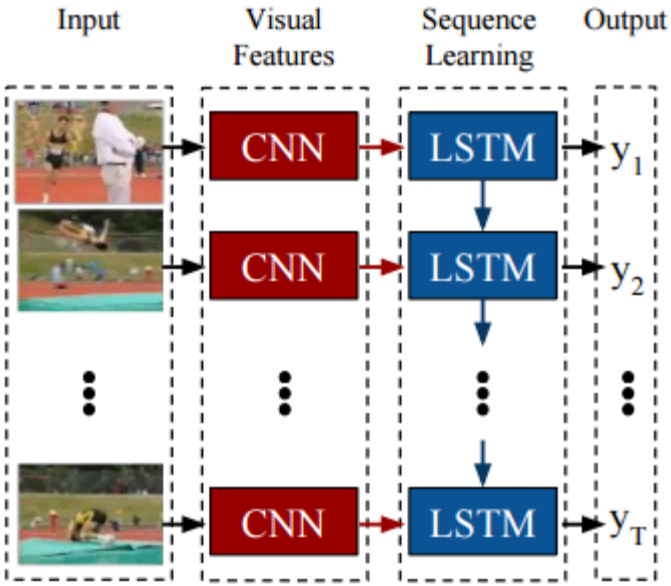
Shot changes

Edge orientation changes

Color changes

CNN ve LSTM





J. Ng, et al. Beyond Short Snippets: Deep Networks for Video Classification, CVPR 2015

Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell, Long-term Recurrent Convolutional Networks for Visual Recognition and Description, CVPR 2015

Çözüm Önerileri

- Video karesi temelinde
  - Kavramlarla ifade
  - Arama motorları ile ölçeklendirme

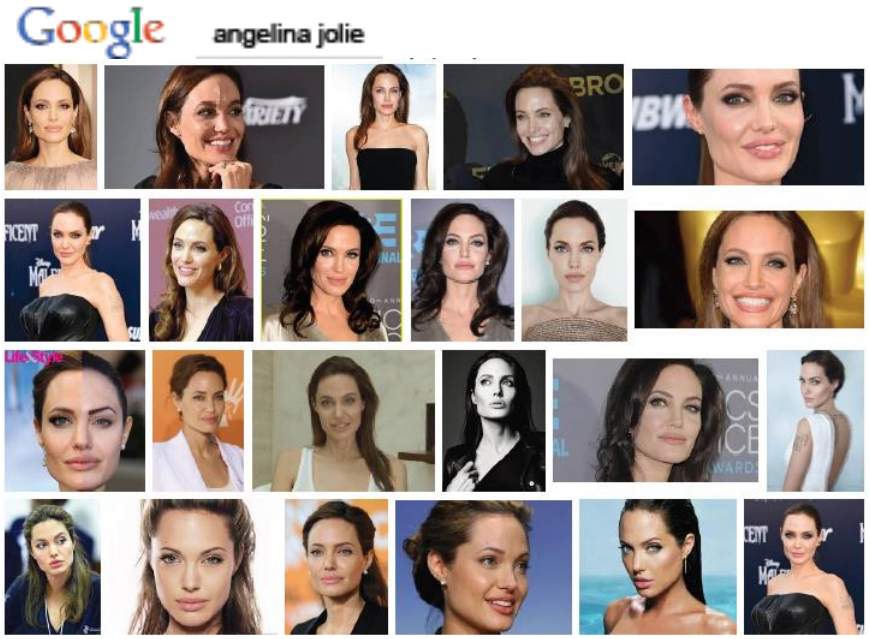
- Videodaki önemli parçaların keşfi
  - Ayırt edici kısımlar

- Hareketleri bulma
  - Daha basit ve hızlı tanımlayıcılar

Internetteki çok sayıda resim ve videodan faydalanarak kavram öğrenme

Pinar Duygulu, Eylul 2016, Ankara

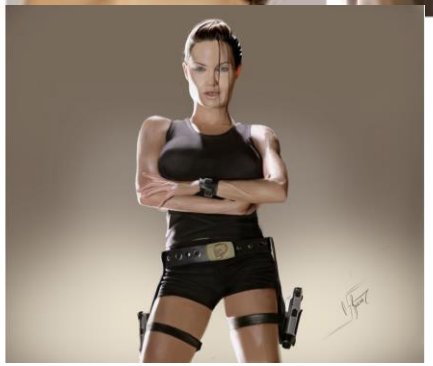
Google arama sonuçları



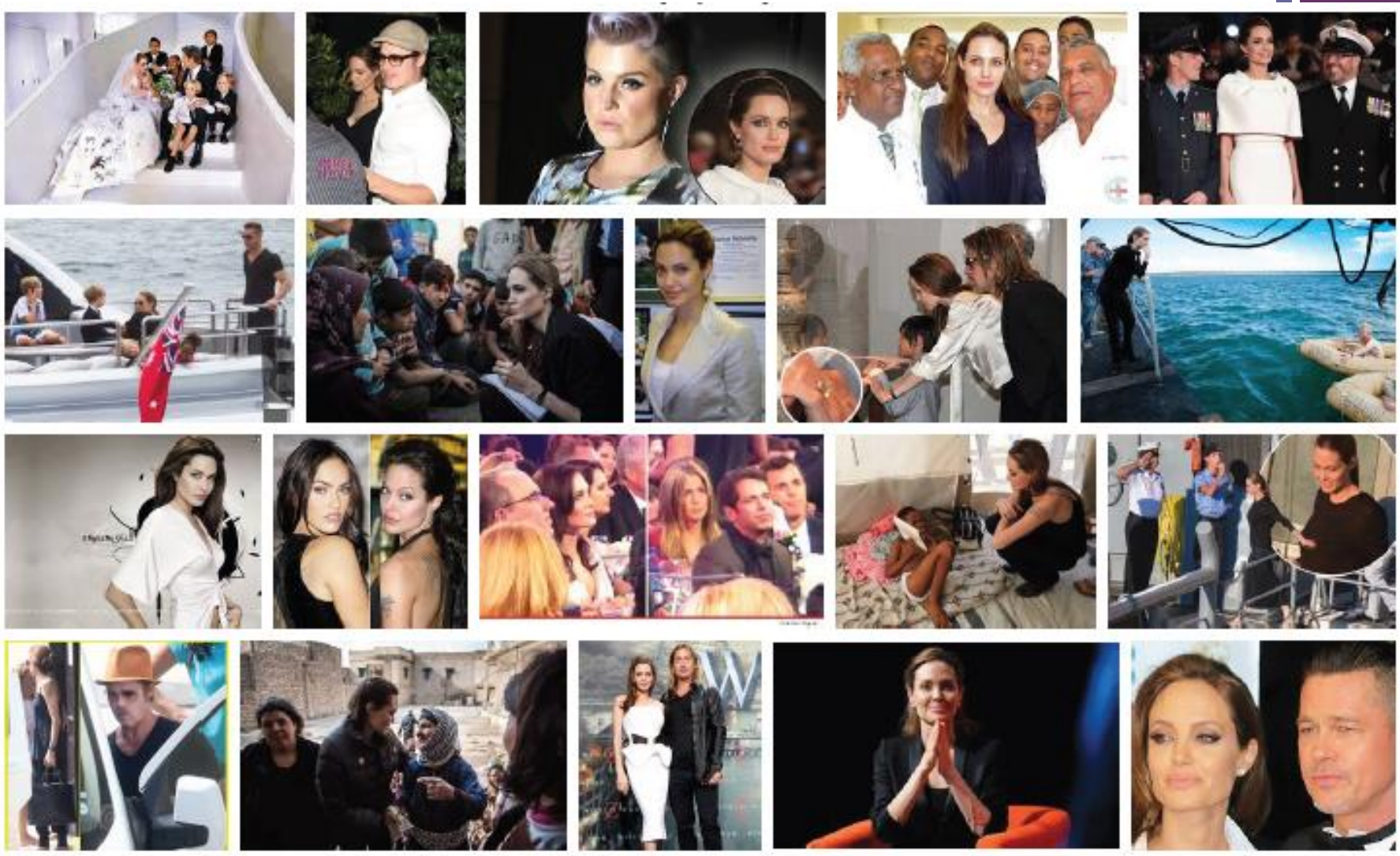
E.Golge, P.Duygulu. FAME: Face Association Through Model Evolution. In CVPR Workshops, 2015

Zayıf etiketli öğrenme

Sınıf içi farklılıklar



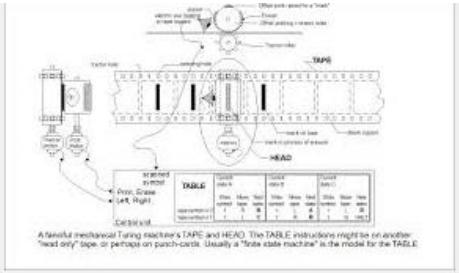
Ayrık elemanlar – sınıfdışı örnekler

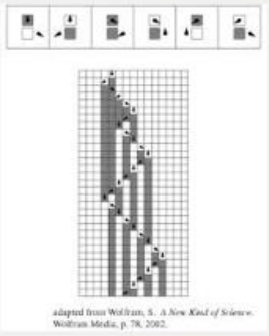


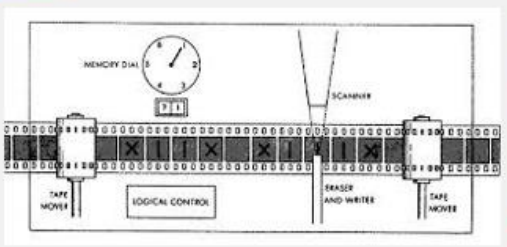
Daha iyi bir model için veri ayıklama

Sorgu tek bir sınıfa karşılık geldiğinde

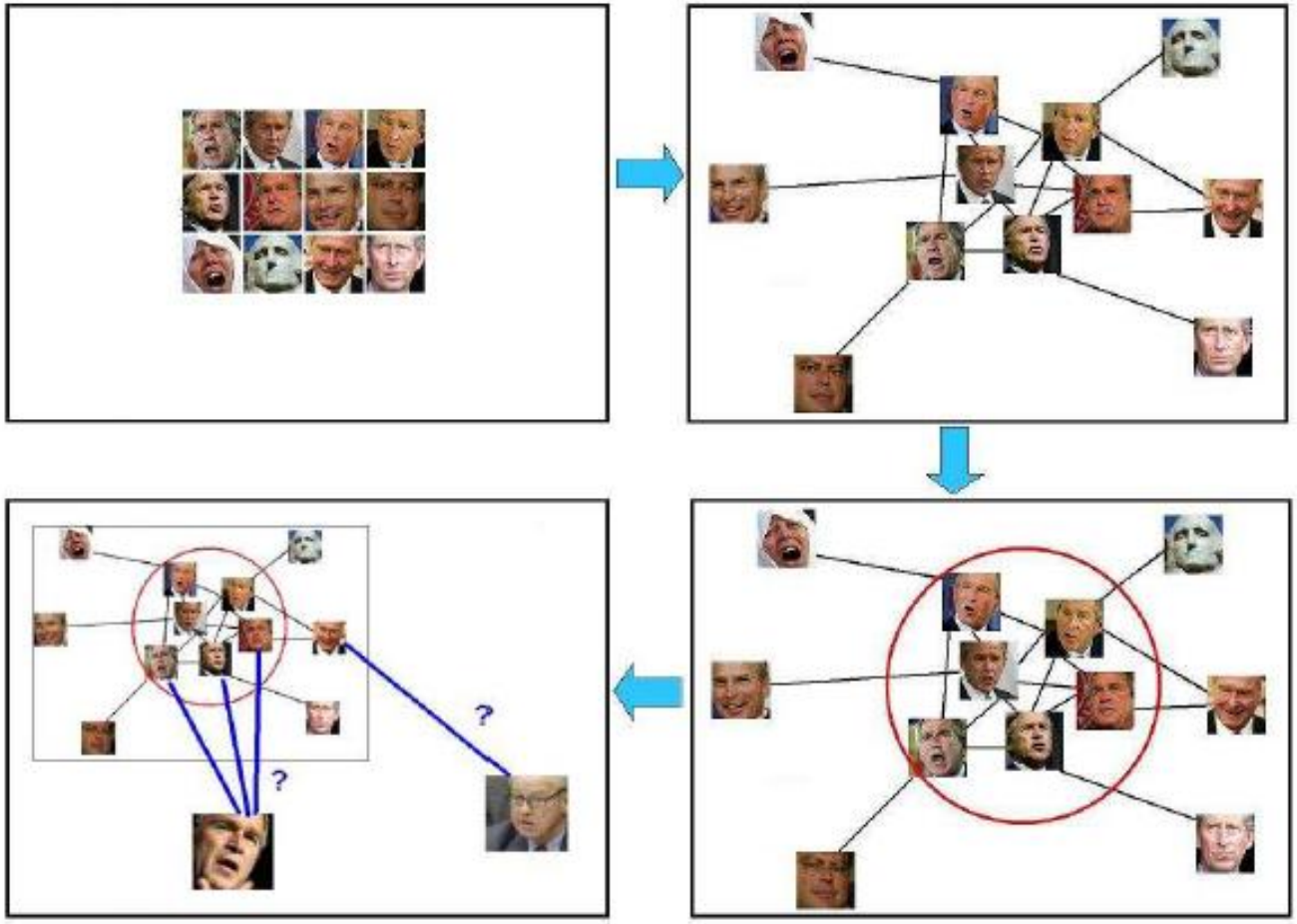
Sorgu : Turing







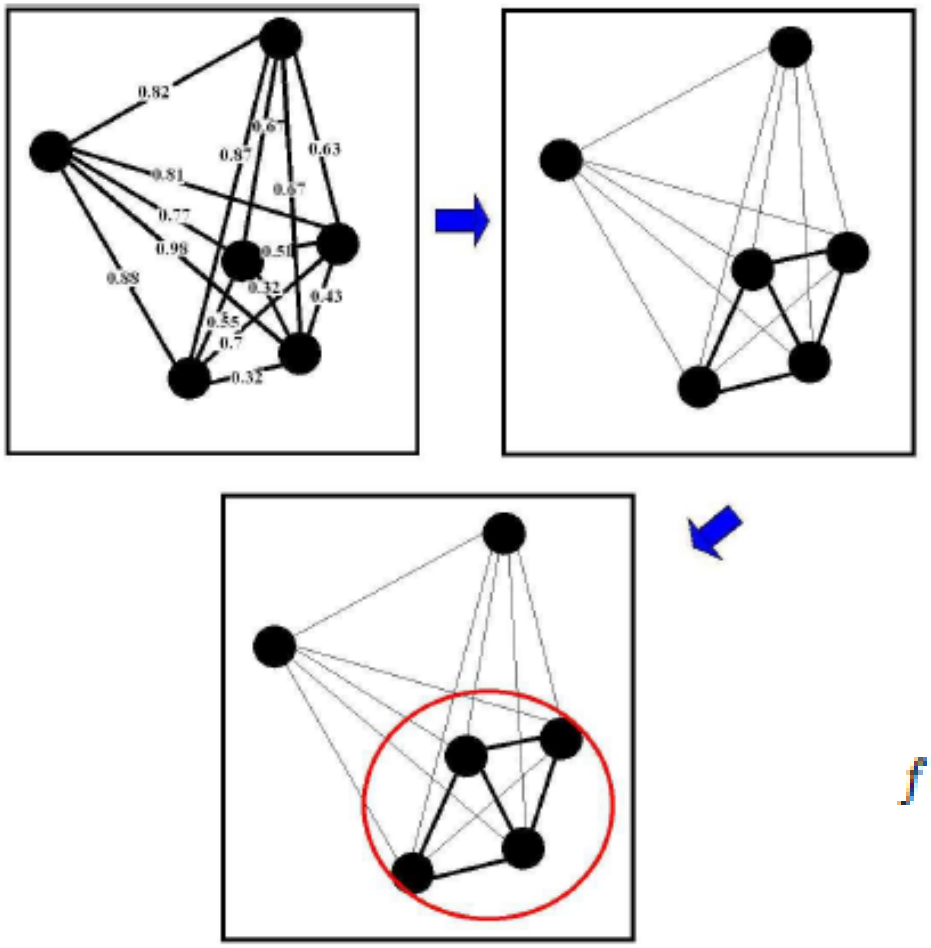
Yüz isimlendirme



Bir isim ile eşleşen yüzler arasında birbirine en yakın alt kümenin bulunması

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

En yoğun alt bileşeni bulma



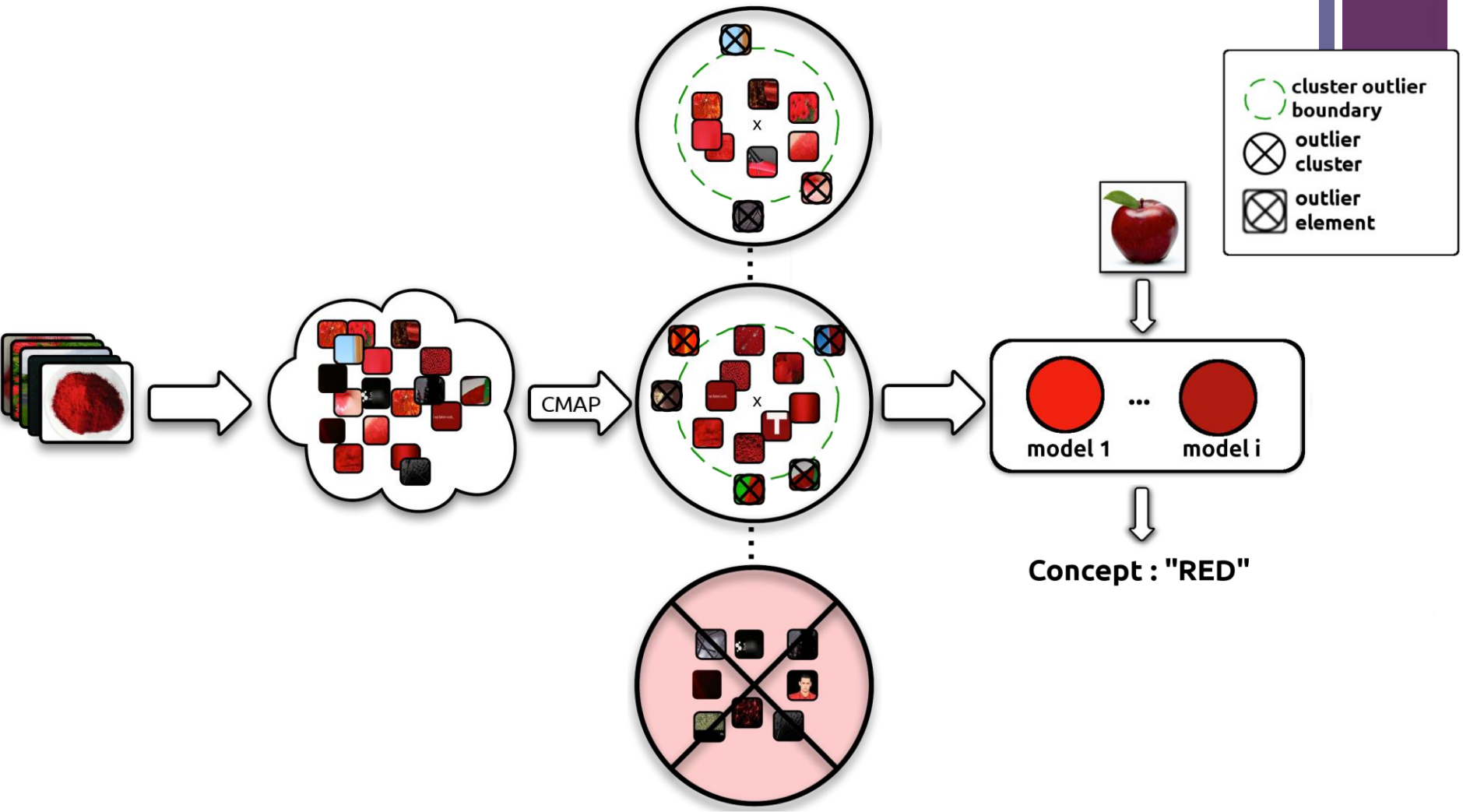
$$f(S) = \frac{|E(S)|}{|S|},$$

Her seferinde minimum dereceye sahip düğüm atılır

(Charikar, 2000)

Birden fazla anlam ya da farklılık olduğunda

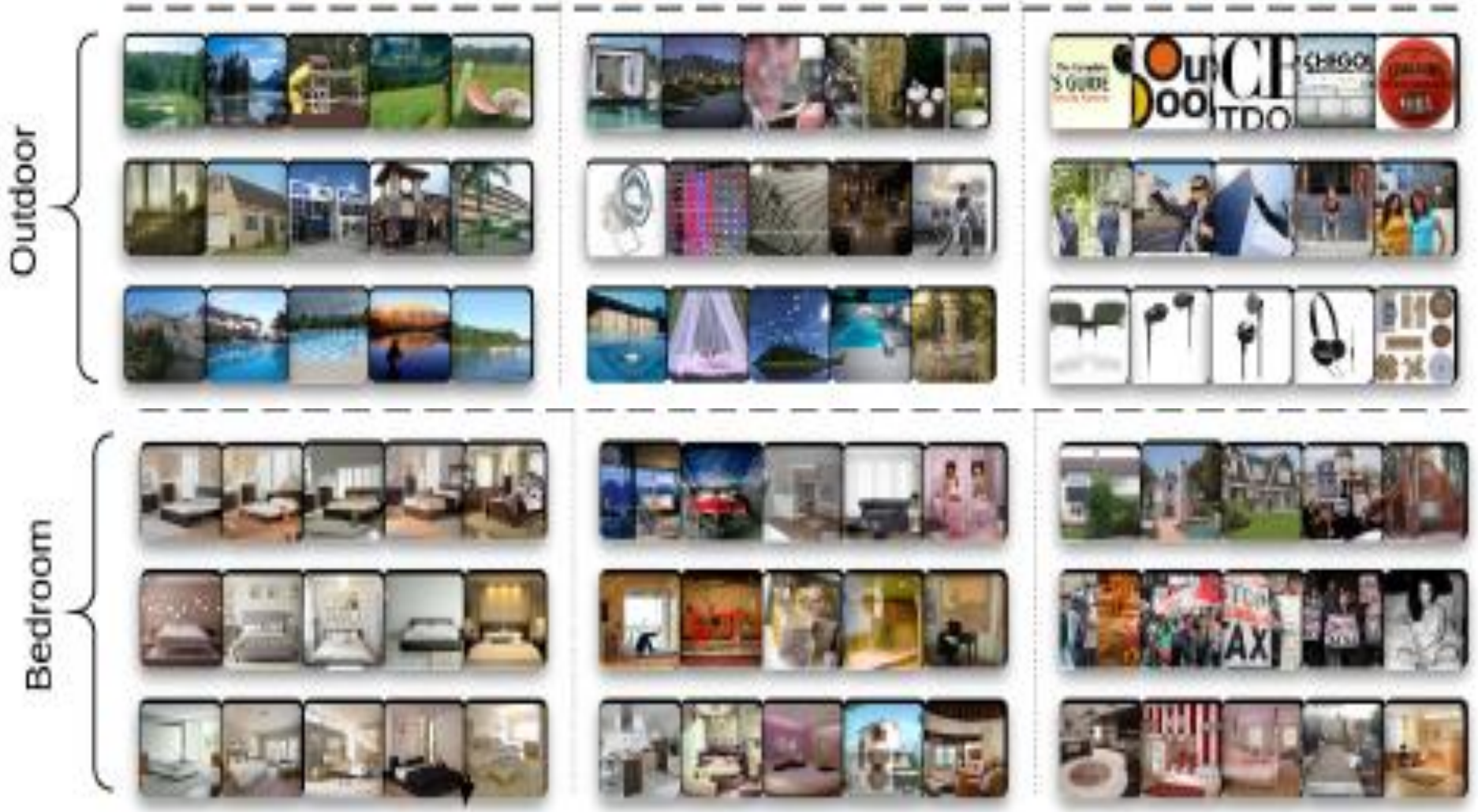
Kavram Haritası



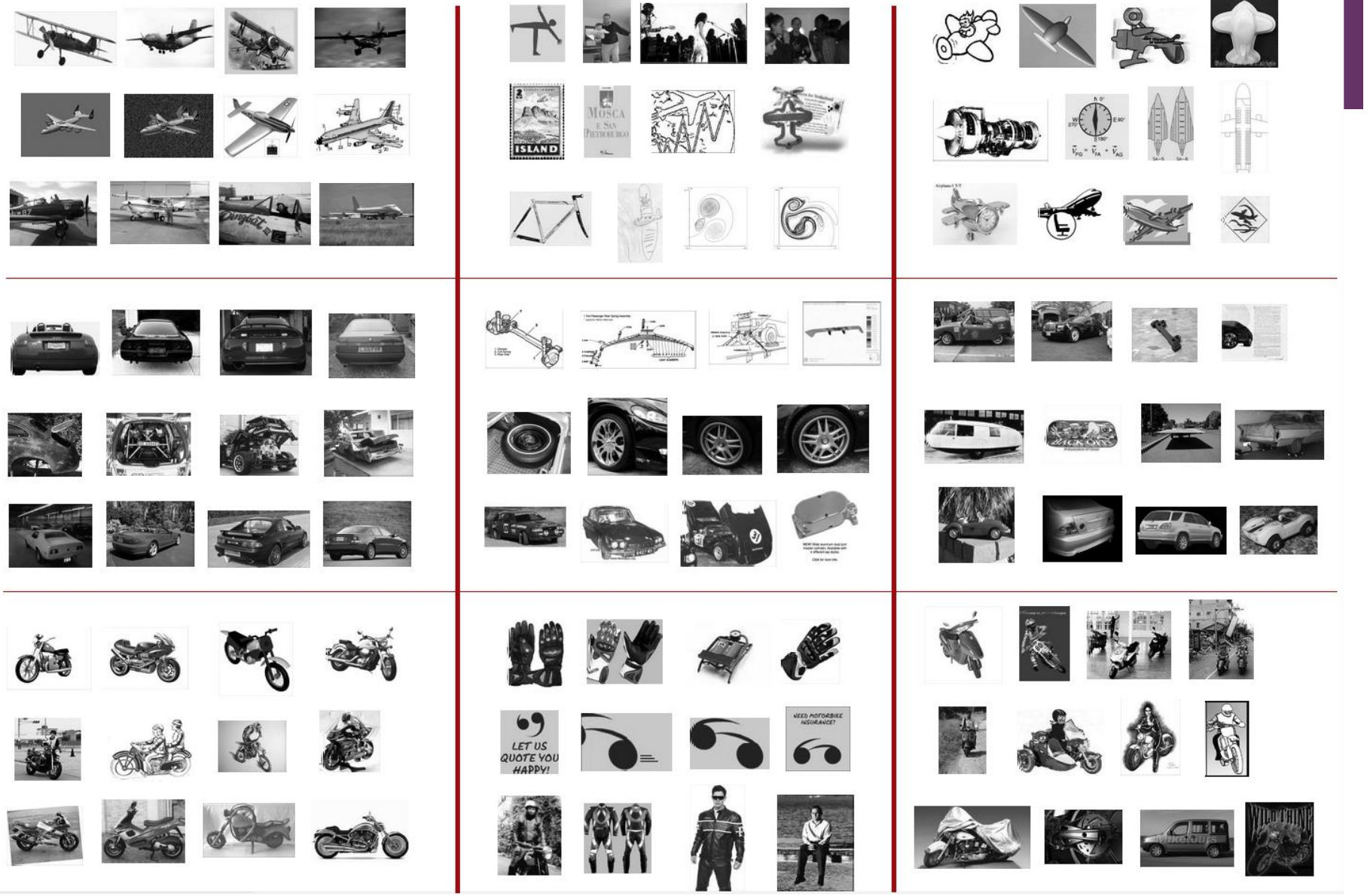
Golge, E., Duygulu, P., "Concept Maps: Mining Noisy Web Data for Concept Learning ", accepted to ECCV 2014

Renk ve doku nitelikleri

Sahne/Ortam

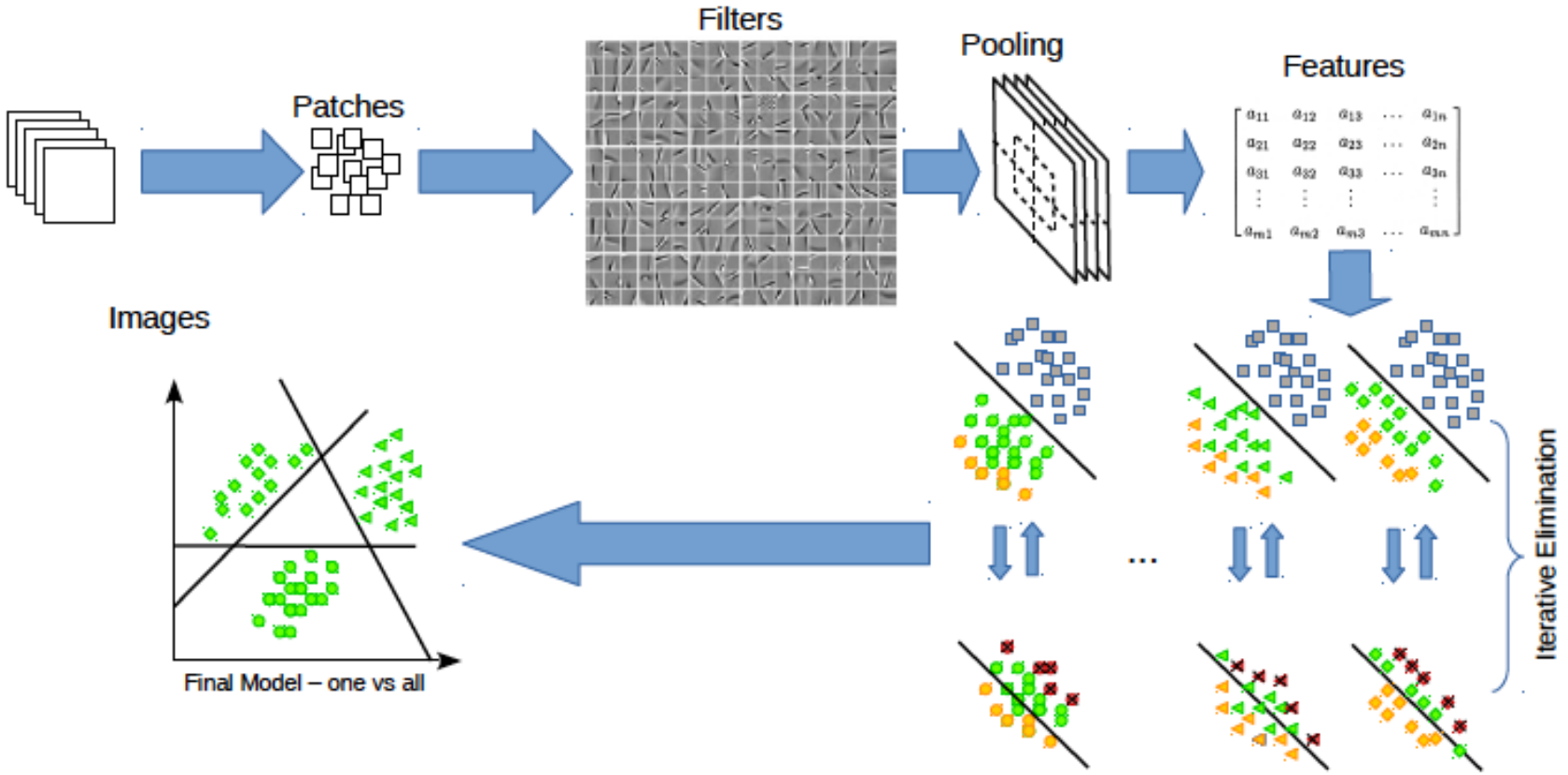


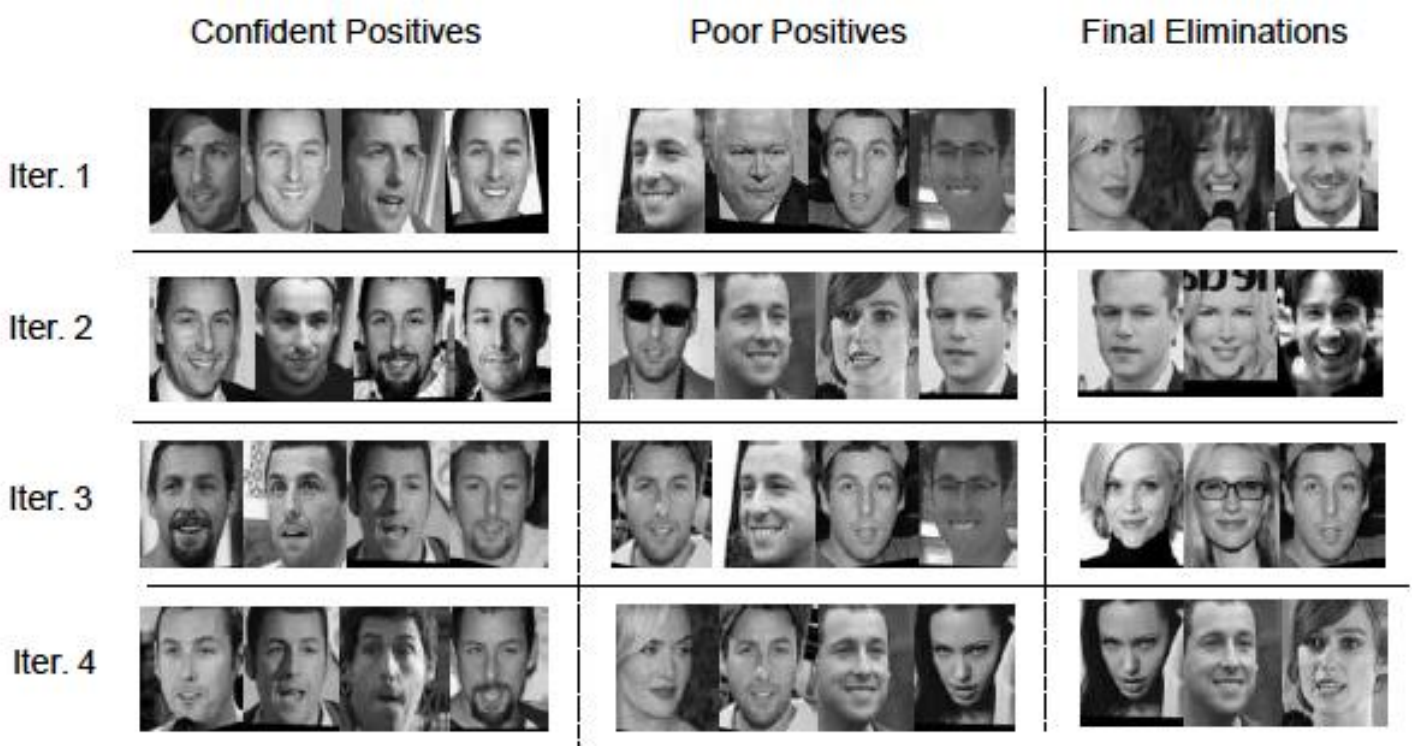
Nesne tanıma



Yüz tanıma

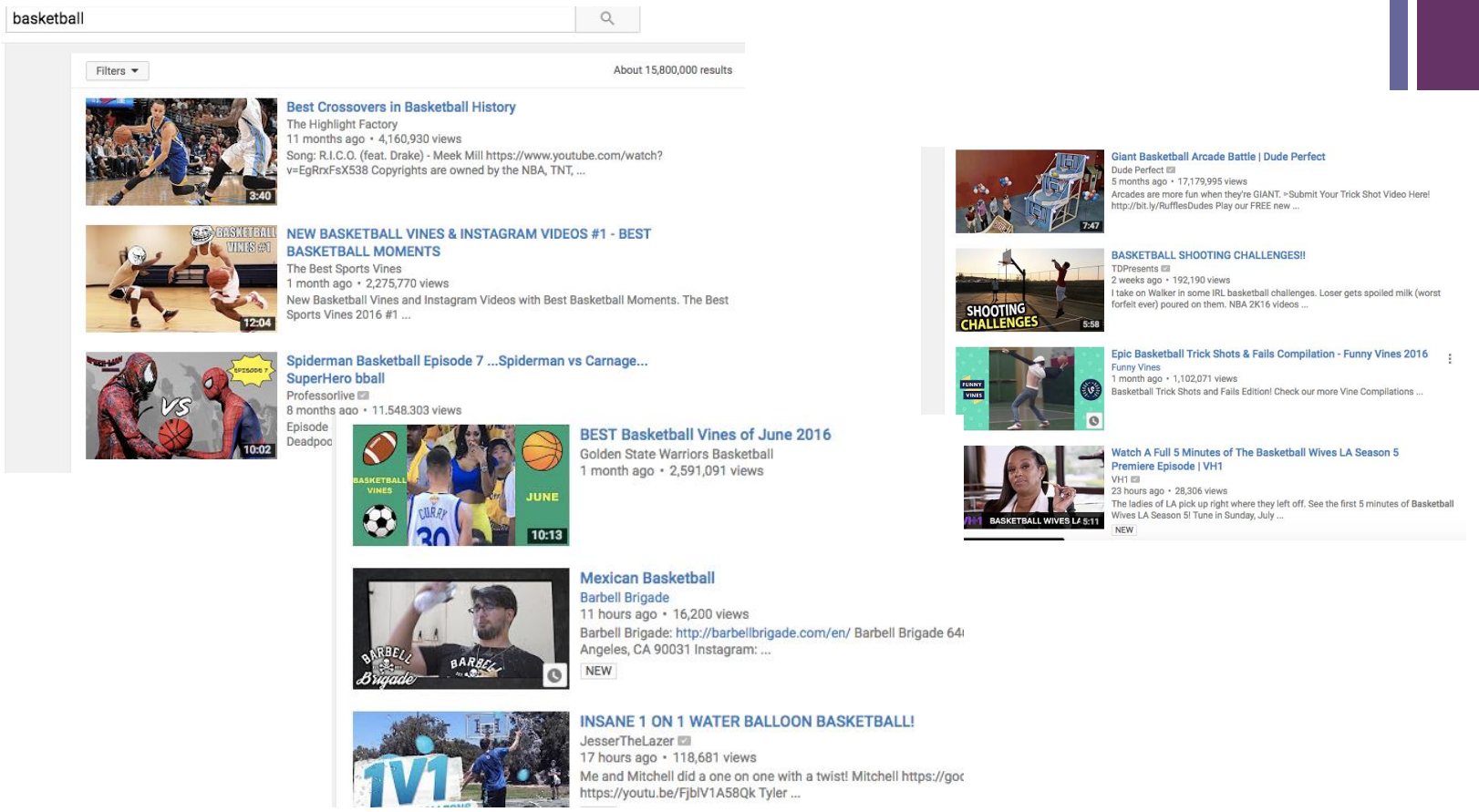
Model Evrimi ile Eşleme (AME)




Confident Positives
Poor Positives
Final Eliminations
Iter. 1
Iter. 2
Iter. 3
Iter. 4

Iteration 1
Iteration 2

Zayıf etiketli videolardan öğrenme



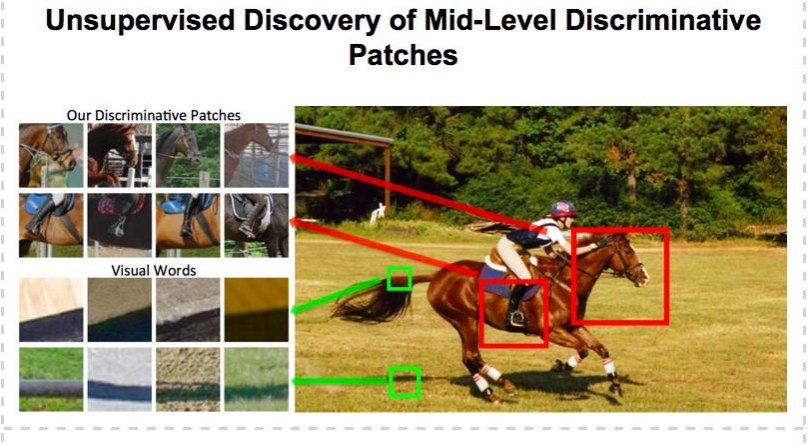
Zayıf etiketli videolardan öğrenme

Çoğul ortam verisinde olay tanıma Doğumgünü

Mum üfleme

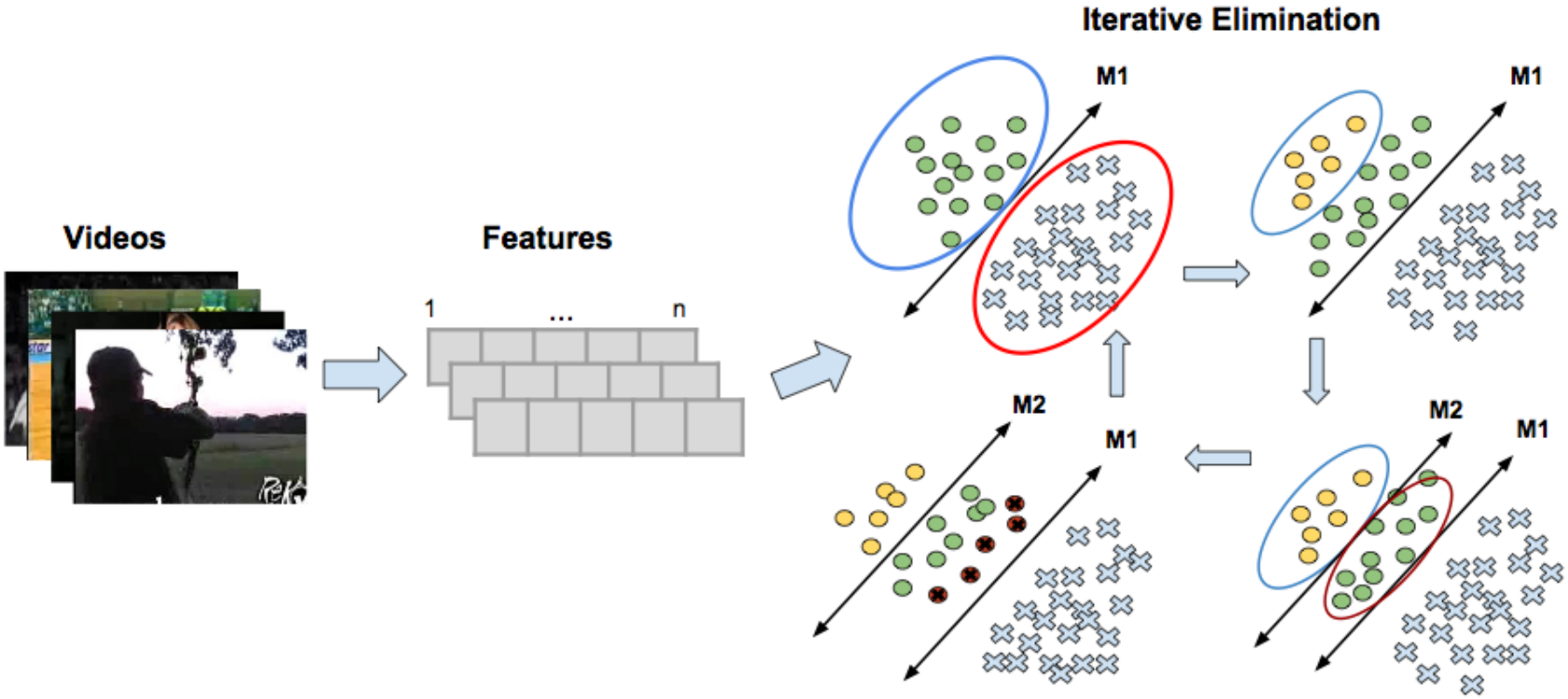
Önemli parçaların bulunması

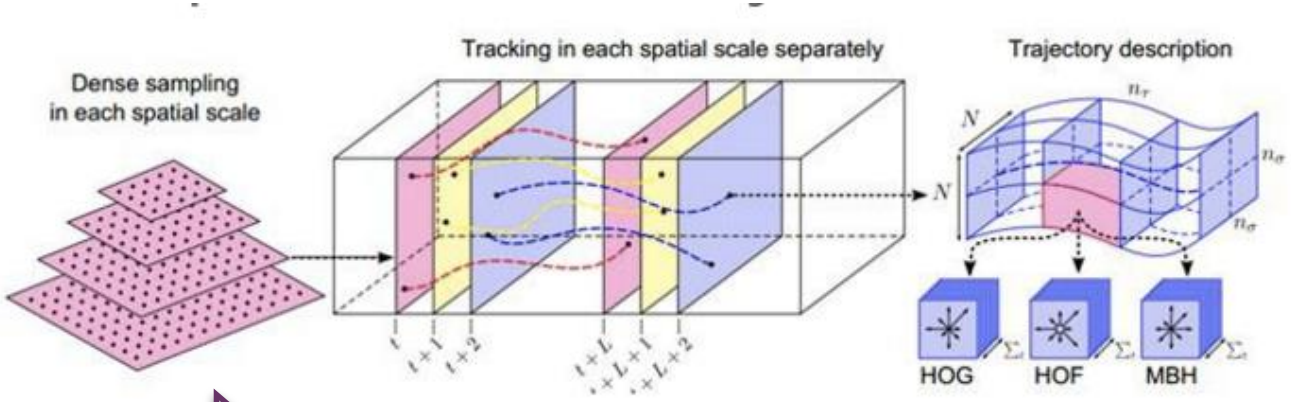
Singh ECCV 2012

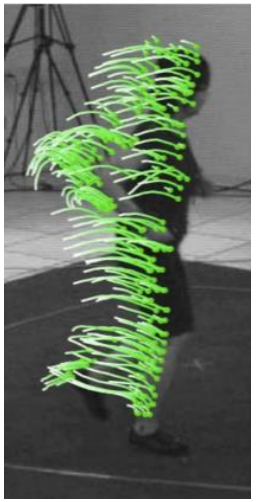


Jain CVPR 2013

AME



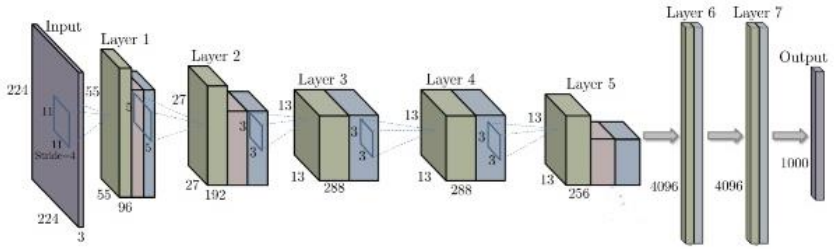




Input Video

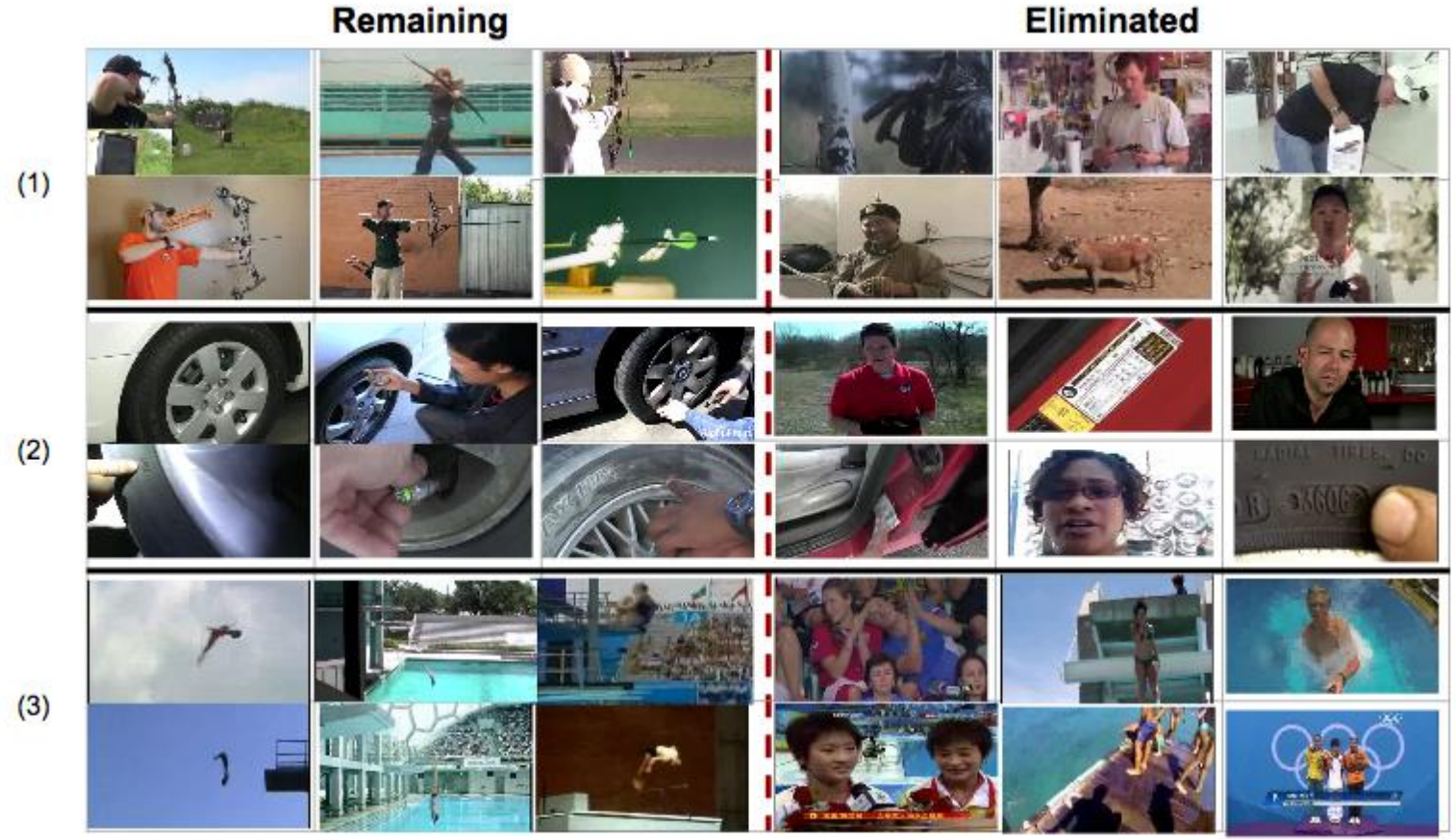
CNN FC-6 Layer Features

Classification



AlexNet [8,9,10]

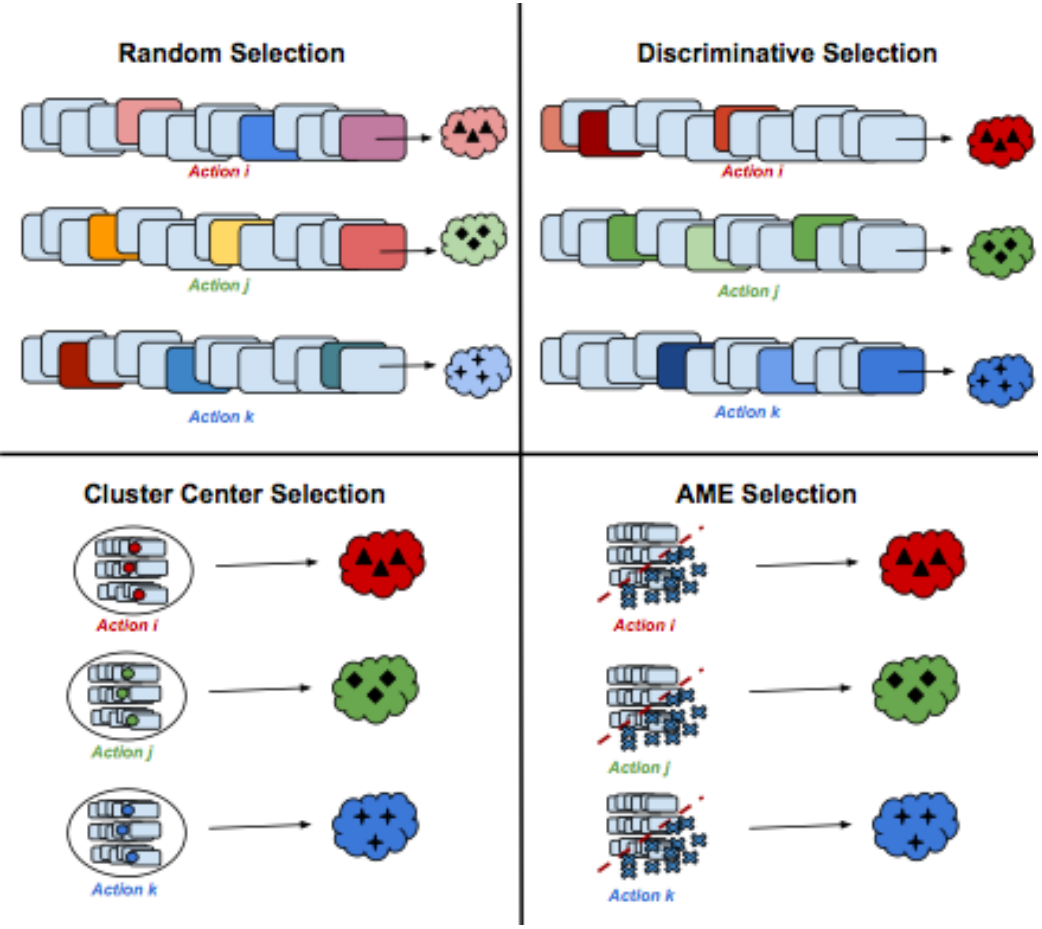
3D ConvNets [11,12]

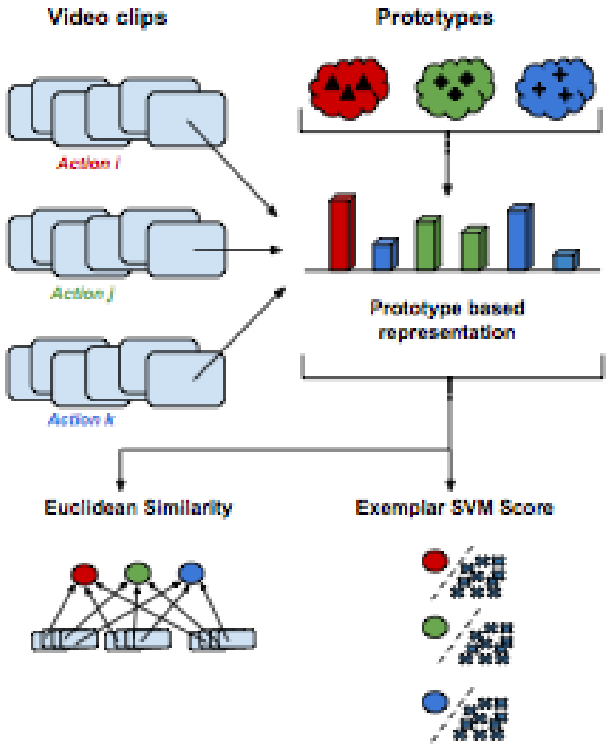


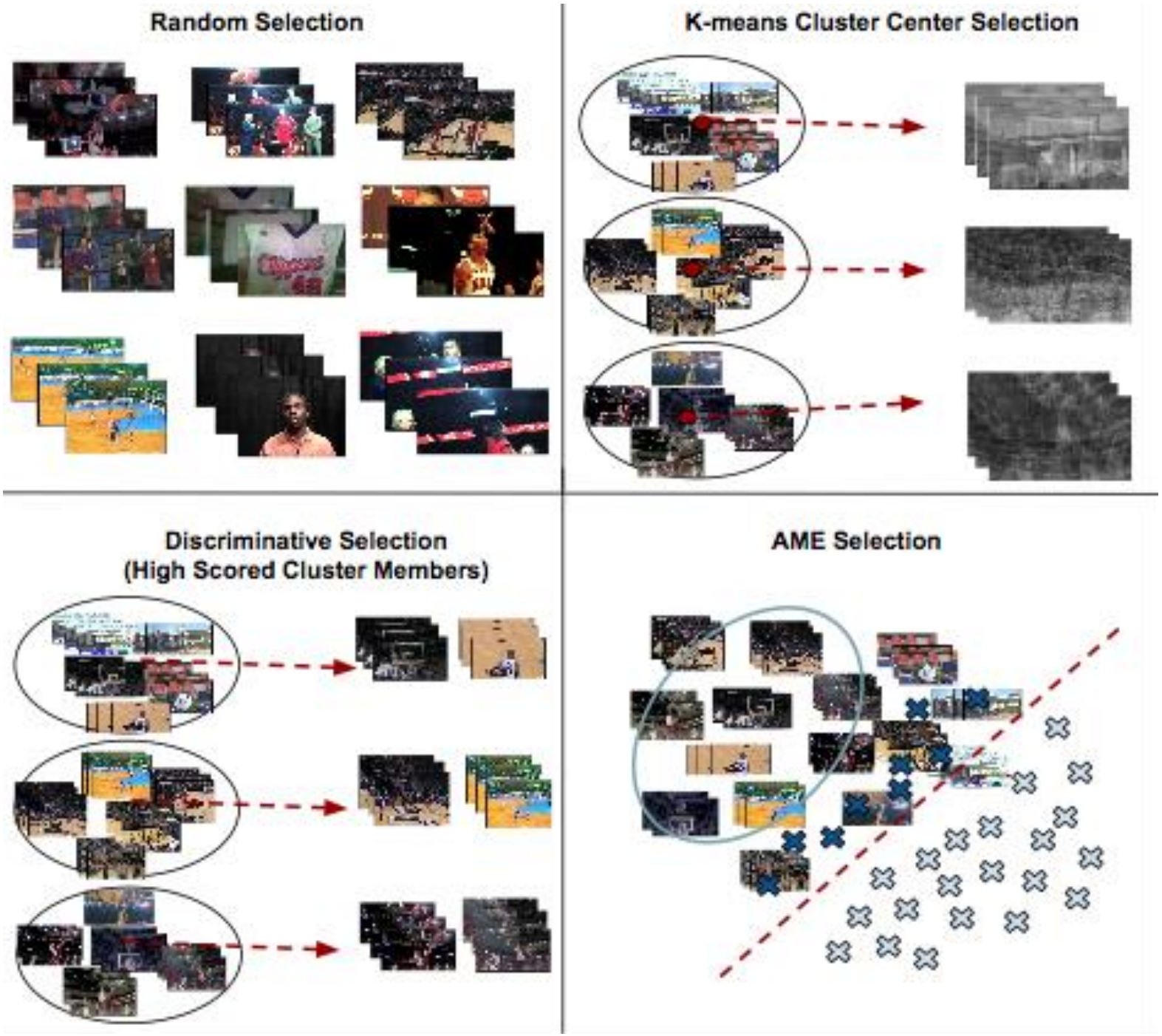
Example of successful eliminated instances by AME[1] for ActivityNet action classes.(1) "Archery" class. Baseline: 31.57%, AME: 44.73%. (2) "Checking tires" class. Baseline: 26.82%, AME: 41.46%. (3) "Platform diving" class. Baseline: 56.66%, AME: 73.33%

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

Prototipler ile gösterim



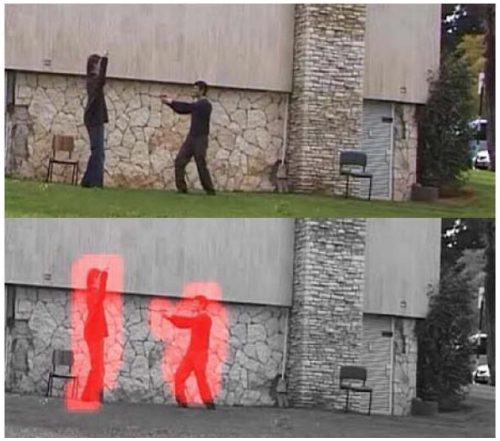



Random Selection
K-means Cluster Center Selection
Discriminative Selection
(High Scored Cluster Members)
AME Selection

Videolardaki beklenmedik durumların keşfi

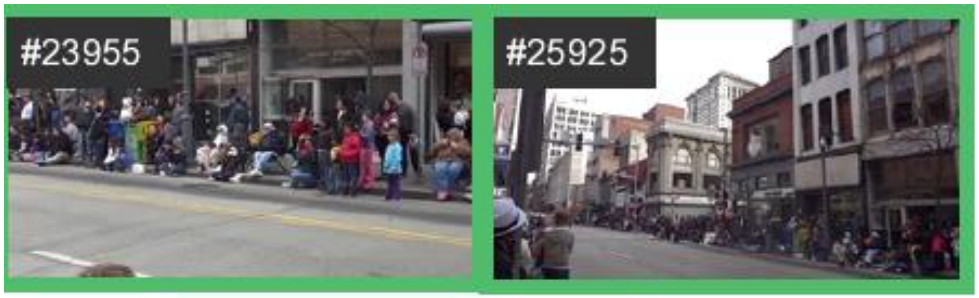
Pinar Duygulu, Eylul 2016, Ankara

three



Boiman and Irani ICCV 2005

Roshtkhari and Levine, CVPR 2013



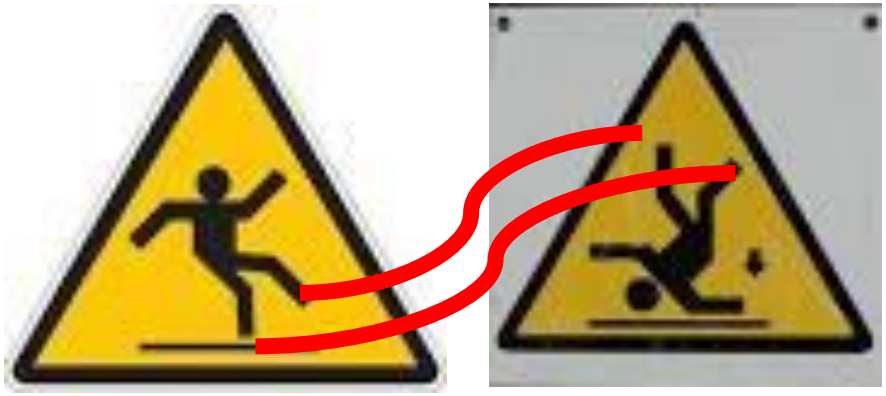
Ito, Kitani, Bagnell, Hebert, 2012

Zhao, Fei-Fei, Xing, CVPR 2011

Olağan yerine olağandışı

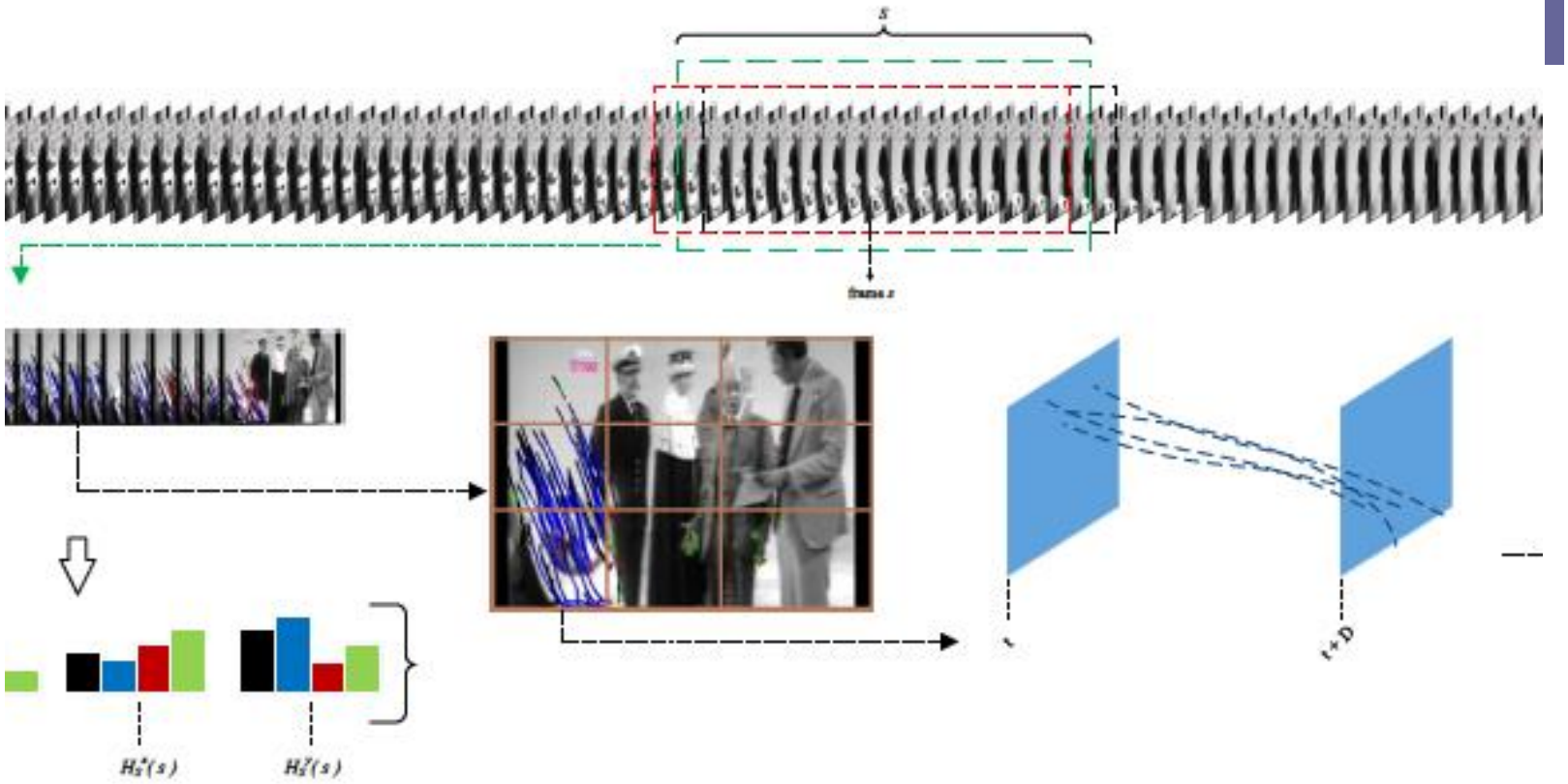
Pinar Duygulu, Eylul 2016, Ankara

Ani hareketler

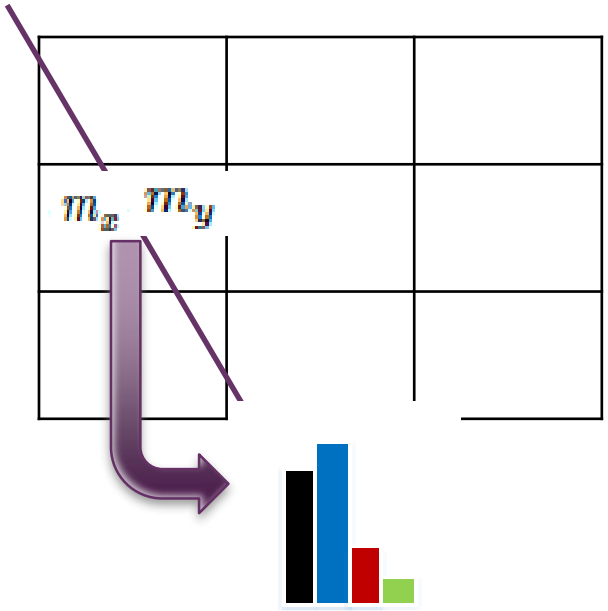


Yüksek hız
Büyük uzamsal kapsam

Yoğun İz Dağılımları



Gösterim



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

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

Velocity and spatial extension of the motion

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

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

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

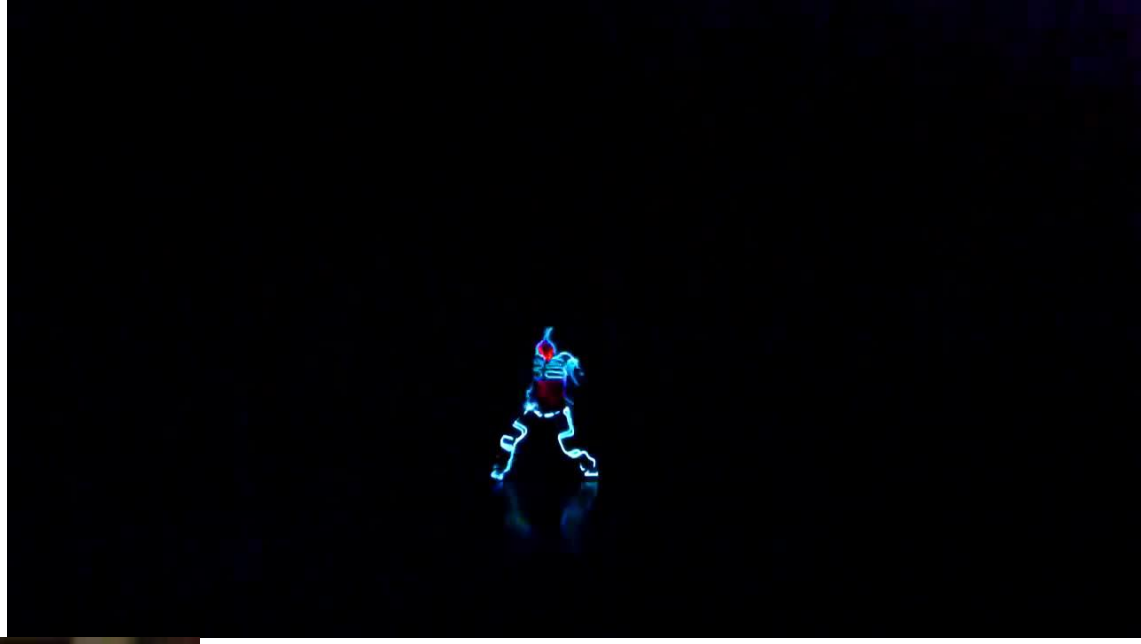
$$l = \sum_t^{t+D-1} \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2}$$

Pinar Duygulu, Eylul 2016, Ankara

Pinar Duygulu, Eylul 2016, Ankara

Pinar Duygulu, Eylul 2016, Ankara

Başarısız durumlar



Pinar Duygulu, Eylul 2016, Ankara

Kısıtlar

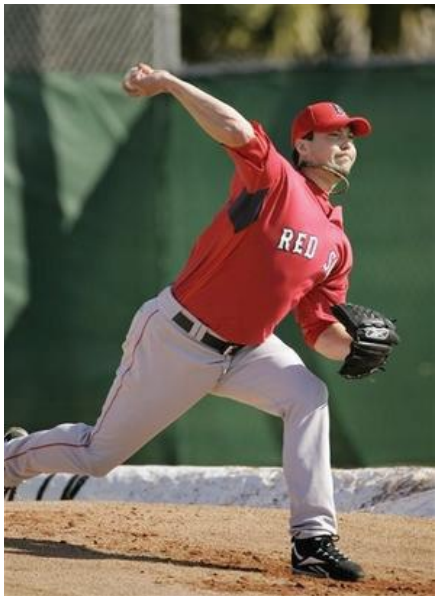
Hareket analizi

Pinar Duygulu, Eylul 2016, Ankara

Bu insanlar ne yapıyor?

running

walking



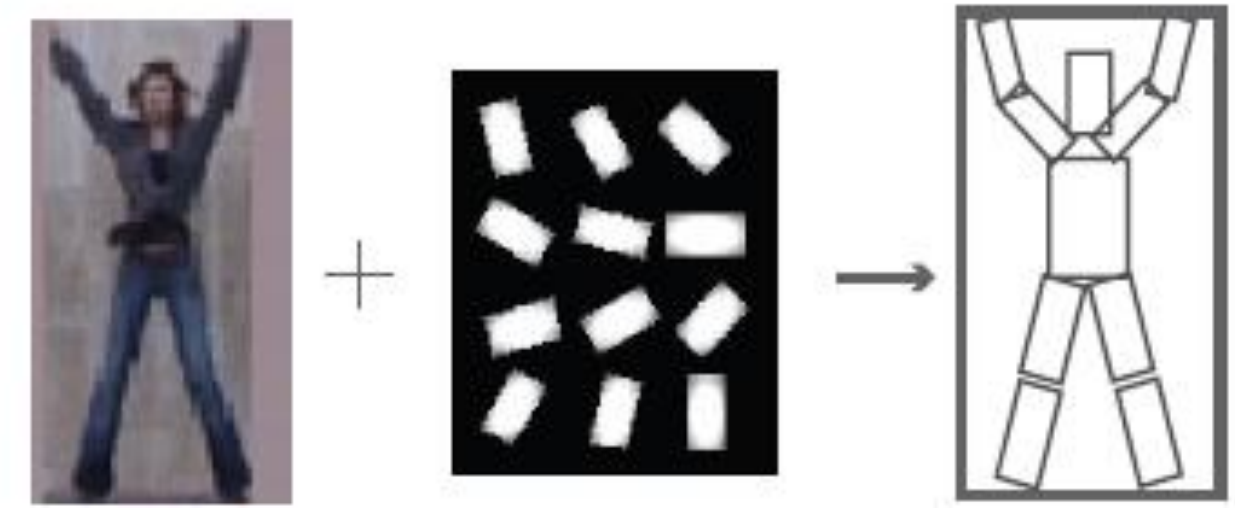
throwing

crouching

- Poz çok bilgi içeriyor ama nasıl tanımlamalı?

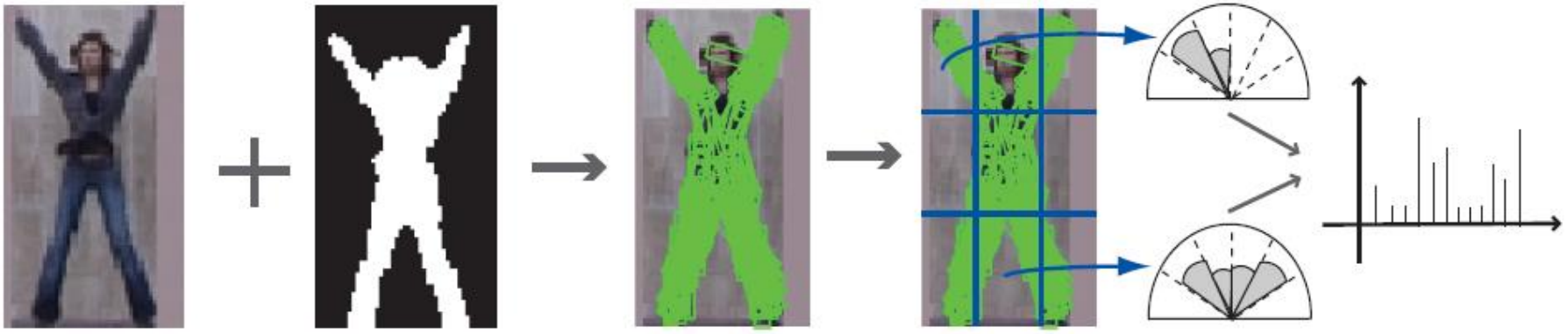
Dikdörtgenlerle pozu tanımlama

- Silindire benzeyen vücut parçalarının izdüşümü olan dikdörtgenlerin açı ve uzunlukları

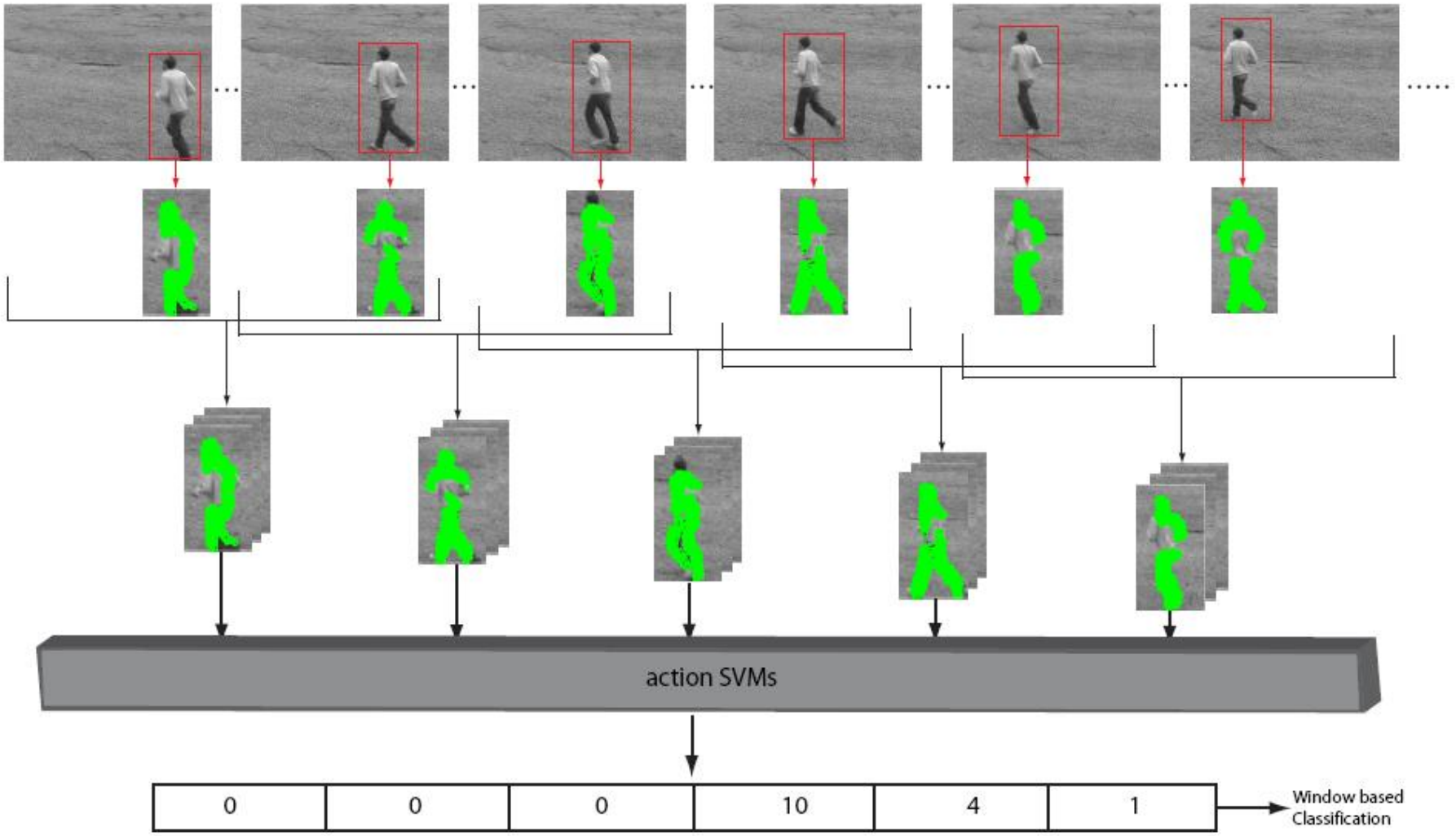


Ikizler, N. Duygulu, P. ''Human Action Recognition Using Distribution of Oriented Rectanguar Patches'', Proc. 2nd Workshop on Human Motion: Understanding, Modeling, Capture and Animation, In conjunction with ICCV2007
Ikizler, N. ve Duygulu P., ''Histogram of Oriented Rectangles: A New Pose descriptor for Human Action Recognition'', Image and Vision Computing, volume 27, Issue 10, pages 1515-1526, September 2009

Histogram of Oriented Rectangles (HOR)

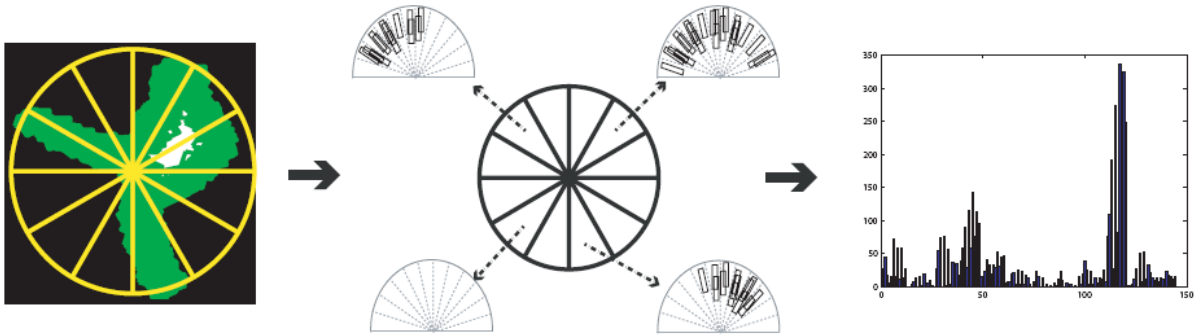


Sınıflandırma



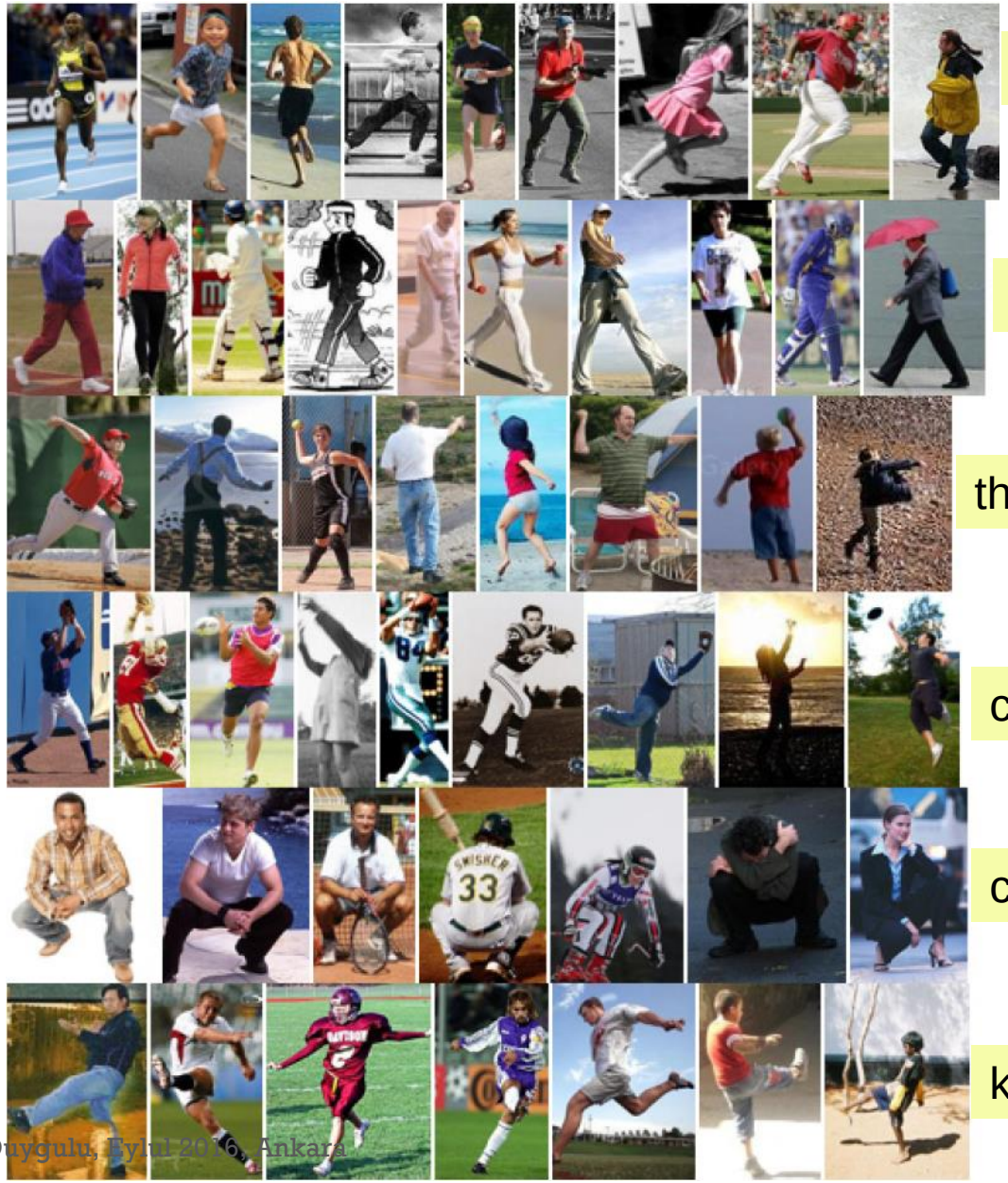
Tek resimde hareket tanıma

- Ramanan'ın yöntemi ile poz tanıma + HOR



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

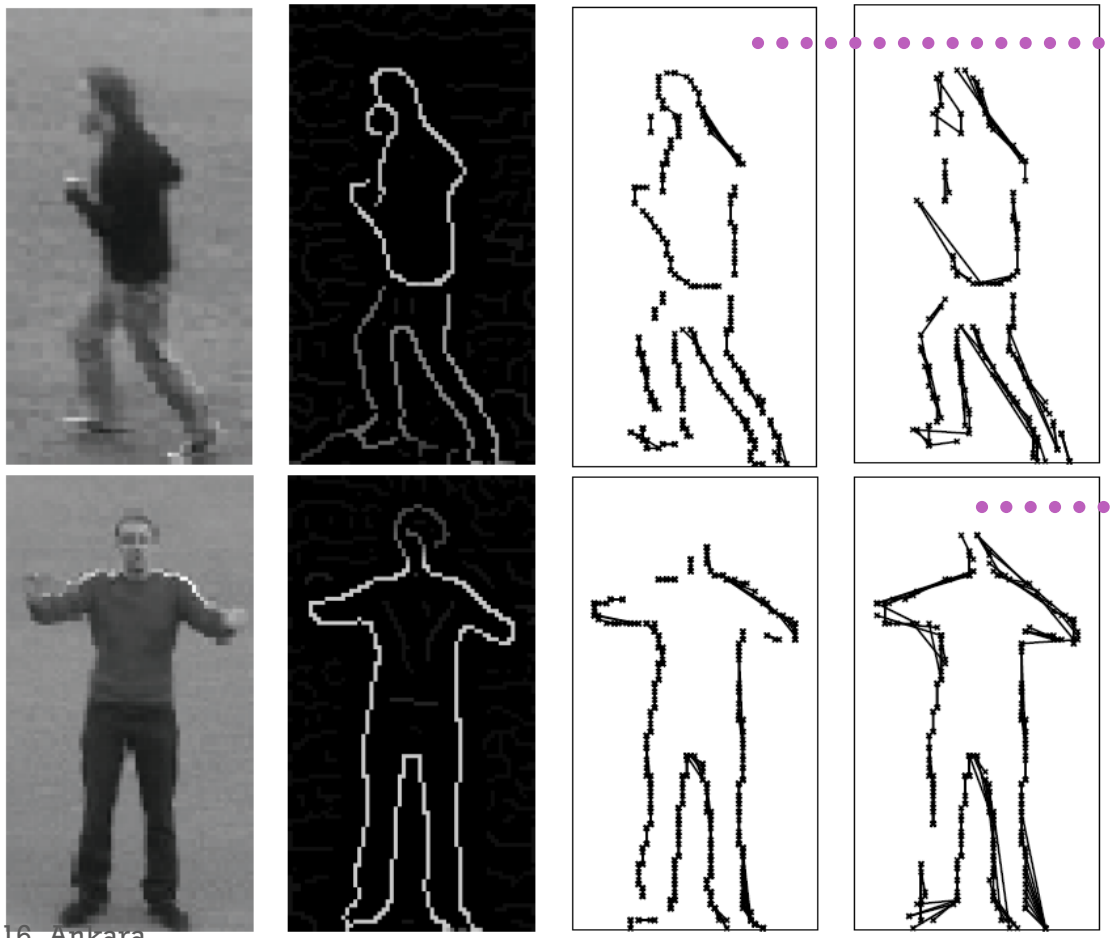
kicking

ActionWeb dataset - 467 images collected from the web

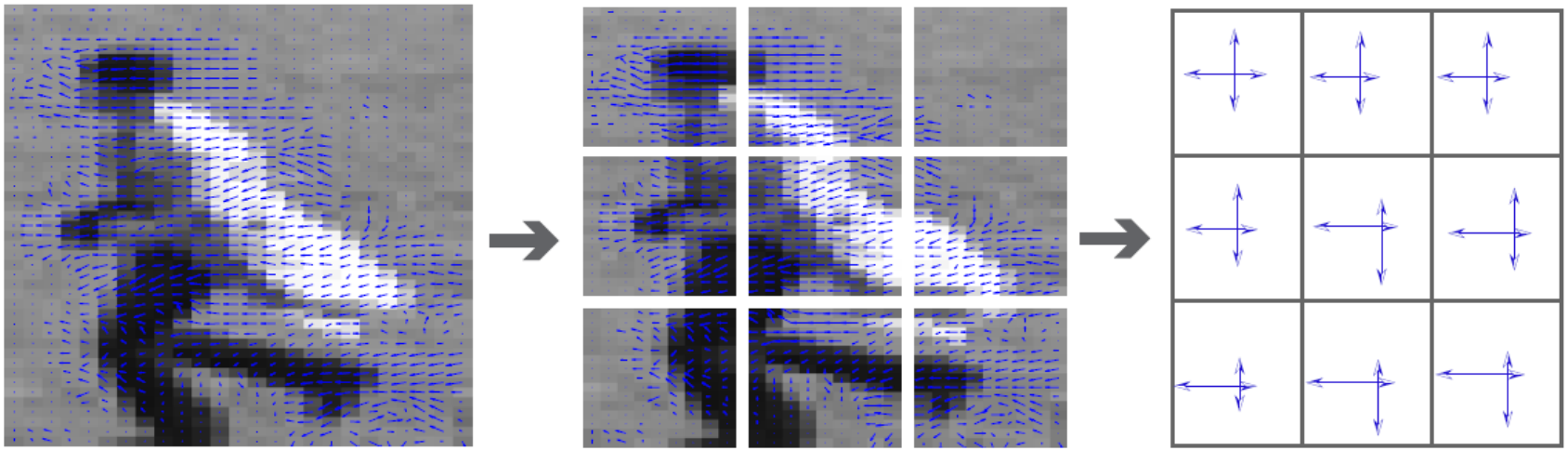
Çevreye oturtulan çizgiler

- Siluet olmadığında çevreye oturtulan çizgiler (Pb)

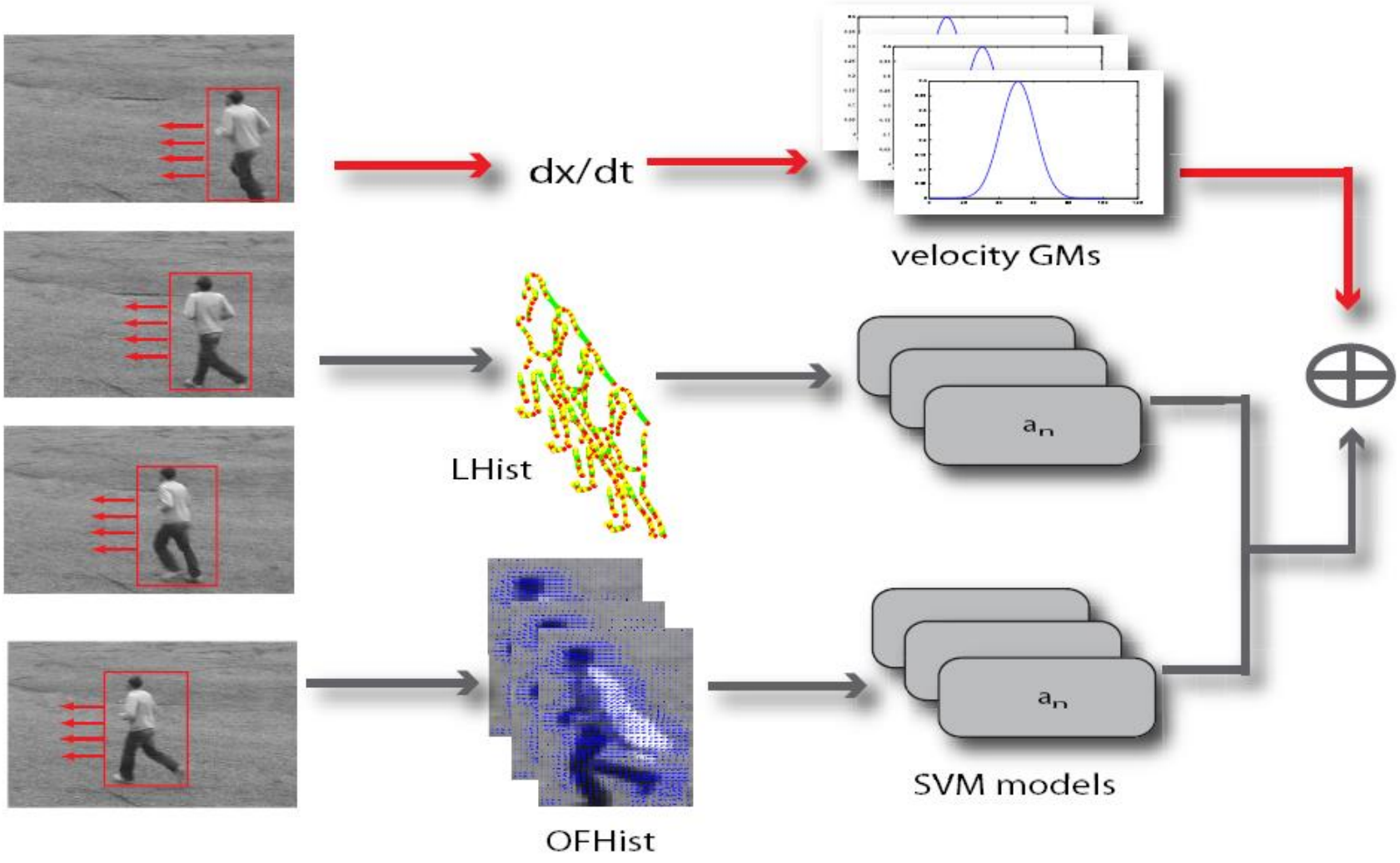
(Martin PAMI2004) yoluyla tanımlama



..ve Optik Akış

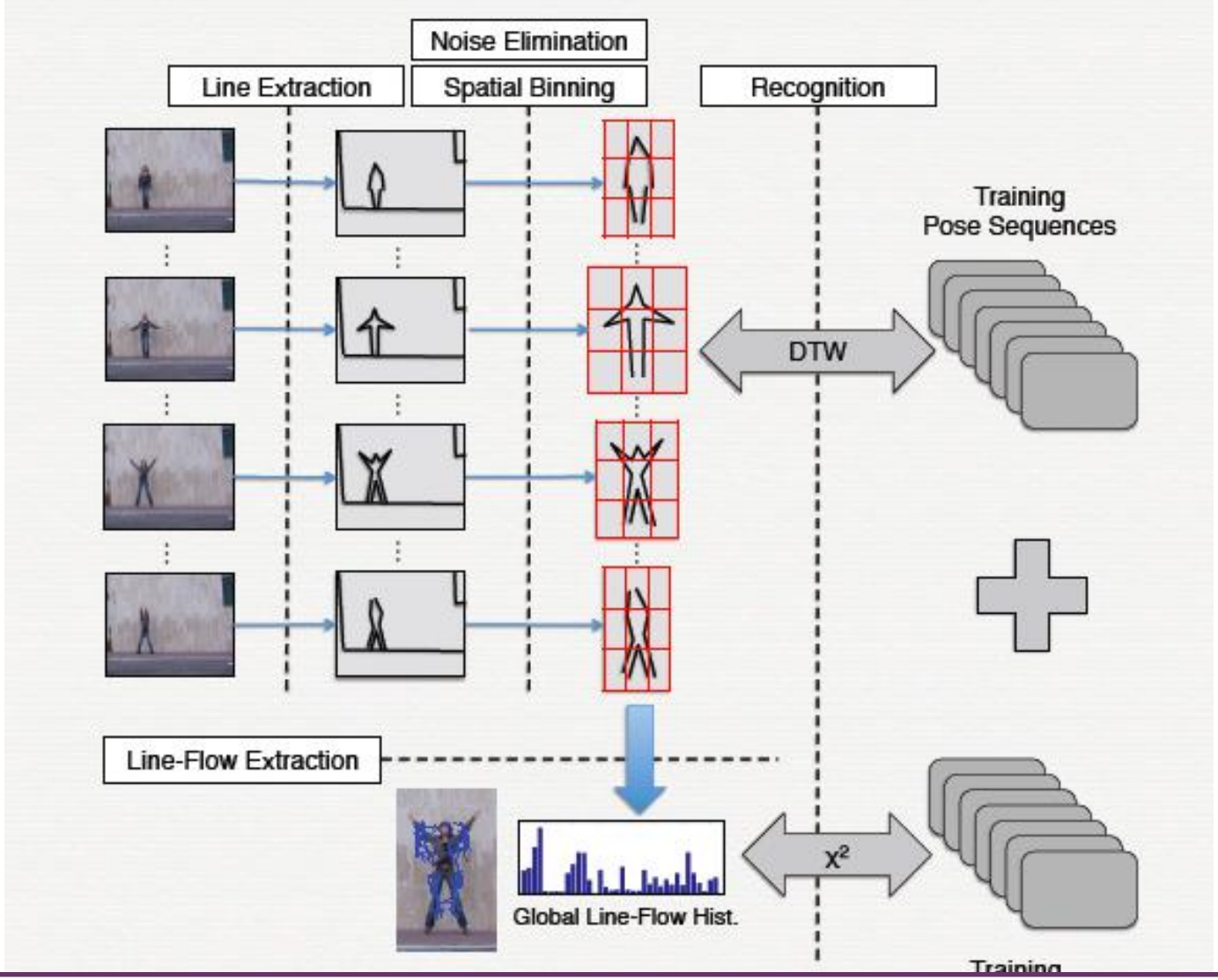


LHist ve OFHist



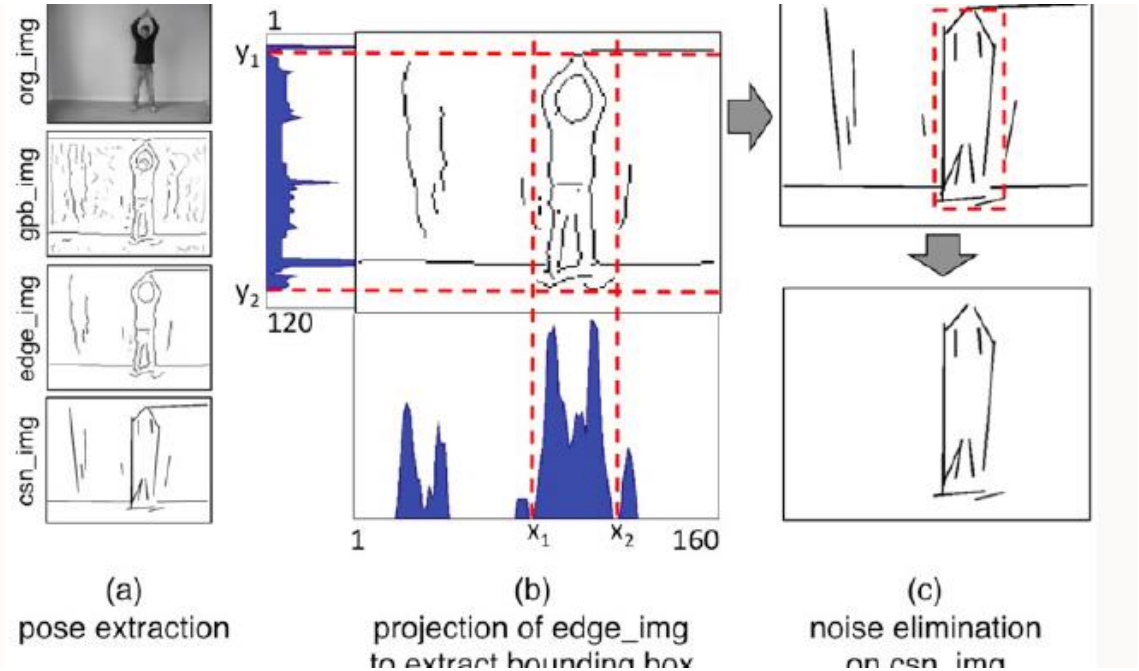
Ikizler, N., Cinbis, R. G., Duygulu, P., ''Human action recognition with line and flow histograms'', Proc. 19th International Conference on Pattern Recognition (ICPR 2008),

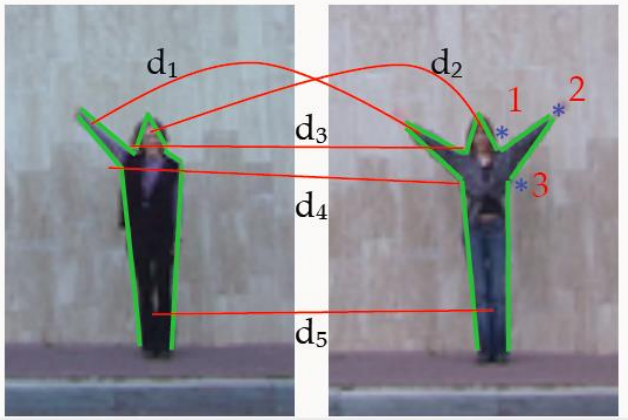
Pozu çizgi parçalarının dağılımı olarak göstermek

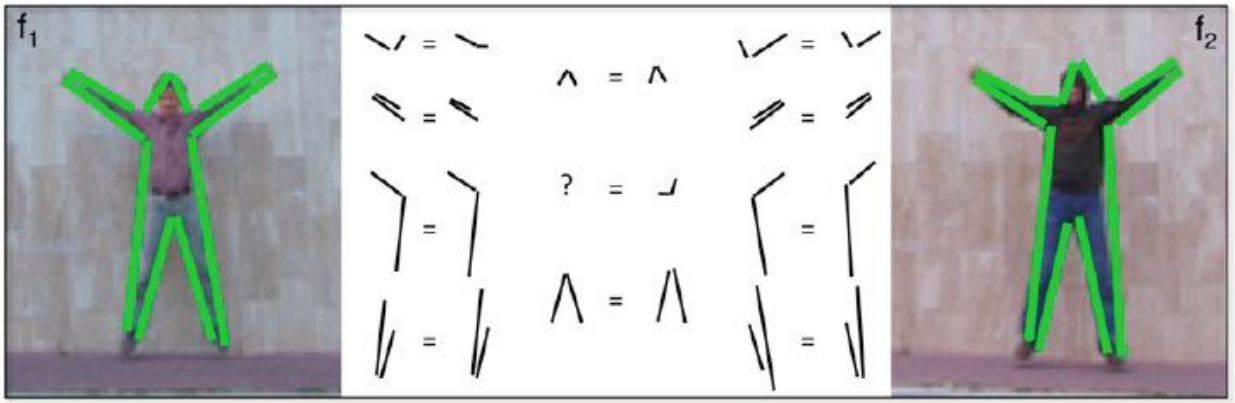


Baysal, S., Duygulu, P., ''A Line Based Pose Representation For Human Action Recognition'', Signal Processing: Image Communication, Volume 28, Issue 5, Pages 458-471, May 2013

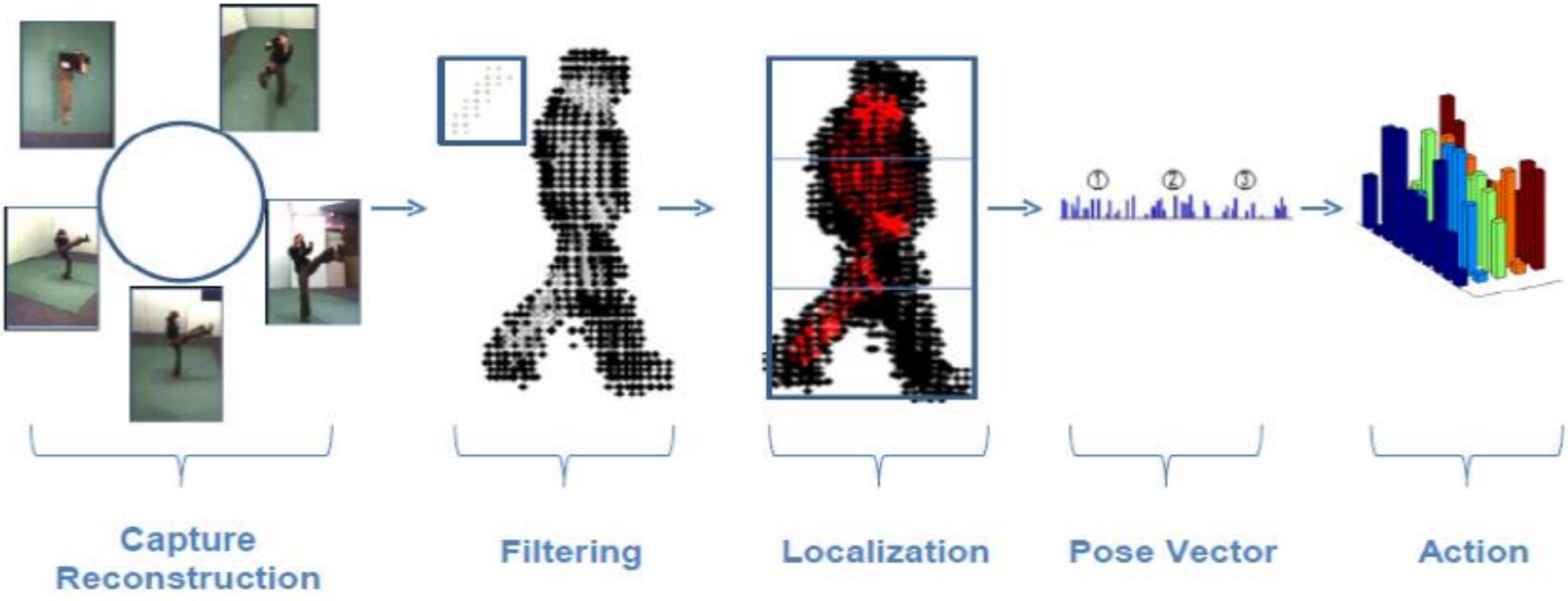
Çizgi eşleri



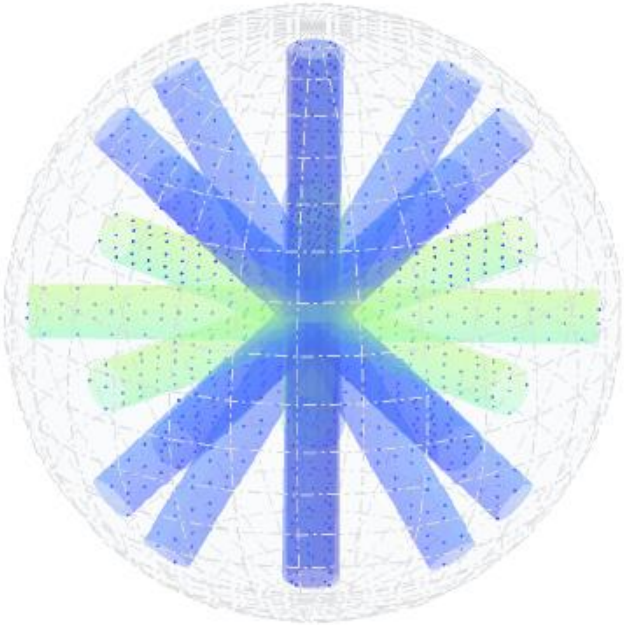


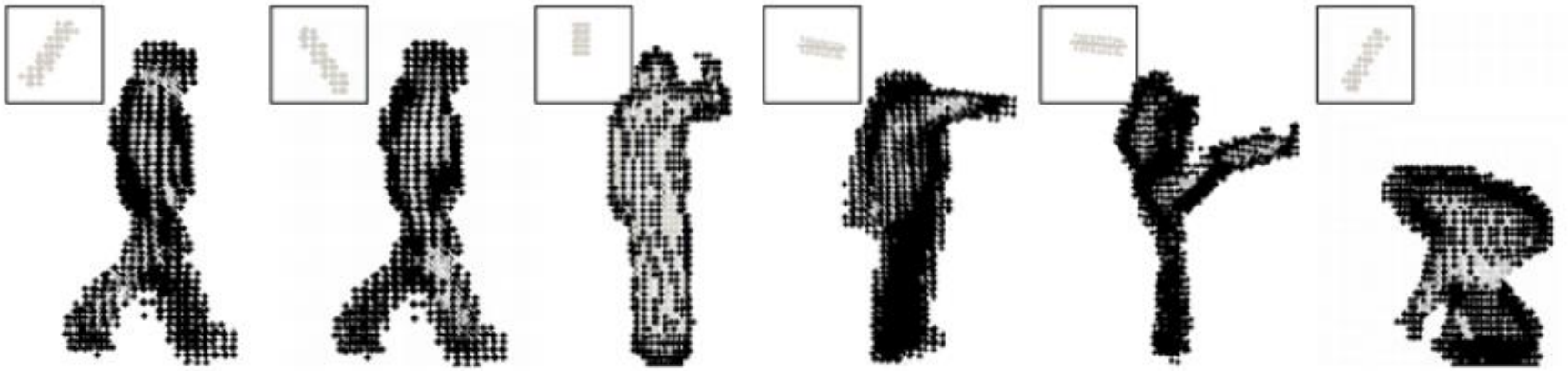


Birden fazla kamera görüntüsü



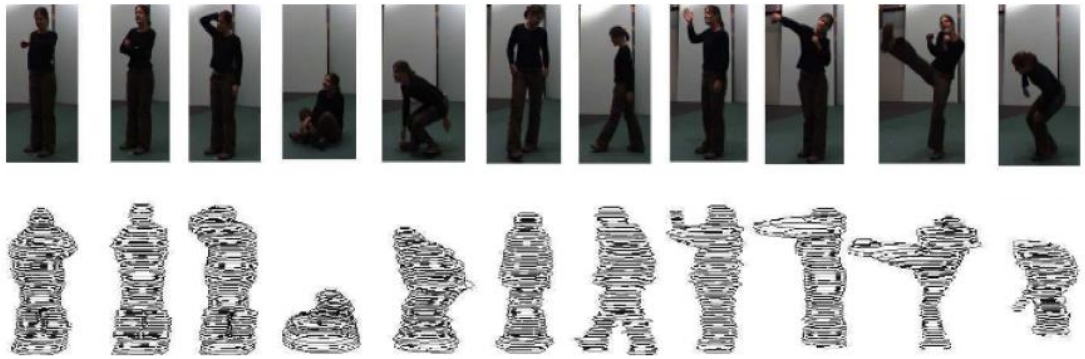
Silindir dağılımı

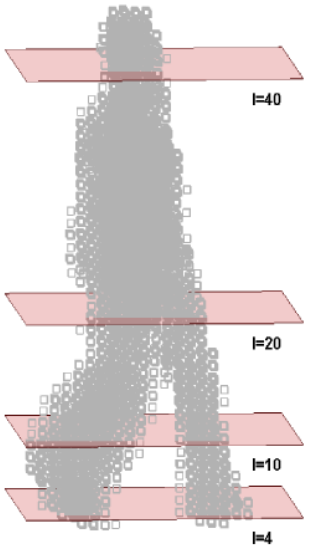


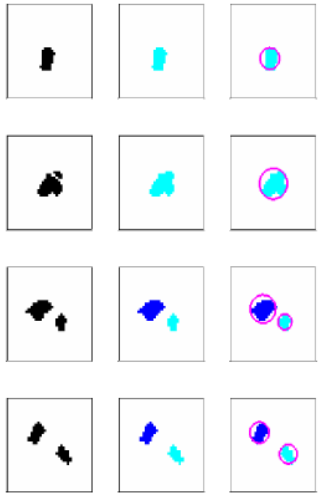


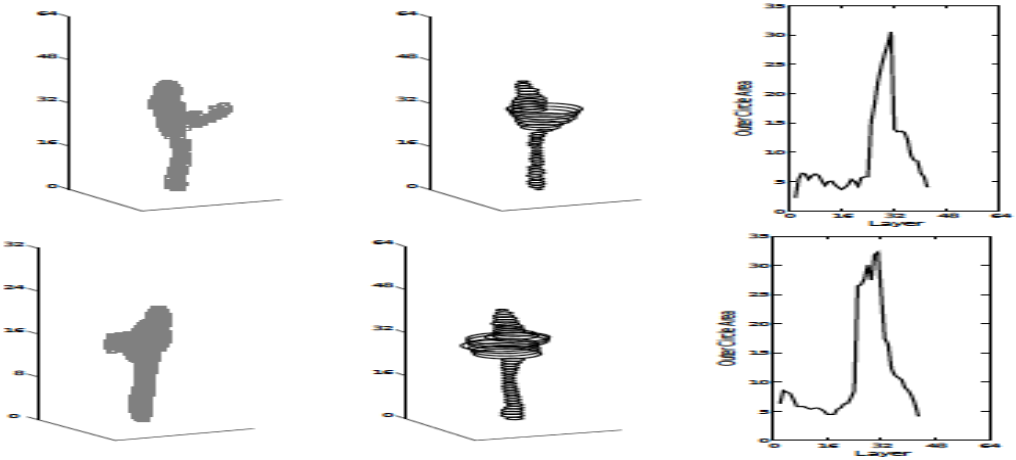
Pehlivan, S., . Duygulu, P. ”3D Human Pose Search using Oriented Cylinders”, IEEE Workshop on Search in 3D and Video (S3DV), in conjunction with ICCV 2009

Daire olarak izdüşüm







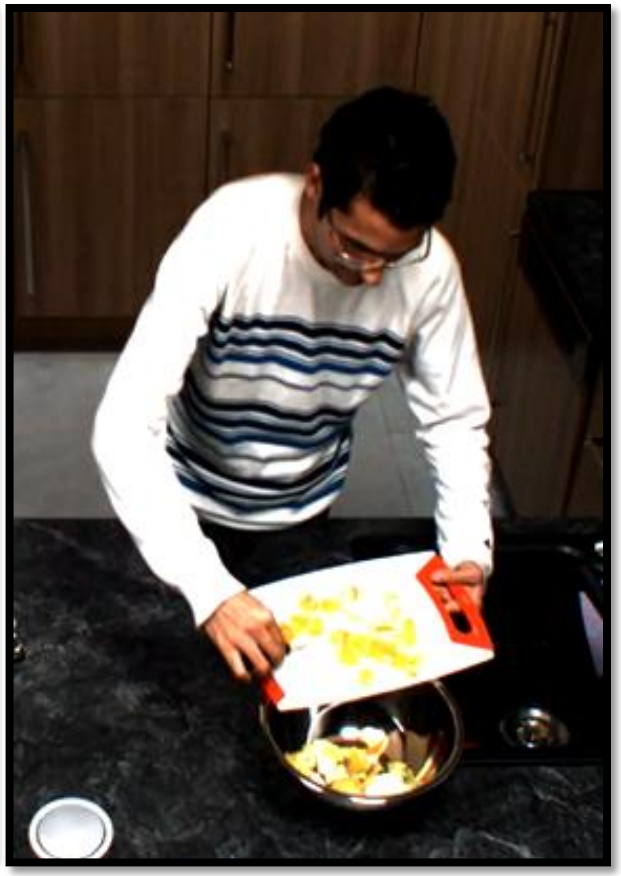


Pehlivan, S., Duygulu, P., ”A new pose-based representation for recognizing actions from multiple cameras”, Computer Vision and Image Understanding, volume 115, number 2, pages 140-151, February 2011

Sağlık uygulamaları

Pinar Duygulu, Eylul 2016, Ankara

Yemek yapma
Yüksek sınıf içi farklılık



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.

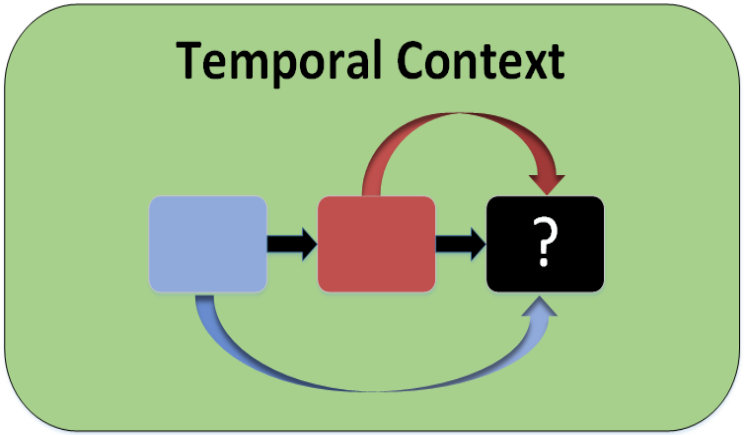
Düşük sınıflarası farklılık

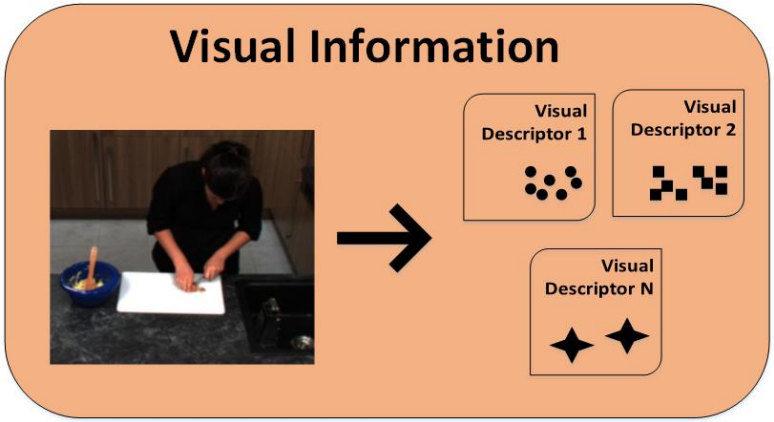
Cut apart, cut ends. cut slices, cut stripes, cut dice

Çözüm

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

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



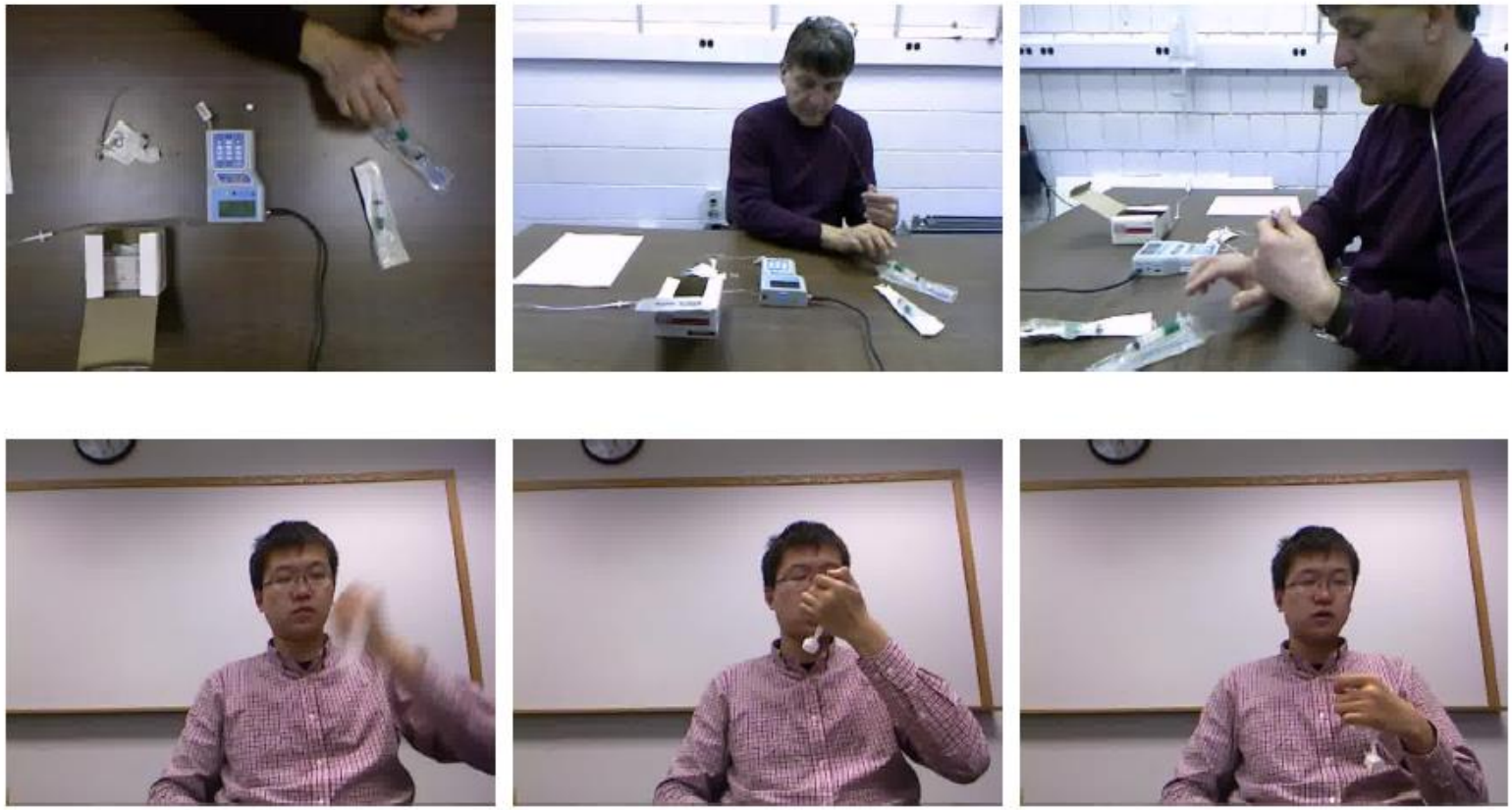


Tavaya koymak mı kaseye koymak mı?

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

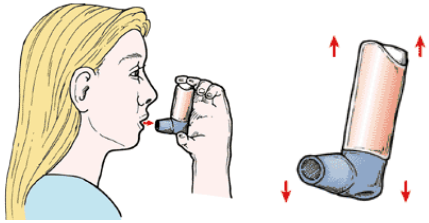
P(“put in pan” | “spread”) >
P(“put in bowl” | “spread”)

İlaç ve alet kullanımı

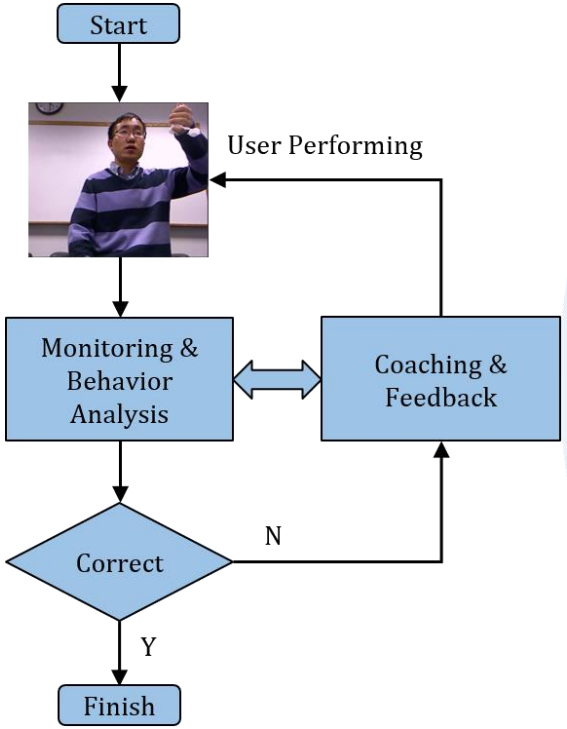


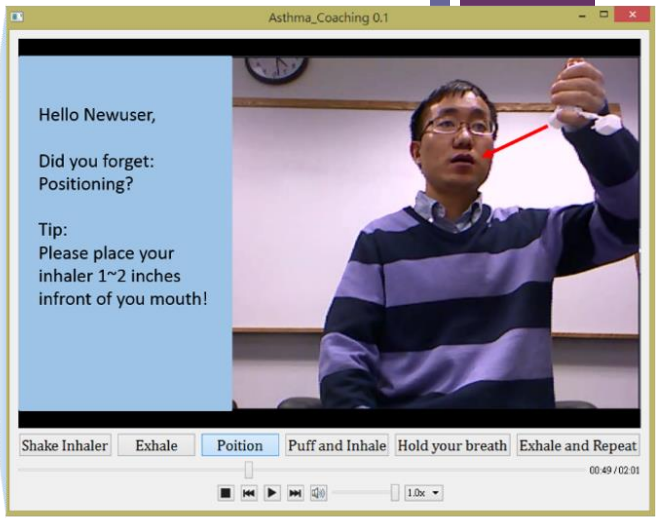
Iscen, A., Duygulu, P., ”Snippet Histograms for Assistive Technologies”, Workshop on Assistive computer Vision and Robotics, in conjunction with ECCV 2014.

Astım için inhaler kullanımı



- 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





User Interface

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

orsm.net