

Social behavior recognition in continuous video

Supplementary Material

Xavier P. Burgos-Artizzu^{*}, Piotr Dollár[†], Dayu Lin⁺, David J. Anderson^{*}, Pietro Perona^{*}

This pdf shows some supplementary information on the benchmarking performed in the paper. For more information go to www.vision.caltech.edu/Video_Datasets/CRIM13.

Detector + Descriptor	Codebook Size			
	250	500	1000	2000
Cuboids+Pca-Sift	24.6%	21.2%	22.3%	21.7%
Cuboids+Hog3D	18.2%	17.9%	15.9%	18.0%
Cuboids+Hog/Hof	18.8%	19.8%	17.1%	19.7%
Harris3D+Pca-Sift	20.9%	16.8%	18.7%	16.1%
Harris3D+Hog3D	16.8%	18.7%	18.2%	17.5%
Harris3D+Hog/Hof	13.6%	15.5%	14.9%	15.1%

Table 1. Codebook size benchmarking for all Spatio temporal interest point detectors+descriptors on validation set. None benefits from a codebook size larger than **500**, in contrast with their use in current human datasets where a Codebook size of **4000** is recommended. This is probably due to the much richer and varied visual information the human body can provide compared with a mouse. Codebook size of 4000 was not benchmarked for computational reasons.

Size	short behaviors			long behaviors		
	attack	chase	walk away	eat	clean	copulation
75 frames	48%	69%	74%	41%	41%	38%
615 frames	30%	54%	60%	51%	44%	57%

Table 2. Two sliding window sizes are used to compute trajectory features. Shorter window size (75) is more suited to detect short behaviors, while longer window size (615) detects better longer behaviors. Results from test on benchmark videos without context. See Figure 5(a) in paper.