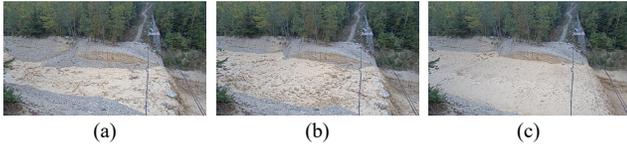


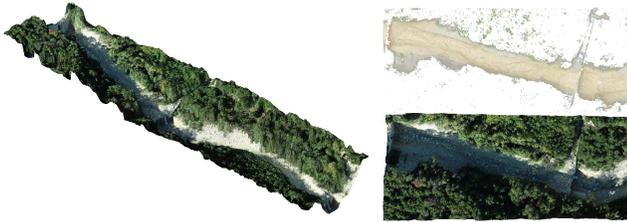
## 1080 A. Appendix

### 1081 A.1. Dataset

1082 In this section, we provide supplementary information  
1083 and visualizations of the debris flow dataset. The repre-  
1084 sentative shapes of the debris-flow surface are summarized  
1085 in Fig. A1. The terrain of the monitoring site is shown in  
1086 Fig. A2.



1087 Figure A1. Three stages of debris flow: (a) pre-event, (b) arrival of  
1088 boulder front and (c) fine-grained slurry fluid.



1089 Figure A2. **3D reconstruction of the debris-flow monitoring**  
1090 **site [11].** Overview of the scene on the left. Reconstruction of the  
1091 channel before the event on the upper right, and after the event on  
1092 the bottom right.

1093 In the debris flow dataset, we release:

- 1094 • 6000 high-resolution images (1920×1080)
- 1095 • 6000 high-accuracy point clouds
- 1096 • Camera calibration matrices
- 1097 • Rotation matrices between camera and LiDAR
- 1098 • Translation vectors between camera and LiDAR

### 1099 A.2. Evaluation Metrics

1100 **SSIM.** Structural similarity index [50] is used in our optical  
1101 flow loss to assess the similarity between the target image  
1102 and warped image:

$$1103 \text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (15)$$

1104 where  $\mu_x$  and  $\mu_y$  represent the mean pixel values of images  
1105  $x$  and  $y$ , and  $\sigma$  denotes the corresponding standard devia-  
1106 tion.

1107 **Census Transform Loss.** We use ternary census transform  
1108 loss [17,33,42] as the second metric to evaluate optical flow

performance:

$$1134 \text{CT}(p, p') = \begin{cases} -1 & \text{if } p' - p \geq \epsilon \\ +1 & \text{if } p - p' \geq \epsilon \\ 0 & \text{if } |p - p'| < \epsilon \end{cases} \quad (16)$$

1135 Given two input images, we compute the corresponding  
1136 census-transformed images and compute the average differ-  
1137 ence between them as the loss.

### 1138 A.3. LiDAR to Range Image

1139 Since point cloud-based networks [37,38,54] are com-  
1140 putationally demanding and complex to train, we convert  
1141 the 3D scan points to sparse range maps with the help of  
1142 the camera-LiDAR transformation by projecting 3D points  
1143 onto the image plane with known camera intrinsics  $\mathbf{K}$  and  
1144 camera pose parameters  $\mathbf{R}$  and  $\mathbf{t}$ :

$$1145 \mathbf{K}[\mathbf{R} | \mathbf{t}] = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{bmatrix}. \quad (17)$$

1146 The 2D projection  $\mathbf{p} = (u, v, 1)^T$  of 3D point  $\mathbf{P} =$   
1147  $(x, y, z, 1)^T$  is computed as

$$1148 \mathbf{p} = \mathbf{K}[\mathbf{R} | \mathbf{t}]\mathbf{P} \quad (18)$$

1149 and rounded to the closest integer pixel coordinate.

### 1150 A.4. Runtime and Model Size

1151 We report the runtimes and model sizes of RAFT,  
1152 DeFlow-Cam, and DeFlow-Fusion in Tab. A1. Our camera-  
1153 only baseline is significantly smaller and faster, since we fol-  
1154 low the lightweight design of PWC-Net [43] and Mono-  
1155 SF [21]. Our fusion model has similar model size and  
1156 runtime as RAFT. The increase in runtime and model size  
1157 compared to the camera-only baseline is caused by the ad-  
1158 ditional depth encoder and the multi-level feature fusion,  
1159 which in return offer marked performance gains.

Method	Runtime [ms]	# params
RAFT [45]	171.3	5.26 M
DeFlow-Cam ( <i>Ours</i> )	54.1	4.16 M
DeFlow-Fusion ( <i>Ours</i> )	191.9	4.99 M

1160 Table A1. Runtime and number of parameters for different mod-  
1161 els.

### 1162 A.5. Additional Qualitative Results

1163 In Fig. A3, we present additional visualizations of opti-  
1164 cal flow results, and of the static pixel masks used to en-  
1165 force a static sensor pose. To see the qualitative behaviour  
1166 of our temporal smoothing module, *readers are encouraged*  
1167 *to watch the video in the supplementary material.*

1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241

1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295

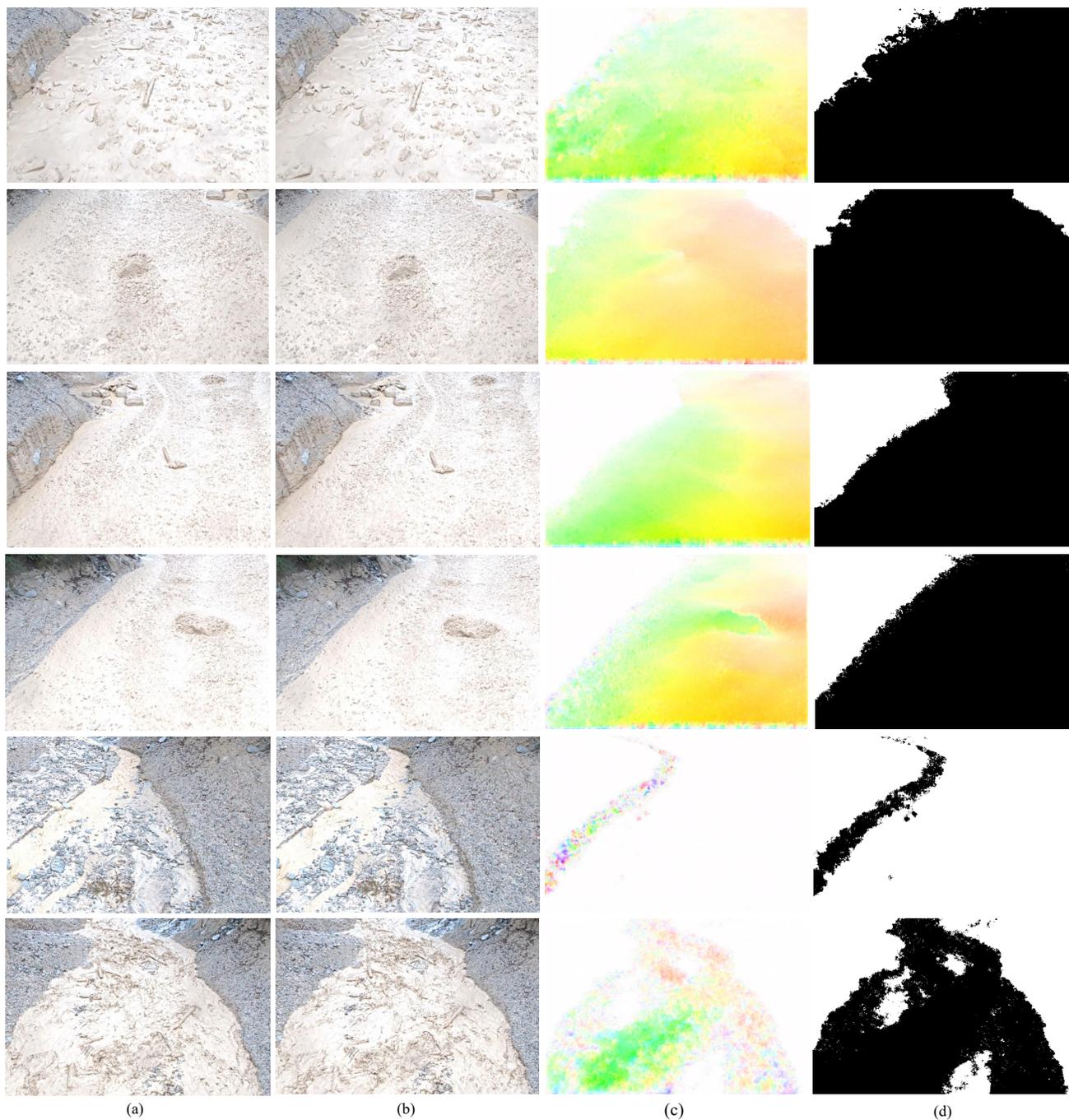


Figure A3. Qualitative results of optical flow (c) and binary (static/dynamic) segmentation (d).