Effective Video Mirror Detection with Inconsistent Motion Cues

Supplementary Material

This supplementary material provides additional details and comparisons of our implementations. These include:

- A more comprehensive description of our data collection process.
- Further statistical analysis of the dataset.
- Quantitative results and analysis of the generalization capabilities of our proposed method and the state-of-the-art models, trained on MMD (Ours) and tested on VMD-D [6].
- Additional qualitative results showcasing the performance of our method, and
- A video highlighting the results of the ablation study.

1. Motion Mirror Dataset

1.1. Dataset Creation

We embarked on the development of a comprehensive video mirror dataset tailored for real-world applicability. This dataset comprises thirty-seven 10-second videos intentionally capturing diverse lighting and environmental conditions. An essential consideration was incorporating consistent motion in these videos to simulate scenarios akin to those encountered in drone footage.

To compile the dataset, we recruited six individuals who voluntarily participated in the data collection process. These participants received explicit instructions to use contemporary mobile phone cameras, ensuring a minimum video resolution of 1080p (1920x1080px). They were directed to record mirror videos in various locations, including homes (specifically bathrooms, living rooms, bedrooms, and hallways), department stores, gyms, and within cars. Additionally, participants were guided to record in different lighting conditions, spanning daytime, natural and unnatural lighting, night-time, and darker environments. The dataset encompasses a wide range of features and conditions, making it relevant for video mirror detection in diverse real-world scenarios.

The manual annotation process was carefully executed. Every third frame of the videos underwent annotation, and interpolation between frames was performed using the method outlined in [1]. Furthermore, a depth-first-search algorithm was computed on each frame to obtain edge features. Each annotated frame, including mirror and edge annotations, underwent manual verification to ensure it attained ground truth quality.

1.2. Dataset Analysis

Table 1 compares our proposed dataset (MMD) with VMD-D [6] in terms of Average Mirror Count Per Frame. The metric highlights the challenges in our dataset. Detecting multiple mirror regions is a significant challenge in video mirror detection, and our MMD poses an even greater challenge with a higher average count compared to VMD-D [6]. Our experimental results in the main paper demonstrate that leveraging inconsistent motion cues in our proposed method MG-VMD contributes to the successful detection of multiple mirror regions within videos.

Dataset	Average Mirror Count Per Frame ↑
MMD (Ours)	1.746326450
VMD-D	1.553479682

Table 1. Table comparing our proposed dataset (MMD) against state-of-the-art VMD-D [6] in terms of Average Mirror Count Per Frame. This metric highlights the challenges within our dataset, especially those with multi-mirror scenes. Red and Blue indicate the best and second-best performances, respectively.

2. Generalization on VMD-D dataset [6]

Model	Accuracy↑	F_β↑	MAE↓
MG-VMD (Ours)	0.685060	0.576727	0.314939
VMD-Net	0.654699	0.389780	0.345300

Table 2. Quantitative results table comparing our proposed method with the state-of-the-art VMD-Net [6]. The two models are trained on our proposed dataset (MMD) and tested on the VMD-D [6] directly. The results demonstrate the generalizability of our approach for the Video Mirror Detection task. Red and Blue indicate the best and second-best performance.

Here, we further compare the generalizing performance of our proposed model, MG-VMD, with VMD-Net [6]. We train both models on our dataset (MMD) and tested on the VMD-D [6] dataset. Table 2 shows that MG-VMD outperforms VMD-Net in Mean Absolute Error (MAE↓), F-measure (F_β↑), and Accuracy↑. The consistent better performance demonstrates the generalizability of our approach for the Video Mirror Detection problem.

3. Further Qualitative Results on MMD

Figure 1 further presents more qualitative results, comparing our proposed model with eight state-of-the-art methods from Video Salient Object Detection and Image-Based/Video Mirror Detection.

When evaluating video mirror detection tasks, our method demonstrates better temporal consistency compared to VMD-Net [6]. This is evident in the 1^{st} and 2^{nd}, 5^{th} and 6^{th}, and 9^{th}–12^{th} rows. This enhanced temporal consistency can be attributed to our model leveraging inconsistent motion cues, and the use of the optical flow coherence block and coherence loss. Single-frame mirror detections (PMD-Net [5], MirrorNet [9]) occasionally fail to detect mirrors in these videos, specifically in the 1^{st}, 7^{th}, and 8^{th} rows.

Furthermore, our technique exhibits greater robustness and stability in identifying mirror regions and their boundaries compared to video salient object detection methods.

3.1. Qualitative Video Showcase

Along with this supplementary material, we present a video titled `"MG-VMD_qaulitative_MMD.avi"`, demonstrating the qualitative performance of our video mirror detection method under ablation study. The video comprises six rows, with the last four each representing a different stage of our method, with the inclusion of the respective Motion-guided Edge Detection Module (MEDM) and Motion Attention Module (MAM):

1. Current Frame
2. Ground Truth
3. Baseline
4. Baseline + MEDM
5. Baseline + MAM
6. MG-VMD (combining MEDM and MAM)

The samples featured in the video are drawn from our proposed MMD dataset. The final exported video maintains a frame rate of 30 fps, and each prediction map has a resolution of (224×224px).

References

[1] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *Eur. Conf. Comput. Vis.*, 2022. 1

[2] Ge-Peng Ji, Keren Fu, Zhe Wu, Deng-Ping Fan, Jianbing Shen, and Ling Shao. Full-duplex strategy for video object segmentation. In *Int. Conf. Comput. Vis.*, 2021. 3

[3] Qingming Huang Jun Wei, Shuhui Wang. F3net: Fusion, feedback and focus for salient object detection. In *AAAI*, 2020. 3

[4] Haofeng Li, Guanqi Chen, Guanbin Li, and Yizhou Yu. Motion guided attention for video salient object detection. In *Int. Conf. Comput. Vis.*, 2019. 3

[5] Jiaying Lin, Guodong Wang, and Rynson W.H. Lau. Progressive mirror detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2, 3

[6] Jiaying Lin, Xin Tan, and Rynson W.H. Lau. Learning to detect mirrors from videos via dual correspondences. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 1, 2, 3

[7] Yukun Su, Jingliang Deng, Ruizhou Sun, Guosheng Lin, and Qingyao Wu. A unified transformer framework for group-based segmentation: Co-segmentation, co-saliency detection and video salient object detection. *IEEE Trans. Multimedia*, 2023. 3

[8] Yi Tang, Yuanman Li, and Guoliang Xing. Video salient object detection via adaptive local-global refinement, 2021. 3

[9] Xin Yang, Haiyang Mei, Ke Xu, Xiaopeng Wei, Baocai Yin, and Rynson W. H. Lau. Where is my mirror? In *Int. Conf. Comput. Vis.*, 2019. 2, 3

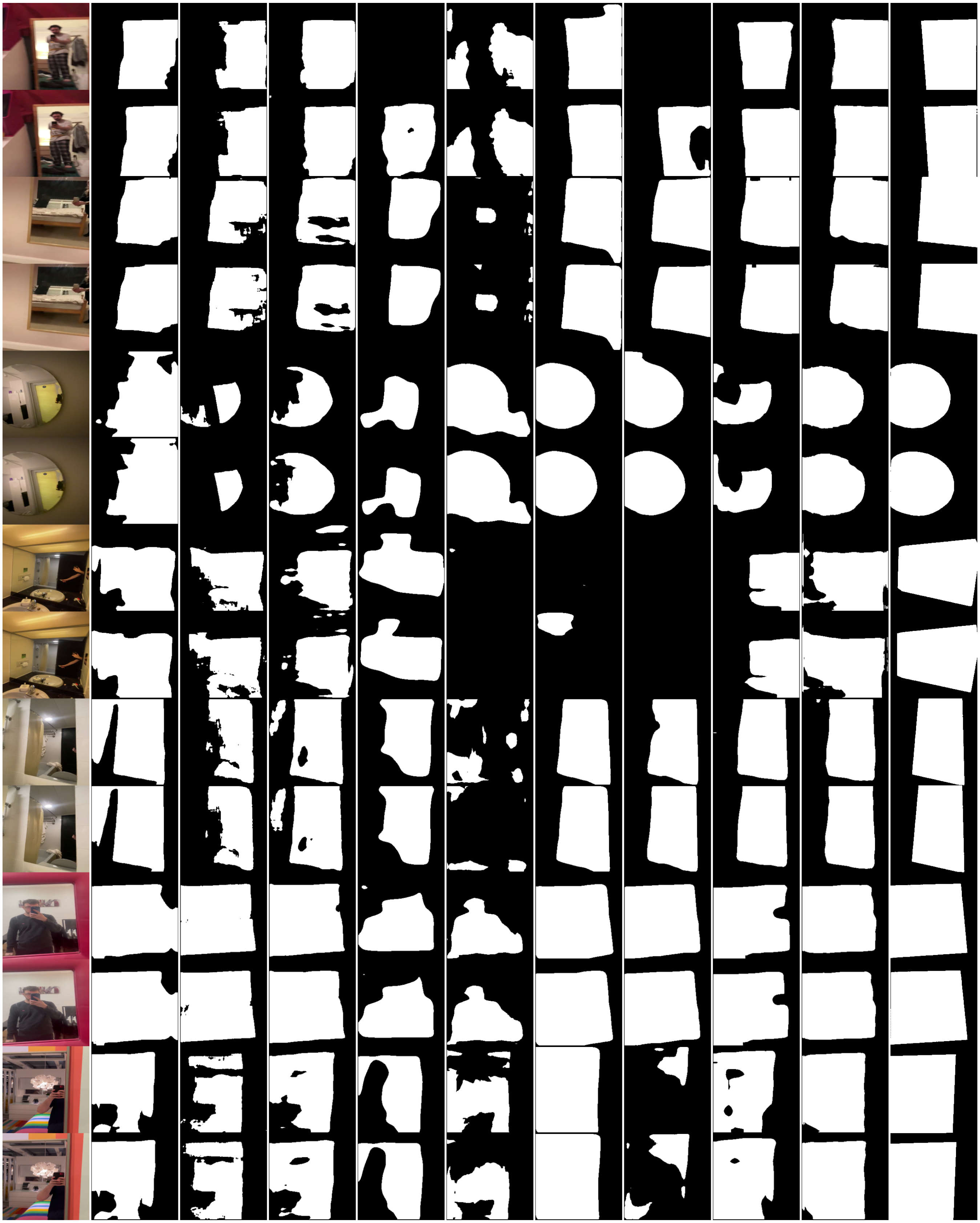

Figure 1. A further comprehensive qualitative results table comparing our proposed method with state-of-the-art video salient object detection, as well as image-based/video mirror detection (namely, FS-Net [2], MGA [4], ALGRF [8], F3Net [3], UFO [7], PMD-Net [5], MirrorNet [9], VMD-Net [6]). The models were trained and validated on our proposed MMD dataset.