

# Supplemental Material

## N-MuPeTS: Event Camera Dataset for Multi-Person Tracking and Instance Segmentation

Tobias Bolten<sup>1</sup> <sup>a</sup>, Christian Neumann<sup>1</sup> <sup>b</sup>, Regina Pohle-Fröhlich<sup>1</sup> and Klaus D. Tönnies<sup>2</sup>

<sup>1</sup>*Institute for Pattern Recognition, Hochschule Niederrhein, Krefeld, Germany*

<sup>2</sup>*Department of Simulation and Graphics, University of Magdeburg, Germany*

{tobias.bolten, christian.neumann, regina.pohle}@hs-niederrhein.de, klaus@isg.cs.uni-magdeburg.de

## 1 Label Mask Generation

Refer to Section 4.2 and 4.3 of main paper.

The main part of the dataset preparation is based on the automated generation of label masks from APS color images using color features provided by the actors' clothing. The quality of this processing step cannot be determined quantitatively due to the lack of references such as manual ground-truth annotations. Yet, subjective and qualitative ratings are possible. Figure 2 shows randomly sampled results of the label generation. These examples show the high quality of the generated labels. The left column shows a cropped APS input image, while the middle column shows the automatically generated label mask.

The mapping quality achieved between the APS and DVS fields of view also plays an important role. Examples are shown in the right column of the figure. The corresponding time window of the DVS event stream is shown as a projection along time. DVS events are shown in gray unless they are under the assigned label mask. In these cases, the events are colored according to the label mask.

The full set of propagated labels for the entire dataset is also visualized in a video<sup>1</sup>.

## 2 Dataset Sequences

Refer to Section 5 of main paper.

The dataset is divided into individual sequences by distinguishing different quality classes. A new sequence is started each time the quality class is changed. The highest quality class 1 contains only

recordings that do not contain unwanted artifacts, e.g. overlap with the static mask, environmental influences, or activity in the background. The Section 6 provides detailed statistics and number of occurrences per annotation in relation to the assigned quality classes within the dataset. The information here is cumulative with respect to higher quality classes. This means, for example, that the information for quality class 2 also includes class 1.

Multiple annotations can be assigned to a single sequence. For example, a single sequence can contain WALKING and OCCLUSION (compare to given annotation file in Listing 1).

Due to the high quality requirement, the scenes of quality class 1 are naturally rather short. Especially with tracking methods, a higher sequence length is of interest. By including quality class 2, temporally longer sequences can be formed. Temporally consecutive sections of quality 1 and 2 can be combined as long as they are not interrupted by quality 3. Figure 1 shows this exemplarily, in which sequences Seq. id00010 to Seq. id00017 can be combined.

An overall listing of all sequences, their quality rating, performed annotation, and durations is given in Table 6.3.

## 3 Dataset Annotations

For each sequence, the annotations created are included with the dataset in a separate JSON file. The start, end and length information in these files refer to a 25 ms sliding time window (analogous to the ‘frame number’ of an APS camera) of the corresponding and plain DVS recording.

The JSON description starts with a block that contains details about the entire scene. In addition to the sequence ID as a counter, the quality class assignment, and the total length of this sequence, the aggregated\_annotation field contains a union of all annotations for everyone in the sequence. An ex-

<sup>a</sup>  <https://orcid.org/0000-0001-5504-8472>

<sup>b</sup>  <https://orcid.org/0000-0002-0871-8629>

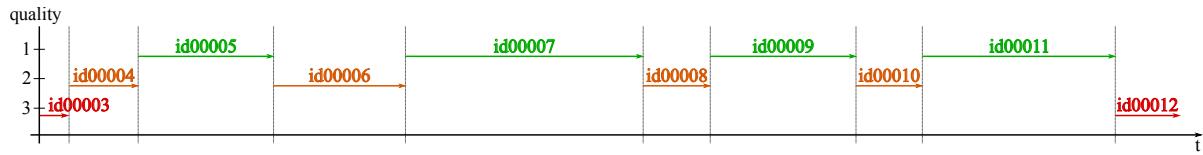


Figure 1: Possible sequence merge from Seq. id00004 to Seq. id00011 when using quality  $\leq 2$  (quality 1 in green, quality 2 in orange, quality 3 in red).

ample is given in the following listing:

```

1  {
2      "id": 338,
3      "quality": 1,
4      "length": 81,
5      "aggregated_annotation": [
6          "occlusion",
7          "walking"
8      ],

```

This is followed by a detailed description per person. For each person, a list of start and end blocks describes the annotation performed in its temporal context:

```

9      "colors": {
10         "CYAN": [
11             {
12                 "start": 13440,
13                 "end": 13520,
14                 "annotation": [
15                     "walking"
16                 ]
17             }
18         ],
19         "BLUE": [],
20         "ORANGE": [],
21         "RED": [
22             {
23                 "start": 13440,
24                 "end": 13514,
25                 "annotation": [
26                     "walking"
27                 ]
28             },
29             {
30                 "start": 13515,
31                 "end": 13520,
32                 "annotation": [
33                     "occlusion",
34                     "walking"
35                 ]
36             }
37         ],
38     }
39 }
```

Listing 1: Single sequence annotation.

These detailed annotations are visualized for the entire dataset in a video<sup>1</sup>.

<sup>1</sup><https://www.youtube.com/watch?v=XpKkfRoDJxA>

## 4 Dataset Files

### 4.1 Structure

The sequence partitioning is mapped to a folder structure where each sequence is operated into its own directory. The name pattern used is as follows:

```

dataset_root
├── id-<$id>-quality<$class>-
    └── len<$duration>-col<$colorIndicator>-
        └── <$aggregatedAnnotation>
            ...
            └── id00338-quality1-len00081-colRC-
                occlusion_walking

```

The used name pattern is as follow:

<\$id> the sequence number, starting at zero, five digits with leading zeros

<\$class> numerical value of quality class indicator, values  $\in \{1, 2, 3\}$

<\$duration> number of time windows

<\$colorIndicator> string encoding the people in this scene by one character if color is included. The order is red (R), orange (O), cyan (C), blue (B)

<\$aggregatedAnnotation> matches the union of person annotations in the given JSON

### 4.2 File Format

The dataset is available for download as a single zip archive<sup>2</sup>. The DVS event data in this archive is available as text files to avoid possible software dependencies and proprietary file formats.

The event data is divided into individual files, each containing a 25 ms time window of the data stream. The file numbering corresponds to the start and end information in the corresponding JSON annotations.

Each line in these text files represents a single DVS event and contains  $(x, y, t, p, l)$ -information as follows:

<sup>2</sup><http://dnt.kr.hsnr.de/DVS-NMuPeTS/>

- $x, y$  Spatial coordinate of the event. The origin is in the upper left corner of the image. The resolution of the used DVS is  $768 \times 640$  pixels.
- $t$  Timestamp of the event, encoded as floating point number, and scaled in milliseconds.
- $p$  Polarity of the DVS event. Values are  $\in \{-1, 0, 1\}$ , indicating a brightness decrease (-1), or brightness increase (1).  
The utilized CeleX-IV sensor determines the polarity of an event based on its transmitted gray value within software. This can lead to events with ‘no change in brightness’ (0).
- $l$  Instance color label. The numerical label values correspond to RED (100), ORANGE (200), CYAN (300) and BLUE (400).

### 4.3 Usage Notes

#### Disk Space Requirements

Be aware of the file size of the dataset. Due to the simple text format used, the included zip archive has a compression ratio of 87.5%. A fully unpacked version of the dataset requires approximately 305 GB. A practical approach is to stream on-the-fly from the archive, or to convert it once to a binary format customized for your application.

#### Sequence Numbering

Technically, the dataset consists of data from two separate recordings. A short temporal interruption of the recording was performed to reconfigure the physical scene (introduction of objects for occlusions). This split is performed at the end of the sequence with Seq. id00326. Note that to indicate this split, the next sequence starts with Seq. id00328.

#### Time Window Numbering

At the start of each recording, an initial temporal synchronization is performed between the DVS and the APS using a visual cue. These time windows are not included in the data. Therefore, the numbering of the files does not start at 0.

The Seq. id00000 starts with a file number of 1103. The first time window of the second recording starts at Seq. id00328 with a file number of 1585. It is important to note that the file names of the text files are only unique within a sequence’s folder structure.

## 5 Shoe Covers

We used cotton shoe covers to cover the shoes of the actors. This allows the actors to wear normal shoes and thus have a natural walking pattern. The color is matched to the suit.

The sewing pattern for the shoe covers is available in the download area. The shoe cover fits typical shoes of up to EU size 45 comfortably. To prevent the shoe cover from separating, we recommend to fold the fabric inside from the heel and stepping onto it. A seam allowance of 10 mm is included.

## 6 Statistics

### 6.1 Quality Class 1

In our paper, we described quality class 1 as including all sequences without any issues.

Table 1: Cumulative durations per color combination in quality class 1

RED	ORANGE	CYAN	BLUE	cumulative duration	
				per color combination	sum per person count
				300	
•				313	
	•			158	
		•		215	882
			•	195	
•	•			213	
•		•		41	
•			•	125	
	•	•		236	701
•			•	76	
	•	•	•	9	
•	•	•		496	
•	•		•	123	
•		•	•	51	687
•	•	•	•	17	
•	•	•	•	350	350

Table 2: Duration statistics per annotation in quality class 1

annotation	number of sequences	mean duration	cumulative duration
RED	137	12.5	1713
ORANGE	123	13.6	1671
CYAN	101	14.0	1416
BLUE	77	12.3	946
BACKGROUND	46	6.5	300
STANDING	96	3.3	312
WALKING	441	5.1	2259
RUNNING	107	4.7	504
RANDOM	18	8.9	160
HELIX	9	7.2	64
FAR	35	4.5	157

Table 3: Occurrence statistics per annotation in quality class 1

annotation	number of occurrences
OCCLUSION	136
EXERCISING	9
KNEELING	13
STOOPED	18
WAVING	9
CROSSING	212
MEET	48
SIDE BY SIDE	94

## 6.2 Quality Class 2

In our paper, we described quality class 2 as corresponding to these minor issues:

- unwanted occlusion, i.e. tree trunks
- one or more person outside of static mask
- incomplete masks, i.e. intersection with static mask

Table 4: Cumulative durations per color combination in quality class 2

RED	ORANGE	CYAN	BLUE	cumulative duration	
				per color combination	sum per person count
				378	
●				385	
	●			182	
		●		259	1052
			●	227	
●	●			222	
●		●		46	
●			●	156	
	●	●		260	782
	●		●	85	
		●	●	14	
●	●	●		589	
●	●		●	154	
●		●	●	57	831
	●	●	●	30	
●	●	●	●	406	406

Table 5: Duration statistics per annotation in quality class 2

annotation	number of sequences	mean duration	cumulative duration
RED	230	8.8	2016
ORANGE	212	9.1	1928
CYAN	173	9.6	1661
BLUE	137	8.2	1127
BACKGROUND	71	5.3	378
STANDING	100	3.2	325
WALKING	658	4.1	2694
RUNNING	142	3.8	542
RANDOM	19	8.7	166
HELIX	14	5.2	72
FAR	79	3.9	305

Table 6: Occurrence statistics per annotation in quality class 2

annotation	number of occurrences
OCCLUSION	239
EXERCISING	9
KNEELING	13
STOOPED	24
WAVING	10
CROSSING	228
MEET	50
SIDE BY SIDE	166

### 6.3 Quality Class 3

In our paper, we described quality class 3 as corresponding to these major issues:

- uninvolved persons in scene
- cars moving in scene
- wildlife in scene, i.e. birds

Table 7: Cumulative durations per color combination in quality class 3

RED	ORANGE	CYAN	BLUE	cumulative duration	
				per color combination	sum per person count
				1140	
●				531	
	●			339	
		●		338	1479
			●	271	
●	●			321	
●		●		73	
●			●	186	
	●	●		279	1004
	●		●	122	
		●	●	23	
●	●	●		765	
●	●		●	177	
●		●	●	73	1048
	●	●	●	34	
●	●	●	●	473	473

Table 8: Duration statistics per annotation in quality class 3

annotation	number of sequences	mean duration	cumulative duration
RED	270	9.6	2599
ORANGE	262	9.6	2510
CYAN	200	10.3	2058
BLUE	163	8.3	1359
BACKGROUND	120	9.5	1140
STANDING	117	3.4	400
WALKING	786	4.4	3465
RUNNING	171	4.1	705
RANDOM	31	7.2	223
HELIX	21	4.6	98
FAR	87	4.3	377

Table 9: Occurrence statistics per annotation in quality class 3

annotation	number of occurrences
OCCLUSION	300
EXERCISING	12
KNEELING	14
STOOPED	37
WAVING	11
CROSSING	282
MEET	55
SIDE BY SIDE	181

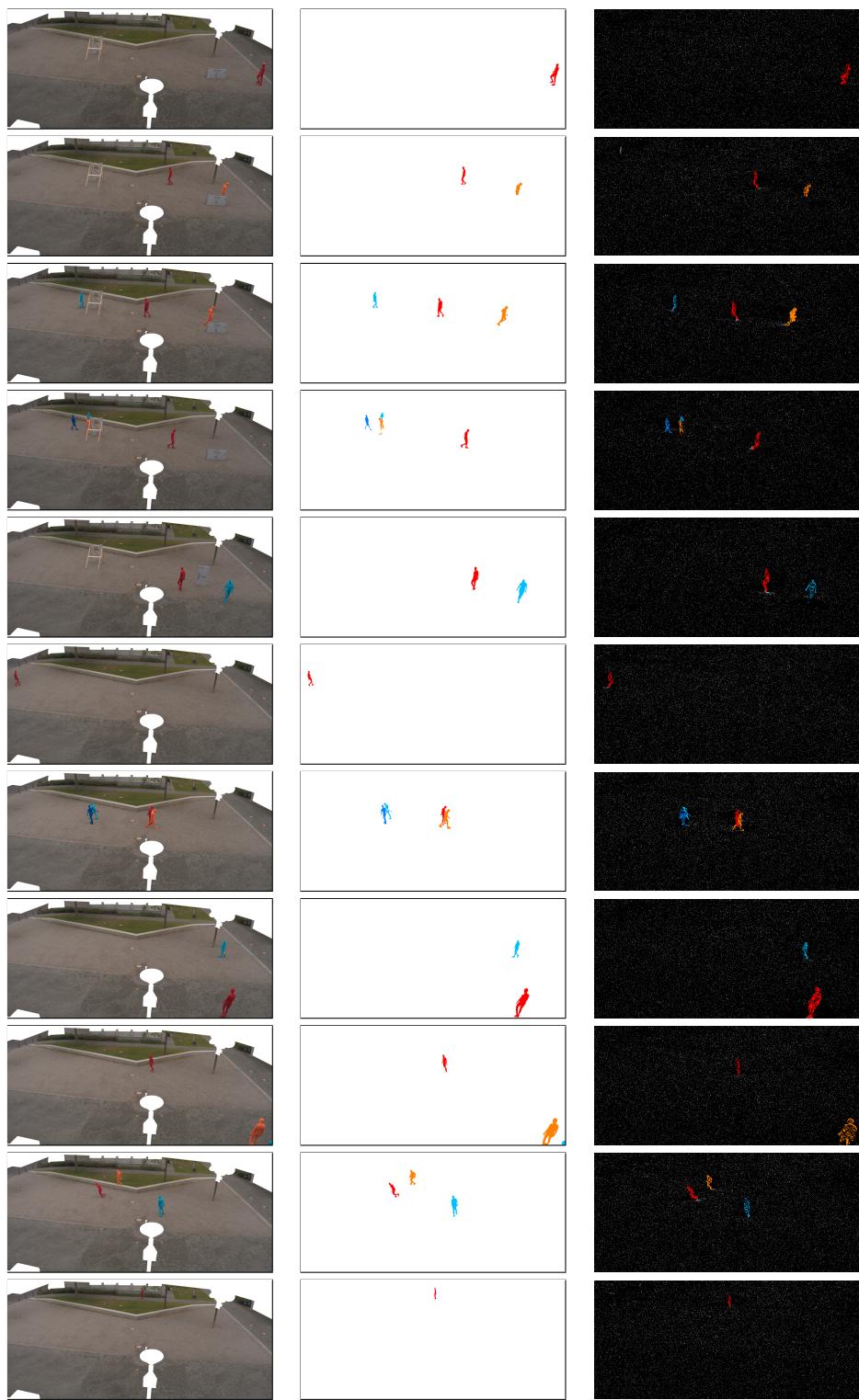


Figure 2: Segmentation examples.

Table 10: Annotation details per sequence

id	quality	persons	empty scene	static occlusion	pose	movement	interaction	different sizes	duration in seconds
00000	3								114.5
00001	3	●							6.0
00002	1	●							6.0
00003	3	●							24.6
00004	2	●							0.9
00005	1	●							17.1
00006	2	●							4.5
00007	1	●							16.0
00008	2	●							1.0
00009	1	●							5.0
00010	2	●							1.0
00011	1	●							13.0
00012	3	●							10.0
00013	1	●							19.5
00014	2	●							5.0
00015	1	●							23.5
00016	2	●							4.6
00017	1	●							7.0
00018	2	●							2.5
00019	1	●							3.5
00020	2	●							4.5
00021	1	●							3.5
00022	2	●							2.0
00023	1	●							6.5
00024	2	●							1.5
00025	1	●							0.5
00026	2	●							7.5
00027	1	●							8.0
00028	2	●							0.5
00029	1	●							1.5
00030	2	●							2.5

id	quality	persons	empty scene	static occlusion	occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds
00031	1		●	●	●						8.5
00032	2		●	●	●						0.9
00033	1		●	●	●						4.6
00034	2		●	●	●						0.5
00035	1		●	●	●						10.0
00036	2		●	●	●						2.0
00037	1		●	●	●						1.0
00038	3		●	●	●						8.0
00039	1		●	●	●						16.5
00040	1		●	●	●						2.5
00041	1		●	●	●						2.0
00042	2		●	●	●						1.5
00043	1		●	●	●						2.0
00044	2		●	●	●						0.6
00045	3		●	●	●						11.1
00046	2		●	●	●						1.2
00047	1		●	●	●						9.7
00048	2		●	●	●						2.0
00049	1		●	●	●						3.5
00050	1		●	●	●						3.6
00051	1		●	●	●						3.5
00052	3		●	●	●						18.6
00053	1		●	●	●						14.0
00054	3		●	●	●						7.5
00055	1		●	●	●						8.0
00056	2		●	●	●						1.2
00057	3		●	●	●						3.3
00058	2		●	●	●						1.0
00059	1		●	●	●						14.0
00060	2		●	●	●						0.5
00061	1		●	●	●						4.0
00062	2		●	●	●						1.5

id	quality	persons				empty scene background	static occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds
		red	orange	cyan	blue								
00063	1	●								●			8.5
00064	3	●	●							●			11.0
00065	2	●	●							●			4.0
00066	1	●	●							●			23.0
00067	2	●	●							●			3.5
00068	1	●	●							●			3.0
00069	2	●	●							●			2.0
00070	1	●	●							●			6.0
00071	2	●	●							●			1.5
00072	3	●	●							●			38.0
00073	1	●	●							●			3.0
00074	3	●	●							●			5.5
00075	1	●	●							●			9.5
00076	2	●	●							●			0.8
00077	3	●	●							●			6.3
00078	2	●	●							●			1.0
00079	1	●	●							●			6.5
00080	3	●	●							●			5.0
00081	1	●	●							●			4.5
00082	3	●	●							●			14.0
00083	1	●	●							●			5.5
00084	3	●	●							●			4.5
00085	1	●	●							●			8.0
00086	3	●	●							●			5.5
00087	1	●	●							●			11.5
00088	2	●	●							●			4.5
00089	1	●	●							●			9.1
00090	1	●	●							●			25.0
00091	1	●	●							●			9.5
00092	2	●	●							●			2.0
00093	1	●	●							●			8.5
00094	2	●	●							●			8.5



id	quality	persons			empty background scene	static occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds
		red	orange	cyan								
00127	2				●						●	2.0
00128	1				●						●	8.0
00129	2				●						●	12.5
00130	1				●						●	7.5
00131	3				●						●	5.0
00132	1				●						●	14.6
00133	3				●						●	22.0
00134	1				●						●	6.0
00135	3				●						●	33.5
00136	1				●						●	3.0
00137	2				●						●	2.0
00138	1				●						●	6.0
00139	1				●						●	58.0
00140	3				●						●	36.5
00141	2				●						●	4.0
00142	1				●						●	13.5
00143	2				●						●	2.0
00144	1				●						●	54.0
00145	2				●						●	2.0
00146	1				●						●	22.5
00147	3				●						●	8.0
00148	1				●						●	9.2
00149	3				●						●	14.4
00150	1				●						●	9.5
00151	2				●						●	0.5
00152	1				●						●	6.4
00153	2				●						●	0.6
00154	1				●						●	5.0
00155	3				●						●	15.5
00156	1				●						●	16.5
00157	2				●						●	2.5
00158	1				●						●	19.5

id	quality	persons	empty scene	static occlusion	occlusion	exercise	pose	movement	interaction	different sizes	duration in seconds	
00159	3					●						7.0
00160	1					●						1.0
00161	3					●						7.5
00162	1					●						15.0
00163	1					●						8.5
00164	3					●						17.0
00165	1					●						8.5
00166	2					●						1.5
00167	1					●						11.0
00168	3					●						5.0
00169	1					●						19.5
00170	3					●						9.1
00171	1					●						21.1
00172	3					●						27.5
00173	1					●						37.5
00174	2					●						1.5
00175	1					●						11.5
00176	2					●						1.5
00177	1					●						15.0
00178	2					●						43.0
00179	3					●						51.1
00180	1					●						35.5
00181	3					●						11.6
00182	2					●						4.0
00183	1					●						3.5
00184	2					●						3.0
00185	1					●						3.5
00186	2					●						1.5
00187	1					●						15.0
00188	3					●						44.1
00189	1					●						1.5
00190	2					●						0.5

id	quality	persons	empty scene	static occlusion	occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds	
00191	1				●							10.5
00192	2				●	●	●	●				1.5
00193	1				●	●	●	●				11.0
00194	2				●	●	●	●				1.0
00195	1				●	●	●	●				1.0
00196	2				●	●	●	●				1.0
00197	1				●	●	●	●				4.2
00198	2				●	●	●	●				1.4
00199	1				●	●	●	●				11.5
00200	2				●	●	●	●				11.5
00201	3				●	●	●	●				3.5
00202	1				●	●	●	●				22.0
00203	3				●	●	●	●				15.6
00204	3				●							240.8
00205	1				●							5.5
00206	2				●							2.5
00207	3				●							15.0
00208	1				●							4.5
00209	2				●							11.0
00210	1				●							8.0
00211	2				●							2.5
00212	1				●							4.0
00213	2				●							7.5
00214	1				●							16.0
00215	1				●							41.7
00216	3				●							19.7
00217	1				●							12.0
00218	3				●							16.1
00219	3				●							5.1
00220	1				●							14.0
00221	2				●							20.5
00222	3				●							3.1

id	quality	persons	empty scene	static occlusion	occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds
00223	1			●							21.6
00224	1			●							10.0
00225	1			●							22.2
00226	2			●							7.0
00227	1			●							22.1
00228	1			●							5.0
00229	1			●							12.0
00230	3			●							22.2
00231	1			●							6.6
00232	2			●							2.6
00233	1			●							3.6
00234	2			●							10.1
00235	1			●							13.1
00236	2			●							17.1
00237	1			●							12.6
00238	2			●							10.6
00239	1			●							15.7
00240	2			●							12.1
00241	1			●							9.6
00242	3			●							2.0
00243	1			●							12.0
00244	2			●							10.1
00245	1			●							21.8
00246	3			●							12.0
00247	1			●							13.9
00248	2			●							17.1
00249	1			●							8.7
00250	2			●							2.1
00251	1			●							8.2
00252	1			●							10.8
00253	1			●							18.9
00254	2			●							3.6

id	quality	persons	empty scene	static occlusion	occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds
00255	2				●						11.7
00256	1				●						7.3
00257	2				●						2.6
00258	1				●						3.6
00259	2				●						9.1
00260	1				●						14.3
00261	2				●						3.1
00262	1				●						25.2
00263	2				●						1.0
00264	1				●						6.2
00265	2				●						1.4
00266	1				●						37.6
00267	2				●						1.6
00268	1				●						2.9
00269	2				●						1.7
00270	1				●						15.1
00271	2				●						4.1
00272	1				●						10.3
00273	2				●						1.6
00274	1				●						4.7
00275	2				●						1.6
00276	1				●						3.1
00277	2				●						1.6
00278	1				●						18.6
00279	2				●						4.8
00280	2				●						3.7
00281	1				●						7.8
00282	2				●						2.6
00283	1				●						14.0
00284	2				●						2.0
00285	1				●						16.6
00286	2				●						4.2

id	quality	persons	empty scene	static occlusion	occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds	
00287	2					●						6.3
00288	1					●						6.7
00289	2					●						1.6
00290	1					●						6.2
00291	2					●						1.0
00292	1					●						4.2
00293	2					●						2.0
00294	1					●						6.2
00295	2					●						1.6
00296	1					●						6.7
00297	2					●						1.1
00298	1					●						4.2
00299	2					●						2.7
00300	3					●						6.8
00301	1					●						6.3
00302	2					●						2.1
00303	1					●						4.2
00304	1					●						8.3
00305	3					●						10.9
00306	1					●						35.4
00307	3					●						16.1
00308	1					●						35.2
00309	2					●						4.3
00310	1					●						14.6
00311	2					●						3.1
00312	1					●						21.1
00313	2					●						2.8
00314	1					●						4.2
00315	2					●						1.0
00316	1					●						56.1
00317	3					●						13.9
00318	1					●						2.6







id	quality	persons				empty scene	static occlusion	exercising	pose	movement	interaction	different sizes	duration in seconds	
		red	orange	cyan	blue									
00415	2	●			●					●				1.5
00416	1	●	●		●					●				9.0
00417	2	●	●		●					●				3.5
00418	3				●					●				23.0
00419	2	●			●					●				4.0
00420	1	●	●		●					●				13.5
00421	1	●	●		●					●				7.6
00422	1	●	●		●					●				14.9
00423	3	●	●		●					●				26.9
00424	1	●	●		●					●				17.6
00425	3	●			●					●				76.6
00426	1				●					●				36.5
00427	3				●					●				62.9