INTERNATIONAL JOURNAL OF TECHNOLOGY AND ACADEMIC RESEARCH

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

A Comparative Review of Geometry-Aware Generative Panoramic Reconstruction in Dynamic Environments

Murodjon Abdukholikov
Department of Applied Informatics, Kimyo International University in Tashkent,
[m.abdukholiqov@gmail.com]
doi:https://doi.org/10.5281/zenodo.23267813

Abstract—Panoramic reconstruction has moved from estimating a single planar transformation between photographs toward learning flexible warps, recovering explicit 3D scene structure, and synthesizing content that no camera observed. This review traces that progression through three methods chosen because each represents a distinct problem formulation: a lightweight homography-and-seam pipeline with explicit moving-object masking (our earlier work), the learned parallax-tolerant warping and composition of UDIS++, and PIS3R, which reprojects a feed-forward 3D reconstruction and refines it with point-conditioned diffusion. Rather than summarizing each paper, we compare them along technical problems: geometric alignment, dynamic objects, large parallax, missing content, computational cost, and temporal consistency. Original experiments on synthetic and real images make two distinctions concrete: depth-induced misalignment versus genuine object motion, and warping observed pixels versus generating unobserved ones. Each generation of methods resolves a different component of the problem, but none jointly models static geometry, independent motion, and generative completion. We close by proposing, as an untested research hypothesis, a geometry- and motion-conditioned generative framework governed by one principle: generate only where reliable observational evidence is insufficient.

Index Terms—image stitching, panoramic reconstruction, parallax, dynamic scenes, 3D reconstruction, diffusion models.

1. Introduction

A panorama is assembled from views captured at different camera positions and, in practice, at different moments. Two physical effects follow. When the camera centre translates, scene points at different depths shift by different amounts between views; this parallax means that no single planar transformation can align the whole overlap. When the scene changes between exposures, the same location may legitimately show different content in each view. The classical pipeline of feature matching, homography estimation and blending [1] assumes that neither effect is significant, and its characteristic failure, ghosting, can arise from either one. Both effects are common in current applications, where images are increasingly captured by mobile and aerial platforms, such as unmanned aerial vehicles (UAVs) used for forest-fire monitoring [20], [21] and cargo delivery [22], and the camera both rotates and translates between frames.

Research has responded in stages. Seam- and mask-based methods avoid compositing inconsistent content [2]–[4]; spatially varying warps [5], [6] and learned deformations [7], [8] reduce misalignment; feed-forward 3D reconstruction [9] now allows images to be fused by reprojection [10]; and diffusion models [11], [12] can supply content that no view observed [10], [13]. These stages are often read as gains in visual quality; we read them as successive changes in what the stitching problem is taken to be: selecting pixels, deforming images, reconstructing a scene, and synthesizing one (Fig. 1).

This paper is a focused comparative review rather than a survey. Three methods are analysed in depth, each representing one formulation: the dynamic-object-aware stitching method

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

of our earlier work [4] (classical geometry with explicit dynamic handling; hereafter DOS), UDIS++ [8] (learned parallax-tolerant alignment), and PIS3R [10] (3D reconstruction with diffusion refinement). The generative stitching method of Tuli et al. [13] serves as a complementary reference point. Throughout, we separate what the original papers report from our interpretation and from the research direction we propose.

Our contributions are: (i) a comparison organized by technical problem rather than by paper; (ii) two distinctions, made concrete by original experiments on synthetic and real images (Figs. 2 and 3): misalignment versus object motion, and warping versus generation; and (iii) a research-gap statement with a conceptual, not yet validated, framework for dynamic panoramas.

2. Evolution of Panoramic Reconstruction

Classical stitching [1] matches local features, estimates a homography H robustly with RANSAC, and blends the warped images. For homogeneous image coordinates $\mathbf{x}$ and $\mathbf{x}'$,

$$\mathbf{x}' \simeq H\,\mathbf{x}, \qquad (1)$$

which is exact when the camera rotates about its optical centre or the scene is planar.

A first line of work kept this model but stopped compositing pixels that disagree. Davis [2] partitioned the mosaic into regions drawn from single sources to avoid ghosts; Mills and Dudek [3] selected seams with Dijkstra's algorithm around moving content; DOS [4] combined a difference-based moving-object mask with a dynamic-programming seam. Gao et al. [14] chose among candidate homographies by the seam each admits. The shared idea is to avoid inconsistent content rather than to explain it.

A second line relaxed the geometric model. APAP [5] estimated spatially varying homographies with moving DLT, and Zhang and Liu [6] combined a homography with content-preserving warps to obtain alignment good enough for a seam. Deep methods then learned this step, following the broader shift of visual and perceptual computing toward learned representations, as seen in face detection and recognition [23], facial and speech emotion recognition [24], [25], and medical image analysis and augmentation [26], [27]: UDIS [7] introduced unsupervised deep homography estimation followed by reconstruction from stitched features to images, and UDIS++ [8] extended the warp to local thin-plate-spline (TPS) motion.

A third line replaces the warp with an explicit scene. Feed-forward networks such as VGGT [9] predict cameras, depth and point maps directly from a few images, and PIS3R [10] stitches by reprojecting such a reconstruction. Radiance-field representations such as 3D Gaussian splatting [15] offer a related route via per-scene optimization from many posed views; dynamic extensions [16] add time. Finally, generative models fill regions that no camera saw: PIS3R uses diffusion as a refinement stage, whereas Tuli et al. [13] make it the core of the stitching process.

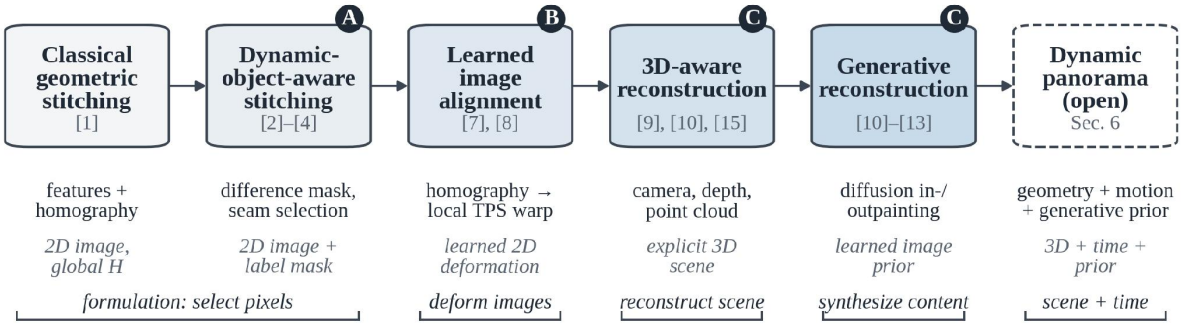


Fig. 1. Methodological evolution of panoramic reconstruction (original diagram created for this review). Boxes give each stage with representative references; the rows beneath give the key operation and, in italics, the scene representation; brackets indicate how each stage formulates the

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

problem. Circled letters mark the approaches compared in depth: A = DOS [4], B = UDIS++ [8], C = PIS3R [10], which spans the 3D-aware and generative stages. The dashed stage is the open direction discussed in Section 6.

3. Representative Approaches

3.1 Classical Geometry with Explicit Dynamic Handling (DOS)

Reported. DOS [4] targets fast panorama construction on mobile devices for image sequences captured by rotating a handheld camera. Images are registered with ORB features and a RANSAC-estimated homography, and the overlap of each warped pair is processed in two branches. One extracts a minimum-cost seam from the squared-difference energy by dynamic programming; the other thresholds and median-filters the absolute difference into a binary mask of possible moving objects. A linear blend constrained by the seam and the mask, with seam-based colour compensation, produces the composite, which is then cropped to a rectangle. Compared with two commercial tools and two published methods, the paper reports fewer ghosting artifacts in scenes with walking people and the second-lowest processing time of the five methods. It also reports three limitations: unsatisfactory results under large parallax, sensitivity to strong illumination differences, and spurious edges in the difference image when registration error exceeds about one pixel.

Interpretation. Abstracted from implementation, DOS combines geometric alignment with a global model, identification of regions where aligned views disagree, and composition that routes the seam and blend around them. Dynamic content is thereby treated as a pixel-selection problem: the method decides which source to show, not what happened in the scene.

3.2 Learned Parallax-Tolerant Alignment (UDIS++)

Reported. UDIS++ [8] is an unsupervised deep stitching method. Its warp parameterizes registration from a global homography to local TPS motion in a unified representation, optimized jointly for alignment in the overlap and shape preservation outside it. An iterative warp-adaptation strategy improves generalization across datasets and resolutions, and a composition network learns seam-driven composition masks without supervision to remove remaining parallax artifacts. The authors report gains over prior methods both quantitatively and qualitatively, with qualitative comparisons that include large-parallax and low-texture scenes.

Interpretation. UDIS++ learns two classical ideas end to end, flexible warping [5], [6] and seam cutting [3], [14], and frees them from hand-crafted geometric features. Its representation nevertheless remains two-dimensional: a deformation field and a composition mask, with no explicit depth and no notion of time.

3.3 3D Reconstruction with Diffusion Refinement (PIS3R)

Reported. PIS3R [10] targets very large parallax and distinguishes it from pure rotation and slight parallax. It applies VGGT [9] to two input images to recover intrinsic and extrinsic camera parameters and a dense 3D reconstruction, reprojects the dense point cloud onto a designated reference view to obtain pixel-wise alignment and an initial stitch, and refines holes and noise with a point-conditioned image diffusion module. The authors state that the result preserves pixel-level geometric integrity, usable for downstream SfM, and report that warp-based methods, including UDIS++, fail on their very-large-parallax cases.

Interpretation. PIS3R replaces alignment with reconstruction: where a pixel lands is decided by estimated depth and pose rather than by a fitted warp. Generation is subordinate to, and conditioned on, geometry. Tuli et al. [13] take the opposite emphasis: they fine-tune a diffusion inpainting model to preserve the content and layout of multiple reference images and then outpaint the full panorama from a single reference, so generation is primary and geometry implicit.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

4. Comparative Analysis

Table 1 summarizes the three approaches; this section examines the trade-offs problem by problem.

Table 1. Methodological comparison of the three approaches analysed in depth. Entries describe trade-offs, not rankings, and no numerical scores are assigned. Unmarked entries summarize what the original papers report; entries marked † are this review's interpretation.

Property	DOS [4]	UDIS++ [8]	PIS3R [10]
Paradigm	Classical feature-based stitching with explicit dynamic handling	Unsupervised learned alignment and composition	Feed-forward 3D reconstruction with diffusion refinement
Input (as evaluated)	Image sequence from a rotating handheld camera; pairwise registration chained	Image pair	Image pair with very large parallax
Scene representation	2D images and a binary moving-object mask	2D images, learned warp field and composition masks	Dense 3D point cloud with recovered intrinsics and extrinsics
Alignment model	One homography per pair (ORB, RANSAC)	Global homography extended to local TPS motion	Reprojection of reconstructed points to a reference view
Parallax regime	Small; large parallax reported as a limitation	Large parallax	Very large parallax
Dynamic objects	Explicit: difference mask, DP seam, masked blending	Not modelled explicitly; affected indirectly through alignment and learned composition†	Not part of the stated problem formulation; moving content is inconsistent with a static reconstruction†
Missing content	Not generated; panorama cropped to a rectangle	Not generated; the warp redistributes observed pixels†	Holes and noise refined by point-conditioned diffusion
Generative prior	No	No	Yes (diffusion)
Main strength	Lightweight, interpretable, explicit dynamic handling	Parallax tolerance without hand-crafted geometric features	Pixel placement consistent with a 3D reconstruction
Main limitation	Large parallax; strong illumination differences (reported)	No explicit depth, motion, or synthesis of unobserved content†	Dynamic and temporal consistency not addressed; high model cost†

4.1 Geometric Alignment

Approximately planar scenes. Eq. (1) is exact when the observed points lie on one plane or the camera only rotates. There the three formulations converge: a homography is the correct model, and flexible warps or 3D reconstruction add little. This explains why DOS, designed for a

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

handheld camera rotated in place, can rely on a single homography: the assumption matches its capture model.

Strong depth variation. For a camera translating by t_x with focal length f, a point at depth Z shifts by

$$\Delta u = f\, t_x / Z. \quad (2)$$

A homography fitted to a dominant plane at depth Z_0 compensates only that plane and leaves the residual

$$\delta u(Z) = f\, t_x\, (1/Z - 1/Z_0), \quad (3)$$

which vanishes only at $Z = Z_0$ and grows with baseline and depth range. Fig. 2(c) shows the result for a facade-aligned homography: the near post is doubled and the slanted ground is misregistered row by row. Real data behave identically: on a stereo pair with a strong depth range, the best RANSAC homography doubles every leaf away from its fitted plane (Fig. 3(e)). A thin-plate spline can absorb δu where it varies smoothly, which is why UDIS++ tolerates far more parallax than a global homography. A warp, however, is a single-valued map; at occlusion boundaries δu is discontinuous and a valid correspondence may not exist, so composition must hide the remainder. Reprojecting a reconstruction removes δu by construction when depth and pose are accurate, as for the post in Fig. 2(e), at the price of making alignment quality depend on reconstruction quality.

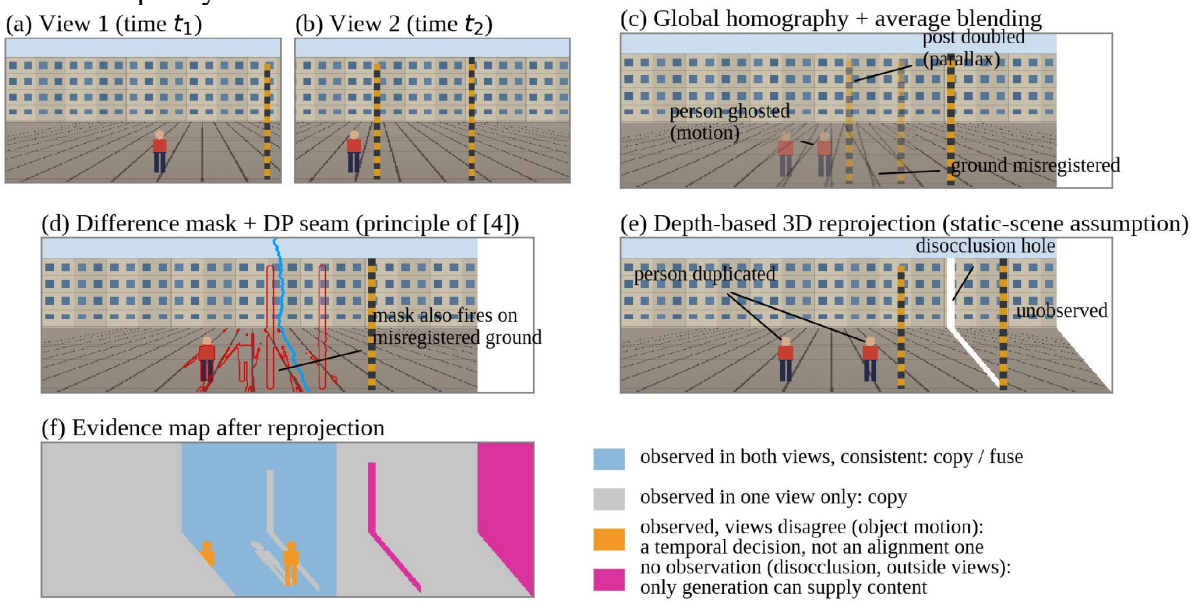


Fig. 2. Original synthetic experiment created for this review; no panel is reproduced from the cited works. A layered scene (facade, slanted ground, two thin posts, and a person who walks between exposures) is rendered from two horizontally translated cameras with known depth, and every composite in (c)–(f) is computed from the two rendered views. (c) A homography aligned to the facade doubles the near post and misregisters the ground (parallax, Eq. (3)) and ghosts the person (motion). (d) Our implementation of the mask-and-seam principle of [4]: red contours show the thresholded difference mask, blue the dynamic-programming seam. One instance of each object survives, but the mask also responds to parallax. (e) Depth-based reprojection into the view-1 frame, in the spirit of [10] but with ground-truth depth and no generative refinement: the post is aligned, the moving person is duplicated, and disoccluded and unobserved regions stay empty. (f) Per-pixel evidence after reprojection.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

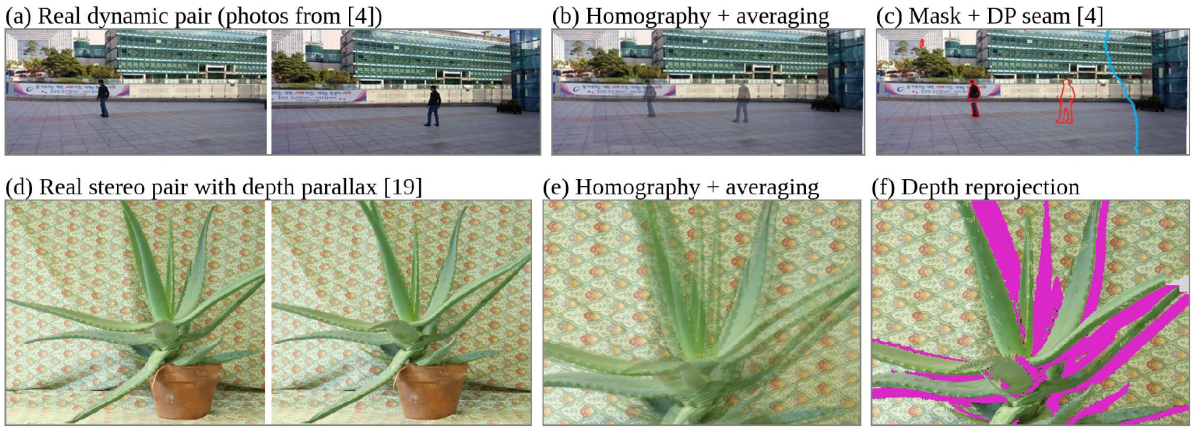


Fig. 3. Real-image examples; all output panels are original results computed for this review. Top: input photographs reproduced from [4] (the authors' own capture; © 2017 KSII, reused with attribution), registered with an ORB + RANSAC homography. (b) Averaging the overlap ghosts the walking person. (c) Our re-implementation of the mask-and-seam principle of [4] (red: thresholded difference mask; blue: dynamic-programming seam) retains a single instance. Bottom: Middlebury 2006 "Aloe" stereo pair [19] at half resolution. (e) A single RANSAC homography doubles leaves lying off its fitted plane. (f) Reprojecting the left view into the right camera with the dataset's ground-truth disparity aligns all depths but leaves pixels with no source: magenta marks disocclusions and pixels of unknown disparity; grey lies outside the left view. No generative completion is applied.

4.2 Dynamic Objects

For a point that also moves between exposures, the observed displacement becomes

$$\Delta u = f\, t_x / Z + \Delta u_m, \qquad (4)$$

where Δu_m is the image projection of the object's own motion. Eq. (4) is the crux of the dynamic problem. The first term is a property of scene geometry and can, in principle, be removed by better alignment; the second is a property of time and cannot. From two images, the two terms are confounded wherever depth is unknown.

The three approaches meet this confound differently. On the near-rotational real pair of [4], averaging ghosts the walker and the DOS mask isolates it cleanly (Fig. 3(b, c)), because Δu there is dominated by motion. DOS, however, detects disagreement, not motion: its mask fires wherever the aligned overlaps differ. The paper itself notes that registration errors above about a pixel produce spurious edges in the difference image, and Fig. 2(d) reproduces this behaviour: the mask outlines the person and also the parallax-misregistered ground. The seam then avoids both while a low-cost path exists. UDIS++ reduces the first term of (4), but its learned composition resolves the second only by selecting one source per pixel, and selecting is not modelling. Optical flow [18] measures Δu but does not decompose it. PIS3R's reconstruction is built for static content: a person observed at two positions yields two sets of 3D points, and reprojection renders both, as in Fig. 2(e). Neither [8] nor [10] states independent object motion as part of its problem formulation.

Hence the central distinction: improved geometric alignment reduces misalignment but does not by itself model object motion. Choosing which instance of a moving person appears in the panorama is a decision about time, and in all three pipelines it is made implicitly, if at all (Fig. 2(f), orange).

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

4.3 Large Parallax

Large parallax stresses each formulation at a different point. For a homography, (3) grows without bound and the seam loses its low-cost paths; DOS reports this case as a limitation, consistent with its rotational capture model. Flexible warps extend the tolerated range but, according to [10], not to very large parallax. Reconstruction handles the geometry yet exposes a consequence that warping conceals: when the baseline is large, regions hidden behind foreground objects in one view and outside the other view have no observation at all. In Fig. 2(e) they appear as a disocclusion strip behind the far post and an unobserved wedge at the boundary; on real data, reprojecting one stereo view with ground-truth disparity leaves such holes beside every leaf (Fig. 3(f)). Large parallax thus converts part of the alignment problem into a missing-content problem.

4.4 Missing Content and Generation

A warping pipeline produces each output pixel by resampling an observation, $I(\mathbf{x}) = I_k(W_k(\mathbf{x}))$: every output value is copied from some view k. If no view observed the scene point, the region is left empty, stretched, or cropped away, as in DOS; the shape-preserving warp of UDIS++ redistributes content without creating it. Generation differs in kind: it samples content from a learned prior conditioned on the available evidence. PIS3R conditions on reprojected points [10]; Tuli et al. condition a diffusion inpainting model on multiple reference images [13]; generic inpainting [12] conditions only on one image.

Generative completion is therefore indispensable where Fig. 2(f) is magenta, but it is not neutral. The mechanism that invents plausible structure can also replace observed detail (loss of source fidelity), contradict scene geometry, vary between samples, and add cost through iterative denoising [11]. These are arguments for constraining generation rather than against it. Geometry supplies two constraints: it indicates where generation is needed (visibility) and what the generated content must agree with (depth, layout, cameras). PIS3R's point conditioning is an instance of the second; the first is rarely made explicit in current pipelines.

4.5 Computational Considerations

Reported timings use different hardware, so we compare cost conceptually. DOS uses operations chosen for low-powered devices [4], a concern shared by lightweight UAV-based vision systems [21]. UDIS++ needs GPU inference plus optional per-input iterative adaptation [8]. PIS3R combines a large feed-forward reconstruction transformer [9] with iterative diffusion sampling [10]. Cost thus grows with the scene structure a method recovers and the content it synthesizes, and for video it multiplies with frame count.

4.6 Temporal Consistency

All three methods produce a still image. Applied per frame to video, each would decide independently in every frame where the seam runs, what the mask removes, or which sample the diffusion model draws. Perazzi et al. [17] reported that blending and minimum-energy seams that succeed on quasi-static input produce visible artifacts on video because they lack temporal coherence. For generative pipelines we expect it to be sharper: generated regions may flicker, object appearance or identity may change between frames, and generated structure may drift from the reconstructed geometry. None of the reviewed papers evaluates video, so we claim nothing further about their temporal behaviour. Dynamic representations such as 4D Gaussian splatting [16] model geometry and motion jointly over time, but target view synthesis from posed captures rather than panoramic completion.

5. Research Gap

The comparison suggests three stages.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Stage 1. Classical dynamic-aware methods [2]–[4] reason cheaply and explicitly about where views disagree, but rest on simplified geometry, so misalignment and motion share one difference signal.

Stage 2. Deep stitching [7], [8] learns flexible deformation and composition, extending parallax tolerance, but remains 2D, without explicit depth, motion, or synthesis.

Stage 3. Geometry-aware generative methods [10], [13] combine 3D reconstruction or reference conditioning with generation, filling unobserved regions consistently, but are formulated for static content and single images.

Remaining problem. A dynamic scene contains a static 3D structure and time-varying content, and the two call for different treatment: the former should be reconstructed and reprojected, the latter tracked and placed explicitly in time, and only what neither can supply should be generated. No reviewed method performs all three jointly. We state the gap as follows:

Existing panoramic reconstruction approaches increasingly incorporate learned alignment, explicit 3D geometry, or generative priors, yet these components are not fully unified for dynamic scenes in which camera motion, scene depth, object motion, occlusion, and missing content interact. This motivates a geometry- and motion-conditioned generative reconstruction framework that preserves reliable observations and uses generation primarily where direct geometric reconstruction is insufficient.

This is a hypothesis bounded by the three methods analysed in depth; a broader survey may reveal partial solutions.

6. Proposed Geometry- and Motion-Aware Generative Framework

The proposed framework is a research hypothesis motivated by the reviewed literature and is not yet experimentally validated. Fig. 4 outlines it. Camera poses and depth are first estimated, e.g., with a feed-forward model [9], and fused into a 3D scene representation. Each panorama location then receives a confidence $c(\mathbf{x}) \in [0, 1]$ derived from visibility (is the point observed?), multi-view consistency (do observations agree after reprojection?), and reconstruction uncertainty. Inconsistency is not treated as error by default: a static–dynamic decomposition separates regions explained by camera motion from those that require the independent term Δu_m of (4), using optical flow [18] together with the reconstructed geometry. Static geometry D and estimated motion M then condition a generator G that acts only on low-confidence regions, and its output is fused with the observations:

$$\hat{I}(\mathbf{x}) = c(\mathbf{x})\, I_{obs}(\mathbf{x}) + (1 - c(\mathbf{x}))\, G(\mathbf{x} \mid D, M, I_{obs}), \quad (5)$$

Eq. (5) formalizes the principle that motivates the framework: generate only where reliable geometric and observational evidence is insufficient. For video, shared geometry and motion estimates would anchor generated content to persistent scene structure instead of re-sampling it per frame.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

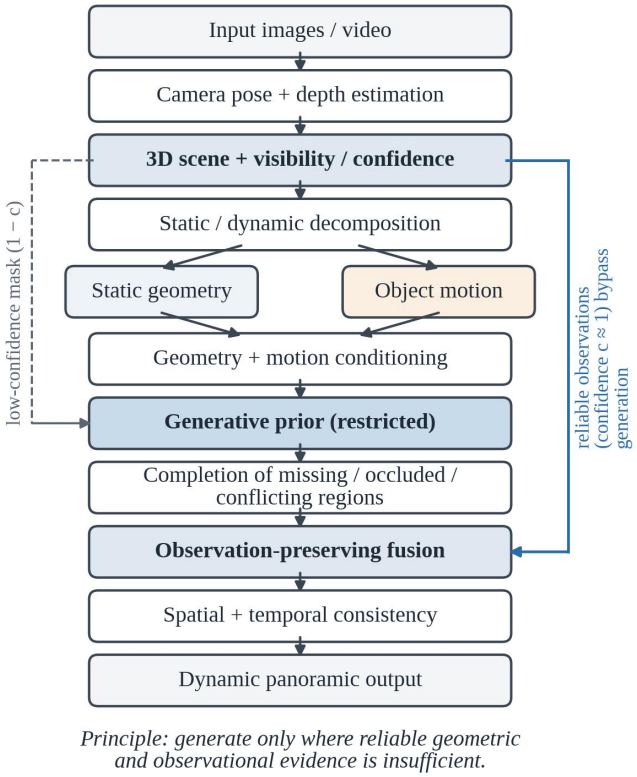


Fig. 4. Proposed geometry- and motion-conditioned generative framework (original diagram; a research hypothesis, not a validated system). Reliable observations bypass generation (blue path); the low-confidence mask (1 − c) restricts where the generative prior acts; Eq. (5) defines the fusion.

The framework raises four questions suitable for doctoral investigation. RQ1 (confidence): how can visibility, reprojection consistency and reconstruction uncertainty be combined into a calibrated c(x)? RQ2 (decomposition): can the two terms of (4) be separated reliably from few views, where depth and motion are jointly ambiguous? RQ3 (constrained generation): how should geometry and motion condition a diffusion model so that generated regions agree with the reconstructed structure? RQ4 (temporal stability): how can generated regions stay consistent across frames at moderate cost? Evaluation should report observation fidelity, geometric consistency of generated regions, and temporal stability separately rather than as one quality score.

7. Conclusion

Panoramic reconstruction has moved from choosing which observed pixels to show, through deforming images, to reconstructing the scene and synthesizing what no camera saw. The three compared formulations resolve, respectively, explicit handling of disagreement, flexible alignment, and geometric consistency with generative completion. None provides the joint treatment of static geometry, object motion and missing content that dynamic scenes require. We have proposed, as an untested hypothesis, a framework in which geometry and motion decide where generation is permitted and what it must match. Its value depends on making confidence estimation and motion decomposition reliable.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

References

1. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. Comput. Vis., vol. 74, no. 1, pp. 59–73, 2007.
2. J. Davis, "Mosaics of scenes with moving objects," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 1998, pp. 354–360.
3. A. Mills and G. Dudek, "Image stitching with dynamic elements," Image Vis. Comput., vol. 27, no. 10, pp. 1593–1602, 2009.
4. M. Abdukholikov and T. Whangbo, "Fast image stitching method for handling dynamic object problems in panoramic images," KSII Trans. Internet Inf. Syst., vol. 11, no. 11, pp. 5419–5435, 2017, doi: 10.3837/tiis.2017.11.013.
5. J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 7, pp. 1285–1298, 2014.
6. F. Zhang and F. Liu, "Parallax-tolerant image stitching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 3262–3269.
7. L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Unsupervised deep image stitching: Reconstructing stitched features to images," IEEE Trans. Image Process., vol. 30, pp. 6184–6197, 2021.
8. L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Parallax-tolerant unsupervised deep image stitching," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023, pp. 7399–7408, doi: 10.1109/ICCV51070.2023.00680.
9. J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, "VGGT: Visual geometry grounded transformer," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025.
10. M. Zhu, X. Jin, C. Wang, Y. Zhang, Y. Xue, T. Ji, and Y. Lao, "PIS3R: Very large parallax image stitching via deep 3D reconstruction," arXiv:2508.04236, 2025.
11. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 10684–10695.
12. A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 11461–11471.
13. M. Tuli, K. Kamali, and D. B. Lindell, "Generative panoramic image stitching," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2026, pp. 5084–5094.
14. J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, "Seam-driven image stitching," in Proc. Eurographics (Short Papers), 2013, pp. 45–48.
15. B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," ACM Trans. Graph., vol. 42, no. 4, Art. no. 139, 2023.
16. G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian, and X. Wang, "4D Gaussian splatting for real-time dynamic scene rendering," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 20310–20320.
17. F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross, "Panoramic video from unstructured camera arrays," Comput. Graph. Forum, vol. 34, no. 2, pp. 57–68, 2015.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

18. Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2020, pp. 402–419.
19. D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2007, doi: 10.1109/CVPR.2007.383191.
20. A. Mamadmurodov, S. Umirzakova, M. Rakhimov, A. Kutlimuratov, Z. Temirov, R. Nasimov, A. Meliboev, A. Abdusalomov, and Y. Im Cho, "A hybrid deep learning model for early forest fire detection," Forests, vol. 16, no. 5, Art. no. 863, 2025, doi: 10.3390/f16050863.
21. A. Abdusalomov, S. Umirzakova, A. Kutlimuratov, D. Mirzaev, A. Dauletov, T. Botirov, M. Zakirova, M. Mukhiddinov, and Y. I. Cho, "Lightweight UAV-based system for early fire-risk identification in wild forests," Fire, vol. 8, no. 8, Art. no. 288, 2025, doi: 10.3390/fire8080288.
22. F. Makhmudov, A. Privalov, S. Egorenkov, A. Pryadkin, A. Kutlimuratov, G. Bekbaev, and Y. I. Cho, "Analytical approach to UAV cargo delivery processes under malicious interference conditions," Mathematics, vol. 13, no. 12, Art. no. 2008, 2025, doi: 10.3390/math13122008.
23. M. U. Haq, M. A. J. Sethi, S. Ahmad, N. Ahmad, M. S. Anwar, et al., "A comprehensive review of face detection/recognition algorithms and competitive datasets to optimize machine vision," Comput. Mater. Contin., vol. 84, no. 1, pp. 1–24, 2025, doi: 10.32604/cmc.2025.063341.
24. F. Safarov, A. Kutlimuratov, U. Khojamuratova, A. Abdusalomov, and Y.-I. Cho, "Enhanced AlexNet with Gabor and local binary pattern features for improved facial emotion recognition," Sensors, vol. 25, no. 12, Art. no. 3832, 2025, doi: 10.3390/s25123832.
25. A. Abdusalomov, A. Kutlimuratov, R. Nasimov, and T. K. Whangbo, "Improved speech emotion recognition focusing on high-level data representations and swift feature extraction calculation," Comput. Mater. Contin., vol. 77, no. 3, pp. 2915–2933, 2023, doi: 10.32604/cmc.2023.044466.
26. A. Abdusalomov, S. Umirzakova, O. Bekmirzaev, A. Dauletov, A. Buriboev, A. Kutlimuratov, A. Nishanov, R. Nasimov, and R. Oh, "From anatomy to genomics using a multi-task deep learning approach for comprehensive glioma profiling," Bioengineering, vol. 12, no. 9, Art. no. 979, 2025, doi: 10.3390/bioengineering12090979.
27. M. J. A. Rasool, A. Abdusalomov, A. Kutlimuratov, M. J. A. Ahamed, S. Mirzakhalilov, A. Shavkatovich Buriboev, and H. S. Jeon, "PixMed-Enhancer: An efficient approach for medical image augmentation," Bioengineering, vol. 12, no. 3, Art. no. 235, 2025, doi: 10.3390/bioengineering12030235.