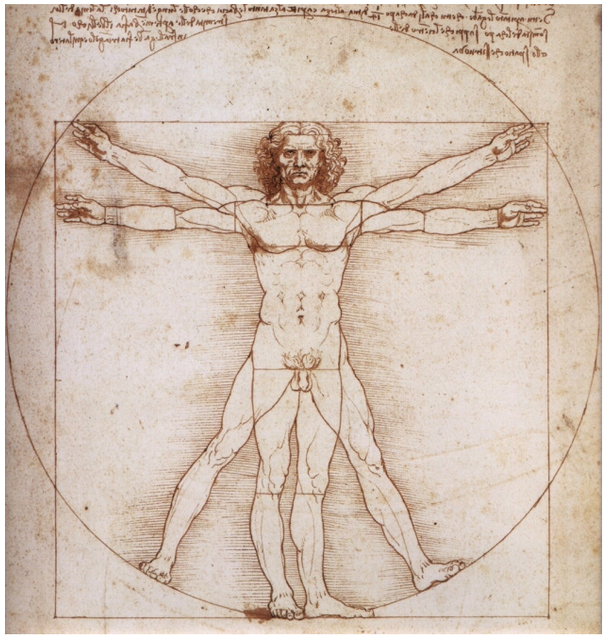
# Fusion of Single View Soft kNN classifiers for Multicamera human action recognition

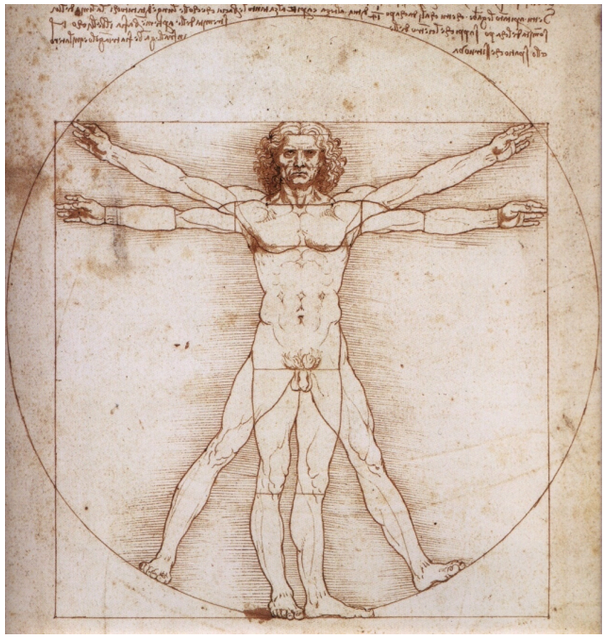
Rodrigo Cilla, Miguel A. Patricio,
Antonio Berlanga, José M. Molina

HAIS 2010

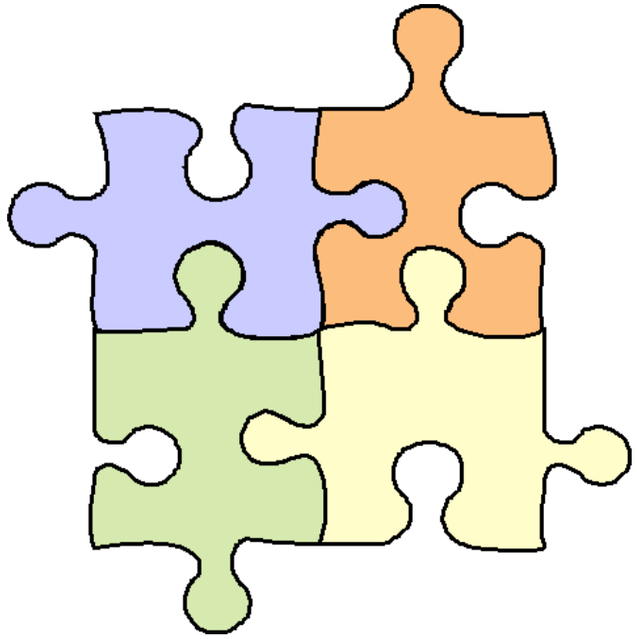
Information Fusion: Frameworks and Architectures



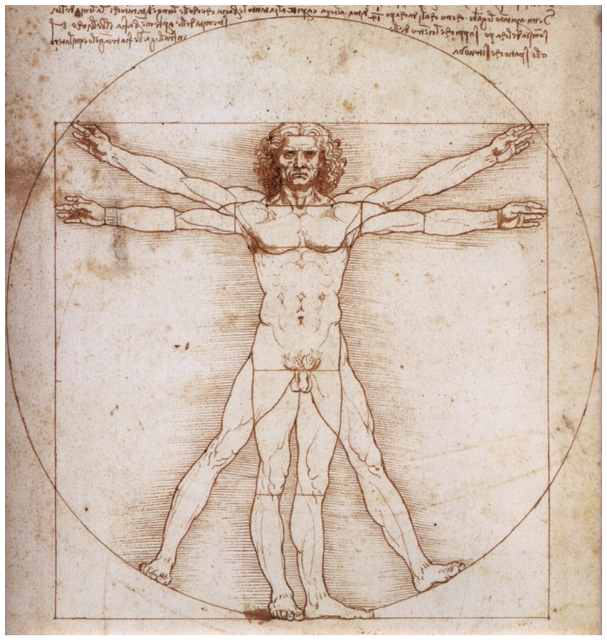
# I. Introduction



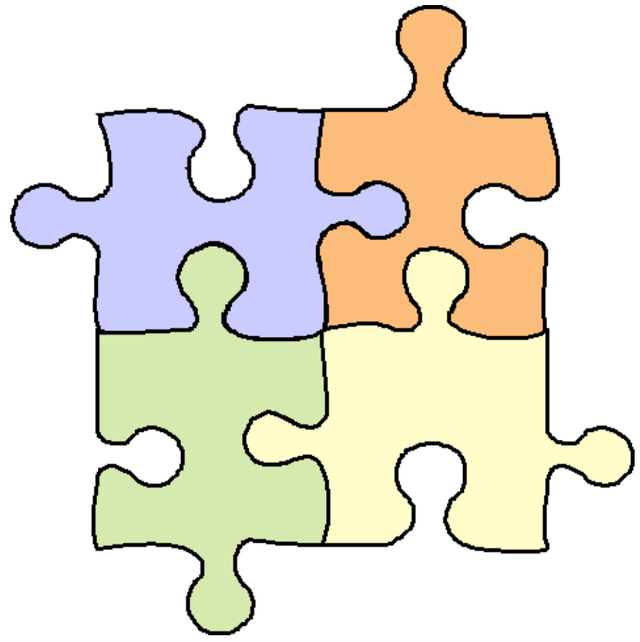
# I. Introduction



# II. Architecture

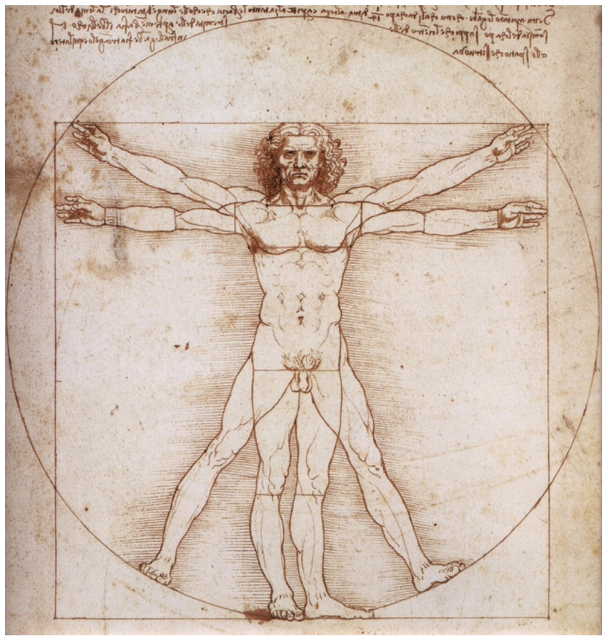


I. Introduction

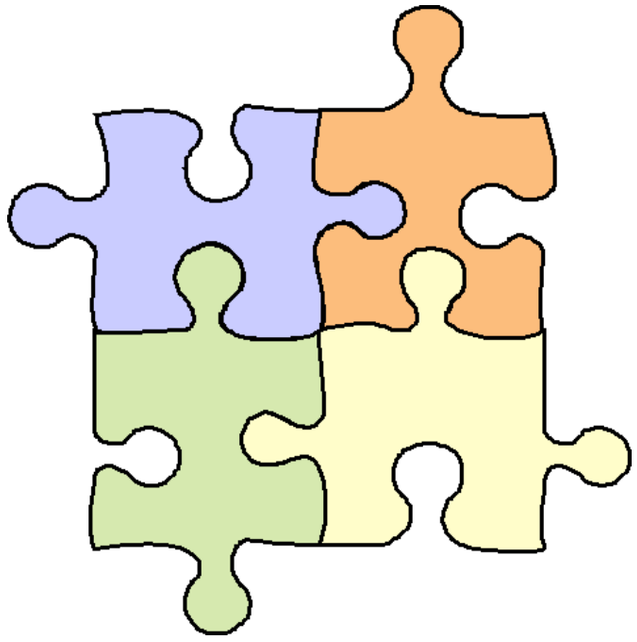


II. Architecture

III. Classifier Fusion



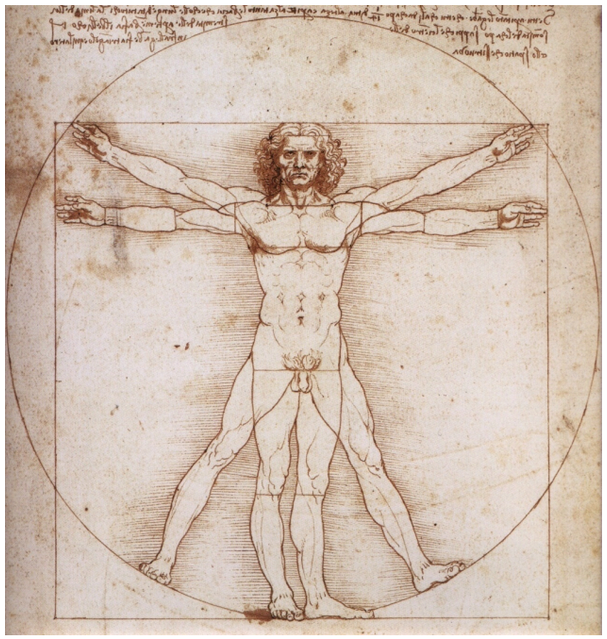
I. Introduction



II. Architecture

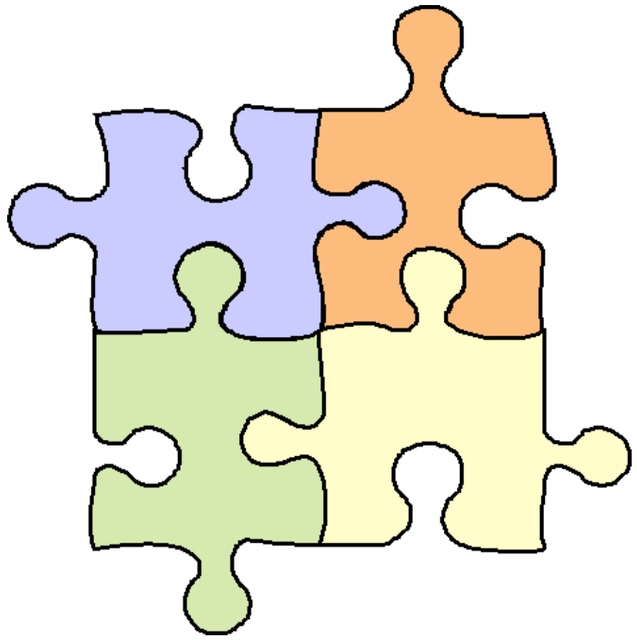
III. Classifier Fusion

IV. Experiments

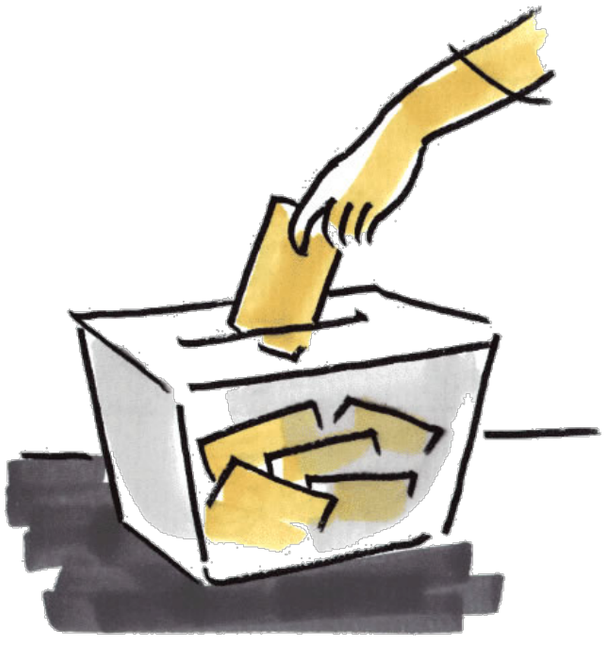


I. Introduction

V. Conclusions

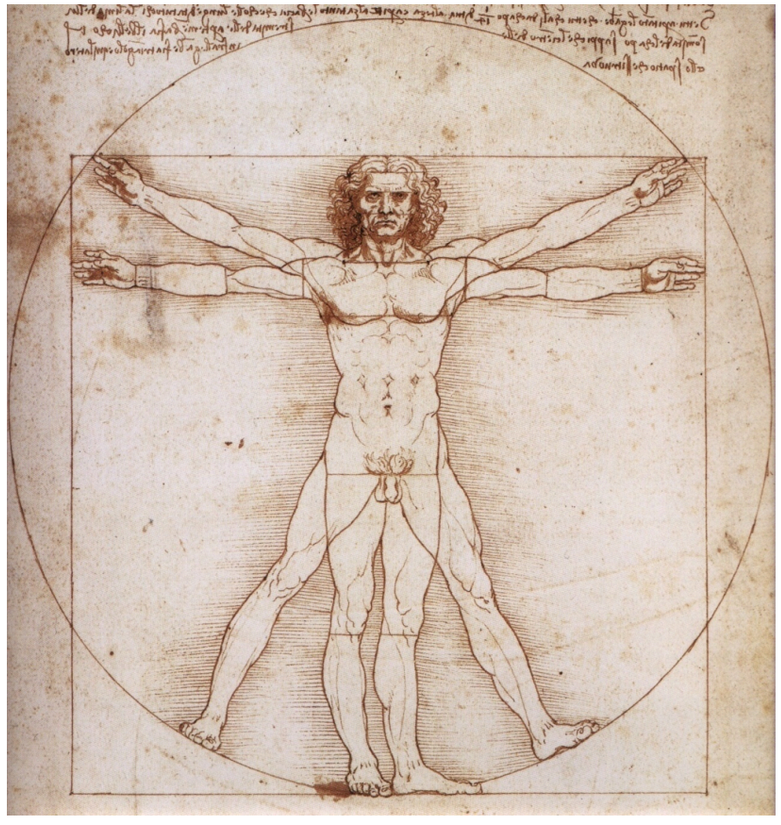


II. Architecture



III. Classifier Fusion

IV. Experiments



# I. Introduction

Understand what Humans do

# Understand what Humans do

Surveillance

Video Annotation

# Understand what Humans do

Surveillance

Video Annotation

Understand what Humans do

Surveillance

Entertainment

Smart Environments

Video Annotation

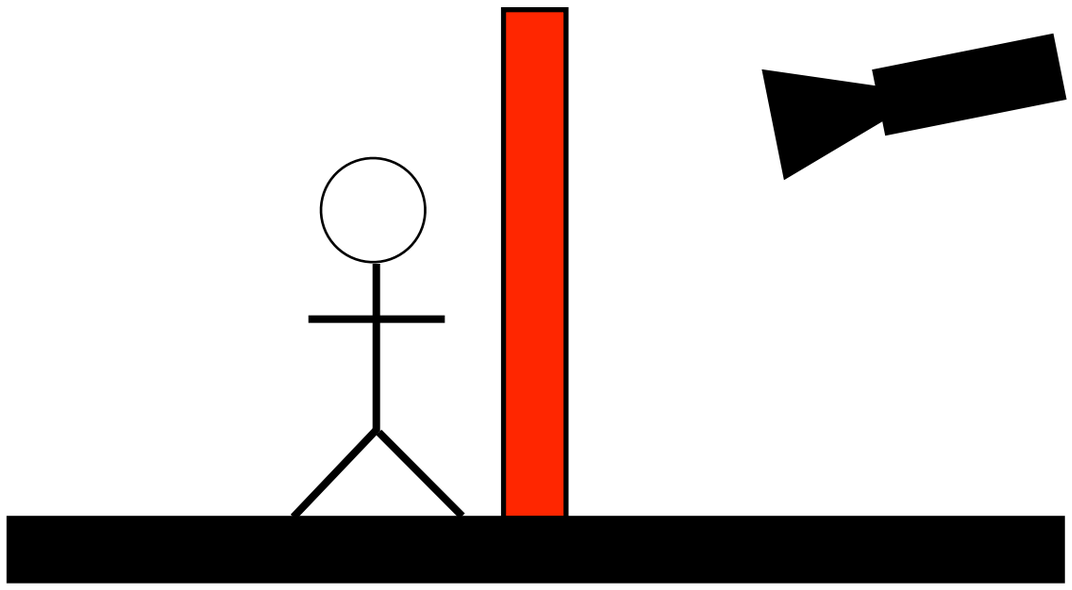
Understand what Humans do

Surveillance

Entertainment

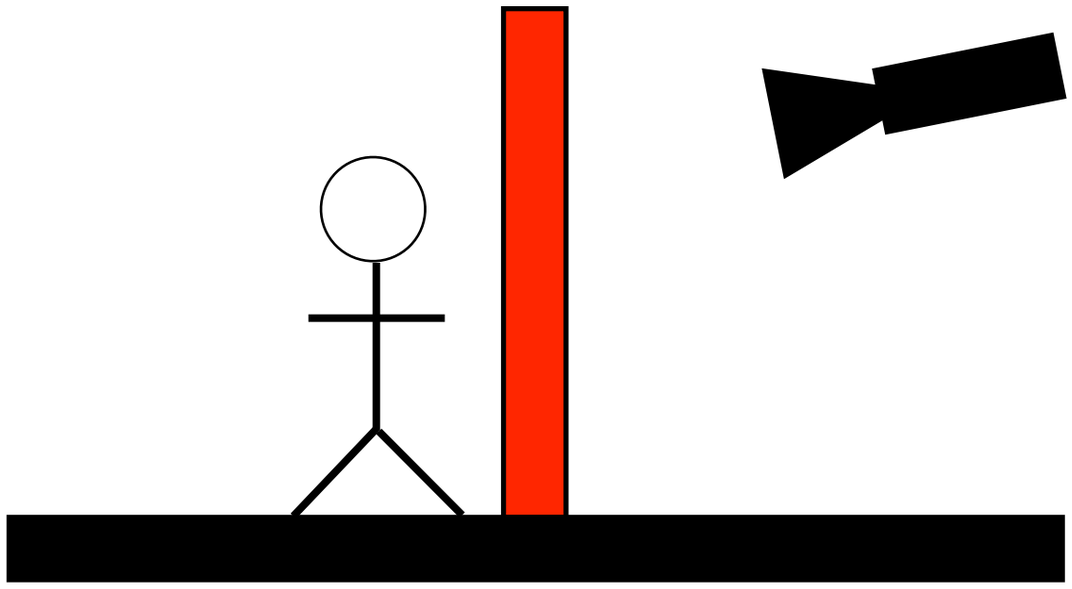
# Problem: Occlusions

# Problem: Occlusions

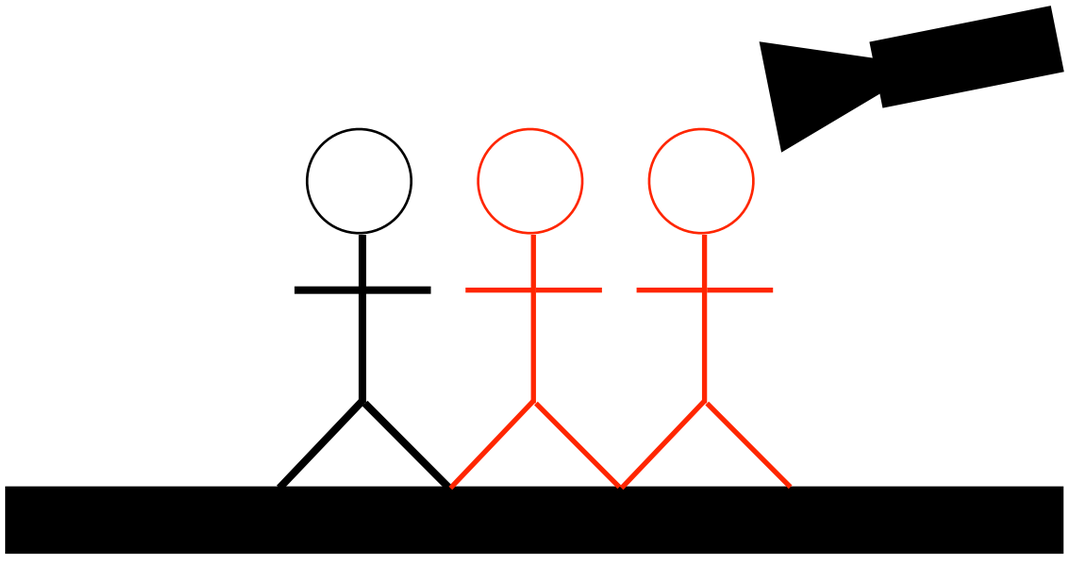


Walls
Furniture

# Problem: Occlusions

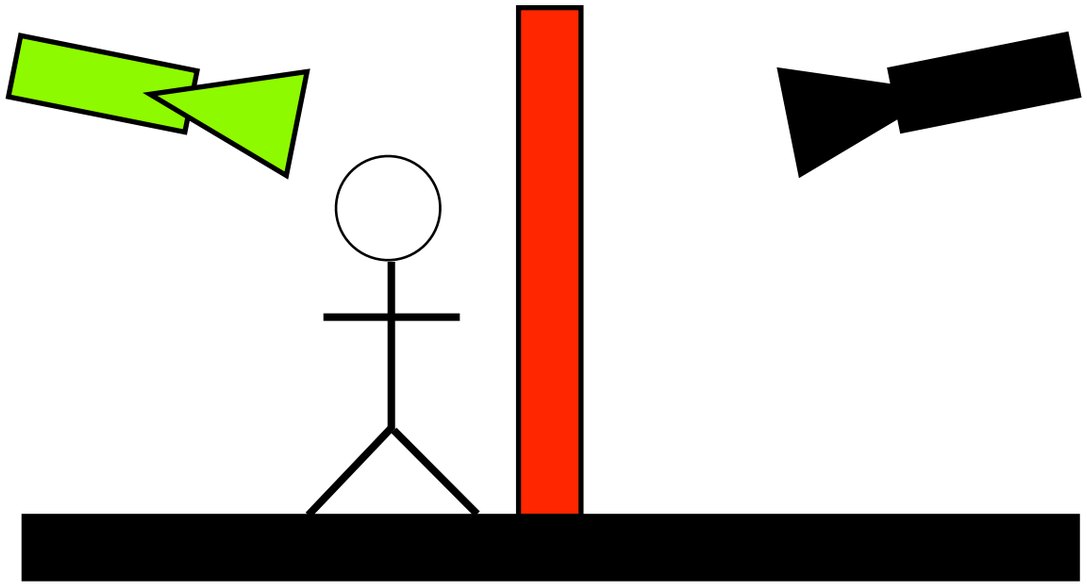


Walls
Furniture

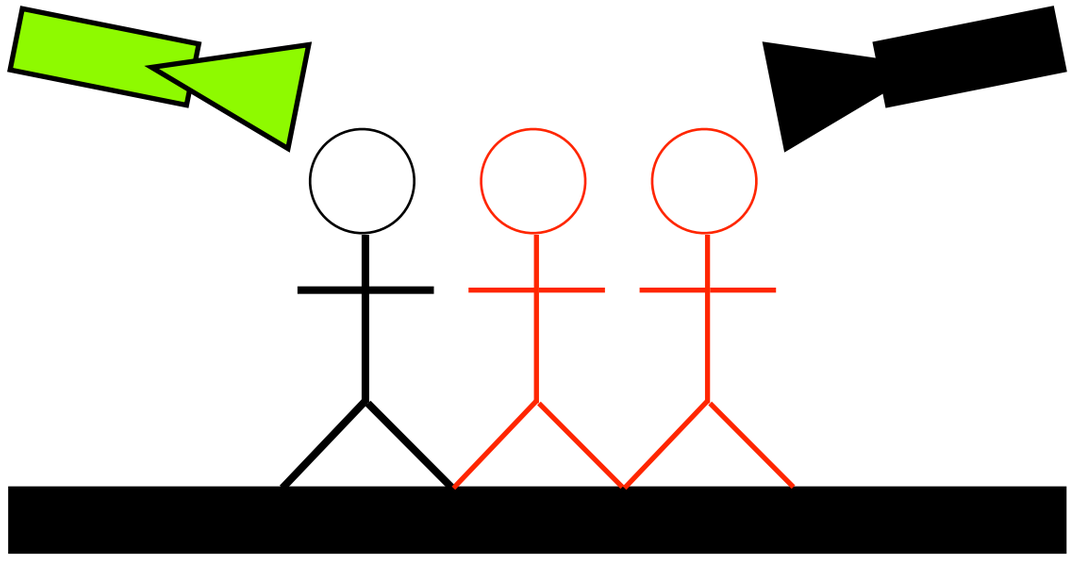


Crowds

# Problem: Occlusions



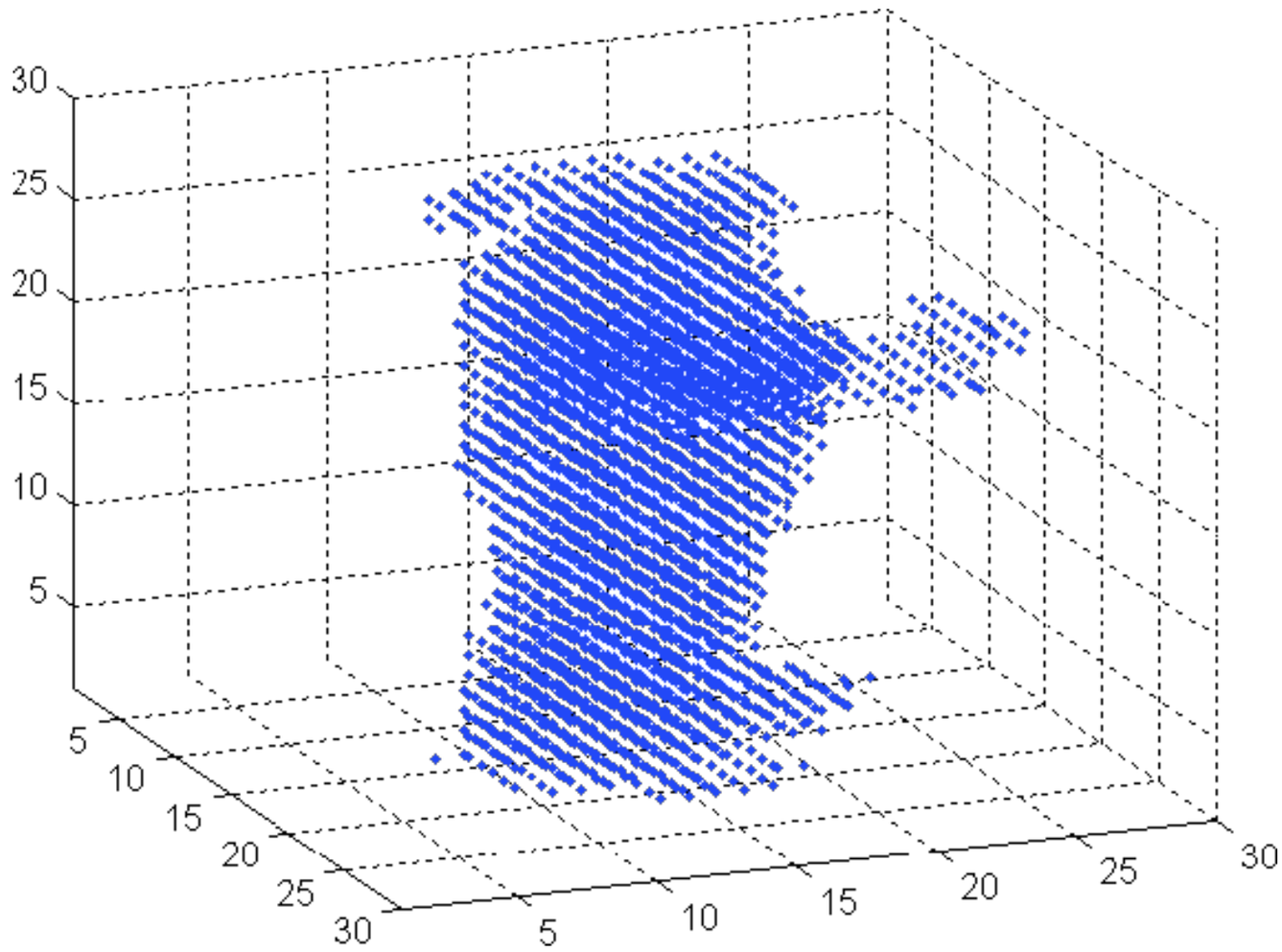
Walls
Furniture



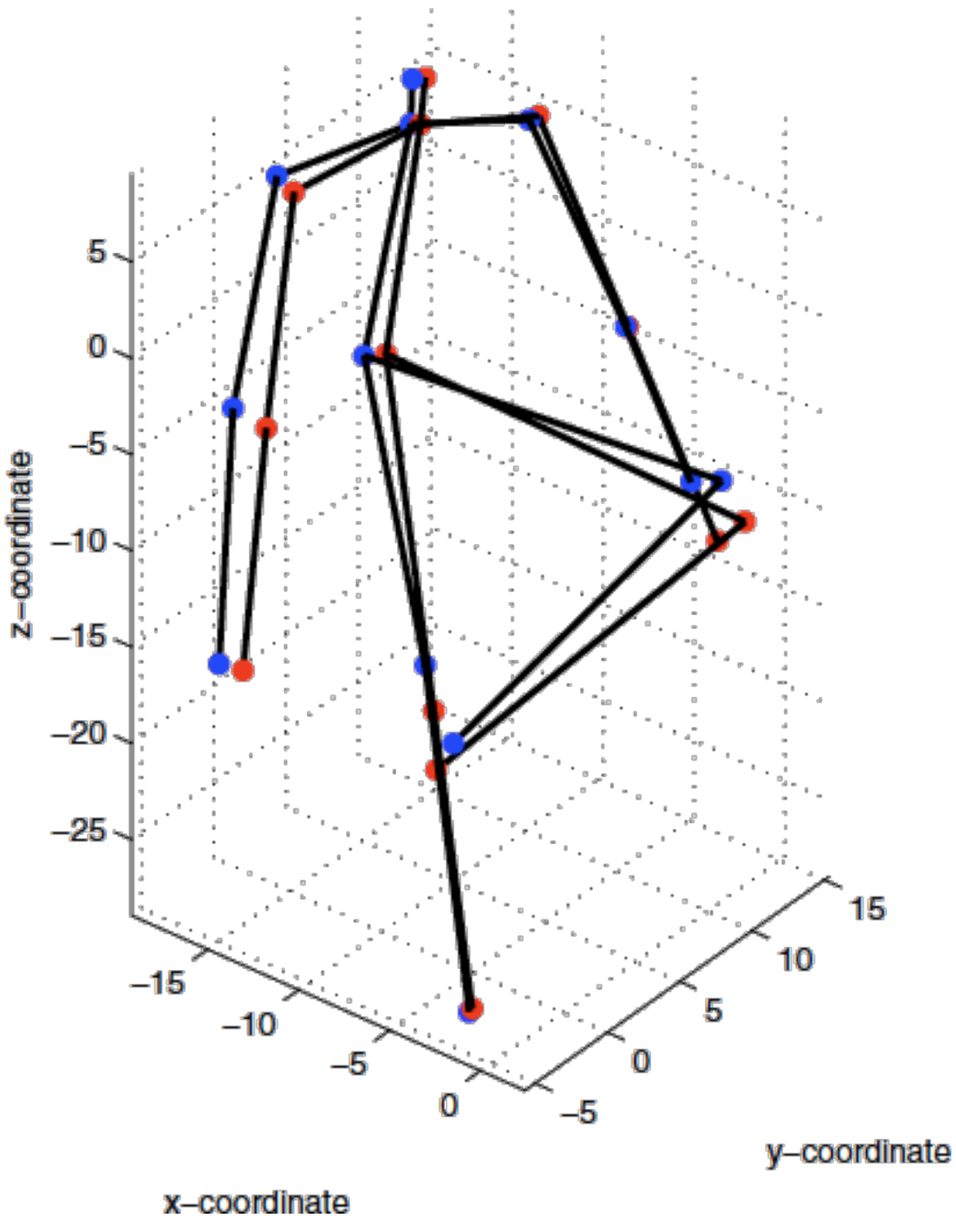
Crowds

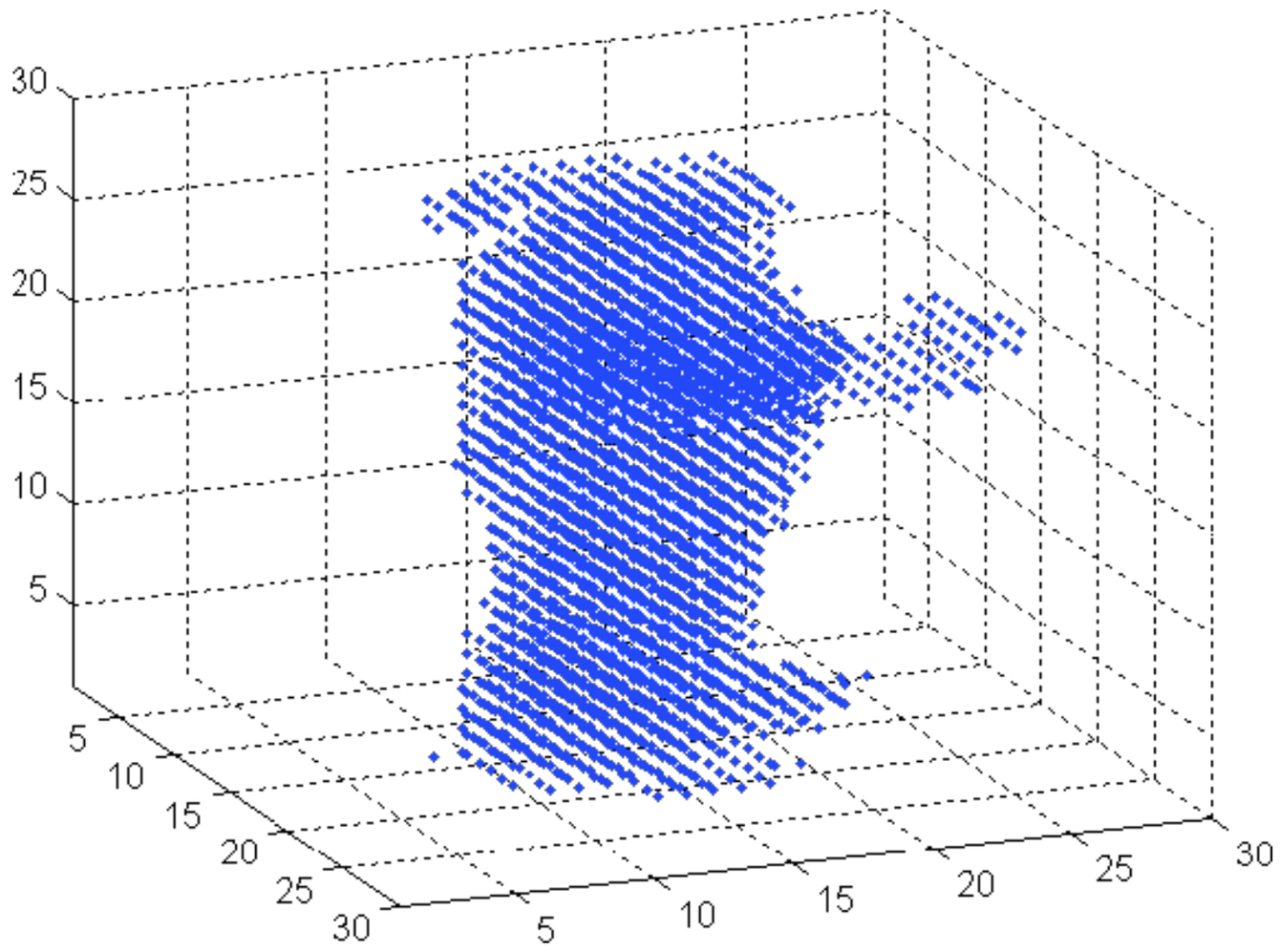
Solution: Multiple cameras

# Visual Hulls



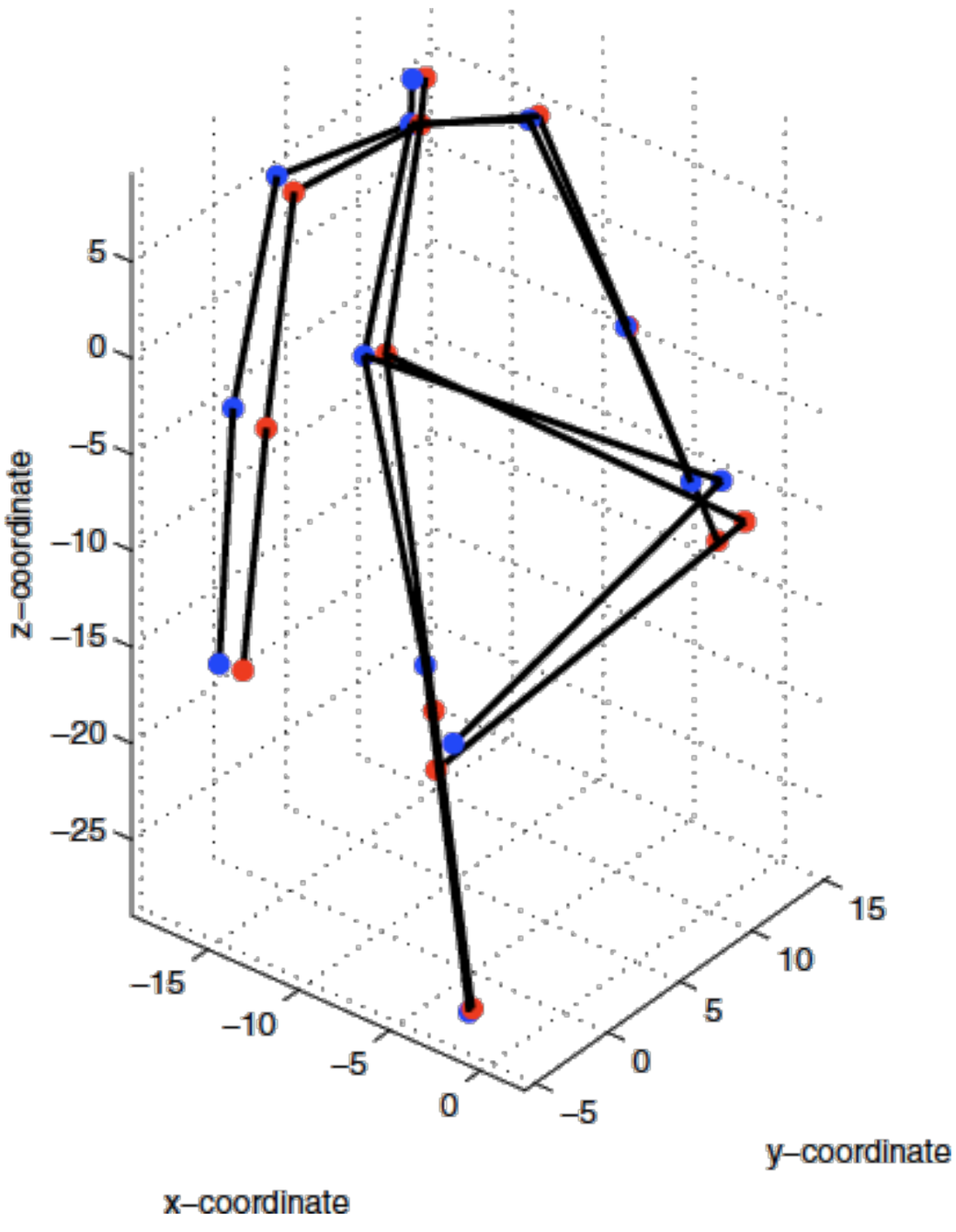
# Visual Hulls

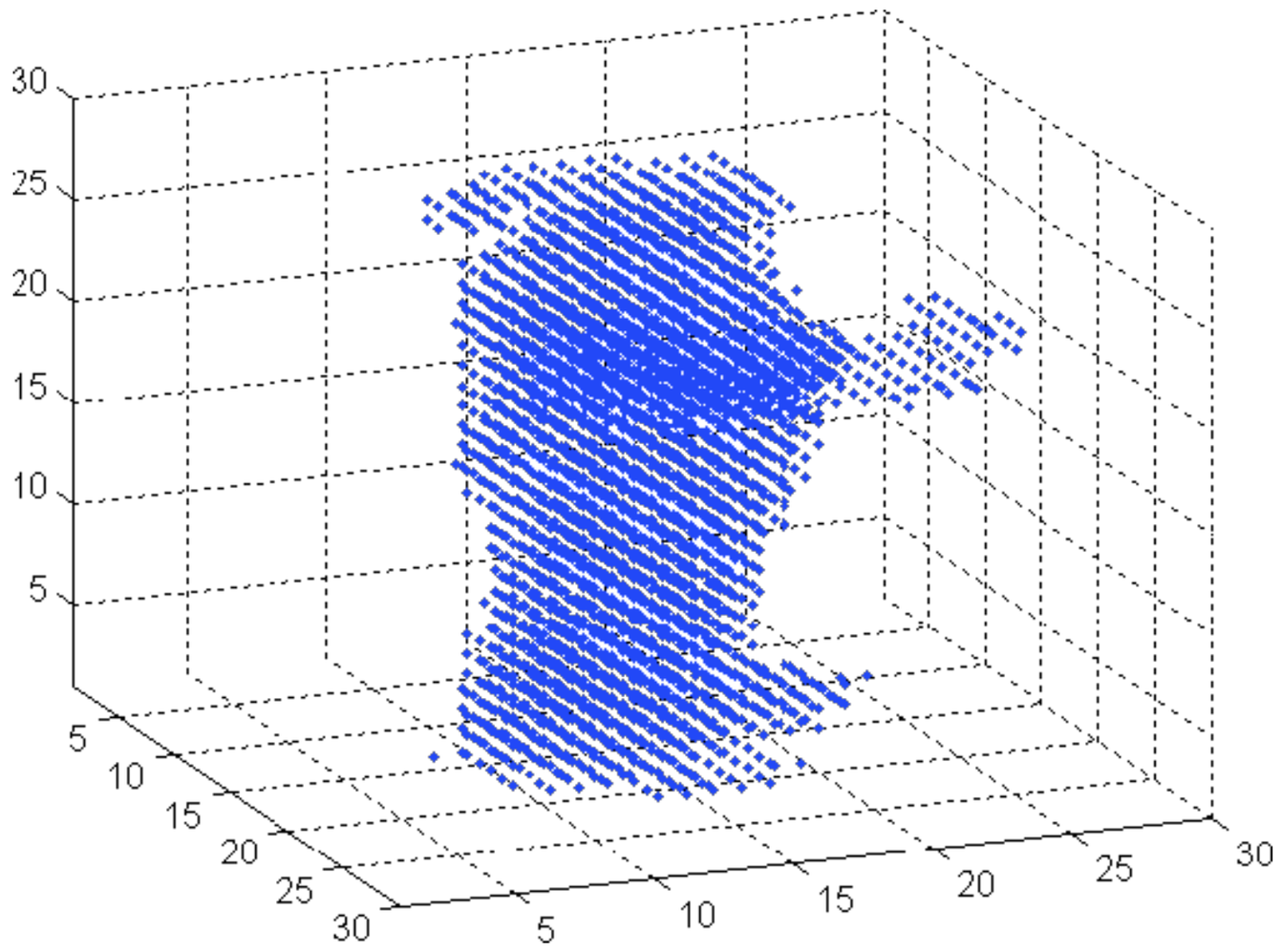



Body limbs configuration

# Visual Hulls

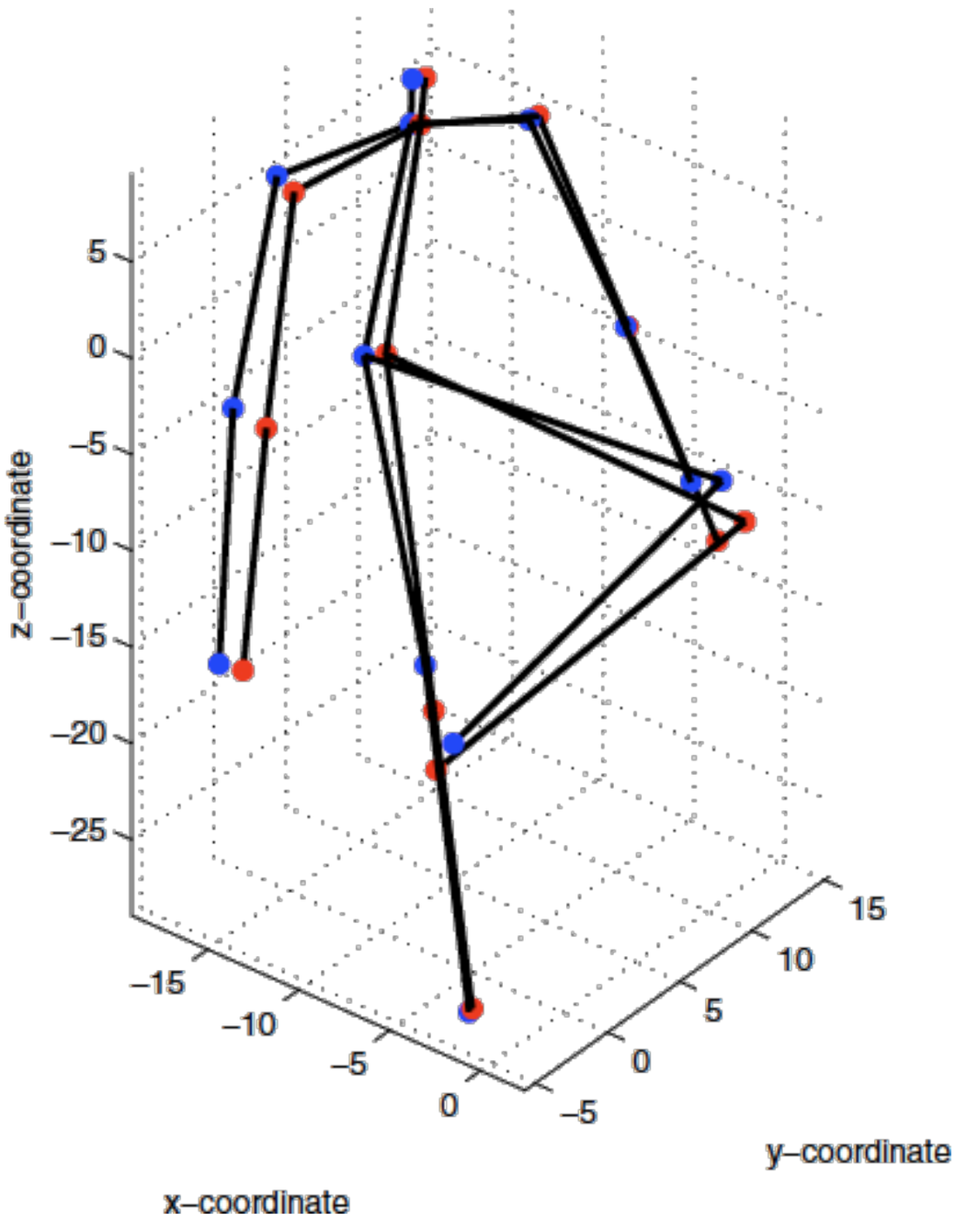


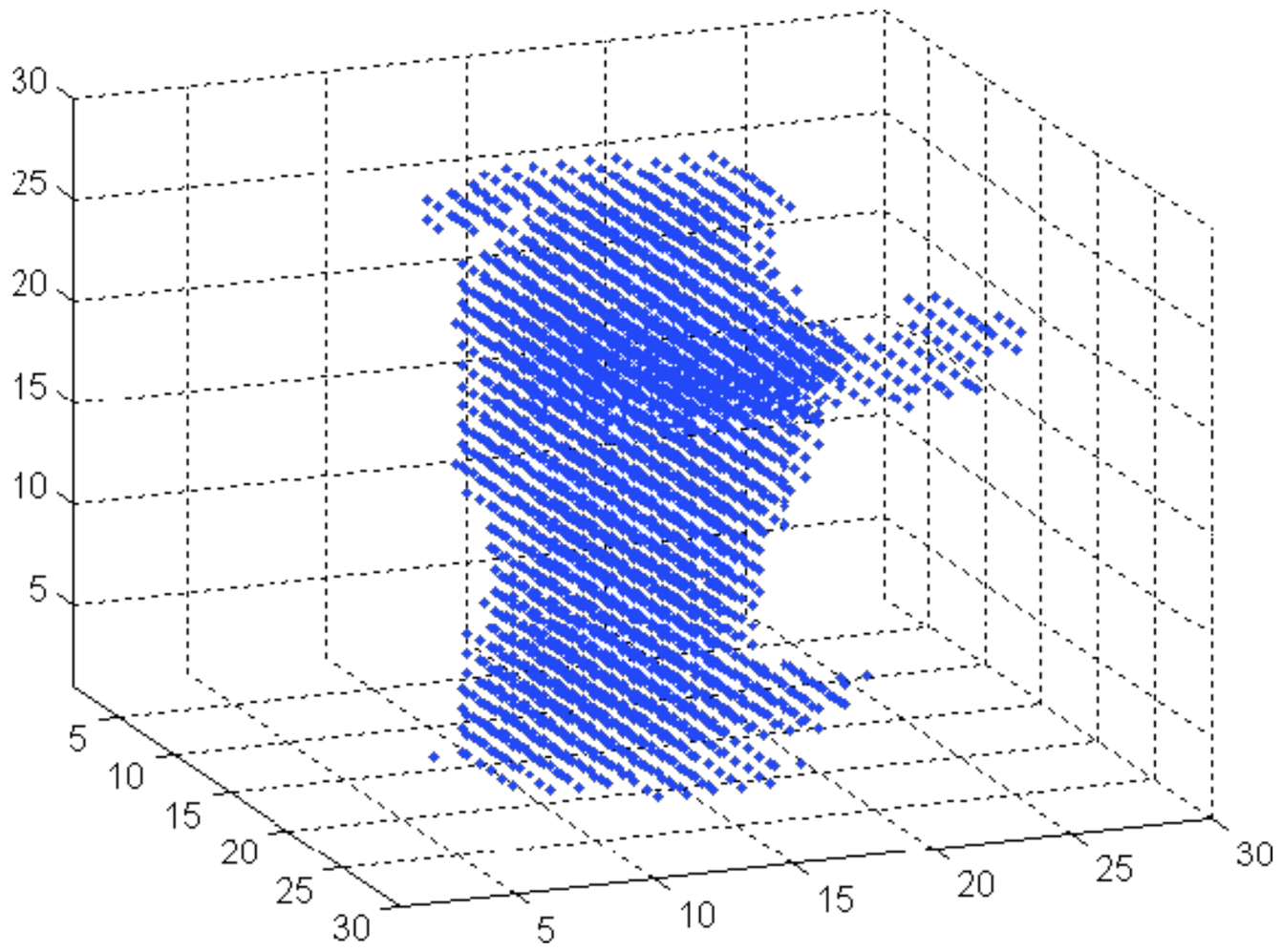



Centralized

Body limbs configuration

# Visual Hulls



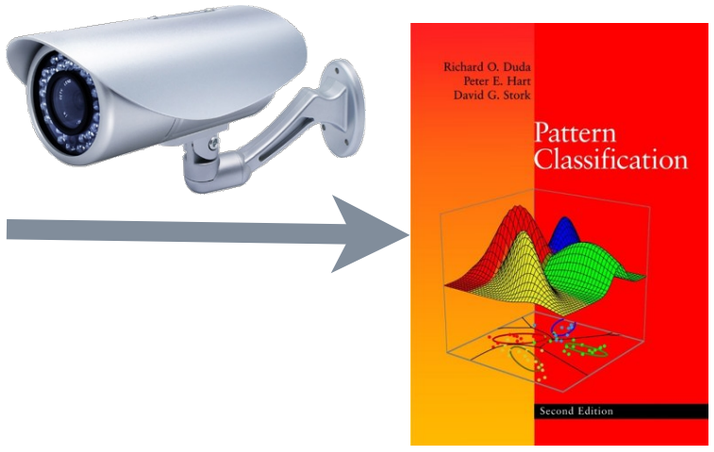
Centralized

Hard to compute

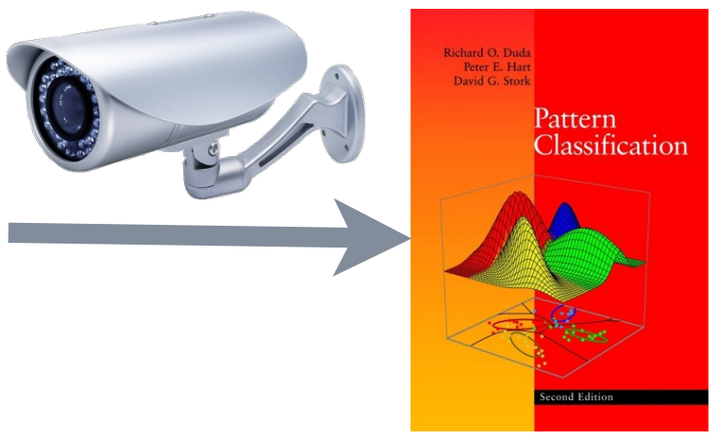
Body limbs configuration

# Proposal: Multicamera 2D processing

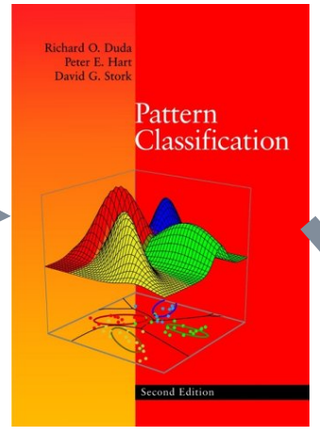
# Proposal: Multicamera 2D processing



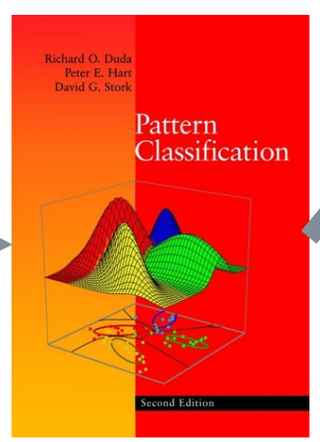

## Single view classification

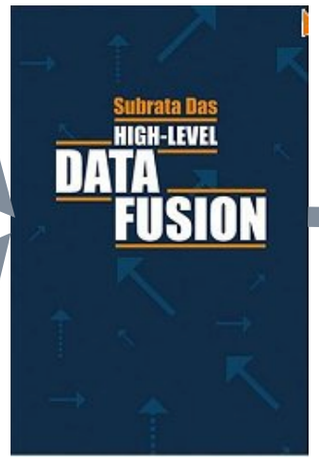


# Proposal: Multicamera 2D processing



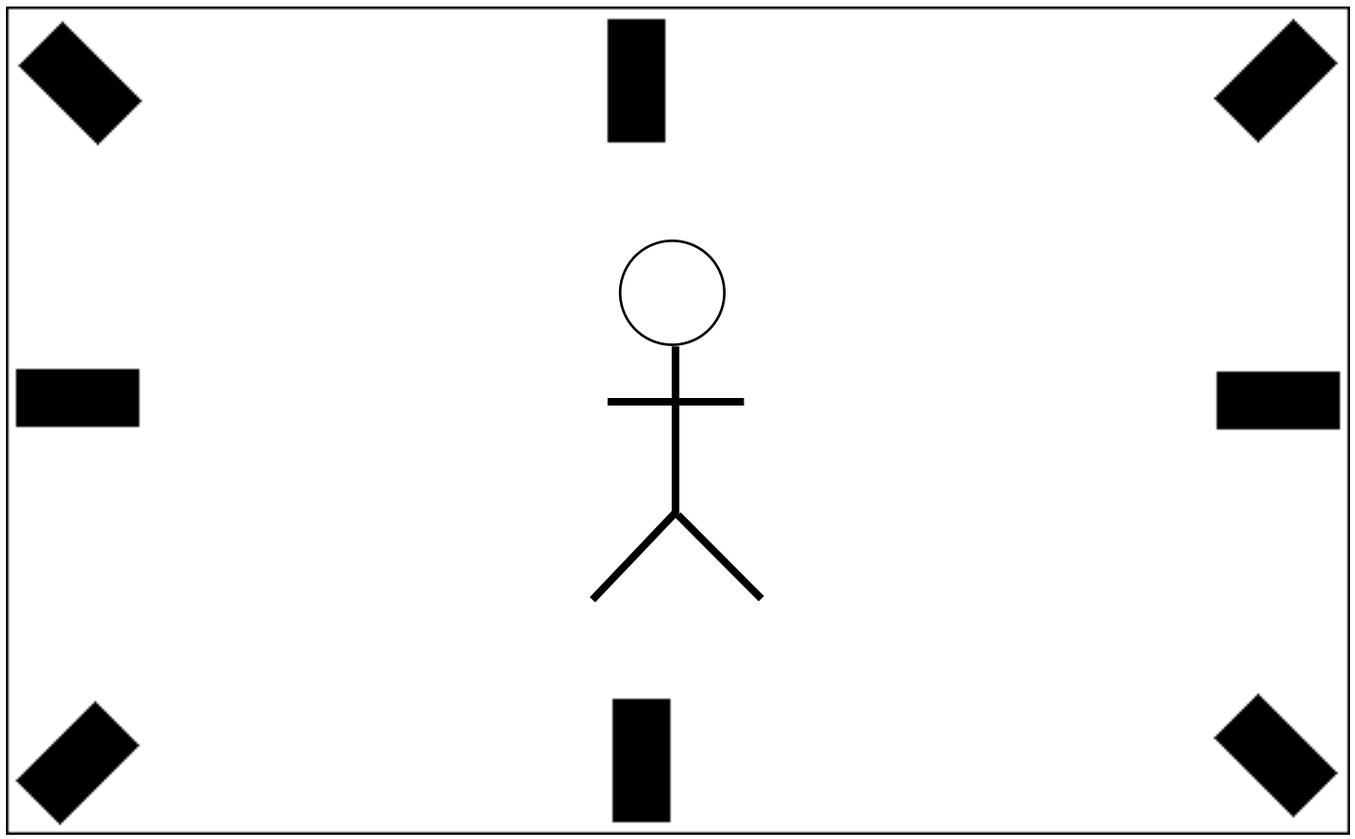

Single view classification



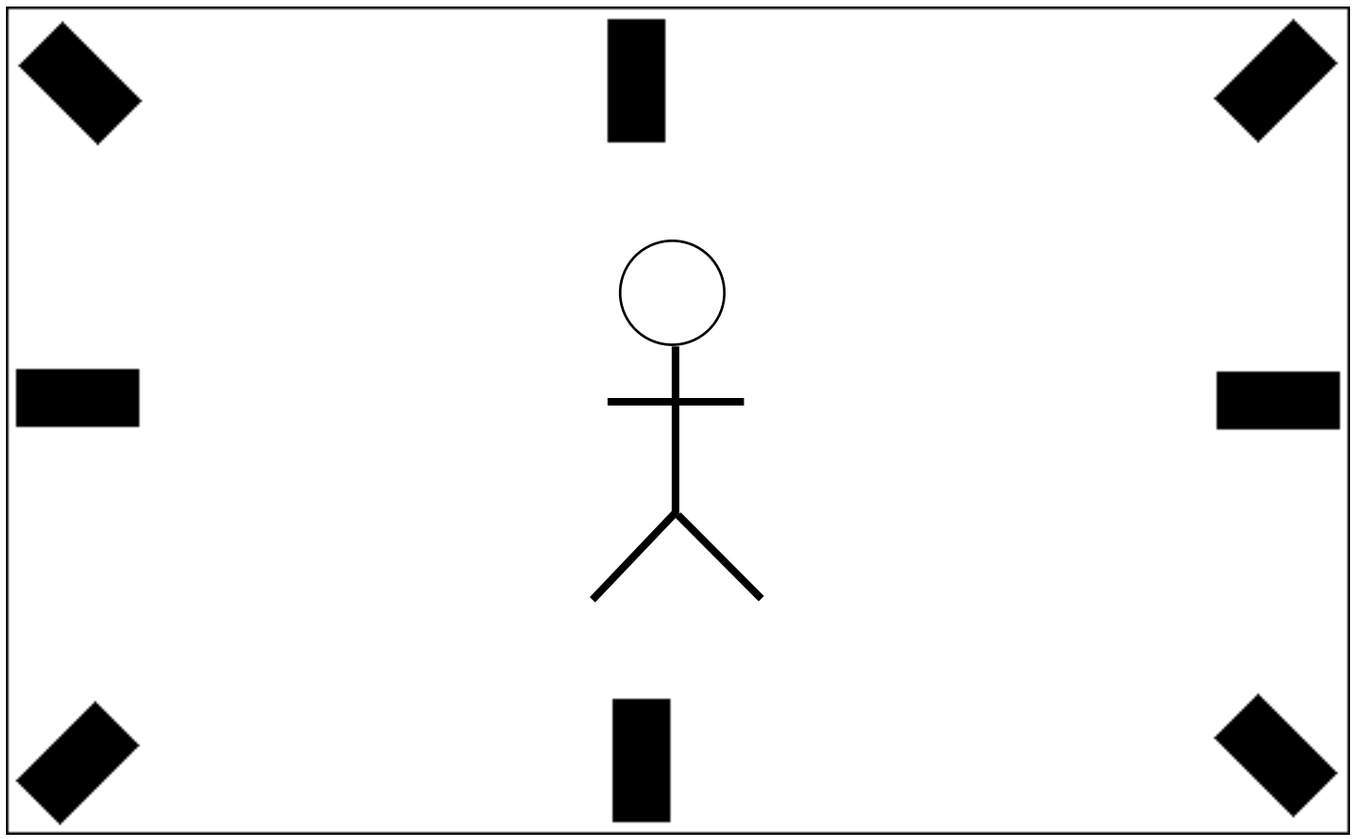

Classification Fusion

# Assumptions



Only one human

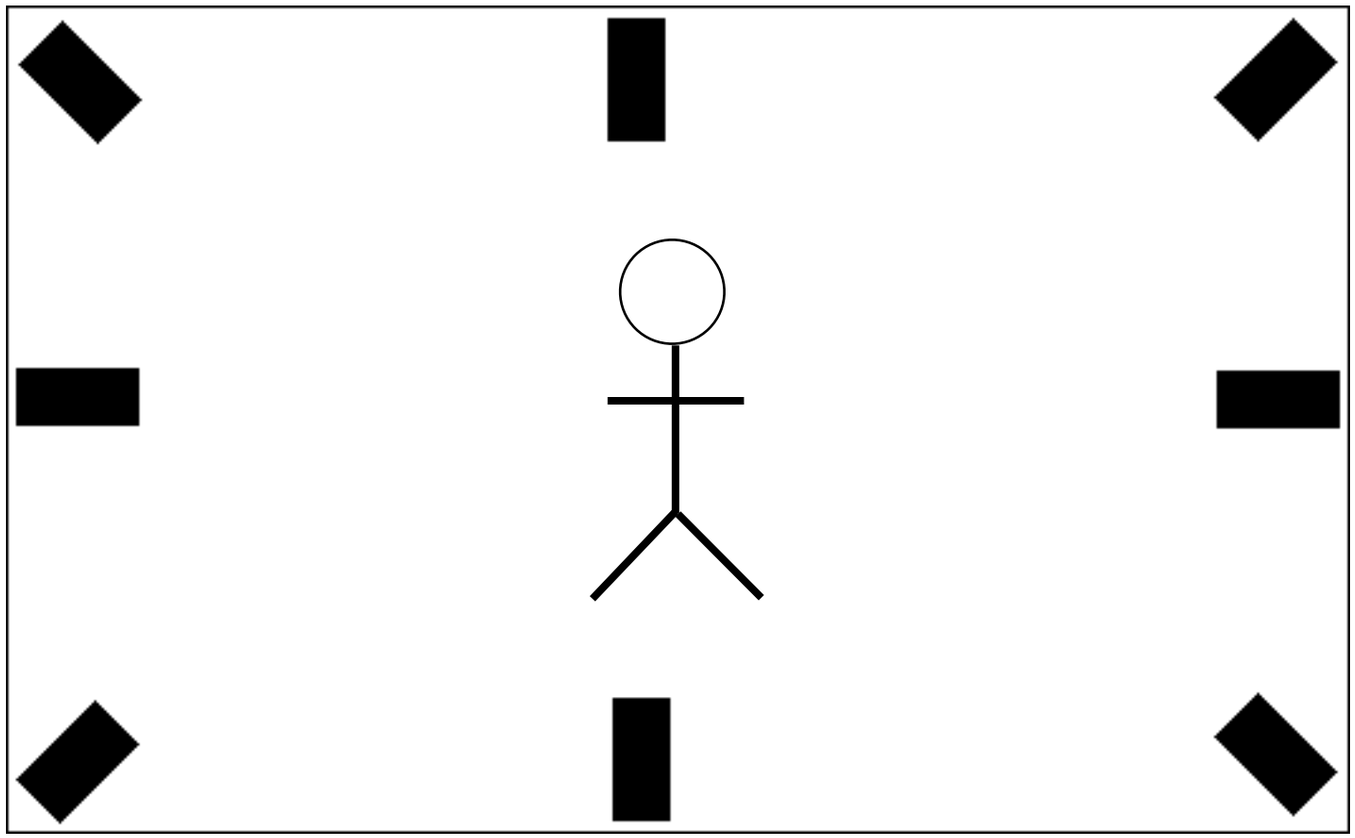
# Assumptions



Only one human

Observed by all the cameras

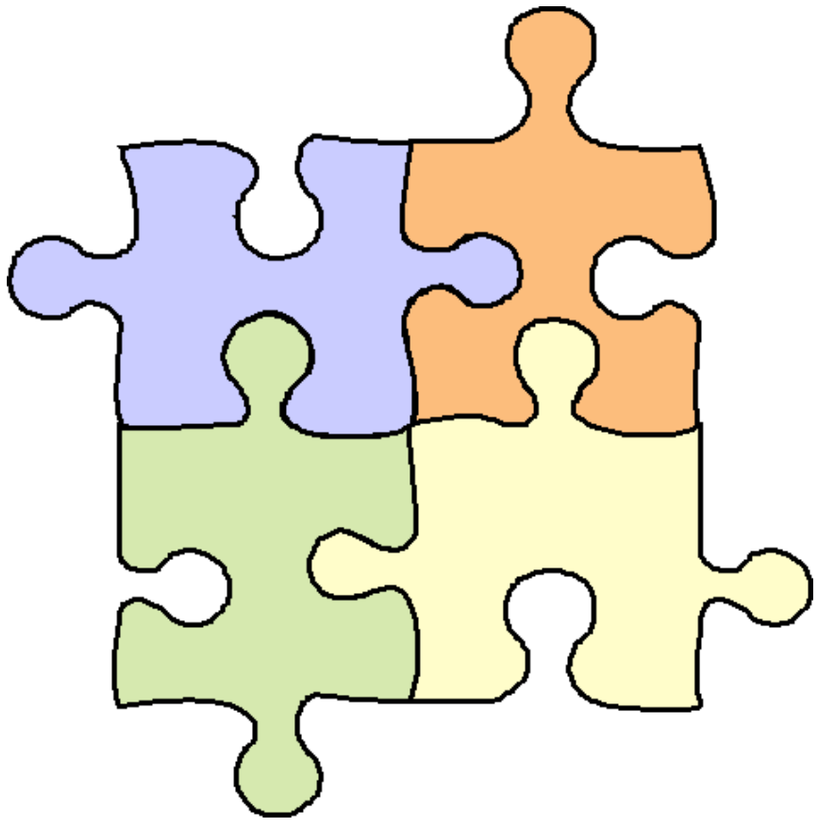
# Assumptions



Only one human

Observed by all the cameras

Only instantaneous processing



# II. Architecture

Feature extraction

Feature extraction

Feature extraction

- Optical flow
- Silhouette
- Spatio-temporal features
- Trajectory
- ...

Feature extraction

Feature extraction

Feature extraction

Feature extraction

Dimensionality Reduction

Feature extraction

Dimensionality Reduction

Feature extraction

Dimensionality Reduction

- Non-supervised: PCA, LPP, ISOMAP,...
- Supervised: LDA,SLPP,...

Feature extraction

Dimensionality Reduction

Feature extraction

Dimensionality Reduction

Feature extraction

Dimensionality Reduction

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

Soft classification:

- Conditional Density Estimation
- Dempster-Shafer belief functions

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

$p(y \mid x_A)$

Feature extraction

Dimensionality Reduction

Local Classification

$p(y \mid x_B)$

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

$p(y \mid x_A)$

$p(y \mid x_B)$

Belief Aggregation

Feature extraction

Dimensionality Reduction

Local Classification

Feature extraction

Dimensionality Reduction

Local Classification

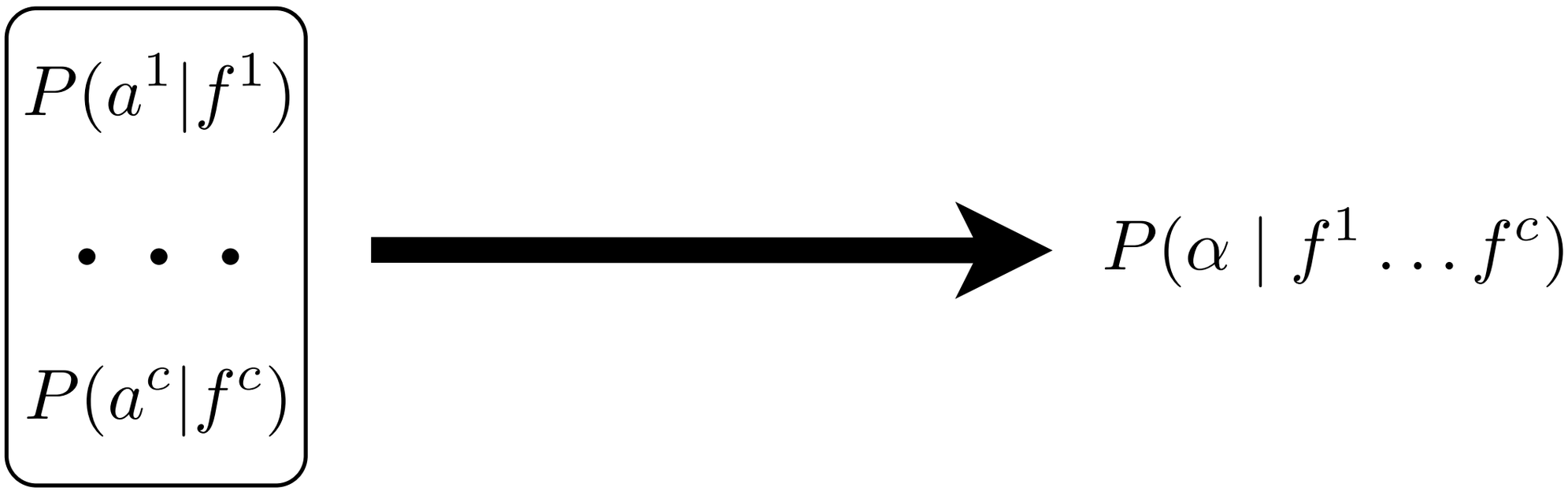
$p(y \mid x_A, x_B)$

$p(y \mid x_A)$

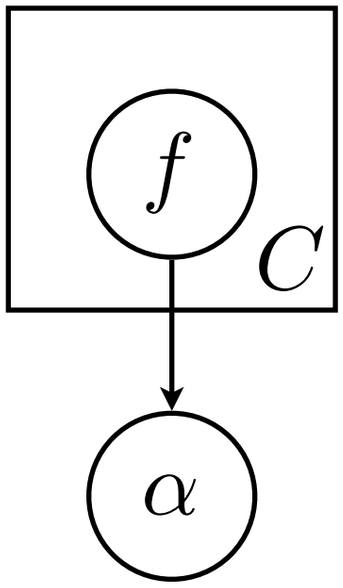
$p(y \mid x_B)$

Belief Aggregation

# III. Classifier Fusion

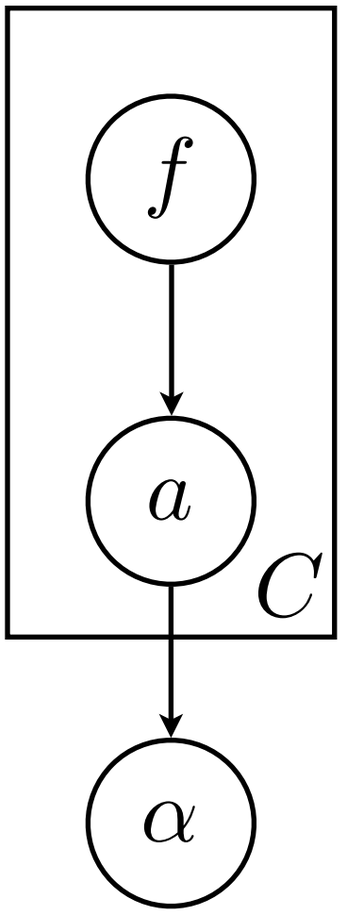

$P(a^1|f^1)$
$\bullet \quad \bullet \quad \bullet$
$P(a^c|f^c)$
$P(\alpha \mid f^1 \ldots f^c)$

# Weighted Voting



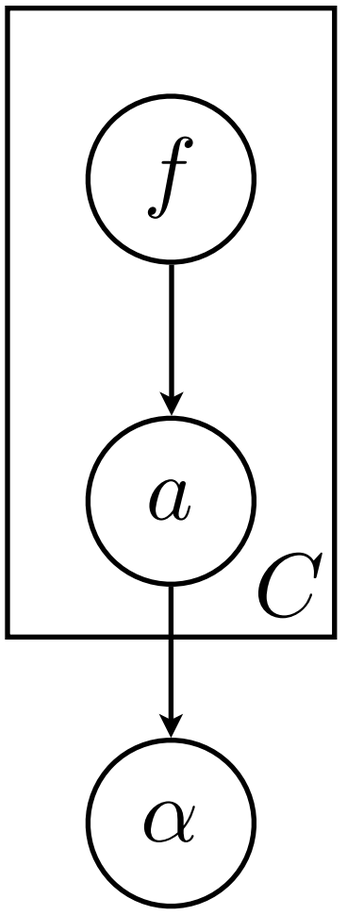

$$P(\alpha \mid \mathbf{f}) \propto \prod_c P(a^c \mid f^c)$$

# Meassuring camera errors




$$P(\alpha, \mathbf{a}, \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})\, P(\mathbf{f})$$

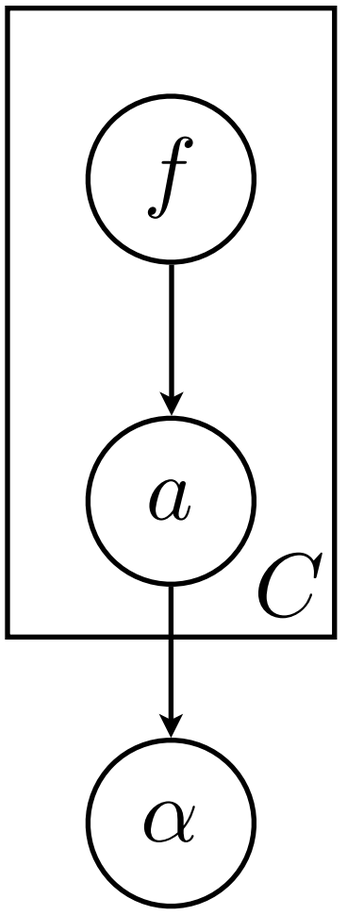
# Meassuring camera errors




$$P(\alpha, \mathbf{a}, \mathbf{f}) = P(\alpha \mid \mathbf{a}) P(\mathbf{a} \mid \mathbf{f}) P(\mathbf{f})$$

$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = P(\alpha \mid \mathbf{a}) P(\mathbf{a} \mid \mathbf{f})$$

# Meassuring camera errors

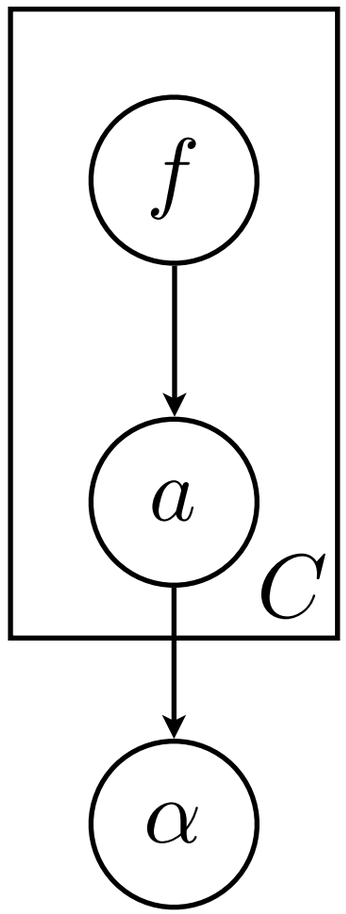



$$P(\alpha, \mathbf{a}, \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})\, P(\mathbf{f})$$

$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})$$

$$P(\alpha \mid \mathbf{a}) \doteq \prod_{c} P(\alpha \mid a^c)$$

# Meassuring camera errors



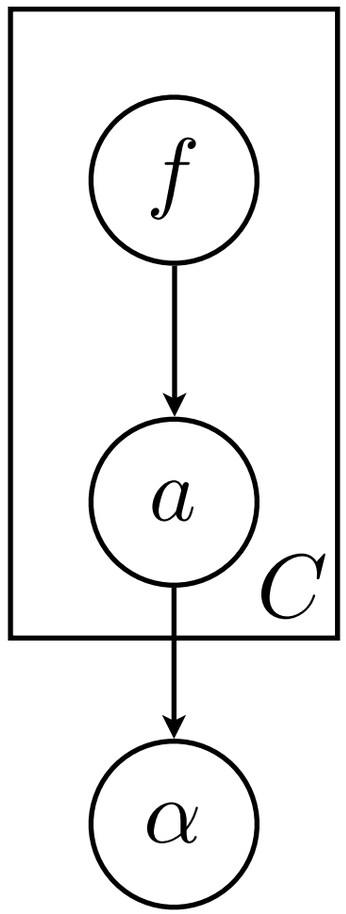

$$P(\alpha, \mathbf{a}, \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})\, P(\mathbf{f})$$

$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})$$

$$P(\alpha \mid \mathbf{a}) \doteq \prod_c P(\alpha \mid a^c)$$

$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = \prod_c P(\alpha \mid a^c)\, P(a^c \mid f^c)$$

# Meassuring camera errors




$$P(\alpha, \mathbf{a}, \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})\, P(\mathbf{f})$$

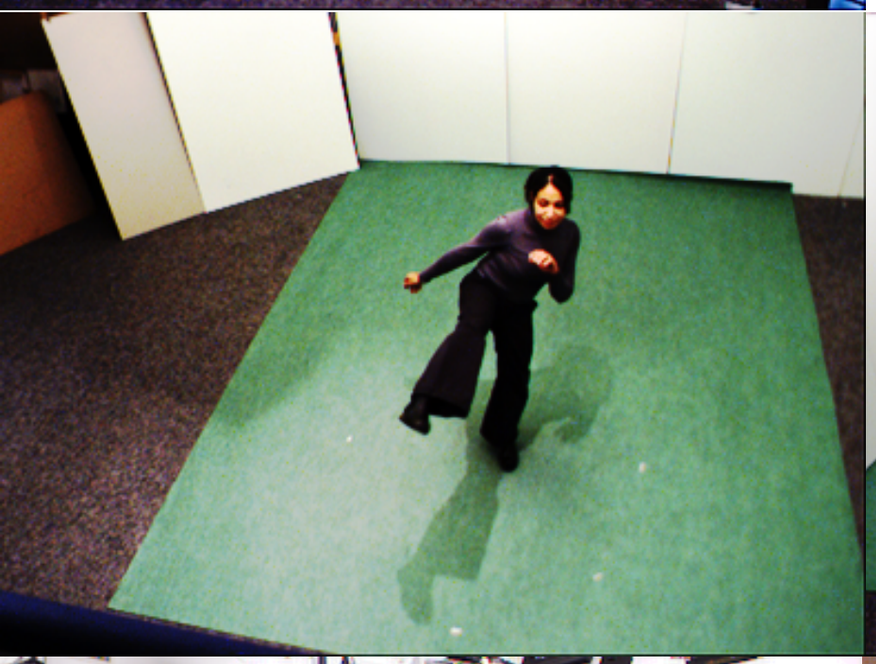
$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = P(\alpha \mid \mathbf{a})\, P(\mathbf{a} \mid \mathbf{f})$$

$$P(\alpha \mid \mathbf{a}) \doteq \prod_c P(\alpha \mid a^c)$$

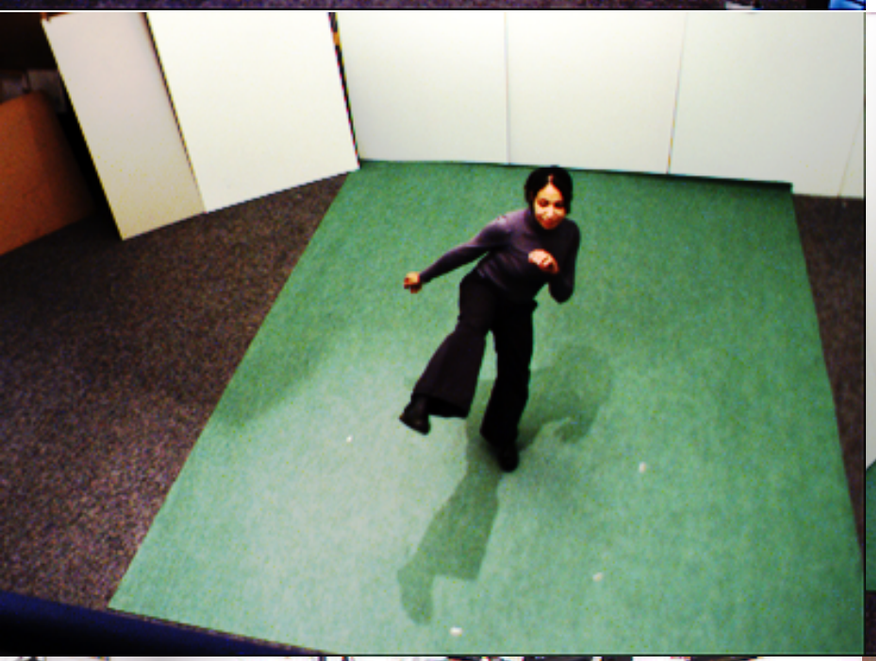
$$P(\alpha, \mathbf{a} \mid \mathbf{f}) = \prod_c P(\alpha \mid a^c)\, P(a^c \mid f^c)$$

$$P(\alpha \mid \mathbf{f}) = \prod_c \sum_{a^c} P(\alpha \mid a^c)\, P(a^c \mid f^c)$$

# IV. Experiments



# IXMAS
# INRIA Xmas Motion Acquisition Sequences



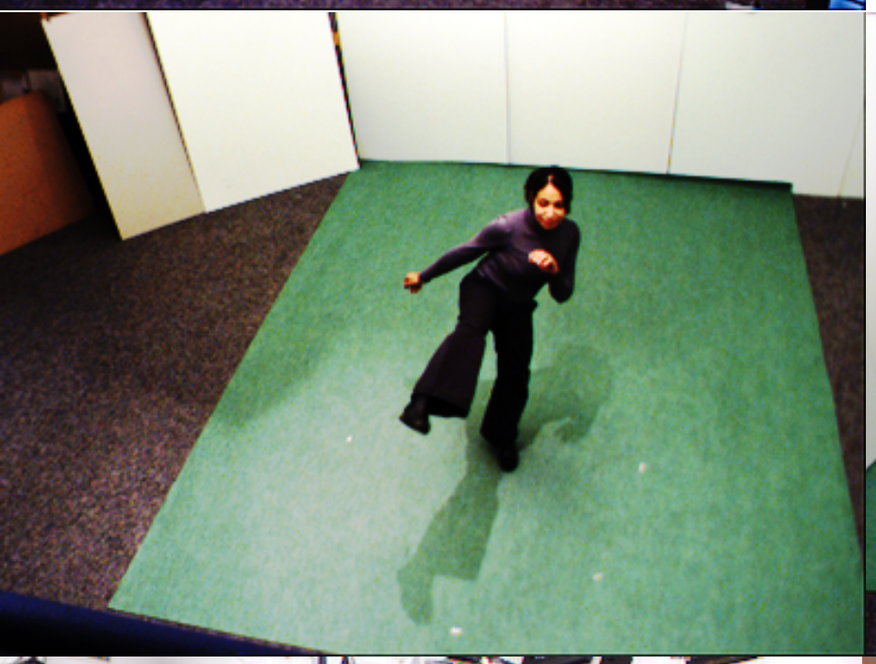
# IXMAS
# INRIA Xmas Motion Acquisition Sequences

5 camera views

12 actors

11 actions

36 clips per action



# IXMAS
# INRIA Xmas Motion Acquisition Sequences

5 camera views

12 actors

11 actions

36 clips per action

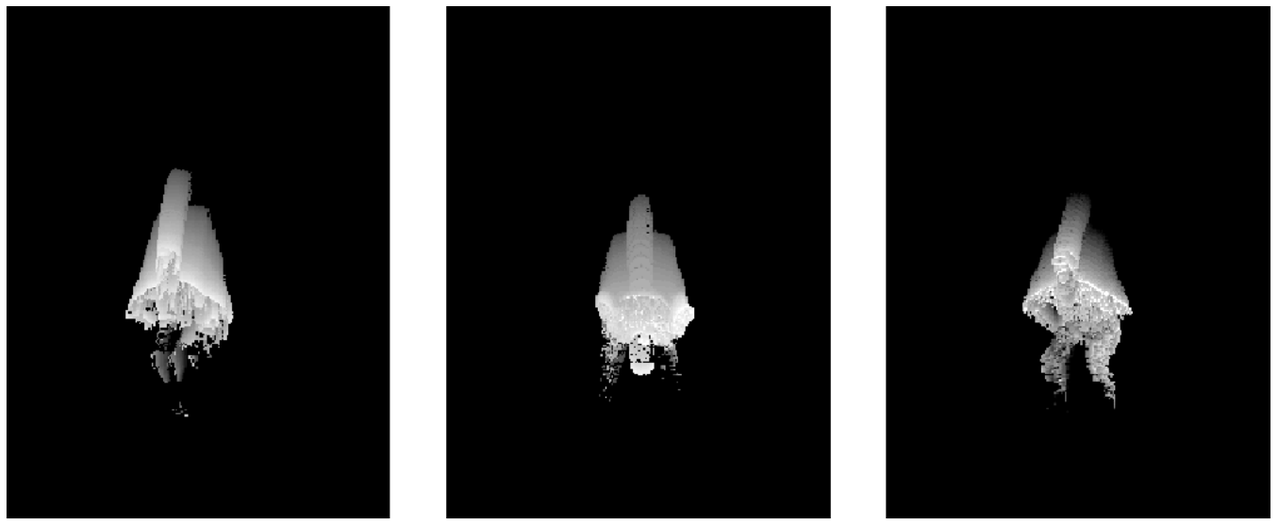
Best Results Reported:

3D: 94.81%

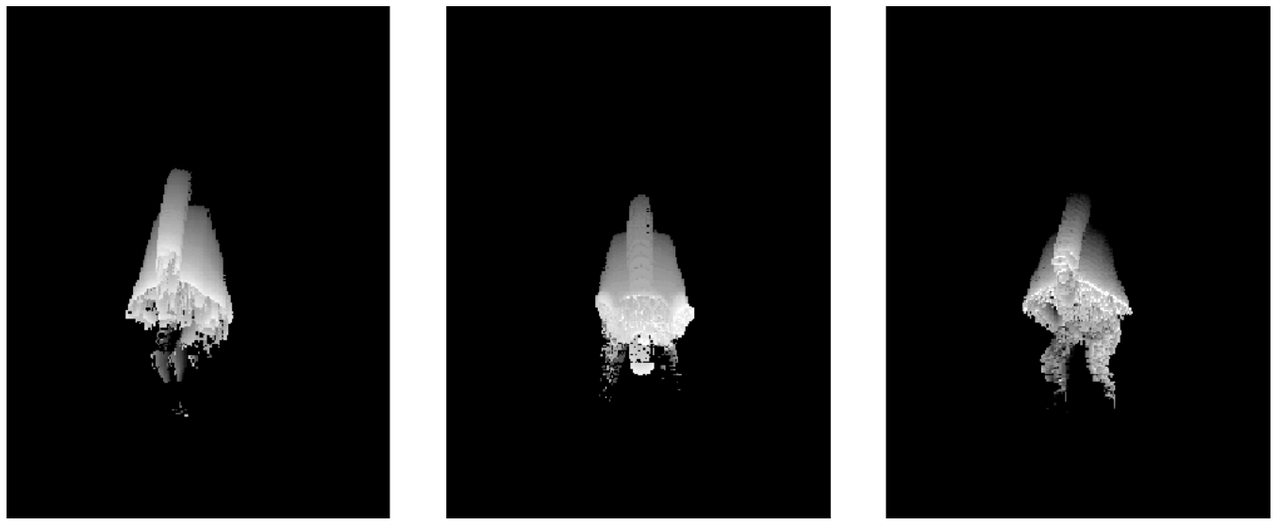
(Peng et al. ICDSC 2009)

2D: 85%

(Srivastava et al. ICDSC 2009)

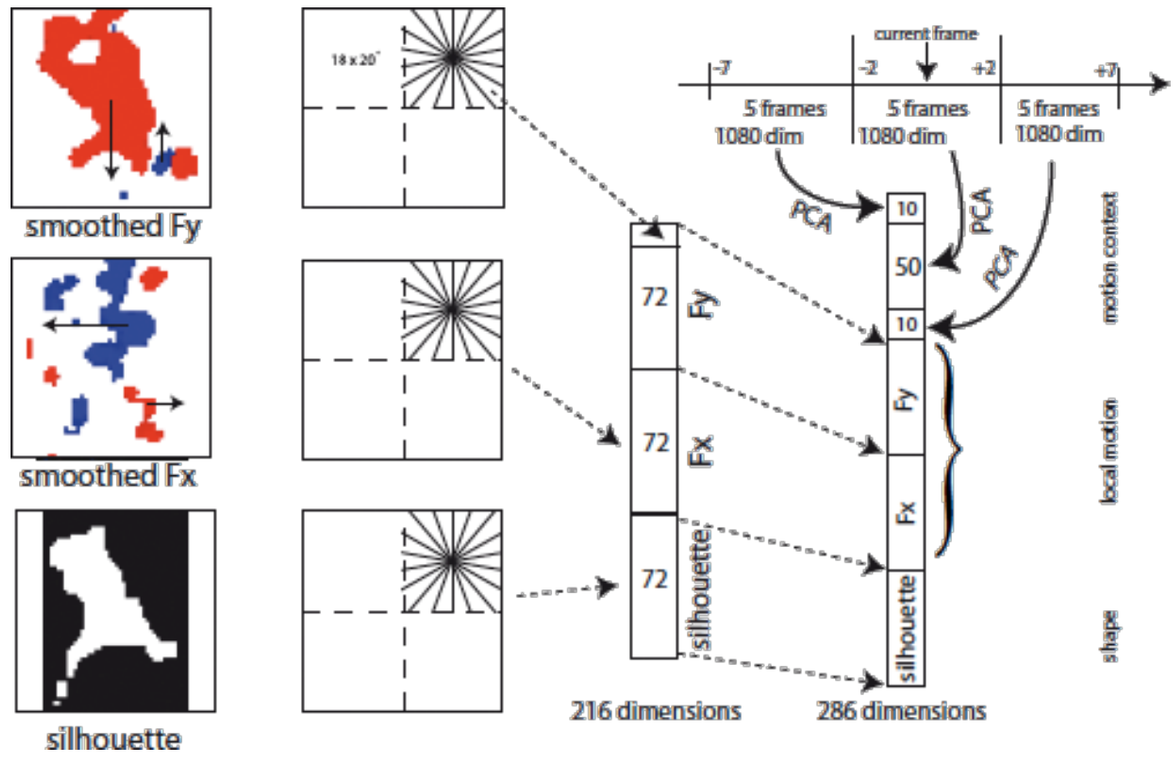


# Motion History Image

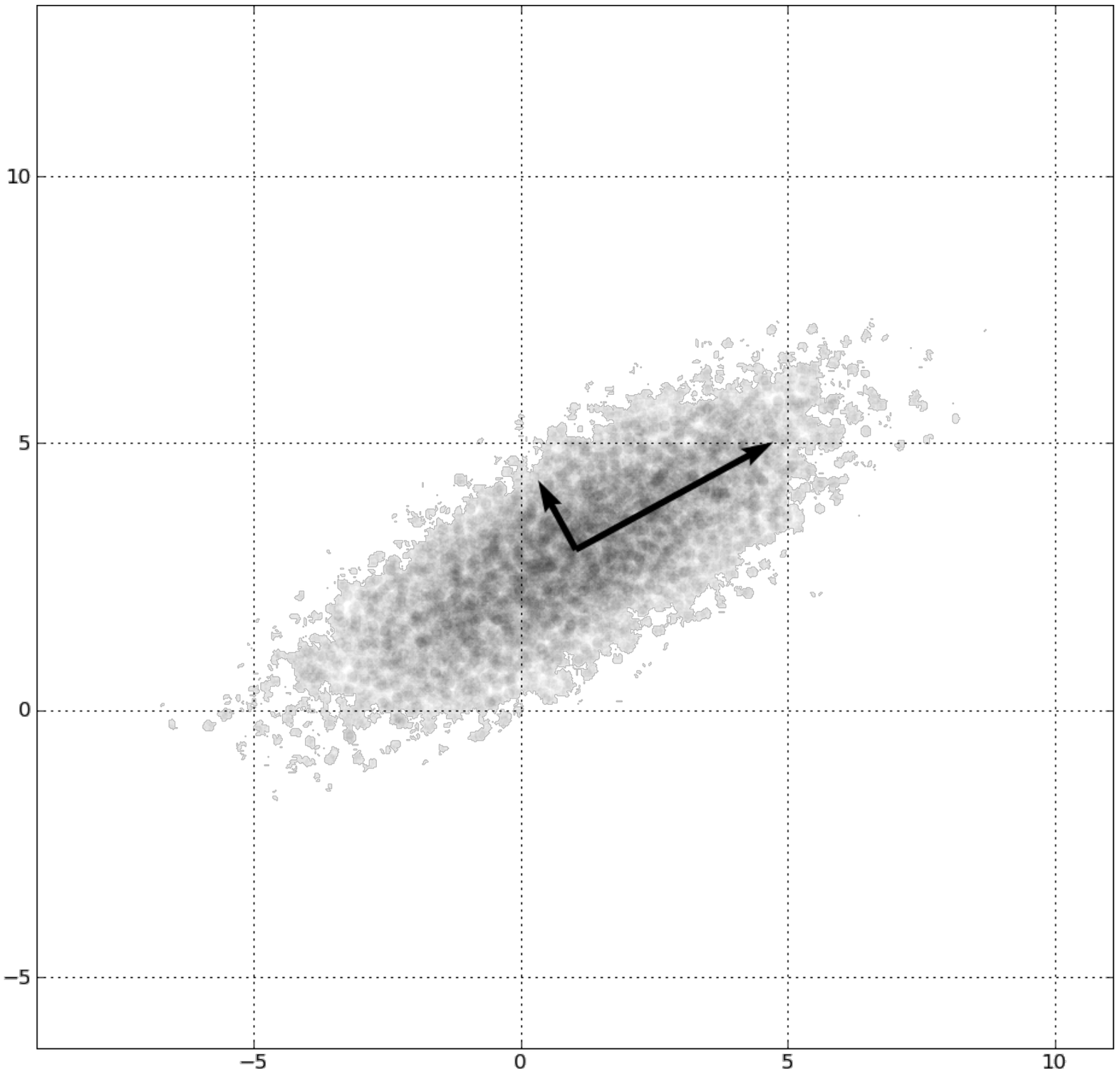


# Motion History Image

# Tran's descriptor: Optical Flow + Shape

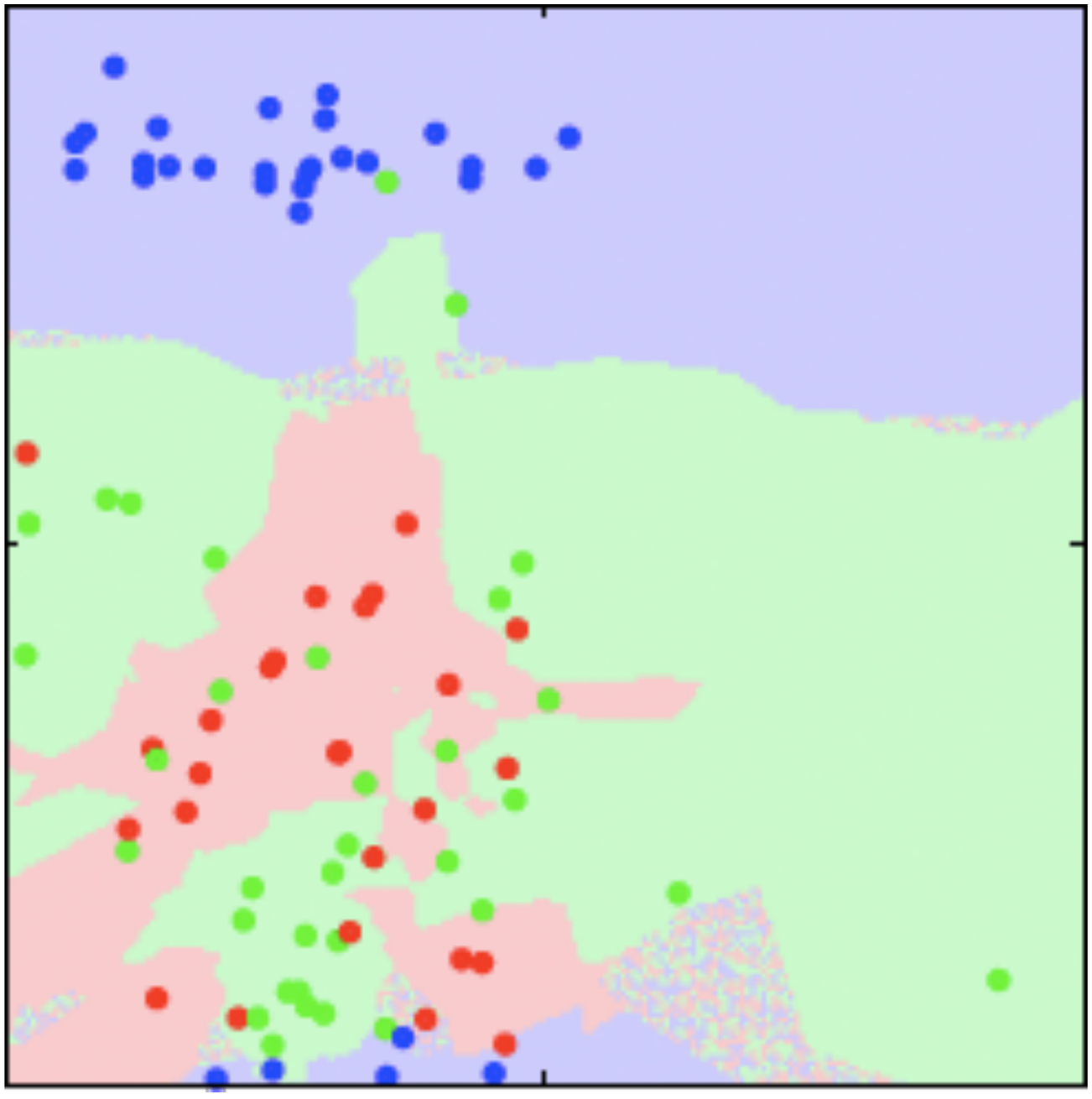


# Dimensionality Reduction: PCA



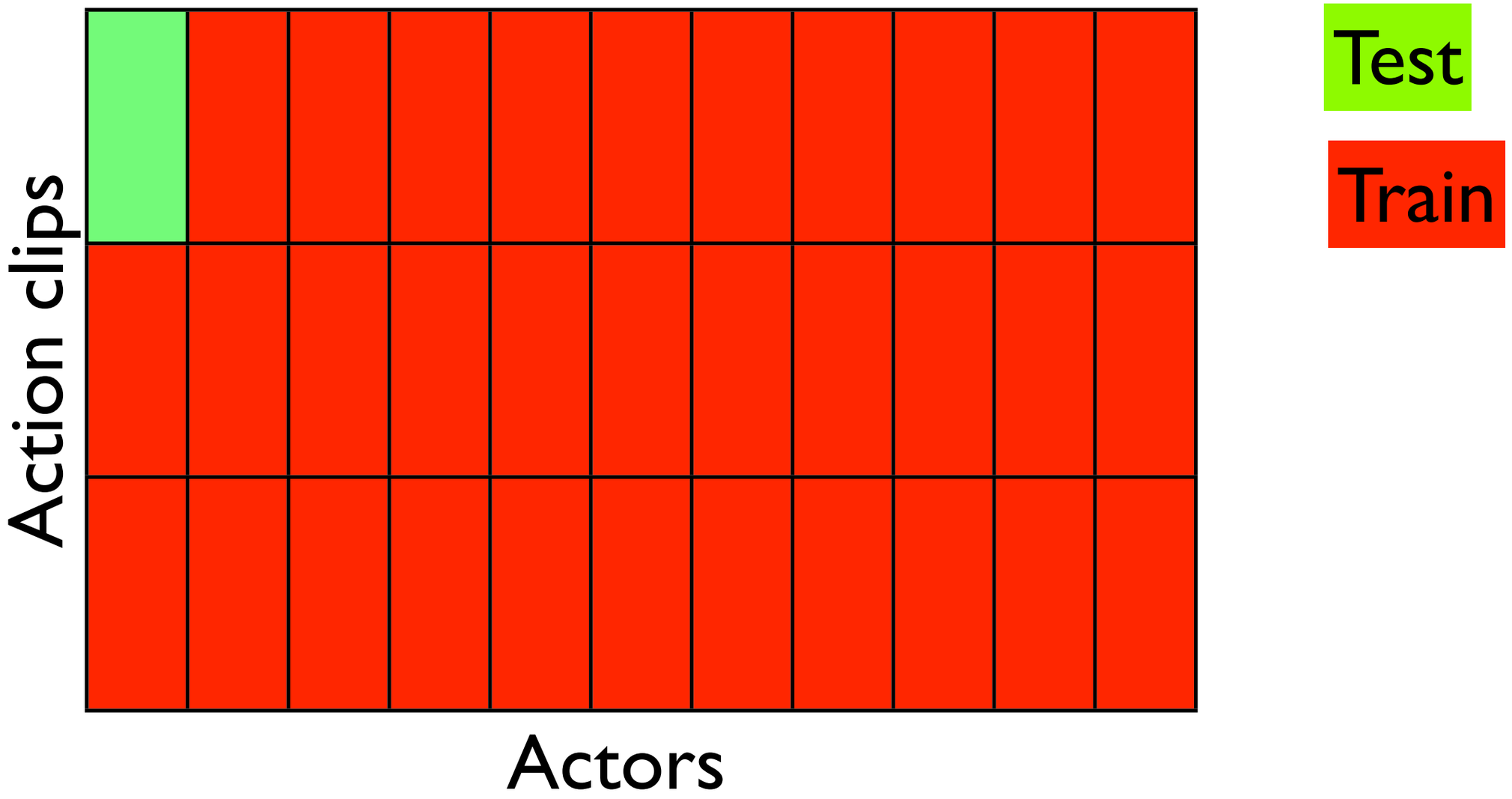

# dims: 10, 15, 20, 25, 30, 35

# Classifier: NN conditional density estimator

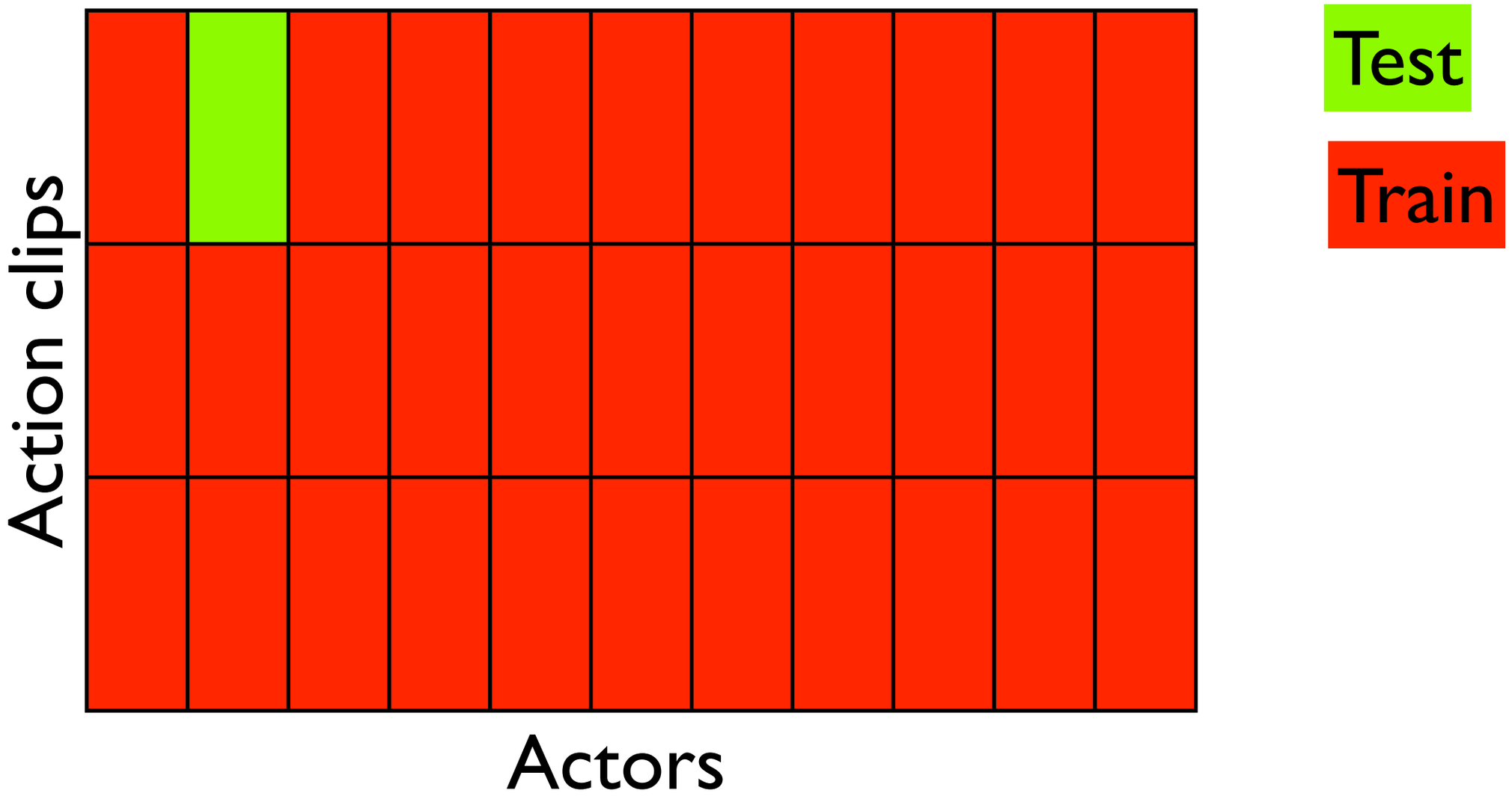


k=3, k=5

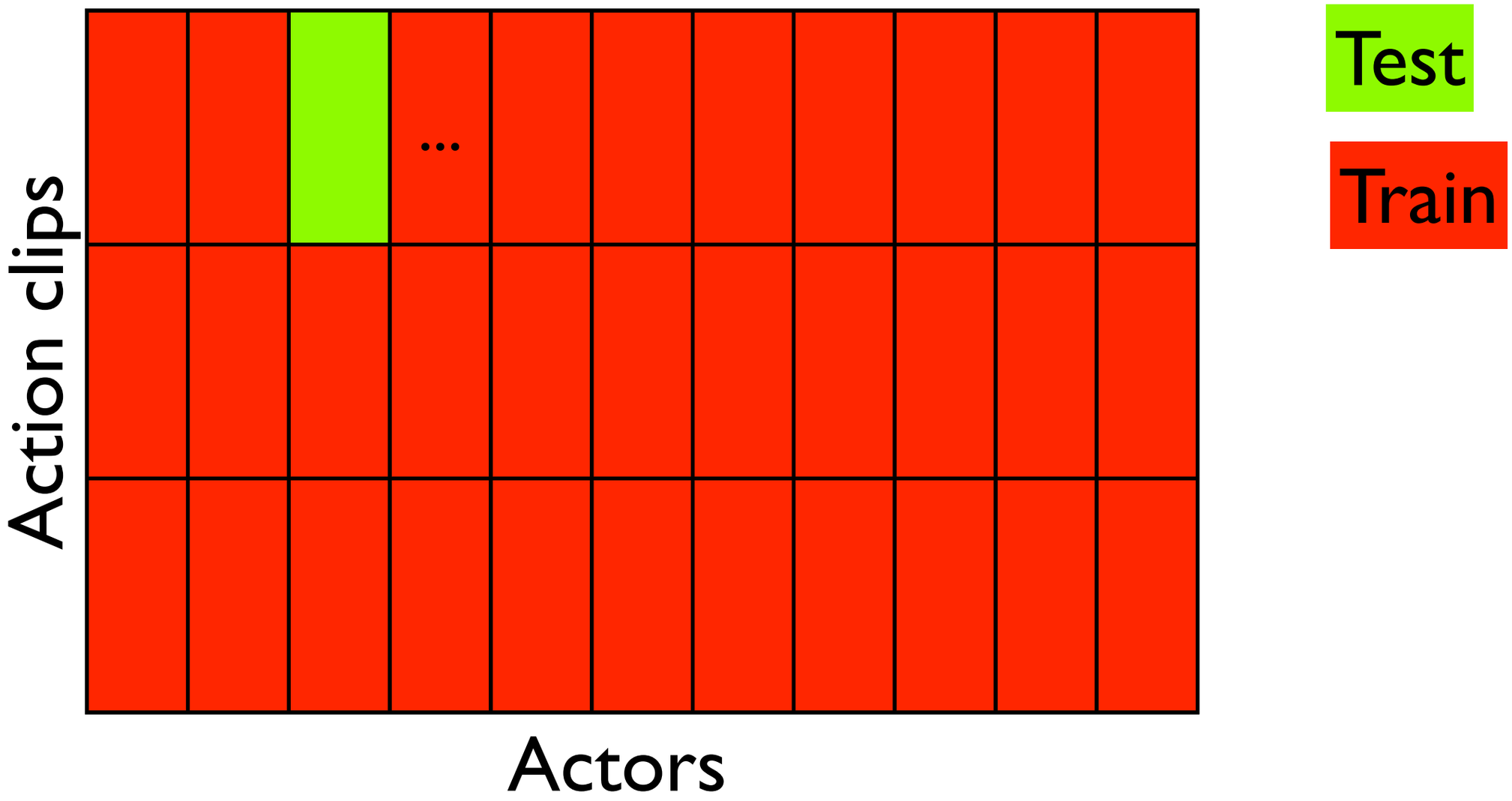
# Evaluation protocol:
# Leave One Sequence Out Cross Validation



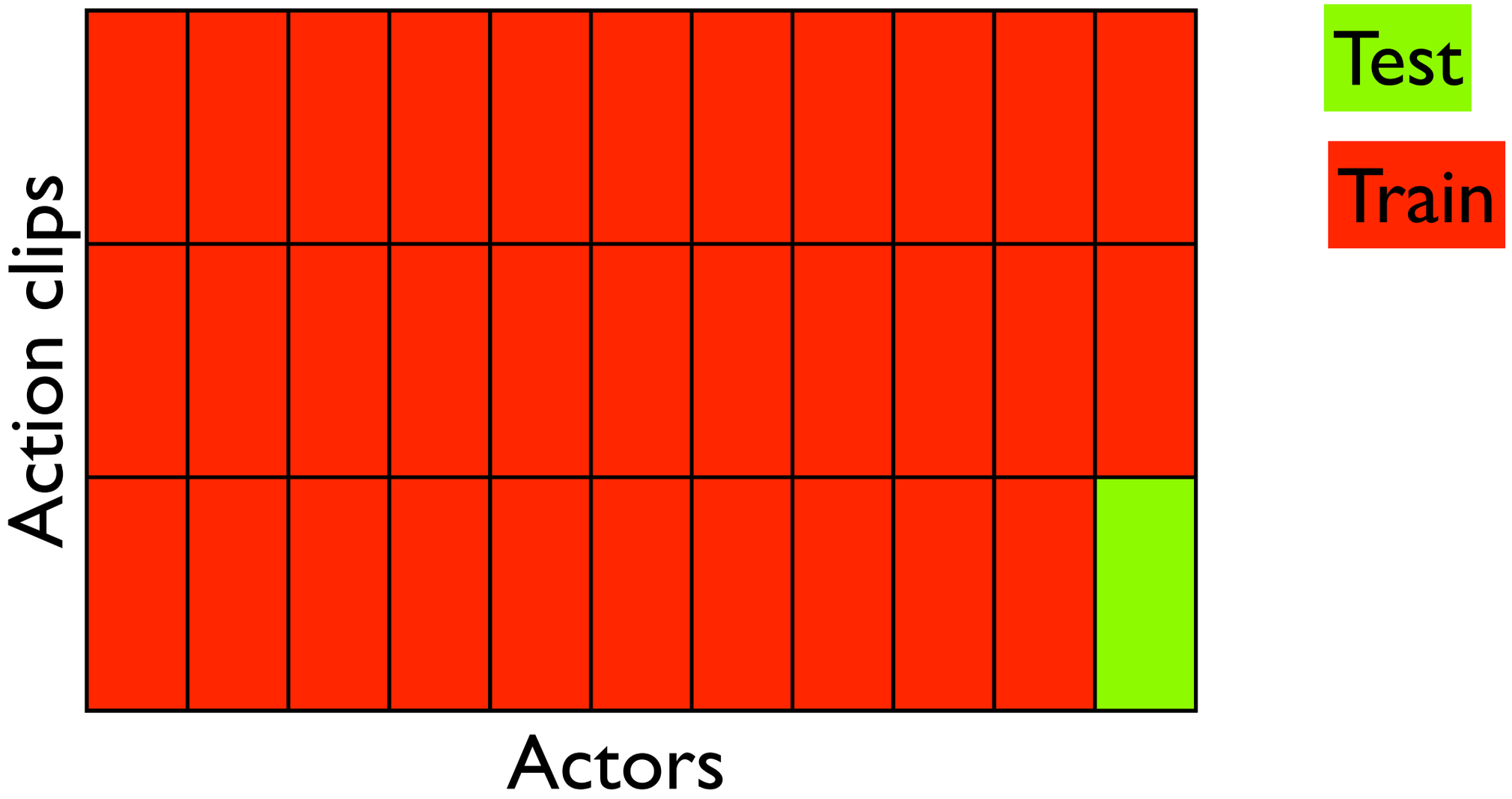
# Evaluation protocol:
# Leave One Sequence Out Cross Validation

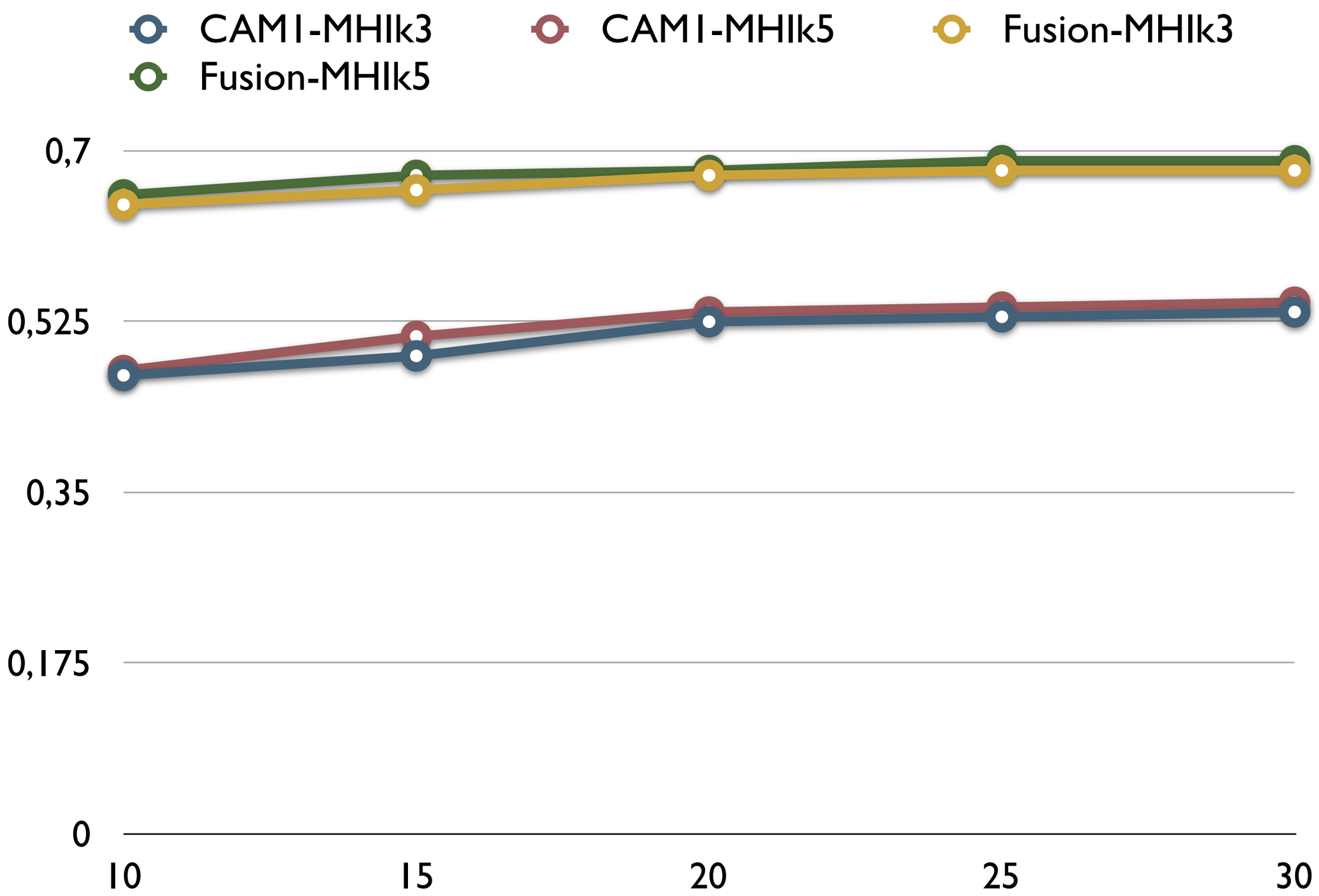


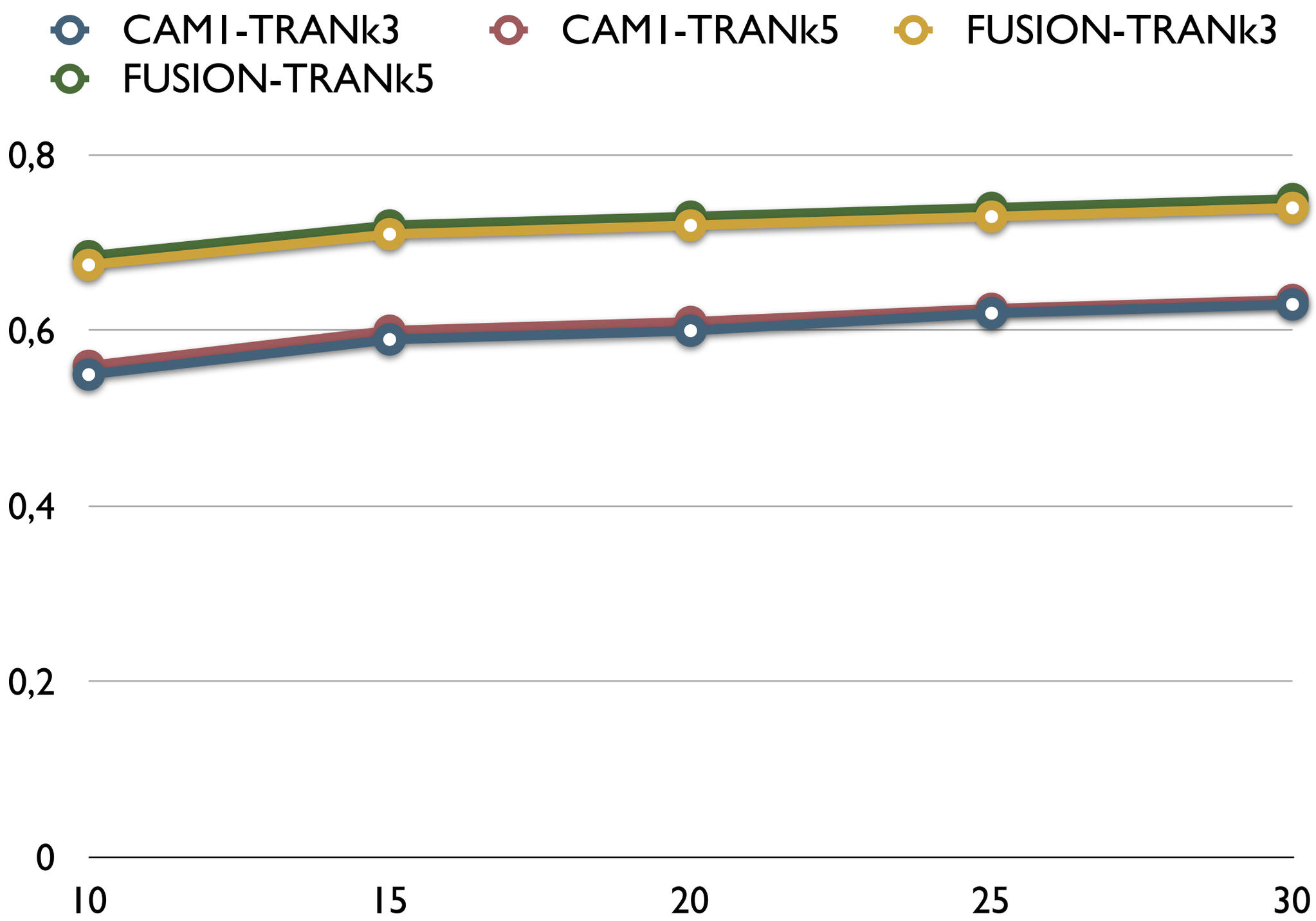
# Evaluation protocol:
# Leave One Sequence Out Cross Validation

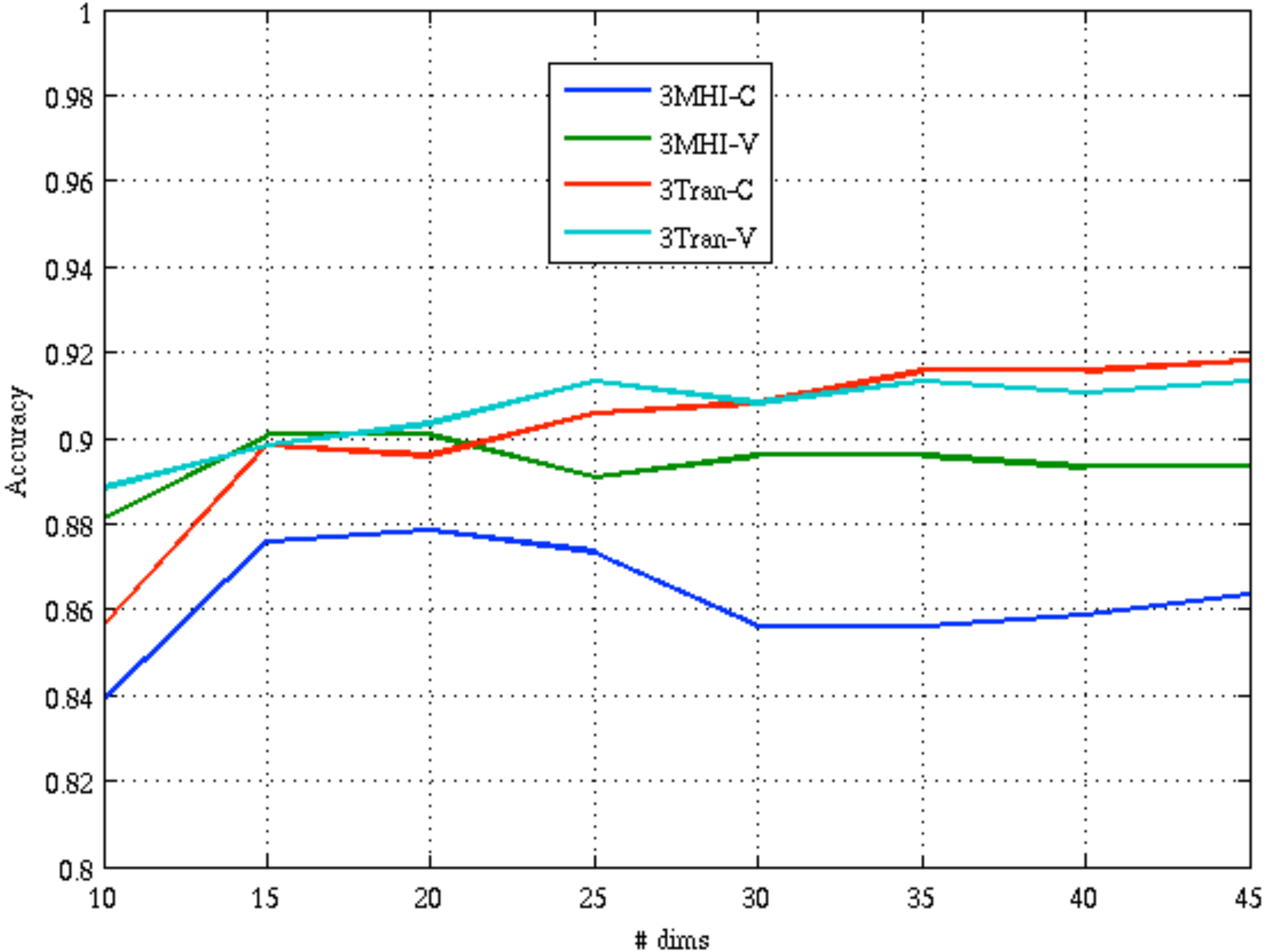


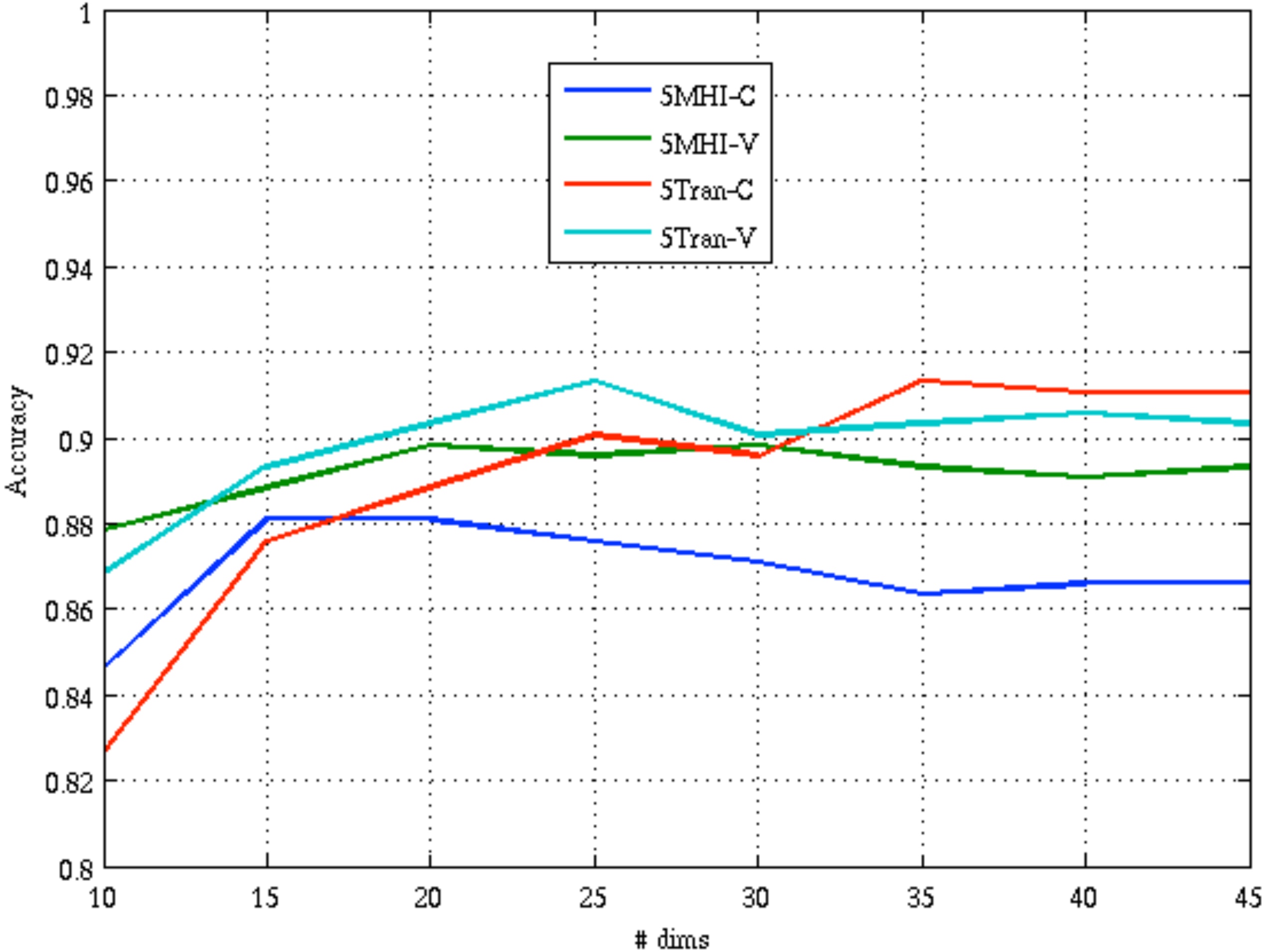
# Evaluation protocol:
# Leave One Sequence Out Cross Validation




CAM1-MHlk3
CAM1-MHlk5
Fusion-MHlk3
Fusion-MHlk5
0,7
0,525
0,35
0,175
0
10
15
20
25
30


CAM1-TRANk3
CAM1-TRANk5
FUSION-TRANk3
FUSION-TRANk5
0,8
0,6
0,4
0,2
0
10
15
20
25
30


3MHI-C
3MHI-V
3Tran-C
3Tran-V
Accuracy
# dims
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
10
15
20
25
30
35
40
45


1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
Accuracy
10
15
20
25
30
35
40
45
# dims
5MHI-C
5MHI-V
5Tran-C
5Tran-V

# V. Conclusions

Single view Classification

+

Fusion

=

State of the art results

# Distributed

# Future

Improve local classification / dimensionality reduction

Future

View Invariance

# Future

## Anthropometry Invariance

# Future

## Multiple People / Data Association

Thank you!

# Questions?

This work was supported in part by Projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/TIC-1485) and DPS2008-07029-C02-02