Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

ISTI Technical Reports

CNR activity in the ESA Extension project

Claudio Vairo, CNR-ISTI, Pisa, Italy

Paolo Bolettieri, CNR-ISTI, Pisa, Italy

Claudio Gennaro, CNR-ISTI, Pisa, Italy

Giuseppe Amato, CNR-ISTI, Pisa, Italy

ISTI-TR-2023/010

CNR activity in the ESA extension project
Vairo C.; Bolettieri P.; Gennaro C.; Amato G.
ISTI-TR-2023/010

Abstract
The CNR activity within the ESA "EXTENSION" project aims to develop an advanced visual recognition system for cultural heritage objects in L'Aquila, using AI techniques such as classifiers. However, this task requires substantial computational resources due to the large amount of data and deep learning-based AI techniques involved. To overcome these challenges, a centralized approach has been adopted, with a central server providing the necessary computational power and storage capacity.

Visual recognition, Artificial Intelligence, Computer vision, Deep Learning, Cultural Heritage
Citation

Vairo C.; Bolettieri P.; Gennaro C.; Amato G. *CNR activity in the ESA extension project* ISTI Technical Reports 2023/010. DOI: 10.32079/ISTI-TR-2023/010

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1
56124 Pisa Italy
http://www.isti.cnr.it

CNR activity in the ESA Extension project

Claudio Vairo[1], Paolo Bolettieri[1], Claudio Gennaro[1], Giuseppe Amato[1]

[1]*ISTI-CNR, via G. Moruzzi, 1, Pisa, 56100, Italy*

Introduction

The primary aim of the CNR activity within the ESA Extension project is to develop a cutting-edge visual recognition system for cultural heritage objects (such as monuments, statues, and paintings) in the city of L'Aquila. This system utilizes advanced Artificial Intelligence (AI) techniques, including classifiers, to analyze video streams or still images, effectively identifying and recognizing the objects of interest depicted in the media.

Visual recognition, however, often demands substantial computational capabilities and resources. The vast amount of data generated by the features of objects to be recognized must be pre-processed and stored efficiently. Furthermore, the extraction of features from query images necessitates the employment of deep learning-based AI techniques, requiring a machine outfitted with a modern graphics card.

To address these challenges, a centralized approach has been adopted, wherein a central server provides the necessary computational power and information storage capacity for effective recognition. This strategy streamlines the recognition process and ensures optimal utilization of resources.

The CNR's efforts primarily focus on work packages WP2100, WP2200, WP3100, and WP4400, the details of which are outlined below.

User Needs (WP2100)

Understanding the needs and preferences of users is crucial for the development of a successful cultural heritage visual recognition system. To ensure our solution effectively addresses these requirements, we have conducted a thorough analysis of existing research, real-world feedback, and our own experiences in similar projects. The following section outlines the user needs identified through this comprehensive investigation, which will guide the design and implementation of our proposed solution.

We have identified a comprehensive list of user needs by examining various scientific publications that focus on cultural heritage applications and drawing on our previous experience developing visual recognition solutions for the POR CREO FESR 2007-2013 project "VISITO Tuscany." This project shares similar usage scenarios with the ones we aim to develop here.

In these scientific documents, researchers conducted interviews with actual tourists to gather insights. The interviews revealed that tourists appreciate not only images of attractions but also contextual textual descriptions to enrich their experience beyond simply observing the monument and taking a picture. They desire more information about the sites they visit but often lack access to such resources. By recognizing the monument in question and providing pre-computed textual and visual information, we can address this need.

Additionally, tourists expressed interest in interactive experiences with the monuments, which can be facilitated through Augmented Reality and 3D model reconstruction. Another common request was for a suggested itinerary featuring the best attractions in a given location, ideally with an optimized path.

Moreover, users showed willingness to receive commercial recommendations for top restaurants, souvenir shops, or local specialty stores, as well as transport information to assist them in visiting attractions. Lastly, some users wanted to be informed about any cultural events scheduled in the nearby area during their stay.

From these sources, the following needs have been identified:

- UN-0300: Users wish to have a more entertaining, immersive and possibly interacting tourist experience.
- UN-0400: Users wish to get instant information, either textual, visual or audio, on their end device (for example their smartphone) about what they are seeing.
- UN-0500: Users wish to receive a proposed itinerary to visit all the most interesting attractions in the city.
- UN-0600: Users wish to receive commercial suggestions about the best restaurants and shops as well as information about transport.
- UN-0700: Users wish to receive information about cultural events that are available in situ.

The list of scientific publications and technical reports used to obtain the presented user needs is:

1. VISITO Tuscany - Documento A1.3.1 - Scenari applicativi e selezione dei requisiti applicativi
2. VISITO Tuscany - Documento A1.2 - Requisiti Utente
3. Damala and N. Stojanovic, “Tailoring the adaptive augmented reality(A^2R) museum visit: Identifying cultural heritage professionals’ motivations and needs,” in the 2012 IEEE international symposium on mixed and augmented reality-arts, media, and humanities (ISMAR-AMH). IEEE,2012, pp. 71–80.
4. M. T. Linaza, D. Marim´on, P. Carrasco, R.´Alvarez, J. Montesa, S. R. Aguilar, and G. Diez, “Evaluation of mobile augmented reality applications for tourism destinations,” in Information and communication technologies in tourism 2012. Springer, 2012, pp. 260–271.
5. D.-I. Han, T. Jung, and A. Gibson, “Dublin AR: implementing augmented reality in tourism,” in Information and communication technologies in tourism 2014. Springer, 2013, pp. 511–523.
6. C. d. l. N. A. Brito, “Augmented reality applied in tourism mobile applications,” in the 2015 Second International Conference on eDemocracy & eGovernment (ICEDEG). IEEE, 2015, pp. 120–125.
7. R. Abd Rashid, H. Mohamed et al., “Mobile augmented reality tourism application framework,” in the International Conference of Reliable Information and Communication Technology. Springer, 2017, pp. 108–115
8. M. Williams, K. K. Yao, and J. R. Nurse, “ToARist: An augmented reality tourism app created through user-centred design,” arXiv preprintarXiv:1807.05759, 2018.
9. D.-I. Han and T. Jung, “Identifying tourist requirements for mobile AR tourism applications in urban heritage tourism,” in Augmented Reality and Virtual Reality. Springer, 2018, pp. 3–20.

System/Service Requirements (WP2200)

Based on the identified user needs, we have compiled a list of corresponding user requirements that detail the functionalities the app should offer upon development.

Specifically, when a user takes a picture of a monument, they should be able to access the monument's name and relevant textual, visual, or audio information, such as its history, historical images, a 3D model, or interesting trivia. Some of this information can be displayed as augmented reality pop-ups, allowing users to interact with the content.

Furthermore, the app should provide users with an optimized route encompassing all or some of the chosen monuments and sites. It should also offer commercial recommendations for the best nearby restaurants, affordable souvenir shops, and helpful transport information to facilitate visiting designated monuments. If any cultural events are scheduled nearby, the app should supply users with instructions on attending these events, including pertinent details like time, location, and price.

To identify user requirements, we employed various methodologies. First, we considered the scientific literature mentioned in the User Needs section. Next, we conducted surveys and held meetings with users. In both cases, user requirements were seen as direct derivatives of the established user needs, allowing us to formulate questions seeking feedback from users based on user needs analysis. This approach enabled us to narrow the research scope and seek additional specifications that align with the previously expressed and identified needs, ensuring consistency between the application call, user needs, and the end-to-end service defined in the Proposal as the project's primary deliverable.

Consequently, we created three lists of questions based on each user need, with one list per user group. The answers to these questions logically lead to the requirements referred to as user requirements. In this report, we only present questions related to CNR's activities within the project. The mapping of these outcomes is summarized below:

- UN - 0300: Users wish to have a more entertaining, immersive and possibly interacting tourist experience. → Related questions: Would you like to visit a monument or museum in Augmented Reality? Would you download an app that allows you to visit a city in Augmented Reality? Do you think it would be useful to offer a solution accessible on site? Do you find it difficult to find information while traveling? What information would you like to receive when traveling? → Answers leading to UR-0301 and UR-0103.
- UN-0400: Users wish to get instant information, either textual, visual or audio, on their end device (for example their smartphone) about what they are seeing. → Related questions: Which kind of information would you like to receive if you download a tourism app? Do you find audio guides useful when visiting a monument or museum? Would you like to enjoy audio/visual/textual content during a city tour? Would you like to participate in a live music event? → Answers leading to UR-0401; UR-1000.
- UN-0500: Users wish to receive a proposed itinerary to visit all the most interesting attractions in the city. → Related questions: Would you like to use an app with personalized itineraries based on your tastes? Would you prefer short and local itineraries or long national itineraries based on your tastes? → Answers leading to UR-0501.
- UN-0600: Users wish to receive commercial suggestions about the best and shops as well as information about transport. → Related questions: Would you like to receive commercial information while travelling? Would you like to receive suggestions about the best places to go to? Would you like to have information regarding the departure time of buses and similar? → Answers leading to UR-0601.
- UN-0700: Users wish to receive information about cultural events that are available on-site. → Related questions: Would you like to receive information regarding a particular kind of cultural event? Would you like to receive information on every kind of cultural event nearby? → Answers leading to UR-0701.

Design Scenario and Contents (WP3100)

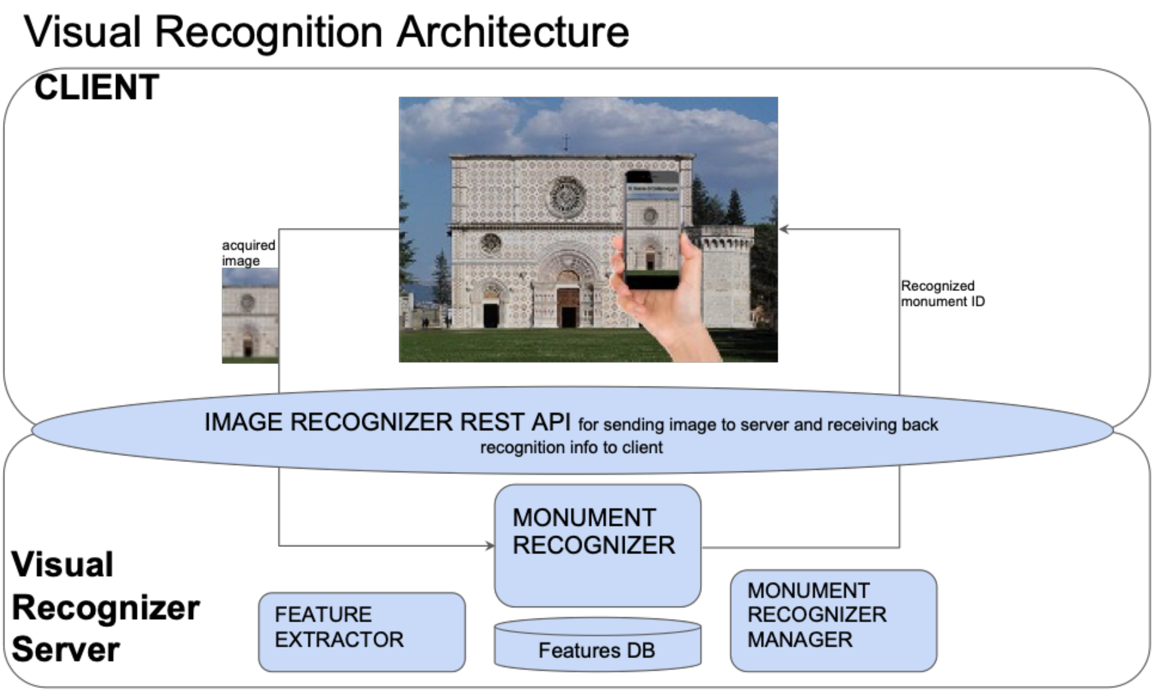
Visual Recognition System/Service Architecture

The Visual Recognition (VR) system is a component responsible for the visual recognition of monuments. In a typical use case, a tourist standing in front of a monument captures an image with their smartphone to obtain additional information and content about the monument.

The recognition process is based on a k-NN classifier, employing artificial intelligence and computer vision techniques to analyze and extract a mathematical representation (feature) of an image depicting a monument. All monument images that make up the training set are analyzed during an offline phase before deploying the application and stored within a component of the VR system. At runtime, when an image is received for a recognition task, the feature extracted from the query image is used to perform a k-NN classification of all the known monument features stored in the VR system to determine if the query monument is among the known ones. In this case, the identifier for that monument is returned. If the query image is not recognized, such as when the query image does not depict one of the known monuments, an 'unknown' message is returned.

The VR system operates on the server-side of the application and is invoked when the user captures an image of the monument they wish to recognize. The mobile application then calls a REST API used to send the acquired image to the VR system. Once the recognition is complete, the same API is used to return the identifier of the recognized monument or "unknown" if the received image has not been recognized as one of the known monuments.

The VR system consists of several internal components responsible for executing different tasks required for monument recognition. Below, we present the architecture of the VR system and a brief description of the functions carried out by each component:



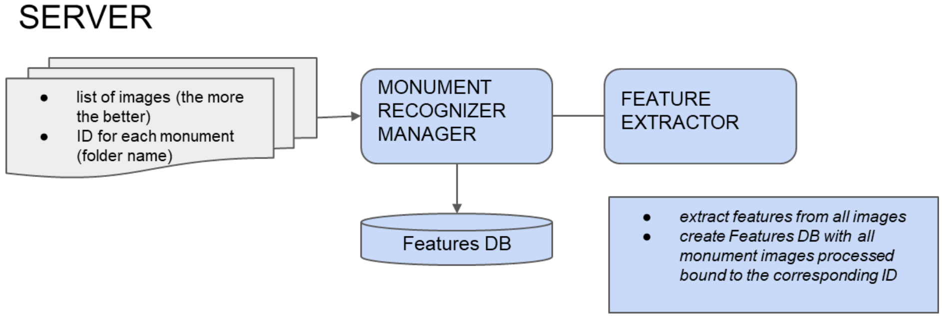
- The *Feature Extractor* component is responsible for analyzing an input image and extracting a mathematical description of the image using state-of-the-art Artificial Intelligence and Computer Vision techniques. This description will be stored and used at runtime to recognize the monument.
- The *Feature DB* stores the features extracted by the Feature Extractor component for all monument images to be recognized, acquired by Carsa, and sent to CNR.
- The *Monument Recognizer* component is responsible for analyzing a user-acquired image at runtime, meaning an image that has never been seen before by the recognizer, and performing the recognition process. This component utilizes the Feature DB and Feature Extractor components.
- The *Image Recognizer REST API* enables the mobile app running on the end device to send an input image to the remote VR system and receive the ID of the recognized monument in return.

Offline Training Process

The Offline Training Process in this architecture is a crucial step for preparing the Visual Recognition (VR) system to recognize monuments effectively. This process involves the collection and analysis of images and their corresponding monument IDs, extraction of features, and population of the Feature DB. The following is a detailed description of the Offline Training Process:

1. Image and ID Collection: The first step in the Offline Training Process is to gather a comprehensive set of images representing the monuments to be recognized. Each image is associated with a unique monument ID, which serves as a reference for identification during the recognition process.
2. Feature Extraction: Once the images and corresponding monument IDs are collected, the Feature Extractor component is employed. This component analyzes each input image using state-of-the-art Artificial Intelligence and Computer Vision techniques. It then extracts a mathematical representation, or feature, of the image that encapsulates its essential visual characteristics. This feature extraction process is vital for enabling the VR system to distinguish and identify different monuments.
3. Populating the Feature DB: After extracting features from all the monument images, the next step is to populate the Feature DB. The Feature DB serves as a storage repository for the extracted features and their corresponding monument IDs. By organizing and indexing the features and IDs, the Feature DB ensures efficient retrieval and comparison during the runtime recognition process.

In summary, the Offline Training Process is a critical aspect of the VR system architecture, enabling it to recognize monuments effectively. The process entails collecting images and their corresponding monument IDs, extracting features using the Feature Extractor component, and populating the Feature DB with these extracted features and associated monument IDs. This pre-processing step lays the foundation for accurate and efficient monument recognition during runtime.



Cultural Heritage Diagnosis (WP4400)

Image Recognizer Rest API

In this section, we report the functions that can be used to start the recognition process of a monument by the Monument Recognizer.

The image to be recognized can be sent directly with a multipart POST call or with a BaseB4 encoded POST, or it can be sent by passing the URL of an image on the Internet with a GET call.

In all cases, it is necessary to include a default API authorization ID that identifies the recognition app for security purposes.

Input functions

Search by multipart image

Description: Recognize an image received by a Multipart POST

Web Method: searchByImg

Request: POST Multipart

URL: *http://bilioso.isti.cnr.it:8190/bcir/searchByImg*

Parameters:

- **image**: query image
- **securityID**: API authorization ID

Return: a JSON document containing the recognized class and the recognition score

Search by Base64 image

Description: Recognize a Base64 encoded image received by a POST

Web Method: searchByImg

Request: POST

URL: *http://bilioso.isti.cnr.it:8190/bcir/searchByImgB64*

Parameters:

- **image**: query image in Base64 format
- **securityID**: API authorization ID

Return: a JSON document containing the recognized class and the recognition score

Search by image URL

Description: Recognize an image received by a GET

Web Method: searchByURL

Request: GET

URL: *http://bilioso.isti.cnr.it:8190/bcir/searchByURL*

Parameters:

- **url**: image query URL
- **securityID**: API authorization ID

Return: a JSON document containing the recognized class and the recognition score

Output Result

As a response to the invocation of the functions described above, the Monument Recognizer at the end of the recognition process will return a JSON document containing the class of the recognized monument (the name of the monument) and a confidence value of the prediction between 0 and 1. In case the analyzed image is not recognized, the class returned in the response JSON will be "unknown".

JSON Format

{

"classname": *name of the recognized class or keyword* "**unknown**",

"conf": *classification confidence. A value between* **0** *and* **1**

}

Example 1

Recognition of the Collemaggio class, with confidence 0.611

{

"classname":"Collemaggio",

"conf":0.611

}

Example 2

Response in case the query image could not be recognized

{

"classname":"unknown",

"conf":1.0

}

Pilot Running (WP5200) and KPIs Assessment (WP5300)

Factory Acceptance Testing (FAT)

The VR system was tested in both the Factory Acceptance Testing (FAT) phase and the Site Acceptance Testing (SAT) phase. However, since the FAT verification was conducted in a laboratory without access to real monuments, the test procedures for the two phases were differentiated.

In the FAT phase, printed images of the monuments to be recognized were prepared. During the test, the user started the recognition functionality inside the mobile application, then took a picture of the image representing the monument to be recognized, initiating the recognition process. Once completed, the recognition result was displayed on the user interface of the mobile application.

In the SAT phase, on the other hand, the user went in front of the monument they wanted to recognize. The procedure was the same as in the FAT phase, with the difference being that instead of taking a picture of the printed image, the user took a picture of the real monument.

The following steps were executed to test the VR system, along with the requirements that were tested and their expected results:

- The recognition functionality was started inside the mobile application.
- The monument (the printed image during the FAT phase, the real monument during the SAT phase) to be recognized was focused and a picture was taken.
- The system automatically triggered the recognition process; once completed, the user saw the recognition result on the mobile application.

The requirements tested included:

1. The VR system ability to recognize the monument represented in the acquired picture if it was one of the monuments belonging to the list of known monuments.
2. The processing time to recognize the monument, excluding network delays to send the image to the server and receive back the result, should be less than two seconds.
3. The VR system ability to correctly recognize at least nine monuments.
4. The recognition accuracy for the known monuments should be at least 70%.

Site Acceptance Test (SAT)

The Extension On-Site Acceptance Test (SAT) focused on testing the VR system ability to recognize monuments during field trials, simulating real-world scenarios. The test aimed to evaluate the system performance in recognizing monuments outside of a controlled laboratory environment.

The following steps were executed to test the VR system during the Extension On-Site Acceptance Test (SAT):

- The user went to the location of a monument to be recognized from the list of known monuments.
- The user launched the mobile application and started the recognition functionality.
- The user focused the camera on the real monument and took a picture.
- The system automatically initiated the recognition process.
- Once the recognition process was completed, the user viewed the recognition result on the mobile application.

The requirements tested during the Extension On-Site Acceptance Test (SAT) included:

1. The VR system ability to recognize the monument represented in the acquired picture, even when taken in varying lighting conditions, angles, and distances.
2. The processing time to recognize the monument, excluding network delays to send the image to the server and receive back the result, should be less than two seconds.
3. The VR system ability to correctly recognize a diverse set of monuments under different environmental conditions.
4. The recognition accuracy for the known monuments should be at least 70% in a real-world setting.

During the Extension On-Site Acceptance Test (SAT), the VR system was evaluated in a field trial to ensure its effectiveness in recognizing monuments under real-world conditions. This test aimed to validate the system performance, accuracy, and efficiency outside of the laboratory environment and to confirm its suitability for practical use.

Achievement of Proposal Objectives

In this section, we highlight the successful achievement of the objectives outlined in the initial proposal and discuss how they were accomplished.

The primary objectives of the proposal were to develop a Visual Recognition (VR) system capable of accurately and efficiently recognizing monuments, enhance user experience in learning about monuments, and ensure the system adaptability and effectiveness under various conditions.

Accurate and Efficient Monument Recognition: The VR system was developed using state-of-the-art Artificial Intelligence and Computer Vision techniques, enabling it to recognize monuments with a high degree of accuracy. A comprehensive and diverse dataset was used in the Offline Training Process to train the system, ensuring its ability to identify a wide range of monuments under different conditions. Continuous optimization and fine-tuning of the algorithms and techniques resulted in improved recognition accuracy and reduced processing time.

Enhanced User Experience: The mobile application's user interface was designed with user experience in mind, allowing for easy navigation and seamless integration of the VR system. Users can quickly access the recognition functionality, capture images of monuments, and receive relevant information about the identified monuments. Rigorous testing procedures and feedback from users helped identify areas of improvement, enhancing the system's usability and overall effectiveness.

Adaptability and Effectiveness under Various Conditions: The Extension On-Site Acceptance Test (SAT) demonstrated the VR system ability to recognize monuments in real-world scenarios, accounting for factors such as lighting, angles, and distances. The system was tested in diverse environments to ensure its adaptability and effectiveness in recognizing monuments under different conditions. Lessons learned from these field trials informed further optimization of the system's algorithms and techniques, ensuring its robust performance in practical use.

By focusing on these objectives and employing an iterative development process, the team successfully achieved the goals outlined in the initial proposal. The VR system's development and testing phases have demonstrated its potential as a valuable tool for users seeking additional information about monuments. As a result, the system is poised to enrich the experiences of tourists and educators, enhancing their understanding and appreciation of historical and cultural landmarks.

Conclusion and Lesson Learned

In conclusion, this document has provided an in-depth overview of the Visual Recognition (VR) system, which is designed to recognize and identify monuments using state-of-the-art Artificial Intelligence and Computer Vision techniques. The architecture of the VR system, along with its various components, has been presented, emphasizing the importance of the Offline Training Process in preparing the system for accurate and efficient monument recognition.

The document also detailed the testing procedures for the Factory Acceptance Testing (FAT) and Site Acceptance Testing (SAT) phases. These testing phases were crucial in evaluating the system performance, accuracy, and overall effectiveness. The FAT phase assessed the system functionality in a controlled laboratory setting using printed images, while the SAT phase evaluated the system ability to recognize real monuments in field trials, simulating real-world scenarios.

Furthermore, the Extension On-Site Acceptance Test (SAT) was conducted to ensure the VR system performance in recognizing monuments under different environmental conditions, such as varying lighting, angles, and distances. This thorough testing process aimed to validate the system suitability for practical use and its adaptability to diverse situations.

Throughout the development and testing process, several lessons were learned, which are valuable for future projects and system improvements:

1. The importance of a comprehensive and diverse dataset for training the system, ensuring its ability to recognize a wide range of monuments under various conditions.
2. The need for continuous optimization and fine-tuning of the algorithms and techniques used in the system to improve recognition accuracy and processing time.
3. The value of rigorous testing procedures that simulate real-world scenarios, allowing for a better understanding of the system performance in practical use.
4. The necessity of considering environmental factors, such as lighting and weather conditions, which can impact the system recognition capabilities.
5. The significance of user experience and feedback in identifying areas of improvement and enhancing the system usability and effectiveness.

In summary, the development and testing of the VR system has demonstrated its potential as a valuable tool for users seeking additional information about monuments. The combination of cutting-edge technology, rigorous testing procedures, and the lessons learned during the development process ensures the system reliability and effectiveness in recognizing and identifying monuments in various conditions. Moving forward, the VR system is expected to enrich the experiences of tourists and educators alike, enhancing their understanding and appreciation of historical and cultural landmarks.