

Fast Boosting based Detection using Scale Invariant Multimodal Multiresolution Filtered Features

Overview

- **Boosting based** sliding window solution for object detection
- Multimodal features from: color, motion and depth based channels
- Multiresolution filtering of channels ⇒ multimodal multiresolution filtered channels (MM-MRFC)
- Feature scale correction scheme \Rightarrow scale invariant features
- Exploitation of context: 2D and 3D context channels
- Running at over **25 FPS** (GPU)



Multimodal Multiresolution Filtered Channels

Each modality captures objects in a different way and provides different invariances:

Computation of multimodal intensity, gradient magnitude and orientation

Filtering scheme for each modality:

- Low pass filtering (iterative box filtering)
- **High pass** filtering (vertical & horizontal difference)



Arthur Costea, Robert Varga,

Technical University of Cluj-Napoca, Romania



Classification features:

- Classifier:

Multiscale Detection using MM-MRFC **3D context channels** Multiscale sliding windows channel samples - grid adapted to window size 4096 x 5 level decision trees (AdaBoost + bootstrapping) One classifier model for all scales! Scale invariant classification features Use of one image scale & multiple sliding window scales **Fast**, but the raw channel features are not scale invariant Results Smoothings and derivatives of multimodal intensity can be easily approximated! Ratio between feature at scale s and at original scale: $r_f(s) = \frac{f(s,x,y)}{s(x,y)}$ $f\left(1,\frac{x}{s},\frac{y}{x}\right)$ Channel Type Caltech MR - reasonable Color MRFC no SC 24.46 Theoretical estimation: Empirical estimation: MRFC E-SC 22.69 Intensity or smoothing: $r_I(s) = 1$ MRFC T-SC 22.84 +orig —orig-fit + 2D spatial 20.80 Gradient magnitude: $r_M(s) = \frac{1}{s}$ 1.4 + s1+ 2D symmetry 18.26 —s1-fit Vertical or horizontal derivative: $r_{Idx}(s) = \frac{1}{2}$ 17.29 Motion + SDt 1.3 -s2-fit+s3 + MM-MRFC 16.11 -s3-fit Feature scale correction: Caltech pedestrian benchmark: 0.8 0.2 0.4 0.6 0.2 0.4 0.6 0.8 = = 22% SpatialPoolin 22% SCCPriors **2D context channels** = = 21% TA-CNN = = 19% CCF 18% Checkerboar 17% CCF+CF 17% Checkerboa 12% Ours-MM-MRF 2D position channels: = = = 12% DeepParts – – 12% CompACT–D = = 10% MS_CNN Normalized vertical and horizontal position 10% RPN+BF false positives per image Classifier can constrain 2D position ChnFtrs
CrossTalk
Roerei
ACF-Caltech
WordChannels 2D symmetry channels: SDN FastCF LFOV SquaresChnFti $S_{r}(x,y) = \sum_{i=r/2}^{r} \left(\frac{D_{x}(x-i,y) - D_{x}(x+i,y)}{D_{x}(x-i,y) + D_{x}(x+i,y)} \right)^{2}$ SpatialPooling
DeepCascade+
Multiresolution
CCF
Multiresolution+S
RotatedFilters
MM-MRFC
CompACT-Deep

$$f(s, x, y) = r_f(s) \cdot f(1, \frac{x}{s}, \frac{y}{s})$$

- Compute S_6 , S_{12} , S_{18} , S_{24} and their sum



Sergiu Nedevschi



Results for different feature subsets on Caltech and KITTI (pedestrian):

Cont	text Type	KITTI AP				
		Easy	Moderate	Hard		
Color	MRFC no SC	62.84	59.98	51.10		
	MRFC	67.14	61.45	52.76		
	+ 2D spatial	69.58	63.83	54.83		
	+ 2D symmetry	70.28	64.75	55.66		
3D stereo	+ 3D spatial	77.88	70.30	60.63		
	+ 3D geometric	77.97	70.61	61.47		
	+ MRFC	82.53	74.82	65.95		
3D LIDAR	+ 3D spatial	$\overline{77.88}$	70.93	61.91		
	+ 3D geometric	79.92	72.48	63.13		
	+ MRFC	84.26	76.34	67.18		
Motion	+MRFC	85.25	77.72	68.28		

KITTI object detection benchmark

Method	Input	Time		Cars			Pedestrians		
	-			Easy	Moderate	Hard	Easy	Moderate	Hard
FusionDPM [34]	C DL	30s	CPU	-	-	-	59.51	46.67	42.05
ACF [15]	С	1s	CPU	-	-	-	60.11	47.29	42.90
VOTE-3Deep [19]	C DL	1.5s	CPU	76.79	68.24	63.23	68.39	55.37	52.59
MV-RGBD-RF [23]	C DL	4s	GPU	-	-	-	73.30	56.59	49.63
FilteredICF [47]	С	2s	CPU	-	-	-	67.65	56.75	51.12
DeepParts [39]	С	1s	GPU	-	-	-	70.49	58.67	52.78
CompACT-Deep [5]	С	1s	GPU	-	-	-	70.69	58.74	52.71
Regionlets [42]	С	1s	CPU	84.75	76.45	59.70	73.14	61.15	55.21
Faster-RCNN [35]	С	2s	GPU	86.71	81.84	71.12	78.86	65.90	61.18
Mono 3D [6]	С	4.2s	GPU	92.33	88.66	78.96	80.35	66.68	63.44
3DOP [7]	C DS	3s	GPU	93.04	88.64	79.10	81.78	67.47	64.70
SDP+RPN [44]	С	0.4s	GPU	90.14	88.85	78.38	80.09	70.16	64.82
MS-CNN [4]	С	0.4s	GPU	90.03	89.02	76.11	83.92	73.70	68.31
MM-MRFC	C DL F	0.05s	GPU	90.63	88.45	78.32	82.18	70.02	64.74

50 + RPN+BF