2010 ICPR Contest on

Semantic Description of Human Activities (SDHA)

M. S. Ryoo*

J. K. Aggarwal

Amit Roy-Chowdhury

THE UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

UNIVERSITY OF CALIFORNIA
UC RIVERSIDE

ETRI
Electronics and Telecommunications Research Institute

SDHA 2010

- The 1st Human Activity Recognition Contest
  - Human activities of general interests
    - Surveillance scenarios
  - Three challenges with three **new datasets**

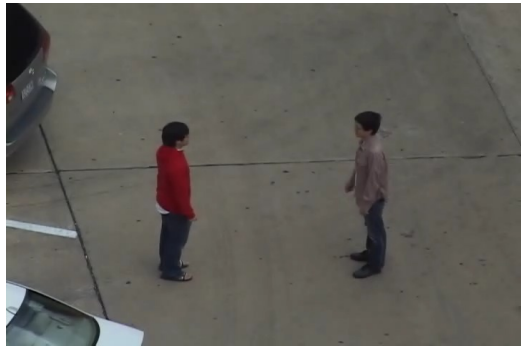
Interaction challenge

Aerial-view challenge

Wide-area challenge

SDHA 2010 challenges

- Interaction ***(UT-Interaction)***
  - Continuous videos
    - ***Detection*** vs. ***classification***
  - Human-human interactions
- Aerial-view ***(UT-Tower)***
  - Low-resolution: small actor
- Wide-area ***(UCR-Videoweb)***
  - Multiple cameras, wide-area
  - Various activities



Results overview

- We have invited the three finalists.

Challenge	TeamName	Authors	Institution	Success	Paper
Interaction	***Team BIWI***	Yao et al.	ETH	△	Variations of a Hough-Voting Action Recognition System
	TU Graz	-	TU Graz	X	-
	SUVARI	-	Sabanci Univ.[1]	X	-
	Panopticon	-	Sabanci Univ.[1]	X	-
Aerial-view	***Imagelab***	Vezzani et al.	Univ. of Modena and Reggio Emilia	O	HMM based Action Recognition with Projection Histogram Features
	ECSI_ISI	Biswas et al.	Indian Statistical Institute	O	-
	BU_Action	Guo et al.	Boston University	O	Aerial View Activity Classification by Covariance Matching of Silhouette Tunnels
	Team BIWI	Yao et al.	ETH	O	Variations of a Hough-Voting Action Recognition System
Wide-area	Vistek	-	Sabanci Univ.[2], Univ. of Amsterdam	X	-

Interaction Challenge

Interaction challenge

- Goal
  - Complex activity recognition from continuous videos
    - Surveillance cameras
  - Interactions
  - Dynamic surveillance-type environments
    - Pedestrians

Previous KTH dataset

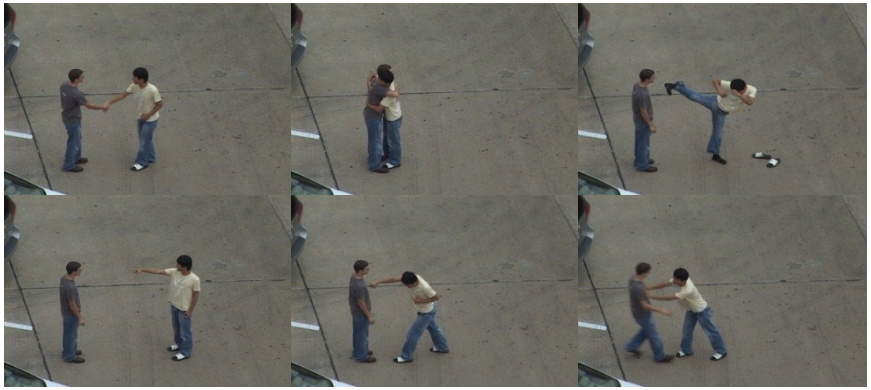
vs.

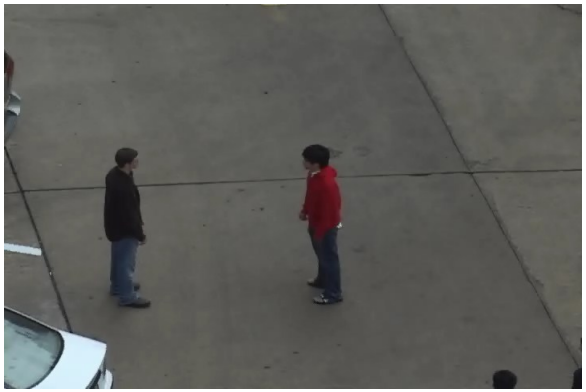
*New **UT-Interaction** dataset*

Human interactions *Pedestrians* *Multiple activities*

UT-Interaction dataset

- Dataset description
  - 720*480
  - Six types of human-human interactions
  - Two different sets
    - Different background: parking lot vs. lawn
    - 10 scenes for each set
    - More than 120 activity executions





Evaluation

- Cross validation
  - 10 scenes, leave-one-out = 10-folds

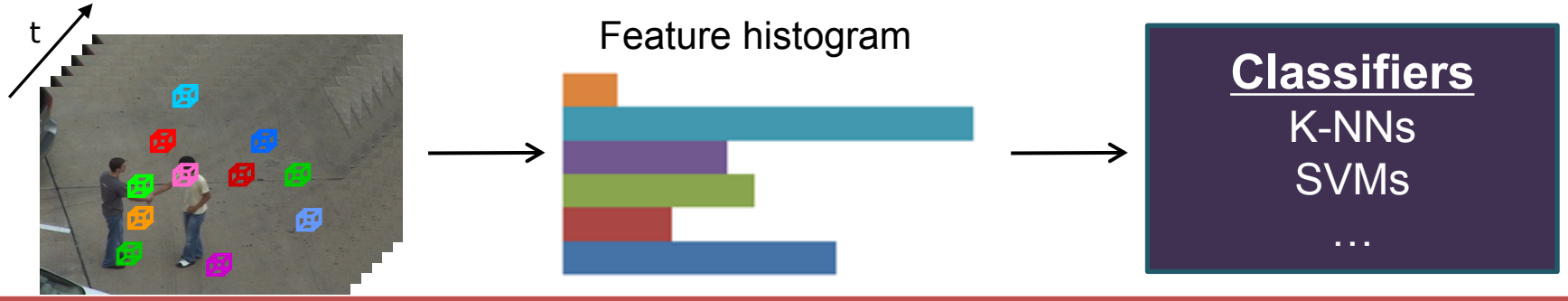
- Two problems
  - Classification
    - Choose activity category given ***segmented*** videos.
  - Detection
    - ***Localization*** in continuous videos

UT-Interaction results – set1

Classification accuracies:

	Shake	Hug	Kick	Point	Punch	Push	Total
Laptev + kNN	0.18	0.49	0.57	0.88	**0.73**	0.57	0.57
Laptev + Bayes.	0.38	0.72	0.47	0.9	0.5	0.52	0.582
Laptev + SVM	0.49	0.79	0.58	0.8	0.6	0.59	0.642
Latpev + SVM (best)	0.5	0.8	0.7	0.8	0.6	0.7	0.683
Cuboid + kNN	0.56	0.85	0.33	0.93	0.39	0.72	0.63
Cuboid + Bayes.	0.49	0.86	**0.72**	**0.96**	0.44	0.53	0.667
Cuboid + SVM	**0.72**	**0.88**	**0.72**	0.92	0.56	**0.73**	**0.755**
Cuboid + SVM (best)	**0.8**	**0.9**	**0.9**	**1**	**0.7**	**0.8**	**0.85**
Team BIWI	**0.7**	**1**	**1**	**1**	**0.7**	**0.9**	**0.88**

Baseline methods:

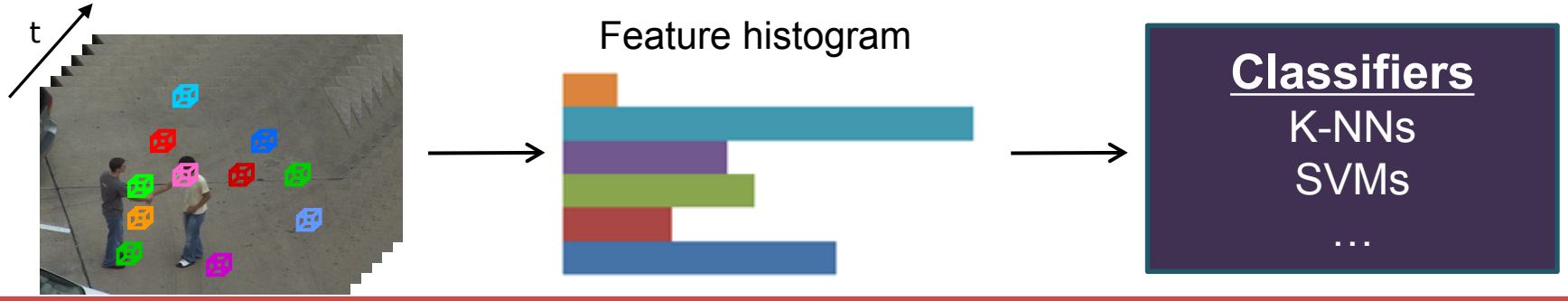


UT-Interaction results – set2

Classification accuracies:

	Shake	Hug	Kick	Point	Punch	Push	Total
Laptev + kNN	0.3	0.38	**0.76**	**0.98**	0.34	0.22	0.497
Laptev + Bayes.	0.36	0.67	0.62	0.9	0.32	**0.4**	0.545
Laptev + SVM	0.49	0.64	0.68	0.9	0.47	**0.4**	0.597
Latpev + SVM (best)	0.5	0.7	**0.8**	0.9	0.5	**0.5**	**0.65**
Cuboid + kNN	**0.65**	**0.75**	0.57	0.9	0.58	0.25	0.617
Cuboid + Bayes.	0.26	0.68	0.72	**0.94**	0.28	0.33	0.535
Cuboid + SVM	**0.61**	**0.75**	0.55	0.9	**0.59**	0.36	0.627
Cuboid + SVM (best)	**0.8**	**0.8**	0.6	0.9	**0.7**	**0.4**	**0.7**
Team BIWI	0.5	**0.9**	**1**	**1**	**0.8**	**0.4**	**0.77**

Baseline methods:



Interaction summary

- ***Classification*** problem
  - Successful results with *UT-Interaction* dataset.
  - Hierarchical approaches
    - Actions of each actor in human-human interaction
- ***Detection*** problem
  - Continuous recognition was requested.
  - **None** among four teams succeeded.
  - Future exploration
    - A hierarchical approach showed its potential.

Aerial-view Challenge

Aerial-view challenge

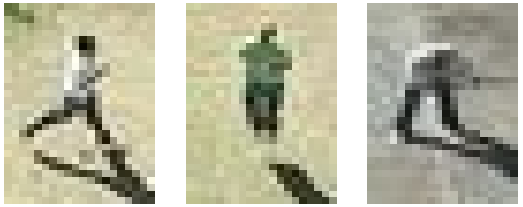
- Goal
  - Classification of human actions from **low-resolution** videos
    - Human height: 20 pixels
  - Top-down viewpoint
    - Unmanned aerial vehicles (UAVs)

UT-Tower dataset

- Dataset description
  - 360*240
  - 9 types of actions
  - Two different settings
    - Lawn vs. square

Evaluation

- Classification problem
  - Segmented videos
    - Only one action per video.



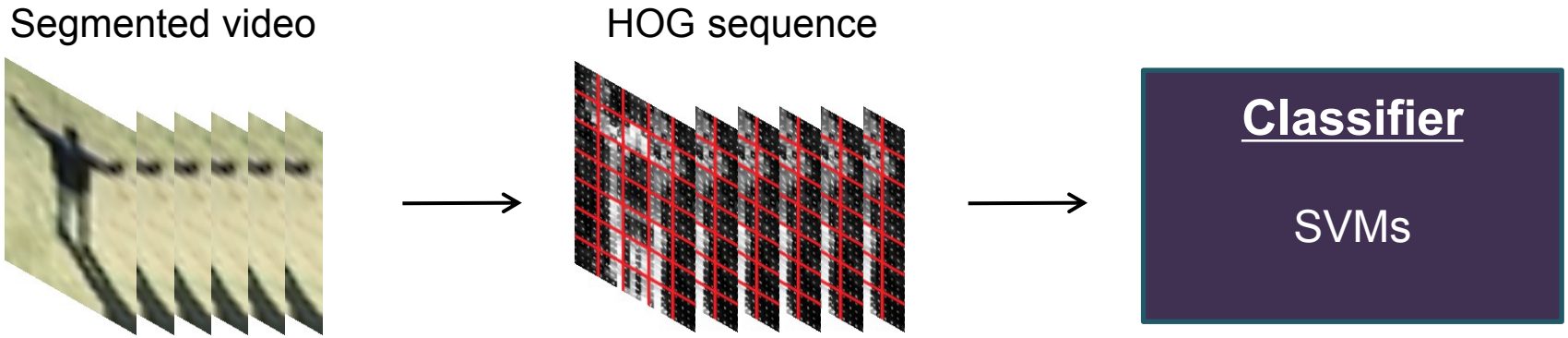
  - Bounding boxes and foreground masks
    - Spatial information provided.
- Cross validation
  - 108 videos, 108 leave-one-out
    - 107 training videos and 1 testing video
    - Abundant training videos

UT-Tower results

Classification accuracies:

	Point	Stand	Dig	Walk	Carry	Run	Wave 1	Wave 2	Jump	Total
Team BIWI	100	91.7	100	100	100	100	83.3	83.3	100	95.4
BU Action	91.7	83.3	100	100	100	100	100	100	100	97.2
ECSU_ISI	100	83.3	91.7	100	100	100	100	91.7	91.7	95.4
Imagelab	83.3	83.3	100	100	100	100	100	100	100	96.3
Baseline	100	83.3	100	100	100	100	83.3	100	100	96.3

Baseline method:



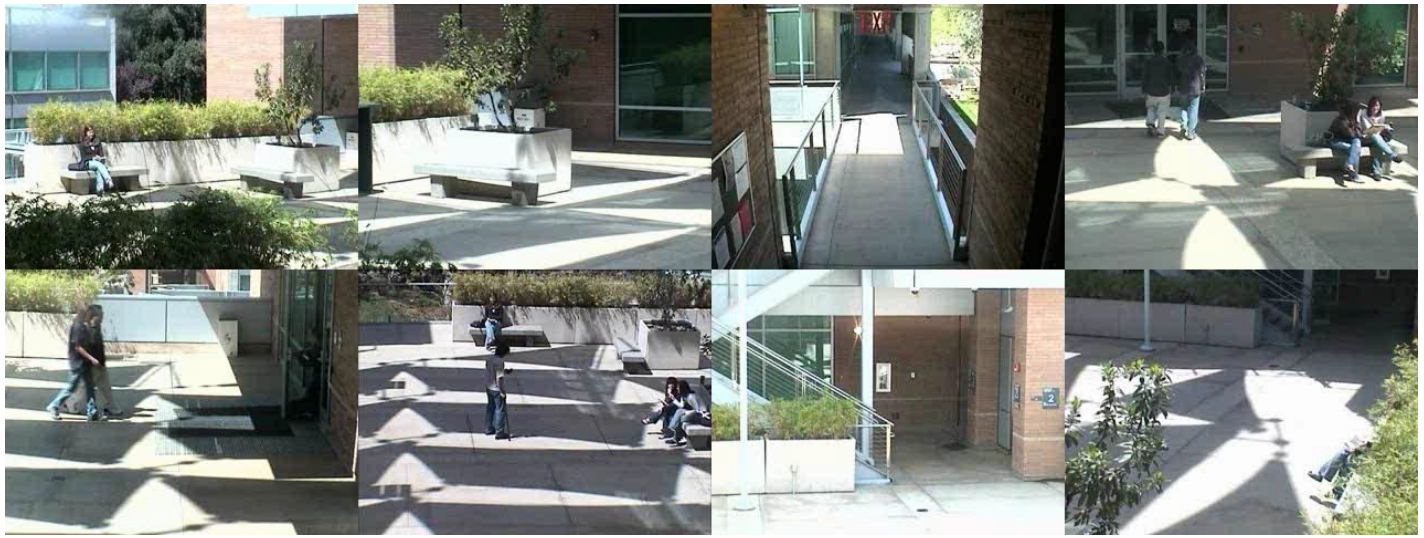
Aerial-view summary

- Most of the teams showed successful results.
  - Abundant training data: 107 training, 1 testing.
  - Baseline method also showed good results.
- Spatial info. provided: Bounding boxes
  - Good segmentation method required.
- Classification vs. detection?
  - Most difficult action: Standing

Wide-area Challenge

Wide-area challenge

- Open challenge using large-scale dataset
  - Multiple cameras observing a wide-area
    - Surveillance
  - Contestants were asked to formulate their own problem.



Open challenge

- Select a portion of the entire dataset
  - 39 possible scenes
- Choose evaluation
  - What activity will the system recognize?
  - Classification? Detection? Multiple cameras?
- Example problems
  - Detecting interactions between two persons
    - Hand-shake
  - Group activities
    - A person joining a group

UCR-Videoweb dataset

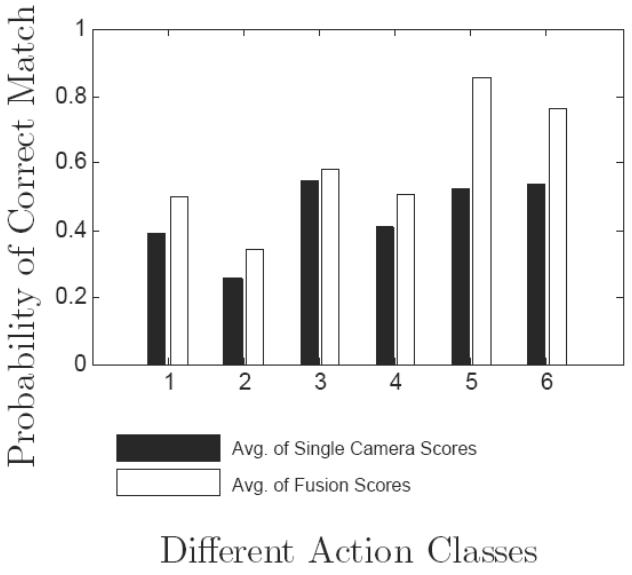
- Continuous dataset
  - 2.5 hours of videos divided into 39 scenes.
  - 4~8 cameras
  - Multiple types of activities
    - Human interactions, group actions, vehicles, …

Example results - 1

- Human interaction detection problem
  - A setting similar to the interaction challenge

Interaction	Our recognition accuracy	False positive rate
Shake hands	0.68	0.57
Hug	0.74	0.55
Point	0.63	0.25

- Multi-camera retrieval problem
  - Retrieving similar activities using multiple cameras.



Example problems - 2

- Group activity detection

Activity	Precision	Recall	Total Fetched	True Pos.	Ground Truth
Person Entering Building	1	1	4	4	4
Person Exiting Building	1	1	2	2	2
Person Entering Vehicle	0.75	0.75	4	3	3
Person Exiting Vehicle	1	1	3	3	3
People Walking Together	1	0.6	3	3	5
People Coming Together	0.7	0.7	7	5	5
People Going Apart	0.8	1	5	4	5
People Milling Together	0.78	0.92	14	11	13
People Meandering Together	0.85	0.92	27	23	25
Group Formation	1	0.78	7	7	9
Group Dispersal	0.8	0.8	5	4	4
Person Joining Group	1	0.95	18	18	19
Person Leaving Group	1	1	11	11	11

Wide-area summary

- Open challenge
  - UCR-Videoweb dataset with 2.5 hours of videos
- Provided a test bed for approaches
  - Difficult problems can be posed.
    - Continuous videos
    - Multiple cameras (4~8)
    - Various activities

[Kamal, A., Sethi, R., Song, B., Fong, A., Roy-Chowdhury, A.: Activity recognition results on UCR Videoweb dataset. In: Technical Report, Video Computing Group, University of California, Riverside (2010)]

Summary

- Introduced 3 new datasets/challenges.
- 8 teams attempted these challenges.
  - We invited 3 finalists based on their algorithms and results.
- No winner for the *interaction* and *wide-area*.
- The winner of *aerial-view* challenge is
  - **Team BU Action Covariance Manifolds**
    - Kai Guo, Prakash Ishwar, and Janusz Konrad
- Remaining problem: continuous recognition

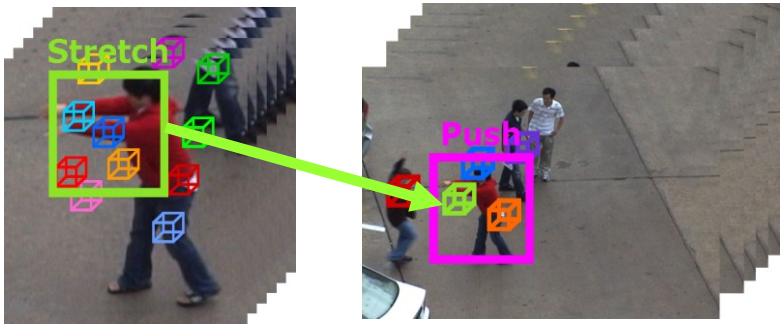
Thank you

- Thank you for your participation!

- The SDHA contest finalists will present their algorithms and results.
  - Imagelab: University of Modena and Reggio Emilia
  - BU Action Covariance Manifolds: Boston University
  - Team BIWI: ETH

Coming up next

- ***UT-Interaction*** dataset version 1.5
  - Sub-event labels for hierarchical recognition



- Result updates
  - Results of other research works
    - e.g. BMVC 2010
    - We will maintain the performance tables.
      - http://cvrc.ece.utexas.edu/SDHA2010
- 2nd SDHA?