A Window to the Past: Connecting Historical Aerial Photographs to Satellite Images with the use of Image Registration and Enhancement

Aikaterini Christopoulou[1,2][0000-0001-9928-141X], Basilis Gatos[1][0000-0001-9873-0826] and Ioannis Kakogeorgiou[1][0000-0001-5200-2620]

[1] Institute of Informatics & Telecommunications, National Centre for Scientific Research "Demokritos", Agia Paraskevi, Greece

[2] Department of Economics & Sustainable Development, Harokopio University, Athens, Greece

achristopoulou@hua.gr

Abstract. Historical aerial photographs are valuable depictions of the past, worth analyzing in order to comprehend the dynamic change of the landscape. Despite their value and importance, historical aerial photographs are hard to further analyze, as in most cases they lack georeferencing, making it impossible to be connected with modern satellite imagery. The task of referencing these images has proven to be a challenge, owing to the large number of differences when compared to satellite images as well as taking into consideration the drastic change of landscape since these photographs were acquired, which is more than 80 years ago. In this work, we propose a workflow which includes the translation of aerial photographs into a modern-like visual domain, using the unpaired image-to-image, CycleGAN framework. These enhanced images exhibit radiometric and textural characteristics closer to modern satellite imagery, while maintaining the original geometric layout. At a next step, we improve image registration robustness by using the enhanced historical images as input to an image alignment pipeline based on improved ORB features and RANSAC algorithm for geometric verification and transform estimation. An additional contribution of our work is the distribution of three publicly available datasets (two for registration and one for image translation) that include historical aerial photographs (1945) and satellite images which cover a larger area, from 2015 and 2025. Our work is characterized by an approximately 50% improvement in registration compared to commonly used SIFT and ORB detectors. An additional evaluation showed that the proposed historical-to-modern appearance translation enhances the performance of template matching techniques.

Keywords: Image Registration, CycleGAN, SIFT, ORB, Template matching.

1 Introduction

Historical aerial photographs constitute one of the most valuable yet underexploited sources of spatial information about the Earth's surface, holding a unique image of the past. Taken between 1935 and 1945, these first historical aerial photographs are an initial real depiction of what the world looked like more than 80 years ago. In many regions of the world, historical aerial imagery represents the earliest spatially explicit record of land use and land cover (LULC). When it comes to studying landscape structure, evolution and dynamics, it is inevitable that one must turn to the past and extract information on the history of a landscape [1]. By looking at a landscape's past, one can detect the effect of human presence and activity on the land cover and measure land use change, thus making the reconstruction of past landscapes via modelling more feasible. Unlike historical maps, historical aerial photographs serve as direct evidence of how extensively the physical and human environment has evolved, long before digital mapping and remote earth observation appeared. Consequently, the motivation for studying and analyzing historical aerial photographs extends well beyond pure curiosity, as they can play a crucial role in disciplines such as historical geography, archaeology, environmental science, urban planning, disaster and risk management and climate-change research.

Historical aerial photographs nonetheless remain difficult to integrate into modern geospatial workflows. Most archival imagery was acquired with analog cameras and film-based sensors, and subsequent scanning under heterogeneous protocols introduced geometric distortions, uneven illumination, and noise. Moreover, historical aerial images are typically not georeferenced, or are only loosely referenced using approximate metadata, preventing their reliable integration into a Geographic Information System (GIS) for further spatial analysis. The georeferencing problem is further compounded by the fact that landscapes present substantial differences to what the land looks like today (see Fig. 1), and is exacerbated by differences in illumination conditions, sensor characteristics and atmospheric elements such as clouds. As a result, direct pixel-wise comparison between historical aerial photographs and modern satellite images is generally not feasible without robust preprocessing and accurate geometric registration.

(a) (b)

Fig. 1. Landscape changes captured at a historical aerial photograph of 1945 (a) and a satellite image of 2015 (b) of Andros Island (Greece). The source of both image types is the Hellenic Cadastre Geoportal.

Image registration, the process of geometrically aligning images of the same scene acquired at different times, viewpoints, or sensors, is the prerequisite step that unlocks the analytical potential of historical aerial photographs. It enables overlaying historical imagery with current satellite products, digital elevation models, and other GIS layers; transferring modern annotations onto historical data; and localizing features visible only in historical imagery (vanished infrastructure, former river channels, obsolete land-use patterns) in modern coordinates. Accurate registration is therefore a prerequisite for large-scale automated analysis of historical imagery, including change analysis and the reuse of modern LULC models. Yet standard registration techniques degrade substantially under the radiometric and geometric disparities between historical aerial photographs and modern satellite imagery.

The aim of this work is twofold. First, we reduce the appearance gap between historical aerial photographs and modern satellite imagery by translating archival images into a modern-like visual domain. Because historical and modern images are not spatially aligned beforehand, paired pixel-aligned training is impractical; we therefore adopt the well-established unpaired image-to-image translation framework Cycle-Consistent Generative Adversarial Networks (CycleGAN) [2] to learn a mapping that normalizes radiometric characteristics (contrast, tone, texture statistics) and produces a colorized/enhanced historical representation that is more consistent with the modern domain. We argue that this appearance normalization simplifies the subsequent registration task, since domain discrepancies that would otherwise burden feature detection, description, and matching are handled by a dedicated translation model. Second, we improve registration robustness by feeding the enhanced historical images into an improved ORB-based alignment pipeline. Specifically, we use an enhanced ORB detector/descriptor with an adaptive patch-size selection per image pair, improving matching reliability under strong radiometric differences and varying scene content, while Random Sample consensus (RANSAC) [3] filters outlier correspondences and estimates the geometric transformation. For evaluation and reproducibility, we openly release three public datasets: (a) RegGR1945_2015, containing 50 historical aerial photographs from 1945, their corresponding satellite images from 2015 (covering a larger area) and ground-truth registration coordinates; (b) RegGR1945_2025, containing 30 historical aerial photographs from 1945, the coordinates of their corresponding satellite images from 2025 (covering a larger area) and ground-truth registration coordinates and (c) CycleGAN_training_patches, containing 1,000 historical aerial photograph patches and 1,000 satellite image patches used for appearance translation training. The first dataset involves only shift misalignment, whereas the second additionally includes anisotropic scale and rotation changes, making it more challenging. Our experiments demonstrate that incorporating CycleGAN-based translation yields a substantial improvement in registration performance, supporting the premise that appearance normalization can materially facilitate cross-temporal and cross-sensor alignment.
Concretely, the contributions of our work are the following:

- We show that unpaired historical-to-modern appearance translation with CycleGAN reduces domain discrepancies, leading to more reliable alignment between historical aerial photographs and modern satellite images.
- We introduce an adaptive ORB patch-size selection strategy that improves the robustness of feature extraction and matching across challenging historical-modern pairs.

- We release three publicly available datasets, RegGR1945_2015 and RegGR1945_2025, with ground-truth registration coordinates for image registration and CycleGAN_training_patches for image translation and enhancement [4].
- We empirically isolate the contribution of appearance translation through a dedicated ablation with multiple template-matching similarity metrics, demonstrating that the translation step provides consistent gains independent of the downstream matching strategy.

2 Related Work

Registration of contemporary satellite images and unmanned aerial vehicles (UAV) imagery is a well-studied topic. In [5], the focus is on the registration between satellite imagery and UAV aerial images of agricultural land, combining nearest-neighbor and brute-force descriptor matching strategies to improve the performance of scale-invariant feature transform (SIFT) [6], Speeded-Up Robust Features (SURF), and Oriented FAST and Rotated BRIEF (ORB) [7] feature points. Approach [8] addresses nonlinear radiometric distortion using a multiscale template matching method based on frequency-domain convolutional maps and rotation-invariant vectors, though it remains constrained by the computational cost of its coarse-to-fine search. Approach [9] proposes a deep learning semantic template-matching framework targeting multimodal images with severe nonlinear radiometric differences; it requires a large training set with known geometric transformations.

More recently, learned local features and learned matchers have substantially advanced wide-baseline and cross-modal image matching. SuperPoint [10] jointly learns keypoint detection and description in a self-supervised manner, LightGlue [11] casts feature matching as a graph-based assignment problem solved with attention, and LoFTR [12] removes the explicit keypoint stage altogether, producing dense matches through a coarse-to-fine transformer. Dense matchers such as RoMA [13] further improve robustness under strong appearance changes. These methods consistently outperform classical handcrafted features on standard benchmarks, but typically require large amounts of training data and substantial computational resources at inference, and their behavior on the specific domain of decades-old, monochromatic, film-scanned aerial photographs remains largely unexplored.

Registering historical aerial photographs, acquired decades before the first satellites, with modern imagery presents additional challenges not addressed by the methods above, including geometric distortions, uneven illumination, and drastic landscape change. In [14], a semi-automatic process is proposed for efficiently matching and preparing scanned historical aerial photographs for georeferencing, combining SIFT feature extraction with RANSAC to remove outliers and obtain good conjugated points between images, but still requiring manual control points for final rectification. The approach of [15] introduces a historical aerial photographs orthorectification workflow based on a conventional photogrammetric procedure integrated as a GIS procedure, illustrated through an application to a rugged landscape in Portugal as part of a study on long-term socioecological dynamics.

In our approach, a CycleGAN [2] is applied to the historical aerial photographs for colorization and image enhancement so that they look closer to modern satellite images. Generative Adversarial Networks (GANs) have emerged as a powerful framework for image synthesis and translation. Among them, CycleGANs have gained significant attention due to their ability to perform unpaired image-to-image translation, which is particularly attractive for applications where paired training data is scarce or unavailable, as is the case for historical imagery. CycleGANs have been widely applied to style transfer, medical imaging, and remote-sensing tasks including image enhancement, grayscale colorization, and cross-sensor translation. Of particular relevance, [16] uses CycleGANs for unpaired translation between recent RGB satellite images and historical monochromatic aerial images, achieving automated multi-class land cover mapping without target-specific annotations. More recent unpaired translation methods based on contrastive learning and diffusion models have also been proposed, achieving strong results on natural-image benchmarks. We adopt CycleGAN in this work because it remains a well-validated, lightweight baseline that trains stably on small unpaired datasets such as ours, and because our objective is to evaluate the effect of appearance translation on the registration pipeline rather than to optimize the translation network itself; the modular design of our framework allows future replacement of CycleGAN with more recent translation models.

In summary, our work differs from related approaches in three respects: (1) it targets the under-studied historical-to-modern aerial setting; (2) it isolates the contribution of appearance translation through a dedicated ablation; and (3) it provides public benchmark datasets that can support future evaluation of learned feature matchers and translation models on this domain.

3 Methodology

The proposed methodology aims to improve the registration of historical aerial photographs to modern satellite imagery by addressing two main challenges: (i) strong radiometric appearance discrepancies between the two domains, and (ii) limited robustness of feature-based registration under such discrepancies. As shown in Fig. 2, our approach combines unpaired image-to-image translation for image enhancement and appearance normalization with an adaptive feature extraction and matching strategy.

3.1 Unpaired Historical-to-Modern Appearance Translation

Historical aerial photographs and modern satellite images differ substantially in radiometry and appearance (e.g., contrast, tone, illumination, texture statistics, and film/scanning artifacts and noise), which often degrades direct feature matching. Our goal is to reduce this appearance gap while preserving scene geometry, so that subsequent registration becomes more reliable. We emphasize that the translation step does more than colorize, as it normalizes contrast, local tone, and texture statistics, all of which directly affect grayscale feature detectors and pixel-wise similarity metrics. The empirical benefit of this normalization is quantified in Section 5.4, where it consistently improves template-matching performance under multiple grayscale similarity measures. Since historical and modern images are typically not pre-aligned, paired pixel-wise supervision is impractical; we therefore adopt an unpaired image-to-image translation approach, based on CycleGANs [2]. The rationale for choosing CycleGAN over more recent alternatives is discussed in Section 2.

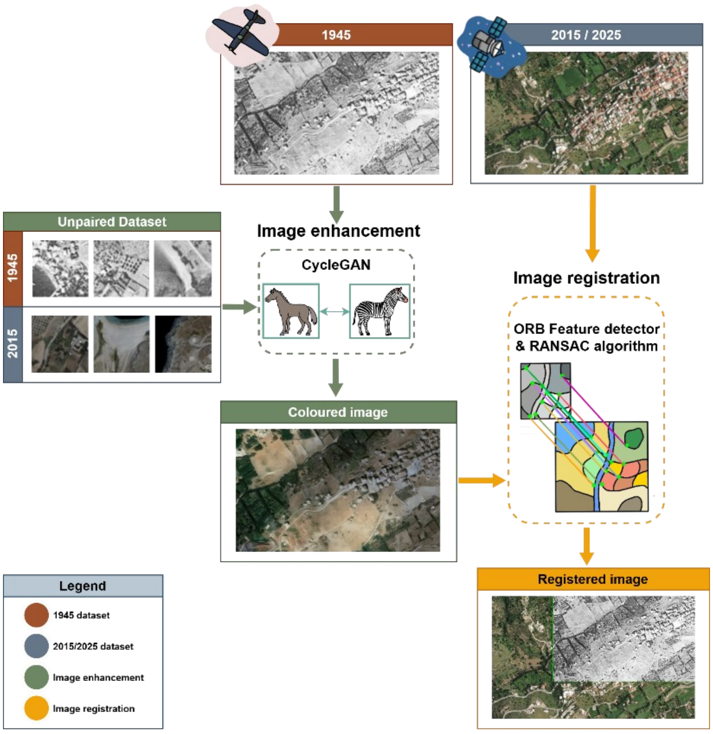

Fig. 2. The pipeline of the proposed method. An unpaired dataset of 256×256 image patches of aerial photographs of 1945 (1,000 patches) and satellite images 2015 (1,000 patches) is used to train the CycleGAN for the colorization of the aerial photograph of 1945. The colored image of 1945 is combined with ORB feature detector and RANSAC algorithm for the registration of the original aerial photograph.

Unpaired, structure-preserving appearance translation. Let $\boldsymbol{H}$ and $\boldsymbol{M}$ denote the domains of historical aerial image patches and modern satellite image patches, respectively. Two generator networks, $\boldsymbol{G_{H\rightarrow M}}$ and $\boldsymbol{G_{M\rightarrow H}}$, are trained jointly with two discriminators, $\boldsymbol{D_M}$ and $\boldsymbol{D_H}$, to learn mappings between the two domains without requiring paired training samples. The training objective combines three terms: an adversarial loss that pushes the generated images toward the target domain; a cycle-consistency loss enforcing $\boldsymbol{G_{M\rightarrow H}(G_{H\rightarrow M}(x)) \approx x}$ (which preserves spatial structure) and an identity

loss (which stabilizes color and tone on inputs already in the target domain). Together, these terms produce outputs that resemble the target domain while preserving the geometric layout, essential for downstream registration.
The CycleGAN is trained using unpaired 256×256 patches extracted from historical aerial photographs and satellite images. Patch-based training is adopted for memory efficiency and to increase the effective size of the training set; it is not an architectural constraint, as the generators are fully convolutional and can be applied to inputs of arbitrary spatial size at inference time. Once trained, the generator $\boldsymbol{G_{H \to M}}$ is used to produce an appearance-normalized (modern-like) version of the historical aerial photographs. These enhanced images exhibit radiometric and textural characteristics closer to modern satellite imagery, while maintaining the original geometric layout.


(a) (b) (c)

Fig. 3. Image enhancement and colorization using CycleGAN: (a) original image, result without (b) and with (c) the overlapping tiling and blending step.

Patch-wise inference with overlap-and-blend reconstruction. Full-resolution inference is performed patch-wise due to GPU memory constraints. Non-overlapping tiling produces visible seams at patch boundaries, since adjacent patches are translated independently with slightly different output statistics. We therefore partition the image into 256×256 patches with a stride of 32 pixels, so that each interior pixel is covered by multiple overlapping patches; predictions in overlap regions are averaged. Averaging suppresses both boundary discontinuities and high-frequency translation noise, since noise is largely uncorrelated across patches while structural content is correlated. Fig. 3 illustrates the effect of overlap-and-blend reconstruction compared to direct patch stitching: the seamed output (b) exhibits a clearly visible block pattern that is removed in the blended result (c).

3.2 Adaptive ORB Feature Matching for Robust Historical–Modern Registration

Despite the radiometric normalization achieved through CycleGAN-based enhancement, residual discrepancies between historical aerial photographs and modern satellite images remain, including variations in local texture, scale, and noise characteristics. Standard feature-based registration pipelines, when applied with fixed parameter settings, often fail to robustly establish correspondences under such conditions. Following [17], we adopt ORB [7] features as the basis of our matching pipeline because of their computational efficiency and their proven effectiveness for remote-sensing image registration. We observe, however, that the performance of the ORB detector and descriptor is sensitive to the choice of descriptor patch size, which directly affects both feature distinctiveness and repeatability: small patches produce many keypoints with limited descriptive context, while large patches produce fewer but more discriminative descriptors that may, however, become unstable in the presence of strong local appearance change. The optimal patch size depends on the specific content, resolution, and degree of appearance change of each image pair, and cannot be set globally. To address this limitation, we introduce an adaptive ORB feature extraction strategy that dynamically selects the descriptor patch size on a per-pair basis.

Adaptive patch-size selection for robust ORB matching. For each historical–satellite pair, ORB features are extracted using 15 candidate descriptor patch sizes drawn from [31, 301] with a step of 20 pixels. Tentative correspondences are obtained by brute-force matching with Hamming distance and cross-check, and each configuration is evaluated by running RANSAC and counting the resulting inliers; the configuration with the largest inlier set is selected. We use the inlier count rather than the raw number of tentative matches, since the latter conflates good correspondences with outliers and propagates them into the transform estimation. Selecting only keypoints repeatable across multiple patch sizes was also considered, but the strong dependence of ORB orientation and descriptors on patch size leaves too few repeatable keypoints across configurations for stable estimation. Once the optimal patch size is fixed, ORB features are re-extracted, matched, and filtered with a distance threshold; sharpening the input with a high-pass filter prior to extraction further improves keypoint detectability in low-contrast regions.

RANSAC-based Geometric Verification and Transform Estimation. Given the presence of noise, occlusions, and scene changes, the initial set of correspondences may include a significant number of outliers. A robust geometric verification step is therefore required to filter incorrect matches. We employ RANSAC [3] to estimate the geometric transformation between the enhanced historical aerial image and the satellite image. RANSAC iteratively samples minimal subsets of feature correspondences to fit a transformation model and evaluates the consensus set of inliers that are consistent with the estimated model. A correspondence is classified as an inlier if the symmetric reprojection error under the estimated transformation is below 10 pixels; this threshold was held fixed across all methods (SIFT, ORB, Adaptive ORB, and the proposed method) to ensure a fair comparison. Depending on the expected geometric distortions and acquisition conditions, similarity or affine transformation models are used: a similarity model is used for RegGR1945_2015, which involves predominantly translational misalignment, and an affine model for RegGR1945_2025, which additionally exhibits anisotropic scale and rotation. The transformation that yields the maximum number of inliers is selected as the final alignment model. The use of adaptive ORB features significantly increases the proportion of inliers identified by RANSAC, resulting in more stable and accurate transformation estimation.

4 Datasets

To support evaluation of the proposed method and to encourage further research on historical–modern image registration, we release three publicly available datasets [4]. The first two (RegGR1945_2015 and RegGR1945_2025) are benchmark datasets that differ in the difficulty of the geometric transformations involved. In RegGR1945_2015, the historical aerial photographs are subject only to shift misalignment with respect to the satellite images, while in RegGR1945_2025 they additionally exhibit anisotropic scale changes and rotation, making the registration substantially more challenging. The first dataset (RegGR1945_2015) (see Fig. 4) contains 50 historical aerial photographs from 1945, the corresponding satellite images from 2015 covering a larger area than the historical footprint, and the ground-truth registration parameters that align each historical photograph to its modern counterpart. The second dataset (RegGR1945_2025) (see Fig. 5) contains 30 historical aerial photographs from 1945, the coordinates of the corresponding satellite images from 2025, and the ground-truth registration parameters. In both datasets, image selection prioritized areas containing permanent structural features — coastlines and primary road networks — which remain largely unchanged across the 80-year temporal gap regardless of broader landscape change.

(a) (b) (c)

Fig. 4. Example images from the RegGR1945_2015 dataset from Amorgos Island: (a) historical aerial photograph, (b) satellite image of 2015, (c) visualization of the correct registration of the aerial photograph to the satellite image of 2015. The source of both image types is the Hellenic Cadastre Geoportal [18].

(a) (b) (c)

Fig. 5. Example images from the RegGR1945_2025 dataset from Rhodos Island: (a) historical aerial photograph, (b) satellite image of 2025, (c) visualization of the correct registration of the aerial photograph to the satellite image of 2025. The source of the satellite images of 2025 is the Satellites pro website [19] and of the historical aerial photographs the Hellenic Cadastre Geoportal [18].

The third dataset contains the image patches used for training the CycleGAN model. The patches are organized into two subfolders, each of which contain 1,000 images, following the standard CycleGAN training convention. These subfolders are trainA (source domain) and trainB (target domain).

5 Experimental Results

This section presents the comparative evaluation of the proposed method together with the implementation details for all methods, an isolated evaluation of the appearance-translation step, and an analysis of failure cases.

5.1 Evaluation Metrics

To quantitatively evaluate the registration performance, we adopt two metrics that capture different aspects of registration quality:

(a) **Average Intersection over Union (Avg IoU)**. For each test pair, IoU is computed as the overlap between the region of the satellite image covered by the predicted transformation of the historical photograph and the corresponding ground-truth region, expressed as Area(Intersection) / Area(Union). The score ranges from 0 (no overlap) to 1 (perfect alignment), and the average over all test pairs summarizes overall registration quality. We deliberately use the term “Average IoU” rather than “mean IoU (mIoU) ”, since the latter is conventionally reserved in the segmentation literature for averaging across semantic classes and may otherwise cause confusion.

(b) **Success rate.** A registration is considered successful if its IoU exceeds a threshold τ, and the success rate is reported as the percentage (and absolute count) of pairs satisfying this criterion. While the two metrics are related, as the success rate is a thresholded summary of the IoU distribution, they offer different views: Average IoU is sensitive to the magnitude of misalignment for borderline cases, whereas the success rate at a strict threshold is robust to outliers and more directly interpretable as a deployment-relevant accuracy figure.

An alternative evaluation strategy would be to define a set of ground-truth point correspondences on each pair and report the mean reprojection error in pixels. We chose the area-based formulation for two reasons: it does not require manually annotating control points on dozens of difficult historical-modern pairs, and it summarizes the full geometric agreement of the registered images rather than the agreement at a sparse set of points. Reprojection error remains a complementary metric that can be computed on the released benchmarks in future work.

5.2 Comparative Results

The proposed method is compared against methods based on SIFT, ORB and Adaptive ORB descriptors. The success rate threshold τ is set to 0.93 for RegGR1945_2015 and to 0.90 for RegGR1945_2025, reflecting the greater geometric complexity of the second dataset, which involves combined transformations of translation, anisotropic scale and rotation, as opposed to the purely translational misalignment of the first. The implementation details for all methods are the following:

(a) **SIFT.** Input images are pre-processed with aggressive smoothing to retain only large-scale structure. SIFT descriptors are computed selecting the best sigma parameter from [0.8,6] based on the largest number of resulting inliers. Correspondences are established using brute-force matching with Euclidean distance and cross-checking enabled, retaining the top 15,000 matches. A robust affine transformation is then estimated using RANSAC with a reprojection error threshold of 10 pixels.

(b) **ORB.** The maximum number of keypoints is set to 50,000 to maximize the likelihood of finding reliable correspondences across domains. The descriptor patch size is left at the OpenCV default of 31 (i.e., the smallest value of our adaptive range) providing the standard ORB configuration as commonly applied in the literature. Because ORB descriptors are binary, matching uses Hamming distance with cross-check; the top 15,000 matches are retained and RANSAC is applied with a reprojection threshold of 10 pixels. The distinction between Euclidean and Hamming distance simply reflects the descriptor type: real-valued descriptors (SIFT) are compared in Euclidean space, while binary descriptors (ORB) are compared by bit-difference.

(c) **Adaptive ORB.** The configuration is identical to ORB, except that the descriptor patch size is selected per pair from the range [31,301] with a step of 20, by the inlier-count criterion described in Section 3.2.

(d) **Proposed method.** We apply Adaptive ORB after performing historical-to-modern appearance translation with CycleGAN. The CycleGAN is trained on 2,000 unpaired patches (1,000 historical, 1,000 satellite) of size 256×256 pixels, drawn exclusively from regions disjoint from the registration benchmarks to prevent data leakage. Training uses the Adam optimizer with learning rate 2×10^{-4} and a batch size of 1 for 200 epochs. The batch-size choice follows the original CycleGAN protocol of Zhu et al. [2] and is standard practice with the instance-normalization layers used in the generator, since instance normalization computes statistics per sample and is most stable with small batches. The generator uses 9 residual blocks, appropriate for the 256×256 patch resolution. Full-resolution inference uses the overlap-and-blend strategy of Section 3.1 with stride 32. The same Adaptive ORB and RANSAC settings as above are then applied to the translated images. Examples of registration results are shown in Fig. 6.

Comparison protocol. The registration pipeline (Adaptive ORB + RANSAC, reprojection threshold 10 px) is held fixed and only the input appearance is varied: raw historical images (the "Adaptive ORB" row) versus CycleGAN-translated images (the "Proposed" row). The improvement between these two rows isolates the translation contribution under an otherwise identical pipeline. SIFT and standard ORB are reported for absolute reference. The translation contribution is further isolated in Section 5.4 under a template-matching backbone, eliminating any effect of the feature detector.

Table 1 and Table 2 report results on the two benchmarks. The main observations are: (a) The proposed method outperforms all baselines on both datasets, with improvements of 3.82 % and 3.94 % in Average IoU over the next-best method (Adaptive ORB) on RegGR1945_2015 and RegGR1945_2025, respectively; (b) Adaptive ORB substantially outperforms standard ORB on both datasets, confirming that adaptive patch-size selection alone provides a large gain over fixed-parameter ORB matching; (c) The further improvement of the proposed method over Adaptive ORB confirms that the appearance-translation step provides a measurable additional benefit on top of the adaptive feature extraction, consistent with the isolated ablation in Section 5.4.

Table 1. Comparative Results on RegGR1945_2015 set (τ=93%)

Method	Avg IoU	#Success (IoU > τ)	Success Rate
SIFT	59.96 %	14	28 %
ORB	39.99 %	5	10 %
Adaptive ORB (Ours)	91.02 %	34	68 %
Proposed (Ours)	94.84 %	42	84 %

Table 2. Comparative Results on RegGR1945_2025 set (τ=90%)

Method	Avg IoU (%)	#Success (IoU > τ)	Success Rate (%)
SIFT	59.44 %	11	36.67 %
ORB	26.26 %	1	3.33 %
Adaptive ORB (Ours)	85.26 %	19	63.33 %
Proposed (Ours)	89.20 %	20	66.67 %

5.3 Analysis of Failure Cases

Registration performance is positively influenced by the presence of stable structural features, particularly coastlines, which provide reliable and distinctive keypoints that persist across the 80-year temporal gap. Conversely, areas that have undergone substantial landscape changes, such as the appearance of new road infrastructure, tend to negatively affect registration accuracy. The quality of the original aerial photograph also plays an important role, as blurry or geometrically distorted acquisitions reduce the effectiveness of both the CycleGAN enhancement and the subsequent feature extraction. Regarding the appearance translation step, CycleGAN produces visually consistent results for natural areas, agricultural land, bare soil, and road networks, but tends to misclassify heavily urbanized regions, rendering them as vegetated areas in some cases. This is likely attributable to the underrepresentation of dense urban fabric in the training patches, where dark tonal values in grayscale historical images, caused by specific illumination conditions, may resemble the appearance of vegetation in the modern target domain. A more balanced sampling of urban patches during CycleGAN training is a natural direction for further reducing this failure mode. For the RegGR1945_2025 dataset, the more complex geometric transformations introduce additional challenges; however, areas with clearly defined building blocks show surprisingly robust registration. Notably, the masked sea regions in the RegGR1945_2015 satellite images appear to have a positive effect, as removing ambiguous open water texture concentrates the matching on coastline features, which are among the most stable and distinctive across time.

Fig. 6. Sample registration results of the colored historical aerial photographs of 1945 with the satellite images of (a) 2015 of Andros Island town, (b) 2015 of Chania town in Crete, (c) 2025 of Agia Marina in Andros Island, and (d) 2025 of Andros Island town.

Table 3. Results on RegGR1945_2015 set using template matching and different similarity metrics with and without the proposed Historical-to-Modern appearance translation (τ=93%)

Similarity Metric	Historical-to-Modern appearance translation	Avg IoU	#Success (IoU > τ)	Success Rate
Cross-correlation Coefficient	NO	95.02 %	43	86 %
	YES	96.71 %	45	90 %
Normalized cross-correlation coefficient	NO	95.34 %	43	86 %
	YES	95.43 %	45	90 %
Cross-correlation	NO	48.24 %	6	12 %
	YES	59.03 %	14	28 %
Normalized cross-correlation	NO	75.11 %	32	64 %
	YES	93.00 %	43	86 %
Squared difference	NO	64.40 %	21	42 %
	YES	78.74 %	34	68 %
Normalized squared difference	NO	61.51 %	18	36 %
	YES	82.97 %	37	74 %

5.4 Further Evaluation of the Enhancement step

To isolate the contribution of appearance translation independently of the choice of feature detector, we evaluate template matching with multiple similarity metrics on the RegGR1945_2015 set, where the misalignment is purely translational. Because the matching backbone is fixed by construction (no detector or descriptor), any difference between with-translation and without-translation conditions is attributable to translation alone. Table 3 reports results across six similarity metrics. Translation improves performance in every case. The gain is marginal for metrics already invariant to monotonic intensity transformations (normalized cross-correlation coefficient) and substantial for metrics sensitive to absolute intensity statistics (normalized cross-correlation, normalized squared difference), consistent with the interpretation that translation principally normalizes contrast, local tone, and texture statistics rather than introducing structural content. Combined with the feature-matching results of Section 5.2, this confirms that the appearance-translation step provides a consistent, detector-agnostic improvement.

6 Conclusions

We have proposed a two-stage pipeline for registering historical aerial photographs to modern satellite imagery: unpaired CycleGAN-based appearance translation followed by adaptive ORB feature matching with RANSAC. The method outperforms classical SIFT and ORB baselines by a substantial margin on both released benchmarks, and the detector-agnostic ablation in Section 5.4 confirms that the appearance-translation step contributes consistently to this gain. Alongside the method, we release three public datasets, RegGR1945_2015, RegGR1945_2025, and CycleGAN_training_patches, which to our knowledge are the first openly available benchmarks for this task. Future work will use these benchmarks to evaluate recent learned matchers (e.g., LightGlue, LoFTR, RoMA) and more recent unpaired translation methods on historical imagery, both standalone and in combination with our translation step. This work is part of a broader project on historical landscape reconstruction; the modern-looking renderings produced here are intended to support downstream semantic segmentation of historical imagery, which is our next step.

Acknowledgments

The authors would like to thank Prof. V. Detsis from the Department of Economics & Sustainable Development, Harokopio University, Athens, Greece for his guidance and support throughout this project. The authors gratefully acknowledge the Hellenic Cadastre [18] for providing access to the 1945 historical aerial photographs and the 2015 satellite imagery used in this work and Satellites Pro [19] for providing access to the 2025 satellite imagery, in accordance with their terms of use. We thank the anonymous reviewers for their constructive feedback, which substantially improved the manuscript.

References

1. Ettehadi Osgouei, P., Sertel, E., Kabadayı, M.E.: Integrated usage of historical geospatial data and modern satellite images reveal long-term land use/cover changes in Bursa/Turkey, 1858–2020. Scientific Reports 12(1), 9077 (2022)

2. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proc. IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. IEEE, Venice (2017)
3. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Fischler, M.A., Firschein, O. (eds.) Readings in Computer Vision, pp. 726–740. Morgan Kaufmann, San Francisco (1987)
4. Zenodo. Accessed 18 June 2026 (2026). https://doi.org/10.5281/zenodo.19955647
5. Abbas, M., Saleem, S., Subhan, F., Bais, A.: Feature points-based image registration between satellite imagery and aerial images of agricultural land. Turkish Journal of Electrical Engineering and Computer Sciences 28(3), 1458–1473 (2020)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: International Conference on Computer Vision (ICCV 2011), pp. 2564–2571. IEEE, Barcelona (2011)
8. Gao, T., Lan, C., Huang, W., Wang, L., Wei, Z., Yao, F.: Multiscale template matching for multimodal remote sensing image. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16, 10132–10147 (2023)
9. Li, L., Han, L., Ding, M., Cao, H., Hu, H.: A deep learning semantic template matching framework for remote sensing image registration. ISPRS Journal of Photogrammetry and Remote Sensing 181, 205–217 (2021)
10. DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 224–236 (2018)
11. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: LightGlue: Local feature matching at light speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 17627–17638 (2023)
12. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8922–8931 (2021)
13. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: RoMa: Robust dense feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19790–19800 (2024)
14. Chen, H.R., Tseng, Y.H.: Study of automatic image rectification and registration of scanned historical aerial photographs. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 41, 1229–1236 (2016)
15. Pinto, A.T., Gonçalves, J.A., Beja, P., Pradinho Honrado, J.: From archived historical aerial imagery to informative orthophotos: A framework for retrieving the past in long-term socioecological research. Remote Sensing 11(11), 1388 (2019)
16. Van den Broeck, W.A., Goedeme, T., Loopmans, M.: Multiclass land cover mapping from historical orthophotos using domain adaptation and spatio-temporal transfer learning. Remote Sensing 14(23), 5911 (2022)
17. Misra, I., Rohil, M.K., Manthira Moorthi, S., Dhar, D.: Feature based remote sensing image registration techniques: A comprehensive and comparative review. International Journal of Remote Sensing 43(12), 4477–4516 (2022)
18. Hellenic Cadastre Geoportal. Accessed 18 June 2026 (2026). https://www.ktimatologio.gr
19. Satellites Pro Greece. Accessed 18 June 2026 (2026). https://satellites.pro/Greece_map#A38.803330,24.570923,8