



Provenance constrained temporal reconstruction of lens obscured images for intelligent vehicles

Abdulazizxon Axmedov Ganijon o'g'li

PhD student, Namangan State University, Namangan, Uzbekistan

ORCID: 0009-0004-6784-4010

Email: abdulazizahmedov97@gmail.com

abdulaziz_axmedov@namdu.uz

ABSTRACT

Earlier camera frames can help recover a view obscured by lens contamination. However, a genuine historical pixel may show an outdated object position. This paper presents a temporal reconstruction rule that accepts only recent, unobstructed source pixels, requires agreement between two historical samples, and excludes potentially dynamic regions. Rejected pixels remain unknown, and reconstructed images never enter the source history. A controlled experiment with computer-generated road sequences tests three motion levels, three obstruction sizes, three alignment biases, and four dynamic-mask retention settings. With exact background alignment and complete dynamic exclusion, the rule reconstructs 9.72% of obscured pixels with zero error on the accepted subset. Clean-source warping reconstructs 63.31%, with a mean absolute RGB error of 0.0596. With a two-pixel alignment bias, constrained coverage rises to 13.79%, but 2.91% of accepted pixels have an error above 0.10. Removing dynamic exclusion also allows outdated pedestrian evidence to pass. The results show the trade-off between recovery coverage and error. Tracking pixel origin alone cannot establish that reused content matches the current scene. Evaluation on recorded driving sequences and downstream perception tasks is still needed

Keywords:

intelligent vehicles;
automotive cameras;
lens obstruction;
temporal
reconstruction; pixel
provenance; motion
compensation;
selective
reconstruction; driver
assistance.

Восстановление изображений с перекрытым объективом по предыдущим кадрам с ограничениями на источники данных для интеллектуальных транспортных средств

АННОТАЦИЯ

Предыдущие кадры камеры могут помочь восстановить часть изображения, закрытую загрязнением объектива. Однако подлинный пиксель из прошлого кадра может показывать устаревшее положение объекта. В статье предложено правило временного восстановления, которое принимает только недавние наблюдения из незакрытых участков, требует согласованности двух исторических образцов и исключает потенциально динамические области. Отклонённые пиксели остаются неизвестными, а восстановленные изображения не пополняют историю исходных наблюдений. В контролируемом эксперименте с созданными на компьютере дорожными сценами проверены три уровня движения, три размера перекрытия, три ошибки совмещения и четыре режима сохранения динамической

Ключевые слова:

интеллектуальные
транспортные
средства;
автомобильные
камеры; перекрытие
объектива;
восстановление по
предыдущим
кадрам;
происхождение
пикселей;
компенсация



маски. При точном совмещении фона и полном исключении динамических областей правило восстанавливает 9,72% закрытых пикселей с нулевой ошибкой на принятом подмножестве. Перенос незакрытых исходных пикселей восстанавливает 63,31% при средней абсолютной ошибке RGB 0,0596. При ошибке совмещения в два пикселя доля восстановления методом с ограничениями достигает 13,79%, но у 2,91% принятых пикселей ошибка превышает 0,10. Отмена исключения динамических областей также допускает устаревшие данные о пешеходе. Результаты показывают компромисс между долей восстановления и ошибкой. Одного отслеживания происхождения пикселя недостаточно, чтобы подтвердить соответствие повторно используемых данных текущей сцене. Необходима оценка на записях дорожного движения и в последующих задачах машинного восприятия.

движения;
выборочное
восстановление;
системы помощи
водителю.

Intellectual transport vositalarida obyektiv to'silganda tasvirlarni manba cheklovlari asosida oldingi kadrlar yordamida tiklash

ANNOTATSIYA

Oldingi kamera kadrlari obyektiv ifloslanishi tufayli to'silgan tasvir qismlarini tiklashga yordam berishi mumkin. Biroq asl manbadan olingan piksel ham obyektning eskirgan holatini ko'rsatishi mumkin. Ushbu maqolada faqat yaqinda olingan, to'silmagan manba piksellerini qabul qiladigan vaqt bo'yicha tiklash qoidasi taklif etiladi. Qoida oldingi ikki kadrda namunalari mos kelishini talab qiladi va harakat bo'lishi mumkin bo'lgan hududlarni istisno etadi. Rad etilgan pikseller noma'lum holatda qoladi, tiklangan tasvirlar esa manba kadrlari tarixiga kiritilmaydi. Kompyuterda yaratilgan yo'l sahnalari bilan o'tkazilgan nazoratli tajribada harakatning uch darajasi, to'siqning uch o'lchami, moslashtirishdagi uch siljish xatosi va dinamik niqobni saqlashning to'rt holati tekshirildi. Fon aniq moslashtirilib, dinamik hududlar to'liq istisno etilganda, qoida to'silgan piksellarning 9,72 foizini tikladi va qabul qilingan qismda xato kuzatilmadi. To'silmagan manba piksellerini ko'chirish usuli 63,31 foizni tikladi; RGB bo'yicha o'rtacha mutlaq xato 0,0596 bo'ldi. Ikki pikselni moslashtirish xatosida cheklovli usulning tiklash qamrovi 13,79 foizga yetdi, ammo qabul qilingan piksellarning 2,91 foizida xato 0,10 dan oshdi. Dinamik hududlarni istisno qilish bekor qilinganda, piyodaning oldingi holatiga oid eskirgan ma'lumotlar ham qabul qilindi. Natijalar tiklash qamrovi va xato o'rtasidagi murosani ko'rsatadi. Piksel manbasini kuzatishning o'zi qayta ishlatilgan ma'lumot joriy sahnaga mosligini tasdiqlamaydi. Usulni transport harakati yozuvlari va keyingi idrok vazifalarida baholash zarur.

Kalit so'zlar:

intellektual transport vositalari; avtomobil kameralari; obyektivning to'silishi; oldingi kadrlar yordamida tiklash; piksel manbasi; harakatni kompensatsiyalash; tanlab tiklash; haydovchiga yordam tizimlari.

Introduction

Vehicle cameras help identify road markings, nearby cars, pedestrians, and traffic conditions. Lens contamination can block the same image coordinates across many frames while the vehicle and surrounding objects move. WoodScape includes soiling detection among its automotive camera tasks [1], and SoilingNet addresses lens contamination directly [2]. Detecting the obstruction tells us where the image is unreliable. It does not reveal what is currently hidden behind it.

An earlier frame may show a scene point before it moves into the obscured image region. Reusing that observation can recover part of the view. Yet a pedestrian or car may have moved, changed direction, or entered the scene since that frame was captured. Copying old road texture may hide a newly arrived object. Copying an earlier pedestrian position may create a false obstacle.

We ask which historical pixels can be reused when some changes behind an obstruction cannot be checked. The proposed rule requires multiple eligible observations, records the source of every accepted value, and leaves unsupported pixels unknown. We evaluate the resulting trade-off between coverage and error, including failures caused by shared alignment bias. The contribution is the selection rule and its evaluation; the study does not introduce an optical-flow network.

Related research and scope

Motion estimation provides the correspondence needed to reuse earlier observations. RAFT estimates dense optical flow through recurrent updates and all-pairs correlations [3]. UnFlow uses occlusion-aware bidirectional estimation for unsupervised learning [4]. Both studies motivate checks on correspondence reliability. We use prescribed background motion and controlled alignment errors to isolate their effect on acceptance, rather than running either network.

E2FGVI combines flow completion, feature propagation, and content synthesis for video inpainting [5]. ProPainter combines propagation with transformer-based completion [6]. Our question concerns the acceptance of historical observations in a vehicle perception interface. These published models are not numerical baselines in this study, so the results do not establish better image quality than either method.

Selective prediction allows a model to reject a case instead of always returning an answer. SelectiveNet learns this behaviour [7]. Here, deterministic tests accept or reject individual obscured pixels. When the available observations fail those tests, the output remains incomplete. The rule provides no calibrated probability or statistical risk guarantee.

The method is intended as an auxiliary camera-processing layer for intelligent vehicles. Its evidence labels may support traceability, but the study does not establish compliance with ISO 21448 [8]. Cleaning, braking, and sensor-fusion decisions lie outside the evaluated method.

Reconstruction method

Original observations and source eligibility

Let $I(t)$ be a normalized RGB image and $M(t)$ its binary obstruction mask, with one marking unreliable locations. The history stores original observations from 1, 3, and 5 frames earlier. $W(t,k,x)$ maps current location x to its corresponding coordinate in frame $t-k$. A source pixel is eligible only if it is inside that frame and outside its obstruction mask. Reconstruction is limited to $M(t)$; other current pixels stay unchanged.

The constrained rule accepts sources no more than three frames old, so only lags 1 and 3 can pass. The comparison method can also use lag 5. At an assumed rate of 30 frames per second, three frames represent 0.10 s. This is an observation age, not measured processing latency or a validated control limit.

Agreement and dynamic exclusion

For each obscured pixel, $C(t,x)$ contains the eligible RGB samples after the age check. The disagreement $\Delta(t,x)$ is the largest channel-wise range across those samples. Acceptance requires at least two samples and $\Delta \leq 0.06$ on the normalized scale. The youngest eligible sample supplies the output value. The rule checks agreement without averaging the samples or generating new content.

$D(t)$ is the region excluded because of possible object motion. In practice, it would need to include current moving objects, uncertain boundaries, and areas those objects could enter. Here, $D(t)$ is built from the known current pedestrian mask and its three warped historical masks. Using these known masks isolates the selection rule from detector errors. We then remove parts of $D(t)$ to test incomplete exclusion.

$$A(t,x) = M(t,x) \cdot 1\{|C(t,x)| \geq 2\} \cdot 1\{\Delta(t,x) \leq 0.06\} \cdot 1\{x \notin D(t)\}. \quad (1)$$

$A(t,x)=1$ means that a pixel is accepted for temporal reuse. A zero inside the obstruction mask means the location remains unknown. The output image retains this acceptance mask. Unknown pixels may be shown in a diagnostic colour, but that colour must not be treated as scene evidence.

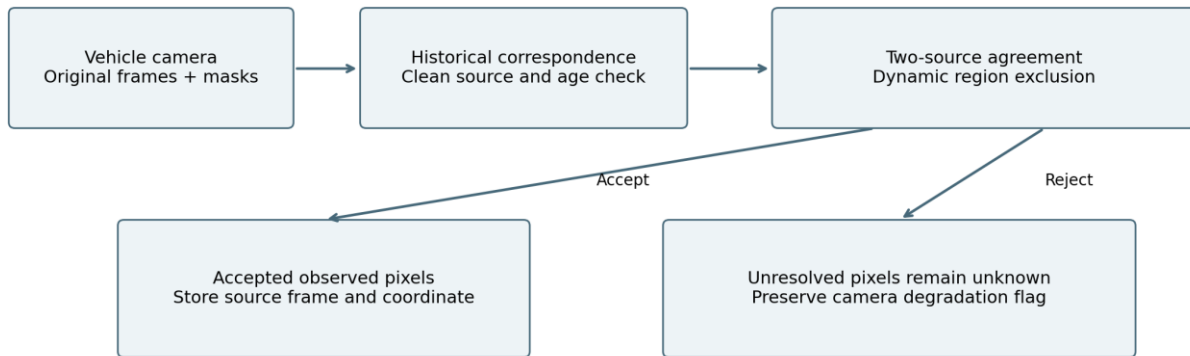


Figure 1. Proposed evidence-selection interface for a vehicle camera. Acceptance retains source identity; rejection preserves an explicit unknown state.

Provenance and the prohibition on recirculation

Every accepted value retains its original frame identifier, source coordinate, age, and mask state. Only original camera observations enter the history buffer. Reconstructed images never return as source frames. Repeated reuse therefore cannot reset a pixel's age or turn reconstructed content into an apparently direct observation.

This property follows from the buffer update rule. The buffer starts with original observations and receives only original observations at each update. Every accepted value can therefore be traced to a recorded source. That trace establishes where the value came from, but not whether it belongs at the mapped location in the current frame. A biased mapping can pass every source check.

Two candidates may agree because they share an alignment error or because an object was absent from both source frames. They are temporally related samples, not independent confirmations. Agreement helps reject some candidates, but cannot rule out a road user entering an obscured region.

Vehicle integration and large obstructions

A vehicle implementation should distinguish directly observed, temporally reused, and unknown pixels. Recovered road markings may support visualization or a separately validated perception function. Partial recovery must not clear the physical camera-health warning. Functions that require current observations should continue to treat unknown regions as unavailable.

The local rule is insufficient when much of the view is blocked. A supervisory gate should disable reconstruction if obstruction coverage exceeds its permitted limit or too little of the required task region remains visible. No universal percentage is proposed: blocking a small pedestrian crossing may matter more than blocking a larger area of sky. The present tests use small circular masks and do not validate extensive lens blockage.

Controlled computational experiment

Image construction and controlled factors

The experiment uses a computer-generated road panorama of 1024×160 pixels. It contains stationary car shapes, buildings, road texture, and lane markings. A moving horizontal crop produces 256×160 RGB frames. Three sequences, each 20 frames long, use displacements of 2, 4, or 8 pixels per frame. These image displacements are not converted to vehicle speed because the scene has no camera calibration, depth, or three-dimensional motion.

A red pedestrian proxy moves independently. Its horizontal centre is $78+9t$ pixels through $t=12$, then $186-12(t-12)$, so it reverses direction. Its head, torso, and legs occupy rows 86–127.



Before drawing the shapes, a deterministic texture offset, $((7x+11y) \bmod 17)-8$, is added to each colour channel of the sky, verge, and road. The frames are generated from these fixed geometry and motion rules.

Table 1.

Controlled experiment settings

Factor	Values and interpretation
Evaluation frames	$t = 5$ to 19; 15 frames per sequence
Background displacement	2, 4, 8 pixels per frame
Obstruction radius	8, 16, 24 pixels; fixed image location
Source history	Lags 1, 3, 5; constrained maximum age 3
Alignment bias	0, 1, 2 pixels; common to all lags
Agreement threshold	0.06; maximum per-channel candidate spread
Dynamic-mask retention	0%, 25%, 50%, 100% of periodic column classes
Design size	108 factor combinations; 1620 constrained frame evaluations

A circular obstruction remains fixed at (128,108). Radii of 8, 16, and 24 pixels cover 197, 797, and 1793 pixels, or 0.48%, 1.95%, and 4.38% of the image. Obscured pixels take the normalized colour (0.30,0.23,0.13). The same mask is applied to source observations. Recovery is possible only where a historical correspondence falls outside that mask.

For displacement s and lag k , the backward mapping samples source column $x+sk+e$, with unchanged row coordinates. The horizontal bias e is 0, 1, or 2 pixels and is shared by all lags. This tests a correlated error that agreement between sources may miss. The mapping follows background translation only; it does not compensate for pedestrian motion.

Dynamic-mask retention q keeps pixels in $D(t)$ where $x \bmod 4 < 4q$, for $q \in \{0,0.25,0.5,1\}$. These settings retain zero, one, two, or four periodic column classes. They represent controlled spatial removal, not measured detector recall. The actual retained fraction depends on object shape. No dilation is applied. This simple perturbation tests missing exclusion evidence without modelling a particular detector.

The design has $3 \times 3 \times 3 \times 4 = 108$ configurations. Each is evaluated at 15 time points, giving 1620 constrained frame evaluations. The same three short sequences are reused across masks and perturbations. The totals describe this deterministic experiment; they are not independent road trials or a basis for statistical significance tests.

Comparison and measures

Clean-source warping accepts the youngest unmasked historical sample from lags 1, 3, and 5. It uses the same mapping and source-mask checks, but requires neither two-source agreement nor dynamic exclusion. The constrained method adds those checks and a three-frame age limit. The comparison tests these conditions together, so it cannot isolate the effect of each one.

Let B contain all evaluated obscured pixels and $A \subseteq B$ the accepted pixels. For each accepted pixel, E is the mean absolute RGB difference from the unobstructed reference frame. Both images use normalized channel values. Coverage and mean absolute error (MAE) are calculated as follows:

$$\text{Coverage} = 100|A| / |B|; \quad \text{MAE} = \sum(E \text{ over } A) / |A|. \quad (2)$$

The large-error rate is the percentage of accepted pixels with $E > 0.10$. This threshold describes image error and has no established vehicle-safety meaning. Pedestrian fill measures the fraction of currently obscured pedestrian pixels given historical content, whether accurate or not. The tested policy aims to leave these dynamic locations unresolved.

$$\text{Large-error rate} = 100|\{x \in A : E(x) > 0.10\}| / |A|. \quad (3)$$

Counts and error sums are pooled before ratios are calculated, so larger masks contribute more pixels. A case with no accepted pixels adds nothing to the accepted count or error sum; it is not assigned a zero frame MAE. Error must be reported alongside coverage because rejecting almost every pixel can produce a low error over a very small accepted area.

Results

Recovery coverage and alignment bias

Table 2.

Results with complete dynamic exclusion for the constrained method

Method	Bias px	Coverage %	MAE	Large error %
Clean-source warp	0	63.31	0.0596	20.71
Constrained reuse	0	9.72	0.0000	0.00
Clean-source warp	1	65.53	0.0846	21.78
Constrained reuse	1	11.73	0.0322	1.42
Clean-source warp	2	67.64	0.0778	22.57
Constrained reuse	2	13.79	0.0246	2.91

With exact background alignment, constrained reuse accepts 9.72% of obscured pixels with zero error on that subset. Clean-source warping accepts 63.31%, with MAE 0.0596 and a large-error rate of 20.71%. The zero error follows from the controlled setup: accepted content is static, background translation is exact, and known dynamic regions are excluded. Recorded driving sequences are needed to test whether this behaviour extends to vehicle cameras.

A one-pixel bias raises constrained coverage to 11.73%, with MAE 0.0322 and a large-error rate of 1.42%. At two pixels, coverage reaches 13.79%, MAE is 0.0246, and the large-error rate is 2.91%. The extra coverage does not by itself show improvement. Bias moves some source coordinates outside the obstruction and makes more candidates eligible, even when their placement is wrong.

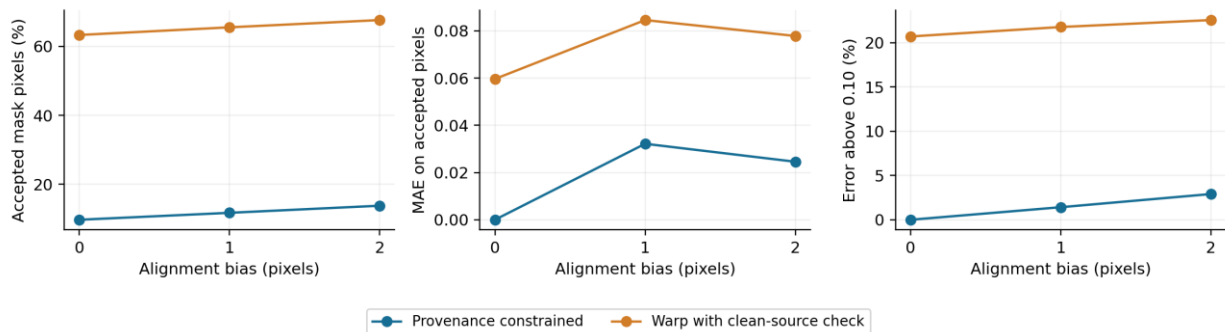


Figure 2. Pooled coverage and accepted-domain error under common alignment bias. Different methods accept different pixel subsets; the curves do not establish equal-coverage superiority.

MAE does not rise steadily with bias. The texture is periodic, and each bias changes the accepted pixel set. However, the large-error rate increases from one to two pixels. A shared mapping error can therefore pass the agreement check while distorting accepted edges or texture.

Where both methods accept a pixel, they select the same youngest historical sample and return the same RGB value. Constrained reuse lowers aggregate error by rejecting more locations. It does not estimate better colours at locations accepted by both methods. The substantial loss of coverage is part of that trade-off.

Dynamic exclusion sensitivity

At zero alignment bias, removing dynamic exclusion raises constrained coverage to 12.30%. MAE rises to 0.0368 and the large-error rate to 12.76%. Historical content fills 14.11% of currently obscured pedestrian pixels. Recent agreement between two sources therefore cannot by itself prevent outdated content from being copied to a moving person's current position.

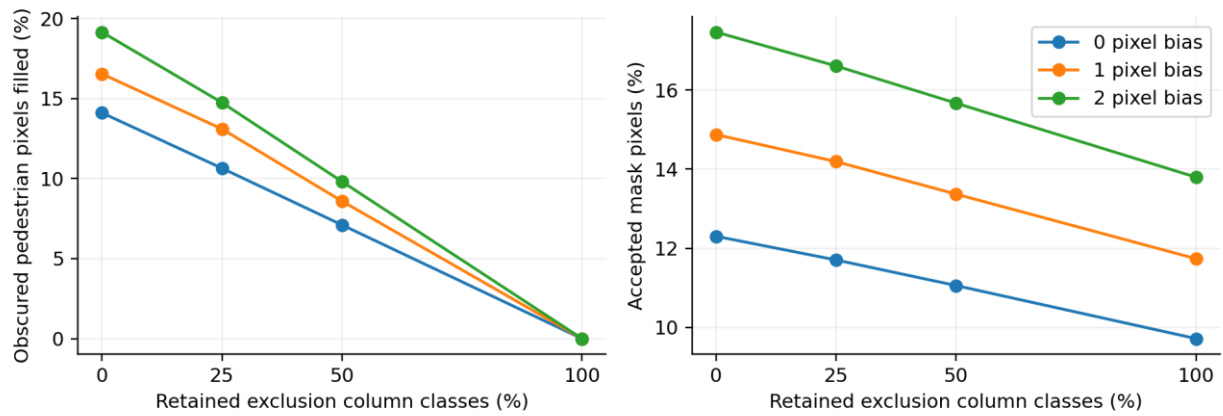


Figure 3. Sensitivity to retained dynamic-exclusion column classes. Pedestrian fill measures reuse at a current pedestrian location, not object-detection accuracy or collision risk.

Keeping two of the four column classes reduces pedestrian fill to 7.10% at exact alignment. With no exclusion and a two-pixel bias, fill reaches 19.15%. With the full current dynamic mask, fill is zero by construction. This confirms that the exclusion rule is enforced. It does not show that a camera can obtain an accurate mask behind an actual obstruction.

Obstruction size and visual behaviour

With zero bias and full exclusion, coverage falls from 18.75% at radius 8 to 11.07% at radius 16 and 8.12% at radius 24, pooled across motion levels. Larger obstructions have fewer recent clean correspondences relative to their area. The three-frame history mainly supports recovery near exposed regions, leaving much of the blocked area unknown.

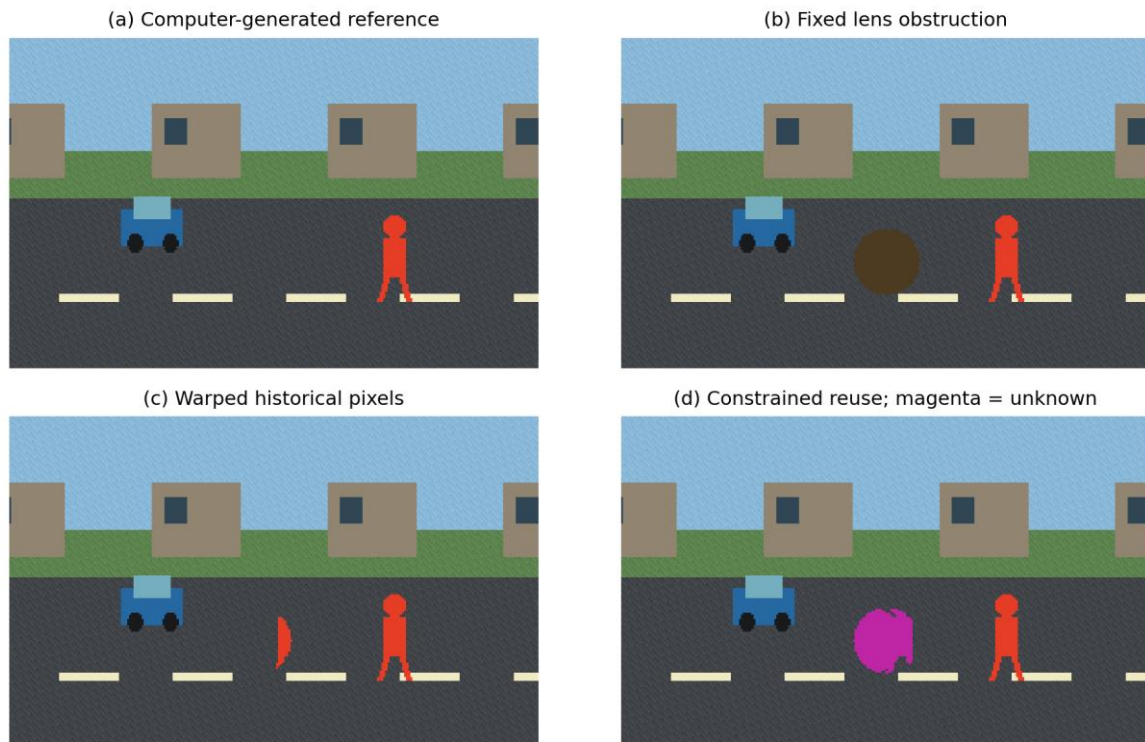


Figure 4. Computer-generated road-scene example at $t=12$, displacement 8 pixels per frame, radius 16, and zero bias. Unrestricted warping introduces an obsolete pedestrian fragment. Magenta marks unresolved pixels in the constrained output; it is a visualization overlay, not recovered scene content.

Figure 4 shows how a visually continuous image can still be misleading. Warping copies part of an earlier pedestrian position into the obscured area. The constrained output leaves unsupported pixels unknown while preserving the visible current pedestrian. It does not recover the whole scene behind the obstruction.

Discussion and limitations

In this experiment, rejecting unsupported pixels reduces error over the accepted area, but leaves most obscured pixels unavailable. Avoiding generated content is not enough to ensure reliable reuse. Genuine historical pixels may show outdated object positions or be misplaced by a shared motion error. Pixel origin, present validity, and usefulness for perception need separate assessment.

The hardest case is a road user entering an area that is fully obscured in the current image. A camera-only detector may have no evidence of that object. The complete-exclusion test avoids this problem by using the known pedestrian mask. A practical system cannot assume that information. It would need wider uncertainty regions around tracked objects, independent sensors, or abstention wherever new occupancy cannot be ruled out. Each approach would require evaluation.

The scenes are schematic and two-dimensional, with horizontal translation instead of full vehicle motion. They omit depth-dependent parallax, rolling shutter, lens distortion, illumination changes, rain transparency, vibration, and motion blur. Obstruction masks are exact, and clean source colours are noiseless. Periodic column removal does not represent a learned detector. These choices isolate the selection rule, but limit what the results say about driving conditions.

The accepted area is small, and we do not measure its effect on lane detection, vehicle detection, or driver decisions. We also report no latency, energy, or hardware throughput measurements. Reliable motion and object-uncertainty estimation may dominate the computational cost. Further tests should isolate each rejection condition while controlling source history and comparing methods at matched coverage.

The next evaluation should use recorded driving sequences, retaining clean current frames only for scoring. Fixed lens masks should be applied over time, with motion estimated from the degraded images. KITTI provides driving sequences and motion benchmarks [9]; automotive fisheye data would also be needed for surround-view geometry [1]. Tests should report results separately for static surfaces, moving cars, pedestrians, newly visible areas, and persistent occlusion. Comparing shared accepted pixels and matching coverage would help separate the benefit of rejection from improved reconstruction.

Conclusion

We presented and computationally evaluated a temporal reconstruction rule that tracks the origin of reused pixels. It accepts recent, unobstructed original observations only when sources agree and the location lies outside the dynamic exclusion region. Other obscured pixels remain unknown. The controlled results show a substantial trade-off between recovery coverage and error. They also expose failures under shared alignment bias and incomplete dynamic exclusion. The method provides a traceable acceptance decision, but its suitability for vehicle functions still requires recorded driving data, estimated motion, and downstream perception tests.

References

1. Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., Uricar, M., Milz, S., Simon, M., Amende, K., Witt, C., Rashed, H., Chennupati, S., Nayak, S., Mansoor, S., Perrotton, X., & Perez, P. (2019). WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9308–9318). IEEE. DOI: 10.1109/ICCV.2019.00940
2. Uricar, M., Krizek, P., Sistu, G., & Yogamani, S. (2019). SoilingNet: Soiling detection on automotive surround-view cameras. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 67–72). IEEE. DOI: 10.1109/ITSC.2019.8917178



3. Teed, Z., & Deng, J. (2020). RAFT: Recurrent all-pairs field transforms for optical flow. In *Computer Vision – ECCV 2020* (pp. 402–419). Springer. DOI: 10.1007/978-3-030-58536-5_24.
4. Meister, S., Hur, J., & Roth, S. (2018). UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). DOI: 10.1609/aaai.v32i1.12276.
5. Li, Z., Lu, C.-Z., Qin, J., Guo, C.-L., & Cheng, M.-M. (2022). Towards an end-to-end framework for flow-guided video inpainting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17541–17550. DOI: 10.1109/CVPR52688.2022.01704.
6. Zhou, S., Li, C., Chan, K. C. K., & Loy, C. C. (2023). ProPainter: Improving propagation and transformer for video inpainting. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10443–10452. DOI: 10.1109/ICCV51070.2023.00961.
7. Geifman, Y., & El-Yaniv, R. (2019). SelectiveNet: A deep neural network with an integrated reject option. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97, pp. 2151–2159)*. PMLR. <https://proceedings.mlr.press/v97/geifman19a.html>
8. International Organization for Standardization. (2022). Road vehicles—Safety of the intended functionality (ISO Standard No. 21448:2022). <https://www.iso.org/standard/77490.html>
9. Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361. DOI: 10.1109/CVPR.2012.6248074.