One-Shot-Similarity Metric Learning for Action Recognition

Orit Kliper-Gross, Tal Hassner, and Lior Wolf

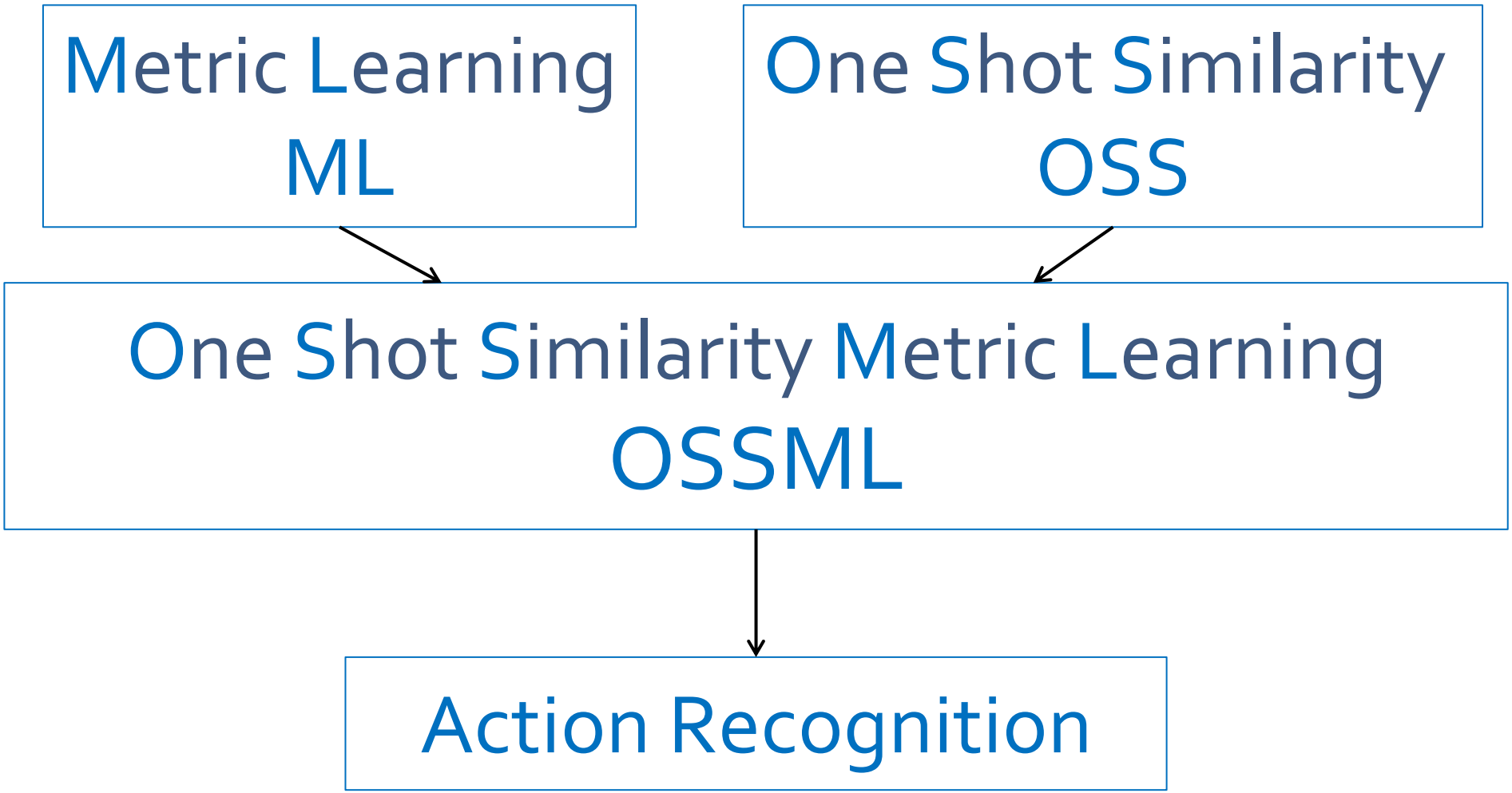
Weizmann Institute of Science

The Open University of Israel

Tel Aviv University

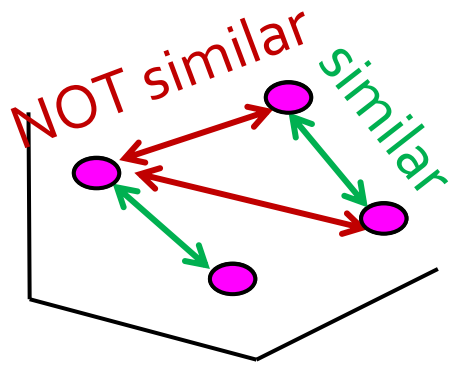
SIMBAD, Sept. 2011

Outline


Metric Learning
ML
One Shot Similarity
OSS
One Shot Similarity Metric Learning
OSSML
Action Recognition

Metric Learning

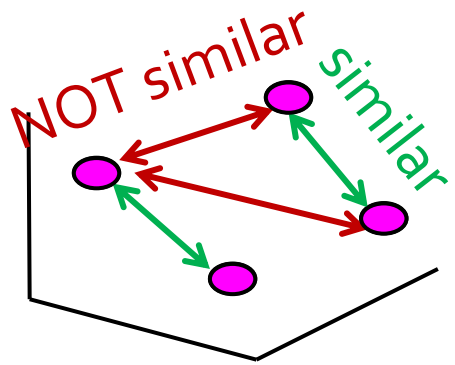
Original Space



- No Labels
- Similar/Not-similar constraints
- Distance metric
  - Not relevant
  - Not meaningful

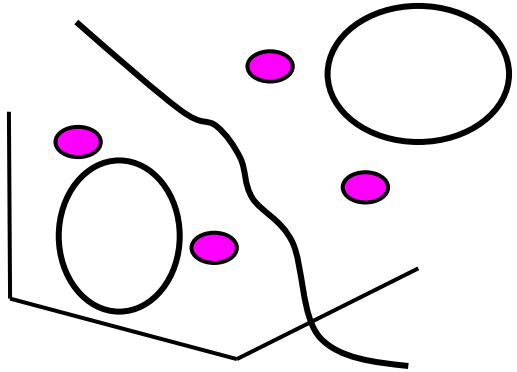
Metric Learning

Original Space



A=?

Projected Space



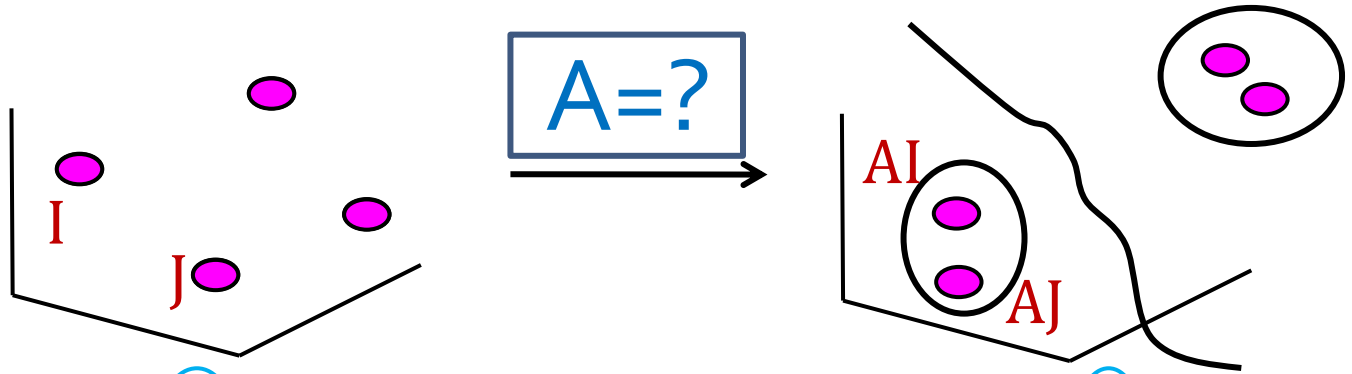
- No Labels
- Similar/Not-similar constraints
- Distance metric
  - Not relevant
  - Not meaningful

- Meaningful distance
- Consistent with constraints
- Easier to perform:
  - Classification
  - Clustering
  - Retrieval tasks
  - Pattern recognition
  - And more…

Metric Learning

Original Space

Projected Space



$$(I-J)^T M (I-J) \equiv (I-J)^T A^T A (I-J)$$

Most Common Approach: *Euclidian Distance*

RCA Bar-Hillel et al. ICML'03

SDP Xing el al. NIPS'03

NCA Goldberger et al. NIPS'05

LMNN Weinberger et al. NIPS'06

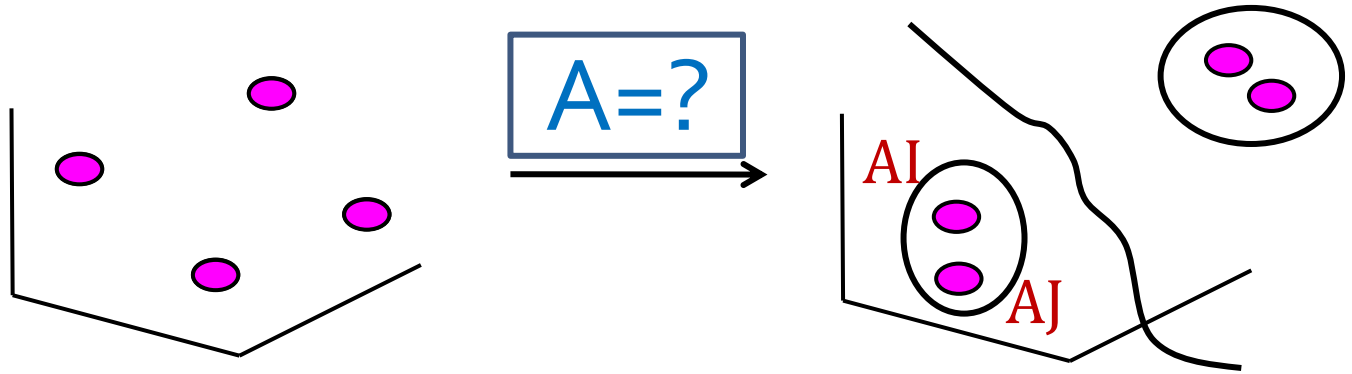
ITML Davis et al. ICML'07

OASIS Chechick et al. JMLR'10 and More…

Metric Learning

Original Space

Projected Space



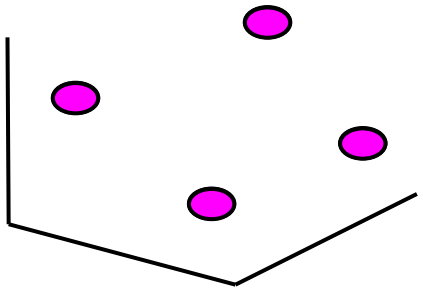
$$CSML \equiv \frac{(AI)^T (AJ)}{\|AI\| \|AJ\|}$$

- *Euclidian Distance*

Recently (Nguyen &Bai'10): *Cosine Similarity*

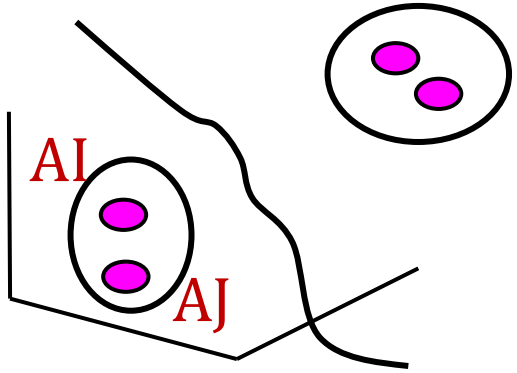
Metric Learning

Original Space



A=?

Projected Space



$OSSML \equiv ???$

- *Euclidian Distance*
- *Cosine Similarity*
- *One-Shot-Similarity*

In This Work:
One-Shot-Similarity Metric Learning (OSSML)

One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

- What:

  A measure of the similarity between two vectors

- Input:

  The two vectors ***I,J***

  A set of "Background samples" ***N***

- How:

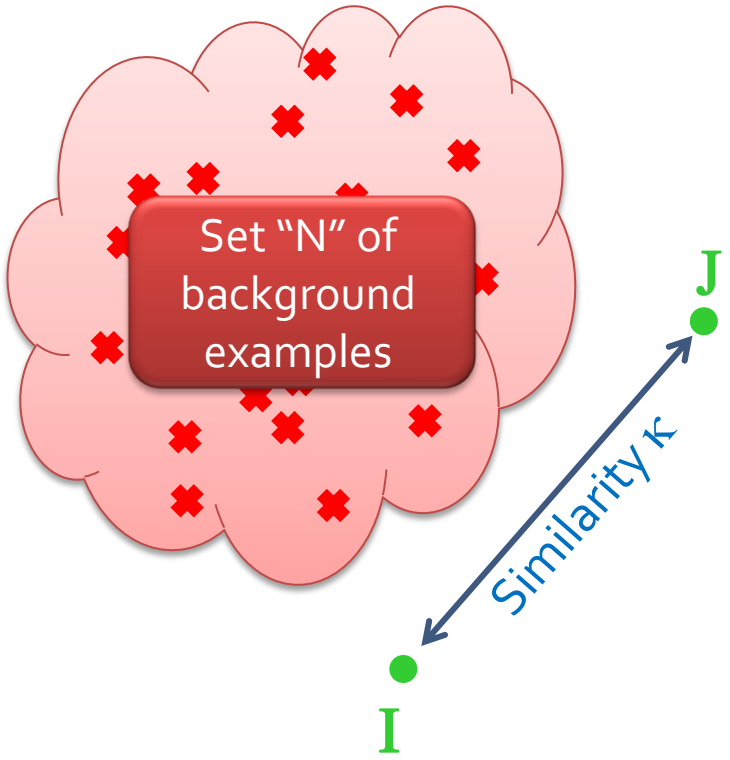
  Use "One-Shot Learning"

  (i.e. classification with one positive example)

One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

OSS Score:



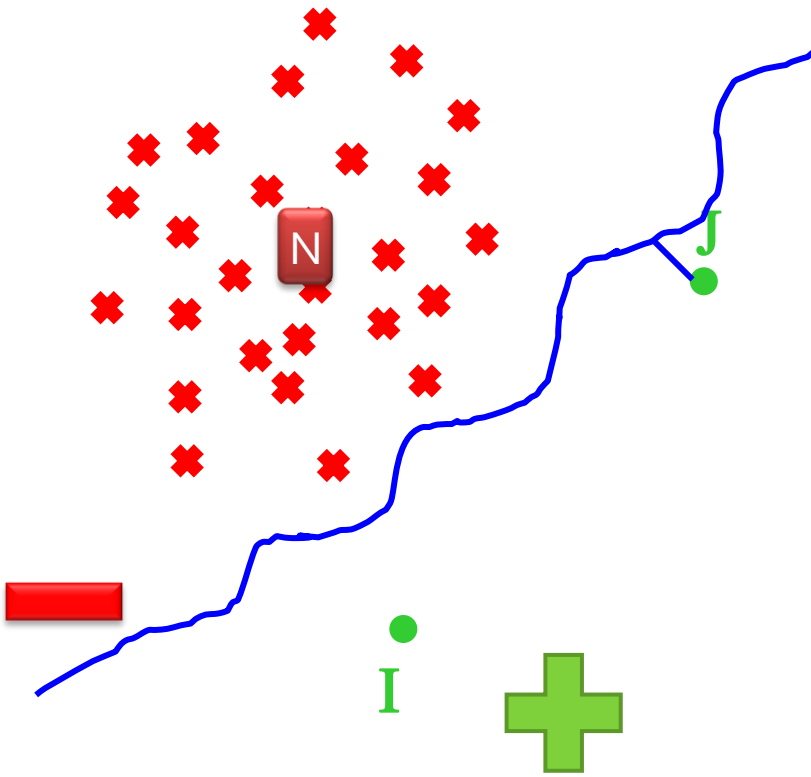
One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

OSS Score:

Model1 = train (I, N)

Score1 = classify(J, Model1)



One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

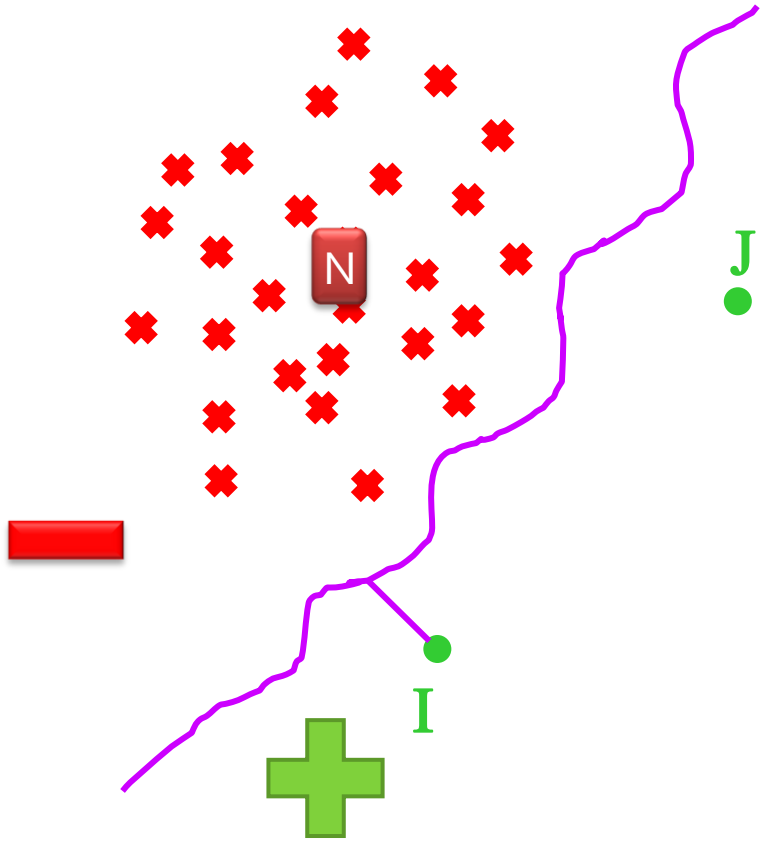
OSS Score:

Model1 = train (I, N)

Score1 = classify(J, Model1)

Model2 = train (J, N)

Score2 = classify(I, Model1)



One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

OSS Score:

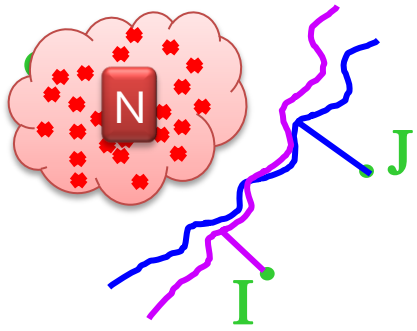
Model1 = train (I, N)

Score1 = classify(J, Model1)

Model2 = train (J, N)

Score2 = classify(I, Model1)

OSS(I,J,N) = (score1 + score2) /2



One-Shot-Similarity (OSS)

Wolf , Hassner, Taigman, ICCV'09A

Properties:

Meta-Similarity

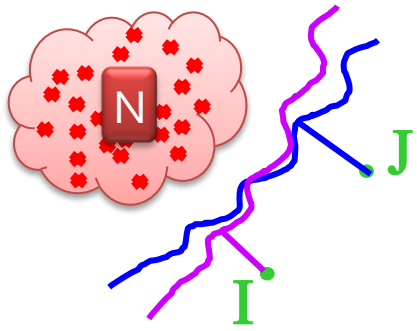
Can use any underlying classifier

Unlabeled training data

Do not know class labels

Efficiently computed

Statistical properties of set N computed once for all pairs I,J



One-Shot-Similarity (OSS)

Face Recognition ☺

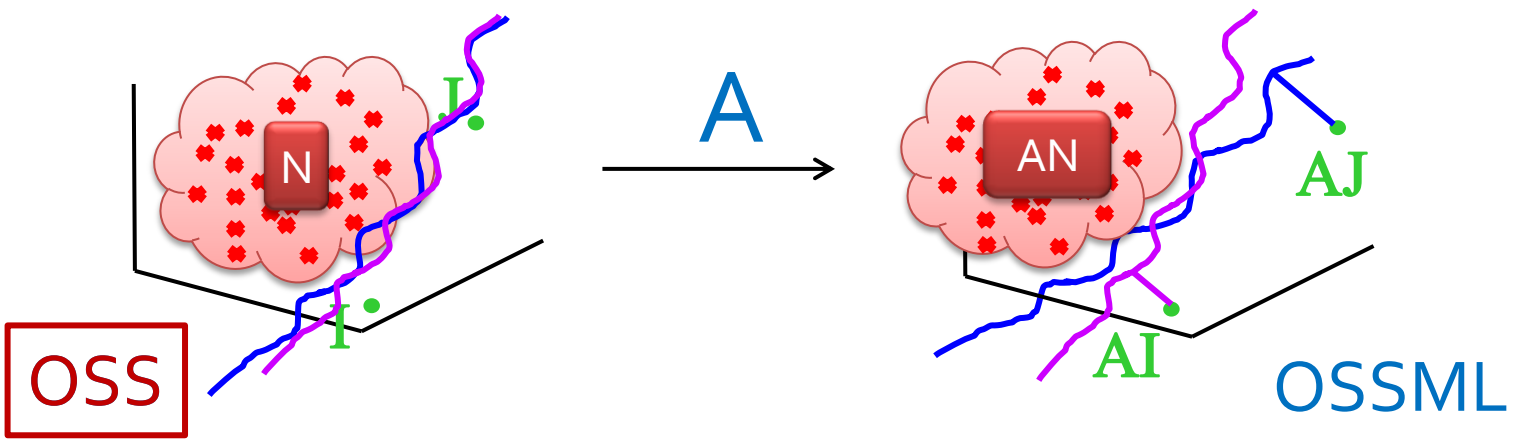
Wolf et-al'09

Documents Analysis ☺

Wolf et-al'10

Action Recognition ☹

One-Shot-Similarity Metric Learning



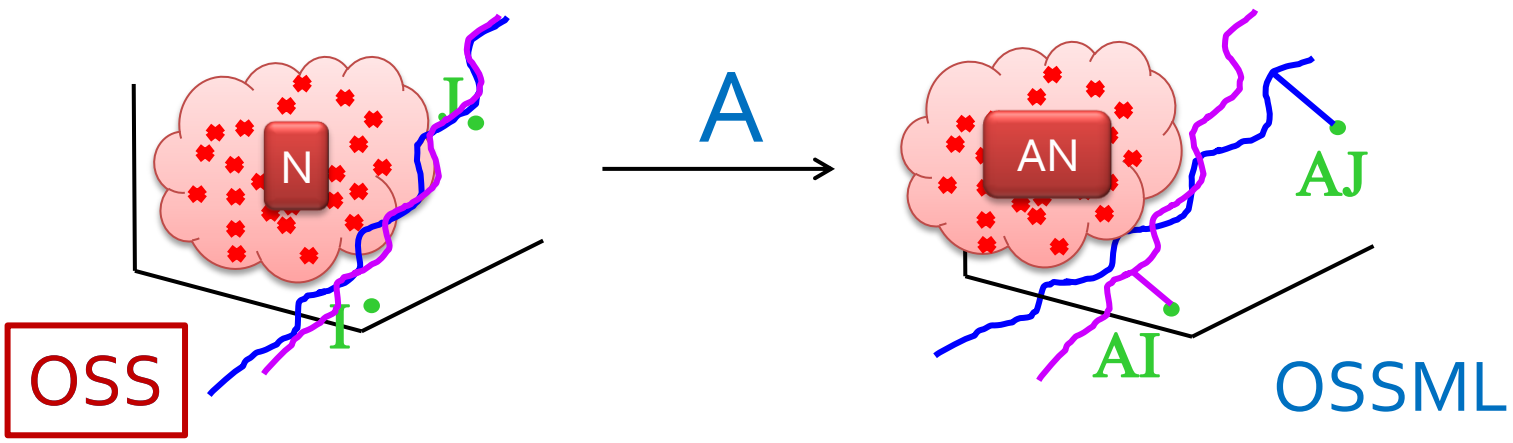
OSS Using *Free-Scale Linear Discriminant Analysis (FS-LDA)*

Variance between classes

LDA separator: $S = \frac{\sigma_b^2}{\sigma_w^2}$

Variance within classes

One-Shot-Similarity Metric Learning



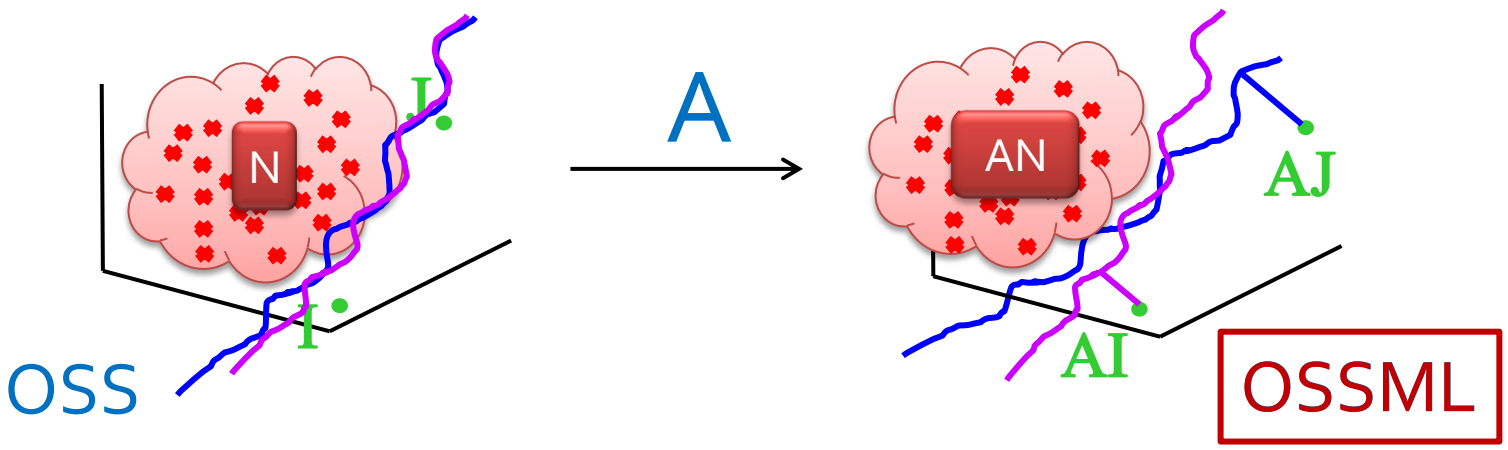
OSS Using *Free-Scale Linear Discriminant Analysis (FS-LDA)*

$$OSS_{FS-LDA}(I,J,N) = (I-\mu_N)^T S_N^{-1}\left(J-\frac{I+\mu_N}{2}\right) + (J-\mu_N)^T S_N^{-1}\left(I-\frac{J+\mu_N}{2}\right)$$

μ_N - mean of the set N

s_N^{-1} - inverse of set N covariance matrix

One-Shot-Similarity Metric Learning

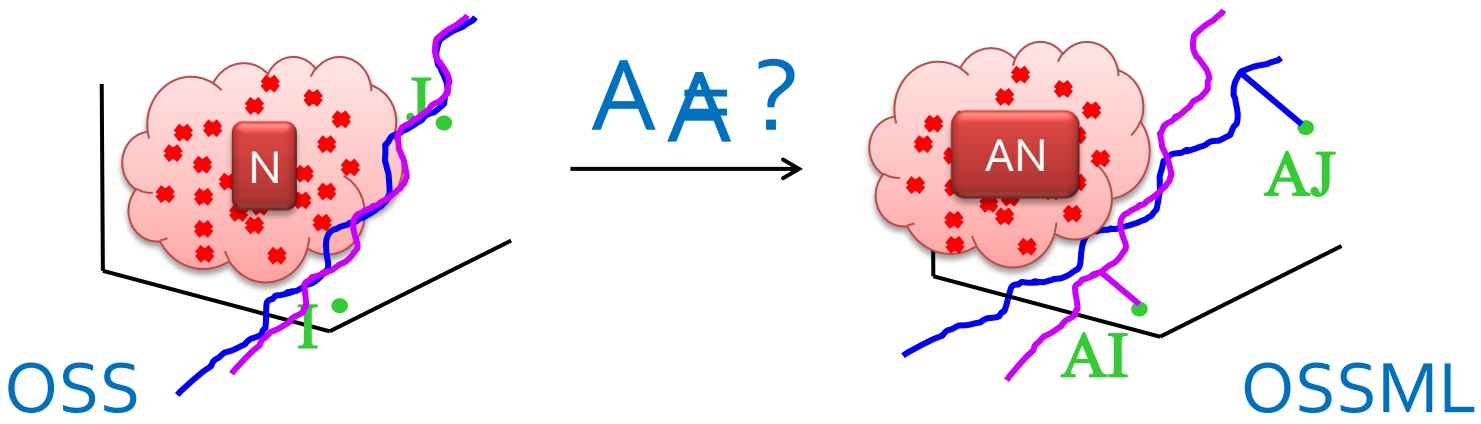


OSSML is simply OSS in the projected space

$$OSS_{ML}(I, J, N, A) = OSS(AI, AJ, AN)$$

$$= (\mathrm{AI} - \mu_{\mathrm{AN}})^T S_{\mathrm{AN}}^{-1} \left(\mathrm{AJ} - \frac{\mathrm{AI} + \mu_{\mathrm{AN}}}{2}\right) + (\mathrm{AJ} - \mu_{\mathrm{AN}})^T S_{\mathrm{AN}}^{-1} \left(\mathrm{AI} - \frac{\mathrm{AJ} + \mu_{\mathrm{AN}}}{2}\right)$$

One-Shot-Similarity Metric Learning



OSSML – Objective Function

Goal: Find ***A*** to optimize:

$$f(A) = \sum_{I,J \in same} OSS_{ML}(I,J,N,A) - \alpha \sum_{I,J \in not-same} OSS_{ML}(I,J,N,A) - \beta \|A - A_0\|^2$$

Balance

Tradeoff

Large separation between *Same* and *Not-same* pairs

Regularization

Gradient has Closed-Form Solution!

We perform the optimization using *gradient descent*

See our paper for more details

OSSML for Action Recognition

WEIZMANN INSTITUTE OF SCIENCE

THE OPEN UNIVERSITY OF ISRAEL

TEL AVIV UNIVERSITY

ASLAN* The Action Similarity Labeling Challenge

- New action recognition database
- > 400 complex action classes
- Web videos: unconstrained, realistic, uncontrolled setting
- Action pair matching

* The Action Similarity Labeling Challenge, Kliper-Gross, Hassner , Wolf, To Appear TPAMI'11

Action pair matching

Input: Video pair

Same / Not-Same action ?

Challenges

Large variability



Class Resolutions

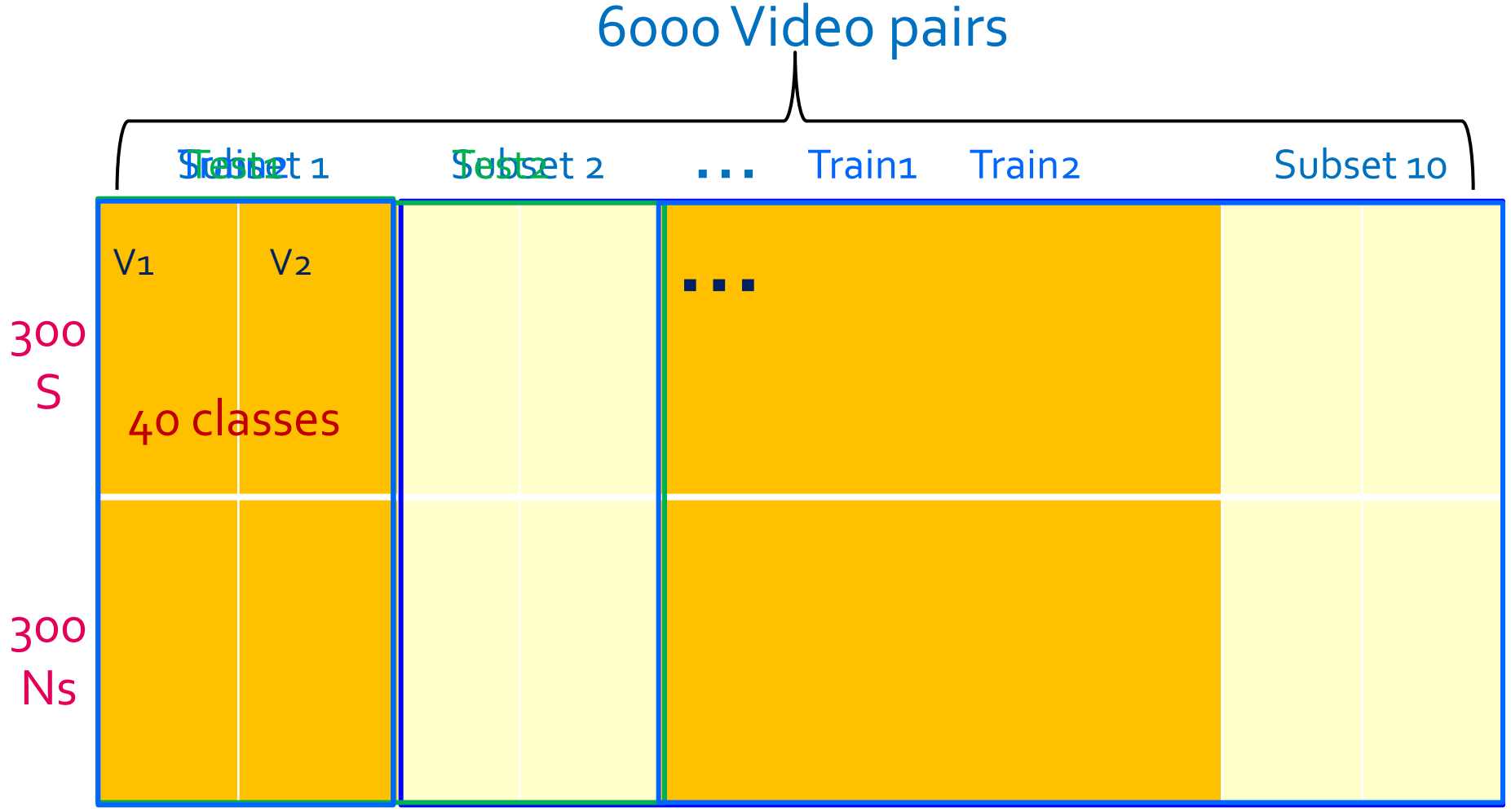

Action Ambiguity



Others

High-dimensional, complicated space

The ASLAN benchmark - Setting



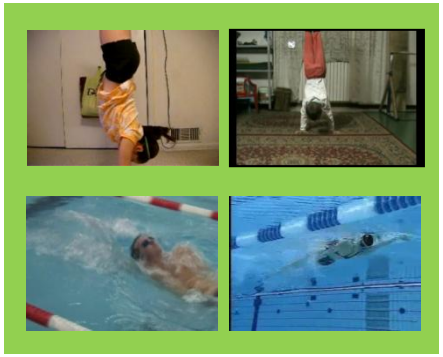
Mutually exclusive subsets

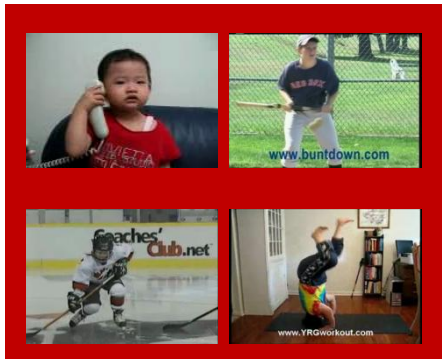
10-fold cross-validation scheme

Baseline Tests

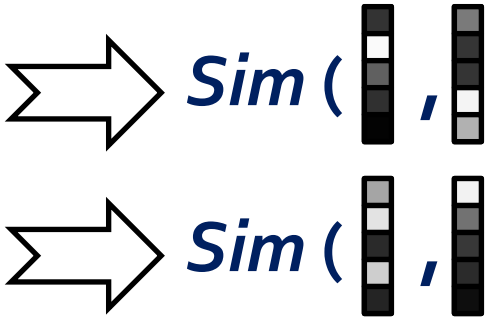
ASLAN. To appear TPAMI'11

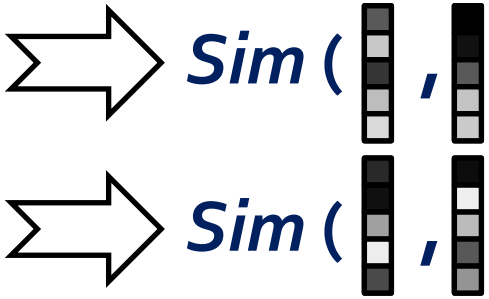
same





not same





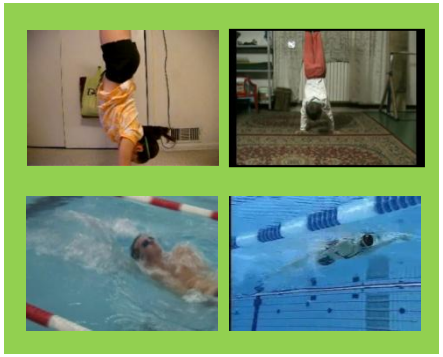
Classifier (e.g. SVM)

Threshold

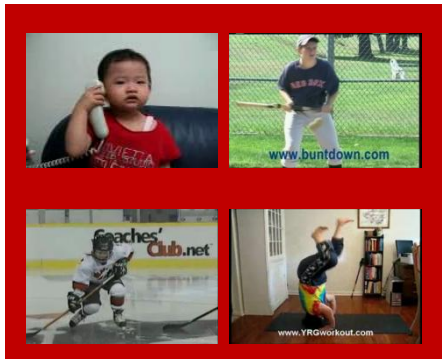
Baseline Tests

ASLAN. To appear TPAMI'11

same



not same



$(\kappa_{1,1}, \kappa_{1,2}, \ldots, \kappa_{1,n})$

$(\kappa_{2,1}, \kappa_{2,2}, \ldots, \kappa_{2,n})$

$(\kappa_{i,1}, \kappa_{i,2}, \ldots, \kappa_{i,n})$

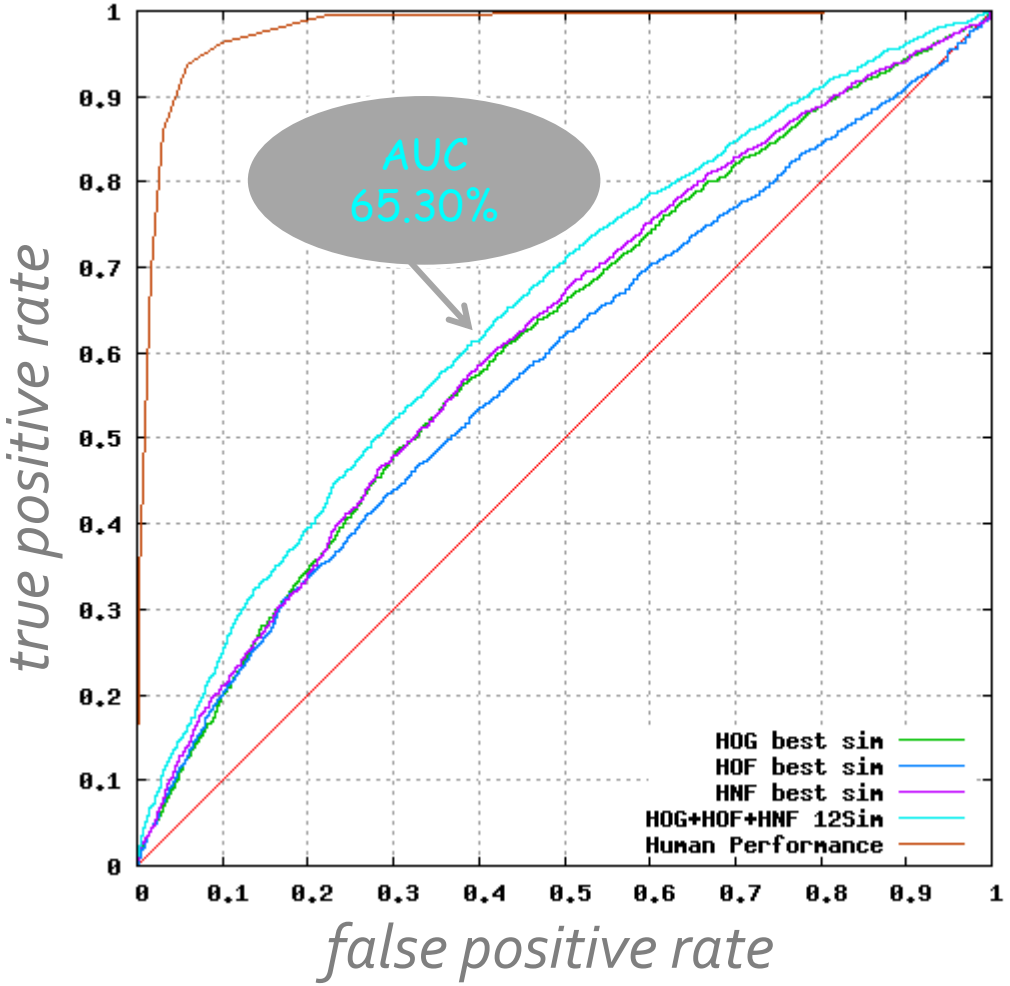
$(\kappa_{i+1,1}, \kappa_{i+1,2}, \ldots, \kappa_{i+1,n})$

Classifier (e.g. SVM)

multiple descriptors \ similarities

Baseline Results on the ASLAN set

ASLAN. To appear TPAMI'11



HOG best: 63.2%

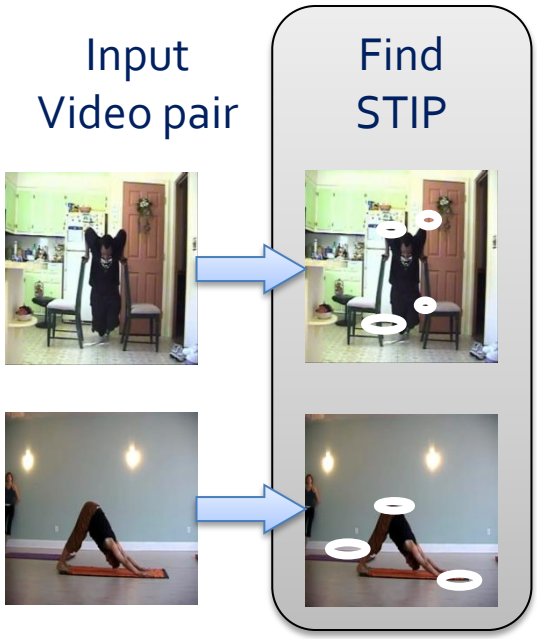
HOF best: 58.9%

HNF best: 63.3%

Human: 97.8%

Multiple descriptors:
AUC: 65.3%
Accuracy: 60.9 ± .77

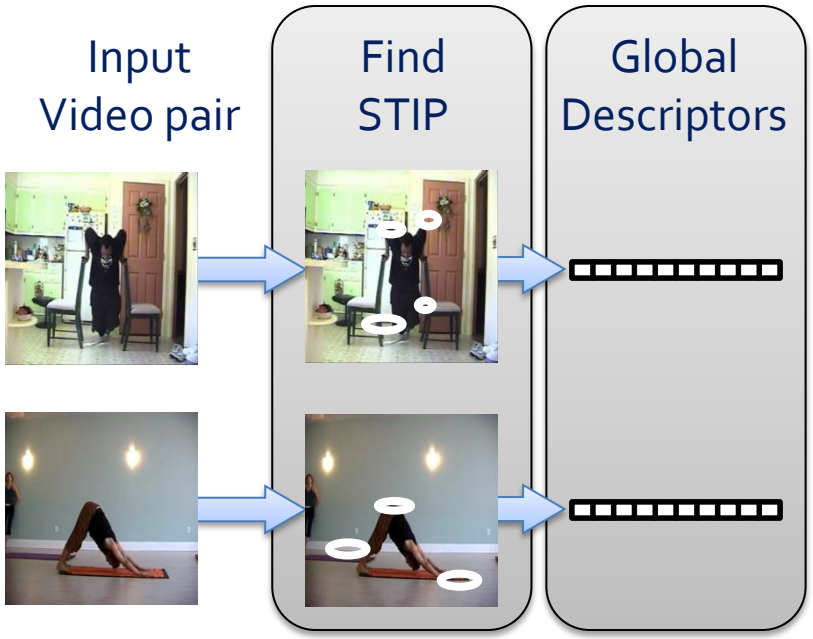
Pipeline



STIP: On Space-Time Interest Points
Laptev IJCV'05

Software available on:
http://www.irisa.fr/vista/Equipe/People/Laptev/download.html#stip

Pipeline



Using:

- HOG, HOF and HNF

With Bag of Words (BOW)

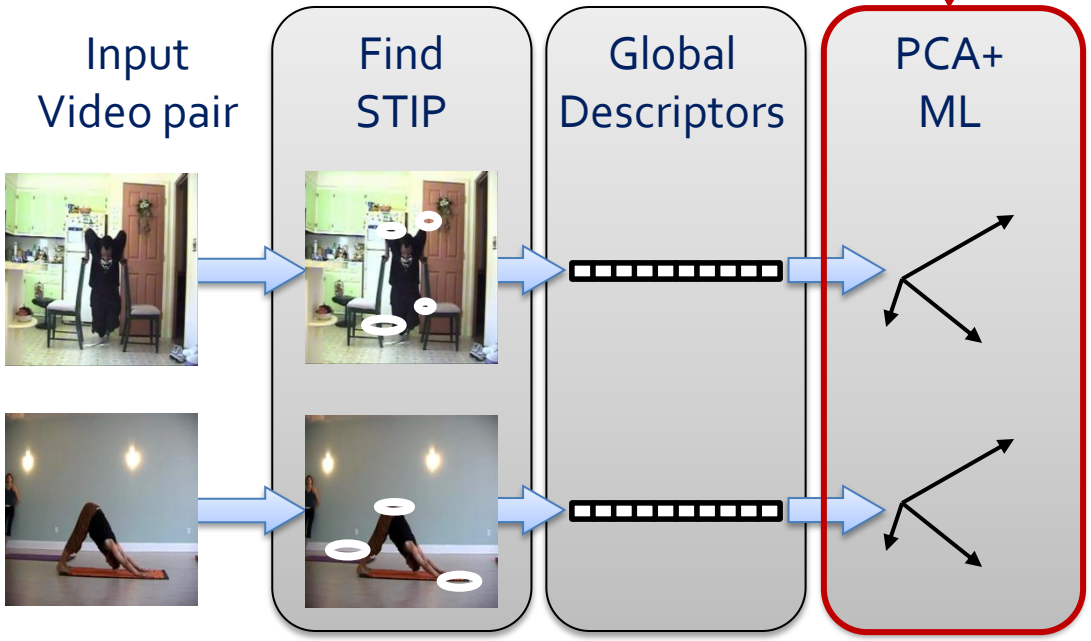
Learning realistic human actions from movies

Laptev et.al. CVPR'08

Descriptor Size: 5000 features

Pipeline

This Work



ML types:

- **CSML**

[Nguyen & Bai'10]

- **OSSML**

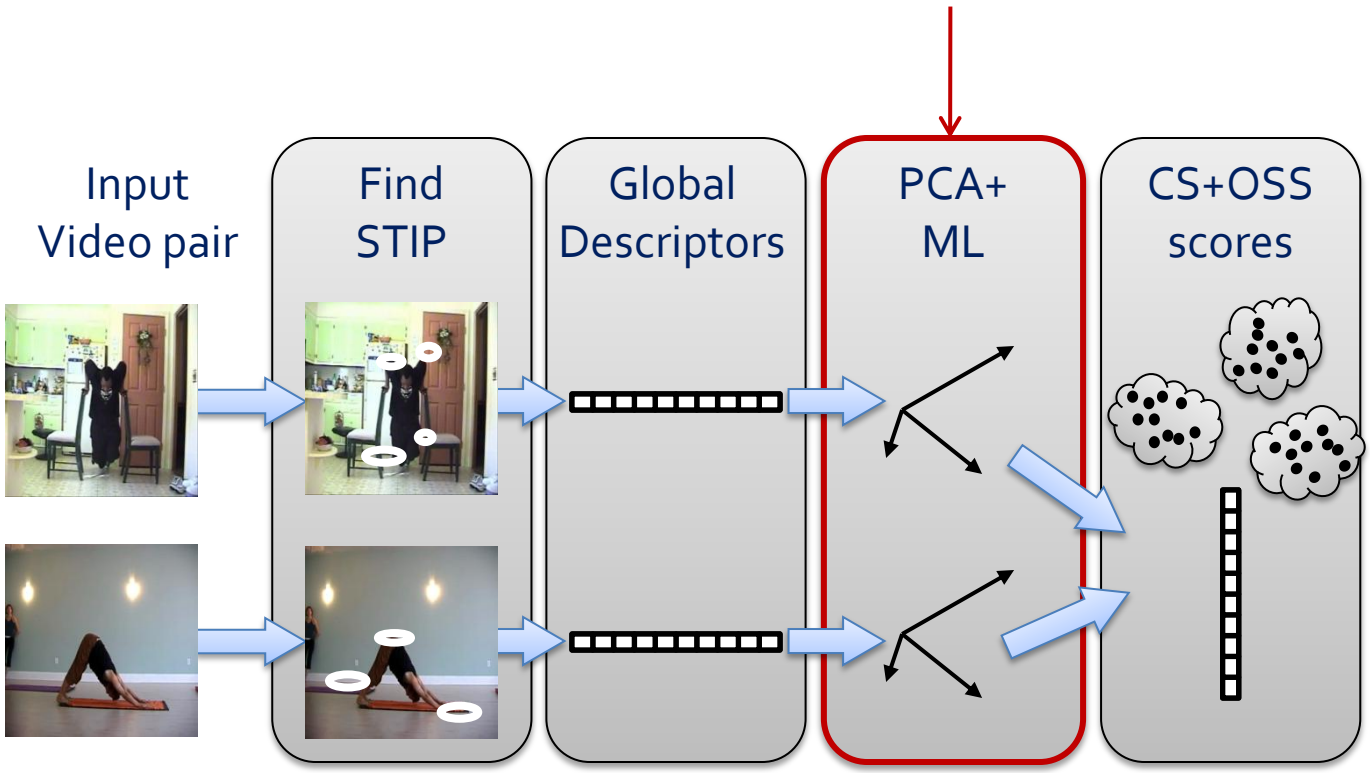
[This work – SIMBAD'11]

and

Dimensionality Reduction: 5000 → 100 → 50

Pipeline

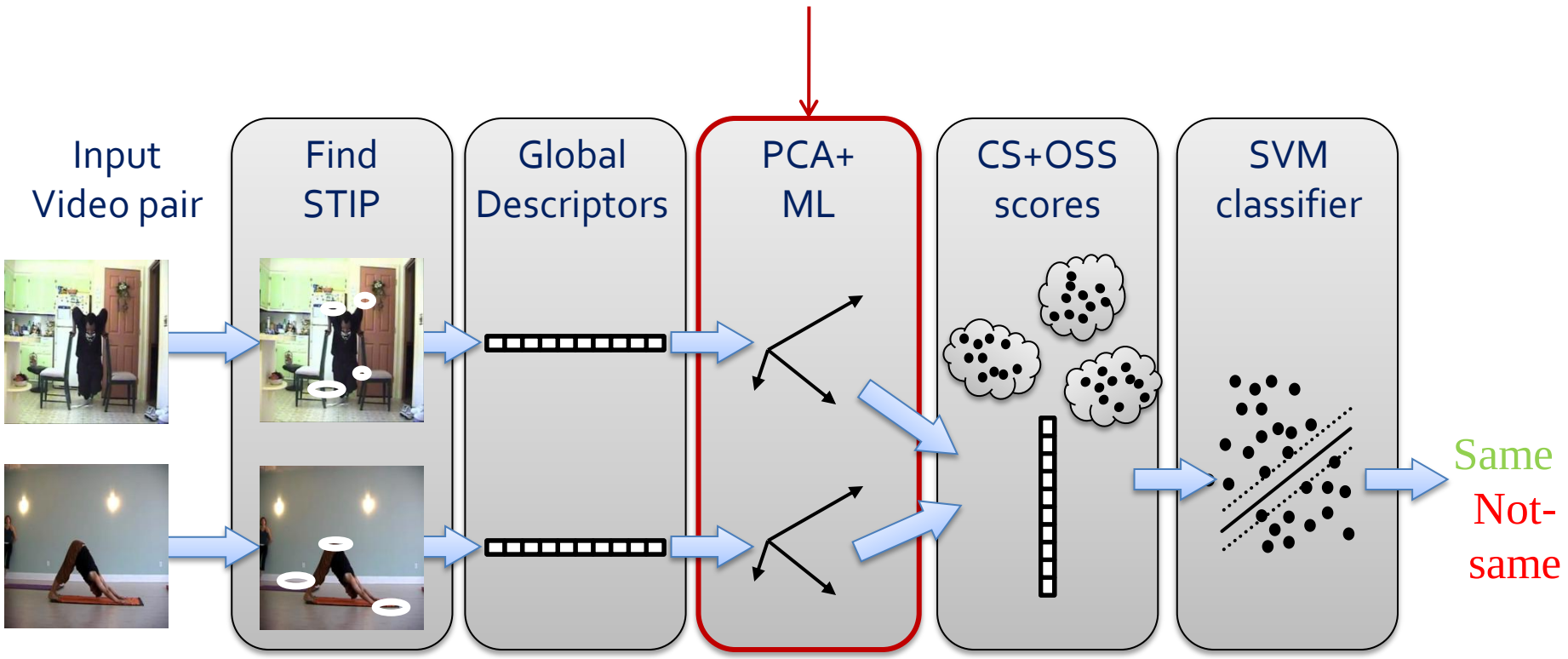
This Work



+ Multiple Similarities
+ Multiple Descriptors

Pipeline

This Work



OSSML Results

Best Baseline Results:

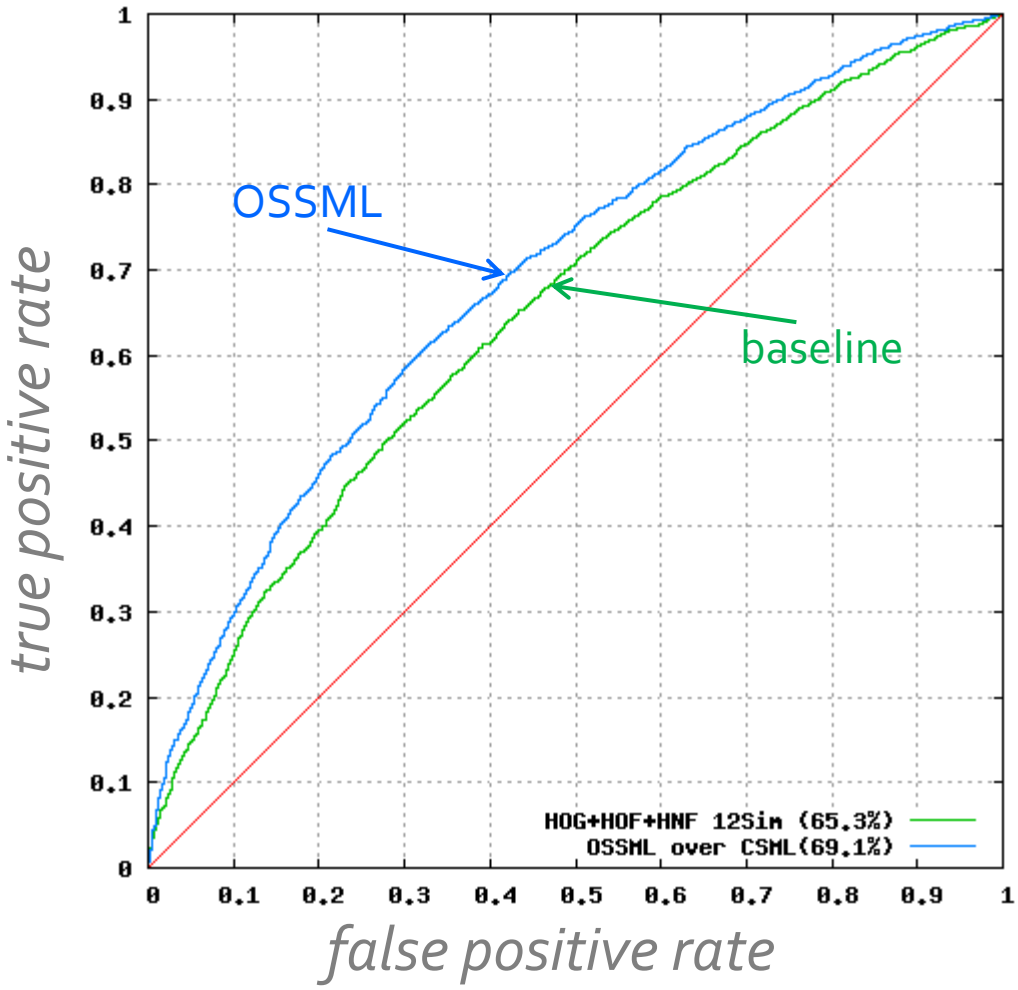
AUC: 65.30%

Accuracy: 60.88 ± .77

Best OSSML Results:

AUC: 69.10%

Accuracy: 64.25 ± .7



State-of-the-art results on the ASLAN benchmark

OSSML Results (AUC)

		HOG	HOF	HNF	ALL Descriptors
PCA initialization	cs	63.9	60.1	64.2	65.2
	oss	63.1	59.4	63.0	64.9
	cs+oss				65.4
CSML after PCA	cs	64.2	61.8	64.3	67.4
	oss	63.8	62.4	63.3	67.6
	cs+oss				68.0
OSSML after PCA	cs	64.1	60.5	64.3	65.7
	oss	63.8	60.7	64.0	66.7
	cs+oss				66.6
OSSML after CSML	cs	65.0	63.6	65.1	68.0
	oss	64.3	63.8	64.1	68.9
	cs+oss				69.1

38% above a random classifier

Same classified as Same (TP) ☺

Not-Same classified as Not-Same (TN) ☺

Same classified as Not-Same (FN) ☹

Not-Same classified as Same (FP) ☹

Summary

- OSSML
  - A metric learning method geared towards improved One-Shot-Similarity (OSS) performance
  - Formulate a cost function using same/not-same labels
  - Gradient descent solution

- Application to Action Recognition
  - Action pair-matching
  - Best reported results on the ASLAN set

Thank You !

SIMBAD, Sept. 2011