FICTIONAL DEMONSTRATION · SYNTHETIC CC0 IMAGES · NOT CLIENT WORK, NOT EVIDENCE OF PERFORMANCE ON ANY REAL LINE

Longlot Vision
FIXED-STATION FEASIBILITY

CASE	CUSTOMER	STATION	ISSUED
DEMO-CAP-001	Fictional Bottling Co. (demo)	Line 2 · capper outfeed · fixed camera	2026-10-31

Vision Feasibility Review

The check: **Is the cap visibly present in the fixed view?** Can it be measured reliably from the recorded footage you provided, before anyone buys cameras or hardware? This report answers that question for one check at one fixed station. It supports a person's decision; it does not stop machines, reject product or make safety decisions.

RECOMMENDATION

✕ NO-GO

DECISION RULE, AGREED BEFORE TESTING

Missed absent cap (FN rate) ≤ 5.0% and false alarms (FP rate) ≤ 10.0%

Agreed by Fictional QA lead on 2026-10-31. GO only if, on held-out recordings, the 95% upper confidence bounds of BOTH the FN rate and the FP rate are within the agreed limits, every observed operating condition meets the limits on its own point estimate with enough examples to be tested, and the label definition passed its check. NO-GO if the overall point estimate already misses a limit, or any condition misses a limit even at its 95% lower bound. Anything else is MORE EVIDENCE: the data neither supports nor rules out the check.

MISSED FAILURES	FALSE ALARMS	HELD-OUT FRAMES	LABEL CHECK
1.3%	**8.1%**	**393**	✓ PASS
2 of 157 · 95% CI 0.4–4.5%	19 of 236 · 95% CI 5.2–12.2%	10 independent recording groups	kappa 0.94 · unsure 3.1%

WHY

- Condition glare=yes: FP rate 35.3% misses the limit even at its 95% lower bound (21.5%).
- Condition camera=shifted: FP rate 17.1% is above 10.0%.
- Condition glare=yes: FN rate 8.0% is above 5.0%.
- Condition lighting=bright: FN rate 5.6% is above 5.0%.
- Condition lighting=normal: FP rate 14.3% is above 10.0%.
- Condition variant=blue: FP rate 10.7% is above 10.0%.
- FP-rate 95% upper bound 12.2% is above 10.0%; even with zero false alarms this needs at least 35 good examples.

1 The question and the labels

CODE	LABEL	DEFINITION USED BY LABELLERS
P	Present	Present: the cap is visible and fully seated in the region of interest.
A	Absent	Absent: the cap is missing from the region of interest (the failure the customer wants flagged).
U	Unsure	Unsure: cannot be decided from the image (occluded, blurred, partly seated, out of view).

The **failure** the check must catch is **absent**. A **false negative (FN)** is a real failure the check would have missed; a **false positive (FP)** is a good part flagged for a person to look at. Unsure and disputed frames are counted but not scored. No single “accuracy” figure is given because it hides missed failures.

2 Dataset inventory

ITEM	VALUE
Source files	400 images, 1 video clips (clips sampled at 2 frames/s)
Frames inventoried	418
Recording sessions	12
Independent groups (after duplicate merge)	10 — merged: S01+S06, S02+S07
Resolutions	640x480 (418)
Date range (file times)	2026-10-31T11:36:43 → 2026-10-31T11:37:03
Exact duplicate copies	1
Near-duplicate pairs	7 within a session, 5 across sessions
Distinct-scene clusters	406 (effective sample is smaller than the frame count when frames repeat)
Image-quality flags	over exposed: 26
Unreadable files	0

Labels used: 236 present, 158 absent, 13 unsure, 11 disputed; 0 frames unlabelled. Scored after exclusions: **393** frames (157 absent, 236 good).

Class balance by operating condition

CONDITION	VALUE	PRESENT	ABSENT	UNSURE / DISPUTED
camera	nominal	195	131	20

CONDITION	VALUE	PRESENT	ABSENT	UNSURE / DISPUTED
camera	shifted	41	27	4
glare	no	202	133	18
glare	yes	34	25	6
lighting	bright	66	36	6
lighting	dim	44	30	2
lighting	normal	126	92	16
variant	blue	103	69	11
variant	red	133	89	13

3 Label-definition check ✓ PASS

Two people labelled the same frames independently: **customer_qa** and **operator** on 418 frames.

MEASURE	RESULT
Raw agreement	94.3%
Cohen's kappa, present vs absent	0.94 (strong)
Cohen's kappa, incl. unsure	0.88 (strong)
Consensus unsure rate	3.1%
Disputed (excluded)	11
Near-duplicates with conflicting labels	0

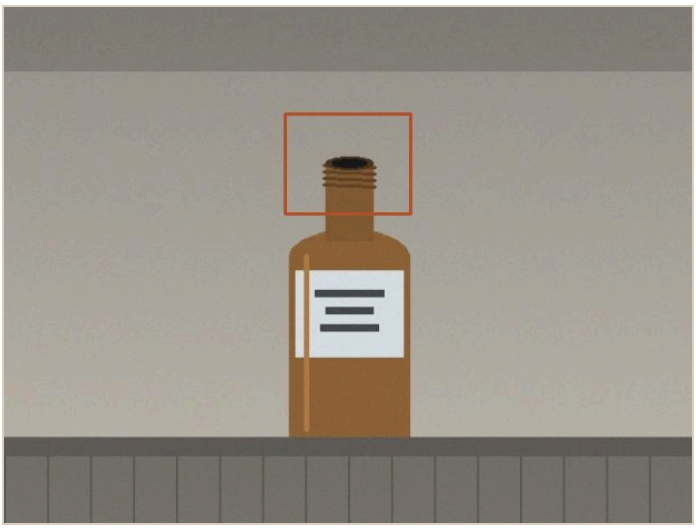
Rater agreement matrix

CUSTOMER_QA ↓ / OPERATOR →	P	A	U
P	236	3	3
A	8	158	1
U	6	3	0

> 11 frame(s) disputed between raters are excluded from evaluation.

Targets: kappa ≥ 0.8, unsure rate ≤ 5%.

4 How it was tested



Reference frame; the outlined box is the fixed region of interest (116×92 px). Only this box is analysed.

1. **Crop** the fixed region of interest from every labelled frame.
2. **Split by recording**: frames are grouped by *session*; sessions sharing near-duplicate frames are merged, so no recording appears in both training and test (5 folds).
3. **Set the threshold on validation only**: inside each training part, a further grouped split produces validation scores; the threshold is chosen there (minimise FN rate / FN limit + FP rate / FP limit on validation scores (a miss weighs FP-limit / FN-limit times a false alarm); ties go to the lower threshold). Held-out recordings never influence it.
4. **Score held-out recordings once** and pool the results. Intervals are Wilson 95%.
5. **Stress-test** the primary baseline with brightness, shift, blur and compression changes.

Primary baseline (declared before testing): **hog_color_logreg**.

5 Results on held-out recordings

Primary baseline: confusion table

	Flagged for review	Not flagged
Real failure (absent)	**TP 155**	**FN 2** missed
Good part	**FP 19** false alarm	**TN 217**

Precision 89.1% (CI 83.6–92.9%) · recall 98.7%.

All baselines

BASELINE	TP/FP/TN/FN	FN RATE	FP RATE
hog_color_linsvm HOG shape + HSV colour histograms, linear SVM (class-balanced)	156/14/222/1	0.6% 0.1–3.5%	5.9% 3.6–9.7%
hog_color_logreg PRIMARY HOG shape + HSV colour histograms, logistic regression (class-balanced)	155/19/217/2	1.3% 0.4–4.5%	8.1% 5.2–12.2%
template_ncc Mean present/absent templates, normalised	155/59/177/2	1.3% 0.4–4.5%	25.0% 19.9–30.9%

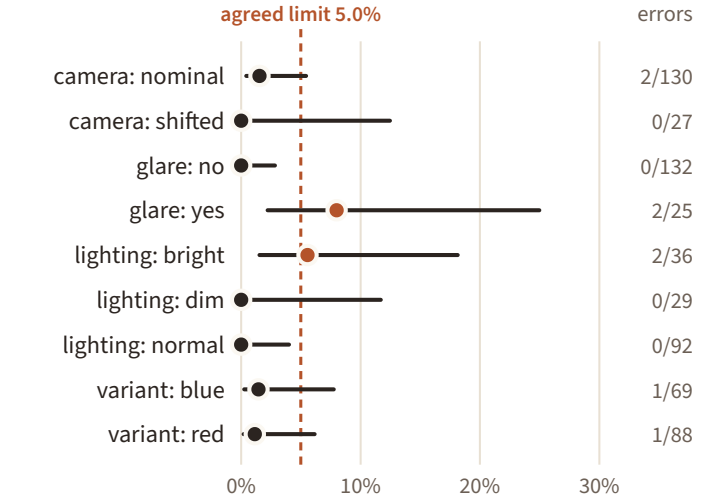
BASELINE	TP/FP/TN/FN	FN RATE	FP RATE
cross-correlation			
difference			

Comparison baselines cannot be promoted to a GO on this same data; a better comparison is a reason to re-test, not to switch.

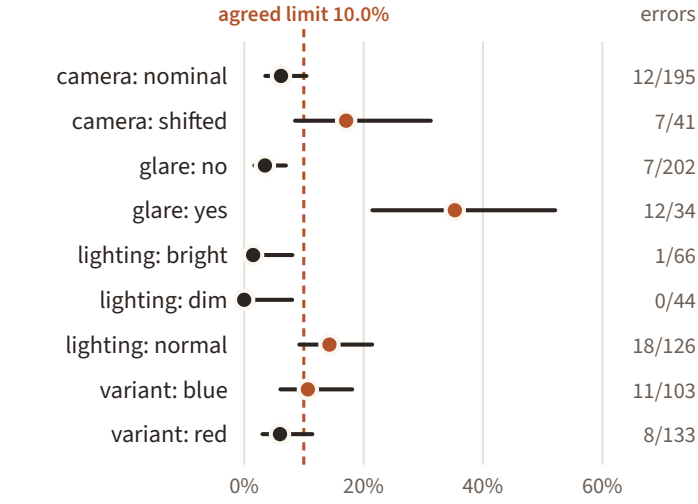
By operating condition (primary baseline)

CONDITION	VALUE	N	MISSED	FN RATE	95% CI	FALSE ALARMS	FP RATE	95% CI	STATUS
camera	nominal	325	2/130	1.5%	0.4–5.4%	12/195	6.2%	3.6–10.4%	✓ MEETS
camera	shifted	68	0/27	0.0%	0.0–12.5%	7/41	17.1%	8.5–31.3%	✕ MISSES
glare	no	334	0/132	0.0%	0.0–2.8%	7/202	3.5%	1.7–7.0%	✓ MEETS
glare	yes	59	2/25	8.0%	2.2–25.0%	12/34	35.3%	21.5–52.1%	✕ MISSES
lighting	bright	102	2/36	5.6%	1.5–18.1%	1/66	1.5%	0.3–8.1%	✕ MISSES
lighting	dim	73	0/29	0.0%	0.0–11.7%	0/44	0.0%	0.0–8.0%	✓ MEETS
lighting	normal	218	0/92	0.0%	0.0–4.0%	18/126	14.3%	9.2–21.5%	✕ MISSES
variant	blue	172	1/69	1.4%	0.3–7.8%	11/103	10.7%	6.1–18.1%	✕ MISSES
variant	red	221	1/88	1.1%	0.2–6.2%	8/133	6.0%	3.1–11.4%	✓ MEETS

Missed failures (FN rate) with 95% CI



False alarms (FP rate) with 95% CI



"Untested" means fewer than 10 failure or 10 good examples in that condition: it neither passes nor fails.

Further slices: image quality, glare inside the ROI, recording session

SLICE	VALUE	N	MISSED	FN RATE	FALSE ALARMS	FP RATE
image quality	ok	368	0/146	0.0%	18/222	8.1%
image quality	over_exposed	25	2/11	18.2%	1/14	7.1%

SLICE	VALUE	N	MISSED	FN RATE	FALSE ALARMS	FP RATE
roi glare	clipped > 3% of ROI	59	2/25	8.0%	12/34	35.3%
roi glare	not clipped	334	0/132	0.0%	7/202	3.5%
session	S01	33	0/17	0.0%	0/16	0.0%
session	S02	31	0/13	0.0%	0/18	0.0%
session	S03	35	0/12	0.0%	0/23	0.0%
session	S04	34	1/12	8.3%	0/22	0.0%
session	S05	33	0/15	0.0%	3/18	16.7%
session	S06	34	0/12	0.0%	4/22	18.2%
session	S07	38	0/17	0.0%	0/21	0.0%
session	S08	33	0/11	0.0%	0/22	0.0%
session	S09	35	0/12	0.0%	4/23	17.4%
session	S10	34	0/15	0.0%	7/19	36.8%
session	S11	35	1/13	7.7%	1/22	4.5%
session	S12	18	0/8	0.0%	0/10	0.0%

6 Stress tests

Held-out crops were altered and re-scored with the same model and threshold. A test is marked **sensitive** if it pushes a rate above the agreed limit or more than 5 points above the unaltered result.

CHANGE	FN RATE	Δ	FP RATE	Δ	
Unaltered	1.3%		8.1%		✓ STABLE
brightness -30%	0.6%	-0.6	7.6%	-0.4	✓ STABLE
brightness +30%	0.6%	-0.6	6.4%	-1.7	✓ STABLE
shift left 6 px (~9% of ROI)	0.6%	-0.6	16.5%	**+8.5**	✕ SENSITIVE
shift right 6 px	0.0%	-1.3	8.1%	+0.0	✓ STABLE
shift up 6 px	1.3%	+0.0	23.3%	**+15.3**	✕ SENSITIVE
shift down 6 px	4.5%	**+3.2**	9.7%	**+1.7**	✓ STABLE
blur sigma 1.5 px	1.3%	+0.0	25.4%	**+17.4**	✕ SENSITIVE
blur sigma 3 px	8.3%	**+7.0**	39.0%	**+30.9**	✕ SENSITIVE
JPEG quality 30	1.3%	+0.0	5.9%	-2.1	✓ STABLE

7 Error gallery

Missed failures (false negatives) 2

Real failures the check did not flag; the costliest error. Most confident errors first.



8ddc4854 · S04
nominal, yes, bright, blue over_exposed
margin -0.036



a5856ad2 · S11
nominal, yes, bright, red over_exposed
margin -0.007

False alarms (false positives) 19

Good parts the check would send to a person. Most confident errors first.



a1164e3e · S10
nominal, yes, normal, blue
margin +0.680



d7441ec4 · S06
nominal, yes, normal, red
margin +0.378



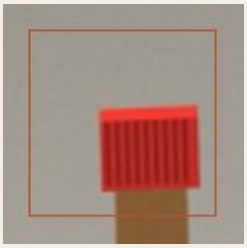
e97c963a · S06
nominal, yes, normal, red
margin +0.298



0d125eb2 · S10
nominal, yes, normal, blue
margin +0.222



b883d713 · S10
nominal, yes, normal, blue
margin +0.134



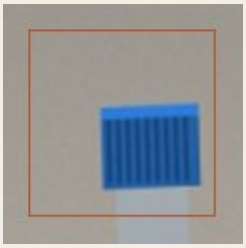
7ba2769b · S05
shifted, no, normal, red
margin +0.109



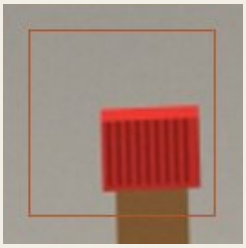
20e4451c · S06
nominal, yes, normal, red
margin +0.098



f1f50b0f · S10
nominal, yes, normal, blue
margin +0.097



c377dfa6 · S09
shifted, no, normal, blue
margin +0.088



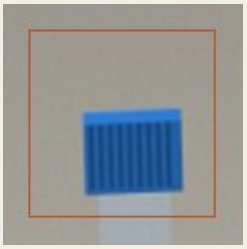
0a24fedc · S05
shifted, no, normal, red
margin +0.087



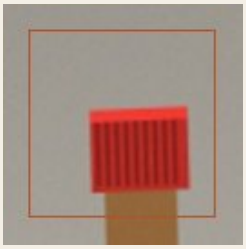
1ec4d981 · S10
nominal, yes, normal, blue
margin +0.042



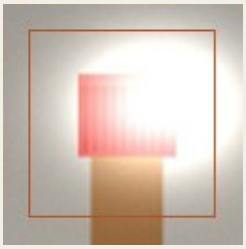
2812f1f4 · S11
nominal, yes, bright, red over_exposed
margin +0.034



eb08dbcf · S09
shifted, no, normal, blue
margin +0.032



efef4855 · S05
shifted, no, normal, red
margin +0.029



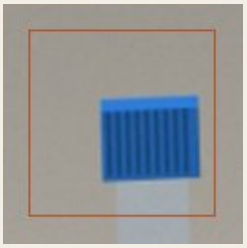
795d4e3a · S06
nominal, yes, normal, red
margin +0.025



5e661411 · S10
nominal, yes, normal, blue
margin +0.018



6beaa548 · S10
nominal, yes, normal, blue
margin +0.009



85c0349c · S09
shifted, no, normal, blue
margin +0.003



18727eff · S09
shifted, no, normal, blue
margin +0.001

8 Environmental and site risks

- **camera = shifted** misses the limits: missed 0.0%, false alarms 17.1%.
- **glare = yes** misses the limits: missed 8.0%, false alarms 35.3%.
- **lighting = bright** misses the limits: missed 5.6%, false alarms 1.5%.
- **lighting = normal** misses the limits: missed 0.0%, false alarms 14.3%.
- **variant = blue** misses the limits: missed 1.4%, false alarms 10.7%.
- **Glare inside the region of interest** (59 frames with saturated pixels) drives errors: FN 8.0%, FP 35.3%.
- **Sensitive to shift left 6 px (~9% of ROI)**: FN 0.6%, FP 16.5%. Camera mount and ROI must be fixed and re-checked after any bump or maintenance.
- **Sensitive to shift up 6 px**: FN 1.3%, FP 23.3%. Camera mount and ROI must be fixed and re-checked after any bump or maintenance.
- **Sensitive to blur sigma 1.5 px**: FN 1.3%, FP 25.4%. Focus, line speed / shutter time and compression settings matter.
- **Sensitive to blur sigma 3 px**: FN 8.3%, FP 39.0%. Focus, line speed / shutter time and compression settings matter.
- Conditions that miss the limits must be controlled physically (shielding, diffuse lighting, fixed exposure, a fixed mount) or excluded from scope before any prototype; model tuning alone is not a fix.
- Image-quality flags in the footage: 26 over exposed.
- Some sessions share near-identical frames and were merged for testing; the footage contains repeats, so it is less varied than its frame count suggests.

- Conditions not present in the footage (other shifts, seasons, product colours, a cleaned or dirty lens) are untested and must be added before a prototype is accepted.
- Any change to the camera, lens, mount, lighting or product after this review invalidates these results until a Reassessment.

9 Privacy and data handling

ITEM	RECORD
Authorization reference	DEMO — synthetic CC0, no authorization needed
Received	—
Storage	Local encrypted disk (BitLocker) on the operator laptop; no cloud copy.
Named access	Demo operator only
Retention until	2026-11-30
At end of retention	Secure deletion with a signed deletion certificate sent to you
Deletion confirmation	Pending: local copies will be deleted by 2026-11-30; a signed deletion certificate will be sent to you.

- All processing ran offline on one local laptop. No cloud service, no telemetry, no live camera or network connection to your site.
- Only the fixed region of interest was analysed. No faces, identities, biometrics, emotions, productivity or any person-level measure were processed or produced.
- This report embeds small crops of the region of interest and one reference frame. Treat it with the same care as the footage.
- You remain responsible for the legal authority, notice and workplace obligations for the footage. This review is not legal certification.

10 Limitations

- Results describe *this footage* under the conditions it contains. Lighting, camera position, product or packaging changes not in the footage are untested.
- Baselines are deliberately simple and run on a laptop CPU. A prototype may do better or worse; nothing here is a performance guarantee.
- Labels define the truth. If the labels are inconsistent, every number inherits that inconsistency.
- To show FN ≤ 5.0% with 95% confidence needs at least 73 failure examples even with zero misses; FP ≤ 10.0% needs at least 35 good examples.
- Recorded footage is not a live line: frame rate, motion blur at line speed, and camera exposure behaviour must be re-checked on site.

- Decision support only. Any future system must fail to **REVIEW REQUIRED**, never silently pass an uncertain part.

11 Next-step options

Prototype scoping

Quoted separately · no table price

Only after a GO. A fixed-scope proposal for one station: camera, lighting and compute sized from this evidence, a written acceptance metric, and PASS / FAIL / REVIEW REQUIRED outputs for a person. No machine stop or auto-reject.

Expanded Evidence Review

$950 CAD [A] · 7 business days

Adds one approved product variation or operating condition, up to the same sample size, and returns a revised error table and recommendation. Best when the answer was MORE EVIDENCE because examples were too few.

RECOMMENDED

Reassessment

$650 CAD [A] · 5 business days

One rerun after you change lighting, camera view or the label definition. Revised error table and go/no-go note. Best after a NO-GO caused by a fixable site condition.

Prices in CAD, excluding HST, 100% up front. [A] = current planning price, confirmed in writing before any work. The $1,250 review fee is credited against a prototype started within 30 days.

12 45-minute walkthrough

MINUTES	TOPIC
0–5	Recommendation and the decision rule
5–15	What footage was used, label definition and agreement
15–30	Results by condition, the error gallery, stress tests
30–40	Site risks, privacy, retention and deletion date
40–45	Next step and who owns it

Appendix · traceability

ITEM	VALUE
Tool	longlot-vision 0.1.0 · Python 3.13.18 · OpenCV 5.0.0 · scikit-learn 1.9.1
Results generated	2026-10-31T11:37:16-04:00
Random seed	20261031
Region of interest	x=262.0, y=100.0, w=116.0, h=92.0 on 640×480 (set 2026-10-31T11:36:58-04:00)
Signed manifest	trace_manifest.json (SHA-256 of every input and output, HMAC-SHA256 signed). Verify with `longlot-vision verify`.
Companion files	acceptance_matrix.md · dataset_summary.csv · predictions_*.csv