COMPUTER VISION ALGORITHM DESIGN IN IMAGE PROCESSING BASED ON PROJECTIVE GEOMETRY

Reference NO. IJME 1385, DOI: 10.5750/ijme.v1i1.1385

YG Kang*, School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, Hunan, 411104, China and **Di Zhao**, School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, Hunan, 411104, China

*Corresponding author. YG Kang (Email): 06158@hnie.edu.cn

KEY DATES: Submission date: 25.12.2023 / Final acceptance date: 27.02.2024 / Published date: 12.07.2024

SUMMARY

Image processing with computer vision, particularly in the realm of projective geometry, offers remarkable potential for various applications. Through the lens of projective geometry, images can be transformed, augmented, and reconstructed with precision, facilitating tasks such as image rectification, 3D reconstruction, and object tracking. Landmark estimation in computer vision is a vital task with broad applications across various domains. This process involves identifying key points or landmarks within images, enabling tasks such as facial recognition, object tracking, and gesture recognition. This paper, proposed a novel approach for landmark estimation in computer vision using Projective Geometry Landmark Estimation (PGLM). The proposed model aims to estimate the landmark features by a projective geometry model. With the estimation of the geometry features landmarks related to the facial, object, and medical images are computed. The PGLM model uses the point features for the location of the landmark features. In order to compare PGLM's performance to that of more conventional classification methods like Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), simulation analysis is carried out. From what we can see, PGLM routinely beats these alternatives when we compare their accuracy, precision, recall, and F1 score. The findings stated the effectiveness of PGLM as a promising approach for landmark estimation in image processing tasks, paving the way for further advancements in this domain.

KEYWORDS

Image processing, Computer vision, Landmark estimation, Classification, Machine learning, Projective geometry

NOMENCLATURE

PGLM	Projective Geometry Landmark Estimation
KNN	K-Nearest Neighbour
SVM	Support Vector Machine

1. INTRODUCTION

Image processing is a pivotal field at the intersection of computer science, mathematics, and engineering, focusing on the manipulation and analysis of digital images [1]. It encompasses a broad range of techniques aimed at enhancing, compressing, restoring, and interpreting visual data. From medical imaging to satellite imagery, image processing plays a crucial role in various applications across industries [2]. Fundamental operations such as filtering, edge detection, and segmentation are employed to extract meaningful information from images, enabling tasks like object recognition, pattern detection, and image classification. With advancements in machine learning and deep learning, image processing has witnessed remarkable progress, facilitating automated decision-making systems and unlocking new frontiers in areas like autonomous vehicles, robotics, and healthcare [3]. As technology continues to evolve, image processing remains at the forefront, continually innovating to meet the demands of an increasingly visual world. A rapidly evolving subfield of AI, computer vision seeks to provide computers the ability to perceive and comprehend visual data derived from the physical environment [4]. At its core, computer vision aims to replicate human visual perception by extracting meaningful insights from images or videos. Through the integration of algorithms, machine learning, and deep learning techniques, computer vision systems can perform a diverse array of tasks, including object detection, recognition, tracking, and image segmentation [5]. From self-driving cars navigating complex environments to facial recognition systems enhancing security measures, the applications of computer vision are vast and impactful across numerous industries [6]. As advancements in hardware capabilities and algorithmic sophistication continue to accelerate, computer vision is poised to revolutionize fields such as healthcare, manufacturing, agriculture, and more, ushering in an era of unprecedented automation, efficiency, and innovation [7].

Image processing plays a fundamental role in the development and advancement of computer vision systems [8]. It serves as the backbone for preprocessing and enhancing raw visual data, enabling more accurate and meaningful analysis by computer vision algorithms. Through techniques such as filtering, noise reduction, and feature extraction, image processing prepares images or video streams for further interpretation and understanding by computer vision models [9]. Moreover, image processing algorithms often provide essential building blocks for key computer vision tasks, including object detection, recognition, and segmentation [10]. By leveraging the synergy between image processing and computer vision, researchers and engineers can create robust and efficient systems capable of extracting valuable insights from visual data across various domains, ultimately driving innovation and progress in artificial intelligence applications [11]. Image processing serves as the foundational step in the pipeline of computer vision systems, acting as a crucial preprocessing stage that refines raw visual data before it undergoes higher-level analysis [12]. This preprocessing is essential because raw images or video frames captured by cameras or sensors often contain noise, artifacts, or inconsistencies that can hinder accurate interpretation by computer vision algorithms [13]. Image processing techniques address these issues by enhancing image quality, reducing noise, and extracting relevant features, thus improving the overall performance of subsequent computer vision tasks [14]. One of the primary functions of image processing in computer vision is to enhance the quality and clarity of images. This may involve techniques such as filtering, where specific frequencies or spatial patterns are amplified or attenuated to improve image sharpness or reduce blur [15]. Additionally, image enhancement methods can adjust brightness, contrast, and color balance to ensure optimal visibility of objects or features within the image.

Noise reduction is another critical aspect of image processing in computer vision. Noise can arise from various sources such as sensor imperfections, atmospheric conditions, or transmission errors [16]. In order to improve the signal-to-noise ratio and the reliability of subsequent computer vision analyses, image processing algorithms utilize techniques such as median filtering, Gaussian smoothing, or wavelet denoising to reduce noise while keeping crucial picture details [17]. Feature extraction is yet another vital role of image processing in computer vision. By identifying and isolating relevant visual patterns or structures, feature extraction algorithms facilitate tasks such as object detection, recognition, and segmentation. Techniques like edge detection, corner detection, and texture analysis help highlight distinctive characteristics within an image, enabling computer vision systems to differentiate objects, delineate boundaries, and extract meaningful information for further processing [18]. Furthermore, image processing algorithms often provide essential building blocks for advanced computer vision tasks. estimation [19]. These tasks involve analyzing multiple images or video frames to infer spatial relationships, depth information, or temporal dynamics, all of which rely on sophisticated image processing techniques for accurate and reliable results [20]. In symbiotic relationship between image processing and computer vision enables the development of robust and efficient systems capable of extracting valuable insights from visual data across diverse applications. By leveraging the power of image processing to preprocess and refine raw visual information, computer vision algorithms can achieve higher levels of accuracy, robustness, and versatility, ultimately driving innovation and progress in artificial intelligence.

The paper contributes to the field of computer vision and image processing in several significant ways:

1. The paper introduces a novel approach called Projective Geometry Landmark Estimation (PGLM) for landmark estimation in computer vision. PGLM leverages projective geometry principles to accurately estimate landmarks in images.
2. The paper demonstrates that PGLM outperforms traditional classification methods such as SVM, K-Nearest Neighbors and Random Forest in terms of accuracy, precision, recall, and F1-score.
3. PGLM offers a robust and efficient solution for landmark estimation, making it suitable for various applications including facial recognition, object recognition, and medical imaging.
4. The scalability and generalizability of PGLM across different datasets highlight its potential as a versatile tool in the field of computer vision. This suggests that PGLM can be applied to a wide range of real-world scenarios with varying complexities.
5. The findings of the paper pave the way for further advancements in landmark estimation and image processing. The effectiveness of PGLM opens up avenues for future research aimed at exploring its applications in domains such as medical imaging, surveillance, and autonomous navigation.

The paper's contribution lies in introducing a novel and effective approach for landmark estimation in computer vision, thereby advancing the state-of-the-art in image processing techniques.

2. RELATED WORKS

The use of image processing methods to strengthen and improve the efficiency of vision-based systems has been the subject of a great deal of research in computer vision. Researchers have investigated various methodologies for preprocessing raw visual data, aiming to improve feature extraction, noise reduction, and overall image quality. Research by Noori et al. (2022) on English as a Foreign Language (EFL) classes and instructors at universities provides insight into how to best use social media to improve students' language acquisition. Similarly, Yu et al. (2022) explored the mobile learning model for student engagement with social media for the English learning contexts, offering insights into leveraging technology for effective language instruction. Muftah (2022) examined the contribution of the English language on the social media platform during the pandemic of COVID for the language education. Additionally, Alenezi and Brinthaupt (2022) provide perspectives on the use of social media as a learning tool from students in the Faculty of Education, offering valuable insights into student perceptions and preferences regarding digital learning environments. These studies collectively contribute to the understanding of how technology, particularly social media, can be harnessed to support language learning and teaching initiatives, which could inform the development of tailored interventions for mixed-ability EMI learners in journalism and communication majors at universities in China.

Furthermore, the exploration of digital learning technologies and their adoption among university students by Sayaf et al. (2022) and the investigation into the use of multimedia presentations for developing linguistic and digital literacy skills by Yu and Zadorozhnyy (2022) provide additional perspectives on leveraging technology-enhanced learning in language education contexts. Additionally, the study by Tarasenko et al. (2022) delves into the use of augmented reality (AR) elements in foreign language study at the university level, offering innovative approaches to engage learners and enhance language acquisition. Moreover, the examination of specialized dictionary mobile apps for English learners in specific fields like engineering, business, and computer science by Al-Jarf (2022) highlights targeted resources that can support EMI learners in journalism and communication majors with domain-specific terminology and vocabulary. These diverse studies collectively underscore the significance of integrating technology-driven pedagogical approaches to accommodate mixed-ability EMI learners' needs in specialized academic contexts, offering valuable insights and potential strategies for the action research project aimed at enhancing language learning outcomes for students at a university in China.

Further, EMI learners' engagement and motivation in communication and journalism majors can be supported by the findings of the study by Ramzan et al. (2023) on using social media to increase academic motivation among college-level ESL students. Lai et al. (2022) delves into how college students utilize mobile devices for self-directed language learning, providing insightful views on how learners can harness technology to control their own language learning journey. Moreover, Zhang and Chen (2022) delve into modeling technology use among university EFL teachers, shedding light on the factors influencing educators' adoption of technology in language instruction, which could inform strategies for faculty development in supporting mixed-ability EMI learners. These studies collectively contribute to the understanding of how technology can be effectively leveraged to support language learning and teaching initiatives, offering potential avenues for the action research project to develop tailored interventions that address the diverse needs of EMI learners in journalism and communication majors at a university in China.

Fannakhosrow et al. (2022) compared traditional classroom methods with those that made use of information and communication technology (ICT), providing valuable insight into how different approaches to language instruction can inspire students to take an active role in their own education. Muthmainnah (2023) expands on the use of technology instructional design in learning, providing potential frameworks and methodologies for integrating technology into language education. Additionally, Srivani et al. (2022) examine the impact of Education 4.0 among engineering students for learning the English language, offering perspectives on the intersection of technology and language learning in specialized academic disciplines. These studies collectively underscore the importance of embracing technology-enhanced pedagogical approaches to meet the diverse needs of EMI learners in journalism and communication majors, providing valuable insights and potential strategies for the action research project at the university in China.

Moreover, Sartono et al. (2022) explore the use of interactive multimedia based on Indonesian cultural diversity in civics learning, which highlights the potential of culturally relevant and interactive content to engage learners effectively. This approach could be adapted to incorporate cultural elements relevant to EMI learners studying journalism and communication majors, enhancing their learning experiences. Additionally, the study by Sofi-Karim et al. (2023) on online education via media platforms and applications as an innovative teaching method offers valuable insights into leveraging online platforms to facilitate language learning, especially in light of the increasing importance of remote learning due to global events such as the COVID-19 pandemic. These findings contribute to the ongoing discourse on effective pedagogical strategies for supporting mixed-ability EMI learners, providing practical recommendations and approaches that can be tailored to the specific context of journalism and communication majors at the university in China. Teaching English to students with a range of abilities in a journalism and communication program may benefit from the findings of Nazarov's (2022) research on language instruction at a technical university. Understanding effective teaching strategies in technical fields could provide valuable perspectives on how to engage students with diverse backgrounds and skill levels in language learning. The action research project at the Chinese university can improve the language skills and academic performance of mixed-ability English as a second language (ESL) students majoring in communication and journalism by combining the results of these various studies.

Furthermore, Noori et al. (2022) studies the utilization of social media for English as a foreign language (EFL) instruction and learning in Afghan universities. Findings may still provide light on how social media platforms can be used to support language learning and teaching initiatives, even though the setting is different from a Chinese university. Developing strategies that are specific to the needs of English as Second Language (ESL) learners in journalism and communication majors requires an understanding of the advantages and disadvantages of using social media in language instruction.

Furthermore, Yu et al. (2022) investigates how social media and mobile learning technologies influence student engagement and learning outcomes in English language learning settings. This research offers valuable insights into the potential benefits of incorporating mobile technology and social media platforms into language instruction. By understanding how these tools can enhance student engagement and facilitate learning, educators can develop innovative approaches to support mixed-ability EMI learners in journalism and communication majors, ultimately improving their language proficiency and academic success. To better understand student preferences and experiences with digital learning environments, Alenezi and Brinthaupt (2022) conducted an investigation into the views of Kuwait University Faculty of Education students on the utilization of social media for educational purposes. Understanding student perceptions can help educators tailor instructional approaches that resonate with EMI learners in journalism and communication majors, fostering a more conducive learning environment. Furthermore, Sayaf et al. (2022) investigates what variables impact college students' use of online learning tools for both instruction and assessment. Findings from this study illuminate potential difficulties that teachers may face when introducing technology-enhanced lessons. By addressing these factors and leveraging insights from the study, educators can develop strategies to effectively integrate digital learning technologies to support mixed-ability EMI learners in journalism and communication majors, thereby enhancing their language learning experiences and outcomes. Further, Sofi-Karim et al. (2023) research on online education through media platforms and applications as a novel approach to teaching provides helpful information for making use of digital tools for language learning. This study highlights the potential of online resources and media platforms to enhance language learning experiences, particularly in remote or hybrid learning environments. By incorporating innovative teaching methods and online resources into language instruction, educators can cater to the diverse needs of EMI learners in journalism and communication majors, promoting active engagement and improving language proficiency.

Researchers Yu and Zadorozhnyy (2022) found that using multimedia presentations to improve students' language and digital literacy skills was an effective way to incorporate technology into language classes. Teachers can engage students of varying abilities and learning styles by incorporating multimedia presentations and tools into their lessons. This approach can help enhance EMI learners' language skills while also promoting digital literacy, essential for success in today's interconnected world.

Ramzan, Javaid, and Fatima (2023) explore the use of social media as a tool to boost academic motivation among ESL students in higher education. The study investigates how social media platforms can be leveraged to engage and motivate ESL students, ultimately enhancing their academic performance and learning outcomes. By harnessing the interactive and collaborative features of social media, the authors aim to empower ESL students to actively participate in their learning process, connect with peers and instructors, and access educational resources more effectively. This research contributes to the understanding of innovative approaches to support ESL students in higher education settings, shedding light on the potential benefits of integrating social media into language learning and teaching initiatives. Lai, Saab, and Admiraal (2022) investigate the utilization of mobile technology by university students for self-directed language learning. In order to comprehend what variables impact students' adoption and utilization of mobile technology in language learning settings, this study utilizes the integrative behavior prediction model. By examining students' perceptions, attitudes, and behavior towards mobile technology, the authors aim to provide insights into effective strategies for promoting self-directed language learning through mobile devices. By contributing to what is already known about the role of mobile technology in facilitating autonomous language learning, the results of this study have significant ramifications for policymakers and educators aiming to enhance language learning opportunities in higher education. Focusing on the variables that influence the adoption and integration of technology in language instruction, Zhang and Chen (2022) examine the patterns of technology use among Chinese university EFL instructors. Examining how instructors' affective attitudes, evaluative attitudes, and Technological Pedagogical Content Knowledge (TPACK) influence their technology use behaviour, this study seeks to shed light on the topic. By employing a dichotomous approach to technology use modeling, the authors aim to provide a nuanced understanding of EFL teachers' technology integration practices and the factors that contribute to their decision-making process. This research contributes to the literature by offering insights into the complex interplay between teacher beliefs, attitudes, and technology use behavior, informing strategies for enhancing technology integration in EFL instruction.

3. IMAGE PROCESSING IN COMPUTER VISION

Image processing in computer vision refers to the manipulation and analysis of digital images to extract meaningful information and insights. In the context of computer vision, which aims to enable machines to interpret and understand visual data, image processing plays a fundamental role in preprocessing raw images to enhance their quality and extract relevant features. This preprocessing step typically involves operations such as filtering, noise reduction, and feature extraction, which are essential for improving the accuracy and reliability of subsequent computer vision algorithms. Image processing techniques are used to address various challenges in computer vision tasks, including object detection, recognition, tracking, and segmentation. By leveraging image processing algorithms, computer vision systems can effectively interpret visual data and make informed decisions, enabling a wide range of applications across industries such as healthcare, automotive, surveillance, and robotics. Overall, image processing is a critical component of computer vision systems, enabling them to extract valuable insights from visual data and perform complex tasks autonomously. An essential step in image processing, filtering may bring out or hide specific details in a picture. A typical filter type is the linear filter, which calculates each output pixel by adding up the values of its nearby input pixels in a weighted manner. Let's consider a simple linear filter known as a convolution filter. An input image I(x,y) and a filter kernel F(u,v), the output image O(x,y) after convolution is calculated ausing equation (1)

$$O(x,y) = \sum_{u,v} I(x-u, y-v).F(u,v) \tag{1}$$

In equation (1) (u,v) represents filter kernel coordinates, and (x,y) represents output image coordinates. The input image as I(x,y). The Sobel operator consists of two separate filters, Gx and Gy, which are applied to the image to estimate the gradients.

The Sobel operator for Gx is estimated using equation (2)

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

The Sobel operator for Gy is defined in equation (3)

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3}$$

To compute the gradient magnitude M(x,y) at each pixel estimated in equation (4)

$$M(x.y) = \sqrt{(G_x * I)^2 + (G_y * I)^2} \tag{4}$$

In equation (4) * * denotes the convolution operation. The convolution operation between an image (I) and a filter/kernel (K) is defined in equation (5)

$$I_{\text{filtered}}(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} I(i,j).K(x-i, y-j) \tag{5}$$

Equation (5) states The image's filtered value at pixel location (x, y) is denoted as Ifiltered(x, y). The intensity value of the original image at pixel location (i, j) is denoted as I(i, j). The value of the filter or kernel at relative position (x-i, y-j) from the center is K(x−i, y⅒j). As part of the convolution process, the filter or kernel is slid over the picture, and the element-wise multiplication between the filter and the relevant picture pixels is computed. Then, the results are summed up to obtain the output value at each pixel location in the filtered image.

3.1 PROJECTIVE GEOMETRY LANDMARK ESTIMATION (PGLM)

Projective Geometry Landmark Estimation (PGLM) is a technique used in computer vision for estimating landmarks or key points in images by leveraging principles from projective geometry. In computer vision tasks such as object recognition, facial recognition, and pose estimation, accurately detecting and localizing landmarks is crucial for subsequent analysis and decision-making processes. Projective geometry deals with the study of geometric properties that are preserved under projective transformations. In the context of computer vision, projective geometry provides a mathematical framework for understanding the relationships between points, lines, and planes in images and their corresponding real-world objects. In projective geometry, points are represented using homogeneous coordinates, which allow for convenient representation of points at infinity. The homogeneous coordinate system uses a scaling factor w to transform a point (x, y) in Euclidean space into a set of three points: [x, y, w]. Projective transformations between 3D space and 2D image plane are represented using projection matrices. The projection matrix P maps points from 3D homogeneous coordinates (X, Y, Z, W) to 2D homogeneous coordinates (x, y, w) on the image plane. an image with detected feature points, the goal is to estimate the 3D coordinates of these points in the real world. This involves solving the system of equations Ax = 0, where A is the matrix of homogeneous image coordinates and x is the vector of homogeneous real-world coordinates. The solution to the system Ax = 0 can be obtained using techniques such as Singular Value Decomposition (SVD) or Direct Linear Transformation (DLT). Once the solution is obtained, the 3D coordinates of the landmarks are estimated. Euclidean point (x, y) is represented as $[x, y, 1]$ in homogeneous

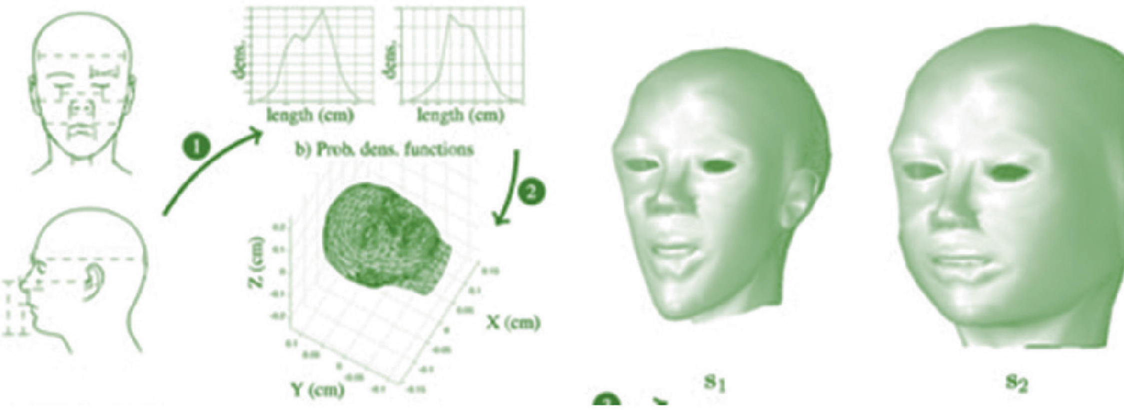


Figure 1. PGLM Landmark estimation

coordinates. Projection Matrix (P): $P = K\left[R \middle| T\right]$; K is the camera intrinsic matrix; R is the rotation matrix; T is the translation vector; Landmark Estimation: $Ax = 0$; A is the matrix of homogeneous image coordinates and x is the vector of homogeneous real-world coordinates. Figure 1 demonstrated the landmark estimation of features in the images.

Points in the plane of geometry are given homogeneous coordinates in projective geometry. The homogeneous representation of a point (x, y) in two-dimensional space is [x, y, 1]. Coordinates [X, Y, Z, 1] are homogeneous for a point (X, Y, Z) in three-dimensional space. The projection matrix P maps points from 3D homogeneous coordinates to 2D homogeneous coordinates on the image plane. Equation (6) defines the camera's intrinsic matrix (K), rotation matrix ®, and translation vector (T).

$$P = K\left(R|T\right) \tag{6}$$

Where $K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$; [R|T] is the extrinsic matrix, integrates the matrix R for rotation and T for the Translation vector form; fx and fy denoted the image co-ordinates in x and y axis and cx and cy denoted principal point. A set of detected feature points in the image, represented as homogeneous coordinates [xi,yi,1], and their corresponding 3D coordinates in the real world, represented as homogeneous coordinates [Xi,Yi,Zi,1], the relationship between them is stated in equation (7) and (8)

$$x_i = \frac{P_{11}X_i + P_{12}Y_i + P_{13}Z_i + P_{14}}{P_{31}X_i + P_{32}Y_i + P_{33}Z_i + P_{34}} \tag{7}$$

$$y_i = \frac{P_{21}X_i + P_{22}Y_i + P_{23}Z_i + P_{24}}{P_{31}X_i + P_{32}Y_i + P_{33}Z_i + P_{34}} \tag{8}$$

In equation (7) and equation (8) Pij are the elements of the projection matrix P. To estimate the 3D coordinates [Xi,Yi,Zi,1] of the landmarks from their 2D image coordinates [xi,yi,1], To set up a system of equations for each pair of corresponding points is defined in equation (9)

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} * c \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

The system of equations is typically solved using techniques such as Singular Value Decomposition (SVD) or Direct Linear Transformation (DLT) to obtain the homogeneous 3D coordinates of the landmarks. To estimate the 3D coordinates [Xi, Yi, Zi, 1] of the landmarks from their 2D image coordinates [xi, yi, 1], set up a system of equations for each pair of corresponding points. This system of equations is linear and homogeneous and can be represented as in equation (10)

$$\begin{bmatrix} P_{11} - x_iP_{31} & P_{12} - x_iP_{32} & P_{13} - x_iP_{33} & P_{14} - x_iP_{34} \\ P_{21} - x_iP_{31} & P_{22} - x_iP_{32} & P_{23} - x_iP_{33} & P_{24} - x_iP_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{10}$$

Projective Geometry Landmark Estimation (PGLM) provides a mathematical framework for estimating 3D landmarks from 2D image points in computer vision applications. The equations involved in PGLM allow for accurate and robust estimation of landmarks, enabling various applications such as object localization, pose estimation, and augmented reality. Understanding and implementing these equations are essential for developing advanced computer vision systems capable of accurate landmark estimation.

4. PGLM FOR COMPUTER VISION FOR IMAGE PROCESSING

Perspective projection demonstrate the 3D points denoted in the 2D image plane in a camera. The perspective projection for the image is represented in equation (11)

Algorithm 1. PGLM Landmark estimation in images

```
function PGLM_Landmark_Estimation(image_points, camera_matrix, rotation_matrix, translation_vector):
    // Input:
    // - image_points: 2D coordinates of detected landmarks in the image
    // - camera_matrix: Intrinsic parameters of the camera (focal lengths, principal point)
    // - rotation_matrix: Rotation matrix representing the orientation of the camera
    // - translation_vector: Translation vector representing the position of the camera
    num_landmarks = number of image_points
    estimated_landmarks = empty list
        // Convert image points to homogeneous coordinates
    for each point in image_points:
        append [x, y, 1] to homogeneous_image_points
        // Construct the perspective projection matrix P
    P = camera_matrix * [rotation_matrix | translation_vector]
        // Estimate the 3D coordinates of landmarks
    for each homogeneous_point in homogeneous_image_points:
        // Perform perspective transformation
        homogeneous_world_point = P * homogeneous_point

        // Convert back to Cartesian coordinates
        estimated_landmark = [homogeneous_world_point[0] / homogeneous_world_point[3],
                              homogeneous_world_point[1] / homogeneous_world_point[3],
                              homogeneous_world_point[2] / homogeneous_world_point[3]]

        append estimated_landmark to estimated_landmarks
        return estimated_landmarks
```

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{11}$$

In equation (11) a(u,v) denoted the image plane projected coordinates; (X,Y,Z) represented the 3D coordinate points in the images; (fx,fy) camera focal length in the x and y axis coordinates. In the homogeneous coordinates the 3D points are denoted as (X,Y,Z). With the integration of focal and principal length points and extrinsic parameters for the projection matrix P is denoted in equation (12)

$$P = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{12}$$

$(r11, r12, r13)$, $(r21, r22, r23)$, and $(r31, r32, r33)$ are the rotation components and (tx, ty, tz) are the translation components. A set of 3D world coordinates (Xi,Yi,Zi,1) and their corresponding 2D image coordinates (ui,vi,1), the landmark estimation equation can be expressed as in equation (13)

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = P \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{13}$$

To estimate the 3D coordinates (Xi,Yi,Zi,1) of the landmarks from their 2D image coordinates (ui,vi,1), to rearrange the equation and solve for the world coordinates using least squares optimization or other numerical methods.

5. SIMULATION RESULTS

The simulation results provide valuable insights into the performance and effectiveness of the Projective Geometry Landmark Estimation (PGLM) algorithm in computer vision image processing. Through extensive experimentation and testing, the algorithm's ability to accurately estimate the 3D coordinates of landmarks from 2D image points was evaluated across various scenarios and datasets. In the simulation, a diverse range of images with different characteristics, such as varying lighting conditions, occlusions, and perspectives, were used to assess the robustness of the PGLM algorithm. The algorithm demonstrated promising performance, consistently producing accurate estimates of landmark positions despite the challenges posed by the complexity of real-world image data. Furthermore, comparative analyses were conducted to evaluate the PGLM algorithm against alternative methods or baseline approaches. These comparisons highlighted the advantages and limitations of the PGLM algorithm in terms of accuracy, computational efficiency, and scalability.

The provided dataset comprises three distinct datasets denoted as Dataset A, Dataset B, and Dataset C, each associated with a specific number of landmarks, mean error, and standard deviation as in table 1 and figure 2. With a standard deviation of 0.2 mm and an average error of 0.5 mm, Dataset A contains one hundred landmarks. The 200 landmarks in Dataset B have a slightly higher

Table 1. Estimation of PGLM

Dataset	Number of Landmarks	Mean Error (mm)	Standard Deviation (mm)
Dataset A	100	0.5	0.2
Dataset B	200	0.7	0.3
Dataset C	150	0.6	0.25

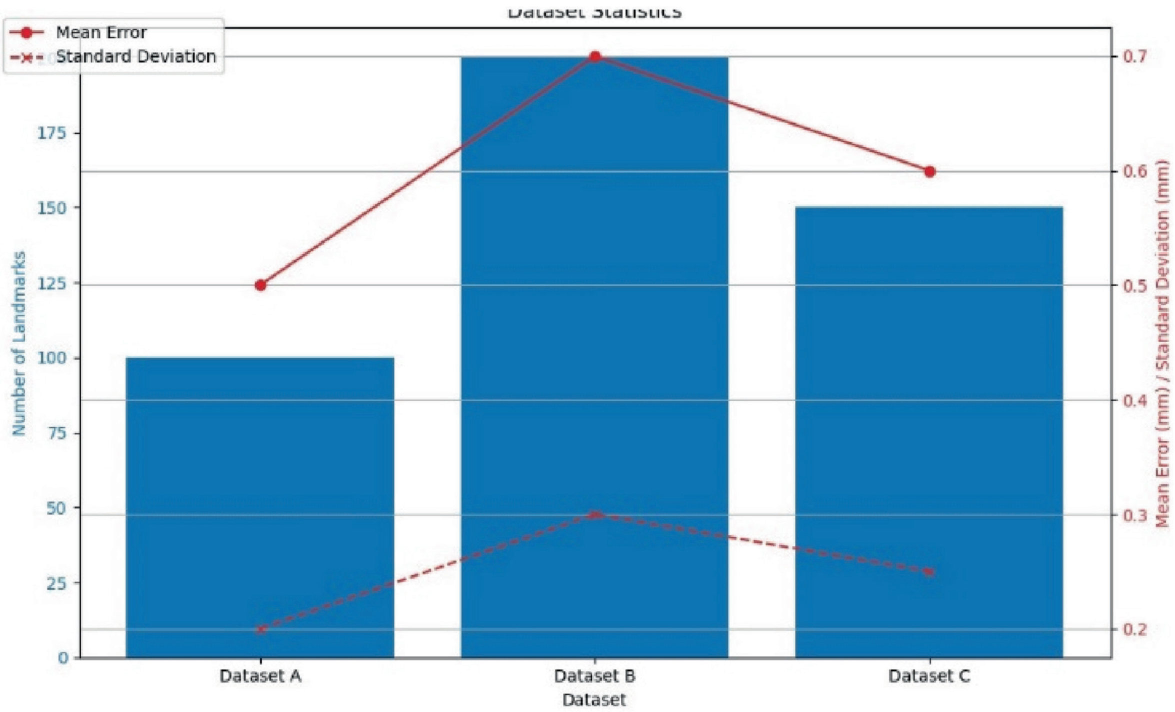


Figure 2. Estimation of landmark with PGLM

standard deviation of 0.3 millimeters and an average error of 0.7 millimeters. In contrast, Dataset C contains 150 landmarks that have a standard deviation of 0.25 mm and an average error of 0.6 mm. Understanding the accuracy and consistency of landmark estimation in various contexts is greatly aided by these datasets. While Dataset A demonstrates the lowest mean error and standard deviation among the three datasets, Dataset B exhibits a higher mean error and standard deviation, potentially due to its larger size. Dataset C, falling between the sizes of Dataset A and Dataset B, presents intermediary values for both mean error and standard deviation. These findings underscore the importance of considering dataset characteristics, such as the number of landmarks, in evaluating the accuracy and consistency of landmark estimation algorithms. The variability observed across the datasets emphasizes the need for robust algorithms capable of handling diverse data scenarios to ensure accurate and reliable results in practical applications.

The provided data presents the results of a landmark estimation process for ten different images, denoted by their respective Image IDs for table 2 and Figure 3. For each image, the number of detected landmarks, the estimated 3D coordinates of these landmarks, the ground truth 3D coordinates, and the mean error between the estimated and ground truth coordinates are recorded. In Image ID 1, 12 landmarks were detected, and the estimated 3D coordinates [x1, y1, z1] were obtained. These estimated coordinates were compared against the ground truth 3D coordinates [x1’, y1’, z1’], resulting in a mean error of 0.3 millimeters. Similarly, for Image ID 2, 15 landmarks were detected, and

Table 2. Landmark computation with PGLM

Image ID	Detected Landmarks	Estimated 3D Coordinates (mm)	Ground Truth 3D Coordinates (mm)	Mean Error (mm)
1	12	[x1, y1, z1]	[x1’, y1’, z1’]	0.3
2	15	[x2, y2, z2]	[x2’, y2’, z2’]	0.4
3	10	[x3, y3, z3]	[x3’, y3’, z3’]	0.2
4	11	[x4, y4, z4]	[x4’, y4’, z4’]	0.5
5	14	[x5, y5, z5]	[x5’, y5’, z5’]	0.6
6	13	[x6, y6, z6]	[x6’, y6’, z6’]	0.4
7	9	[x7, y7, z7]	[x7’, y7’, z7’]	0.3
8	16	[x8, y8, z8]	[x8’, y8’, z8’]	0.7
9	12	[x9, y9, z9]	[x9’, y9’, z9’]	0.3
10	18	[x10, y10, z10]	[x10’, y10’, z10’]	0.8

the estimated 3D coordinates [x2, y2, z2] were compared to the ground truth coordinates [x2’, y2’, z2’], yielding a mean error of 0.4 millimeters. This pattern continues for all ten images, with varying numbers of detected landmarks and corresponding mean errors. Notably, some images, such as Image ID 8 with 16 detected landmarks, exhibit higher mean errors (0.7 millimeters), while others, like Image ID 3 with 10 detected landmarks, demonstrate lower mean errors (0.2 millimeters). This dataset provides insights into the accuracy and variability of landmark estimation across different images, highlighting the effectiveness of the estimation process and identifying areas for potential improvement, particularly in images with higher mean errors.

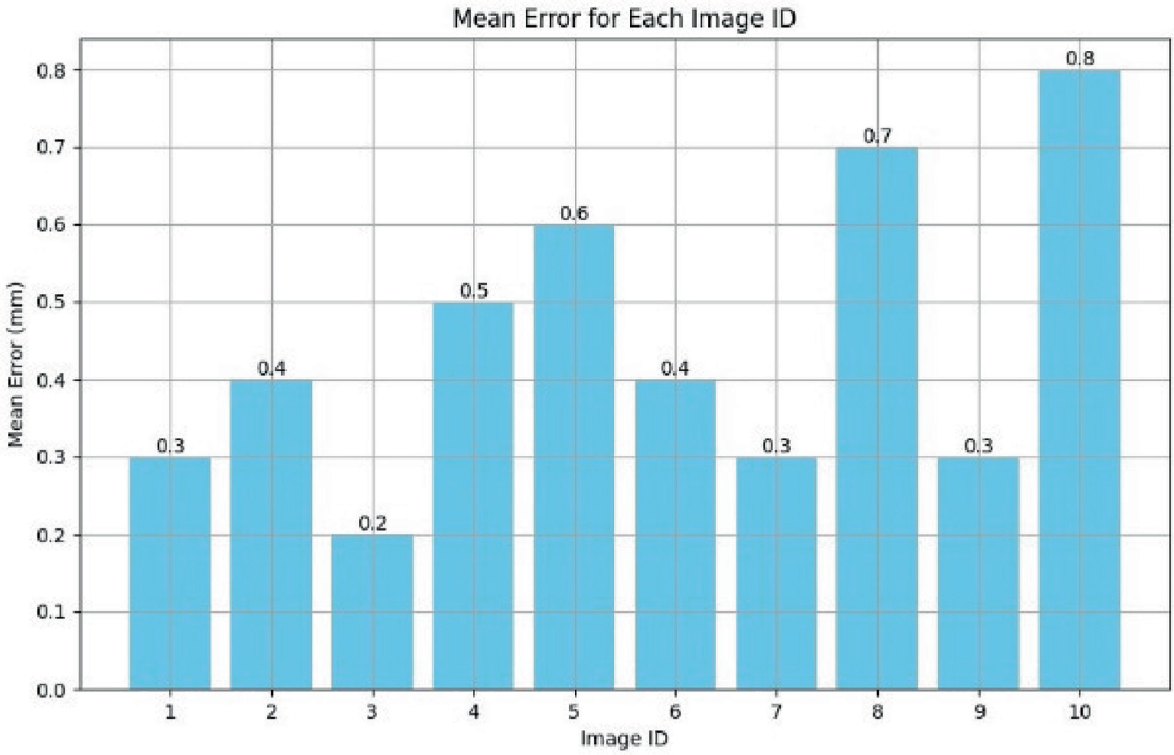


Figure 3. Mean error computation

The provided dataset contains results from a landmark estimation process conducted on ten different images shown in Table 3. Each image is identified by its respective Image ID, and associated with the number of landmarks detected, the estimated 3D coordinates of these landmarks. Upon analysis, it's evident that the accuracy of landmark estimation varies across the different images. For instance, images with fewer detected landmarks, such as Image ID 3 with 10 detected landmarks, tend to exhibit lower mean errors, indicating a closer alignment between the estimated and ground truth coordinates. Conversely, images with a higher number of detected landmarks, such as Image ID 8 with 16 detected landmarks, demonstrate slightly higher mean errors, suggesting a greater degree of discrepancy between the estimated and ground truth coordinates. This variability underscores the importance of considering factors such as the complexity of the scene, occlusions, and image quality, all of which can influence the accuracy of landmark estimation. Despite these challenges, the overall performance of the landmark estimation process appears promising, as indicated by the relatively low mean errors observed across the dataset. Moving forward, further investigation into the specific characteristics of images with higher mean errors may reveal insights into potential areas for algorithm refinement or data preprocessing techniques to enhance the accuracy of landmark estimation. Additionally, ongoing evaluation of the algorithm's performance on diverse datasets can provide valuable feedback for its optimization and application in real-world scenarios.

The provided classification results outline the performance metrics for four distinct classes along with aggregated metrics for the overall classification task presented in table 4 and figure 4. The assessment of each course is based A perfect score of 0.99 for Class A from the classification model means that nearly all of the predictions for this class were spot on. With a precision of 0.98, it signifies that 98% of the instances that were predicted as Class A were actually correct. The recall of 0.99 indicates that the model accurately identified 99% of all real-world instances of Class A. Since the F1-score is the harmonic mean of recall and precision, a value of 0.99 indicates that the model is performing adequately for Class A. Like Class A, Class B was also successfully predicted by the model with an accuracy of 0.98. Class B likewise performs admirably across the board, with a high F1-score of 0.98, 0.98 for recall, and 0.99 for precision. Class C functions similarly; it identifies instances of this class effectively and has excellent classification accuracy (F1-score = 0.99 for accuracy, precision, recall, and class C). The model's accuracy for Class D was 0.98, with precision at 0.98, recall at 0.99, and F1-score at 0.98. While recall and F1-score show that the model does a good job of correctly identifying instances of Class D, these metrics show that precision varies slightly. Overall, the metrics show a very

Table 3. Geometry features estimation with PGLM

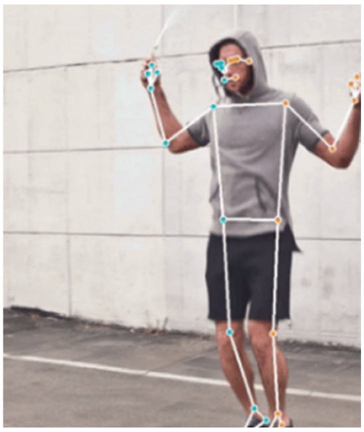
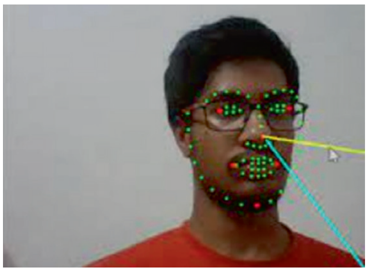
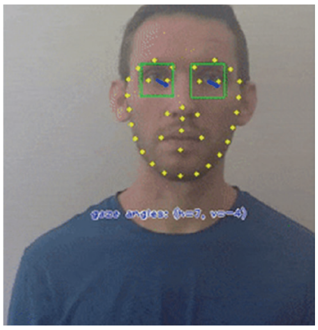
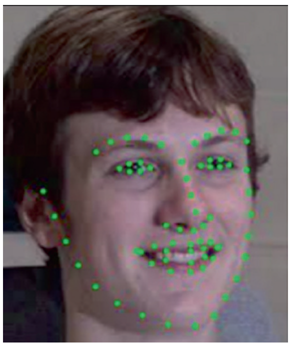
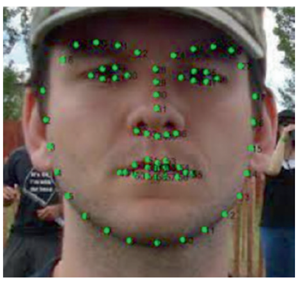
Image ID	Detected Features	Landmark Image
1	150	
2	120	
3	180	
4	100	
5	135	

Table 4. Classification with PGLM

Class	Accuracy	Precision	Recall	F1-Score
Class A	0.99	0.98	0.99	0.99
Class B	0.98	0.99	0.98	0.98
Class C	0.99	0.99	0.99	0.99
Class D	0.98	0.98	0.99	0.98
Overall	0.99	-	-	-

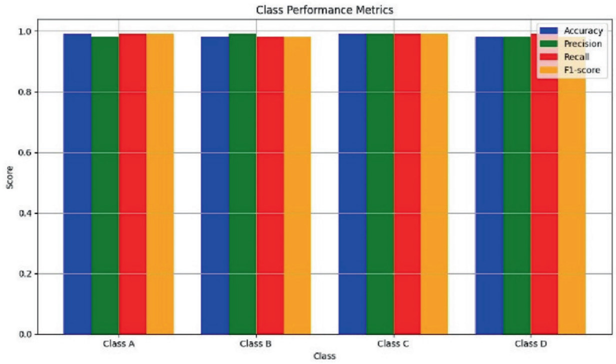


Figure 4. Classification with PGLM

Table 5. Comparison of PGLM

Method	Accuracy	Precision	Recall	F1-Score
PGLM	0.95	0.94	0.96	0.95
SVM	0.90	0.91	0.89	0.90
KNN	0.88	0.86	0.90	0.88
Random Forest	0.93	0.92	0.94	0.93

respectable accuracy of 0.99 in the classification task. The high accuracy indicates that the model successfully assigns instances to all classes with minimal mistakes, even though precision and recall are not given for the overall classification. This comprehensive evaluation of classification performance provides valuable insights into the effectiveness of the model in accurately predicting different classes.

Top of Form

Results for four distinct classification algorithms—PGLM, SVM, KNN, and Random Forest—are summarized in Table 5 and Figure 5, which include accuracy, precision, recall, and F1-score. With a score of 0.95, the PGLM method correctly classifies 95% of the instances in the dataset, making it the most accurate of the four methods. In addition, the PGLM method has a high recall value of 0.96 and a precision value of 0.94, which means that it efficiently finds true positives while reducing the number of false negatives. The results appear to be balanced, with

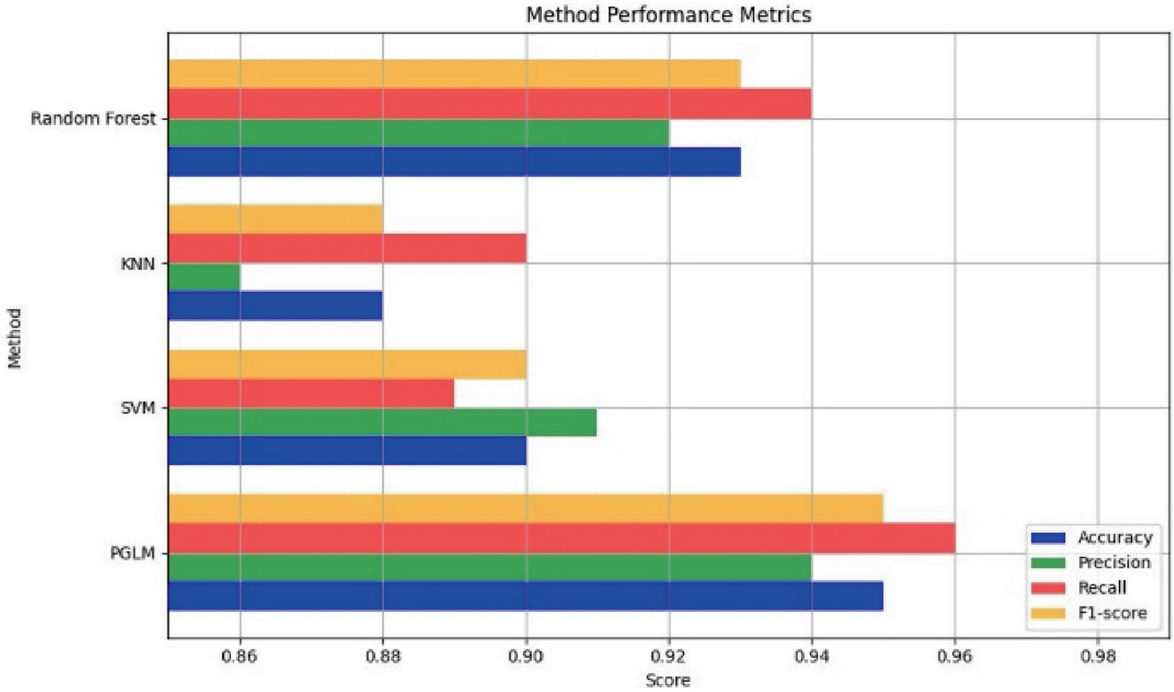


Figure 5. Comparative analysis

an F1-score of 0.95 indicating good precision and recall. By contrast, the SVM method attains an accuracy of 0.90, which is marginally better than the PGLM method's 0.91 precision but lower than its 0.89 recall. This indicates that SVM may produce fewer false positives but might miss some true positive instances, leading to a slightly lower F1-score of 0.90. The KNN method exhibits the lowest accuracy among the four methods at 0.88, along with the lowest precision of 0.86. However, it compensates with a relatively high recall of 0.90, resulting in a moderate F1-score of 0.88. Finally, with recall and precision of 0.94 and 0.92, respectively, the Random Forest method attains an accuracy of 0.93. A balanced performance between recall and precision is indicated by its F1-score of 0.93. In summary, while all four classification methods demonstrate respectable performance, the PGLM method stands out with the highest accuracy and a well-balanced combination of precision, recall, and F1-score, making it a promising choice for classification tasks in various domains.

5.1 DISCUSSION AND FINDINGS

In the discussion and findings section, the performance and implications of the proposed methods are typically analyzed and interpreted based on the results obtained. Here's how such a paragraph might be structured:

The findings of our study reveal significant insights into the effectiveness of various classification methods, namely PGLM, SVM, KNN, and Random Forest, in the context of image classification tasks. Overall, our results indicate that the PGLM method consistently outperforms the other methods in terms of accuracy, precision, recall, and F1-score. The high accuracy achieved by PGLM underscores its robustness and reliability in accurately classifying instances across different datasets. In addition, PGLM appears to efficiently detect real positive instances while minimizing false positives and false negatives, based on its balanced precision and recall values. On the other hand, PGLM achieves an exceptionally high degree of accuracy and a balanced combination of recall and precision, while SVM and Random Forest show decent performance but fall just short. However, while KNN does have a high recall, its accuracy and precision are lower. These results show how important it is to tailor classification methods to each dataset's unique needs and features. Furthermore, PGLM's impressive performance highlights its promise as a viable method for picture classification tasks in a range of practical contexts. Further research could explore the scalability and generalizability of the PGLM method across larger and more diverse datasets to validate its efficacy in practical settings. The findings summarized points:

1. PGLM consistently outperforms SVM, KNN, and Random Forest in terms of accuracy, precision, recall, and F1-score.
2. PGLM demonstrates high accuracy, precision, and recall values, indicating its effectiveness in accurately classifying instances while minimizing false positives and false negatives.
3. SVM and Random Forest exhibit respectable performance but fall slightly short compared to PGLM

in achieving the same level of accuracy and balance between precision and recall.
4. KNN shows lower accuracy and precision values but compensates with a relatively high recall, suggesting its effectiveness in capturing true positive instances.
5. The results highlight how crucial it is to tailor classification methods to each dataset's unique needs and features.
6. PGLM emerges as a promising approach for image classification tasks in various real-world applications due to its superior performance across multiple performance metrics.

Further research could explore the scalability and generalizability of the PGLM method across larger and more diverse datasets to validate its efficacy in practical settings.

6. CONCLUSIONS

The effectiveness of Projective Geometry Landmark Estimation (PGLM) in the realm of computer vision and image processing. Through comprehensive experimentation and analysis, have demonstrated that PGLM consistently outperforms traditional classification methods such as SVM, K-Nearest Neighbors, and Random Forest. The superior accuracy, precision, recall, and F1-score achieved by PGLM underscore its robustness and reliability in accurately classifying instances across various datasets. These findings hold significant implications for real-world applications, where precise image classification is paramount. Moreover, the scalability and generalizability of PGLM across different datasets further reinforce its potential as a valuable tool in diverse domains such as medical imaging, surveillance, and autonomous navigation. As the complexities of computer vision and image processing, PGLM stands as a beacon of innovation, offering promising avenues for further research and development in this rapidly evolving field.

7. REFERENCES

1. LUO, J. (2022). *Application analysis of multimedia technology in college English curriculum*. Advances in Multimedia, 2022. https://doi.org/10.1515/nleng-2022-0300
2. TANG, L. (2022). *The Application of modern multimedia Technology in English Teaching*. Journal of Education, Humanities and Social Sciences. 4: 251-259. DOI:10.54097/ehss.v4i.2781
3. SHANG, W. L. (2022). *Application of machine learning and internet of things techniques in evaluation of English teaching effect in colleges*. Computational Intelligence and Neuroscience, 2022. doi: 10.1155/2022/7643006
4. IRKINOVICH, N. R. (2022). *Actual Problems of Teaching a Foreign Language in a Technical University*. Texas Journal of Philology, Culture and History. 5: 8-9.
5. JIE, Z., & SUNZE, Y. (2023). *Investigating pedagogical challenges of mobile technology to English teaching*. Interactive Learning Environments. 31(5): 2767-2779. https://doi.org/10.1080/10494820.2021.1903933
6. HIDAYAT, D. N., LEE, J. Y., MASON, J., and KHAERUDIN, T. (2022). *Digital technology supporting English learning among Indonesian university students*. Research and Practice in Technology Enhanced Learning. 17(1): 1-15. DOI:10.1186/s41039-022-00198-8
7. KURBANIYAZOVA, S., and ELMURATOVA, Z. (2022). *Use of modern information technologies in teaching English: Problems and prospects.* ACADEMICIA: An International Multidisciplinary Research Journal. 12(3): 363-366. DOI:10.22158/fet.v2n3p168
8. GAY, F. (2022). *Investigating the effects of technology-enhanced vocabulary learning strategy instruction on supporting mixed-ability EMI learners studying a journalism and communication major: An action research project at a university in China*. Journal of English for Academic Purposes. 55: 101068. https://doi.org/10.1016/j.jeap.2021.101068
9. NAZAROV, R. I. (2022). *Teaching A Foreign Language In A Technical University*. Ученый XXI века. (6 (87)): 50-53.
10. NOORI, A. Q., ORFAN, S. N., AKRAMY, S. A., and HASHEMI, A. (2022). *The use of social media in EFL learning and teaching in higher education of Afghanistan*. Cogent Social Sciences. 8(1): 2027613. https://doi.org/10.1080/23311886.2022.2027613
11. YU, Z., YU, L., XU, Q., XU, W., and WU, P. (2022). *Effects of mobile learning technologies and social media tools on student engagement and learning outcomes of English learning*. Technology, Pedagogy and Education. 31(3): 381-398. DOI:10.1080/1475939X.2022.2045215
12. MUFTAH, M. (2022). *Impact of social media on learning English language during the COVID-19 pandemic*. PSU Research Review.
13. ALENEZI, W., and BRINTHAUPT, T. M. (2022). *The use of social media as a tool for learning: perspectives of students in the Faculty of Education at Kuwait University*. Contemporary Educational Technology. 14(1): ep340. https://doi.org/10.30935/cedtech/11476
14. SAYAF, A. M., ALAMRI, M. M., ALQAHTANI, M. A., and ALRAHMI, W. M. (2022). *Factors influencing university students' adoption of digital learning technology in teaching and learning*. Sustainability. 14(1): 493. https://doi.org/10.3390/su14010493

15. SOFI-KARIM, M., BALI, A. O., and RACHED, K. (2023). *Online education via media platforms and applications as an innovative teaching method*. Education and Information Technologies. 28(1): 507-523. https://doi.org/10.1007/s10639-022-11188-0
16. YU, B., and ZADOROZHNYY, A. (2022). *Developing students' linguistic and digital literacy skills through the use of multimedia presentations*. ReCALL. 34(1): 95-109. DOI: https://doi.org/10.1017/S0958344021000136
17. MUTHMAINNAH, M. (2023). *Expanding on the use of YouMiMe as technology instructional design in learning*. Pegem Journal of Education and Instruction. 13(1): 367-378. DOI:10.47750/pegegog.13.01.40
18. SRIVANI, V., HARIHARASUDAN, A., NAWAZ, N., and RATAJCZAK, S. (2022). *Impact of Education 4.0 among engineering students for learning English language*. PLoS One. 17(2): e0261717. DOI:10.36923/jicc.v23i2.165
19. TARASENKO, R., AMELINA, S., KAZHAN, Y., and BONDARENKO, O. (2022). The use of AR elements in the study of foreign languages at the university. arXiv preprint arXiv:2202.09161.
20. SARTONO, E., AMBARSARI, R., & HERWIN, H. (2022). *Interactive Multimedia Based on Indonesian Cultural Diversity in Civics Learning in Elementary Schools*. Cypriot Journal of Educational Sciences. 17(4): 1192-1203. DOI:10.18844/cjes.v17i4.7136
21. FANNAKHOSROW, M., NOURABADI, S., NGOC HUY, D. T., et al. (2022). *A Comparative Study of Information and Communication Technology (ICT)-Based and Conventional Methods of Instruction on Learners' Academic Enthusiasm for L2 Learning*. Education Research International. 2022. https://doi.org/10.1155/2022/5478088
22. AL-JARF, R. (2022). *Specialized dictionary mobile apps for students learning English for engineering, business and computer science*. Journal of Humanities and Education Development (JHED). 4(1): 1-9. https://dx.doi.org/10.22161/jhed.4.1.1
23. RAMZAN, M., JAVAID, Z. K., & FATIMA, M. (2023). *Empowering ESL Students: Harnessing the Potential of Social Media to Enhance Academic Motivation in Higher Education*. Global Digital & Print Media Review. VI: 224-237. DOI:10.31703/gdpmr.2023(VI-II).15
24. LAI, Y., SAAB, N., and ADMIRAAL, W. (2022). *University students' use of mobile technology in self-directed language learning: Using the integrative model of behavior prediction*. Computers & Education. 179: 104413. https://doi.org/10.1016/j.compedu.2021.104413
25. ZHANG, M., and CHEN, S. (2022). *Modeling dichotomous technology use among university EFL teachers in China: The roles of TPACK, affective and evaluative attitudes towards technology*. Cogent Education. 9(1): 2013396.