JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 5, Page 508-513 | Article No. 60
DOI: 10.4898/jener.v2i5.a60
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

A Robust Offline Paper-based Exam Grading System with Marker-guided Coordinate Reconstruction

Myong Jin Kim[1], Nam Il Pak[2], Yong Jin Choe[2], Myong Song Ji[1], Myong Hyok Hwang[3] & Kwang Hui Kim[2]

[1]Department of Computer Science, Kim Chaek University of Technology, Pyongyang, DPR of Korea
[2]Department of Education Science, Kim Hyong Jik University of Education, Pyongyang, DPR of Korea
[3]Department of Mathematics, O Jong Hub Chongjin University of Education, Chongjin, DPR of Korea

ARTICLE INFORMATION

Article history:
Published: May 2026

Keywords:
Offline Exam Grading
Paper-Based Assessment
Automatic grading system
ArUco Marker
Optical Mark Recognition (OMR)

ABSTRACT

Paper-based examinations remain widely used due to their low infrastructure requirements and compatibility with conventional classrooms, but manual grading is time-consuming and error-prone. Existing automatic grading methods often rely on rigid templates or unstable feature-based alignment, which fails under low-resolution scanning, mobile capture, and perspective distortion. This paper proposes a robust offline grading system using marker-guided coordinate reconstruction and template-aware answer extraction. The framework integrates exam generation and grading through shared JSON-based metadata. During generation, ArUco markers and normalized answer coordinates are embedded. During grading, the system reconstructs the document coordinate system via homography estimation, enabling resolution-independent and distortion-robust answer extraction. The system fully supports multiple-choice and true/false questions via OMR, and provides semi-automatic grading for matching, ordering, fill-in-the-blank, and short-answer questions using OCR and similarity matching with instructor review. Experiments on 1,200 sheets under various conditions show that the proposed method achieves an IoU of 0.96, a reprojection error of 2.1 pixels, and up to 99.1% grading accuracy for objective questions, outperforming conventional approaches.

1. Introduction

Despite advances in digital learning, paper-based exams remain common due to their simplicity and accessibility [1,2]. However, manual grading is labor-intensive and inconsistent. Existing automatic systems based on OMR or OCR suffer from limitations: OMR requires fixed templates and fails under distortion, while OCR is sensitive to handwriting variation and image quality [3,4]. Furthermore, conventional feature-based alignment methods like SIFT, SURF, or ORB become unstable under low resolution, repetitive text, or sparse features [5–7].

Another critical gap is that most systems treat exam generation and grading as separate processes, missing the opportunity to use template metadata for robust alignment. This paper proposes an integrated framework that tightly couples generation and grading through shared structural metadata. Answer regions are stored as normalized coordinates, and ArUco markers enable reliable homography estimation under geometric distortion [8].. The system is fully offline and supports both automatic and semi-automatic grading with human-in-the-loop review.

The main contributions are:

- An integrated framework coupling exam generation and grading via shared metadata.
- A marker-guided calibration method using ArUco for robust alignment.
- A normalized coordinate reconstruction strategy for resolution-independent extraction.
- A modular offline pipeline supporting multiple question types and instructor review.

2. Related Work

Research on automatic examination grading has primarily focused on three major areas: Optical Mark Recognition (OMR), Optical Character Recognition (OCR), and document alignment or registration techniques [1–4]. These technologies constitute the core components of most existing paper-based examination processing systems.

Traditional OMR-based systems are highly effective for structured objective-type examinations such as multiple-choice and true/false questions. Commercial OMR solutions typically achieve high grading accuracy under controlled scanning environments and standardized answer sheet layouts. However, such systems generally depend on predefined templates, dedicated scanners, and carefully constrained acquisition conditions [12,13]. Even minor geometric distortions, paper deformation, rotation, or illumination variation can significantly degrade recognition performance. Furthermore, conventional OMR systems usually require specialized hardware and are difficult to adapt to flexible examination formats or mobile image acquisition environments.

To overcome the limitations of purely mark-based approaches, OCR-based grading systems have been studied for fill-in-the-blank, short-answer, and descriptive question evaluation [2,14,15]. OCR enables automatic extraction of textual responses and supports semi-automatic grading through keyword matching or similarity analysis. Modern OCR engines such as Tesseract provide practical recognition performance for printed text and constrained handwriting scenarios [2]. Nevertheless, OCR-based grading remains sensitive to handwriting variability, low image quality, shadows, blur, and background noise. In educational environments, handwritten answers frequently exhibit inconsistent writing styles and irregular spacing, which substantially reduce recognition reliability. Consequently, many existing OCR-assisted grading systems still require human verification for low-confidence predictions.

Another important research direction concerns document alignment and geometric correction. Since scanned or photographed examination sheets often contain rotation, perspective distortion, scaling variation, and partial displacement, accurate answer region localization requires robust document registration [5–8]. Early approaches commonly relied on template matching, projection profiles, or edge-based alignment methods. While computationally efficient, these methods are sensitive to noise and often fail under non-uniform geometric transformations.

Feature-based registration techniques such as SIFT [5], SURF [6], and ORB [4] have therefore become widely adopted for document alignment tasks. These methods estimate geometric transformations by detecting and matching local image features between template and target images. Feature-based approaches generally provide improved robustness compared with simple template matching and can tolerate moderate viewpoint and scale changes. However, examination documents often contain repetitive text structures, sparse distinctive features, and low-texture regions, making stable feature correspondence difficult. In low-resolution scanning or mobile camera environments, feature detection performance may further deteriorate, leading to unstable homography estimation and inaccurate answer localization.

To address these limitations, recent studies have explored marker-based calibration approaches using fiducial markers such as ArUco [3]. Fiducial markers provide explicitly designed visual reference points that can be robustly detected under varying illumination, perspective distortion, and partial occlusion conditions. Compared with generic feature-based matching, marker-guided calibration offers more stable and computationally efficient homography estimation for structured documents. Marker-based systems have demonstrated strong robustness in augmented reality, camera calibration, and document rectification applications. However, existing educational grading systems rarely integrate marker-guided alignment with examination metadata and answer region reconstruction within a unified framework.

In addition to alignment techniques, image preprocessing plays a critical role in examination grading systems. Document preprocessing methods such as grayscale normalization, denoising, adaptive thresholding, and binarization are widely used to improve OCR and OMR performance under noisy acquisition conditions [9,10]. Otsu thresholding [11] and Sauvola adaptive binarization [10] are among the most commonly applied techniques for document image enhancement. These preprocessing methods help improve marker detection, text segmentation, and answer extraction reliability when dealing with uneven illumination, low contrast, or scanning artifacts.

Several recent studies have proposed integrated examination processing systems combining image preprocessing, OCR, and answer analysis [12–15]. Nevertheless, most existing systems still rely heavily on fixed templates, manually configured coordinates, or cloud-based processing pipelines. Such approaches reduce flexibility, complicate deployment in privacy-sensitive educational environments, and limit robustness under practical acquisition conditions involving smartphones, low-resolution scanners, or unconstrained document capture.

The proposed framework differs from previous work in several important aspects. First, it tightly integrates examination generation and grading through shared JSON-based structural metadata. Second, it employs marker-guided homography estimation using ArUco markers for robust geometric calibration. Third, normalized coordinate reconstruction enables resolution-independent answer extraction without relying on hardcoded pixel coordinates. Finally, the entire pipeline operates fully offline while supporting both fully automatic and semi-automatic grading workflows. By combining marker-guided calibration, normalized coordinates, and metadata-driven reconstruction within a unified architecture, the proposed system addresses several limitations of existing paper-based examination grading approaches.

3. System Overview

The proposed framework consists of two integrated modules: examination generation and examination grading, sharing a JSON-based metadata structure. The generator produces a Question Paper, an Answer Sheet, and `exam_structure.json`. The Answer Sheet contains answer regions and ArUco markers. During grading, only the scanned Answer Sheet and metadata are required.

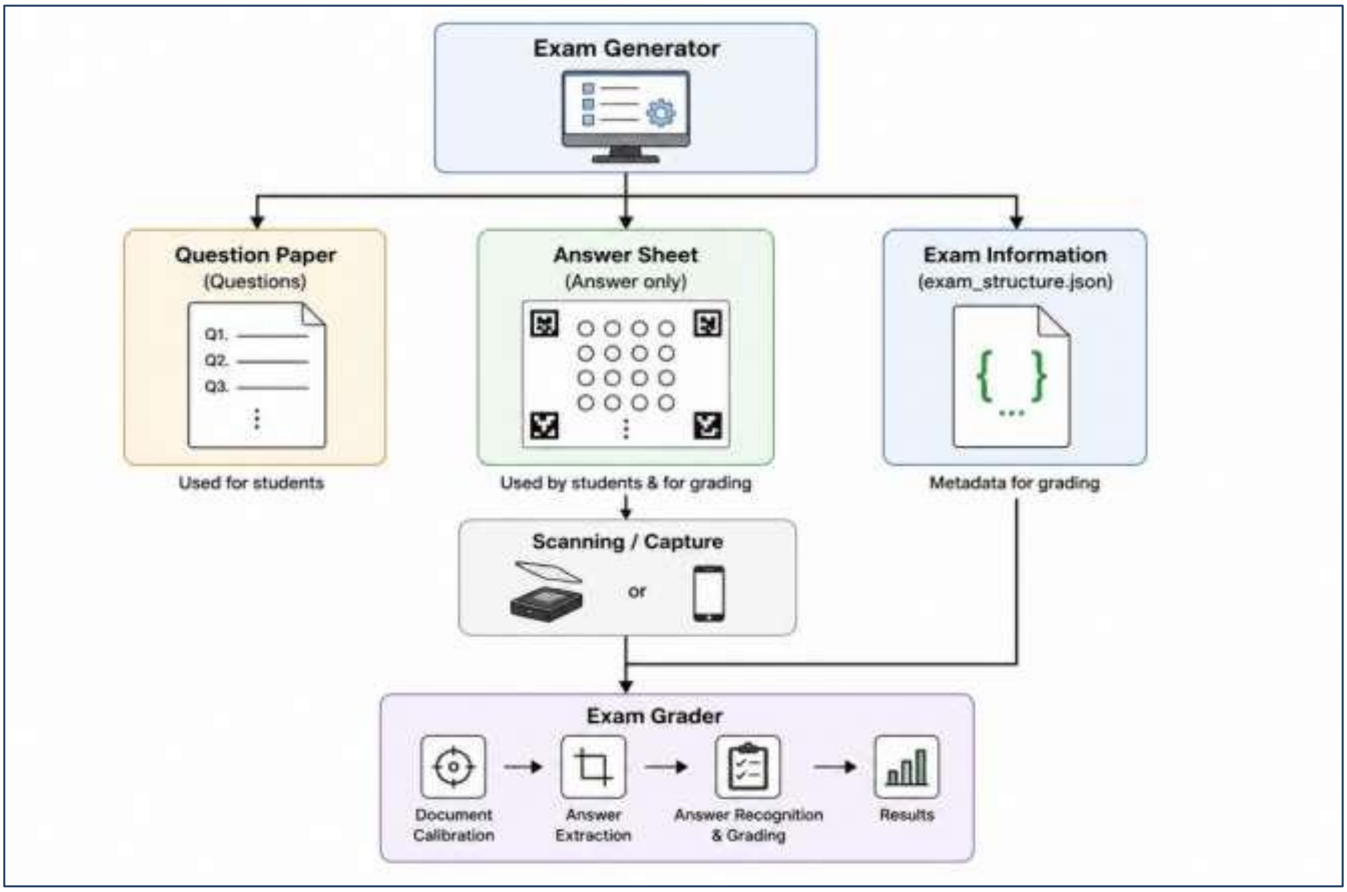

Figure 1: Overall architecture of the proposed offline grading framework

The grading pipeline includes preprocessing, marker detection, coordinate reconstruction, answer extraction, recognition, grading, and review management. The system operates entirely offline using computer vision and OCR technologies [2–4].
Marker-guided coordinate reconstruction uses detected ArUco markers to estimate homography, mapping normalized coordinates from metadata to actual image coordinates [8, 11], enabling resolution-independent extraction and distortion correction.
Question types are categorized by automation level: fully automatic (multiple-choice, true/false) using OMR; semi-automatic (matching, ordering, fill-in-the-blank, short answer) using OCR and similarity matching with optional instructor verification; and extensible types (essay, calculation, code) for future work.

4. Proposed Method

The pipeline consists of examination structure generation, document calibration, answer extraction, recognition, and grading.

4.1 Examination Structure Generation

The system generates a Question Paper, an Answer Sheet, and `exam_structure.json`. The metadata stores question types, correct answers, normalized answer coordinates, and score information.

4.2 Marker-guided Document Calibration

The system generates a Question Paper, an Answer Sheet, and `exam_structure.json`. The metadata stores question types, correct answers, normalized answer coordinates, and score information.

4.3 Image Preprocessing

The scanned Answer Sheet is converted to grayscale, denoised, and adaptively thresholded to improve marker detection and answer extraction [9, 10].

4.4 Normalized Coordinate Reconstruction

Answer regions are stored as normalized coordinates: $u = (x - x_{TL}) / (x_{BR} - x_{TL})$, $v = (y - y_{TL}) / (y_{BR} - y_{TL})$. After calibration, these are mapped to actual image coordinates, ensuring robustness to resolution and scaling changes [11, 12].

4.5 Answer Extraction and Recognition

Reconstructed answer regions are cropped and sent to OMR or OCR modules according to question type [1, 3, 4].

4.6 Question-type-aware Grading

Multiple-choice and true/false questions are graded automatically via OMR. Fill-in-the-blank and short-answer questions use OCR with similarity matching and optional instructor verification [3, 13]. Matching and ordering questions are also semi-automatic. Essay, calculation, and code writing are reserved for future extension.

4.7 Human Review Mechanism
Low-confidence OCR results are forwarded to instructors for manual verification, improving reliability.

4.8 Offline Modular Architecture
All processing — calibration, extraction, OCR/OMR, grading — runs locally without cloud services.

5. Implementation
The system is implemented in Python with OpenCV (for preprocessing, marker detection, homography, and extraction) [2], PyQt (for GUI) [14], and Tesseract OCR (for printed and short-answer text) [4]. It operates fully offline.
The central data structure is `exam_structure.json`, which contains question identifiers, question types, correct answers, score information, normalized answer region coordinates, and marker configuration information. Normalized coordinates eliminate hardcoded templates and enable device-independent grading.
The architecture is modular, with separate components for preprocessing, marker detection, coordinate reconstruction, answer extraction, OMR, OCR, grading, and review management. This design improves maintainability and extensibility.
The implementation source code, examination templates, and grading pipeline are publicly available through the project repository: https://github.com/kwolf-212/Auto-grading-offline-system

6. Experiments
We constructed a dataset of 1,200 examination sheets (180 students, 12 templates, 28,800 total questions) covering Multiple Choice, True/False, Matching, Ordering/Ranking, Fill in the Blank, and Short Answer question types. The dataset was designed to reflect realistic examination diversity without explicitly fixing a rigid per-type distribution at the dataset level, ensuring balanced evaluation across heterogeneous question formats. Images were acquired under various conditions, including flatbed and document scanners (72/150/300 DPI) and smartphone cameras (1280×720 to 4032×3024), with indoor/outdoor lighting variations, perspective distortion, rotation (±15°), and varying shooting distances. Essay, code-writing, and calculation questions were excluded from quantitative evaluation because they are reserved for future system extension.
Evaluation metrics were divided into fully automatic evaluation and semi-automatic evaluation depending on whether human intervention is required.
Fully Automatic Metrics: Grading accuracy, Intersection over Union (IoU) for region accuracy, Reprojection error (in pixels), Processing time
Semi-automatic Metrics: OCR accuracy, Answer similarity accuracy (for short-answer normalization and matching), Instructor correction rate (human-in-the-loop verification ratio)
Similar evaluation strategies are commonly used in computer vision and document analysis research [11,12].
Baselines for objective grading were selected from traditional objective-question processing methods, including conventional OMR and template matching [1, 7], both primarily designed for multiple-choice and other structured response formats, as well as ORB-based homography for document alignment and geometric correction.

Table 1: Answer Region Reconstruction Performance

Method	IoU	Reprojection Error
Template Matching	0.81	12.6
ORB-based Homography	0.88	7.4
Proposed Method	0.96	2.1

Answer region reconstruction performance is summarized in Table 1. The proposed method achieved the highest localization accuracy with an IoU of 0.96 and reprojection error of 2.1 pixels.

Table 2 presents grading accuracy across different question types. The proposed framework consistently outperformed conventional OMR and ORB-based approaches in all categories.

Table 2. Grading Accuracy by Question Type (%)

Question Type	Conventional OMR	ORB-based	Proposed Method
Multiple Choice / True-False	96.2	97.8	99.1
Matching	90.3	92.7	96.8
Ordering / Ranking	88.9	91.2	95.6
Fill in the Blank	91.5	93.4	97.3
Short Answer	82.4	86.7	92.5

Average processing time per examination sheet is reported in Table 3. The total processing time was 205 ms per sheet, demonstrating practical applicability for real-time offline grading environments.

Table 3. Average Processing Time per Sheet

Processing Stage	Time (ms)
Preprocessing	41
Marker Detection	28
Coordinate Reconstruction	16
Answer Extraction	35
OCR / OMR	73

Grading	12
Total	205

Robustness evaluation results under low-resolution scanning and mobile camera acquisition conditions are summarized in Table 4. The proposed method maintained stable performance under severe geometric and imaging variations.

Table 4. Robustness Evaluation under Challenging Acquisition Conditions (%)

Acquisition Condition	Template Matching	ORB-based	Proposed Method
72 DPI Low-resolution Scan	79.5	86.2	94.8
Mobile Camera Acquisition	74.2	85.7	95.4

Table 5 presents ablation study results. Removing ArUco-based calibration or normalized coordinate reconstruction significantly reduced localization and grading performance, confirming the importance of both components in the proposed framework.

Table 5. Ablation Study Results

Configuration	IoU	Reprojection Error
Full Proposed System	0.96	99.1
Without ArUco Calibration	0.84	88.7
Without Normalized Coordinates	0.87	90.4

7. Discussion

7.1 Advantages of the Proposed Framework

The proposed framework provides several key advantages over conventional systems.

First, separating question and answer documents simplifies grading and improves answer localization.

Second, marker-guided document calibration ensures robust alignment even under perspective distortion, low resolution, or repetitive document patterns. Unlike feature-based methods, the explicit ArUco references produce stable homography estimation [5-8].

Third, resolution-independent answer extraction via normalized coordinates allows the same metadata to operate across scanners, cameras, and varying DPIs without redesign [11, 12].

Fourth, integrating exam generation and grading through shared metadata eliminates manual template registration and reduces errors.

Fifth, fully offline operation preserves student privacy, avoids cloud dependency, and enables deployment in restricted or disconnected network environments.

Sixth, the modular and extensible architecture also allows future extension toward semantic grading and advanced handwritten recognition systems based on machine learning approaches [13,15].

Finally, the proposed framework distinguishes between fully automatic and semi-automatic grading tasks, enabling practical deployment while maintaining extensibility toward more advanced assessment types.

7.2 Limitations

The system depends on marker visibility; occlusion or severe blur may degrade calibration. Handwriting recognition remains challenging due to style variability and image quality, though human review mitigates this. The system lacks semantic understanding of descriptive answers, so lexically different but semantically equivalent answers may require manual review. Also, the exam must be generated within the framework, limiting use with arbitrary external documents.

In addition, the current system does not support essay-type responses or code-writing questions, as these require deeper semantic evaluation and programmatic correctness checking, which are beyond the scope of the current rule-based and OCR/OMR-centered grading pipeline.

7.3 Practical Applicability

The system is suitable for schools, universities, and offline exam centers. It works with low-cost scanners or smartphone cameras, reduces manual grading burden, and preserves data privacy via local processing. The semi-automatic workflow requests human review only for low-confidence cases.

7.4 Future Work

Future research will explore advanced handwritten recognition, semantic grading models, adaptive marker recovery for severe degradation, and multi-page/distributed grading support, all while maintaining offline operation.

8. Conclusion

This paper presented an offline paper-based exam grading system using marker-guided coordinate reconstruction and template-aware answer extraction. By integrating exam generation and grading through shared metadata, and by employing ArUco markers and normalized coordinates, the system achieves robust answer localization under practical acquisition conditions. Experiments demonstrated superior reconstruction accuracy (IoU 0.96, reprojection error 2.1 px) and grading accuracy (up to 99.1% for objective questions) compared to conventional methods. The system is fully offline, modular, and supports multiple question types with human-in-the-loop review, making it suitable for real educational environments where dedicated hardware or cloud infrastructure is unavailable.

The current implementation demonstrates robust automatic grading for objective paper-based question types and provides OCR-assisted grading support for short textual answers. In addition, the modular architecture establishes a scalable foundation for future extension toward essay, calculation, and code-writing assessment.

The proposed framework separates question presentation and grading by independently generating Question Papers and Answer Sheets while performing grading using only scanned Answer Sheets and structured examination metadata.

References

[1] O'Gorman, L. (1993). The document spectrum for page layout analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(11), 1162–1173.

[2] Smith, R. (2007). An overview of the Tesseract OCR engine. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)** (pp. 629–633).

[3] Salvi, D., Barata, J.-C., Cobos, S., & Diaz, M. (2014). Automatic localization and recognition of fiducial markers for augmented reality applications. *Pattern Recognition, 47*(6), 2280–2292.

[4] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2564–2571).

[5] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

[6] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In *European Conference on Computer Vision (ECCV)* (pp. 404–417).

[7] Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press.

[8] Szeliski, R. (2011). *Computer vision: Algorithms and applications*. Springer.

[9] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

[10] Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition, 33*(2), 225–236.

[11] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 62–66.

[12] Yalçınkaya, Ü. Z., & Eren, T. (2012). An automated paper-based exam evaluation system. *Computers & Education, 59*(2), 596–603.

[13] Bulut, M. F., Erol, Y. K., & Akin, H. K. (2014). Optical mark recognition and processing system for examination papers. *International Journal of Computer Applications, 95*(10), 1–7.

[14] Sharma, S. K., & Nandgaonkar, A. N. (2017). Automated answer sheet evaluation system using image processing. *International Journal of Advanced Research in Computer Science, 8*(5), 1124–1129.

[15] Chinnasamy, R., & Suresh, P. (2018). Automatic evaluation of objective answer scripts using computer vision techniques. *Journal of Electronic Imaging, 27*(5), 053021-1–053021-11.

[16] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

[17] Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.

[18] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). Wiley.

[19] Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media.

[20] Wolf, C., & Jolion, J.-M. (2004). Extraction and recognition of artificial text in multimedia documents. *Pattern Analysis & Applications, 6*(4), 309–326.

[21] Ha, T. M., & Bunke, H. (1997). Off-line handwritten text recognition using hidden Markov models. *Pattern Recognition, 30*(12), 2157–2168.