

# 360VSOD

Yi Zhang

## 1 Categories

**5,300 key frames (aiming to 10K) from 69 videos (28 seconds on average, 64,354 frames in total) are currently annotated with a key frame interval of 6.** Note that the 7 videos with both front/rear-initialization-based w/-w/o-sound fixation maps are marked with \* in Table. 1.

### 1.1 Environment

#### 1.1.1 rural: 7 videos

#### 1.1.2 urban: 11 videos

### 1.2 Social Events

#### 1.2.1 conversation A/B/C: 8/10/6 videos

The videos with conversation scenes are divided into three sub-classes (conversation A/B/C), based on the ranking (ascending) of presenting of foreground people.

#### 1.2.2 speech A/B: 12/7 videos

The videos with speech scenes are divided into two sub-classes (speech A/B), based on the ranking (ascending) of presenting of foreground people.

#### 1.2.3 music: singing: 8 videos

#### 1.2.4 music: instrumental performance: 8 videos

#### 1.2.5 music: street music: 4 videos

#### 1.2.6 sports: racing: 9 videos

#### 1.2.7 sports: ball game: 4 videos

#### 1.2.8 sports: others: 7 videos

#### 1.2.9 crowd: 8 videos

## 2 Attributes

### 2.1 Illumination

**nL:** Natural light

**aL:** Artificial light

**mL:** Miscellaneous lights

### 2.2 Camera Status

**sA:** Shaking

**mS:** Moving slow

**mF:** Moving fast

**sT:** Static

### 2.3 Language Involving

**w/:** Language-involved

**w/o:** No language-involved

Table 1: *360vSOD statistics*  
109 videos with both w/-w/o-sound fixation maps.

Name	Frames	Duration	Resolution	Categories	Attributes	Task	Key Frames	No.
–1LM <sub>1</sub>	913	15s	3840*2048	conversation/A	nL/mS/w/	-	149	1
–1LM <sub>2</sub>	990	17s	3840*2048	conversation/A	nL/sT/w/	-	161	2
–1LM <sub>3</sub>	1,320	22s	3840*2048	conversation/A	nL/mS/w/	-	219	3
–JBr <sub>1</sub>	817	27s	3840*2160	conversation/A	nL/sT/w/	-	137	4
–UpF	965	32s	3840*2048	conversation/A	aL/sT/w/	-	161	5
–ZuX	936	31s	3840*2048	conversation/A	nL/sT/w/	-	155	6
con.	922	31s	3840*2048	conversation/A	aL/sT/w/	-	153	7
con.2	977	33s	3840*2048	conversation/A	aL/sT/w/	-	103	8
–Ocf <sub>1</sub>	870	29s	3840*1920	conversation/B	nL/sT/w/	-	145	9
–Ocf <sub>2</sub>	900	30s	3840*1920	conversation/B	nL/sT/w/	-	149	10
–ey9	1,830	37s	3840*2048	conversation/B	aL/sT/w/	-	305	11
–Mzc <sub>4</sub>	886	35s	3840*1920	conversation/B	mL/sT/w/	-	95	12
–Ngj <sub>1</sub>	960	32s	3840*2160	conversation/B	mL/sT/w/	-	135	13
–P4K	642	21s	3840*2160	conversation/B	nL/sA/w/	-	107	14
–PT4	1,110	37s	3840*1920	conversation/B	nL/sT/w/	-	-	15
–Uy5 <sub>1</sub>	300	10s	3840*2048	conversation/B	aL/mF/w/	-	50	16
–Uy5 <sub>2</sub>	642	21s	3840*2048	conversation/B	aL/sT/w/	-	107	17
Ellen2	900	30s	3840*1920	conversation/B	aL/sT/w/	-	149	18
–MYm	1,125	45s	3840*2048	conversation/C	aL/sT/w/	-	101	19
–Mzc <sub>1</sub>	875	35s	2048*1024	conversation/C	nL/sT/w/	-	-	20
–Mzc <sub>2</sub>	750	30s	3840*1920	conversation/C	mL/sT/w/	-	-	21
–Mzc <sub>3</sub>	775	31s	3840*1920	conversation/C	mL/sT/w/	-	-	22
–Mzc <sub>5</sub>	900	36s	3840*1920	conversation/C	nL/sT/w/	-	-	23
–Mzc <sub>6</sub>	750	30s	3840*1920	conversation/C	nL/sT/w/	-	-	24
–osu	780	26s	3840*1920	speech/A	nL/sT/w/	-	130	25
–4Si	900	30s	3840*1920	speech/A	nL/sT/w/	-	149	26
–72f <sub>1</sub>	1,150	23s	3840*1920	speech/A	nL/sT/w/	-	191	27
–72f <sub>2</sub>	982	20s	3840*1920	speech/A	nL/sT/w/	-	163	28
–72f <sub>3</sub>	1,204	24s	3840*1920	speech/A	nL/sT/w/	-	201	29
–72f <sub>4</sub>	1,612	32s	3840*1920	speech/A	nL/sT/w/	-	101	30
–idL	825	28s	3840*1920	speech/A	aL/sT/w/	-	137	31
–IRG	964	32s	3840*1920	speech/A	nL/sT/w/	-	161	32
–J0Q <sub>2</sub>	309	10s	3840*1920	speech/A	aL/sT/w/	-	49	33
–jb5	750	30s	3772*1886	speech/A	mL/sT/w/	-	125	34
–n52	780	26s	3840*2160	speech/A	nL/sA/w/	-	129	35
Ellen	900	30s	3840*1920	speech/A	aL/sT/w/	-	143	36
–5h9	878	37s	3840*2000	speech/B	aL/sT/w/	-	147	37
–ey9 <sub>2</sub>	2,080	42s	3840*2048	speech/B	aL/sT/w/	-	347	38
–FAe	1,313	22s	3840*2048	speech/B	aL/sT/w/	-	219	39
–g4f	840	28s	3840*2048	speech/B	mL/sT/w/	-	-	40
–Gq4	840	35s	3840*2000	speech/B	aL/sT/w/	-	139	41
–J0Q <sub>1</sub>	690	23s	3840*1920	speech/B	aL/sT/w/	-	115	42
–o7J	870	29s	3840*1920	speech/B	nL/sT/w/	-	145	43
–1An	1,533	61s	3840*2048	singing	aL/sT/w/	-	169	44
–6QU	1,710	29s	3840*2048	singing	aL/sT/w/	-	255	45
–dBM	778	31s	3840*2048	singing	aL/sT/w/	-	129	46
–fJf	775	31s	3840*2048	singing	aL/sT/w/	-	-	47
–gTB	870	29s	3840*1920	singing	aL/sT/w/	-	145	48
–Oak	1,050	35s	3840*1920	singing	aL/sT/w/	-	175	49
–SdG	828	28s	3840*2160	singing	aL/sT/w/	-	137	50
band	840	28s	3840*1920	singing	mL/sT/w/	-	139	51
–4fx	650	21s	3840*1920	instruments	mL/sT/w/o	-	89	52

Name	Frames	Duration	Resolution	Categories	Attributes	Task	Key Frames	No.
– <i>Bvu</i>	1,035	35s	3840*1920	instruments	aL/sT/w/o	-	173	53
– <i>eGG</i>	780	26s	3840*2048	instruments	nL/sT/w/o	-	129	54
– <i>MDw</i>	960	32s	3840*1920	instruments	aL/sT/w/	-	159	55
– <i>SBB</i>	819	33s	3840*1920	instruments	aL/sT/w/o	-	-	56
– <i>SJI</i>	900	30s	3840*1920	instruments	mL/sT/w/o	-	-	57
– <i>TCU</i>	625	25s	3840*1920	instruments	mL/sT/w/o	-	105	58
– <i>V3z</i>	915	31s	3840*1920	instruments	aL/sT/w/o	-	153	59
– <i>69A</i>	515	18s	3840*2160	street music	nL/sT/w/o	-	77	60
– <i>bO4</i>	990	33s	3840*1920	street music	aL/sA/w/o	-	165	61
– <i>JJ–</i>	900	30s	3840*1920	street music	nL/sA/w/o	-	-	62
– <i>MFV</i>	750	30s	1920*960	street music	nL/sT/w/o	-	125	63
– <i>0d8</i>	920	31s	3840*2160	crowd	aL/sT/w/	-	-	64
– <i>7P3</i>	691	23s	3840*1920	crowd	aL/sT/w/o	-	-	65
– <i>Di3<sub>1</sub></i>	910	30s	3840*1920	crowd	mL/sA/w/o	-	-	66
– <i>Di3<sub>2</sub></i>	770	26s	3840*1920	crowd	aL/sA/w/o	-	129	67
– <i>dQa</i>	900	30s	3840*2160	crowd	nL/sT/w/o	-	-	68
– <i>eqm</i>	750	25s	7680*3840	crowd	mL/sT/w/o	-	125	69
– <i>j7D</i>	780	26s	3840*1920	crowd	nL/sA/w/o	-	-	70
– <i>Rbg</i>	871	29s	3840*1920	crowd	aL/sT/w/	-	145	71
– <i>3DM</i>	750	30s	3840*2160	racing	nL/sA/w/	-	-	72
*– <i>72f<sub>6</sub></i>	1,632	33s	3840*1920	racing	nL/sA/w/	-	269	73
– <i>fNV</i>	819	27s	3840*2160	racing	nL/sA/w/	-	-	74
– <i>fRM</i>	613	20s	3840*1920	racing	nL/mF/w/o	-	-	75
*– <i>I6E</i>	750	25s	3840*2160	racing	nL/mF/w/o	-	-	76
– <i>IDR</i>	900	30s	3840*2160	racing	nL/mF/w/o	-	-	77
– <i>nwE</i>	870	29s	3840*2160	racing	nL/mF/w/o	-	145	78
– <i>3LH</i>	960	32s	3840*1920	ball game	nL/mS/w/	-	-	79
– <i>HNQ</i>	900	30s	3840*2048	ball game	nL/sT/w/	-	149	80
– <i>Ngj<sub>2</sub></i>	900	30s	3840*2160	ball game	mL/sT/w/o	-	149	81
– <i>nZJ</i>	670	22s	3840*1920	ball game	nL/sT/w/o	-	77	82
– <i>e8D<sub>1</sub></i>	880	29s	3840*2160	other sports	nL/mS/w/o	-	-	83
– <i>gSu</i>	829	28s	1920*960	other sports	nL/sT/w/o	-	-	84
– <i>kZB</i>	900	30s	3840*1920	other sports	aL/sT/w/	-	147	85
– <i>RrG</i>	840	28s	3840*1920	other sports	nL/mS/w/	-	139	86
– <i>SZY</i>	750	25s	3840*2160	other sports	nL/sA/w/	-	85	87
*– <i>ZOq<sub>8</sub></i>	975	33s	3840*2160	other sports	nL/mF/w/o	-	-	88
*– <i>ByB</i>	1,040	35s	3840*2032	rural	nL/sT/w/o	-	-	89
– <i>gy4</i>	750	25s	3840*1920	rural	nL/mF/w/	-	125	90
– <i>I–<sub>1</sub></i>	900	30s	3840*1920	rural	nL/sT/w/o	-	-	91
*– <i>JbG</i>	1,150	23s	3840*2048	rural	nL/sT/w/o	-	-	92
– <i>nDu</i>	780	26s	7680*3840	rural	nL/sT/w/o	-	-	93
– <i>dd3</i>	750	25s	5760*2880	urban	nL/sT/w/o	-	-	94
– <i>dpf</i>	750	25s	3840*2160	urban	nL/mF/w/o	-	125	95
*– <i>G8p</i>	840	28s	2560*1440	urban	nL/sT/w/o	-	139	96
– <i>gqL</i>	810	27s	5760*2880	urban	nL/sT/w/o	-	-	97
– <i>LTh</i>	900	30s	3840*1920	urban	nL/sT/w/o	-	99	98
– <i>N6U</i>	900	30s	3840*2048	urban	nL/sT/w/o	-	-	99
– <i>Oue</i>	944	31s	3840*1920	urban	nL/sT/w/o	-	-	100
– <i>Rbg<sub>2</sub></i>	692	23s	3840*1920	urban	nL/sT/w/o	-	-	101
*– <i>RSY</i>	810	27s	3840*1920	urban	nL/sT/w/o	-	117	102