Marker-based Augmented Reality

Basic Computer Vision | Wang Zhaofeng | 04/11/2026

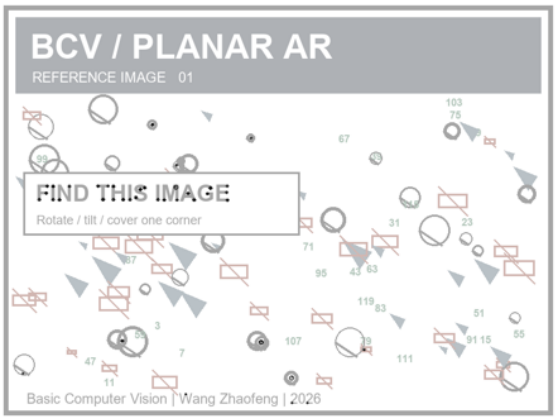
The program detects a planar reference image in a webcam frame and replaces it with another image. It uses local features, not a coded-marker detector. The tests include generated scenes and a real phone display.

1. Method and choices

Block	Implementation and reason
Descriptor extraction	SIFT uses scale-selected, oriented gradient patches [1]. It is the default because it retained more useful matches in the webcam test. ORB uses oriented binary tests [2] and provides a faster comparison. A feature budget of 1800 limits work; it is not a tuned optimum.
Matching	Brute-force kNN, k=2: the second neighbour provides an ambiguity check. L2 fits SIFT vectors; Hamming fits ORB bits [3]. Reference descriptors are cached.
Filtering	Keep d1 < r*d2, then allow each target keypoint only once. RANSAC removes pairs inconsistent with a common plane. This combines descriptor and geometric evidence.
Alignment	A homography models a plane under perspective, unlike an affine transform. At least four non-degenerate pairs are needed [4]; extra inliers and spatial spread make the estimate less fragile. Implausible outlines are rejected.
Augmentation	warpPerspective maps the substitute to the detected plane. A warped mask restricts blending to that region. Rejected frames keep their original pixels.

Settings: ratio 0.72 (SIFT), 0.78 (ORB); RANSAC 3 px; at least 10 inliers, inlier ratio 0.35 and reference coverage 3%. These were practical starting values. Section 2 checks their sensitivity after the original tests.

Reference: strongest 60 SIFT points



Real frame: strongest 60 inside marker

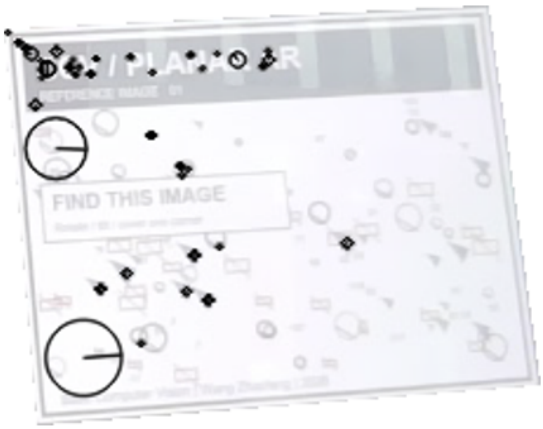


Figure 1. drawKeypoints with DRAW_RICH_KEYPOINTS. Circles show scale and ticks show orientation. The strongest 60 points are shown; artwork is faded for readability. Real-frame detection used 1800 as its feature budget.

2. Parameter sensitivity

A later one-factor-at-a-time study used seeds 300 and 301, separate from the original seeds 100-102. Each setting had 18 detailed-marker positives, 18 low-detail positives and four absence trials per method (two backgrounds paired with both references). Other settings stayed fixed. There were 1920 parameter evaluations, with descriptors cached for 40 marker/scene cases.

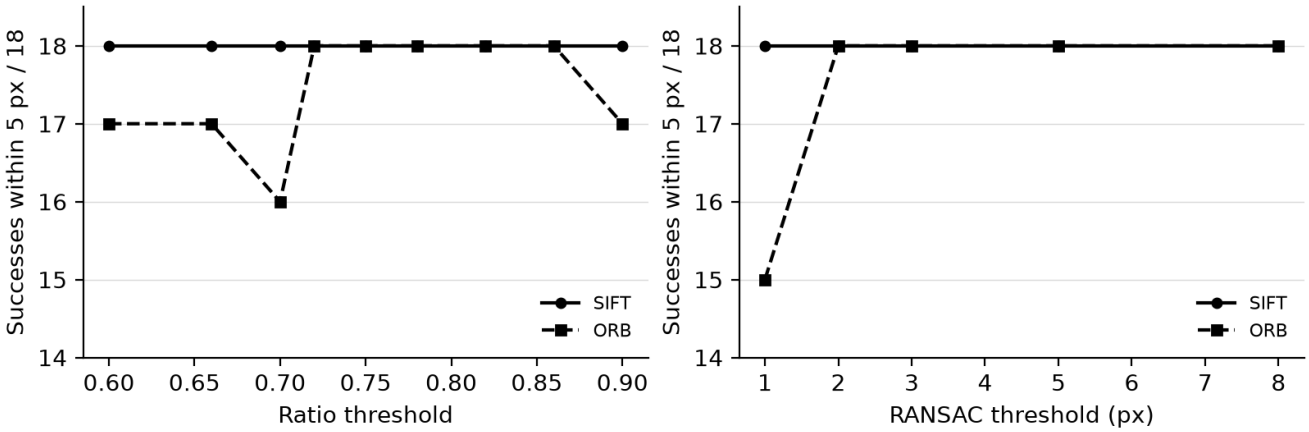


Figure 2. Success requires an accepted homography with mean corner error <=5 px. Each point uses the same 18 detailed-marker images. This small synthetic study does not measure generalization.

Parameter	Values tested	Observed result
Ratio r	0.60, 0.66, 0.70, 0.72, 0.75, 0.78, 0.82, 0.86, 0.90	SIFT: 18/18 throughout. ORB: 16-17/18 at 0.60-0.70 and 0.90; 18/18 at 0.72-0.86.
RANSAC distance	1, 2, 3, 5, 8 px	SIFT: 18/18 throughout. ORB: 15/18 at 1 px; 18/18 at 2-8 px.
Minimum inliers	4, 6, 10, 15, 20	Both methods: 18/18 at every value.
Minimum coverage	0%, 1%, 3%, 6%, 10%	Both methods: 18/18 at every value.

The original ratios, 0.72 and 0.78, both lie in a stable part of this scan. Distance distributions differ between descriptors, so their ratio cutoffs need not match. These data do not show that different cutoffs are necessary or that either value is best. They are retained to keep the original comparison reproducible.

A 3 px RANSAC threshold tolerates small feature-localization errors while limiting geometric disagreement. The scan shows that 1 px is too strict for some ORB cases; 2-8 px all pass. ORB mean corner error at 3 px was 1.40 px, versus 0.89 px at 5 px. Thus 3 px is a conservative operating choice, not the measured optimum.

Ten inliers exceeds the four-pair minimum. Coverage discourages fitting a tiny cluster then extrapolating across the marker. The scan cannot distinguish these guards: changing them had no effect here. All low-detail cases and all four negatives per setting were rejected; this does not establish which guard prevents false detections. The 0.35 inlier-ratio check was held fixed and remains heuristic.

The cached evaluator was checked against the production pipeline on 12 baseline combinations, including normal, perspective and absent scenes. Original results and timestamps were kept unchanged.

3. Original synthetic and webcam results

The original synthetic set contains 2 markers x 2 methods x 10 conditions x 3 trials = 120 runs at 800 x 600. There are nine positive conditions and one absence condition. Known transformed corners provide ground truth. Success means acceptance and mean four-corner error <=5 px.

Synthetic condition	SIFT pass	Mean px	ORB pass	Mean px
normal	3/3	0.13	3/3	0.62
rotation	3/3	0.20	3/3	0.56
perspective	3/3	0.49	3/3	3.50
dim	3/3	0.13	3/3	0.75
shadow	3/3	0.14	3/3	0.61
blur	3/3	0.56	3/3	2.14
occlusion 25	3/3	0.12	3/3	0.62
occlusion 50	3/3	0.38	3/3	1.38
clutter	3/3	0.14	3/3	0.60

Each method passes all 27 detailed-marker positives. Neither passes any of the 27 grey-rectangle positives: the system refuses these feature-poor examples. Across both marker types and methods, the 12 absence runs produce no false acceptance. SIFT has lower corner error in the perspective and blur cases.

Real phone test

The initial frontal run accepted 203/203 frames over 12.5 s. A guided handheld recording then tested harder views at 640 x 480. Both methods replayed the same every-fourth-frame sample. Different capture rates produce different sample counts per timed stage. Actions were qualitative, not calibrated angles or occlusion percentages.

Real condition	SIFT accepted	ORB accepted	SIFT ms	ORB ms
Frontal	23/23	22/23	159.4	61.7
Tilt	20/22	14/22	134.8	45.0
Rotation	21/23	8/23	119.1	49.1
Dim screen	3/29	0/29	80.4	30.7
Cover step 1	14/26	0/26	94.6	43.4
Cover step 2	23/23	0/23	105.0	30.6
Changed background	19/24	3/24	123.7	45.6
Phone absent	0/21	0/21	96.1	47.1

Time is median processing time per sampled frame, excluding drawing and display, from one replay. ORB is faster; SIFT keeps more difficult views. Acceptance is a gate decision, not a true-location score. Phone removal before 79 s is a transition, excluded from the reviewed 21-frame absence segment.

4. Failure analysis and real corner checks

Replay diagnostics separate feature detection from useful correspondence. Values below are stage medians; keypoint counts cover the whole frame, including the background.

Stage	SIFT keypoints / matches / inliers	ORB keypoints / matches / inliers
Frontal	882 / 111 / 100	1800 / 45 / 29
Rotation	927 / 48 / 35	1800 / 19 / 5
Dim screen	547 / 10 / 4	1772 / 12 / 4
Cover step 2	673 / 41 / 32	1800 / 12 / 5

Dimming reduces SIFT keypoints and good matches. Its 26 rejected dim frames fail because filtered matches are insufficient (14) or RANSAC inliers are insufficient (12). ORB still finds many points, but only about four form an inlier set. More detected points therefore do not guarantee a usable pose.

Under rotation, 14 of ORB's 15 rejections have too few inliers. In cover step 2, SIFT retains 32 median inliers while ORB retains five. SIFT's scale-normalized gradient patches and ORB's steered binary tests respond differently to appearance changes [1,2]. The measured bottleneck is correspondence consistency; this sequence cannot isolate orientation estimation, blur, reflections or viewpoint as the single cause.

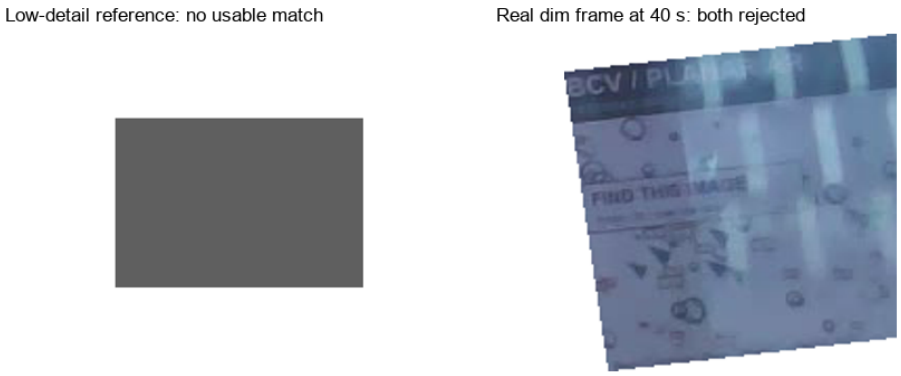


Figure 3. Rejected examples. Left: low-detail reference. Right: inner artwork crop from a dim real frame at 40 s. This crop is manually bounded for privacy; it is not a detected outline.

The synthetic-to-real gap is substantial: both methods pass synthetic dimming and occlusion, while the phone recording has many losses. Synthetic warps reuse the exact reference pixels and omit screen reflections, changing focus and motion. Cover step 2 also differs in distance and hand placement from step 1; its better SIFT count does not mean that more occlusion helps.

Independent corner check

Five times were fixed before inspecting estimates: 8, 11, 18, 21 and 29 s. The nearest scored frames were viewed with enlarged pixel grids. The last two have clipped artwork corners and were excluded. The assistant visually annotated the four outer artwork corners in the other three frames, independently of either homography.

Requested time	SIFT mean / max px	ORB mean / max px
8 s	1.51 / 2.23	2.66 / 6.10
11 s	0.83 / 1.39	2.53 / 4.31
18 s	1.65 / 3.10	2.14 / 3.16

Both methods accept these frames. Averaging all 12 corner distances gives 1.33 px for SIFT and 2.44 px for ORB. A single visual annotation has roughly 2 px uncertainty, so these are approximate checks on three easy views, not a precise ranking or a whole-video accuracy claim. Coordinates and per-corner distances are supplied.

5. Blending, interface and delivery

Let S be the substitute warped by H and T the current frame. Inside the marker mask, the course formula is:

fusion = (1 - alpha) * H(source) + alpha * target

The GUI uses source opacity beta: fusion = beta*S + (1-beta)*T. Thus beta = 1-alpha. Opacity 1 gives full replacement (course alpha=0); opacity 0.5 mixes equal parts. Outside the mask, the target is unchanged.

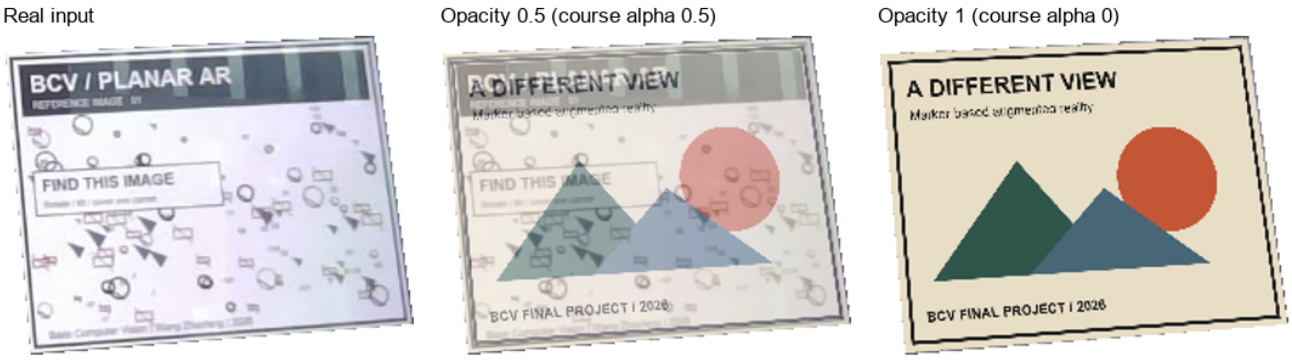


Figure 4. The same recorded frame with no overlay, opacity 0.5 and opacity 1. All pixels outside the manually bounded artwork are whitened. The source is the sample substitute image.

Interface

The delivered Tk GUI accepts a webcam or video file, switches between SIFT and ORB, and adjusts the feature budget, ratio test, RANSAC threshold, minimum inliers and opacity. It shows input/outline or rich keypoints, augmentation, and filtered inlier matches using drawMatches. Evidence can be saved locally. Capture runs separately so a stalled camera can be stopped.

The current feature slider supports 1800 (step 100). An earlier demo screenshot showed 1750 because its old slider used steps of 250; it is not the setting used in the reported experiments. Figures 1 and 4 were generated offline from saved frames with a budget of 1800.

Limits

There is no temporal tracking or smoothing. If registration is rejected, the overlay disappears. Replacement covers the full marker, including an occluding hand; foreground-aware rendering is not implemented. Optional residual alignment refinement is omitted. The small synthetic and handheld tests do not establish general reliability or a jitter score.

Delivered files

The bundle includes the Python application, evaluation scripts and CSV results, assets/marker.png and assets/substitute.png. Run ar_app.py in the opencv environment; requirements.txt lists dependencies. The teacher utility is optional for image reading; replacement is implemented in ar_core.py. Raw camera frames remain private and are excluded from the bundle and blog.

References

[1] OpenCV: Introduction to SIFT

[2] OpenCV: ORB (Oriented FAST and Rotated BRIEF)

[3] OpenCV: Feature Matching

[4] OpenCV: Feature Matching + Homography to find Objects

Course source: Final Project, Marker-based augmented reality, Nicolas Rougon, 2026. Tested with Python 3.11.16 and OpenCV 5.0.0. The linked 4.12 tutorials match the version specified in the assignment.