Frame size = 540x960

Total frames = 590

I ran the code in loops to get various combinations. Area Threshold had 5 levels: 20K, 30K, 50K, 70K, and 90K.

WinStride and Padding each had 4 levels: 4x4, 8x8, 16x16, and 32x32.

Scale had 7 level from 1 to 1.06 with a step of 0.01.

This resulted in 560 rows. Full table is attached in excel file. Table below represents top 10 combinations with the biggest number of frames where person was detected.

Table 1. Top 10 combinations with the biggest number of frames where person was detected

Row #	Area Threshold	WinStride	padding	scale	# of frames w/people
36	20000	(4, 4)	(32, 32)	1.02	428
20	20000	(4, 4)	(32, 32)	1.01	410
52	20000	(4, 4)	(32, 32)	1.03	402
35	20000	(4, 4)	(16, 16)	1.02	398
19	20000	(4, 4)	(16, 16)	1.01	389
68	20000	(4, 4)	(32, 32)	1.04	381
148	30000	(4, 4)	(32, 32)	1.02	379
51	20000	(4, 4)	(16, 16)	1.03	369
84	20000	(4, 4)	(32, 32)	1.05	363
132	30000	(4, 4)	(32, 32)	1.01	359

From the table, where the # of frames where the person was detected equals 0, only WinStride = 4x4 did not appear. The greatest number of frames is produced by combination of any area threshold, WinStride = 4x4, padding = 32x32, scale = 1.02.

From what I've noticed in this video, the smaller the WinStride and greater the padding the better is the prediction. Scale plays lesser role, but the smaller the scale (except 1) the better the prediction and smaller the false positive rate. Area Threshold played the smallest role and mostly influenced only the time required to process the frame.