Special Session: Dynamic Driving Environment Perception Based on Multi-Sensor Fusion, Tracking and Classification II

Dynamic Road Scene Classification: Combining motion with a visual vocabulary model

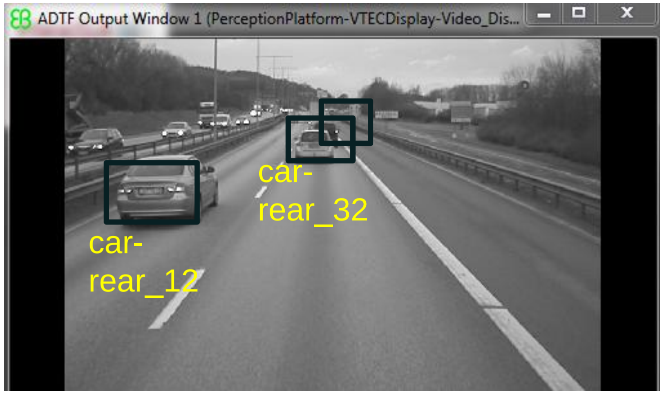
Anastasia Bolovinou
(research engineer- ICCS, phd cand. -NKUoA)

>joint work with C.Kotsiourou and A. Amditis during interactIVe IP

Askeri Museum, Istanbul Turkey, July 11th 2013

Motivation

…Tracking and classification of road objects already part of interactIVe Perception Platform

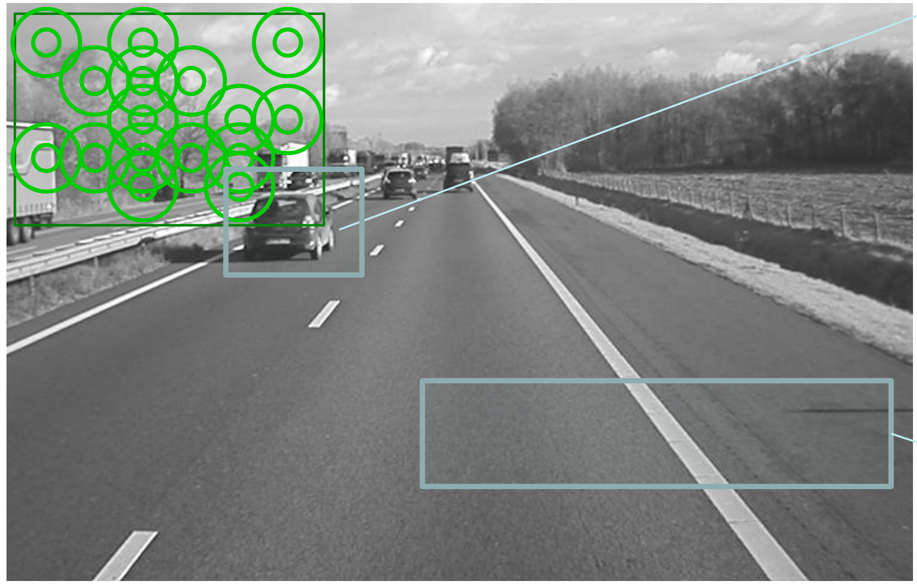


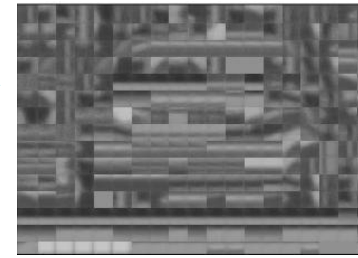
Add scene label information based on a cost-effective monochrome camera system: holistic scene understanding

…**Static** scene classification by learning appearance of local features through image pyramids in scale-space

- Cope with lower quality images coming from a moving vehicle
- Select efficient visual features for fast processing
- Exploit as much information we can get from a camera sensor

Problem setting (scene description)





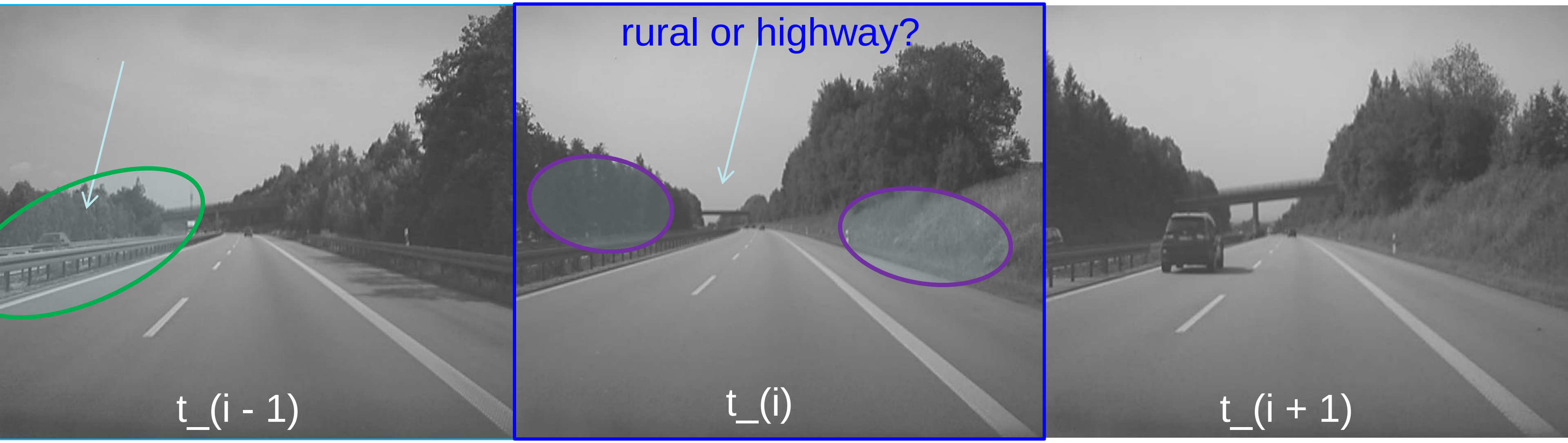
occlusions

restricted field of view

lack of textured surfaces

Core idea

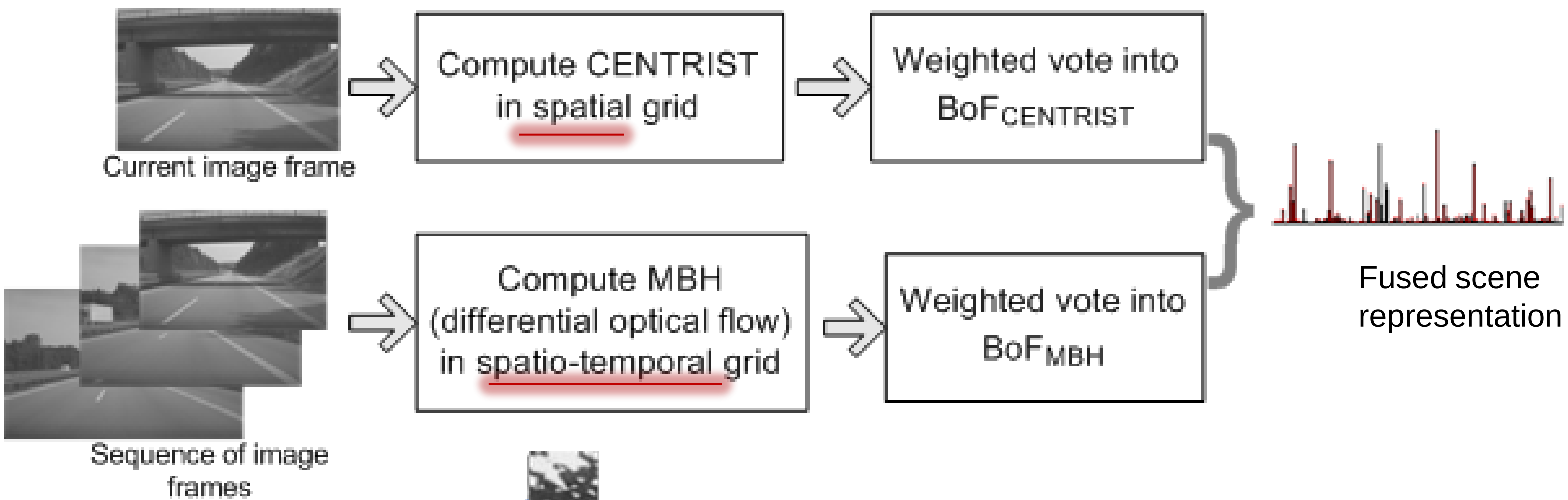
➧ **motion features** inherent in the frames' sequence can help disambiguate visually similar scenes

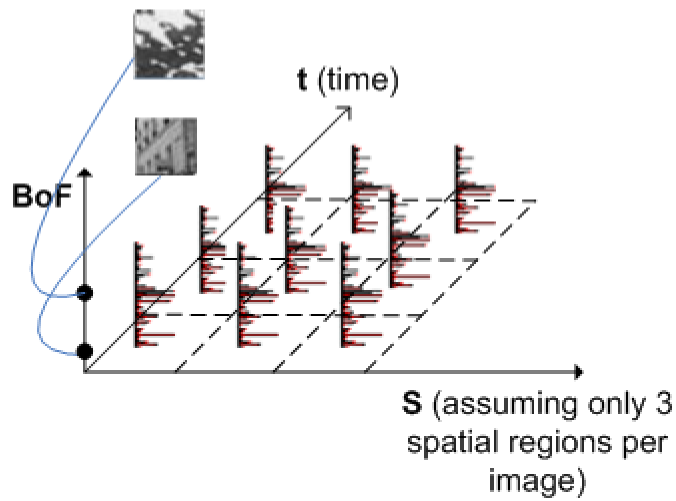


Note: Motion attributes can also show different properties in different time or spatial scale space since >> local degree of busyness varies
>> optical flow granularity varies

➢ *Inspired by work of [Derpanis, Lecce, Daniilides, Dynamic Scene Understanding, CVPR2012] and [Shroff, Turaga, Chellappa, Exploitig Motion for describing scenes, CVPR 2010] …dedicated to natural scene surveillance.*

Method overview



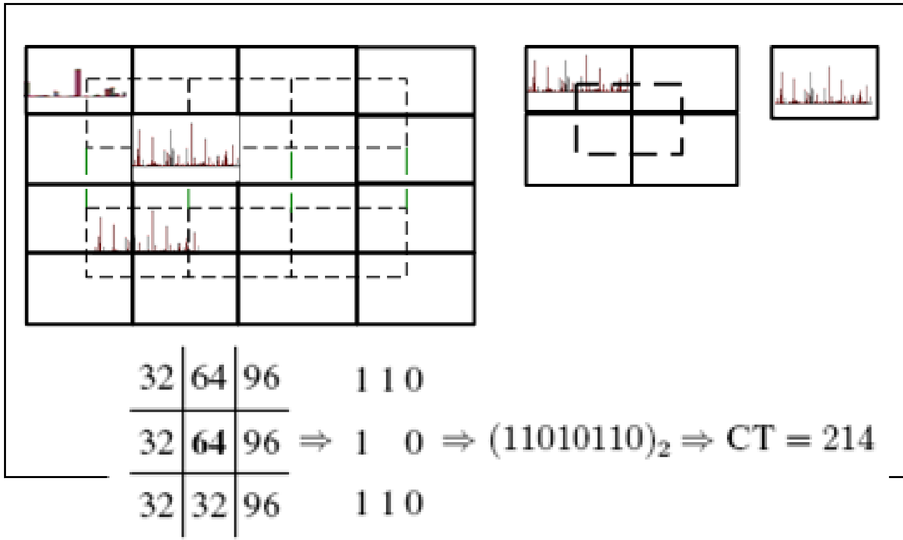


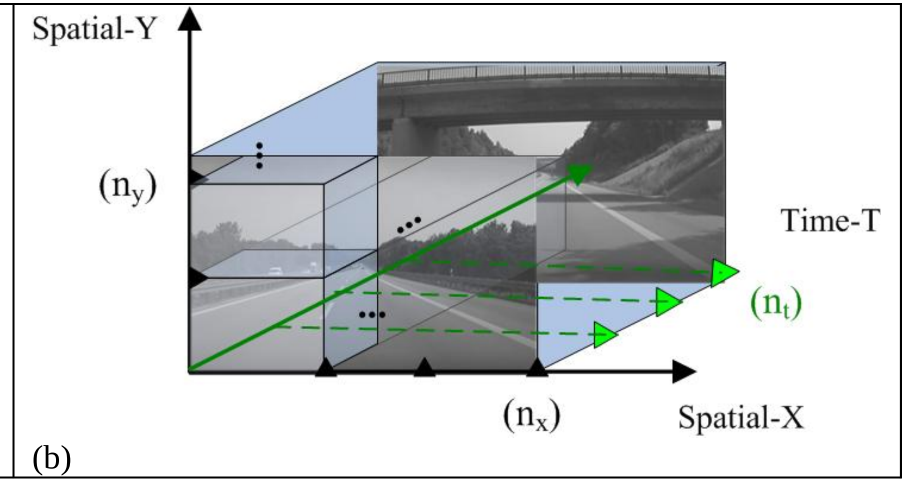
(Time-Space-Appearance representation)

Video Scene representation step

- Feature extraction from grid pyramids in time and space

(static) CENTRIST + (dynamic) MotionBoundaryHist_x,y





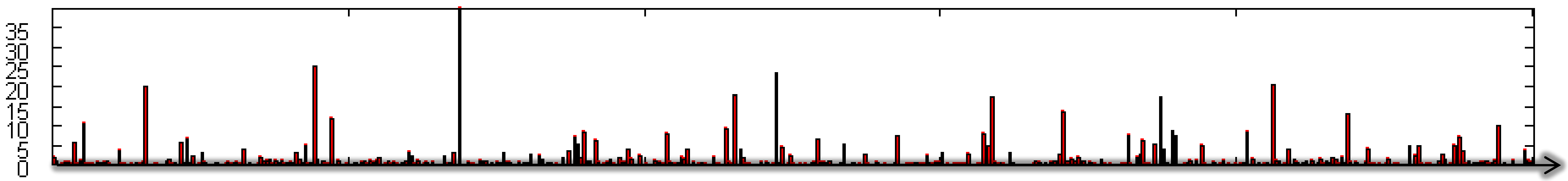
(31 spatial sectors of different sizes ,using grids in different scales
→If voc_length =200,
6200-d image representation)

(3x3 spatial sectors of the same size x 3 frame subsampling rates
→ If voc_length =200,
16200-d image representation)

Video Scene representation step

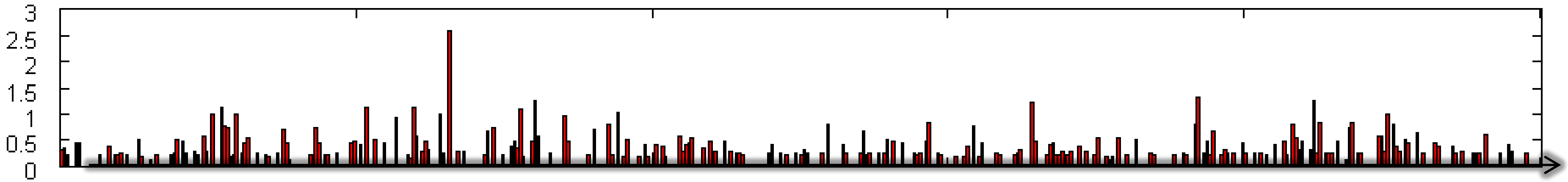
- Bag of Features for video (bag of MBH) and scene (bag of CENTRIST) through Histogram Intersection k-means clustering (better for histogram-based features of big dimensionality) and 4NN weighted voting into H-MBH, H-CENTRIST.

H-MBH, example histogram of a video record of 90 frames (9 frames history with R = 10)



K=200 visual words

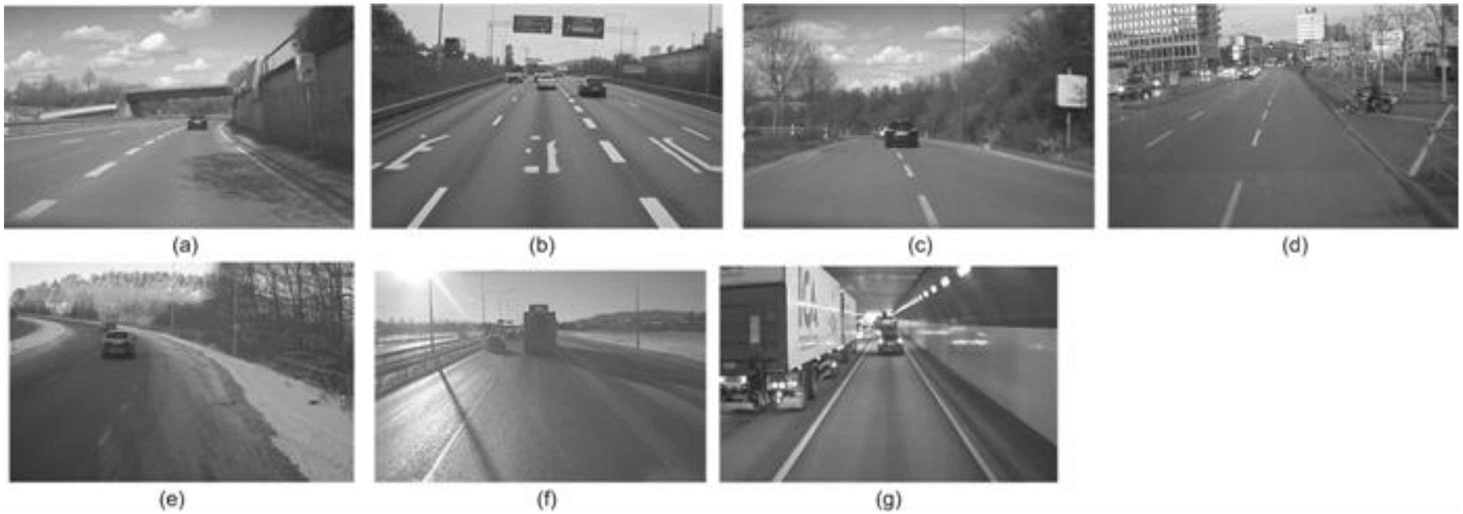
H-CENTRIST, example histogram representation for one image (90th frame)



K'=200 visual words

Experimental setup 1/2: Dataset + parameterization

- Video database was split in 7 classes:



Fusion: vector concatenation
SVM kernels for comparison: X2 radial-basis kernel, HI kernel
Grid partitions for comparison: $\{[n_x] \times [n_y] \times [n_t]\} = \{[1,2,3], [1,2,3], [3,6,9]\}$.

Experimental setup 2/2: input format + libs

➢ technical characteristics:
- CMOS HDR camera: wide 752x480 resolution and about 40° horizontal field of view optics. @30fps.

→Subsampling applied: R= 10 (3fps)

→ The average dimensions of the video data corresponding to cropped videos with duration of 2 minutes are therefore 752x480x3600 (frames before sampling).

Libs publicly available used:

- MBH computation: http://lear.inrialpes.fr/people/wang/dense_trajectories
- CENTRIST, HIK clustering: https://sites.google.com/site/wujx2001/home/libhik
- Classification: LibSVM, http://www.csie.ntu.edu.tw/˜cjlin/libsvm

Aux:
OpenCV library (tested with OpenCV-2.4.2)
ffmpeg library (tested with ffmpeg-0.11.1)
boost libraries (tested with boost_1_49_0)

Dynamic Scene classification results (1/3)

Scene Classes	**Mean Performance (%) per scene class**				
	Static (CENTRIST)	***Dynamics***			***Static+ Dynamics***
		MBH_x	***MBH_y***	***MBH***	
highway-smooth	83.6	71.6	73.2	74.8	86.2
highway-traffic	82.4	69.6	70.9	72.0	88.6
rural	73.3	63.4	66.1	67.9	74.8
urban	**85.2**	72.2	**74.6**	78.2	**89.1**
snow	71.2	60.5	69.3	70.7	73.8
back-lighting	72.3	34.6	41.2	43.8	68.4
tunnel	84.1	**77.2**	74.1	**79.5**	88.9
Avg (%)	78.9	64.2	67.0	69.5	81.4

Dynamic Scene classification results (2/3)

$[n_x \times n_y \times n_t]$ grid	**Mean Performance over all classes (%)**		
	MBH$_x$	***MBH$_y$***	***MBH***
1x1x3 (1 sec history)	34.9	41.8	44.5
1x1x6 (2 secs history)	38.2	42.9	48.4
1x1x9 (3 secs history)	51.2	54.1	56.2
3x3x3 (1 sec history)	46.7	48.8	50.9
3x3x6 (2 secs history)	49.3	52.1	59.7
3x3x9 (3 secs history)	64.2	67.0	**69.7**

SVM kernel	**Mean Performance (%) over entire dataset**				
	Static (CENTRIST)	***Dynamics***			***Static+ Dynamics***
		MBH$_x$	***MBH$_y$***	***MBH***	
RBF-Chi_sq	75.4	60.8	63.9	65.8	77.2
HI	78.9	64.2	67.0	69.7	81.4

Dynamic Scene classification results (3/3)

<table>
<tr><th rowspan="2">Total time (secs)</th><th colspan="5">Percentages of time spent during training</th></tr>
<tr><th colspan="3">Descriptors</th><th rowspan="2">Save features</th><th rowspan="2">Clustering</th></tr>
<tr><td rowspan="2">428.6</td><td>Opt. Flow</td><td>CENTRIST</td><td>$MBH_{y,y}$</td></tr>
<tr><td>29%</td><td>9%</td><td>19%</td><td>15%</td><td>28%</td></tr>
<tr><td></td><th colspan="5">Percentages of time spent during testing per image</th></tr>
<tr><td rowspan="2">1.95</td><td colspan="3">Descriptors extraction and assignment</td><td colspan="2">Classification</td></tr>
<tr><td colspan="3">85%</td><td colspan="2">15%</td></tr>
</table>

Results summary

- Best Algorithms for BoF creation:
  - CENTRIST on spatial grid -- [31 x 200] = >6200 dimensions
  - MBHx, MBHy on spatio-temporal --- {n_x=3 x n_y=3 x n_t=9 x 200} = >16200 dimensions

  - Histogram Intersection kernel k-means for clustering into 200-length codebook
  - SVM classifier with HI kernel

- Empirical observations:
  - motion analysis in different directions can help
  - motion helsp mores in busy scenes
  - faster motion feature extraction is needed or regions of interest should be selected.

Future work

- large dataset evaluation in order to quantify empirical observations
- investigate other motion features (faster than optical flow)
- include other motion compensation
- investigate robustness of the algorithm in fast scene changes

This is the final event

20-21 November 2013

EUROGRESS, Aachen (Germany)

For further information:

interactIVe
www.interactive-ip.eu
Anja Winzer: anja.winzer@eict.de **Evi Brousta:** p.brousta@iccs.gr

eCoMove
www.ecomove-project.eu
Julie Castermans: j.castermans@mail.ertico.com

SEVENTH FRAMEWORK PROGRAMME

Projects co-funded and supported by the European Union

Thank you.

Contact us:
abolov@iccs.gr
http://i-sense.iccs.ntua.gr/members/members-list/54/189