

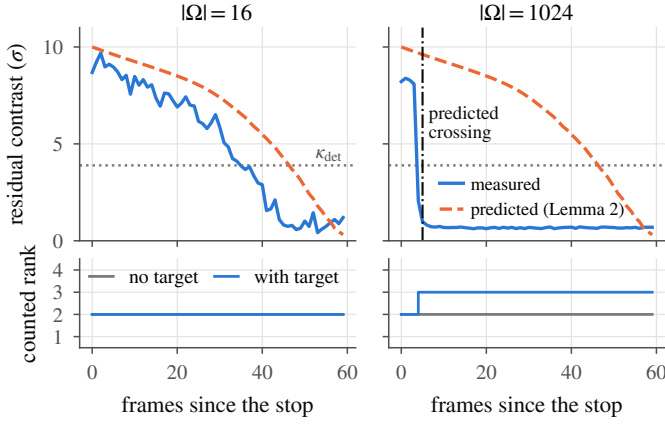


Fig. 2. The two mechanisms on the implemented estimator (synthetic, $f = 0.97$, contrast 10σ , soft cleaning, one run). Top: residual contrast of the stopped target and the ramp of Lemma 2 predicted from the clean background; bottom: counted rank. The small target ($|\Omega| = 16$) is absorbed by the ramp (below κ_{det} at frame 35; predicted 47); the large one ($|\Omega| = 1024$) when a component carrying it is counted, at the predicted crossing (frame 5).

Proof: With fixed regressors, (1) solves a weighted least-squares problem with information matrix R_t ; Sherman–Morrison gives the recursion and $z_t^\top (w_{p,t} - w_{p,t-1}) = h_t \iota_{t,p}$. In the model’s basis $R_{t-1} = S_r^2$, so $z_t^\top (f S_r^2 + z_t z_t^\top)^{-1} z_t = \eta_t / (f + \eta_t)$; for stationary scores $R_t \approx \mathbb{E}[z z^\top] / (1 - f)$. ■

The row-wise form is that of PAST and PETRELS [12], [13]; its consequence for absorption is new. The model can read the leverage at every frame, and the cleaning operator decides the innovation that drives the update.

Proposition 1 (Absorption law): A target of contrast c (noise units) stops on Ω ; let δ_t be its share in the fit. (a) *Leverage ramp:* by Lemma 2,

$$\delta_{t+1} = \delta_t + h_t \iota_t, \quad (3)$$

with $\iota_t = \kappa$ while $c - \delta_t > \kappa$ for soft thresholding at $\kappa\hat{\sigma}$, $\iota_t = c - \delta_t$ for raw learning and $\iota_t = 0$ for imputed or masked pixels; the target is absorbed when $c - \delta_t < \kappa_{\text{det}}$, independently of $|\Omega|$. (b) *Rank crossing:* a component carrying the target is counted once $\sum_k f^{t-1-k} |\Omega| \iota_k^2 + d \geq \tau / \hat{\sigma}^2$, where d is the noise energy of the uncounted slot that the leak joins. The absorption time is the earlier of the two.

Part (a) follows from Lemma 2; part (b) is an approximation whose noise term was calibrated on the estimator. Two corollaries contradict the accumulated-leak rule (crossing without the noise term): soft-threshold trackers absorb *every* stopped target in a time linear in contrast, and retention is a design parameter only for operators with zero innovation (Fig. 2).

Verification. On 4,320 synthetic episodes of the implemented estimator (truncated incremental SVD [14]; 10 seeds, $f \in \{0.90, \dots, 0.99\}$, $|\Omega|$ from 4 to 1024 cells, soft cleaning and imputation with release), we predicted each absorption time from the clean background alone (Fig. 3). On the 2,880 runs with contrast at least 10σ the median error is 2 frames (68 % within 3) and the median relative error 0.130 (95 % CI

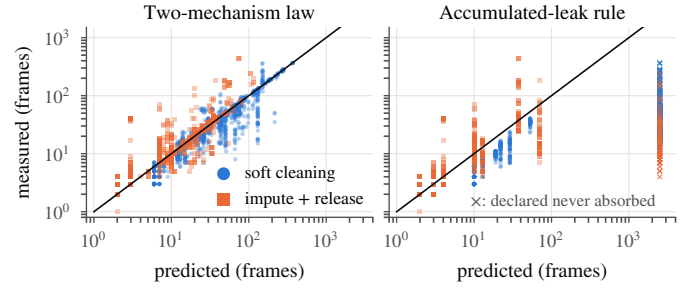


Fig. 3. Predicted vs. measured absorption time of the implemented estimator (2,880 synthetic runs, contrast $\geq 10\sigma$; frames after learning starts, plus one). Left: Proposition 1. Right: accumulated-leak rule; crosses mark runs it declares never absorbed.

[0.114, 0.143]); the crossing time has median error 0 frames. The accumulated-leak rule predicts no absorption in 1,200 absorbed runs. An exact sliding-window SVD absorbs like the truncated one (mean difference 0.2 frames, 108 pairs).

III. TRACKER IN THE LOOP

Masks. A tracker supplies a mask M of pixels not to be learned, and the coefficients are fitted on M^c . Let $x = Ua + \varphi + n$ with orthonormal U , noise of variance σ^2 and a target φ on Ω .

Proposition 2 (Masked coefficients): With $G = I - U_M^\top U_M$, eigenvalues ρ_i of $U_M^\top U_M$, leakage $\bar{\varphi} = \varphi_{\Omega \setminus M}$ and $\hat{a} = G^{-1} U_{M^c}^\top x_{M^c}$,

$$\mathbb{E} \|\hat{a} - a\|^2 = \sum_i \frac{\sigma^2}{1 - \rho_i} + \|G^{-1} U_{M^c}^\top \bar{\varphi}\|^2. \quad (4)$$

Imputed (masked) pixels have zero innovation: they are never absorbed, and their loadings freeze.

Proof: $U^\top U = I$ gives $\hat{a} - a = G^{-1} U_{M^c}^\top (n_{M^c} + \bar{\varphi})$; the noise term has covariance $\sigma^2 G^{-1} U_{M^c}^\top U_{M^c} G^{-1} = \sigma^2 G^{-1}$ and the cross term zero mean. Substituting $c_{t,p} = w_{p,t-1}^\top z_t$ in Lemma 2 makes the update vanish. ■

Monte Carlo confirms (4), and the deterministic bound $\|\hat{a} - a\| \leq (\|U_{M^c}^\top n_{M^c}\| + \|U_{M^c}^\top \bar{\varphi}\|) / (1 - \max_i \rho_i)$ held in all 24,000 trials; leakage costs far more than masking, so masks should err large.

Lock-in. Protecting a model error (a ghost, a light change) keeps it alive. Protection of an unverified region ends when an expiry statistic reaches the horizon T_p ; release then takes T_{rel} frames and learning D_{raw} . On CDnet *parking*, a clock driven by the tracker’s speed estimate kept a track on ground-truth background (0.29 to 0.67 of its box static) protected for at least 66 s against a horizon of 15 s: centroid jitter restarted the clock.

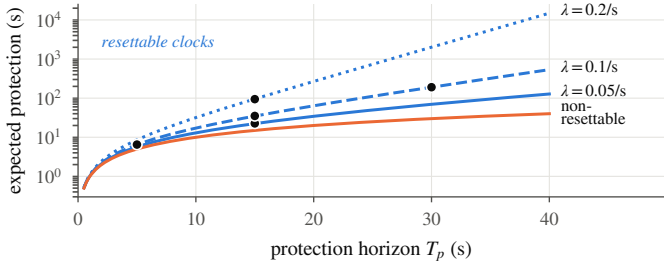


Fig. 4. Expected protection of a model error vs. horizon: resettable clocks (blue, Proposition 3(ii)), Poisson restart rate λ) grow exponentially, the non-resettable one (orange) linearly; dots: simulation at 30 frames per second.

Proposition 3 (Persistence): For a protected error that never travels: (i) if the expiry statistic grows at unit rate (non-resettable), persistence is at most $T_p + T_{\text{rel}} + D_{\text{raw}}$; (ii) if it restarts at the events of a Poisson process of rate λ (e.g. centroid jitter read as motion), the expected protection time is

$$m(T_p) = \frac{e^{\lambda T_p} - 1}{\lambda}; \quad (5)$$

(iii) with a verifier that confirms an error with probability β per look before each expiry, the expected persistence is at most $T_p/(1 - \beta) + T_{\text{rel}} + D_{\text{raw}}$.

Proof: (ii) Let m be the expected time to the first gap of length T_p . Conditioning on the first event $X \sim \text{Exp}(\lambda)$, $m = T_p e^{-\lambda T_p} + \mathbb{E}[(X + m)\mathbf{1}\{X < T_p\}]$, whence (5). (iii) The number of protection periods is geometric with parameter $1 - \beta$. (i) is immediate. ■

Proposition 4 (Horizon): With n_o stops of duration $D \sim F_D$ and n_e ghosts per unit time, and costs c_a per unit time a stopped object is unprotected and c_e per unit time a ghost is protected, $J(T) = c_a n_o \mathbb{E}(D - T)_+ + c_e n_e T$ is convex and minimised at

$$T^* = F_D^{-1}(1 - q), \quad q = \frac{c_e n_e}{c_a n_o}. \quad (6)$$

Proof: $J'(T) = -c_a n_o P(D > T) + c_e n_e$ and $J''(T) = c_a n_o f_D(T) \geq 0$. ■

In simulation, (ii) gives 94.3 s against the predicted 95.4 s at $\lambda = 0.2/\text{s}$ and $T_p = 15$ s (Fig. 4), and the empirical minimiser of J equals the quantile in 6 of 6 cases. Our system therefore measures stillness by displacement from an anchor, a clock that jitter cannot reset.

Policy. Confirmed tracks are protected (boxes enlarged by 20%). A pretrained detector (YOLOv10s [15], COCO [16], ONNX on one CPU core) verifies tracks under a token-bucket budget of half the frame period; verified tracks stay protected while they live, rejected ones lose protection at once, and tracks not yet looked at expire after T_p . In our runs no confirmed track stood still for more than 3.6 s before its first look, so the horizon never bound: the verifier turns the

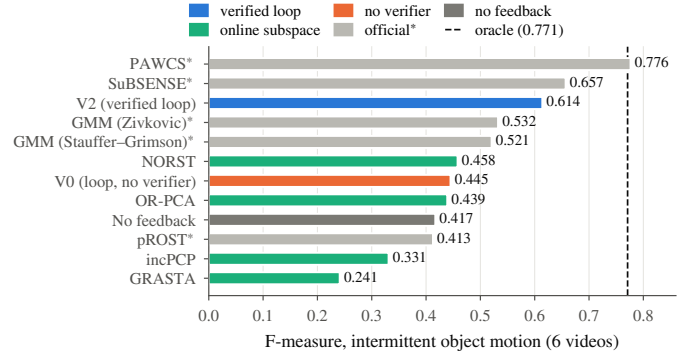


Fig. 5. IOM F-measure of every method. *Official changedetection.net results (accessed 23 October 2026); the online subspace baselines ran on our frames and grid.

horizon problem of Proposition 4 into a classification problem governed by Proposition 3(iii).

IV. EXPERIMENTS ON CDNET 2014

Protocol. Our evaluator reproduces the confusion matrices of the official CDnet MATLAB code exactly (10 of 10 videos checked). The front end (work grid, false-alarm rate, mask filter, stabilisation) was selected among 36 configurations on a development split of 25 videos (baseline, dynamic background, camera jitter, shadow, thermal; mean F-measure 0.526 to 0.614) and frozen. Under a protocol fixed before the runs, a policy must not be worse than V0 (feedback without verifier) on the development split (one-sided paired Wilcoxon test), and the intermittent-object-motion (IOM) result is leave-one-video-out. The 22 videos of the five categories added in 2014 were held out and evaluated once. Four online baselines ran streamed on the same frames, reduced to at most 20000 cells: GRASTA, NORST and OR-PCA from the authors' code, and incremental PCP as our re-implementation of [5], which reproduces its published F-measure on *highway* (0.763 vs. 0.7455). The sparsity weights of the last two were selected on the development split.

Results. On IOM (Fig. 5) the verified loop raises the mean F-measure from 0.445 (V0) and 0.417 (no feedback) to 0.614, closing 52% of the gap to oracle masks (0.771), and it exceeds the online subspace methods (NORST 0.458, OR-PCA 0.439, incremental PCP 0.331, GRASTA 0.241; official pROST 0.413) and mixture models (official 0.521–0.532), while SuBSENSE (0.657) and PAWCS (0.776) remain ahead. On the development split it also improves on V0 (0.614 to 0.651, two-sided paired Wilcoxon $p = 0.030$). Recall on pixels that have been foreground for 10–30 s rises from 0.04 without feedback to 0.66 (Fig. 7), while the false-positive rate on pixels vacated 10–30 s earlier rises from 0.007 to 0.013.

Failures. Two videos fail (Fig. 6). In abandonedBox, whose stopped box has no COCO class, only 19% of the pixels of objects stopped for 3 s or more lie in protection boxes (at least 56% in the other videos). In tramstop, feedback multiplies false positives by 10 relative to no feedback, and 88% of them lie inside protection boxes, where the model does not learn.

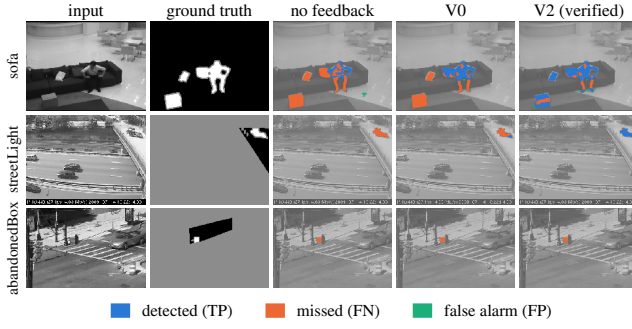


Fig. 6. Error maps at the frame with the most ground-truth pixels stopped for at least 3 s, chosen without looking at any output (frames 1344, 2153, 2787). In *sofa* and *streetLight* the verified loop keeps the stopped objects that the other settings absorb; in *abandonedBox* the box, outside the detector’s vocabulary, is absorbed by all three.

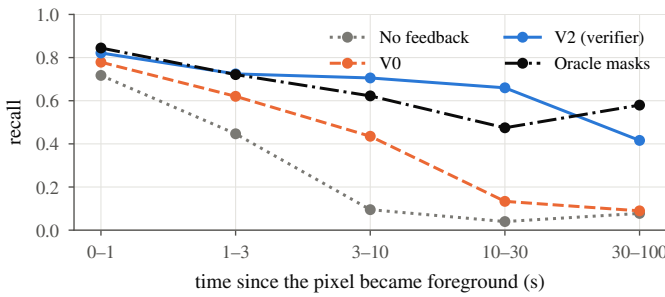


Fig. 7. Recall of ground-truth foreground pixels by the time since they became foreground (CDnet IOM, pooled over the 6 videos).

Held-out categories and cost. On the 22 videos added in 2014, evaluated once, the verified loop reaches 0.552 (V0: 0.526; no feedback: 0.546), on a par with no feedback, and 0.603 over all 11 categories (Table I). On one idle desktop core at 320×240 it takes 27.4 ms per frame on average (8.3 ms without the verifier; 95th percentile 231.9 ms, the synchronous verifier’s stall); the online baselines take 0.3 (incremental PCP) to 79 ms (NORST) at 160×120 .

V. CONCLUSION

Absorption in forgetting subspace background models is a per-pixel recursive least-squares process plus a rank-crossing event, and it can be predicted from the background alone. Tracker feedback removes it but must be designed as a closed loop: expiry clocks that errors can reset lock errors in, and a budgeted verifier both protects long stops and releases errors. Limitations: part (b) of Proposition 1 is a calibrated approximation; the verifier is a pretrained network, and comparisons should account for that; the real-data front end trails the best per-pixel methods; timings were measured on a desktop core, not an embedded board, and the synchronous verifier stalls the frames on which it runs. A random-matrix treatment of part (b) and an asynchronous verifier are the next steps.

REFERENCES

[1] J. He, L. Balzano, and A. Szlam, “Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1568–1575.

TABLE I
ALL 11 CDNET 2014 CATEGORIES, MEAN F-MEASURE. THE FIVE CATEGORIES ADDED IN 2014 WERE EVALUATED ONCE, AFTER FREEZING. NO FB.: NO FEEDBACK.

Category	V2	V0	No fb.	GRASTA	NORST	OR-PCA	incPCP
Baseline	0.748	0.687	0.611	0.610	0.759	0.765	0.681
Dyn. bg.	0.564	0.573	0.550	0.374	0.122	0.230	0.399
Jitter	0.656	0.640	0.616	0.489	0.694	0.732	0.515
Shadow	0.681	0.559	0.559	0.530	0.582	0.683	0.617
Thermal	0.606	0.610	0.485	0.306	0.617	0.560	0.391
IOM	0.614	0.445	0.417	0.241	0.458	0.439	0.331
Bad weather	0.746	0.680	0.687	0.393	0.381	0.660	0.566
Low fr.	0.613	0.611	0.606	0.399	0.506	0.499	0.432
Night	0.381	0.368	0.394	0.361	0.284	0.445	0.377
PTZ	0.242	0.210	0.270	0.211	0.125	0.218	0.156
Turbul.	0.777	0.759	0.770	0.699	0.125	0.265	0.646
Overall (11 cat.)	0.603	0.558	0.542	0.419	0.423	0.500	0.465

[2] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, “Recursive robust PCA or recursive sparse recovery in large but structured noise,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5007–5039, 2014.

[3] P. Narayanamurthy and N. Vaswani, “Nearly optimal robust subspace tracking,” in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, vol. 80, 2018, pp. 3701–3709.

[4] J. Feng, H. Xu, and S. Yan, “Online robust PCA via stochastic optimization,” in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013, pp. 404–412.

[5] P. Rodríguez and B. Wohlberg, “Incremental principal component pursuit for video background modeling,” *Journal of Mathematical Imaging and Vision*, vol. 55, no. 1, pp. 1–18, 2016.

[6] F. Seidel, C. Hage, and M. Kleinsteuber, “pROST: a smoothed ℓ_p -norm robust online subspace tracking method for background subtraction in video,” *Machine Vision and Applications*, vol. 25, no. 5, pp. 1227–1240, 2014.

[7] Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[8] M. Harville, “A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models,” in *Computer Vision – ECCV 2002*, ser. Lecture Notes in Computer Science, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds. Springer, 2002, vol. 2352, pp. 543–560.

[9] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.

[10] L. Taycher, J. W. Fisher III, and T. Darrell, “Incorporating object tracking feedback into background maintenance framework,” in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION)*, 2005, pp. 120–125.

[11] Y. Wang, P.-M. Jodoin, F. Porikil, J. Konrad, Y. Benezeth, and P. Ishwar, “CDnet 2014: An expanded change detection benchmark dataset,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 393–400.

[12] B. Yang, “Projection approximation subspace tracking,” *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[13] Y. Chi, Y. C. Eldar, and R. Calderbank, “PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5947–5959, 2013.

[14] M. Brand, “Incremental singular value decomposition of uncertain data with missing values,” in *Computer Vision – ECCV 2002*, ser. Lecture Notes in Computer Science. Springer, 2002, pp. 707–720.

[15] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, “YOLOv10: Real-time end-to-end object detection,” arXiv:2405.14458, 2024, verified on arXiv (v2) on 2026-10-24.

[16] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft COCO: Common objects in context,” arXiv:1405.0312, 2014, verified on arXiv (v3) on 2026-10-24.