Cross-view Activity Recognition using Hankelets

Binlong Li, Octavia I. Camps and Mario Sznaier

Dept. of Electrical and Computer Engineering

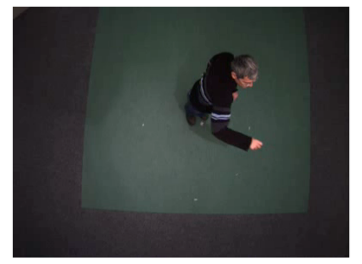
Northeastern University

Abstract

*We introduce a new feature for cross-view activity recognition: the "**Hankelet**". This type of feature captures dynamic properties of short tracklets that are invariant to viewpoint changes and time shifts. Experiments using Hankelets on the IXMAS database show a **20% improvement** over the state of the art.*

GOAL

To recognize an activity from a different viewpoint than the one used for training.



Testing Data

EXISTING APPROACHES

- Geometric constraints [32]
- Track body joints [21,22]
- 3D Models [8,15,30,31]
- Quasi-invariant geometric features [10,11]
- **Transfer features across views [7,16]** (75.3% accuracy on IXMAS)

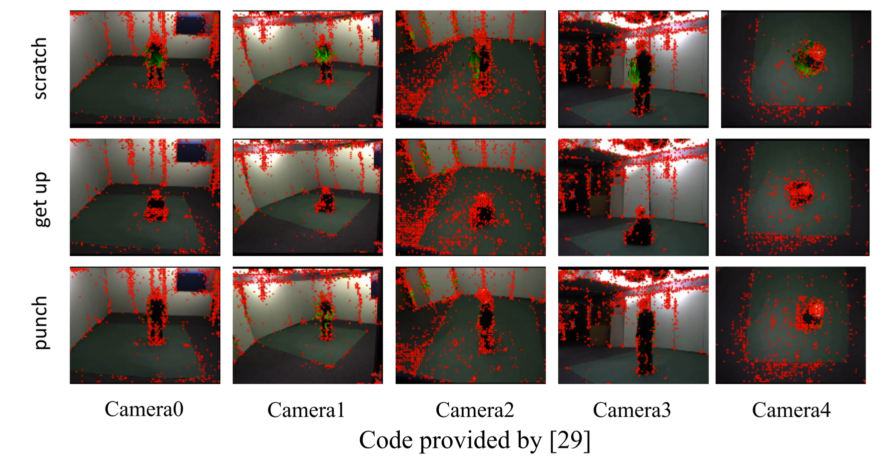
Best performance is far below the state of art performance for single view activity recognition.

This work was supported in part by NSF grants IIS–0713003 and ECCS–0901433, AFOSR grant FA9550–09–1–0253, and the Alert DHS Center of Excellence under Award Number 2008-ST-061-ED0001.

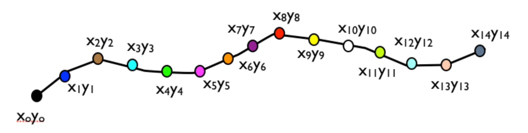
HANKELETS

A Dynamics-based Feature: Hankelet

Tracklets Detection (shown in green)



Detected Tracklet

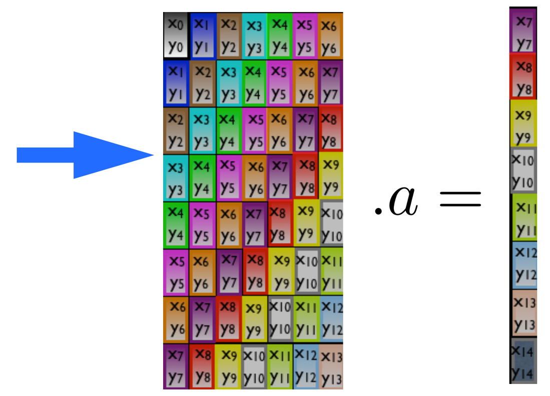


$$\mathbf{p}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \mathbf{C}\mathbf{X}_k$$

$$\mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} \quad \mathbf{X}_o \text{ given}$$

$$\sum_{i=1}^{rank(H)} a_i \begin{bmatrix} x_{k-i} \\ y_{k-i} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

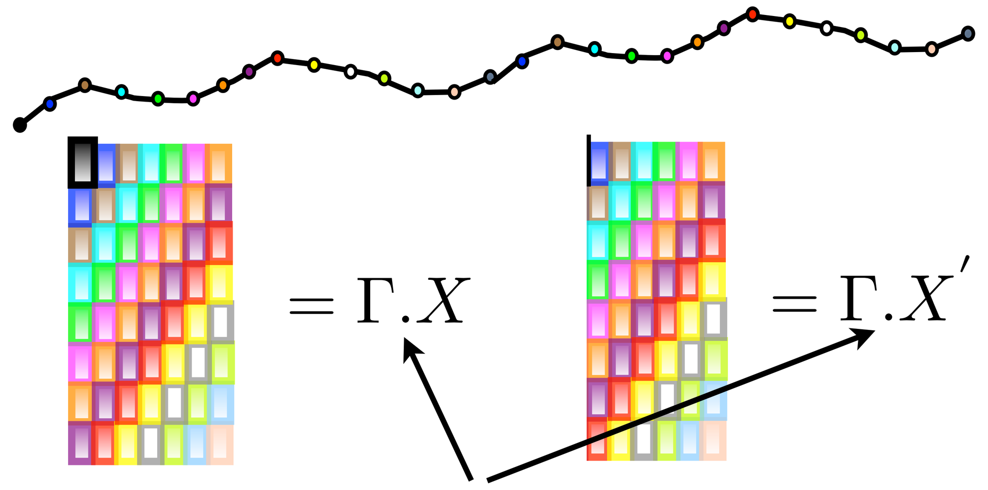
n^th order regressor



Hankelet

$$\hat{\mathbf{H}} = \frac{\mathbf{H}}{\|\mathbf{H}\mathbf{H}^T\|_F^{1/2}}$$

Hankelets: Initial Frame Invariance



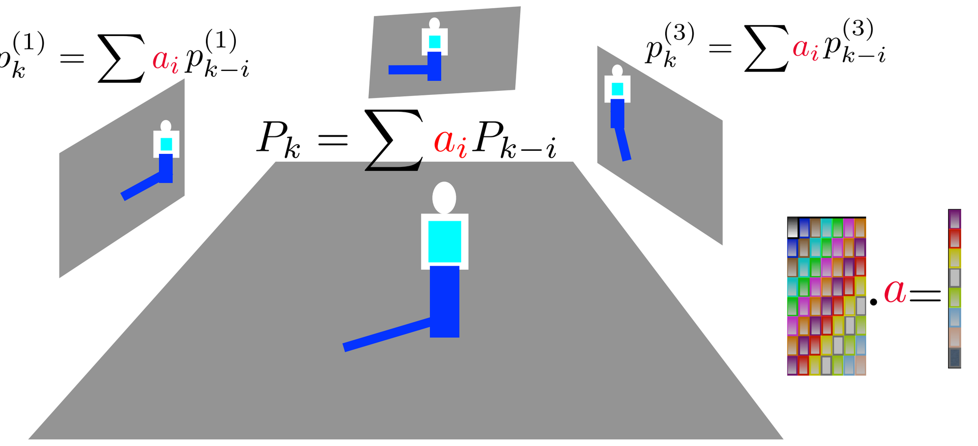
$$= \Gamma.X \qquad = \Gamma.X'$$

The columns of the Hankelets span the same space, regardless of the initial frame.

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{C}^T & \ldots & (\mathbf{C}\mathbf{A}^m)^T \end{bmatrix}^T$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_o & \ldots & \mathbf{X}_m \end{bmatrix}^T$$

Hankelets: Viewpoint invariance



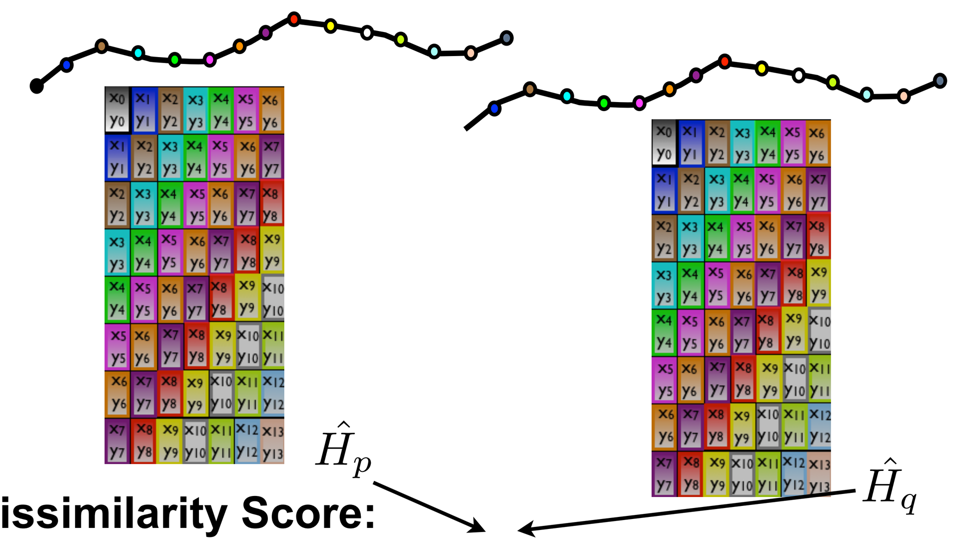
$$p_k^{(1)} = \sum a_i p_{k-i}^{(1)} \qquad p_k^{(2)} = \sum a_i p_{k-i}^{(2)} \qquad p_k^{(3)} = \sum a_i p_{k-i}^{(3)}$$

$$P_k = \sum a_i P_{k-i}$$

Regressor is invariant to affine transformations

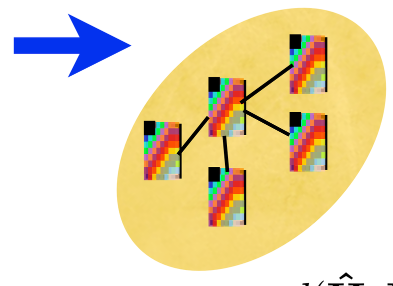
$$p_k^{(j)} = \Pi_j P_k = \Pi_j \sum a_i P_{k-i} = \sum a_i \Pi_j P_{k-i} = \sum a_i p_{k-i}^{(j)}$$

Hankelet Codebook: Clustering



Dissimilarity Score:

$$d(\hat{\mathbf{H}}_p \hat{\mathbf{H}}_q) = 2 - \|\hat{\mathbf{H}}_p \hat{\mathbf{H}}_p^T + \hat{\mathbf{H}}_q \hat{\mathbf{H}}_q^T\|_F$$



$$\hat{\mathbf{H}}_p = \mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^T$$

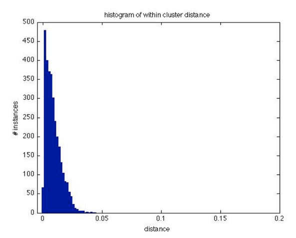
$$\hat{\mathbf{H}}_q = \mathbf{U}_q \mathbf{\Sigma}_q \mathbf{V}_q^T$$

$$\mathbf{\Theta}_{pq} = (\mathbf{U}_p \mathbf{\Sigma}_p)^T (\mathbf{U}_q \mathbf{\Sigma}_q)$$

$$d(\hat{\mathbf{H}}_p \hat{\mathbf{H}}_q) = 2 - \sqrt{2 + 2\|\mathbf{\Theta}_{pq}\|_F^2}$$

- Captures "alignment" between the column spaces.
- Robust to noisy measurements.
- Computationally efficient.
- Dissimilarity within cluster follows a Gamma pdf:

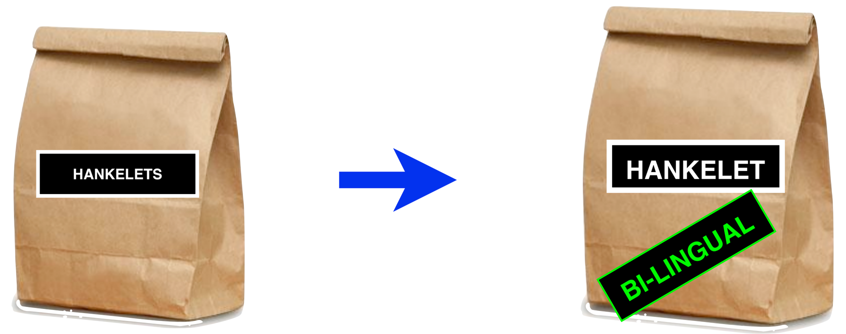
$$p(d|w) = \begin{cases} \frac{a^b d^{b-1}}{(b-1)!} e^{-ad} & \text{for } d > 0 \\ 0 & \text{otherwise} \end{cases}$$

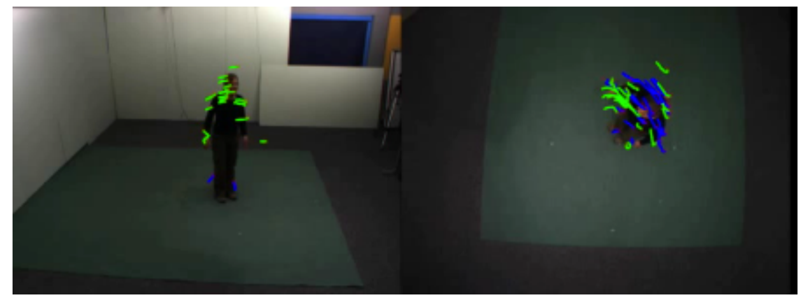


BAGS of HANKELETS

While Hankelets carry viewpoint invariance, they are not immune to self-occlusions and limited field of view.

Improving Cross-view Performance: Bag of BI-LINGUAL Hankelets





Bi-lingual tracklets marked in green

Bi-lingual Hankelets: subset visible from different viewpoints

Experimental Results

Single View: KTH Dataset

Algorithm	*Perf*
Ours	95.89
Cao et al. [3]	95.02
Wang et al.[29]	94.2
Le et al. [13]	93.9
Li et al [14]	93.6

Act Type	Acc
Boxing	95.71
HClapping	95.48
HWaving	99.09
Walking	99.52
Running	91.52
Jogging	94.05

Cross-View: IXMAS Dataset

Train classifier with data from one viewpoint, test with data from a different one

	Cam 0		Cam 1		Cam 2		Cam 3		Cam 4	
	BoHks	cuboids	BoHks	cuboids	BoHks	cuboids	BoHks	cuboids	BoHks	cuboids
Cam 0			83.70	14.40	59.20	10.69	57.37	10.61	33.62	19.09
Cam 1	84.27	16.12			61.58	11.11	62.75	7.41	26.93	9.22
Cam 2	62.52	10.27	65.17	65.17			71.96	12.90	60.14	8.08
Cam 3	57.05	11.15	61.45	61.45	71.04	9.98			31.24	9.30
Cam 4	39.60	8.80	32.84	32.84	68.12	9.22	37.36	10.06		

Algorithm	*Perf.*
Ours	56.4%
cuboids [16]	10.9%

More than 400% accuracy improvement

Cross-View: IXMAS Dataset

Comparing against state-of-the-art cross-view techniques

Algorithm	*Perf*
Ours	90.57%
Liu et al. [16]	75.3%
Farhadi et al. [7]	58.1%
Junejo et al. [10]	59.5%
Farhadi et al. [6]	74.4%

Over 20% improvement

*See Table 5 in paper for detailed comparisons