Supplementary material

Into the Fog: Evaluating Robustness of Multiple Object Tracking

Nadezda Kirillova
nadezda.kirillova@tugraz.at
M. Jehanzeb Mirza
mirza@tugraz.at
Horst Bischof
horst.bischof@tugraz.at
Horst Possegger
possegger@tugraz.at

Institute of Computer Graphics and Vision
Graz University of Technology
Graz, Austria

In this supplementary material, we provide further details on volumetric homogeneous and heterogeneous fog simulation in real-world arbitrary video datasets, along with additional visualizations and an advanced analysis of MOT performance. Furthermore, we present detailed results of fog appearance validation, including a description of our user study.

1 Metric Depth and Meteorology Visibility

The MiDaS [5] monocular depth estimation achieves strong results in single-image *relative inverse depth* $\mathbf{d}(\mathbf{x})$ estimation without providing metric values, whereas *scale* s and *shift* t factors remain unknown. If 3D ground truth reference points of the captured scene are available, such as minimum D_{min} and maximum D_{max} distances, the actual *metric depth* $\mathbf{D}(\mathbf{x})$ in $[m]$ can be calculated from:

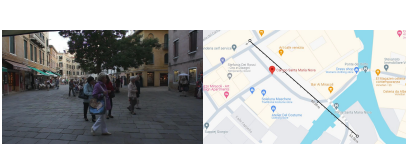
$$\mathbf{D}(\mathbf{x}) = \frac{1}{s\,\mathbf{d}(\mathbf{x}) + t}, \qquad where \quad s = \frac{1}{D_{min}} - \frac{1}{D_{max}}, \quad t = \frac{1}{D_{max}}. \tag{1}$$

For example, in the MOT17-02 sequence, captured at the Venetian square Campo Santa Maria Nova, we have the advantage of accessing absolute distances via publicy available map providers (such as Google Maps, see Fig.1(a)) and obtain metric depth information.

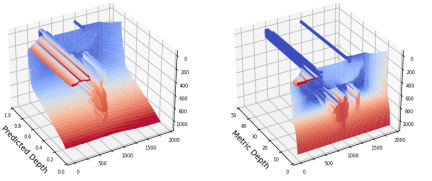
In scenarios with sky areas, the maximum distance is considered infinite, and we approximate it as $D_{max} = 10^6\ m$. Fig.1(b) illustrates the scaling difference between predicted and metric depths.

Metric depth map improve the accuracy of volumetric fog rendering by aligning its severity with meteorological *visibility* V measured in meters. The *attenuation coefficient* β, which regulates fog intensity by rendering, is inversely proportional to the visibility. Thus, for a visibility less than 1 km ($V = 1000\,m$), we obtain an attenuation coefficient of

$$\beta = \frac{-ln(0.05)}{V} \approx \frac{2.9957}{1000} \approx 0.003\ [m^{-1}]. \tag{2}$$



(a) Reference point estimation (minimum and maximum scene distances) using Google Maps. (Better seen zoomed in on screen.)



(b) Differences in scaling between predicted (left) and metric (right) depth maps. In the metric map, distant sky pixels are clearly visible, while in the predicted map, scene disparity is evident.

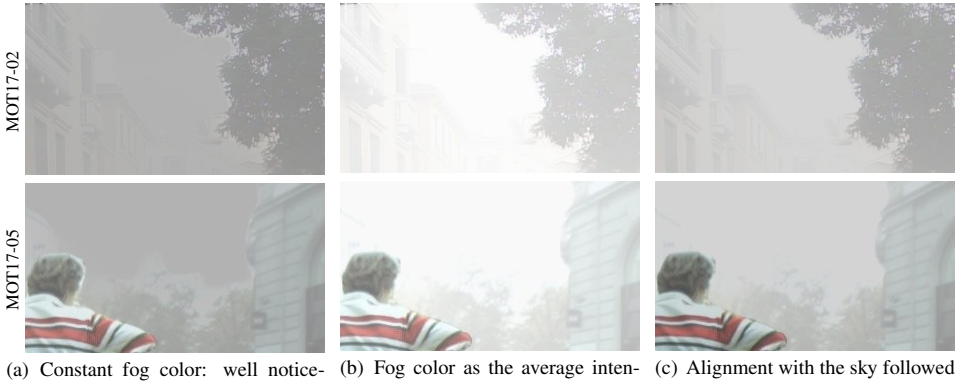
Figure 1: An example of metric depth map estimation for one frame of MOT17-02 sequence.

For fog (outdoors) and smoke (indoors) simulations, we apply different visibility ranges. We employ β values of $[0.03,\ 0.06,\ 0.15,\ 0.3]$ for outdoor scenes with visibility less than $100\,m$, $50\,m$, $20\,m$ and $10\,m$ respectively, and β values of $[0.15,\ 0.3,\ 0.6,\ 1]$ for indoors with visibility less than $20\,m$, $10\,m$, $5\,m$ and $3\,m$.

2 Atmospheric Light and Fog Color

We compared the appearance of simulated fog based on different estimation methods of atmospheric light at the horizon L_∞ (see Fig. 2 and Fig. 3).

In sky-visible scenarios, defining the fog color as the average intensity of sky pixels, we additionally reduce the final foggy image brightness by 20% to mitigate the unnatural whitish fog appearance caused by the high brightness of original sunny scenes. This adjustment better reflects typical fog conditions, where the sky is constantly overcast, resulting in lower atmospheric light (see Fig. 2). Our user study (see Sec. 4) confirms that such brightness reduction enhances the realism of fog simulation.

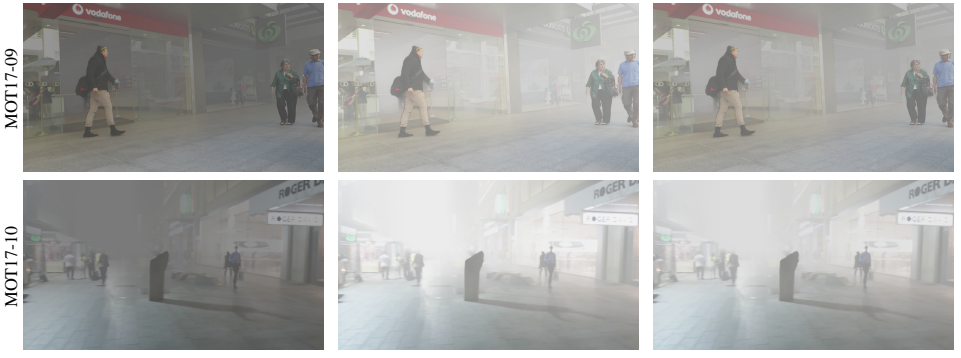


(a) Constant fog color: well noticeable artifacts in form of unrealistic sky area borders due to a lack of alignments with the sky color.

(b) Fog color as the average intensity of sky pixels: no noticeable artifacts, but unnaturally bright whitish fog appearance due to intensity estimation in clear conditions.

(c) Alignment with the sky followed by brightness reduction (Ours): the most realistic simulation, reminiscent of fog often observed after rain or during the early morning hours, when the sky is overcast and atmospheric light is low.

Figure 2: Comparison of fog appearance from different simulation methods in daytime outdoor scenarios with visible sky (Best viewed zoomed in on screen.)



(a) Fog color derived from the average intensity of all image pixels [3] leads to the formation of unrealistically dark image regions.

(b) Fog color computed from the top ten percent of the brightest image pixels [2] appears unrealistically whitish for indoor or night scenes.

(c) Fog color estimated using DCP (Ours) produces the most realistic fog appearance with stable results across arbitrary images.

Figure 3: Comparison of fog appearance from different simulation methods in indoor (top row) and night (bottom row) scenarios. (Best viewed on screen.)

In indoor scenes or surveillance scenarios with downward oriented cameras, where the sky is not visible, experiments show that our method, based on the dark channel prior (DCP) [2], yields better results compared to other approaches (see Fig. 3).

3 Homogeneous vs Heterogeneous Fog Effects

Applying the fog formation optical model and turbulence noise, we render both homogeneous and heterogeneous fog effects to achieve more diversity in representing the intricate nature. Firstly, we generate turbulence texture $\tau(\mathbf{x})$ with the shape of 640×640 by summing up 5 levels (octaves) of pseudo-random gradient perlin noise $\mathbf{P_n}(\mathbf{x})$. We apply 4 iterations (periods) of noise generation along each axis for one octave, and a scaling factor (persistence) of 0.5 with a frequency factor (lacunarity) of 2 between two octaves. Then, we interpolate the texture $\tau(\mathbf{x})$ to match the image size, resulting in the turbulence map depicted in Fig. 4(b). To ensure consistency, we use the same turbulence map throughout the entire frame sequence to prevent fog flickering in the video.

Depending on factors such as camera position, field of view, focal length and light, the visibility of the heterogeneous effect in photography may vary. We found that reducing the brightness of the turbulence map by 20% or even by 50% yields a more photorealistic fog (see Fig. 4(a)). These findings are consistent with our user study (see Sec. 4).

(a) Left to right: homogeneous fog, heterogeneous fog with 50% turbulence brightness reduction, heterogeneous fog with 20% turbulence brightness reduction. (Best viewed zoomed in on screen.)

(b) Turbulence map generated from 5-octave perlin noise

Figure 4: Heterogeneous fog formation and comparison of its appearances demonstrated on a surveillance scene.

4 Fog Simulation Validation

4.1 Qualitative Evaluation

Fig. 5 presents our results of fog simulation with increasing intensity levels across multi-domain scenarios, starting with clear weather. Our approach effectively captures the complexity of natural fog, gradually increasing its density with distance. The method demonstrates its ability to reproduce this intricate natural phenomenon, achieving photorealistic results and accurately simulating real-world conditions. The simulations are particularly notable for their consistency across various environments, from urban landscapes with different camera positions and lighting conditions to indoor settings, highlighting the robustness and versatility of our technique.

Figure 5: Increasing fog simulation (from left to right) in multi-domain scenarios starting with clear weather. From top to bottom: outdoor daytime static frontal view (MOT17-02); nighttime surveillance view of a heavily illuminated, crowded square (MOT17-04); outdoor daytime view from a moving camera (MOT17-05); semi-indoor view of a covered pedestrian street, filmed by a static camera from a low angle position (MOT17-09); outdoor night view from a moving camera (MOT17-10); scene captured by a forward-moving camera in shopping mall (MOT17-11); daytime road view from a shaking, moving double-decker bus (MOT17-13).

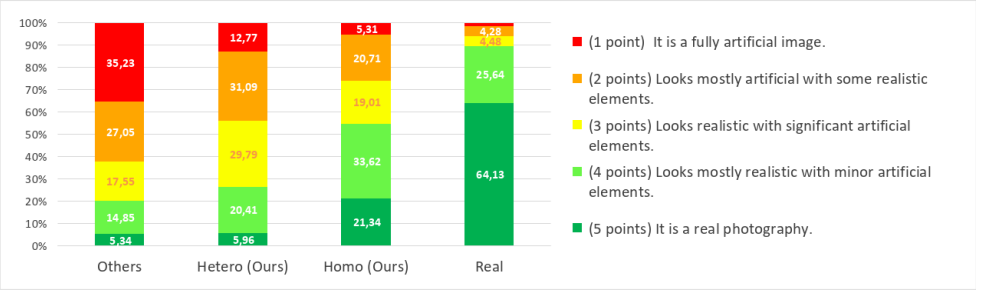


Figure 6: Bar chart of the first part of our user study on realism perception of foggy images. The split of each bar displays the percentage of users assigning a specific rating (denoted by color) to each method.

4.2 User Study

To validate our fog simulation approach quantitatively, we conducted a user study consisting of two parts. We recruited 48 participants with varying levels of experience in computer graphics and image processing and employed Mean Opinion Score to compare and evaluate all methods.

In the first part of our study, participants are asked to evaluate one image at a time on a 5-point scale of realism, ranging from "It is a fully artificial image" to "It is a real photography." (see Fig. 6). The study includes 40 images divided into four equal parts: images from previous approaches (Others) [1, 3, 4, 6, 8], our heterogeneous fog simulation (Hetero), our homogeneous fog simulation (Homo), and real photographs of foggy weather from the web (Real). The images cover diverse scenarios, such as indoors, outdoors, with varying levels of visibility and lighting conditions. Examples of previous methods include images from foggy Cityscapes [6], NuScenes [4], 3D Common Corruptions [3], as well as fully synthetic images from the CARLA simulator [1] and the FRIDA dataset [8].

Fig. 6 and Tab. 1 demonstrate the results of our user study on images. Evaluating Real photographs, only 64.13% of participants rated them with the highest realism score (5 points), perceiving them as authentic images. Other 25.64% identified minor artificial elements, while 5.77% (comprising 1.49% and 4.28%) rated the real images with 1 and 2 points, considering them mostly or completely artificial. Our Homogeneous method was per-

Method	Fully artificial ⟷ Real photography Rating					Mean Rating (points)	Realism (%)
	1	2	3	4	5		
Others	35.23	27.05	17.55	14.85	5.34	2.28	32
Hetero (Ours)	12.77	31.09	29.79	20.41	5.96	2.76	44
Homo (Ours)	5.31	20.71	19.01	33.62	21.34	**3.45**	**61**
Real	1.49	4.28	4.48	25.64	64.13	4.47	87

Table 1: Results from the first part of our user study on realism perception of foggy images. For each rating (in a 5-point scale), the percentage of users assigning that rating is provided. The rating with the most votes for each method is marked in blue. The mean rating score shows the total rating for each method. The best mean rating score and the higher percentage of fog realism across all methods are highlighted in bold.

Video	Homogeneous ⟵ Rating ⟶ Heterogeneous					Mean Rating (points)	Realism (%)
	−2	−1	0	1	2		
MOT17-02	4.3	21.3	40.4	21.3	12.8	0.17	8
MOT17-04	10.6	27.7	10.6	38.3	12.8	0.15	7
MOT17-05	14.9	36.2	12.8	21.3	14.9	−0.15	−7
MOT17-09	0	15.2	30.4	39.1	15.2	**0.54**	**27**
Total	5.56	19.88	18.84	24.2	11.54	0.18	8.93

Table 2: Results from the second part of our user study, in which participants were asked to compare homogeneous and heterogeneous fog types. For each rating on a scale from −2 to 2 (with absolute values indicating realism and a sign indicating the fog type: minus for homogeneous, plus for heterogeneous), the percentage of users assigning that rating is provided. The rating with the most votes for each video pair is marked in blue. The mean rating score shows the total rating for each video pair. The best absolute values of mean rating score and percentage of realism across all video pairs are highlighted in bold.

ceived as the most realistic among the simulated methods, garnering 55.94% in the top two rating scores (green bars in Fig. 6) and 73.95% in overall "realistic" split (scoring between 3 to 5 points). Only 5.31% of participants regarded our images as fully artificial. Tab. 1 provides further validation of these findings, with Real photographs receiving the highest mean rating score (4.47 points out of 5), followed by our Homogeneous (3.45 points) and Heterogeneous (2.76 points) methods. Converting mean rating scores from a 5-point scale into a $[0,100]$ interval, we obtain a metric for assessing the realism of fog in percentage. Accordingly, Real images achieve 87% of realism, while Homogeneous and Heterogeneous methods gain 61% and 44% of realism, respectively. This means, our approach, when evaluated on images, was judged to be 12% more realistic for heterogeneous fog and 29% more realistic for homogeneous fog compared to the state-of-the-art methods.

Interestingly, people tend to perceive heterogeneous fog appearance in images as more artificial than homogeneous fog. This perception might be caused by several factors. First, viewers may be more accustomed to seeing homogeneous fog in real-life photographs and heterogeneous fog in video games, leading them to perceive heterogeneous fog as less realistic. Second, heterogeneous fog may introduce inconsistencies in visibility or lighting conditions, resulting in a less coherent overall appearance. Third, random difference in fog intensity in heterogeneous fog may create more noticeable contrasts between foggy and clear areas, potentially appearing artificial to viewers. However, it's important to note that heterogeneous fog is indeed a phenomenon that exists in nature. For instance, it is commonly observed in mountainous regions, where fog gathers in valleys, or near bodies of water such as lakes or rivers. Also in urban scenes, buildings, streets, and other structures can lead to variations in temperature, humidity, and airflow, resulting in complex fog patterns that include heterogeneity.

In the second part of our user study, participants compared pairs of videos – one with homogeneous fog and the other with heterogeneous fog. They rated which type of fog appeared more realistic on a scale from −2 to 2, ranging from "Left video looks definitely more realistic than right" to "Right video looks definitely more realistic than left". If both videos looked identical, the test received 0 points. Absolute values of rating scores ($|\pm 1|$ or $|\pm 2|$) indicated the magnitude of fog realism, while a sign (minus or plus) denoted homogeneous and heterogeneous fog, respectively. We randomized the left-right location of homogeneous

(a) Three tracks was found in homogeneous fog (left) and one more mostly occluded track was additionally found in heterogeneous fog (right).

(b) Five tracks was found in homogeneous fog (left) and one more very distant track was additionally found in heterogeneous fog (right).

Figure 7: Comparison of MOT performance (FairMOT tracker) in homogeneous and heterogeneous fog of the third intensity level (*Fog 3*) using two scenes from MOT17-02 as an example. Heterogeneous scenarios demonstrate slightly better results compared to homogeneous fog. (Best viewed zoomed in on screen.)

and heterogeneous fog simulations for independent evaluation.

The study validated fog simulation in the MOTChallenge (third release: MOT17 dataset) videos across various scenarios, including outdoor scenes with visible sky, indoor settings, and nighttime surveillance scenarios. Tab. 2 presents the results of our user study on videos. In outdoor daytime static frontal view scenario (MOT17-02), a higher percentage of participants (40.4%) found both homogeneous and heterogeneous fog equally realistic. In semi-indoor view of a covered pedestrian street, filmed by a static camera from a low angle position (MOT17-09), 39.1% of participants rated heterogeneous fog as slightly more realistic than homogeneous. Conversely, in the outdoor daytime view from a moving camera (MOT17-05), 36.2% of participants thought that homogeneous fog looked a little better. Overall, across all video pairs, we have a mean rating score of 0.18, which is positive. This indicates that heterogeneous fog is perceived as more realistic (by 8.93%) than homogeneous.

Remarkably, heterogeneous fog simulation in videos was perceived as more realistic than homogeneous, contrary to the findings in image augmentation. This finding highlights the importance of heterogeneous fog rendering in videos, which, to our knowledge, has not been done before and actually makes our method more suitable for MOT tasks.

5 Further Analysis of MOT Robustness to Fog

Tab. 3 illustrates the degradation in MOT performance evaluated on the augmented MOT17 dataset with both homogeneous and heterogeneous fog. For each tracker (ByteTrack, Tracktor++, CenterTrack, FairMOT and TransCenter) and fog intensity level (*Fog 1*, *Fog 2*, *Fog 3*, and *Fog 4*), we compute the percentage drop in performance compared to clear weather conditions. We consider HOTA, MOTA, and IDF1 metrics.

ByteTrack and TransCenter trackers demonstrate better robustness to foggy atmospheric conditions. Moreover, across all trackers, there is slightly less degradation in performance under heterogeneous fog compared to homogeneous. These results may be attributed to the presence of transparent regions between clouds, which improve overall visibility. We found that in heterogeneous fog, trackers can handle occlusions slightly better (see Fig. 7(a)) and occasionally detect distant objects more effectively (see. Fig. 7(b)).

Method	Metric	Homogeneous Fog (%)				Heterogeneous Fog (%)			
		Fog 1	Fog 2	Fog 3	Fog 4	Fog 1	Fog 2	Fog 3	Fog 4
ByteTrack	HOTA	**−0.4**	**−4.8**	**−25.9**	**−64.9**	**−0.1**	**−2.6**	**−11.7**	−61.6
	MOTA	**−0.7**	**−4.6**	**−37.9**	**−85.0**	**−0.1**	**−2.4**	**−18.6**	**−74.9**
	IDF1	−1.6	**−5.2**	**−27.5**	**−75.3**	**−0.7**	**−3.2**	**−11.7**	**−60.7**
Tracktor++	HOTA	−14.4	−38.1	−71.4	−85.9	−9.6	−26.3	−59.9	−84.6
	MOTA	−15.4	−48.7	−86.7	−96.0	−9.5	−29.8	−78.0	−96.1
	IDF1	−18.3	−45.5	−82.2	−94.3	−12.3	−31.2	−71.0	−94.5
CenterTrack	HOTA	−3.6	−20.0	−51.3	−80.9	−1.8	−14.8	−38.2	−71.5
	MOTA	−6.1	−32.2	−72.9	−94.3	−3.7	−21.8	−59.4	−85.6
	IDF1	−3.5	−19.9	−57.3	−89.5	−1.6	−15.7	−42.9	−79.6
FairMOT	HOTA	−2.7	−11.9	−48.0	−82.7	−0.4	−8.4	−29.5	−66.9
	MOTA	−5.2	−17.8	−65.9	−94.4	−3.7	−23.7	−39.6	−84.0
	IDF1	−3.8	−13.4	−54.8	−90.5	−0.9	−9.8	−32.4	−75.6
TransCenter	HOTA	−1.9	−7.1	−39.9	−68.5	−0.9	−4.9	−19.8	**−58.4**
	MOTA	−5.9	−12.8	−59.9	−92.7	−4.2	−9.6	−32.9	−80.9
	IDF1	**−1.2**	−6.5	−44.4	−85.2	−0.8	−4.2	−20.2	−67.8

Table 3: MOT performance degradation scores (in percent) compared to clear weather conditions, evaluated on the augmented MOT17 dataset with homogeneous and heterogeneous fog at four intensity levels (*Fog 1* to *Fog 4*). The best scores for each metric across all trackers are highlighted in blue bold, the second best – in blue.

References

[1] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An Open Urban Driving Simulator. In *Conference on Robot Learning (CoRL)*, 2017.

[2] Kaiming He, Jian Sun, and Xiaoou Tang. Single Image Haze Removal Using Dark Channel Prior. *IEEE TPAMI*, 2011.

[3] Oğuzhan Fatih Kar, Teresa Yeo, Andrei Atanov, and Amir Zamir. 3D Common Corruptions and Data Augmentation. In *CVPR*, 2022. Oral.

[4] Dipkumar Patel. Nuscenes fog augmented samples for bad weather pre. https://www.kaggle.com/datasets/dipkumar/nuscenes-fog-augmented-samples-for-bad-weather-pre.

[5] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. *IEEE TPAMI*, 2022.

[6] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic Foggy Scene Understanding with Synthetic Data. *IJCV*, 2018.

[7] Robby T. Tan. Visibility in bad weather from a single image. In *CVPR*, 2008.

[8] Jean-Philippe Tarel, Nicolas Hautière, Aurélien Cord, Dominique Gruyer, and Houssam Halmaoui. Improved visibility of road scene images under heterogeneous fog. In *Intelligent Vehicles Symposium (IVS)*, 2010.