Learning the Manifolds of Local Features and Their Spatial Arrangement

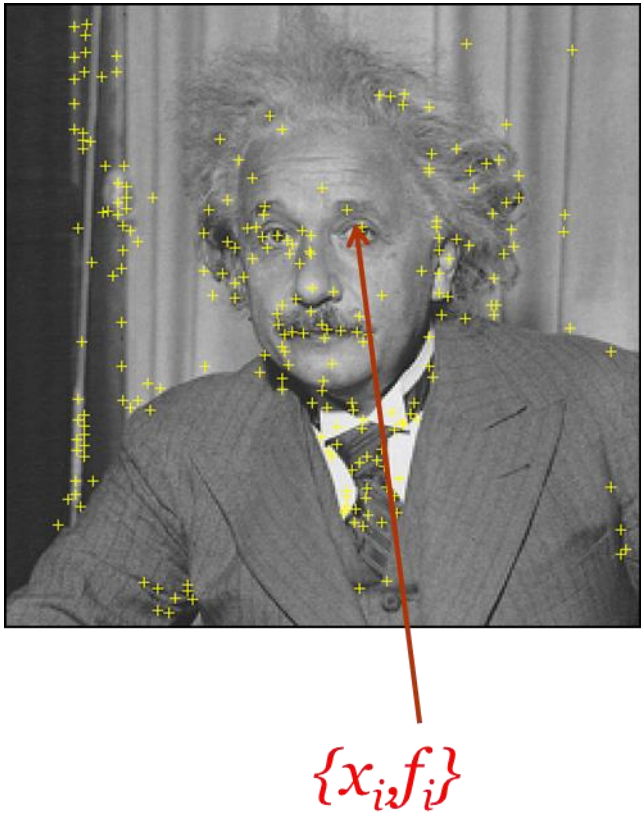
By: Marwan A. Torki
Rutgers University
PHD Defense June 13th
mtorki@cs.rutgers.edu

Outline

- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

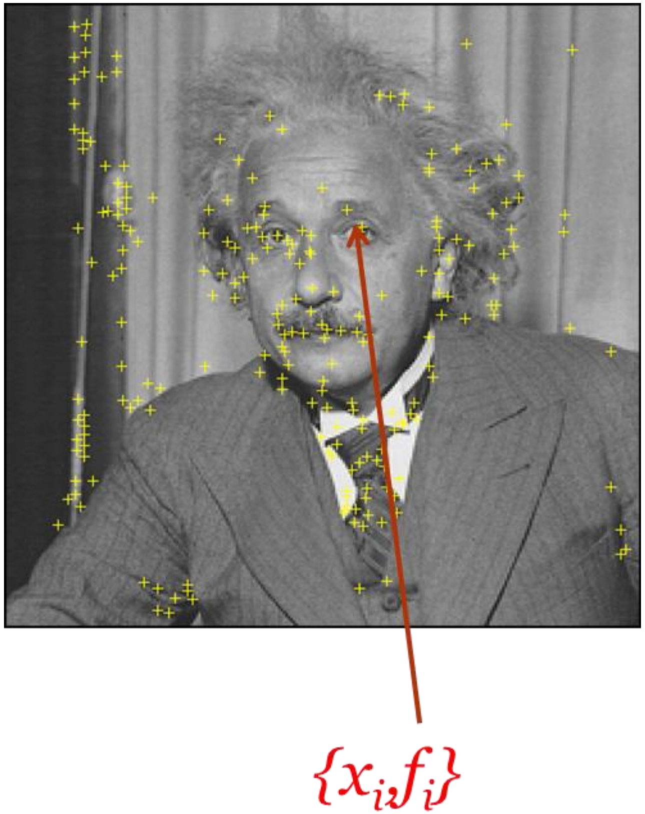
Local Features

- An interest point (local feature) is a point in the image which in general can be characterized as follows:
  - Invariant: invariant to scale, rotation, affine, illumination and noise for robust matching across different imaging conditions.
  - Distinctive: a single feature should be rich in information that can be matched with high probability.
  - Repeatable: produced whenever it appears.



Local Features

- Every local feature is having:
  - Feature Descriptor f_i and spatial location x_i
- Feature descriptor usually describes the appearance of the local feature.
  - SIFT (Lowe 99)
  - Geometric Blur(Berg et al. 01)
  - HOG (Dalal et al. 05)
  - ,…etc.



Applications of Local Features

- Local features are the core of the current state-of-the-art research in object recognition for the last decade.
  - Image retrieval.
  - Object classification.
  - Object detection.
  - Unsupervised category discovery.
  - Scene understanding.
  - …,etc

Spatial Structure for Shape Representation

- The spatial structure, or the arrangement of the local features plays an important role in perception since it encodes the shape.

The role of Spatial Arrangement

- No Spatial Structure
  - Bag-of-visual words model (Csurka et al. 2004)
- Spatial Partitioning
  - Pyramid Matching Kernel (Grauman et al. 05).
  - Spatial Pyramid of Histograms (Lazebnik et al. 06).
- Part Models
  - Constellation model (Weber et al 00).
  - Pictorial Structure (Felzenszwalb et al 05).
- Descriptor Free (shape only)
  - Point based matching approaches (Scott and Longuett-Higgins91).

Graph-Based Methods for Manifold Learning

- Manifold Assumption
  - The high-dimensional data lie on a low-dimensional manifold.
- Dimensionality reduction methods are examples of graph based methods
  - Laplacian EigenMaps(Belkin et al 2001)
  - LLE (Roweis et al 2000)
  - …,etc.
- Representing data as graph nodes.
- Edges are labeled with the pairwise distances of the incident nodes.
- Approximates the geodesic between two points with respect to the manifold of data points.

Manifold Learning in Recognition Context

- Eigenface for face recognition (M. Turk and A. Pentland 1991).
- Linear Dimensionality reduction using PCA to learn appearance manifolds (Murase et al 95).
- Dimensionality reduction in activity recognition.

Holistic Representations

- Using holistic approaches; image as a vector
  - Silhouettes
  - Whole image
- Full correspondence of point sets (Land mark based)
  - Active appearance model (Cootes et al. 1998).

Learning the Manifolds of Local Features and Their Spatial Arrangement

- Goals
  - From collections of local features from different images we want to
    - Learn a smooth manifold capturing appearance and spatial arrangement of local features .
    - Model within-class variations and object view manifolds.
    - Address computer vision problems like recognition, detection, regression and matching.

Local Features and Manifold Structures

- Appearance described by descriptors is having a manifold structure.
  - Descriptors of similar feature should be close to each other in the descriptor space.
- Spatial configuration of the local features gives another manifold structure.
  - Spatially neighboring features describe the shape of an object.
- Putting both manifold structures together is important to gain strengths from both structures.

Outline

- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

Contributions

- We propose a novel framework for learning manifold representations from local features and their spatial arrangement in a smooth way to achieve a feature-spatial embedding.
- We learn a manifold representation of the images that is suitable for recognition tasks.

Contributions

- We propose a novel solution for regression based on sets of local features.
- We propose novel solution for matching multiple sets of local features with scalable approach.
- We propose to deal with the joint structure implicitly for recognition tasks.

Outline

- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

Elements of Graph-Based Methods

- A graph $g = (V, E)$with set of vertices V and real edge weight E.
- $\boldsymbol{W}$ is the weighted adjacency matrix of g
- $\boldsymbol{W}_{ij} = \begin{cases} \omega(e), & e \in E(i,j) \\ 0, & e \notin E(i,j) \end{cases}$
- The diagonal matrix $\boldsymbol{D}$ defined by $\boldsymbol{D}_{ii} = \sum_j \boldsymbol{W}_{ij}$is the degree matrix of g.
- The normalized graph Laplacian$\hat{L}$and the unnormalized Graph Laplacian$L$ are defined:
  - $\hat{\boldsymbol{L}} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$
  - $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$

Framework for Learning Joint Feature-Spatial Embedding

- The goal is to learn an explicit representation for joint feature-spatial structure for local features in images for different tasks.
- Given K sets of feature points, $X^1, X^2, \ldots X^K$ in K images where $X^K = \{ \left(x_1^k, f_1^k\right), \ldots, (x_{Nk}^k, f_{Nk}^k) \}$
- Each feature point $\left(x_i^k, f_i^k\right)$ has Spatial location $x_i^k \in \mathbb{R}^2$and its feature descriptor $f_i^k \in \mathbb{R}^D$

Torki and Elgammal CVPR10a

Framework for Learning Joint Feature-Spatial Embedding

- The intra-image spatial (structure) in each image can be represented by a weight matrix S^k.
  - $S_{ij}^k = K_s(x_i^k, x_j^k)$and $K_s(.,.)$ is a spatial kernel local to the $k - th$ image that measures the spatial proximity.
- The inter-image feature affinity between features in image p and q can be represented by the weight matrix U^{pq}.
  - $U_{ij}^{pq} = K_f\left(f_i^p, f_j^q\right)$ and $K_f(.,.)$is a feature kernel that measures the similarity in the descriptor domain between the $i - th$ feature in image p and the $j - th$ feature in image q.

Framework for Learning Joint Feature-Spatial Embedding

- The embedded representation of local features should capture both kinds of affinities.
- Let $y_i^k \in \mathbb{R}^d$ denotes the embedding coordinate of point $\left(x_i^k, f_i^k\right)$
- We are seeking a set of embedded point coordinates $Y^k = \{y_1^k, \cdots, y_{Nk}^k\}$ for each input feature set X^k.

Framework: Objective Function

- $\Phi(Y)$ $= \sum_k \sum_{i,j} \left(y_i^k - y_j^k\right)^2 S_{ij}^k + \sum_{p,q} \sum_{i,j} \left(y_i^p - y_j^q\right)^2 U_{ij}^{pq}$
- The first term preserves the spatial arrangement within each set.
- The second term of the objective function tries to bring close the embedded points y_i^p and y_j^q if their feature similarity kernel U_{ij}^{pq} is high.

Framework: Objective Function

- Rewrite the objective function
  - $\Phi(Y) = \sum_k \sum_{i,j} \left(y_i^k - \mathrm{y}_{\mathrm{j}}^k\right)^2 S_{ij}^k + \sum_{p,q} \sum_{i,j} \left(y_i^p - y_j^q\right)^2 U_{ij}^{pq}$
- Use one sets of weights
  - $\Phi(Y) = \sum_{p,q} \sum_{i,j} \left(y_i^p - y_j^q\right)^2 A_{ij}^{pq}$
- Where the matrix A is defined as
- $$A_{ij}^{pq} = \begin{cases} S_{ij}^k & p = q = k \\ U_{ij}^{pq} & p \neq q \end{cases}, A = \begin{pmatrix} S^1 & U^{12} & \vdots & U^{1K} \\ U^{21} & S^2 & & \\ \cdots & & \ddots & \\ U^{K1} & & \cdots & S^K \end{pmatrix}$$
- The |A| is linear in the number of input points.

Framework: Objective Function

- The problem reduces to Laplacian Embedding of the point set defined by the weight matrix A.
- The Solution is $\boldsymbol{Y}^* = arg \min_{\boldsymbol{Y^T D Y = I}} tr(\boldsymbol{Y^T L Y})$
- Where $\boldsymbol{L}$ is Laplacian of matrix $\boldsymbol{A}$.
- Minimizing this objective function is a straight forward generalized eigenvector problem:
  - $\boldsymbol{L}y = \lambda \boldsymbol{D}y.$
- The $N \times d$ matrix $\boldsymbol{Y}$ is the stacking of the desired embedding coordinates such that
- $Y = [y_1^1, \cdots, y_{N1}^1, y_1^2, \cdots, y_{N2}^2, \cdots, y_1^K, \cdots, y_{NK}^K]$.

Framework: Spatial Structure Weights S^k

- Euclidean-based weights:
  - Based on the Euclidean distances between features defined in each image coordinate system.
  - Weights are invariant to translation and rotations.
  - Examples:
    - Gaussian Kernel $\boldsymbol{S}_{ij}^{k} = e^{-\left\|x_i^k - x_j^k\right\|^2 / 2\sigma^2}$
    - Double Exponential Kernel $\boldsymbol{S}_{ij}^{k} = e^{-\left\|x_i^k - x_j^k\right\| / \sigma}$

Framework: Spatial Structure Weights S^k

- Affine invariant-based weights:
  - A configuration matrix of the features in a given set $\boldsymbol{X} = [\,\mathrm{x}_1\ \mathrm{x}_2 \cdots\ \mathrm{x}_\mathrm{N}\,] \in \mathbb{R}^{N\times 3}$.
  - Where x_i is the homogenous coordinate of point x_i.
  - The range space of such configuration matrix is invariant under affine transformation (Wang et al 09).
- An affine representation can be achieved by $\boldsymbol{QR}$ decomposition of the projection matrix of $\boldsymbol{X}$.
  - $\boldsymbol{QR} = \boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}$
- The columns of $\boldsymbol{Q}$ give an affine invariant representation of the points.
- Kernels can be used on the computed affine representation using Gaussian or other kernel as before.

Framework: Feature Similarity Weights U^{pq}

- Weights should be soft encoded.
- $\boldsymbol{U}_{ij}^{pq} = e^{-\left\|f_i^p - f_j^q\right\|^2 / 2\sigma^2}$
- Nearest neighbor can also be used to bound the weights as well

Outline

- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

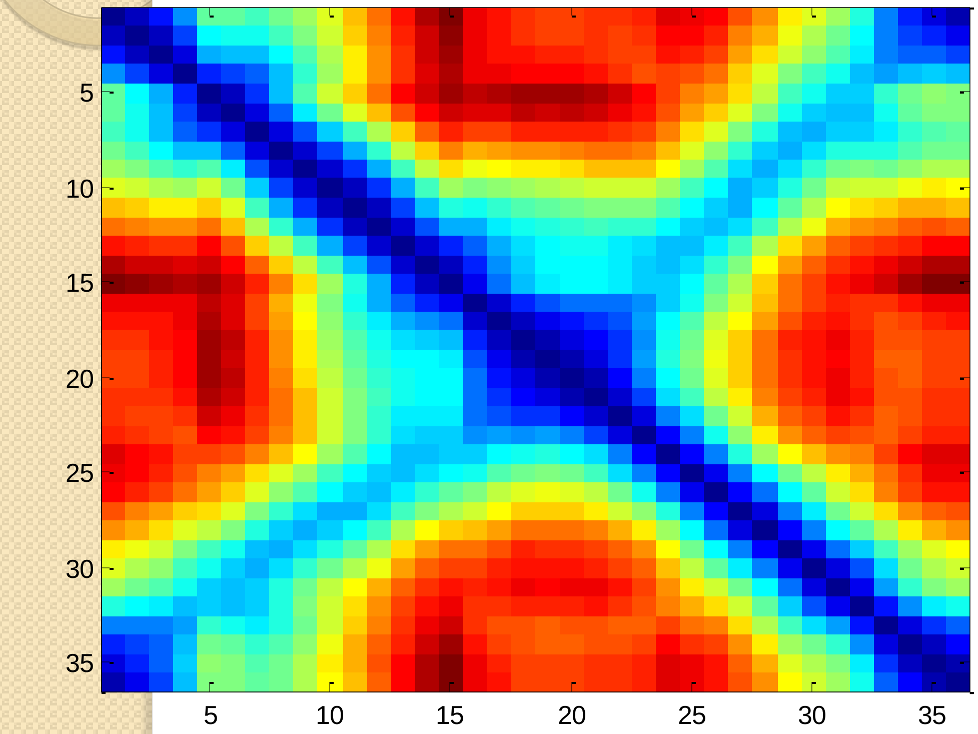
Learning Image Manifolds from Local Features

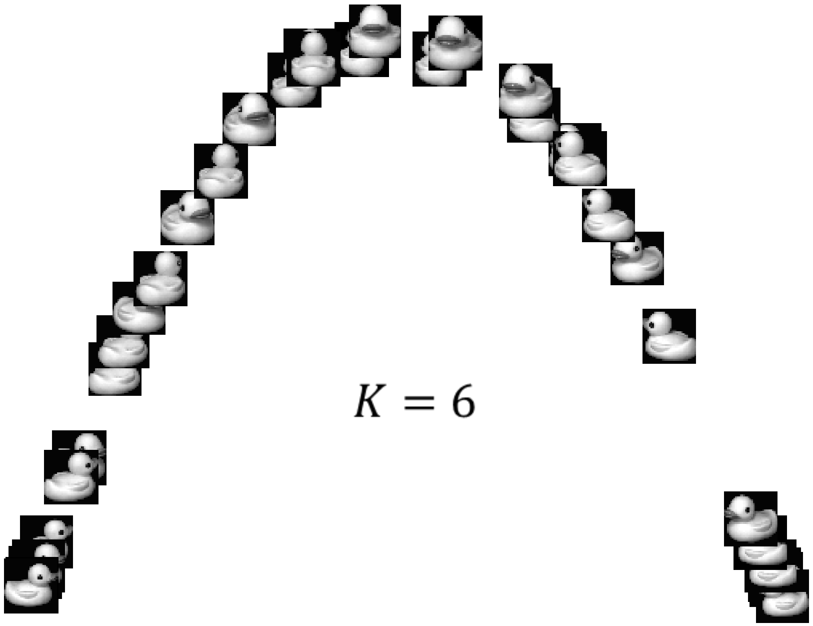
- Manifold Learning From Local Features
  - Current manifold learning methods use holistic representation like whole images or silhouettes to learn visual manifolds
  - Using our framework we can replace the holistic representations representation by local features extracted.
- By embedding a bulk of images using the proposed feature-spatial embedding we can compute image to image distance .
- The distance matrix can be used to learn image manifolds for recognition tasks.

Image Embedding

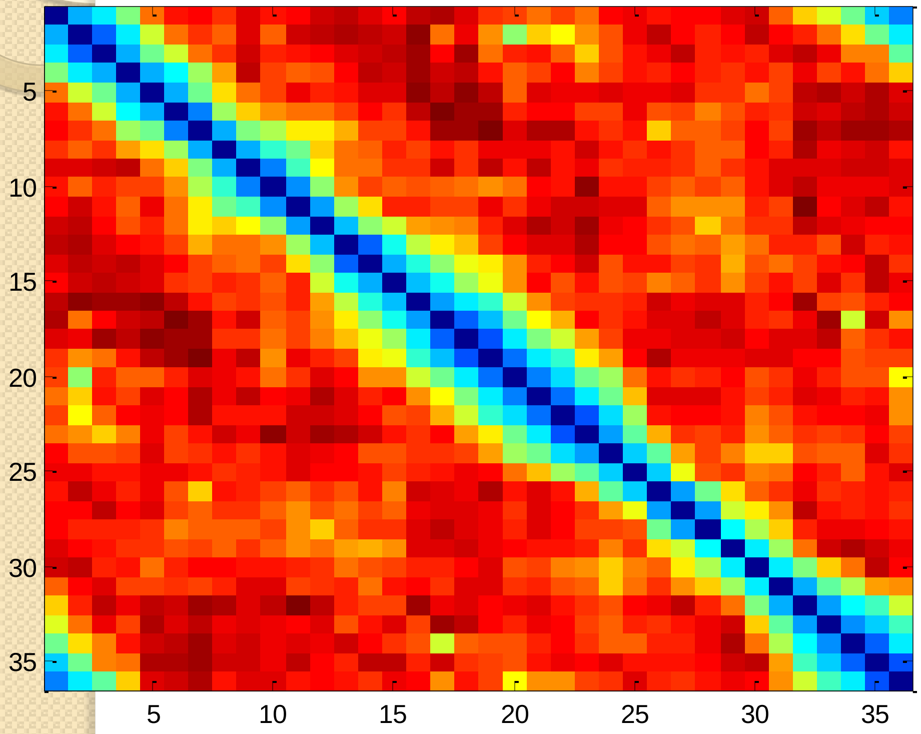
- Measure the similarity between two images in the feature-spatial space.
- For robustness, we use a percentile-based Hausdorff similarity based measure between two sets of features from two images
- $\mathbf{H}(\mathbf{X}^{\mathrm{p}}, \mathbf{X}^{\mathrm{q}})$ $= \max\{\max_j^{l\%} \min_i \left\| y_i^p - y_j^q \right\|, \max_i^{l\%} \min_j \left\| y_i^p - y_j^q \right\|\}$
- Once a distance measure between images is defined, any manifold embedding techniques, such as MDS ,LLE, Laplacian Eigen maps, etc., can be used to achieve an embedding of the image manifold.

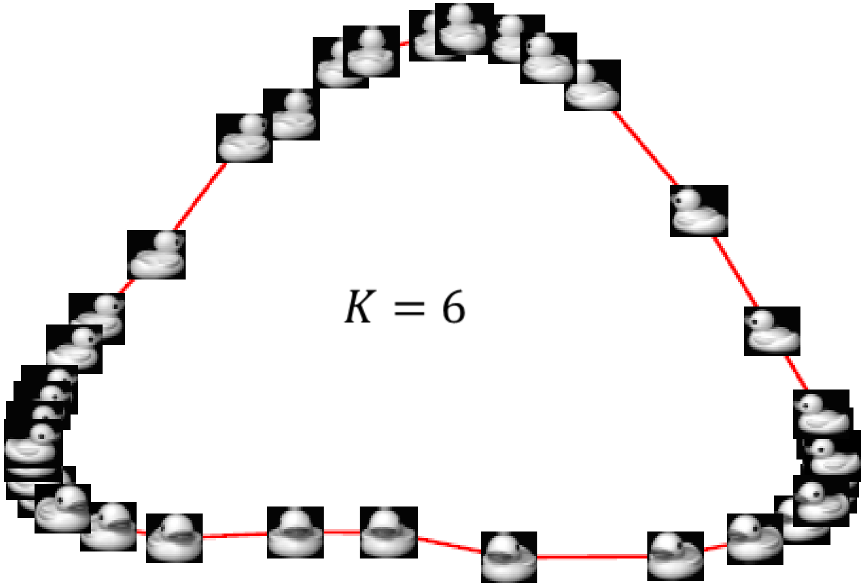
Sample View Manifold Using Image Intensity from COIL-20 (Murase et al 95)



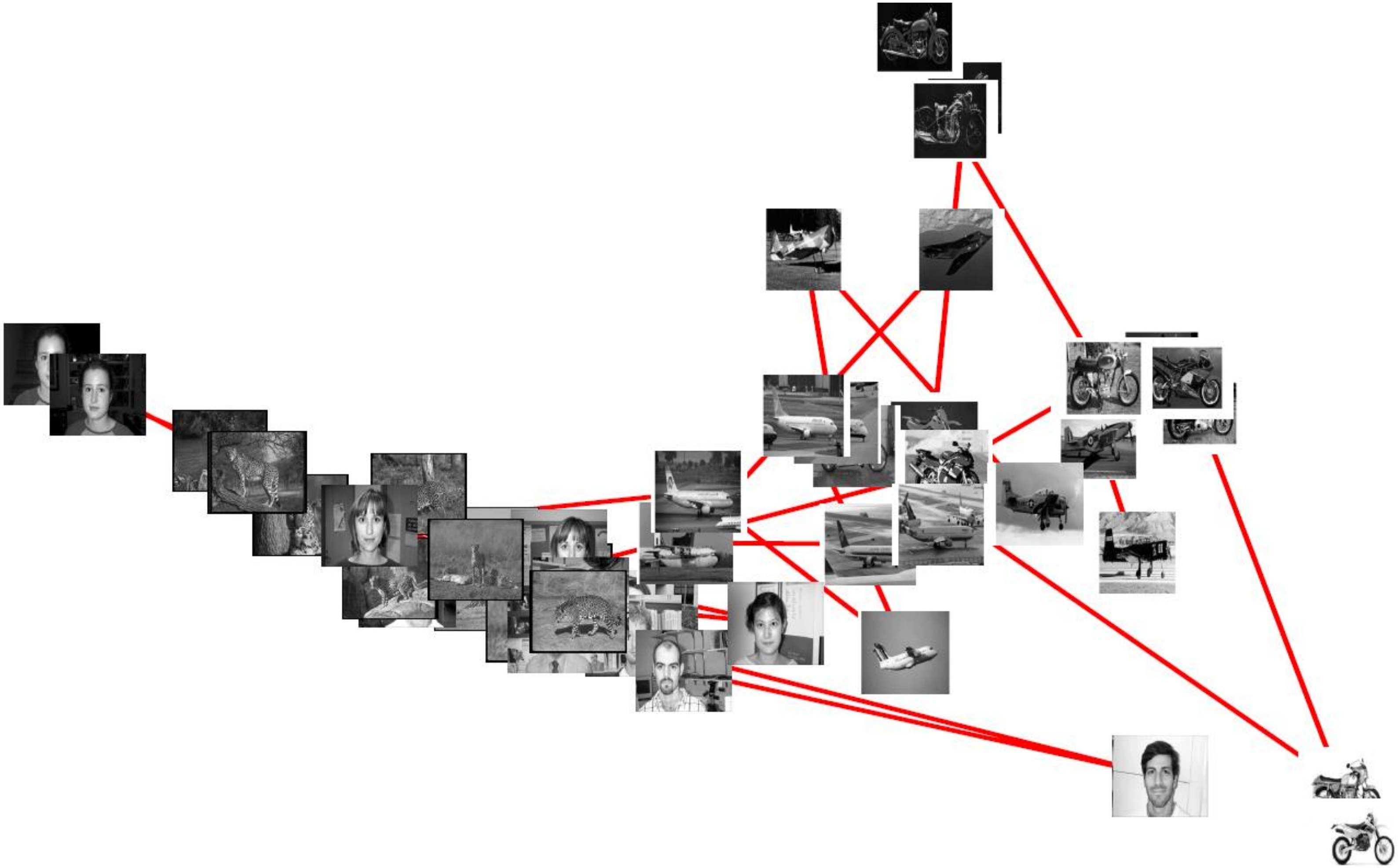

$K = 6$

Sample View Manifold Using Our Framework from COIL-20

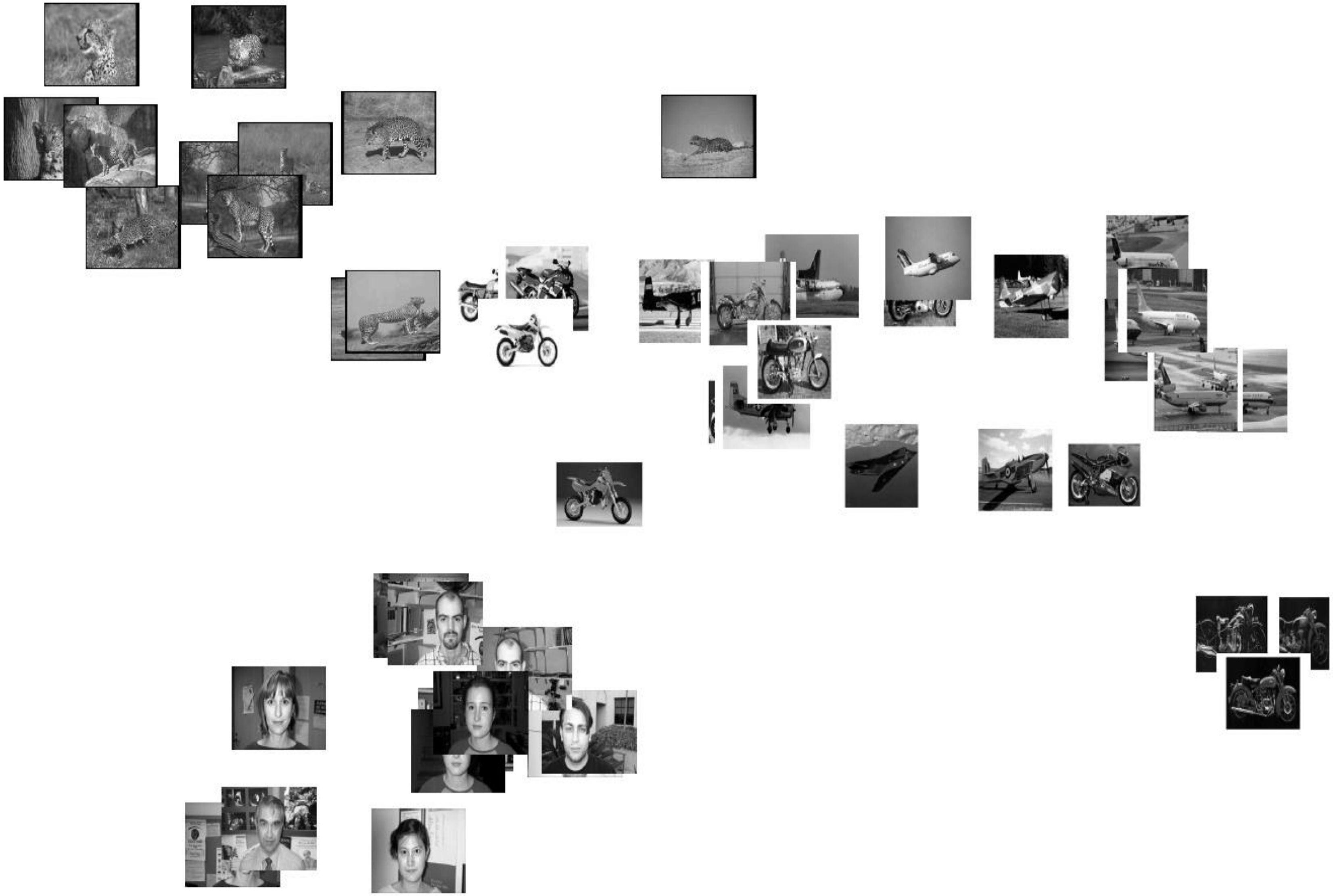




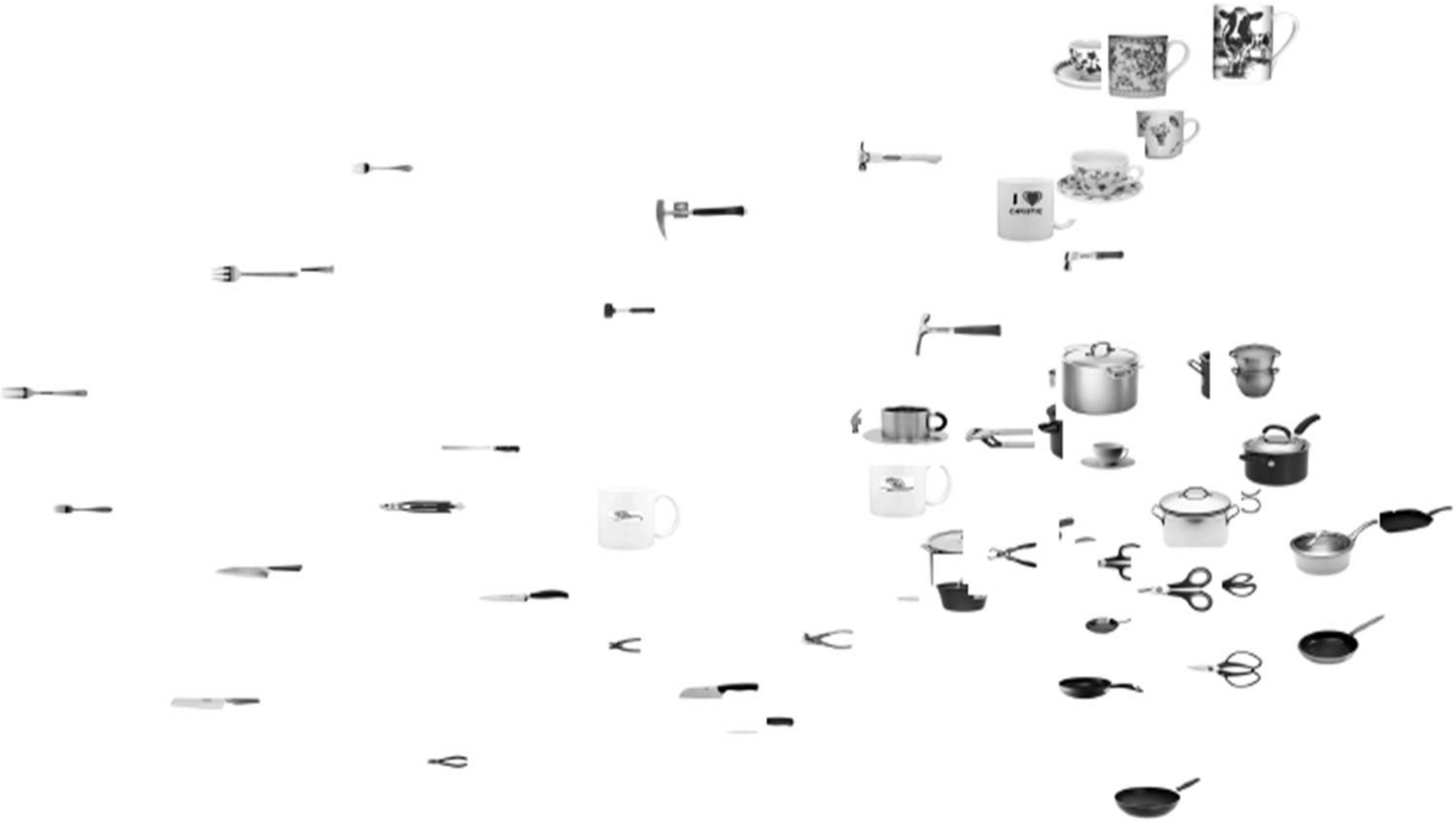
Sample View Manifold Using Image Intensity from Caltech-101(Li et al 04)



Sample View Manifold Using Our Framework from Caltech-101



Example Image Embedding from Shape Dataset of Stark et al 07



Embedding captures shape similarity as well as appearance similarity

Out-of-Sampling for Features from New Image Train/Test

- The out-of-sample is essential to
  - Embed large number of images with large number of features.
  - Embed features from a new image for classification purpose.

Out-of-Sampling Solution

- For every new image instance compute extended affinity matrix A.
- $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}^{\tau} & \boldsymbol{U}^{\nu^{T}} \\ \boldsymbol{U}^{\nu} & \boldsymbol{S}^{\nu} \end{pmatrix} \boldsymbol{U}^{\nu} = [\boldsymbol{U}^{\nu,1}, \boldsymbol{U}^{\nu,2}, \cdots, \boldsymbol{U}^{\nu,K}]$
- The objective function we use for embedding new points is
- $\boldsymbol{Y}^{*} = arg \min tr(\boldsymbol{Y}^{\boldsymbol{T}} \boldsymbol{L} \boldsymbol{Y})$
- $s.t\, y_i^k = \hat{y}_i^k, i = 1, \cdots Nk\ and\ k = 1, \cdots K$
- $\boldsymbol{Y}^{\nu} = (\boldsymbol{L}^{\nu})^{-1} \boldsymbol{U}^{\nu} \boldsymbol{Y}^{\boldsymbol{\tau}}$

Feature-Spatial Embedding Framework for Object Classification.

- Putting the features from different images in the same embedding space.
  - Very problematic because the size of the eigenvector problem will increase rapidly with the number of features in the datasets.
  - Solution is to use two step approach
    - Initial Embedding.
    - Populating Embedding.

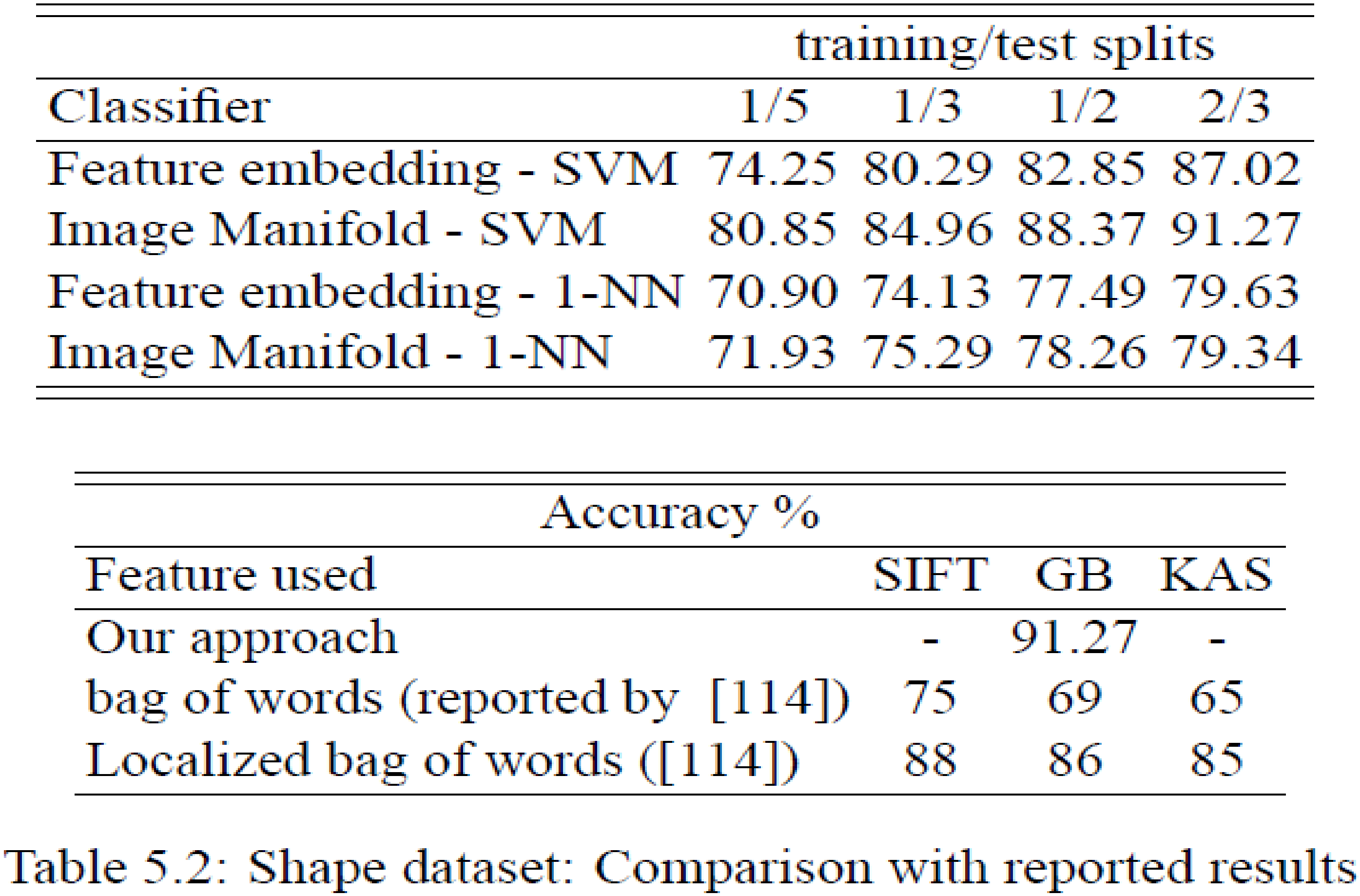
Populate Embedding

- Embed the whole training data with a larger number of features per image, one image at a time by solving an out-of-sample problem using the initial embedding solution.

Results for Object Classification

- Shape Dataset
  - 10 classes
  - 724 Images
  - Comparative evaluation to baseline of bag of words and localized bag of words. With different splits.
- Caltech 4I, 4II, 6 subsets from Caltech-101
  - Larger datasets with clutter.
  - Compare SVM to 1-NN classifiers for different training sizes.

Shape Dataset

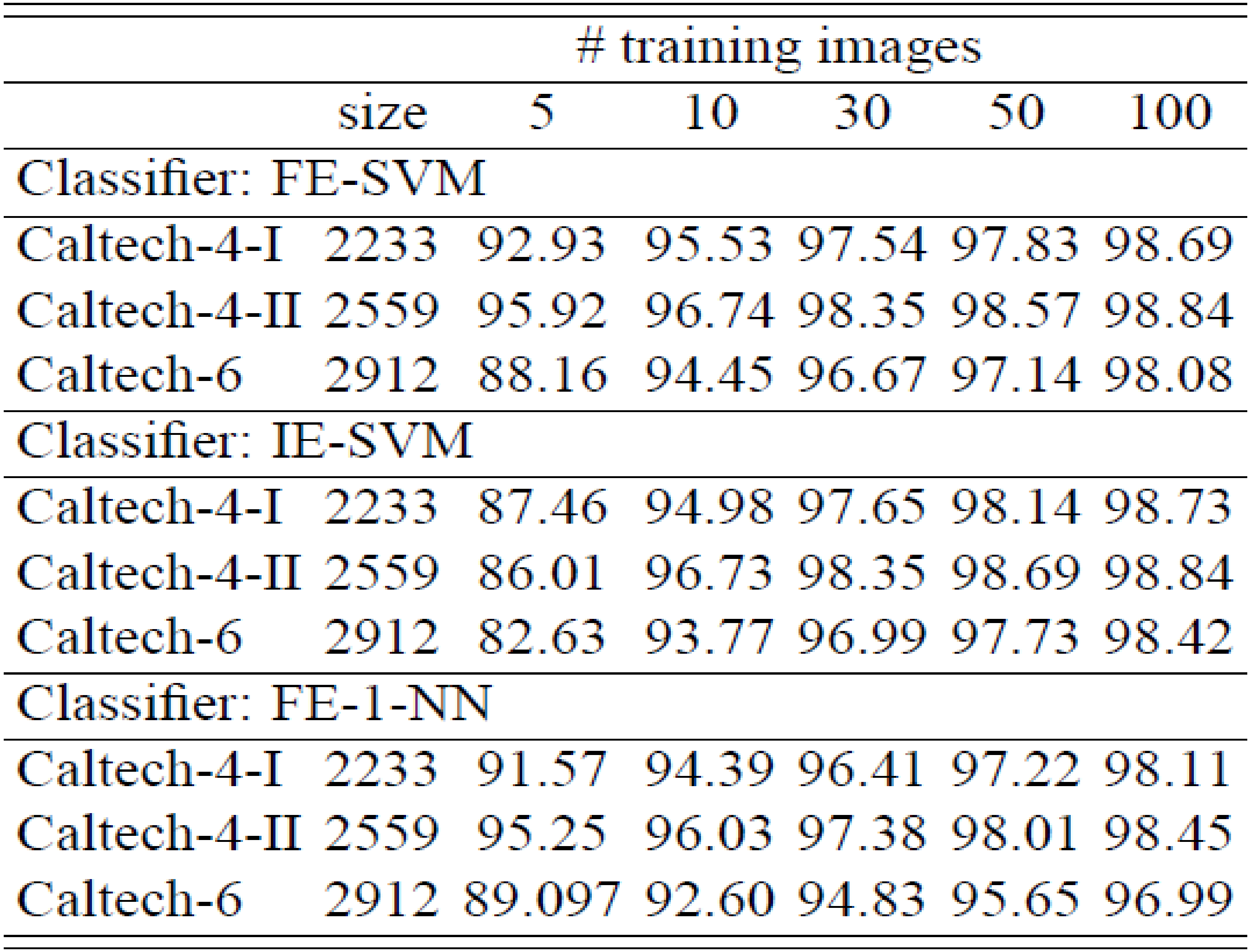


	training/test splits			
Classifier	1/5	1/3	1/2	2/3
Feature embedding - SVM	74.25	80.29	82.85	87.02
Image Manifold - SVM	80.85	84.96	88.37	91.27
Feature embedding - 1-NN	70.90	74.13	77.49	79.63
Image Manifold - 1-NN	71.93	75.29	78.26	79.34

Accuracy %			
Feature used	SIFT	GB	KAS
Our approach	-	91.27	-
bag of words (reported by [114])	75	69	65
Localized bag of words ([114])	88	86	85

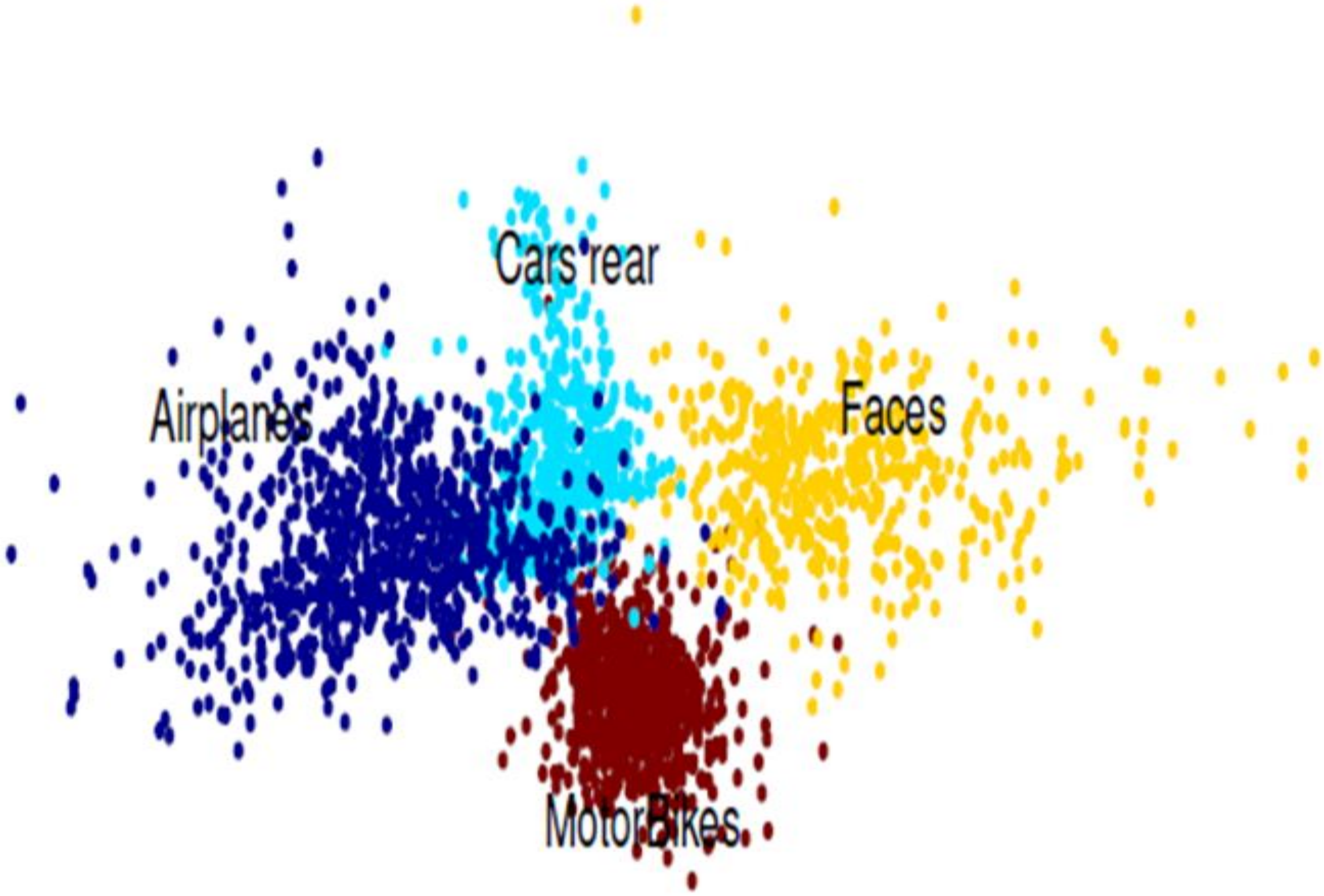
Table 5.2: Shape dataset: Comparison with reported results

Caltech subsets

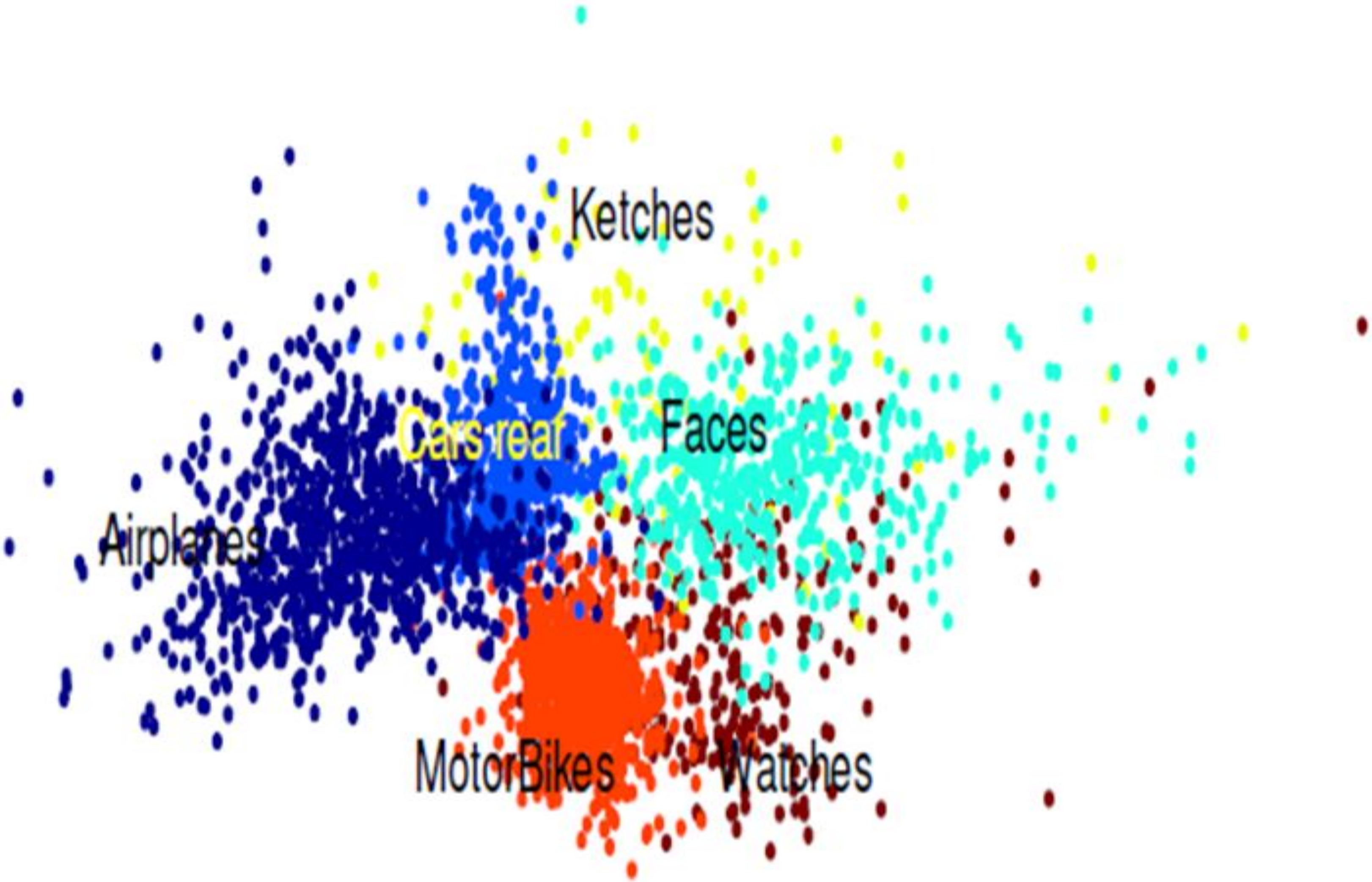


		# training images				
	size	5	10	30	50	100
Classifier: FE-SVM						
Caltech-4-I	2233	92.93	95.53	97.54	97.83	98.69
Caltech-4-II	2559	95.92	96.74	98.35	98.57	98.84
Caltech-6	2912	88.16	94.45	96.67	97.14	98.08
Classifier: IE-SVM						
Caltech-4-I	2233	87.46	94.98	97.65	98.14	98.73
Caltech-4-II	2559	86.01	96.73	98.35	98.69	98.84
Caltech-6	2912	82.63	93.77	96.99	97.73	98.42
Classifier: FE-1-NN						
Caltech-4-I	2233	91.57	94.39	96.41	97.22	98.11
Caltech-4-II	2559	95.25	96.03	97.38	98.01	98.45
Caltech-6	2912	89.097	92.60	94.83	95.65	96.99

Caltech 4 Image Embedding after Out-of-Sampling for All Features in All Images



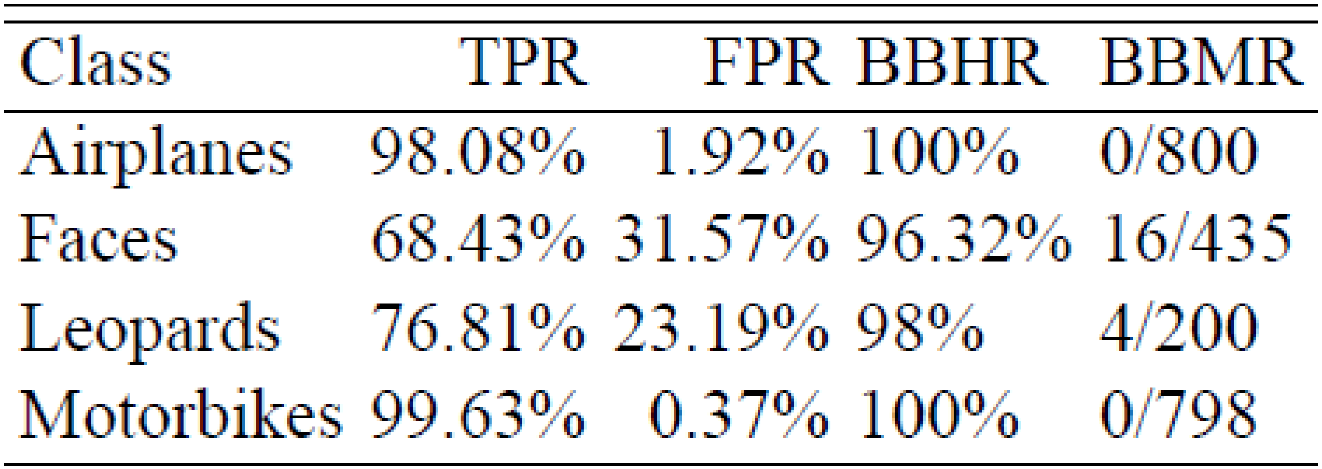
Caltech-6 Image Embedding in 2D



Feature localization on Caltech 41

- We used Caltech-41 subset for evaluation.
- We learned the feature embedding from four classes, using only 12 images per class.
- For evaluation we used 120 features in each query image and embed them by out-of-sample. The object is localized by finding the top 20% features closer to the training data

Feature localization on Caltech 4I



Class	TPR	FPR	BBHR	BBMR
Airplanes	98.08%	1.92%	100%	0/800
Faces	68.43%	31.57%	96.32%	16/435
Leopards	76.81%	23.19%	98%	4/200
Motorbikes	99.63%	0.37%	100%	0/798

Table 5.4: Object localization results - Caltech101-4

Outline

- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

Regression from Local Features

- Using feature-spatial Embedding framework for regression.
  - Without vectorized representation of images we obtain embedding as before.
  - The regression is achieved by defining a proper kernel in the embedding space.
  - Further we enforce manifold locality constraint on the embedding using label information in train set.

Torki and Elgammal ICCV11

Regression from Local Features

- Input is pairs in the form $(X^k, v^k), v^k \in \mathbb{R}$.
- learn a regularized mapping function to minimize a regularized risk criteria, which can be definedas

$$\sum_k \|g(X^k) - v^k\| + \lambda\Phi\,[g]$$

- Where the first term measured the error in the approximation, the second term is a smoothness function on g for regularization, and λ is a regularization parameter.
- We seek a regression in the form

$$v = \hat{g}(X) = \sum_j b_j K(X, X^j)$$

- Therefore, it suffices to define a suitable kernel K(·, ·) that measures the similarity between images.

Regression from Local Features

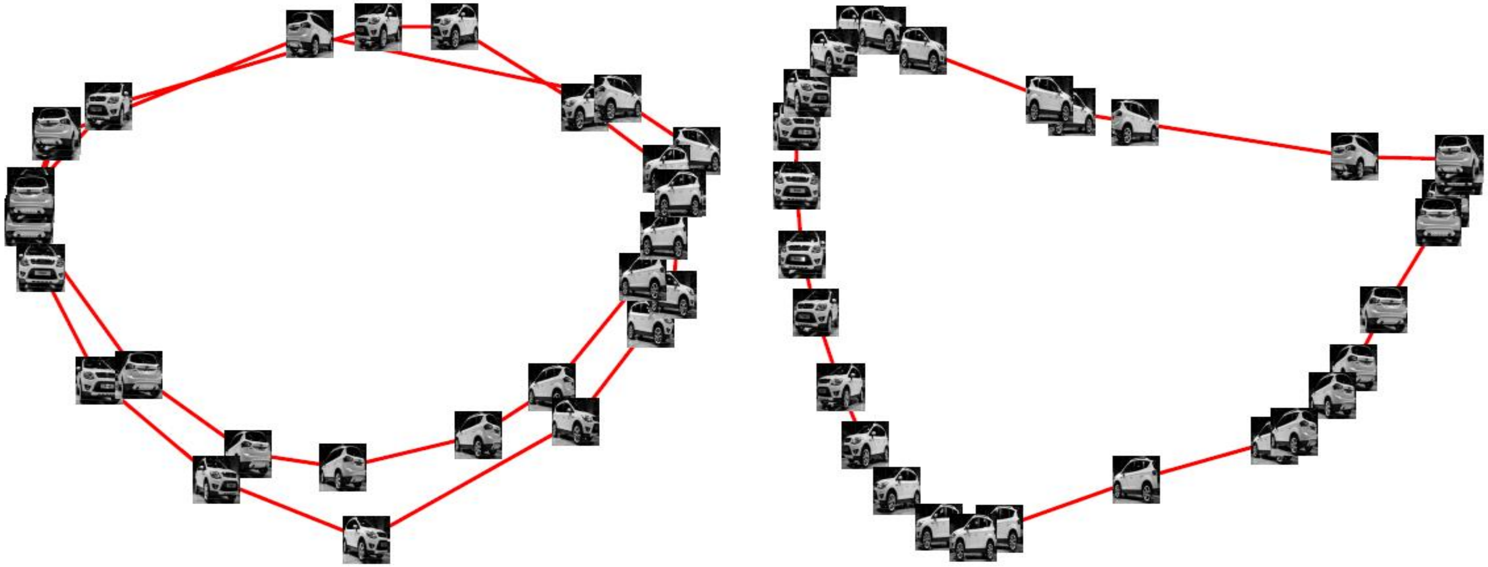
- $K(\mathbf{X}^{\mathrm{p}}, \mathbf{X}^{\mathrm{q}}) = \max\{\max_j^{l\%} \min_i \left\| y_i^p - y_j^q \right\|, \max_i^{l\%} \min_j \left\| y_i^p - y_j^q \right\|\}$
- Kernel is measured in the feature embedding space, so it reflects both feature similarity and shape similarity
- Radial Basis Function (RBF) kernels can be used.
- The features in new image has to be mapped first to the embedding space $v = \hat{g}(X) = \sum_j b_j K(\mathcal{O}(X), Y^j)$
- Where O(X) is a function that maps the features in a test image X into a set of coordinates in the embedding space,

$$\mathcal{O}(X) : \{(x_i, f_i)\} \longrightarrow \{y_i\}$$

- The out of sample solution described earlier used to obtain such a function.
- $\mathrm{O}(\boldsymbol{X}) = (\boldsymbol{L}^{\nu})^{-1}\boldsymbol{U}^{\nu}\boldsymbol{Y}^{\boldsymbol{\tau}}$

Enforcing Manifold Locality Constraint

- Feature-spatial embedding preserves Inter-image feature affinity and Intra image spatial structure.
- We add a third constraint that enforces manifold locality.
- Supervised Manifold Locality Constraint.



(a) Unsupervised Manifold Embedding

(b) Supervised Manifold Embedding

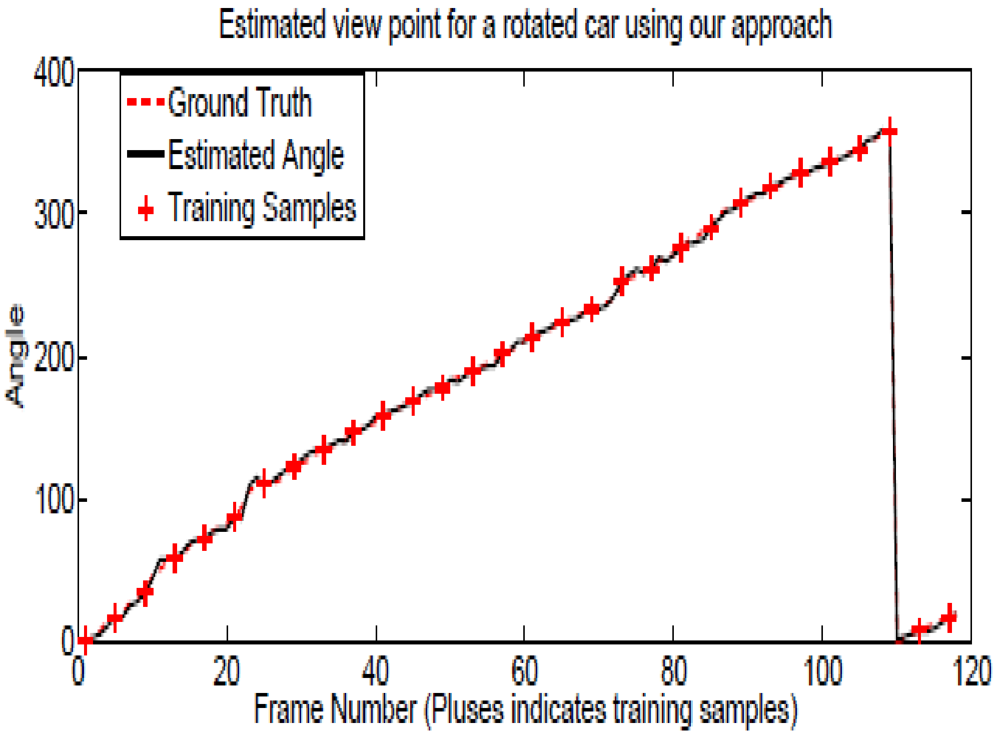
Enforcing Manifold Locality Constraint

- We can enforce the manifold constraint in a supervised way from the labels v^k.
- $\Phi(Y) = \sum_k \sum_{i,j} \left(y_i^k - y_j^k\right)^2 S_{ij}^k + \sum_{p,q} \sum_{i,j} \left(y_i^p - y_j^q\right)^2 \omega(p,q) U_{ij}^{pq}$
- $\omega(p,q)$=$g(v^p - v^q)$
- Gaussian function can be used or alternatively, a uniform window kernel
- $A_{ij}^{pq} = \begin{cases} S_{ij}^k & p = q = k \\ g(v^p - v^q).U_{ij}^{pq} & p \neq q \end{cases}$

Datasets for Evaluation

- Multiview Car Dataset Ozuysal et al 09
  - 20 cars rotating in car show.
  - Hard instances (some are really odd cars).
  - Suitable for regression.
  - Total 2137 images covering the range of view angle.
  - Comparisons to Ozuysal et al 09 where 16 bins of the view angle are used as 16 classifiers. They used spatial pyramid of histograms.

Multiview Car Dataset Single Car

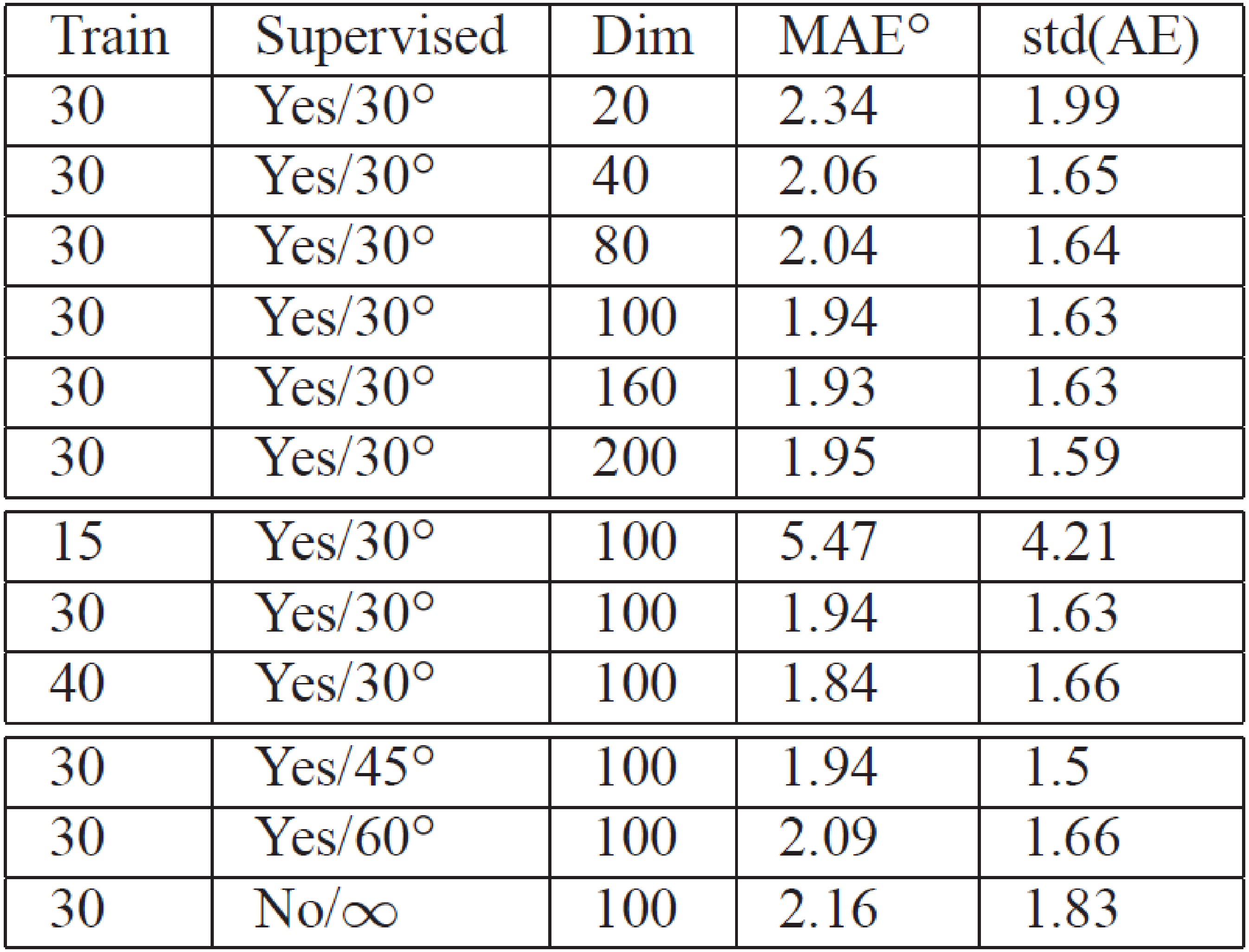


(a) Estimated Angle and Ground Truth

(b) Sample Views

Training samples are 12° apart
MAE < 2°

Multiview Car Dataset Single Car



Train	Supervised	Dim	MAE°	std(AE)
30	Yes/30°	20	2.34	1.99
30	Yes/30°	40	2.06	1.65
30	Yes/30°	80	2.04	1.64
30	Yes/30°	100	1.94	1.63
30	Yes/30°	160	1.93	1.63
30	Yes/30°	200	1.95	1.59
15	Yes/30°	100	5.47	4.21
30	Yes/30°	100	1.94	1.63
40	Yes/30°	100	1.84	1.66
30	Yes/45°	100	1.94	1.5
30	Yes/60°	100	2.09	1.66
30	No/∞	100	2.16	1.83

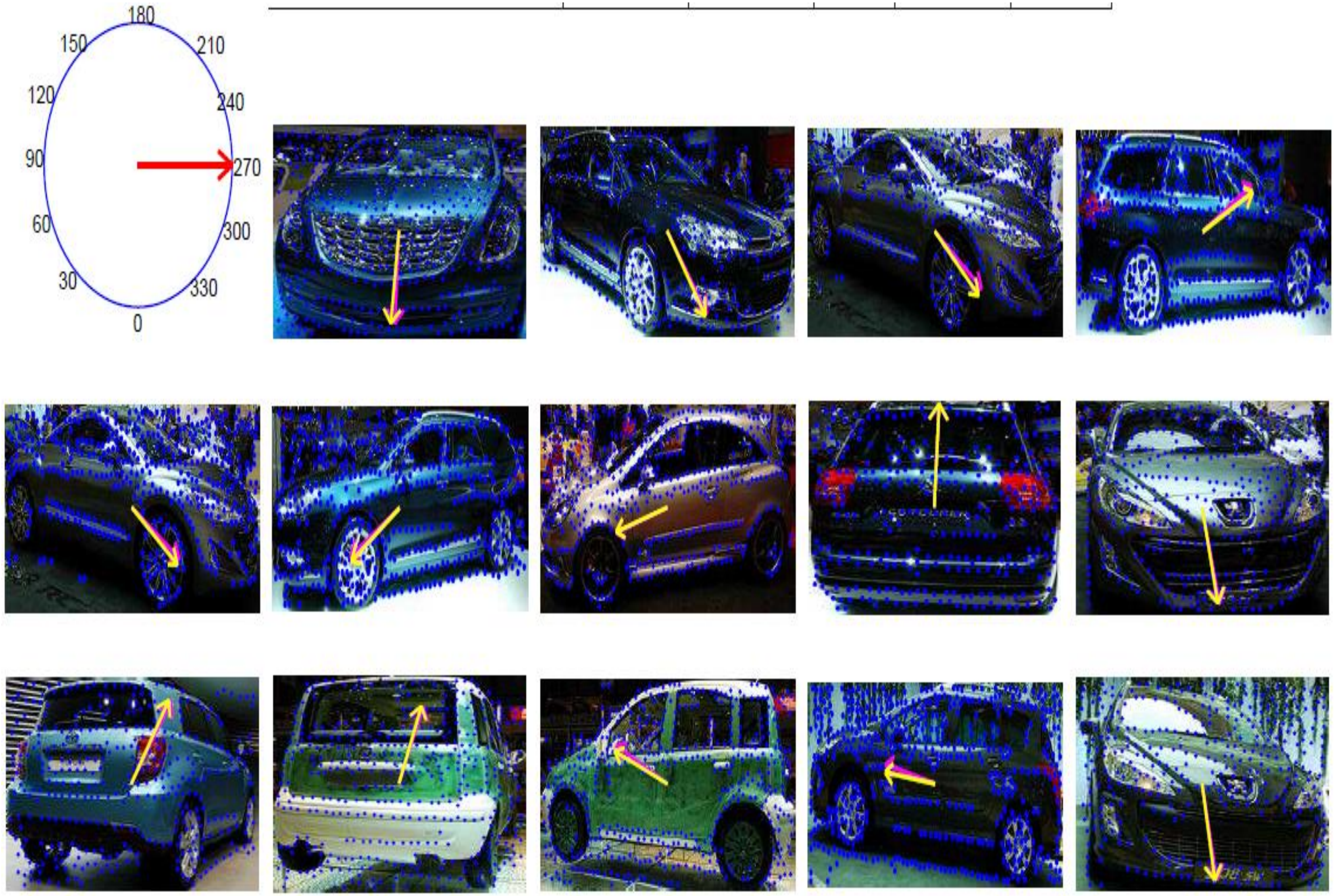
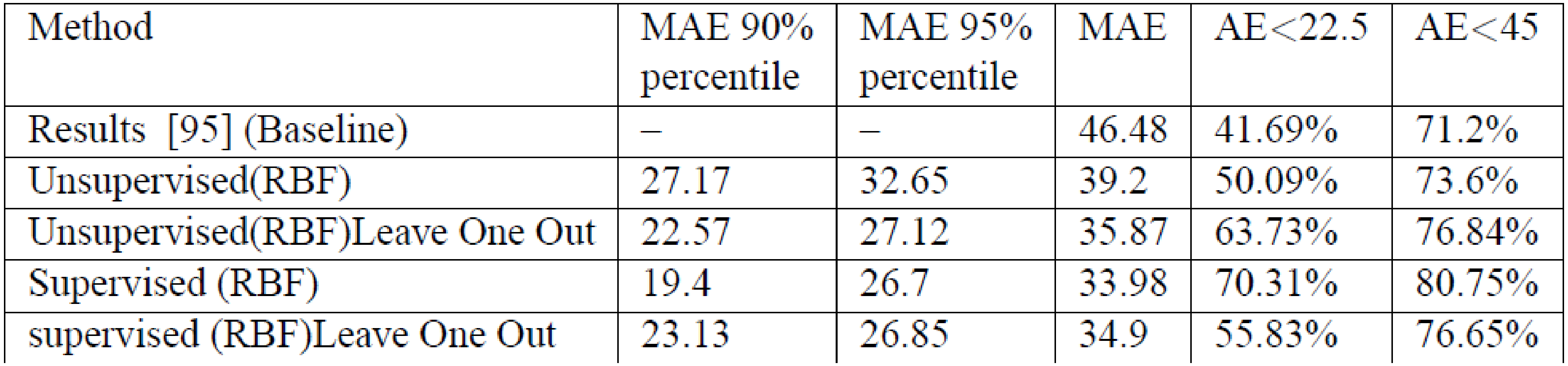
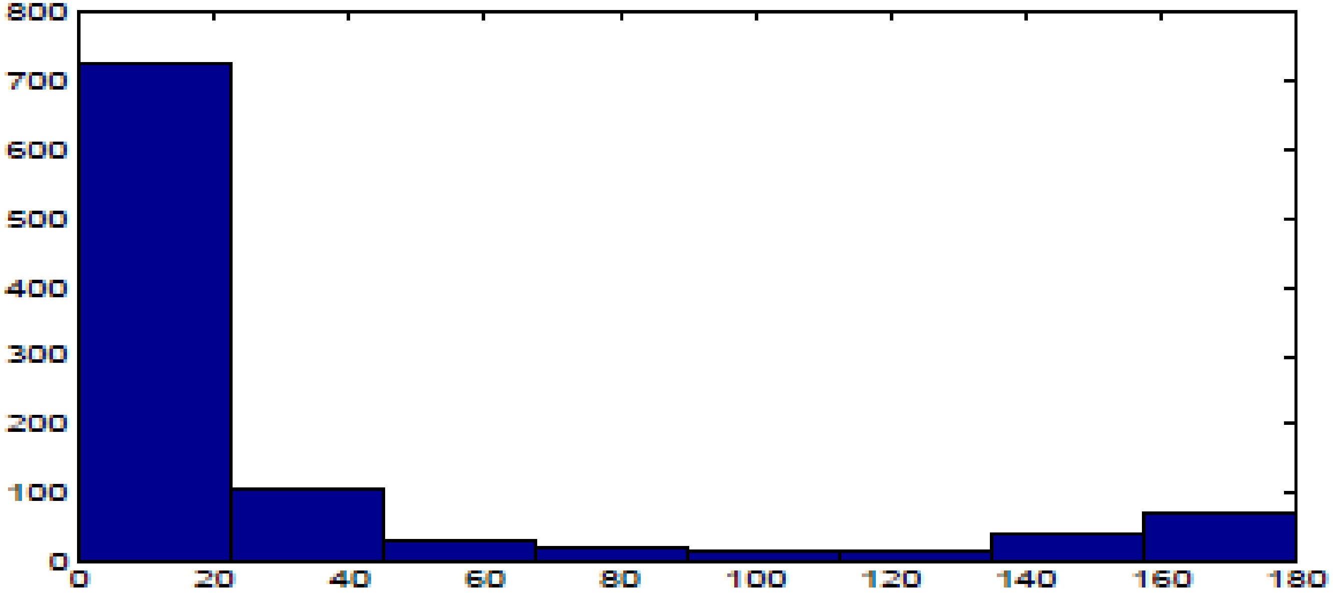


Figure 2: Regression on a Multi-view car dataset: Top left corner shows how the arrows reflect he estimated angle. The ground truth is shown along with the estimated angle. Yellow arrows for ground truth and Magenta for our results, features are shown as blue dots(Best viewed in color)

MutiView Car Dataset



Method	MAE 90% percentile	MAE 95% percentile	MAE	AE<22.5	AE<45
Results [95] (Baseline)	–	–	46.48	41.69%	71.2%
Unsupervised(RBF)	27.17	32.65	39.2	50.09%	73.6%
Unsupervised(RBF)Leave One Out	22.57	27.12	35.87	63.73%	76.84%
Supervised (RBF)	19.4	26.7	33.98	70.31%	80.75%
supervised (RBF)Leave One Out	23.13	26.85	34.9	55.83%	76.65%



Datasets for Evaluation

- Face pose data set Aghajanian et al 09
  - Uncontrolled environment
  - Different illumination, expression, occlusions, pose.
  - Comparisons to Aghajanian et al 09
    - Ours supervised $MAE = 11.15°$
    - Ours unsupervised $MAE = 10.92°$
    - Aghajanian et al 09 $MAE = 13.21°$

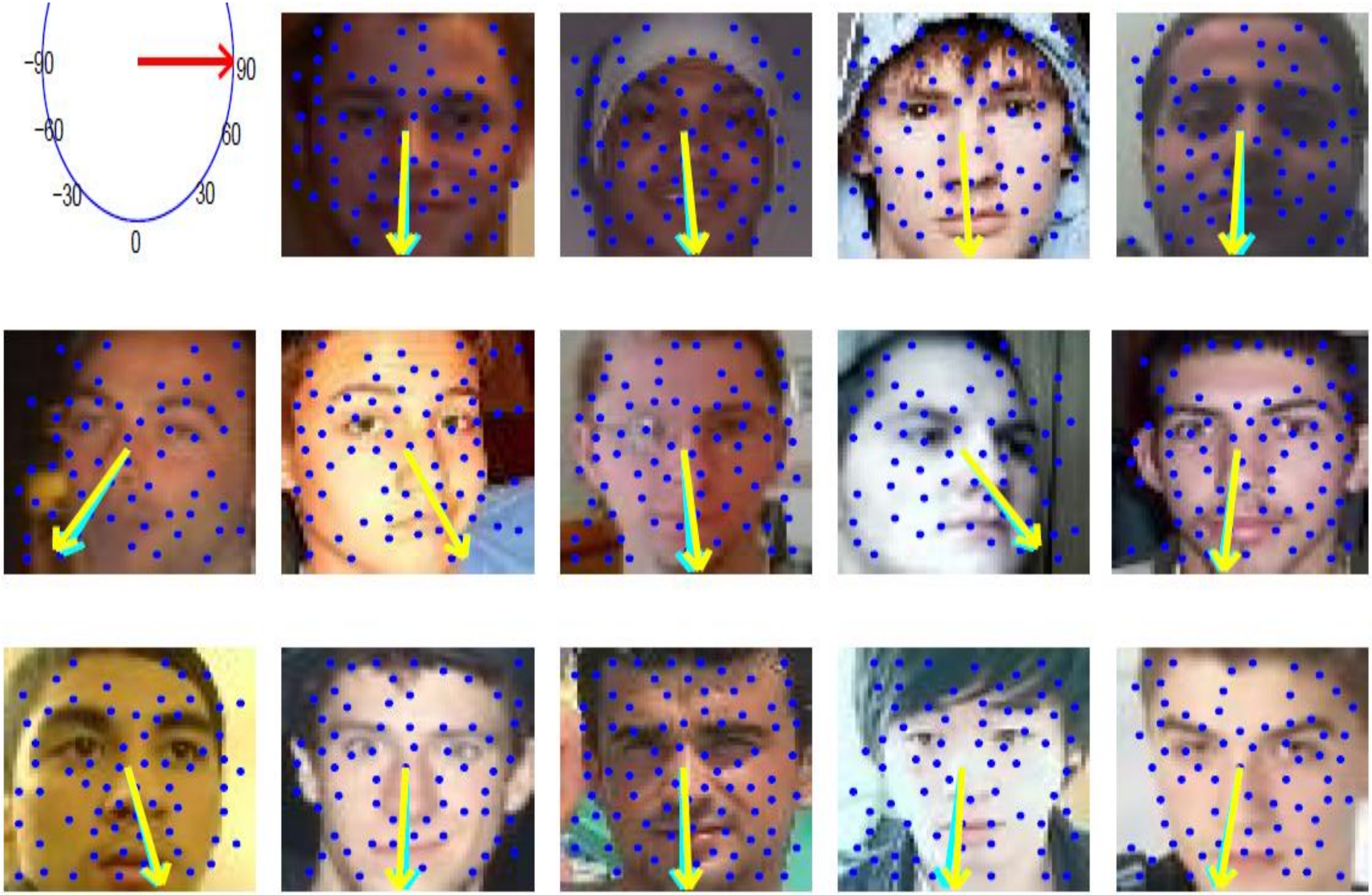


Figure 4: Regression on a Face Pose estimation dataset: Top left corner shows how the arrows reflect he estimated angle. The ground truth is shown along with the estimated angle. Green arrows for ground truth and Yellow for our results, features are shown as blue dots(Best viewed in color)

Body Posture Estimation

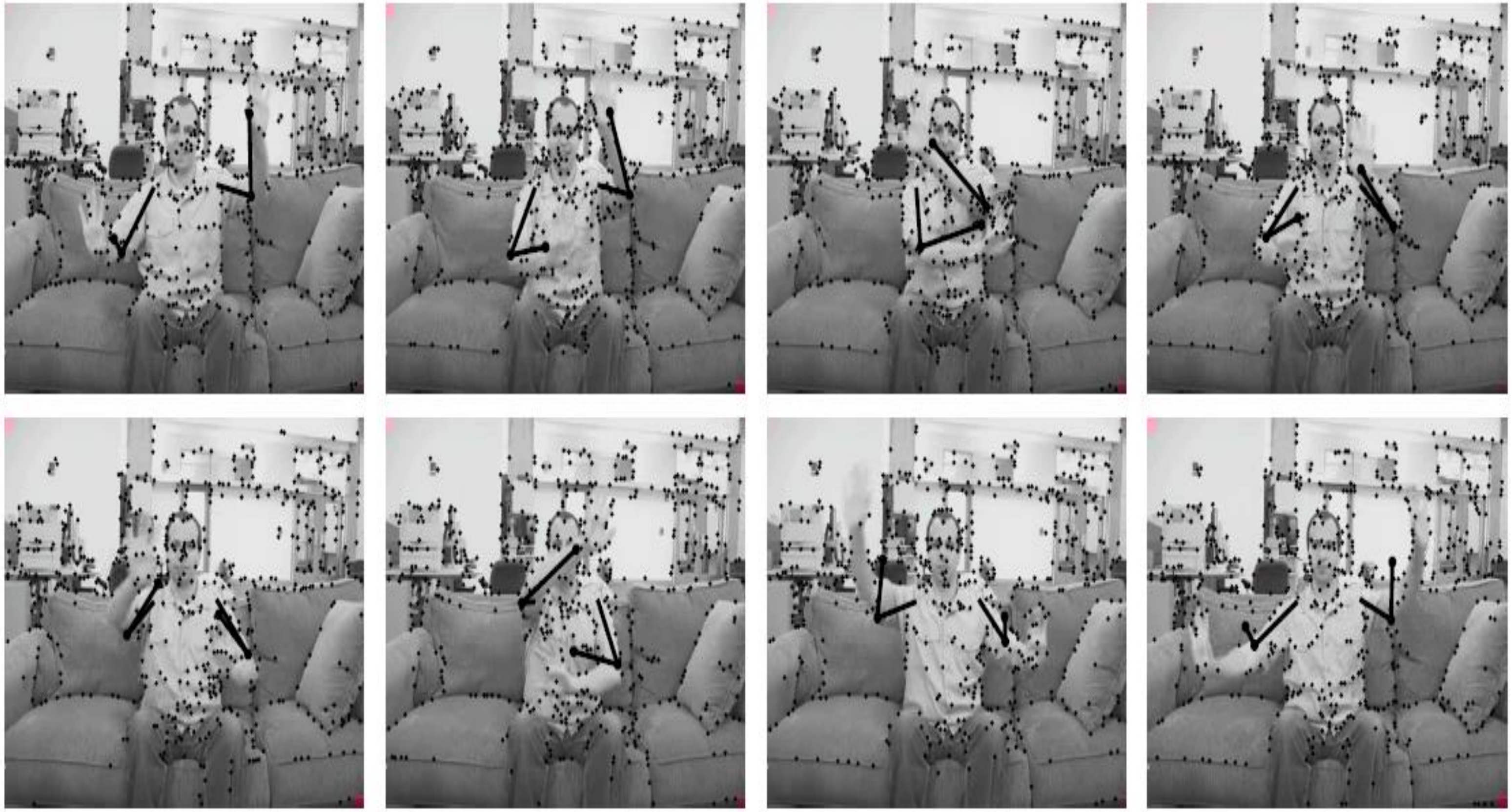


Figure 5: Regression example for articulated body posture estimation: shown are frames 20,40,60,80,100,120,140,160

Outline

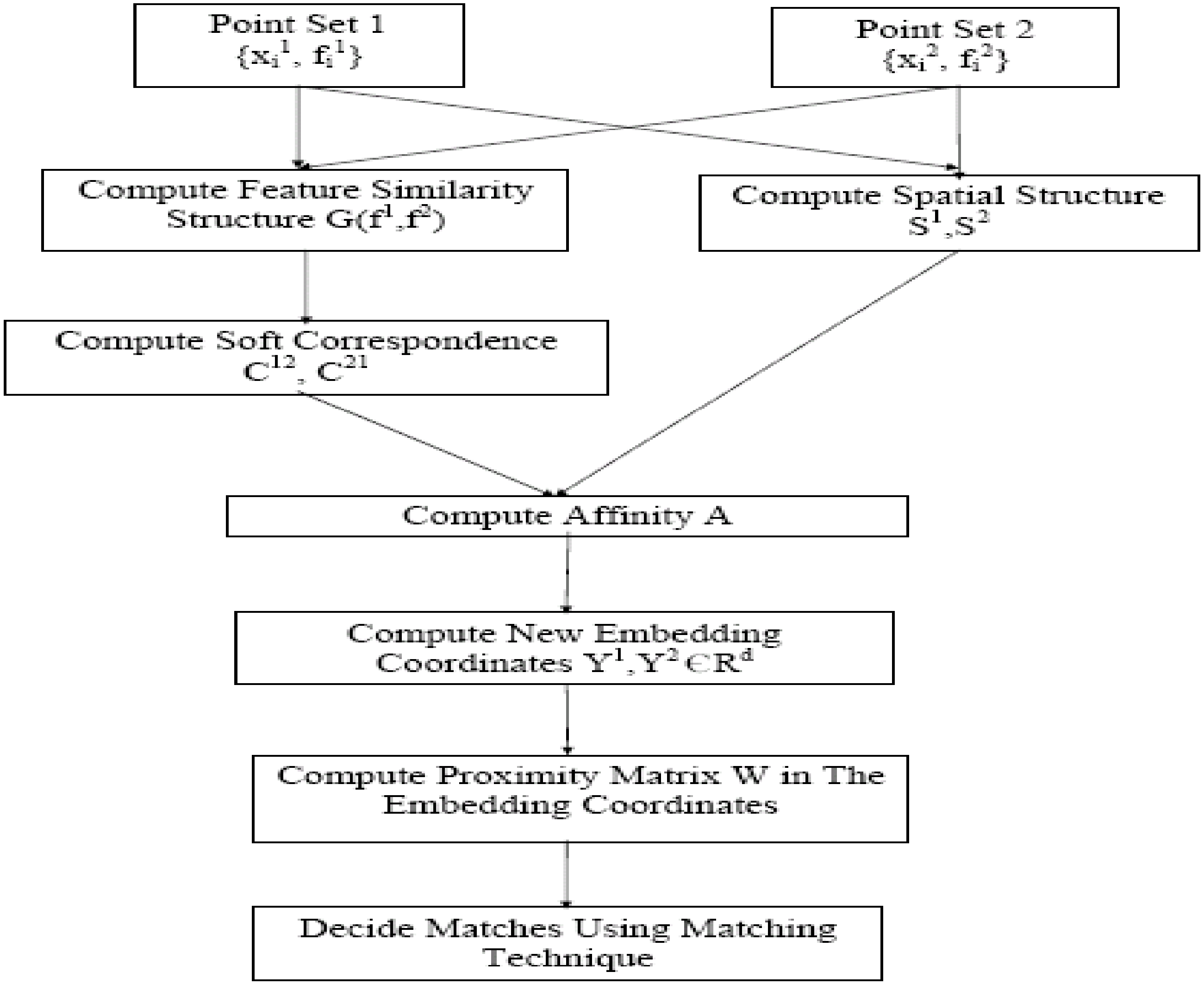
- Introduction
- Contributions
- Feature-Spatial Embedding Framework
- Image Embedding from Local Features
- Regression from Local features
- Multi-set Feature Matching.
- Implicit Feature-Spatial Manifold
- Conclusions

Multi-set Feature Matching

- Graph Matching through Embedding
- Matching Multiple sets in one shot
- Scalability

Torki and Elgammal CVPR10b

Block Diagram for Matching Two Images



Point Set 1
$\{x_i^1, f_i^1\}$

Point Set 2
$\{x_i^2, f_i^2\}$

Compute Feature Similarity Structure $G(f^1,f^2)$

Compute Spatial Structure S^1,S^2

Compute Soft Correspondence C^{12}, C^{21}

Compute Affinity A

Compute New Embedding Coordinates $Y^1,Y^2 \in R^d$

Compute Proximity Matrix W in The Embedding Coordinates

Decide Matches Using Matching Technique

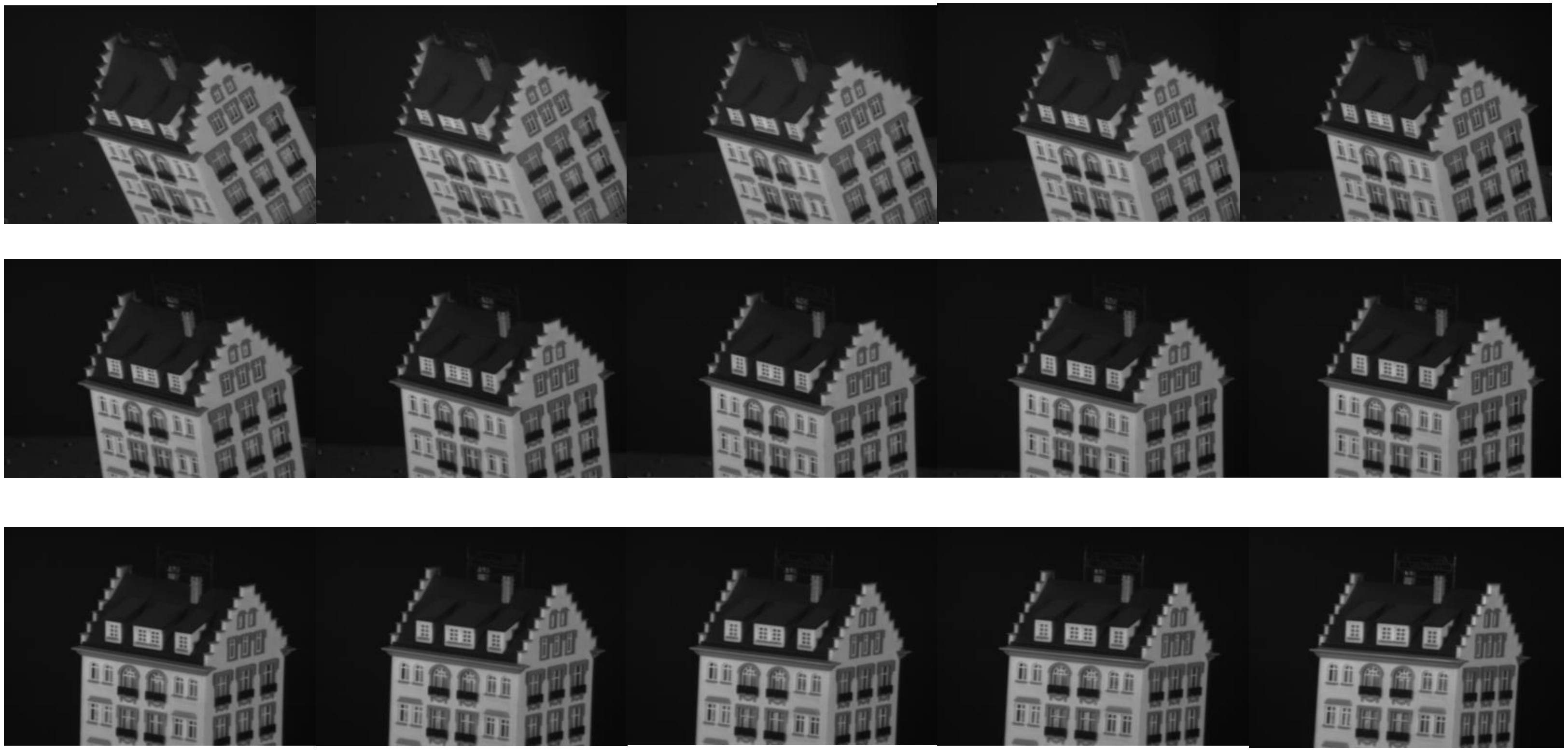
Compute Soft Correspondences $\boldsymbol{C}$

- Weights should follow exclusion principle.
- Weights should be soft encoded.
- $\boldsymbol{G}_{ij}^{pq} = e^{-\left\| f_i^p - f_j^q \right\|^2 / 2\sigma^2}$
- Gaussian kernel in feature space is not suitable for our formulation.
- Use Scott and Longuet-Higgins 91 algorithm in the feature space.

Matching Settings

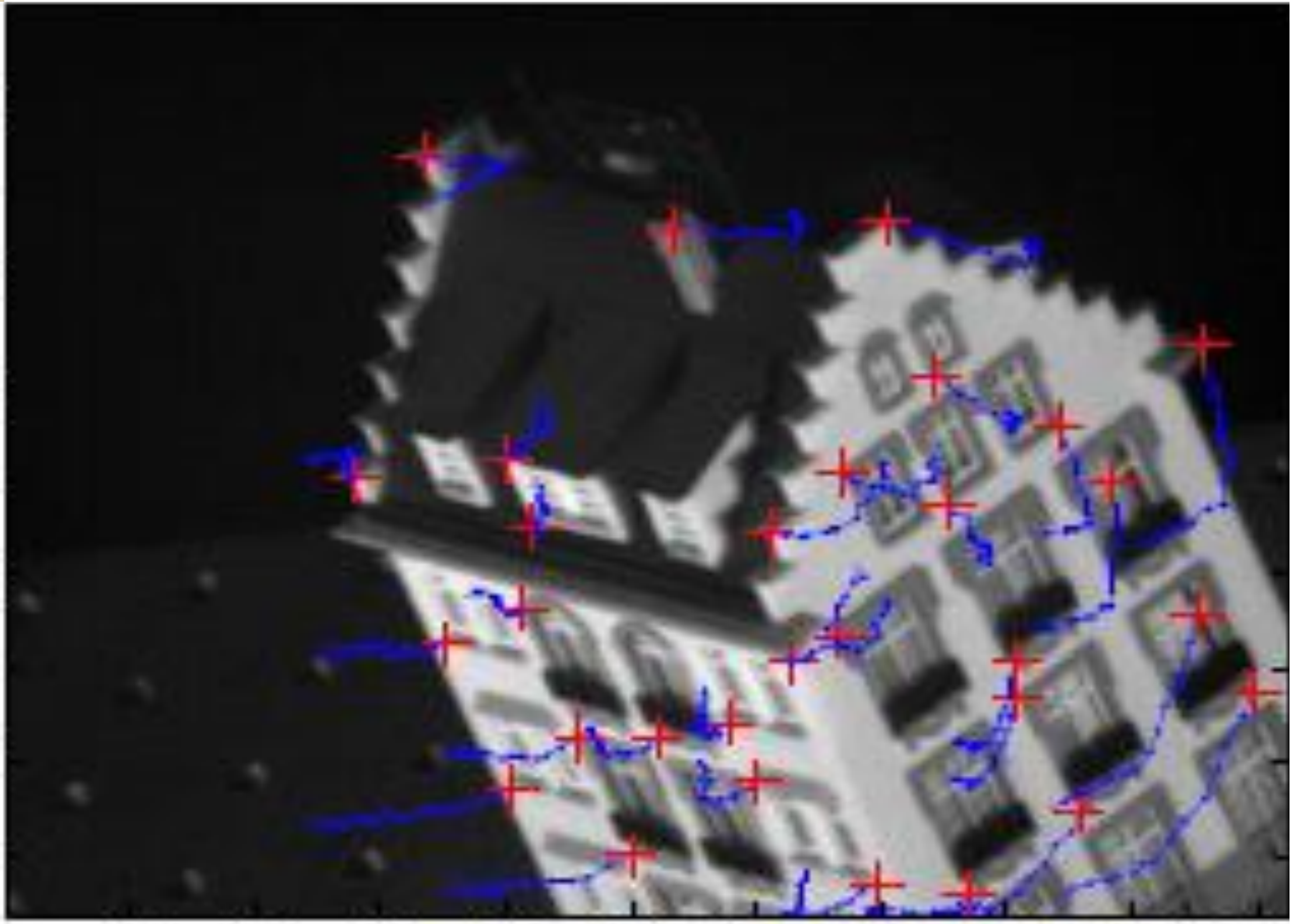
- Pairwise Matching (PW):
  - Only two images are matched
- Multiset Pairwise Matching (MP):
  - Embedding all features from K sets together. Use Embedding coordinates to compute pairwise matches.
- Multiset Clustering (MC):
  - Embedding all features from K sets together. Use Embedding coordinates to cluster the feature points in M clusters every cluster describes matched features.

Comparative Evaluation (Hotel Sequence)

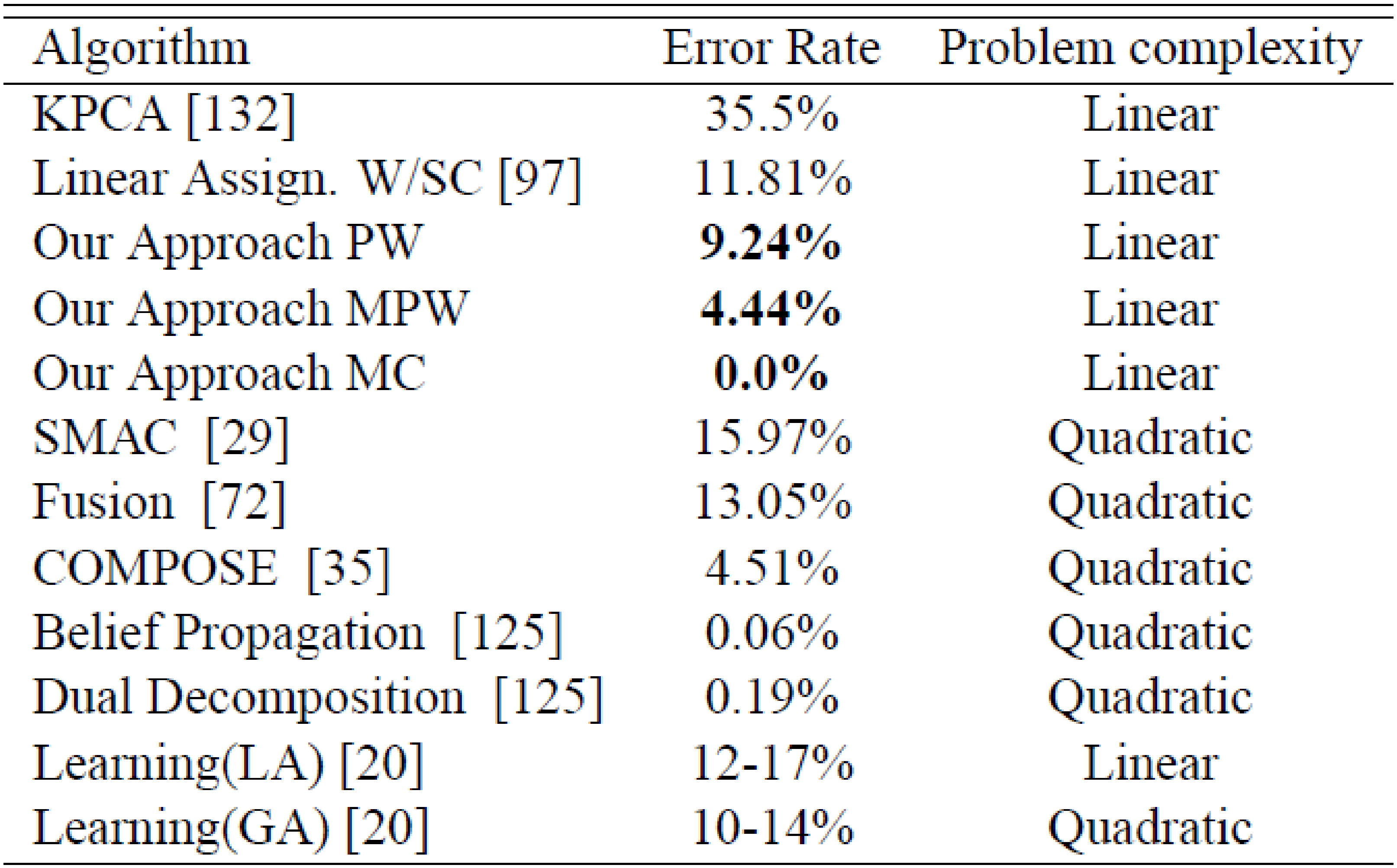


- 101 frames sampled every 7 frames.
- Each frame contains 30 manually labeled features.
- 105 matching pairs.

Comparative Evaluation (Hotel Sequence)



Comparative Evaluation (Hotel Sequence)

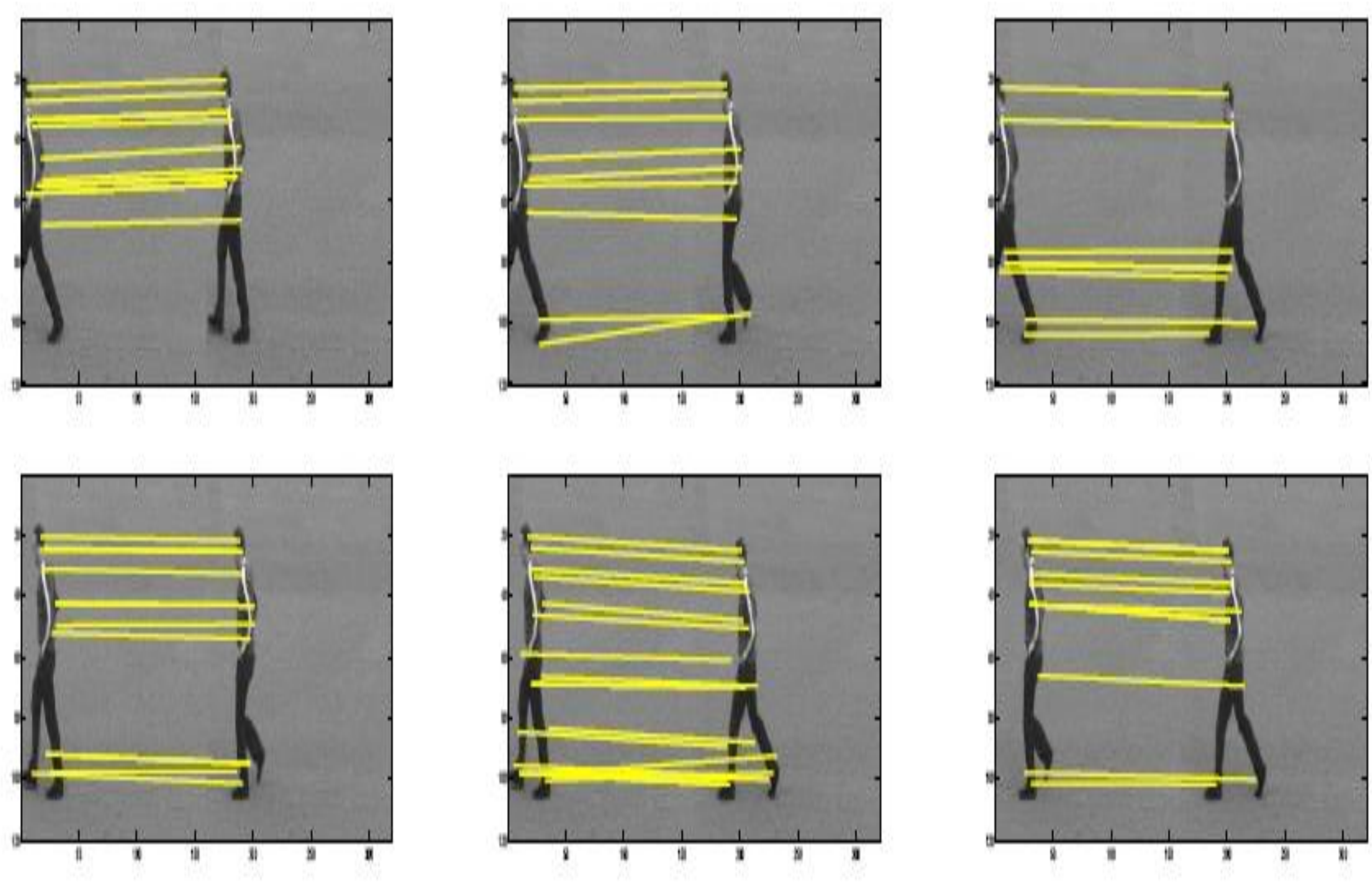


Algorithm	Error Rate	Problem complexity
KPCA [132]	35.5%	Linear
Linear Assign. W/SC [97]	11.81%	Linear
Our Approach PW	**9.24%**	Linear
Our Approach MPW	**4.44%**	Linear
Our Approach MC	**0.0%**	Linear
SMAC [29]	15.97%	Quadratic
Fusion [72]	13.05%	Quadratic
COMPOSE [35]	4.51%	Quadratic
Belief Propagation [125]	0.06%	Quadratic
Dual Decomposition [125]	0.19%	Quadratic
Learning(LA) [20]	12-17%	Linear
Learning(GA) [20]	10-14%	Quadratic

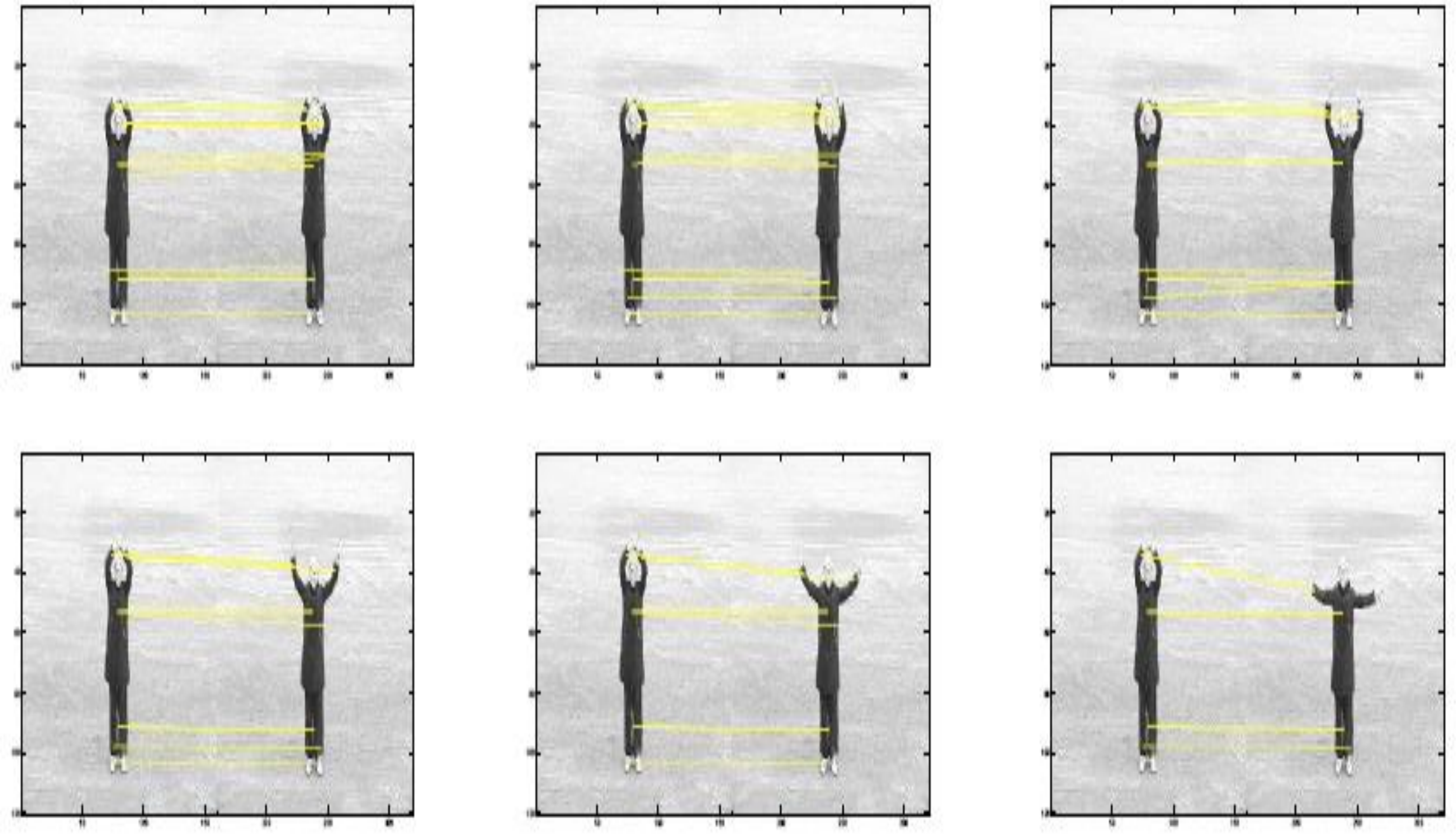
Comparative Evaluation (Hotel Sequence)

- Using our multiset**MPW** and **MC** we reach *95.56%* and *100%* accuracy, which is not reached by any of the competing algorithms.
- The size of our affinity matrix A in the case of the multiset of *15* frames is just *450* × *450* and for the case of the pairwise matching is *60* × *60*,
- The size for one edge compatibility matrix for any of the quadratic assignment approaches is *900* × *900*.

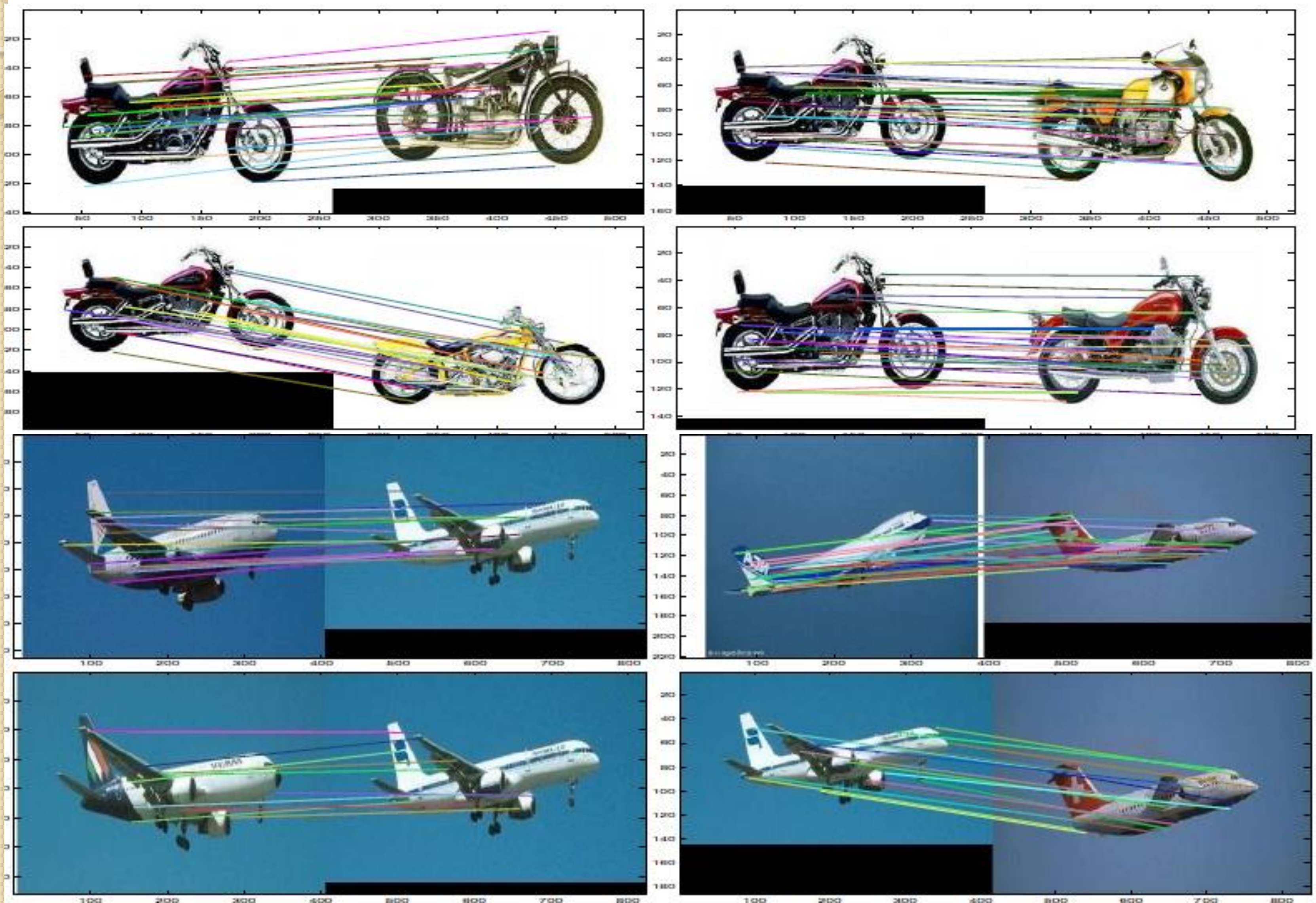
Non-Rigid Matching (Walking) PW



Non-Rigid Matching (HandWaving) MC



Non-rigid Matching: within Class Variation (MPW)



Conclusions

- We showed that we can find explicit representation of the local features with their spatial arrangement in the form of embedded coordinates.
- The joint manifold representation enabled us to find proper kernel between images that can be used for recognition.

Conclusions

- Also, the joint manifold representation enabled us to do regression from local features without assuming holistic representation of the image instances.
- The joint manifold also gives excellent matching results compared to state-of-the-art methods.

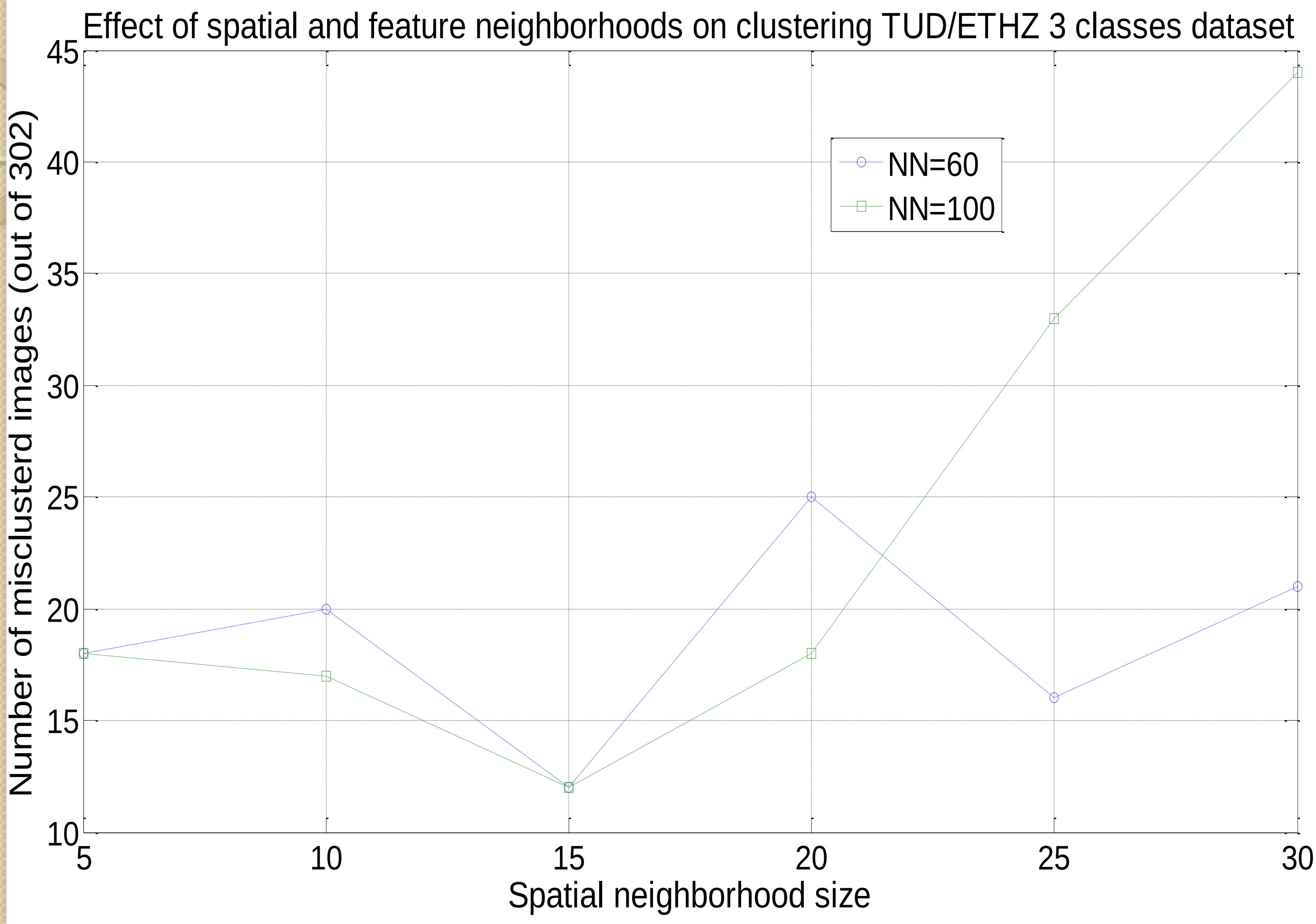
Thank You

Acknowledgement

- Advisor Dr. Ahmed Elgammal
- Committee
  - Dr. Kulikowski
  - Dr. Pavlovic
  - Dr. Kumar
- Family (Parents, Esraa, Menna, Abdelrahman)
- Friends(Ali, Amr, Imdad, Tarek,…)
- Labmates (CBIM)

Boody

Menna



Effect of spatial and feature neighborhoods on clustering TUD/ETHZ 3 classes dataset