Generating Humanless Environment Walkthroughs from Egocentric Walking Tour Videos

Supplementary Material

Figure 9. **Full baseline comparison for the scenes in Figure 1.** Red boxes indicate failures in foreground removal or shadow handling, while yellow boxes highlight blurry inpainting artifacts. ProPainter [48] and DiffuEraser [18] struggle with sharp cast shadows and often blur or lose details in large masked regions. Casper achieves more reliable effect association but exhibits noticeable hallucinations when masks become large. In contrast, *CrowdEraser* remains robust under significant mask sizes, preserving background structure and producing more visually plausible results.

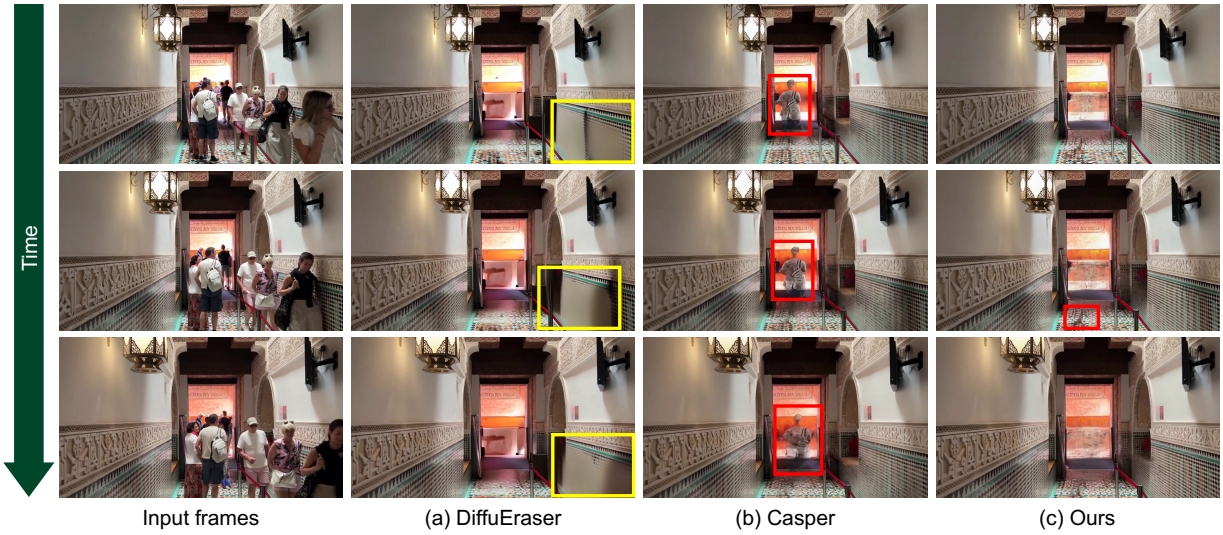

Figure 10. **Baseline comparison across temporal frames for the "Marrakech" scene in Figure 5.** Red boxes indicate failures in foreground removal or shadow handling, while yellow boxes mark regions where the background is over-smoothed instead of plausibly inpainted. Casper [24] struggles with larger masks, producing noticeable hallucinations within masked areas, whereas DiffuEraser [18] tends to over-smooth patterns in occluded regions.

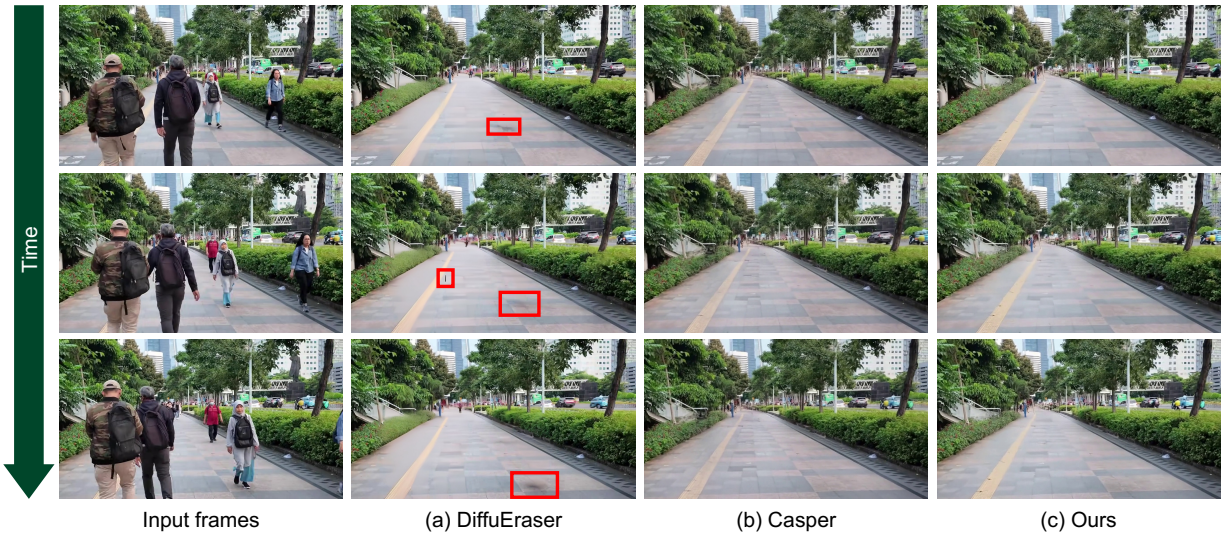

Figure 11. **Baseline comparison across temporal frames for the 'Jakarta' scene in Figure 5.** Red boxes highlight failures in foreground removal or shadow handling. DiffuEraser [18] struggles to capture shadows, leading to floating shadows on the pathway.

We provide per-scene quantitative results in Table 3. Our *CrowdEraser* consistently achieves the best DreamSim scores, reflecting better perceptual quality aligned with human judgment. While ProPainter attains higher PSNR in some cases, this mainly stems from its strict adherence to the mask rather than improved inpainting, as PSNR is affected by unmasked, and thus unchanged, pixels. Notably, our method achieves higher in-mask PSNR, indicating more accurate inpainting within the masked regions.

We present additional qualitative examples, including comprehensive baseline comparisons in Figure 9, temporal dynamics, and comparisons with two recent methods in Figures 10–13, as well as extended 4D reconstruction results in Figures 14–15.

For reproducibility, we provide a comprehensive list of videos used to construct our dataset in Tables 4–6.

Table 3. **Quantitative comparison across cities.** Best results are highlighted in red and second-best in yellow. Our *CrowdEraser* consistently achieves the best DreamSim, reflecting superior perceptual quality.

Scene	Birmingham		Boston		Capetown		Chicago		Dubai		Rome		Zurich	
	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓	PSNR↑	DreamSim↓
ProPainter [48]	**27.77**	0.052	**26.00**	0.049	**25.86**	0.059	25.91	0.053	22.34	0.050	29.97	0.019	25.76	0.059
DiffuEraser [18]	27.25	0.034	25.49	0.042	25.43	0.050	25.59	0.036	22.65	0.042	29.86	0.013	25.14	0.049
Casper [17]	26.15	0.028	25.12	0.031	24.20	0.035	26.02	0.022	23.14	0.027	29.44	0.012	25.21	0.027
Ours	26.58	**0.022**	**26.00**	**0.025**	25.50	**0.026**	**27.08**	**0.019**	**24.81**	**0.019**	**30.31**	**0.009**	**25.99**	**0.022**

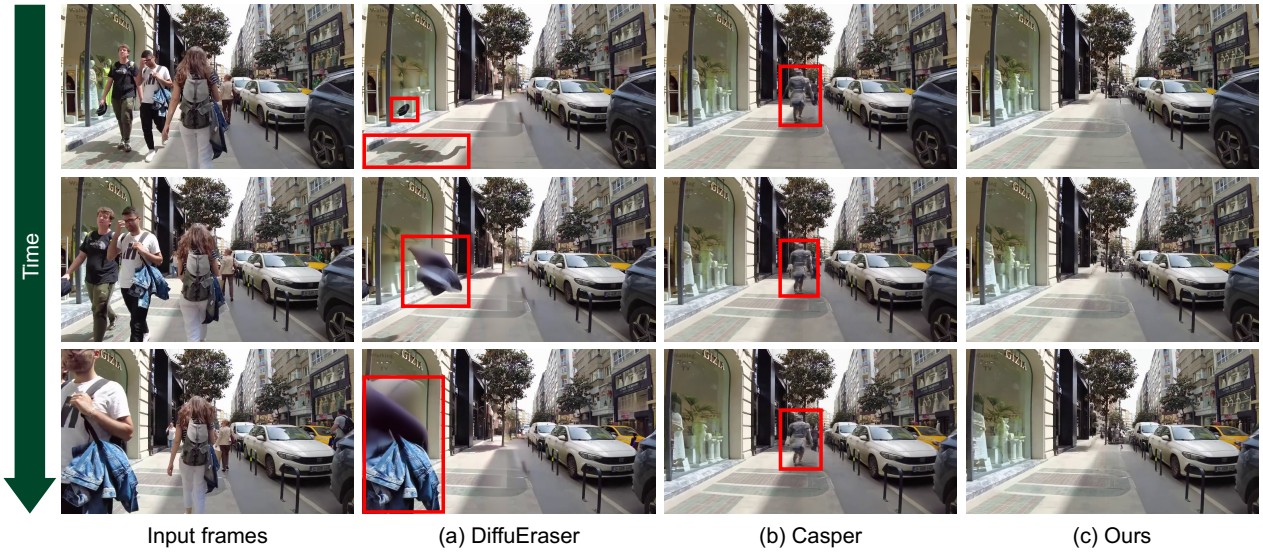

Figure 12. **Baseline comparison across temporal frames for the "Istanbul" scene in Figure 5.** Red boxes indicate failures in foreground removal or shadow handling. Casper [24] struggles with larger masks, producing noticeable hallucinations within masked areas, while DiffuEraser [18] has difficulty handling shadows and removing associated objects.

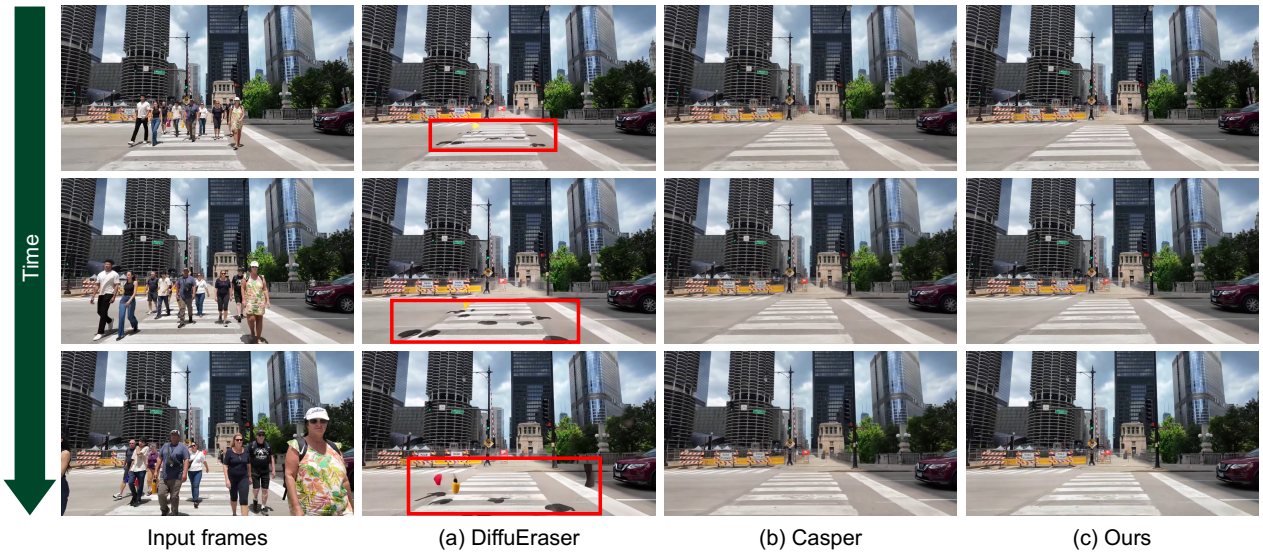

Figure 13. **Baseline comparison across temporal frames for the "Chicago" scene in Figure 5.** Red boxes highlight failures in foreground removal or shadow handling. DiffuEraser [18] struggles to associate shadows and obejects, leading to floating shadows and objects.

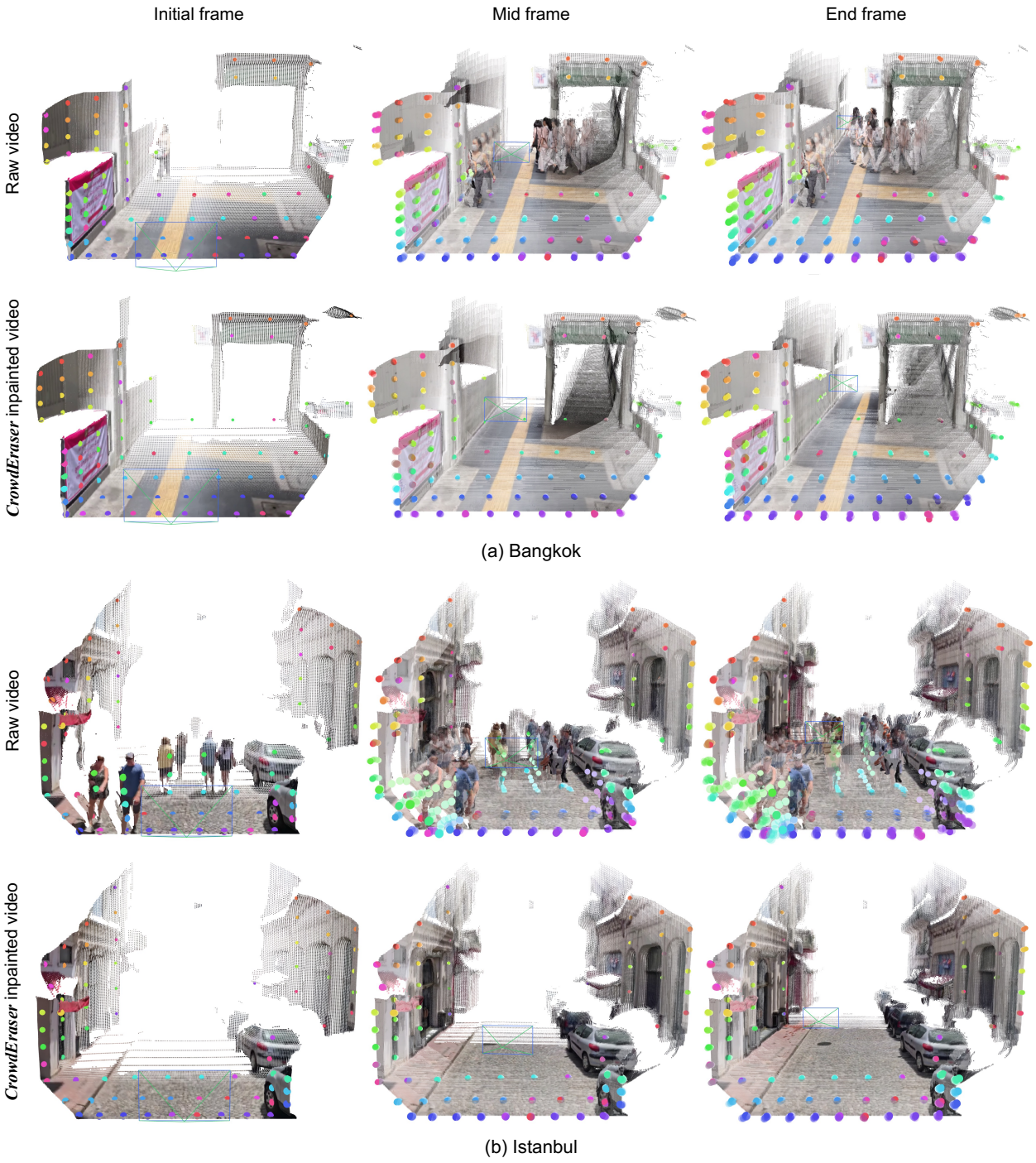

Figure 14. **SpatialTrackerV2 4D reconstruction results.** We compare results using raw walking tour video inputs (top) versus our crowd-removed versions (bottom). Each image displays the inferred 3D point clouds for the scene visualized from the camera viewpoint of the initial, middle, and final video frames, with overlaid colored circles corresponding to point tracks in the 3D space. Tracking points remain more stable in static background regions, indicating that our crowd removal leads to more reliable and robust reconstruction. Moreover, the resulting point clouds are denser and more consistent, benefiting downstream tasks such as scene modeling and 3D novel view synthesis.

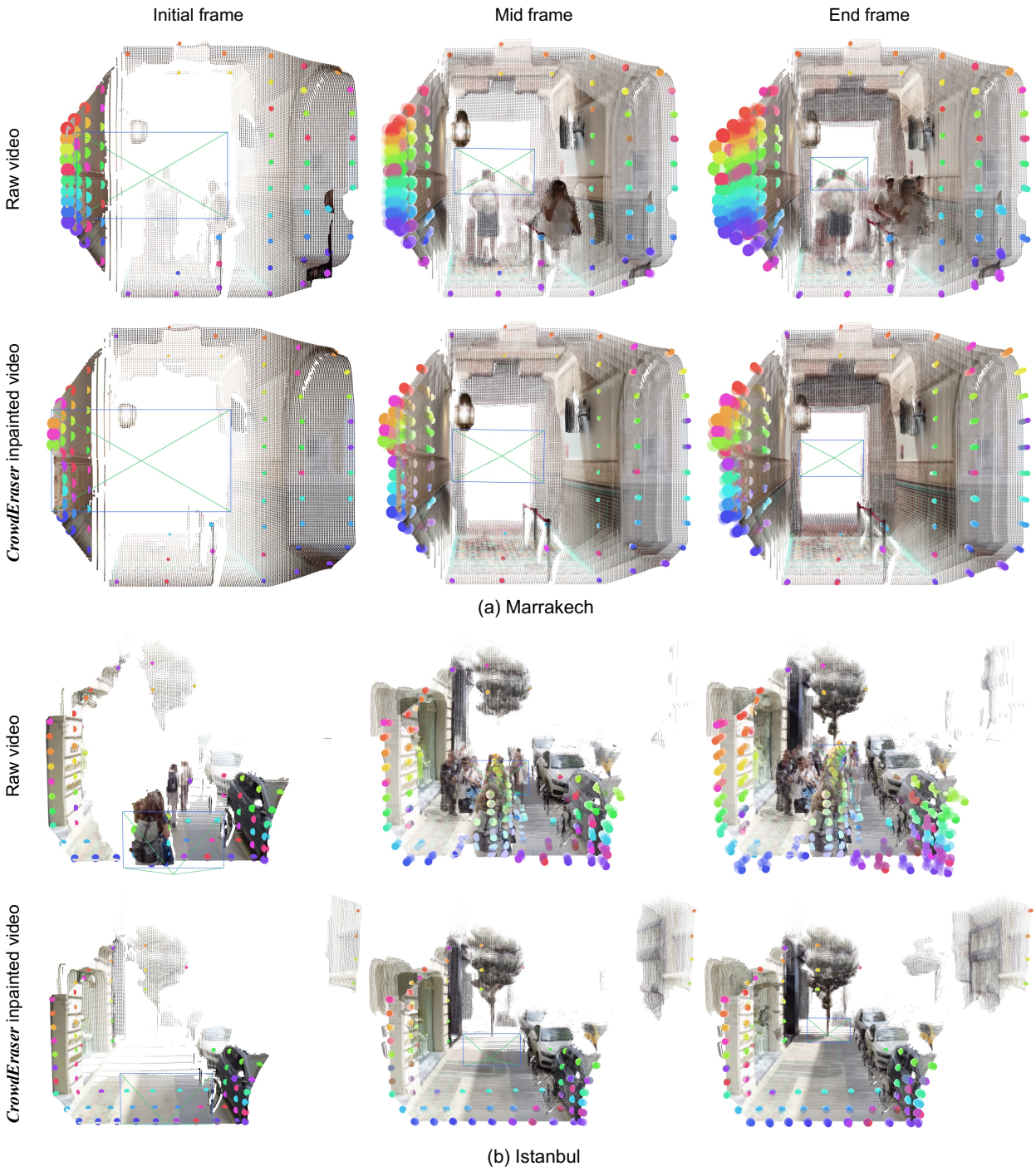

Figure 15. **SpatialTrackerV2 4D reconstruction results.** We compare results using raw walking tour video inputs (top) versus our crowd-removed versions (bottom). Each image displays the inferred 3D point clouds for the scene visualized from the camera viewpoint of the initial, middle, and final video frames, with overlaid colored circles corresponding to point tracks in the 3D space. Tracking points remain more stable in static background regions, indicating that our crowd removal leads to more reliable and robust reconstruction. Moreover, the resulting point clouds are denser and more consistent, benefiting downstream tasks such as scene modeling and 3D novel view synthesis.

Continent	Country	City / Area	URL(s)
Train Background Video Sources			
Africa	Egypt	Cairo	https://youtu.be/TVe7Th_EfrM
Asia	China	Beijing	https://youtu.be/FNK5UEObcEg, https://youtu.be/MU-obosH1ow
Asia	China	Great Wall	https://youtu.be/cVmM6sUcdwg
Asia	China	Shanghai	https://youtu.be/Z1i0v6wsGLI, https://youtu.be/DUANyWsER_o
Asia	South Korea	Cheongju	https://youtu.be/kqkUJZllt0U
Asia	South Korea	Daejeon	https://youtu.be/n1V69LjdNQg, https://youtu.be/mGLh9Ss_OEo
Asia	South Korea	Seoul	https://youtu.be/LJaIGjruqtE
Europe	Austria	Vienna	https://youtu.be/LKNyOXwooKo
Europe	France	Paris	https://youtu.be/HJgflqZvTl0
Europe	Germany	Berlin	https://youtu.be/qgNKZBQW0hA
Europe	Germany	Würzburg	https://youtu.be/hFzqKwrRdi8
Europe	Italy	Pompeii	https://youtu.be/9L1jrC2-BTE
Europe	Italy	Positano	https://youtu.be/UgYMsj4dDfE
Europe	Spain	Majorca	https://youtu.be/ufPda6XaA7E
Europe	Sweden	Stockholm	https://youtu.be/HJgflqZvTl0
Europe	UK	London	https://youtu.be/VkFpqAG6mm8
North America	Canada	Guelph, ON	https://youtu.be/mLk9-S6hpsU
North America	Canada	North Vancouver, BC	https://youtu.be/IAhJvjeM9SI
North America	Canada	Surrey, BC	https://youtu.be/mjr1r190-VQ
North America	Canada	Toronto, ON	https://youtu.be/5I-WYTYLD0o
North America	USA	Atlanta, GA	https://youtu.be/mseo6t1hiYs
North America	USA	Austin, TX	https://youtu.be/7cVQAs-c2Lg
North America	USA	Boston, MA	https://youtu.be/6ZwGo49Dce8, https://youtu.be/2MHqXI-j-zY
North America	USA	Cambridge, MA	https://youtu.be/uHRHlba3CyQ
North America	USA	Charleston, SC	https://youtu.be/JdPkO2iIvfg
North America	USA	Hayward, CA	https://youtu.be/9EDA2IHtJFM
North America	USA	Honolulu, HI	https://youtu.be/JTQdSKz9wEc
North America	USA	Houston, TX	https://youtu.be/t2ojP7lrfXw
North America	USA	Las Vegas, NV	https://youtu.be/GH25Pzv0WNo
North America	USA	Los Angeles, CA	https://youtu.be/kiyUR7xPkAM, https://youtu.be/EM1XQfC1Vdw
North America	USA	Miami, FL	https://youtu.be/ruXuOM1PAJY
North America	USA	New Haven, CT	https://youtu.be/r0_sbCxgP58
North America	USA	New York, NY	https://youtu.be/3koOEPntvqk, https://youtu.be/2UXhhyNYpLc https://youtu.be/MheS3NBAZJ0, https://youtu.be/YbiCtAdiS6U https://youtu.be/fYY7uEgPw1c
North America	USA	Pine Bluff, AR	https://youtu.be/3FDjNp77wGo
North America	USA	Portland, OR	https://youtu.be/TkZU-yfUqe8, https://youtu.be/KiZ36s2IUi0
North America	USA	Provo, UT	https://youtu.be/653tnKwzNdg
North America	USA	Sacramento, CA	https://youtu.be/W5XSfxIZdMg
North America	USA	San Diego, CA	https://youtu.be/m13-S2HEl6E
North America	USA	San Francisco, CA	https://youtu.be/SX-2d1VyTUw
North America	USA	San Jose, CA	https://youtu.be/DNnP60oi-mc, https://youtu.be/Kc_NWFQrzpo
North America	USA	State College, PA	https://youtu.be/R81NaRZISTU
North America	USA	Syracuse, NY	https://youtu.be/FIg579AzSUg
North America	USA	Washington, DC	https://youtu.be/secTBj63dcI
North America	USA	Wellesley, MA	https://youtu.be/dCxAuWK5gLw
North America	USA	North Dakota	https://youtu.be/mr02QEJooOQ
Train Foreground Video Sources			
Asia	India	Varanasi	https://youtu.be/Odh_7dQwzYQ
Asia	South Korea	Seoul	https://youtu.be/D-F4L5Gfhik, https://youtu.be/DF8KDaUn1TA https://youtu.be/KisjSKv53FA
Europe	Germany	Hamburg	https://youtu.be/aqgRc-sne8g
Europe	Netherlands	Amsterdam	https://youtu.be/7Ttc3AaPNZs
North America	USA	Anaheim, CA	https://youtu.be/Eo8q61Xtc50
North America	USA	Honolulu, HI	https://youtu.be/eSSrUot4yhQ
North America	USA	New York, NY	https://youtu.be/bCoqUaLHjy0, https://youtu.be/C_nK_-ZI6Zo

Table 4. **Training dataset sources.** Background and foreground video sources used to construct *EgoCrowds*. For cities where a single video did not provide sufficient clips, multiple videos were collected to ensure sufficient coverage.

Continent	Country	City / Area	URL(s)
Test Data Background			
Africa	South Africa	Cape Town	https://www.youtube.com/watch?v=eG_SV5aSBqQ
Asia	United Arab Emirates	Dubai	https://www.youtube.com/watch?v=mElSLruob6c
Europe	Italy	Rome	https://www.youtube.com/watch?v=xpRDEoEQpwk
Europe	Switzerland	Zurich	https://www.youtube.com/watch?v=UcRW2OHqC2o
Europe	UK	Birmingham	https://www.youtube.com/watch?v=IlzFH4yE2Z8
North America	USA	Boston, MA	https://www.youtube.com/watch?v=8NVjJsjFLEA
North America	USA	Chicago, IL	https://www.youtube.com/watch?v=R9VGInHbKik
Test Data Foreground			
Asia	China	Hong Kong	https://www.youtube.com/watch?v=JRvQ_pM87ik
Asia	Switzerland	Zurich	https://www.youtube.com/watch?v=65KsVRG1ao8
Europe	Austria	Vienna	https://www.youtube.com/watch?v=TCRD9Dz6k88
North America	USA	Houston, TX	https://www.youtube.com/watch?v=r6cLF5s2B_g
North America	USA	Los Angeles, CA	https://www.youtube.com/watch?v=Oifxr_fLfNE

Table 5. **Quantitative test dataset sources.** Background and foreground video sources used in the quantitative evaluation dataset.

Continent	Country	City	URL(s)
Crowd Walking Tour Video Sources			
Asia	India	Mumbai	https://youtu.be/_2GM4gVlors
Asia	Japan	Tokyo	https://youtu.be/jeQd-n7Rot0
Asia	Japan	Kyoto	https://youtu.be/OhOlwqjt_Lg
Asia	Thailand	Bangkok	https://youtu.be/sWRoDRYi1Lk
Asia	Indonesia	Jakarta	https://youtu.be/2lSUV5KZgwI
Africa	South Africa	Cape Town	https://youtu.be/pL-5CjB0hf8
Africa	Morocco	Marrakech	https://youtu.be/OvN1numZqqU
Africa	Nigeria	Lagos	https://youtu.be/LZJ000F-CLc
North America	USA	New York City	https://youtu.be/o012OW9vej8, https://youtu.be/77EXFlRLbiM
North America	USA	Denver	https://youtu.be/L3Uz0O1pO3k
North America	USA	Chicago	https://youtu.be/75OQ94gCOeI
North America	Mexico	Cancun	https://youtu.be/3mU1CbBTIJ8, https://youtu.be/RlcMESpoHs8
South America	Brazil	Rio de Janeiro	https://youtu.be/RYxqpz5XS0A
South America	Argentina	Buenos Aires	https://youtu.be/Hug4u_7ZYxE, https://youtu.be/QVyueY43tA8
Europe	Croatia	Dubrovnik	https://youtu.be/_93zEDEYBf0
Europe	Sweden	Stockholm	https://youtu.be/FuFe9WC3rjg
Europe	Spain	Seville	https://youtu.be/m4AmnRnWcRk, https://youtu.be/1o_V2qXNyUM
Europe/Asia	Turkey	Istanbul	https://youtu.be/ZOGEYkHyNWU, https://youtu.be/Hp3gETuZa8o

Table 6. **Qualitative test dataset sources.** Real walking tour video sources used for qualitative evaluation. For cities where a single video did not provide sufficient clips, multiple videos were collected to ensure sufficient coverage.