Computational Visual Media

Electronic Supplementary Material

View suggestion for interactive segmentation of indoor scenes

Sheng Yang[1] (✉), Jie Xu[2], Kang Chen[1], and Hongbo Fu[3]

Supporting information to DOI 10.1007/s41095-017-0078-4

This supplementary document contains 4 sections. Section S1 explains the attributes we considered for the preference view model but found dispensable. We also list the learned weights and show the average score of each view. Section S2 gives our graph-cut-based optimization formulation for point cloud segmentation. Section S3 explains our patch similarity measure. The last section gives the recorded information from the user study of interactions (for comparing our approach to the alternative approaches).

S1 View attributes and scores

Here we list the attributes evaluated in lasso regression but got relatively low weights. They were excluded from the final preference view model.

A_{dd}: **depth distribution**. We import depth distribution from Secord et al. [S1]. This attribute is supposed to distinguish between planar distributed views and head on views.

A_{dc}: **depth continuity**. A complicated scene often consists of isolated objects in an interior perspective view, which may result in a wide separated distribution of depth values. To capture this effect we calculate the variance of the weighted interval of depth histogram to describe radial distribution.

A_{o}: **object layout**. For aesthetic consideration, a view with good photographic composition possibly tells the user where he/she should concentrate on. We follow the layout optimization rules summarized by Zhang et al. [S2], and take each discrete patch smaller than 15% on the full image as object regions. We then calculate the distance from power points, the distance from diagonal lines, and visual balance as A_{op}, A_{od}, and A_{ov}, respectively.

A_{rp}: **radial patches**. A view may contain surfaces which are nearly parallel to the view direction. Such radial surfaces emerge as extraordinarily dense points located in parts which may be illegible. Denoting $\mathcal{N}_{\mathcal{P}}$ as the view direction, this attribute is then calculated as

$$A_{\mathrm{rp}} = \frac{\sum_{P_n \in \mathcal{P}} (1 - |\mathcal{N}_n \cdot \mathcal{N}_{\mathcal{P}}|) S_n}{S_{\mathcal{P}}}$$

Figures S1 and S2 show the average scores of some sampled views from 2 scenes by 16 participants on the scale from 1 (worst) to 5 (best). Table S1 shows the learned weights (with attributes normalized) from lasso regression.

S2 Graph-cut segmentation

Our graph-cut optimization [S3] for point cloud segmentation considers color, position, normal, and support relations. Below we give the detailed formulation:

$$E = \sum_{P_i} E_{\mathrm{D}}(P_i, v_i) + \lambda \sum_{P_i, P_j} E_{\mathrm{S}}(P_i, v_i, P_j, v_j)$$

Table S1 Learned weights from lasso regression

Description	Note	Weight
View entropy	β_{ve}	2.4611
Point density	β_{pd}	2.9857
Depth distribution	β_{dd}	0.2246
Depth continuity	β_{dc}	0.5843
Layout center	β_{op}	0.9678
Layout diagonal	β_{od}	−0.3017
Layout balance	β_{ov}	−1.0460
Parallel patches	β_{pp}	0.0127

1 Tsinghua University, Beijing, China. E-mail: S. Yang, shengyang93fs@gmail.com (✉); K. Chen, chenkangnobel@hotmail.com.

2 Massachusetts Institute of Technology, Cambridge, USA. E-mail: eternal_answer@126.com.

3 City University of Hong Kong, Hong Kong, China. E-mail: fuplus@gmail.com.

where E_{D} is the data term measuring the cost of each patch P_i belonging to v_i. E_{S} is the smooth term measuring the cost of each adjacent pair $(P_i, P_j) \in G_1$ belonging to different labels v_i and v_j. We set $\lambda = 1$ for the entire experimentation.

Data term. We consider the data term with the following features $\mathcal{D}_n$. Let $\mathcal{U}_i$ denote the seeds set of v_i.

1. **Distance**. The same as previous works [S4], we define the distance data term as $\mathcal{D}_{\mathrm{D}} = \exp\left(\frac{D'(P_i,\mathcal{U}_i)}{2\sigma_{\mathrm{D}}^2}\right)$, where $D'(P_i, \mathcal{U}_i)$ is the shortest Euclidian distance between the patch centroid of P_i and the patches in $\mathcal{U}_i$.
2. **Color**. We compare the difference of color distibution to define the color data term as $\mathcal{D}_{\mathrm{C}} = -\log \mathcal{T}_{\mathrm{C}}(P_i, \mathcal{U}_i)$. Here $\mathcal{T}_{\mathrm{C}}$ is defined as $\mathcal{T}_{\mathrm{C}}(P_i, P_j) = \exp\left(\frac{\chi^2(I_i,I_j)}{2\sigma_{\mathrm{C}}^2}\right)$, where $\chi^2(I_i, I_j)$ is the χ-squared distance between the normalized HSV histograms I of the two patches P_i and P_j.
3. **Plane**. We compare the difference of plane fitting to define the plane data term as $\mathcal{D}_{\mathrm{P}} = -\log(\max_{P_j \in \mathcal{U}_i} \mathcal{T}_{\mathrm{P}}(P_i, P_j))$. $\mathcal{T}_{\mathrm{P}}$ is defined as $\mathcal{T}_{\mathrm{P}}(P_i, P_j) = \min\left(n_i \cdot n_j, \ \exp\left(\frac{-\rho(P_i,P_j)}{2\sigma_{\mathrm{P}}^2}\right)\right)$, where $\rho(P_i, P_j)$ is the squared projected distance between their patch centroid on both their normals, assigned as the longer.
4. **Support**. The support data term is computed as $\mathcal{D}_{\mathrm{S}} = -\log\left(\max_{P_j \in \mathcal{U}_i} \mathcal{T}_{\mathrm{S}}(P_i, P_j)\right)$, where $\mathcal{T}_{\mathrm{S}}$ is the support relationship defined in the paper.

The data term E_{D} is then calculated as $E_{\mathrm{D}} = \sum_n \lambda_n \mathcal{D}_n$, with $\lambda_{\mathrm{D}} = 0.15$, $\lambda_{\mathrm{C}} = 0.15$, $\lambda_{\mathrm{P}} = 0.3$, $\lambda_{\mathrm{S}} = 0.4$, $\sigma_{\mathrm{D}} = 1.0$, $\sigma_{\mathrm{C}} = 0.7$, $\sigma_{\mathrm{P}} = 0.1$ for all scenes.

Smooth term. We consider *color* ($\mathcal{T}_{\mathrm{C}}$), *plane* ($\mathcal{T}_{\mathrm{P}}$), and *support* ($\mathcal{T}_{\mathrm{S}}$) costs between adjacency pairs. The smooth cost is computed as $E_{\mathrm{S}} = \sum_n \lambda'_n \mathcal{T}_n$, with $\lambda'_{\mathrm{C}} = 0.3$, $\lambda'_{\mathrm{P}} = 0.4$, $\lambda'_{\mathrm{S}} = 0.3$ for all scenes.

S3 Patch similarity measure

We use the following features (ξ_n) to measure the similarity between a pair of patches (P_i, P_j):

1. **Color histogram**. ξ_{ch} is computed as the χ-squared distance between their normalized color histograms in the HSV colorspace.
2. **Normal**. Let ϑ be the angle between their fitting planes. ξ_{nl} is defined as $\xi_{\mathrm{nl}} = \exp\left(-\frac{\vartheta^2}{2\sigma_\vartheta^2}\right)$.
3. **Vertical location**. This measures the location similarity between the two patches. For a quasi-vertical pair we compare their range in z-axis as ψ, while for a quasi-horizontal pair we consider the vertical distance δ_{h} between their centroid point. ξ_{vl} is then defined as
$$\xi_{\mathrm{vl}} = \begin{cases} \exp\left(-\frac{\delta_{\mathrm{h}}^2}{2\sigma_{\mathrm{h}}^2}\right), & P_i, P_j \in Q_{\mathrm{h}} \\ 1 - \frac{|\delta_{\mathrm{h}}|}{\min(\psi_i, \psi_j)}, & P_i, P_j \in Q_{\mathrm{v}} \end{cases}$$
4. **Area coverage**. Let ζ_i denote the ratio of the areas of the convex hull of patch P_i. ξ_{ad} is defined as $\xi_{\mathrm{ad}} = 1 - (1 - \frac{\min(\zeta_i, \zeta_j)}{\max(\zeta_i, \zeta_j)})^2$.

The final patch similarity is calculated as $\mathcal{A}_{i,j} = \prod_n \xi_n$, with parameters $\sigma_\vartheta = 10°$ and $\sigma_{\mathrm{h}} = 0.15$ for all scenes.

S4 Results of user study for interactions

We compared our method to interactive segmentation of point clouds without view suggestion and interactive segmentation of RGB-D images in a user study. Tables S2–S9 give all the recorded information.

References

[S1] Secord, A.; Lu, J.; Finkelstein, A.; Singh, M.; Nealen, A. Perceptual models of viewpoint preference. *ACM Transactions on Graphics* Vol. 30, No. 5, Article No. 109, 2011.

[S2] Zhang, F. L.; Wang, M.; Hu, S.-M. Aesthetic image enhancement by dependence-aware object recomposition. *IEEE Transactions on Multimedia* Vol. 15, No. 7, 1480–1490, 2013.

[S3] Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 23, No. 11, 1222–1239, 2001.

[S4] Golovinskiy, A.; Funkhouser, T. Min-cut based segmentation of point clouds. In: Proceedings of the IEEE 12th International Conference on Computer Vision Workshops, 39–46, 2009.

Table S2 Record of user study. Scene A with view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	Wheel	GC	Sim.	Ske. Len.	Progress
246.466	10.407	25.852	5	2	1	39	33	18	3	1075	0.903339
187.825	7.763	10.309	4	1	1	43	12	12	4	1184	0.943761
142.897	5.16	6.3	3	0	3	40	12	8	3	1230	0.938489
91.198	4.235	3.681	5	1	1	41	7	7	2	2184	0.906854
94.63	1.623	3.729	3	1	0	39	16	3	1	1152	0.912127
89.56	2.703	1.295	2	0	0	39	0	4	1	1267	0.903339
202.627	7.631	20.378	6	1	1	39	19	13	2	1376	0.910369
176.873	6.843	11.732	4	0	0	41	16	11	3	1482	0.912127
160.625	2.586	8.627	3	1	0	42	20	4	1	1346	0.910369
133.983	4.683	4.319	3	0	0	40	8	9	1	1863	0.920914
118.828	2.045	3.359	4	1	0	41	6	3	1	1740	0.919156
96.895	4.426	2.681	2	0	1	39	5	7	2	1326	0.931459

Table S3 Record of user study. Scene A without view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	Wheel	GC	Sim.	Ske. Len.	Progress
298.695	8.938	44.615	—	—	3	51	46	16	1	2810	0.913884
177.342	6.788	19.284	—	—	0	40	0	11	3	1836	0.917399
167.264	7.341	17.847	—	—	3	41	16	12	3	1360	0.917399
164.503	6.579	8.331	—	—	5	44	20	11	2	2192	0.910369
105.691	0.796	9.641	—	—	0	43	4	5	0	920	0.908612
147.218	3.63	25.959	—	—	0	41	14	6	2	1105	0.903339
245.921	7.35	25.054	—	—	0	47	33	14	1	2435	0.917399
210.986	6.683	12.562	—	—	3	42	9	12	2	1684	0.927944
135.623	5.625	9.973	—	—	1	43	18	11	0	1950	0.919156
114.975	6.201	10.882	—	—	6	42	4	10	3	1245	0.912127
92.361	4.26	8.675	—	—	2	44	6	7	2	1304	0.920914
120.682	5.526	11.223	—	—	4	41	7	9	2	1763	0.917399

Table S4 Record of user study. Scene A by RGB-D

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	Wheel	GC	Sim.	Ske. Len.	Progress
229.291	9.063	—	—	—	—	51	—	17	2	3126	0.920914
204.845	9.244	—	—	—	—	50	—	16	2	4455	0.920914
246.731	6.648	—	—	—	—	44	—	12	3	4537	0.917399
228.698	10.021	—	—	—	—	48	—	19	3	6115	0.913884
161.274	7.746	—	—	—	—	42	—	16	0	1919	0.908612
124.27	5.105	—	—	—	—	42	—	10	0	2196	0.901582
193.564	10.913	—	—	—	—	49	—	20	2	3374	0.903339
213.681	9.719	—	—	—	—	40	—	18	2	2843	0.908612
226.163	8.98	—	—	—	—	38	—	15	4	2869	0.922671
197.616	10.761	—	—	—	—	46	—	21	1	1769	0.920914
189.453	10.176	—	—	—	—	42	—	19	2	4385	0.927944
172.16	8.684	—	—	—	—	43	—	16	1	2486	0.910369

Table S5 Record of user study. Scene B with view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	WheelEvent	GC	Sim.	Ske. Len.	Progress
69.015	1.416	5.897	1	0	0	20	0	3	1	1572	0.937143
180.852	4.443	12.339	10	4	0	21	31	10	1	3036	0.917143
101.51	2.434	7.473	1	0	0	21	13	5	1	1546	0.922857
128.436	4.352	10.048	7	4	1	18	0	9	1	2554	0.917143
185.625	3.496	17.129	1	0	0	22	1	7	1	1035	0.922857
48.08	1.249	0	1	0	0	18	0	2	1	1432	0.9
87.832	1.638	4.619	2	0	0	18	0	3	1	1376	0.908571
135.163	2.863	8.168	3	1	1	23	14	6	0	1941	0.922857
126.301	1.983	10.643	6	0	0	18	0	4	1	2195	0.928571
132.605	2.682	7.354	5	2	0	16	3	5	2	1456	0.914286
114.681	3.061	11.732	4	0	0	19	0	6	1	2327	0.92
78.68	1.213	6.935	1	0	0	18	6	2	1	1467	0.914286

Table S6 Record of user study. Scene B without view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	WheelEvent	GC	Sim.	Ske. Len.	Progress
37.159	0.39	3.904	—	—	0	23	0	1	0	1684	0.908571
207.45	6.162	35.758	—	—	3	25	23	14	1	2238	0.908571
149.792	3.626	31.321	—	—	2	24	45	8	0	2943	0.905714
168.731	5.757	20.234	—	—	0	26	15	13	1	2061	0.914286
155.268	4.791	14.898	—	—	0	27	30	11	0	1382	0.931429
90.137	1.824	17.13	—	—	3	20	23	5	1	2232	0.922857
104.687	0.975	16.198	—	—	0	24	0	2	1	1864	0.902857
137.749	2.354	20.186	—	—	4	23	21	6	1	2347	0.9
143.568	2.578	25.674	—	—	0	24	26	6	2	2490	0.931429
129.381	3.811	18.647	—	—	2	25	14	8	1	1570	0.92
168.577	4.687	23.149	—	—	0	28	17	10	1	2869	0.917143
62.163	4.25	7.684	—	—	1	21	18	9	1	1861	0.905714

Table S7 Record of user study. Scene B by RGB-D

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	WheelEvent	GC	Sim.	Ske. Len.	Progress
88.78	3.529	—	—	—	—	22	0	11	1	2761	0.905714
205.984	5.709	—	—	—	—	27	0	14	1	2048	0.9
140.229	3.621	—	—	—	—	21	0	9	2	2854	0.925714
115.472	3.196	—	—	—	—	23	0	10	1	3052	0.902857
235.219	6.114	—	—	—	—	24	0	15	1	1710	0.92
110.043	3.182	—	—	—	—	21	0	10	1	1631	0.902857
113.199	4.61	—	—	—	—	25	0	12	1	2468	0.928571
168.319	3.125	—	—	—	—	22	0	8	2	2946	0.917143
101.683	5.978	—	—	—	—	23	0	14	1	1686	0.9
196.263	6.368	—	—	—	—	26	0	16	1	1726	0.928571
146.168	5.354	—	—	—	—	28	0	13	1	2237	0.902857
96.165	4.577	—	—	—	—	21	0	12	2	1680	0.905714

Table S8 Record of user study. Scene C with view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	WheelEvent	GC	Sim.	Ske. Len.	Progress
56.192	0.843	0	1	0	1	23	0	2	0	1730	0.906367
63.336	1.389	0.827	4	1	0	22	0	4	0	2319	0.925094
78.048	2.543	0	2	0	1	24	0	6	0	2353	0.902622
112.508	2.395	1.794	2	0	0	26	0	5	0	2005	0.906367
77.829	1.267	0.858	1	0	0	23	0	2	0	1611	0.910112
78.281	2.512	0	4	0	0	24	0	5	0	2811	0.902622
82.961	1.068	0.624	2	0	0	25	1	3	0	1495	0.910112
90.616	1.683	0	3	0	0	27	0	4	0	1876	0.906367
115.639	2.437	0	4	1	0	28	0	5	0	2062	0.932584
146.316	2.768	2.453	3	1	0	30	0	6	0	2471	0.902622
120.684	1.616	0	2	0	1	24	0	4	0	1896	0.910112
106.637	1.275	0.438	3	0	1	23	0	3	0	1945	0.921348

Table S9 Record of user study. Scene C without view suggestion

TotalTime	ProcTime	NaviTime	Sug.	Rej.	Minimap	Ske.	WheelEvent	GC	Sim.	Ske. Len.	Progress
88.812	1.841	7.925	—	—	4	25	17	5	0	1457	0.909367
119.793	2.949	5.008	—	—	0	25	11	6	0	1411	0.902622
175.095	4.353	43.12	—	—	0	25	30	10	0	3477	0.93633
213.066	3.847	29.566	—	—	0	29	37	8	0	2389	0.925094
198.184	4.09	18.499	—	—	0	24	29	9	0	1383	0.910112
91.775	2.355	7.395	—	—	4	25	11	7	0	2796	0.902622
117.169	2.436	12.683	—	—	3	25	18	7	0	1862	0.925094
134.647	3.438	14.168	—	—	0	26	13	8	0	2292	0.902622
203.197	3.84	35.744	—	—	2	30	32	9	0	2761	0.925094
244.005	4.689	47.556	—	—	0	27	40	11	0	2904	0.902622
226.126	4.292	24.16	—	—	2	24	27	10	0	1753	0.962547
134.068	2.894	18.737	—	—	1	28	16	6	0	1699	0.951311

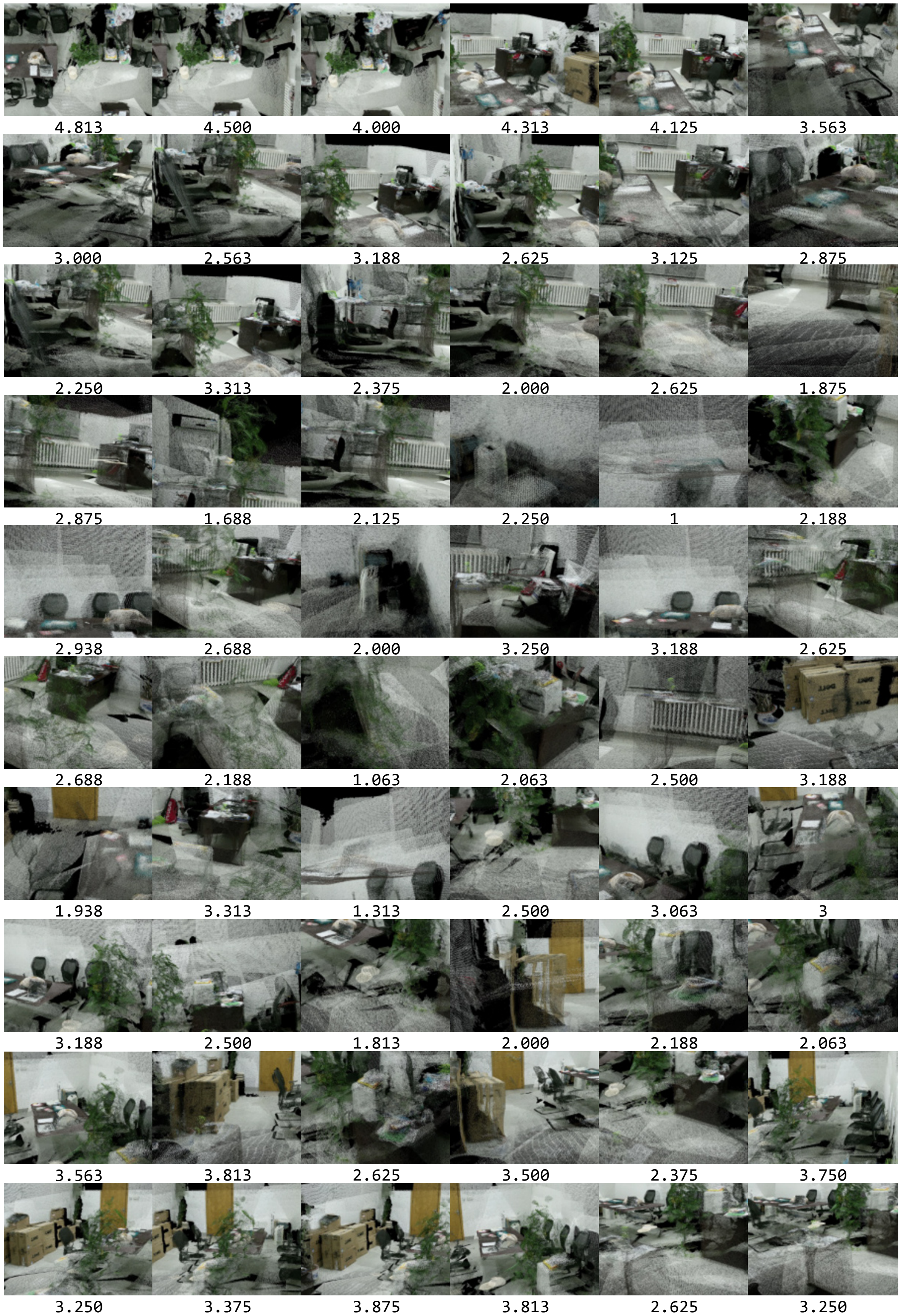

Fig. S1 Average score for every selected view of Scene 1 over the participants.

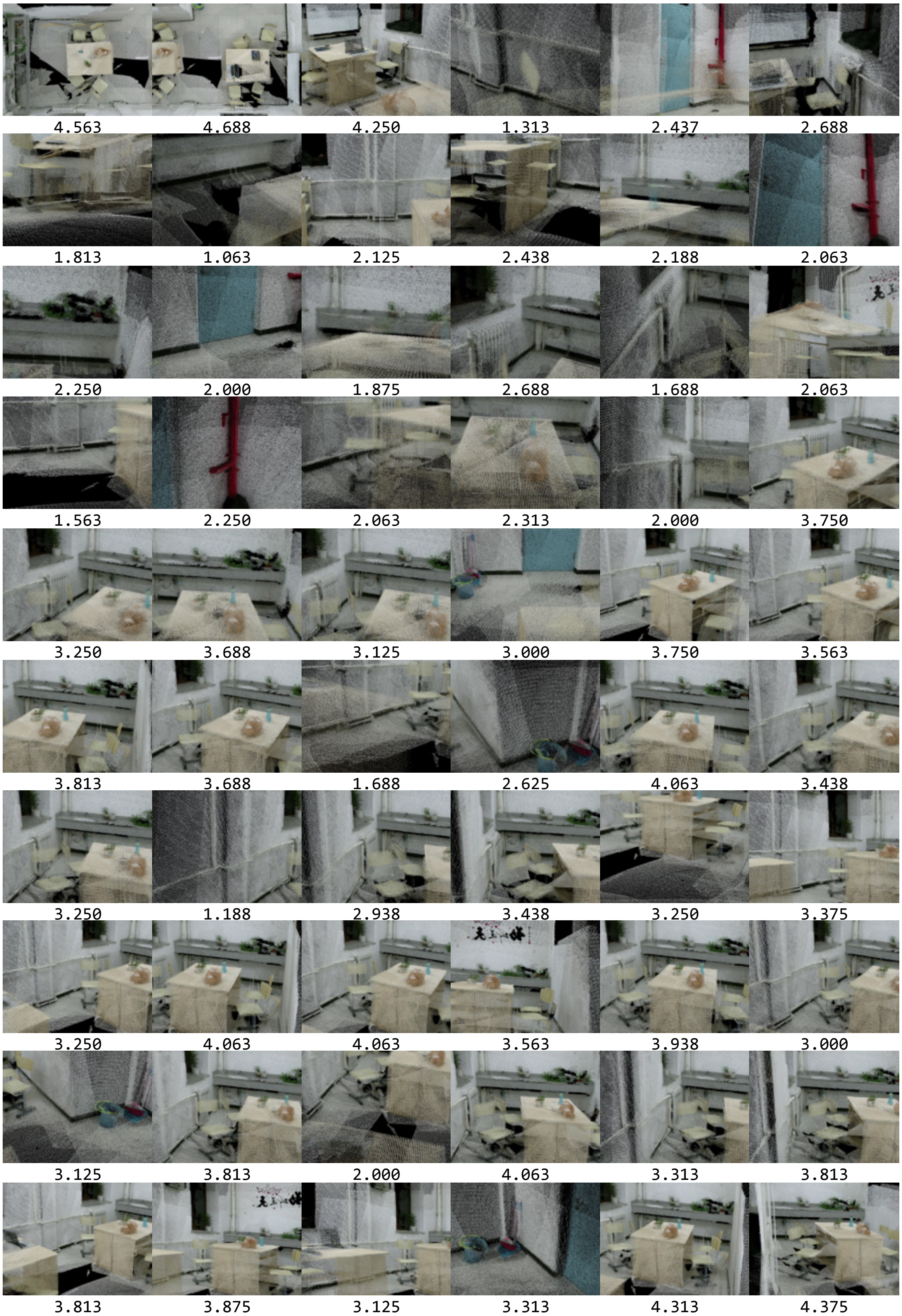

Fig. S2 Average score for every selected view of Scene 3 over the participants.*