Designing A Visual Control Framework Based on Feature Point Analysis to Guide Robots in Uncharted Areas

Researcher Isam Sadiq Rasham AL-Ghadhbawi
Islamic Azad University, Faculty of International
Computer Engineering-Software
Isfahan (Khorasgan) Branch
Islamic Republic of Iran
essamsadik2@gmail.com

تصميم إطار تحكم بصري يعتمد على تحليل نقاط الميزة لتوجيه الروبوتات في المناطق غير المستكشفة

الباحث عصام صادق رشم الغضباوي
جامعة آزاد الاسلامية - الهيئة العالمية لهندسة الحاسوب
اصفهان - جمهورية ايران الاسلامية

Abstract

Image processing may be made more accurate and efficient with the use of optical processing technologies. In order to make better design judgments early in the package design process, visual packaging art simulation aims to swiftly develop realistic packaging effects. Conventional simulation techniques frequently yield findings that are neither realistic or detailed, and they also take a lot of time and human resources. Optical processing-based machine vision technology can fully utilize computer vision and image processing techniques to simulate package designs accurately and efficiently. This research proposes a foundation for machine vision technologies based on image optical processing. Real package pictures are captured using high-resolution cameras during the image collection stage, and during the optical processing step, algorithms are employed to enhance and reduce noise and analyze fine details in the images. After that, advanced image processing techniques are used to accurately extract and analyze the main features of the design.

Keywords: Computer Vision, Geometric properties, Posture, Motion, Human visual functions, Visual packaging art design, Machine learning algorithms.

المستخلص

يمكن أن تصبح معالجة الصور أكثر دقة وكفاءة باستخدام تقنيات المعالجة البصرية. من أجل إصدار أحكام تصميمية أفضل في وقت مبكر من عملية تصميم ، تهدف محاكاة فن التغليف المرئي إلى تطوير تأثيرات تغليف واقعية بسرعة. غالبًا ما تسفر تقنيات المحاكاة التقليدية عن نتائج ليست واقعية أو مفصلة، كما أنها تستغرق الكثير من الوقت والموارد البشرية. يمكن لتقنية الرؤية الآلية القائمة على المعالجة البصرية الاستفادة الكاملة من تقنيات الرؤية الحاسوبية ومعالجة الصور لمحاكاة تصميمات العبوات بدقة وكفاءة. يقترح هذا البحث أساسًا لتقنيات الرؤية الآلية القائمة على المعالجة البصرية للصور. يتم التقاط صور العبوات الحقيقية باستخدام كاميرات عالية الدقة أثناء مرحلة جمع الصور، وأثناء خطوة المعالجة البصرية، يتم استخدام الخوارزميات لتعزيز وتقليل الضوضاء وتحليل التفاصيل الدقيقة في الصور. بعد ذلك، يتم استخدام تقنيات معالجة الصور المتقدمة لاستخراج وتحليل السمات الرئيسية للتصميم بدقة.

الكلمات المفتاحية: الرؤية الحاسوبية، الخصائص الهندسية، الوضعية، الحركة، الوظائف البصرية البشرية، تصميم فن التغليف البصري، خوارزميات التعلم الآلي.

Introduction

The word "robot" can mean different things to different people. Many people's expectations of robots and their capabilities are largely influenced by science fiction literature and movies. Unfortunately, this common perception does not reflect the state of robotics practice. But one thing is certain: robotics will play a major role in the technological landscape of this century. Items such as vacuum cleaners will soon be among the many smart devices that will appear in our homes and offices, and the first will be cleaning robots. Automata like the duck depicted in Vaucanson (Figure 1.1a) fascinated 17th-century Europeans.

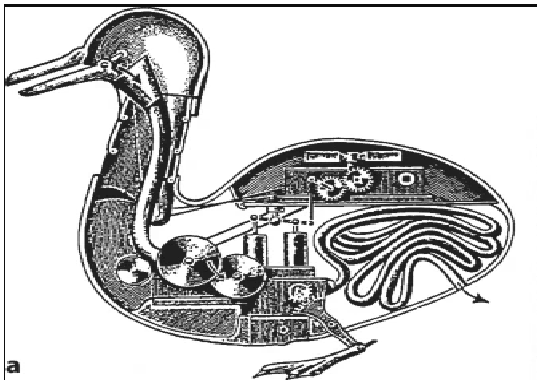

Fig 1.1a: Early programmable machines a Vaucanson's duck (1739) .

These robots exhibited behavior that seemed lifelike at the time, despite their sophistication for the period. Vaucanson then went on to study the automation of silk weaving, and the duck uses a cam mechanism to synchronize its movements. Jacquard further developed these ideas and built a loom that looks like a programmed loom (Fig. 1.1b).

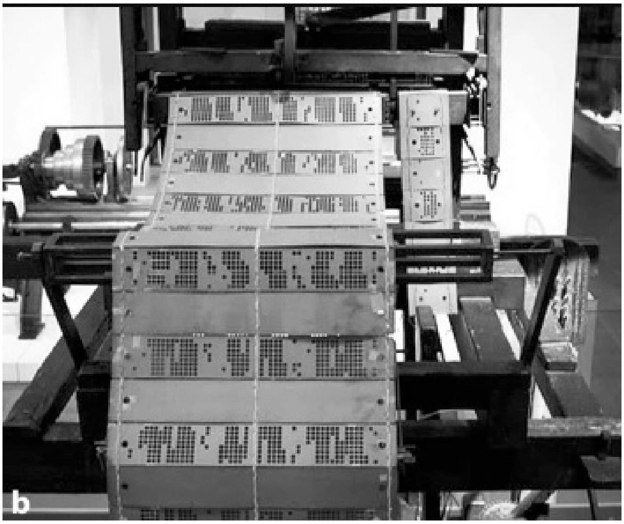

Fig 1.1 b: Programmed mechanism system

The coded pattern to be woven was represented by a series of holes on a punch card. The device had several features of a modern robot: it could perform physical tasks and was programmed. The word "robot" first appeared in the Czech science fiction play "Rossum's Universal Robot" by Karel Čapek in 1921. The word "robot" alludes to an artificial person or robot and comes from the Czech word for "slave." As in many subsequent robot stories, the robot uprising in the play ends tragically for humans. Isaac Asimov's Robot Trilogy deals with ethics and human-machine relations. He was involved in several novels and short stories published between 1950 and 1985. The "positronic brain" possessed by these robots is where the "Three Laws of Robotics" are kept. These stories influenced public perception of robots and also later novels and films. Also in the mid-20th century, the field of cybernetics emerged, a name rarely heard today but once an exciting science at the forefront of understanding life and creating intelligent machines. The first robot patent was filed by George C. Devol in 1954 and granted in 1961. The device was a robotic arm with a gripper attached to a track and its movements recorded by a camera. It was designed as a magnet stored on a rotating drum. The first robotics company, Unimation, was founded in 1956 by Devol and Joseph Engelberger. Their first industrial robot, the Unimate (Figure 1.2), was put into operation in 1961.

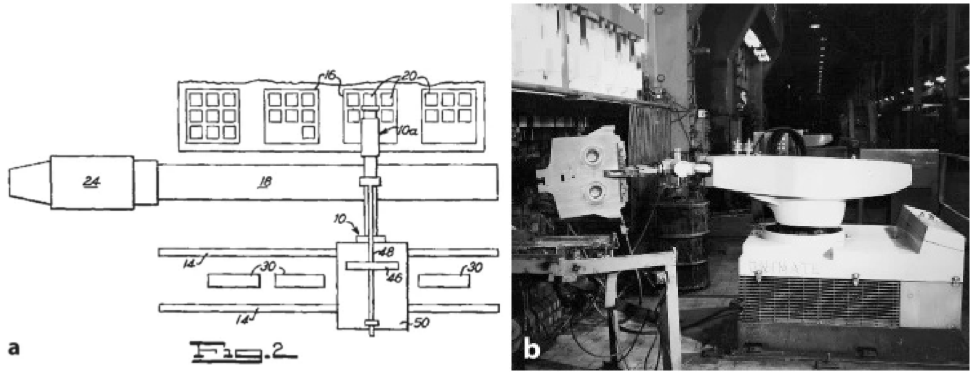

Fig. (1.2a & b): Universal automation plan view of the machine from Devol’s patent

Devol and Engelberger's original idea of robotic automation has been realized, and millions of arm robots, such as the one in (Figure 1.3), have been built and are currently used for a range of tasks, such as: B, welding, painting, loaders, electronic palletizing, and packaging assembly.

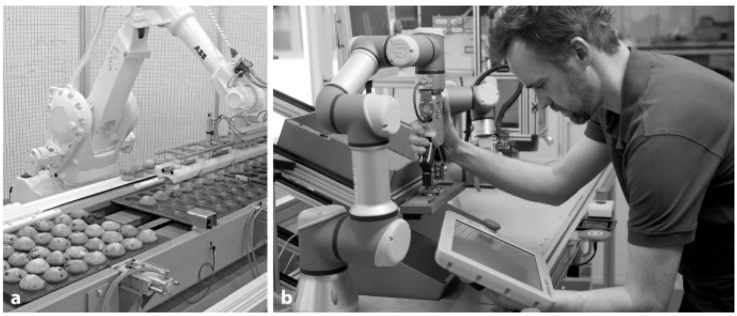

Fig (1.3) Arm robots

The use of robots has increased productivity and improved product quality. Rather than eliminating jobs, it has helped manufacturers in countries with high labor costs to remain profitable. Many of the products we buy today are made or operated by robots. Today, these early robots fall under the category of manufacturing robots, a branch of robotics. Other types include field robots for outdoor work, humanoid robots that resemble real people (Figure 1.4), and service robots that perform tasks such as household chores, personal assistance, or medical rehabilitation.

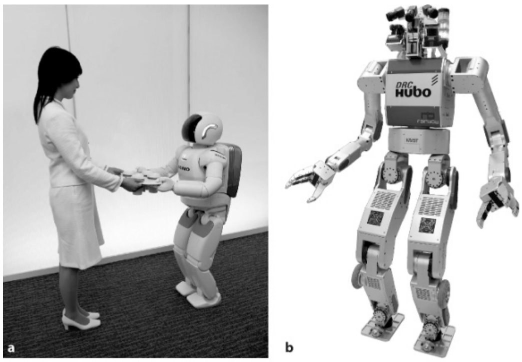

Fig. (1.4 a & b): Non-terrestrial mobile robots. a) SeaBed Autonomous Underwater Vehicle (AUV) operated by the Australian Centre for Field Robotics (Photo: Roger T. Hanlon), b) Global Hawk Unmanned Aerial Vehicle (UAV) - (Photo courtesy of NASA)

A factory robot is typically a small work cell manipulator in the form of a fixed arm that performs repetitive tasks. Parts are systematically fed to the robot so it can maximize its superior speed and precision. Robotic work environments should be avoided because their high pace can be dangerous. Field and service area robots face huge obstacles requiring robots to work and navigate in complex, crowded, and dynamic environments, a requirement that is required of delivery robots in hospitals that must move between crowds and constantly moving parked carts and carts. A Mars rover must navigate small craters and boulders because it does not have an accurate local map before setting off. Robotic cars must obey traffic rules, traffic signals, and road regulations.

The DARPA Grand Challenge showcases examples of these vehicles (Buehler, Iagnemma, and Singh, 2007). The second challenge requires these robots to work safely in human environments. Hospital delivery robots operate inside the human body, robotic surgical instruments operate inside the human body, and robotic cars operate inside the human body. " **How do you define a robot**? Not all of the definitions are very useful, yet there are plenty. This book will use the following definition to guide us: a goal-oriented machine with sense, plan, and action capabilities. A robot senses its surroundings and plans its actions based on that information and a predetermined purpose. The task could involve manipulating an object with an arm-robot's tool or guiding a mobile robot to a specific spot. For robots, sensing is essential. Proprioceptive sensors detect the actual condition of the robot, such as the current drawn by an electric motor or the angle of the joints on a mobile robot with wheels. Outside-in sensors Calculate the current status of the world in relation to the robot. For instance, to detect collisions, it might make use of a robot vacuum cleaner. It could be a compass that determines the direction of the Earth's magnetic field in relation to the robot's heading, or it could be a GPS receiver that calculates the distance to a constellation of satellites. An active sensor could possibly be the source.1956–1982, Unimation Inc. Devol first met with Joseph Engelberger in order to secure funding for the development of his animation technology. He received financial support from Moore, Lanning, and Maxwell. In 1956, they co-founded Unimation, the first robotics company, in Danbury, Connecticut. Consolidated acquired the company in 1970. Diesel Corp. (CDC) developed it into the Condec division. Their first robot was used in a General Motors die casting plant in New Jersey. In 1968, they licensed their technology to Kawasaki Heavy Industries, allowing it to produce the country's first industrial robot. Engelberger ran the company until Westinghouse Electric acquired it in 1982.

Employees in the company's engineering department volunteered to help operate the cars. Peter Corke's book Robotics, Vision, and Control: Basic Algorithms in MATLAB Mobility: We have already mentioned the types of mobile robots and how they move. This section is about mobility, which is determined by how far the vehicle moves in space. It can be characterized by a scalar parameter, called the generalized coordinates of the train. In addition, the train has a position in configuration space, or C space, specified by this space CC. For each state in C, the robot can achieve a certain amount of free motion, which means that it can move in any direction along this space. It is worth noting that the people on the surface also had a certain advantage in creating a chart that can be used to travel around Mars, although they only provided the most general information. In other words, the robot chooses its own path to the destination after the human tells it where to go. Considering that the communication is delayed by several minutes, the time difference in communication is clearly visible. Camera Calibration Some robots are hybrid controllers, where the control system is shared or exchanged with the human operator. In tracked hybrids, the control system is shared by the human operator and the robot. For example, the aircraft pilot can act like an autopilot and regain control. With shared control, the human operator has the authority to keep the car in its lane and avoid collisions, while the human operator is only responsible for steering. "Robotics, Vision, and Control: Fundamental Algorithms in MATLAB" by Peter Corke The location of the camera's center of mass is a problem caused by external factors, namely the position of the camera.

Determining the intrinsic and extrinsic properties of a camera with respect to a world coordinate system is called camera calibration. The set of world points on which calibration techniques rely has known relative coordinates and associated image plane coordinates. Advanced methods such as Bouguet's Calibration Toolbox for MATLAB® (Bouguet 2010) require only many images of a planar chessboard target. From this, the intrinsic parameters (including distortion parameters) and the relative positions of the chessboard in each image can be computed, as explained in Section 11.2.4. Although they can estimate the distortion model, classical calibration methods can only be applied to views of the calibration target in three dimensions. Nevertheless, these methods are easy to understand and serve as the basis for our discussion in the following sections. "Robotics, Vision, and Control: Fundamental Algorithms in MATLAB" by Peter Corke Camera Class: The abstract superclass Camera is the predecessor of the toolbox camera classes Central Camera, Fisheye Camera, and Spherical Camera. The public methods of all classes are shown in

Table (1) The behavior of camera views is related to MATLAB® figures behave similarly. By default, plots and grids redraw the camera's view. If a camera view does not already exist, it is created. The MATLAB commands clf and hold correspond to the procedures clf and hold.

Ethod	Description
p = cam.project(P)	Project the world point P to the image plane p
cam.plot(P)	Plot the world points defined by the columns of P
cam.mesh (X, Y, Z)	Plot the mesh defined by X, Y, and Z
cam.showpose (T)	Show the camera at specified pose
cam.clf	Clear the current camera view
cam.hold	Hold the current camera view, future calls to plot add to the camera view
cam.rpy(r, p, y)	Set the camera transform to specified roll-pitch-yaw angles
cam.name	Property: name of camera
cam.Tcam	Property: default camera transform (read and write)

Specific camera classes: Table (2) lists the option parameters that the constructor of each camera type accepts. possess special choices that are explained in the web documentation. The standard central camera specifications when building cameras are 1024×1024, 8mm focal length, 10 square pixels. Unless explicitly specified, the principal point is assumed to be at the center of the image. "Robotics, Vision, and Control: Basic Algorithms in MATLAB" by Peter Corke.

Option	Description
'name', name	The name of the camera which is displayed in the window's title bar
('resolution', npix	The dimensions of the image. npix is a scalar for a square image or a 2-vector
'centre', pp	The coordinate of the principal point
'pixel', rho	The dimensions of the pixel. rho is a scalar for a square pixel or a 2-vector
'noise', sigma	The standard-deviation of Gaussian noise added to the image-plane coordinates
'pose', T	The default pose of the camera. Default is as shown in Fig. 11.2
'image', im	Set camera image-plane dimensions according to the image dimensions and display the image

Computer vision: What is it? People can easily see and understand the three-dimensional structure of their surroundings. Imagine how clearly we perceive three dimensions when we look at a vase of flowers on a table. Imagine how clearly we perceive three dimensions when we look at a vase of flowers on a table. The flowers can be easily distinguished from the background due to the delicate surface patterns of light and shadow on the petals (Fig. 1.1a). It is easy to count (or say) the names of all the people in a group photo and to infer their emotions from their facial expressions. Although perceptual psychologists have successfully created illusions1 to illuminate some of the principles of the visual system, they have spent decades trying to understand how the visual system works. (Fig. 1.3) They have not yet completely solved this puzzle (Marr 1982; Palmer 1999; Livingstone 2008). Accurate methods are currently being developed to compute partial 3D models of the environment from hundreds of partially overlapping photos (Fig. 1.2). When enough images of a particular object or facade are available, stereo matching can generate dense and accurate 3D surface models.

Additionally, it may follow a moving subject against a complicated backdrop. Additionally, by fusing face, clothes, With facial hair identification, it is able to identify and identify each individual in a picture . The goal of making a computer comprehend visuals at a level comparable to that of a two-year-old—for example, counting every animal in a picture—has not yet been accomplished, despite recent advancements. Why is eyesight so challenging? One reason is the inverse nature of vision. When there is not enough data to be completely certain of an answer, we seek to discover the unknown. Therefore, we are forced to use physics-based probabilistic models. However, modeling the complexity of the visual world is far more demanding than simulating the vocal tract that produces spoken language. Forward models used in computer vision are usually created in computer graphics or physics (optics, radiometry, sensor design). Both areas model how objects move, how light is scattered by the atmosphere, how it reflects off surfaces, and how it is refracted by camera lenses. (or the human eye), and lastly projected onto an image plane that is flat or curved. Even if entirely computer-animated movies including human characters haven't been able to overcome the uncanny valley2, which divides actual people from android robots and computer-animated humans, computer visuals are still improving every day.

The illusion of reality works perfectly in certain contexts, such as depictions of static scenes made up of objects or animations of extinct animals like dinosaurs, that is, using one or more photographs to describe the world we live in and reconstructing its features, such as shape, lighting, and color distribution. It's interesting to note that computer vision algorithms frequently make mistakes, whereas humans and animals can perform this task without any issues. Individuals without prior experience in this sector sometimes underestimate the complexity of this issue. (My coworkers frequently ask me for software that can identify and identify every person in a picture.) This is ...Algorithms and Applications in Computer Vision by Richard (Szeliski, 2022, p. 3).

Computer Vision's Objectives: The digital implementation of human visual functions, such as detecting, recognizing, processing, and interpreting three-dimensional scenes in real-world scenarios, is known as computer vision. Eventually, to accomplish typical visual tasks, a more universal system that is akin to the human visual system should be developed. The second study objective is to use this work as a means of investigating the systems of the human brain involved in visual tasks and to gain knowledge and comprehension of these mechanisms (e.g., computational neurology). Here, research on biological mechanisms is the main focus. Long-term study of the human brain's visual system from the perspectives of physiology, psychology, neurology, and cognition has yielded a wealth of knowledge, but it has not yet fully unraveled all the mysteries surrounding visual processes, particularly with regard to the study and comprehension of visual mechanisms. When it comes to the research and proficiency of visual information processing, we are still far behind. A thorough study of computer dream would be made calmer with a full grasp of how the humanoid intelligence processes visual.. We mainly address the first research goal in this text. As previously said, computer vision research is heavily inspired by human visual functions and is a computerized embodiment of those functions. from the view of humans. Typical examples include motion detection through filtering techniques, the application of the concept of local orientation, the efficient data structure of pyramids, and more recently, artificial neural networks. The creation of novel computer vision algorithms may also result from an understanding of an investigation into the operation of the human visual system. Applications and research in computer vision have a long history. In general, 2D projected images of 3D objective scenes were the main source of information for early computer vision systems. The primary objective of computer vision research was to enhance the of the image and enabling the user to obtain information more easily and clearly, or on automatically gathering several types of characteristic information from the picture so that the user may examine and identify the scene. This feature is related to 2D computer vision, which is a reasonably developed field with a wide range of current uses. As theory and technology have advanced, an increasing amount of study has concentrated on fully utilizing 3D spatial information (typically in conjunction with temporal information). derived from objective landscapes to automatically assess, comprehend, and make decisions about the objective world. This involves deriving extra depth information from 2D projection images in order to have a thorough grasp of the 3D environment. Computer vision is currently focused on artificial intelligence, however this subject requires the introduction of other technologies. research that is currently being conducted. This book focuses on recent related research, which falls under the category of 3D computer vision. Basics of 3D Computer Vision and Advanced Techniques. Automation Robots are machines that have the ability to be programmed to carry out certain activities. These chores can be as easy as transferring a piece of paper from one place to another. These can also be more difficult assignments, such traveling the globe or completing activities that are doable by humans, like figuring out a Rubik's cube. Constructing a robot that can handle difficult jobs, however it can be difficult because a lot of training data is needed for the robot to behave human-like. One way to generate this data is through simulation, which is a technique for modeling a robot's behavior. Robotics requires synthetic data for a number of reasons. **First**, there's typically a lack of real-world data. This is particularly valid for the data required for model training in machine learning. This is particularly valid for the data required for machine learning model Training, an important part of robotics. In some cases, real data can be completely replaced or supplemented by synthetic data. Second, noisy real-world data is common.

This noise can be generated by the environment, actuators, sensors, and other sources. Machine learning models can be trained using noise-free data generated by synthetic data. Third, collecting real data is often costly. This is especially true for the data required for model training in machine learning. We can generate data that can be collected more cost-effectively by using synthetic data. Fourth, real data is often biased. There are many possible reasons for this distortion, including the environment, actuators, and sensors. To help train machine learning models, unbiased data can be generated using synthetic data. Real data is often unrepresentative, which is the fifth reason why synthetic data is needed in robotics. This is especially true for the information required to train machine learning models. Because synthetic data is used, machine learning models can be trained using data that more accurately represents the real world. (Basic methods of deep learning: One of the straightforward methods of deep learning is reinforcement learning, transition learning, data amplification, hyper-parameter optimization, and network design.

1.Reinforcement Learning Model project based on reinforcement knowledge has two components: One for creating models and the other for verifying them. Initially, a collection of sub-networks is generated by the model generation unit using a random starting technique. The income feedback generation unit is the proper rate of validation after training using the sub-networks on the indicator dataset. The generating unit conducts fresh trials and modifies the design strategy in light of the model effects. The design action that happens automatically Space encompasses functions like pooling, group convolution, residuals, and convolution. Global network topologies can be constructed macroscopically, while repeated substructures of convolutional networks can be designed microscopically. There are numerous brain structures in the search space. many networks. Deep neural networks can be used to integrate insight, decision-making, or perceptual decision-making finished end-to-end knowledge by marrying reinforcement information and executive skills.

2. Transfer learning Assignment learning: Transfer learning is a method to maximize one's own objective by using related support work. This task optimization method works well when there is little labeled data for the target job. In order to optimize the model for the target task or accelerate learning, transfer learning techniques are usually applied after training the source model using the source task with a large number of training examples.

Training the Source Model: Training a model on a source task with a sizable amount of labeled training data is the first step in the process. The target task and this source task usually include similar kinds of data. As an illustration, a model may be taught to identify a wide range of common things in photos (source task), which may subsequently be helpful for identifying particular objects, such as different species of plants or animals (target job).

Knowledge Transfer: The knowledge (features, patterns, etc.) acquired from the source task is transferred to the target task once the model has been trained on it. This is accomplished by fine-tuning the trained model—or portions of it—using the dataset of the target task. The theory behind this is that general properties like edges, textures, and forms are captured by the neural network's early layers, which can be used again. It is only necessary to modify or retrain the later layers—which are more task-specific—using the target data.

Fine-Tuning: The pre-trained model is further trained on the dataset of the target task during the fine-tuning phase. Through this procedure, the model's parameters are incrementally modified to better match the unique properties of the target data. By fine-tuning, the model is able to match the more specialized features required for the target task to the broader traits it learnt from the source task. Although transfer learning has many advantages, there are drawbacks as well. Making sure the source and target tasks are appropriately related is one of the main challenges.; In the absence of these, the knowledge transfer may be ineffective or even detrimental to the model's performance (a phenomena referred to as negative transfer). Furthermore, especially in cases when the target dataset is limited, fine-tuning needs to be carefully controlled to prevent over-fitting. A major development in machine learning is transfer learning, which offers a framework to support models' effective and efficient learning even in the face of sparse data. Through the purposeful utilization of pre-existing knowledge, transfer learning not only expedites the creation of intelligent systems but also expands the range of sectors in which AI may be applied.

3. Data Augmentation Deep learning: A large quantity of labeled exercise information is required. Though, there are few general exercise datasets for practical situations. Moderate amounts of data can be processed efficiently through data augmentation. Through processes such as transformation and synthesis, data augmentation creates a new dataset that is comparable to the training dataset. Adding noisy data can also enhance the resilience of the model. Common data augmentation can be divided into three main categories: traditional data augmentation techniques (such as copying, rotation, ascending, translation, random thinning, adding clamor,

etc.); recently, direct and efficient data augmentation techniques (such as safety device, mixing, sample pairing, chance pruning, etc.); and automatic data increase methods (such as reinforcement learning-based methods).

4. Hyper- parameter **optimization Deep learning model design requires data**: Model design, model training, feature selection, preprocessing, and hyper-parameter tuning. Hyper-parameters are crucial for deep knowledge. Manual parameter tuning requires expertise and machine learning knowledge. Automated hyper--parameter optimization based on bottomless knowledge can efficiently replace physical parameter change. The main idea here is to control the relationship amid the loss drive of the authentication set and the hyper-parameters, then compute the function originated on the hyper-parameters and smear the incline descent method to enhance the hyper-:parameters. Though, hyperparameter derivation involves tremendously complex calculations.

5. **Network Project Network design**: It is an essential skill for deep learning. Extended Short-Term Reminiscence Networks, Multilayer Perceptron Networks, Perceptron Networks, Auto-Coupled Systems, Bidirectional Regular Neural Systems, Convolutional Neural Systems, Profound Convolutional Neural Grids, Copiously Coupled Networks, Reproductive Combative Webs, Graph Convolutional Neural Networks, Self-Learning Network Coding Networks, Recurrent Neural Networks, and Recurrent Neural Networks. The list includes many more. 3D Computer Vision Fundamentals and Advanced Techniques Page 35.36

Deep Learning in Computer Vision: Subterranean erudition procedures have demonstrated remarkable performance in several tasks such as semantic segmentation, depth estimation, object detection, object tracking, picture and video classification, and more since 2012. Deep learning has steadily supplanted classical statistical learning to become the de facto standard for computer vision frameworks and techniques. Image classification is an attempt to place a picture in a specific category. Classical image classification models, such as Alex-Net, VGG, Google-Net, Res-Net, and Dense-Net, have worked as the foundational net for various errands such as discovery and division. Neural net models are made up of layers upon layers, ranging in depth from a few to thousands. Classification of videos: Two-stream convolution network is another effective deep learning method that was used earlier to combine motion and apparent data. A 2D convolution kernel serves as the foundation for this two-stream convolution network. Many academics have suggested using 3D CNNs in recent years to achieve video categorization by either merging 2D and 3D or expanding the 2D convolution kernel to 3D. I3D, C3D, P3D, and other ones are some of these. These suggestions for video motion detection include the generalized compact nonlocal network, boundary sensitive network, and attention clustering network, among others.

Object tracking: Object tracking is the act of figuring out how one or more objects move within a certain scene. In multi-object tracking, a video image is captured with numerous trajectories of interest, and temporal correlation is used to derive information about the moving trajectory. Object tracking can be broadly classified into two groups: Both discriminative and generative approaches are used. In order to reduce the reconstruction error, generative approaches primarily employ a reproductive perfect to describe the object's seeming topographies and then look for other potential items. Tracking via detection refers to the second category, which is a discriminative strategy that uses the classifier's training to separate the target from the back-ground. Due to its more reliable performance, object tracking research has gradually focused on this method. In recent years, some tracking techniques based on Siamese networks semantic segmentation have attracted attention: semantic segmentation requires labeling the semantic class of each pixel in the double. A common method to bottomless education is to custom a fully convolutional network (FCN). FCN implements end-to-end semantic image segmentation by inputting an image and immediately obtaining the group of apiece pixel as output. U-Net, advanced convolution, Deep Lab series, PSP-Net and many other advances are examples of further progress. Monocular depth estimation: Image data is usually used as input to predict the depth value consistent to each pixel in the image. This is the basis of deep learning for monocular complexity guesstimate. The characteristic elementary construction of CNN uses unfathomable erudition for monocular deepness estimation methods. It is believed that this solitary view stereophonic corresponding (SVSM) model can be trained based on a minimum amount of depth annotated images, because monocular depth estimation is usually data-intensive for a big quantity of depth footnote data, and data collection is expensive. Another is the creation of images and videos, which is more in line with computer graphics technology. The intellectual traits of the double are the participation, and the related image circulation is the output. The automatic creation of photos and movies, the growth of databases, and the improvement of image information have received more consideration in the development of deep learning. VAE and GAN are two popular deep generative models. GAN is an unsupervised deep learning technique that learns through a game amid two neural nets and can partly solve the problematic of data scarcity. Starting from Pix2Pix, which prepares data in pairs, to Cycle-GAN, which processes unpaired data, to Star-GAN, which can handle multiple domains, this research direction is steadily approaching reality. Consider an AI anchor. 3D Computer Vision Fundamentals and Advanced Methods, pp. 37-38"

Traditional Calibration Methods: "In order to establish the corresponding relationship between the calibration target points and the corresponding points on the captured image for determining the internal and external parameters of the camera, the traditional method of camera calibration requires a known calibration target. This can determine the size and shape of the calibration target. These offer good calibration accuracy for the calibration target, a straightforward solution, and theoretical clarity. The drawback is that the

The precision of the calibration target is relatively high, despite the somewhat complex calibration method. Fundamental Steps and Conditions: With reference to the full space imaging model presented in Sect. 2.2.4 and the nonlinear camera model presented in Sect. 2.3.2, the camera can be calibrated in the direction of transformation from 3D world coordinates to computer image coordinates. illustrates the four processes involved in converting the world coordinate system to the computer picture coordinate system, each of which requires some parameter calibration. The parameters that need to be calibrated are the translation matrix T and the rotation matrix R. First step:. Step 2: The lens's focal length, or λ, serves as the calibrating parameter. Step 3: The lens's radial distortion coefficient (k), eccentric distortion coefficient (l), and thin prism distortion coefficient (m) are the calibrating parameters. Step 4: The uncertainty image scale factor, μ, is the calibrating parameter. Two-Step Calibration Process A classic example of a traditional calibration method is the two-level method. The reason for its name is that the calibration is split. divided into two parts: determining the camera's exterior parameters (without accounting for translation along the optical axis) and determining additional parameters. Because it makes use of radial alignment constraints (RAC), this technique gets its name. The technique involves a large number of linear equations, making parameter solving reasonably simple. This technique is used in industrial vision systems; the depth direction accuracy can approach 1/8000 and the average 3D measurement accuracy can reach 1/4000. The calibrating procedure might be separated into two scenarios:. Since k is the lens's radial distortion coefficient, it is not possible to take k into account when calculating R using the methods described above. If f µ is known, then only one image comprising a collection of coplanar datum points needs to be utilized for calibration. Right now, R, Tx, and Ty are calculated in the first step, and λ, k, and Tz are calculated in the second. Similarly, k is not taken into consideration when calculating Tx and Ty, but it is when calculating Tz (the impact of a changeTz on the picture is comparable to k), therefore it moves on to the second phase. Foundations and Advanced Methodologies for 3D Computer Vision, page 40.

Predictive Tracking: The KLT approach, as was previously mentioned, operates under the premise that there are slight displacements between images captured at subsequent instants. When these are large, using spatiotemporal coherence to forecast the locations of the points in the subsequent frame can be a way to offset the motion. The following three steps are repeated in a conventional scheme: First, extraction; then, association, prediction. The method usually employs a Kalman filter, which makes use of a motion model to forecast a point's location based on its past positions at a given time instant. Additionally, it preserves an estimate of the prediction's uncertainty, enabling the definition of a search window. Within the search window, features are identified; one of these is linked to the tracked point; the filter then uses the found point's position and uncertainty to update the position estimate. There could be uncertainties in the association operation. More complex filters, such as the Joint Probabilistic Data Association Filter, are required when many points are identified in the search window. Bar-Shalom and Fortmann (1988) provide an in-depth analysis of the issue. Computer Vision: Techniques for Three-Dimensional Reconstruction page155" **Scale Invariant Feature Transform** "Blob-like feature locations with high contrast are identified by the Scale Invariant Feature Transform (SIFT) operator, which was introduced by Lowe (2004). It does this by approximating the LoG filter (Laplacian of Gaussian) and applying it to a scale-space representation of the picture. Scale-Space Recall that at a given scale, feature points of differing types are detected by both the HS operator and the LoG. Although it is not given much thought at first, the choice of scale is not a secondary one and has to be addressed soon. In reality, real scenes are composed of numerous structures, each with its own inherent scales. frequently distinct from one another. In addition, the distance from the camera affects how big these different parts are projected into an image. This implies further that the perception of a genuine thing will vary based on other situational conditions. One technique to deal with the fact that it is impossible to determine beforehand which size is best to explain each of the intriguing structures in the photographs is to include them all in a multi-scale description.. Thus, we present the concept of scale-space. The representation of the image consists of a set of smoothed copies of the image, which are determined by the smoothing kernel's size. Together with the two spatial variables, (u, v), the parameter sigma, often known as scale, locates a point (u, v, σ) in scale-space. We'll make use of the linear scale-space, which a Gaussian kernel yields. This decision is not capricious; rather, it comes about as an inevitable outcome when the of because when formalizing the transition from a fine to a coarser scale, filtering shouldn't produce additional bogus structures. For more information, the reader is directed to Lindeberg (2012). Computer Vision: Techniques for Three-dimensional Reconstruction page156""

Description and Matching of Points of Interest: Similar to the gradient data produced by SIFT, the SURF descriptor characterizes the brightness distribution near the key point. The difference is that SURF container brand full use of the essential copy to haste up the cunning because it is based on the response of the first-order Haar wavelet in the X and Y directions instead of the gradient. In addition, its descriptor length is only 64, which saves time and improves flexibility in the feature matching and calculation process. The matching process of the SURF descriptor includes three steps: creating a square rectangle and finding the direction at the interest point based on the previously selected direction, obtaining the SURF descriptor at this location, and (3) comparing the properties of the descriptors in the two regions. Direction determination: To ensure the invariance of image rotation, the direction is specified. First, the X and Y hair wavelet responses are calculated in a 6σ circular neighborhood around the detected interest point, where σ is the scale at which the interest point is detected.

The sampling step size relative to the scale is defined by the value of σ. To ensure compatibility with other components, the size of the wavelet is scale-dependent and is determined by the side length of 4σ. Therefore, the integral image is reused for fast filtering. Based on the properties of the Haar wavelet template, the responses in the X and Y guidelines at any scale container be calculated with only six operations. Then, after weighting the wavelet response with a Gaussian delivery placed at the opinion of attention, the reply is got as a opinion in a establish interplanetary anywhere the flat coordinate agrees to the flat response concentration and the perpendicular manage corresponds to the perpendicular response concentration. As shown in Character 1.5, the course is determined by summing all responses in a descendent fan-shaped gap of size π/3 radians (stage scope π/18).

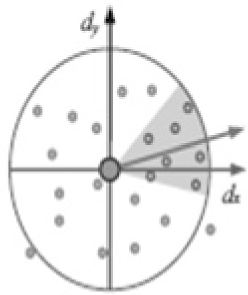

Fig. 1.5: Schematic diagram of determining orientation

A local orientation vector is created by adding the individual sums of all of the replies—both vertical and horizontal—in the window. The point of interest's orientation is determined by the biggest vector across all windows. The moving window's dimension is an input that needs to be carefully set because high dimensions often result in nonsensical maxims in the vector, whereas small dimensions are more likely to reflect the gradient of a single benefit. Characteristic derived from the total of the Haar wavelet responses A square window of that orientation is centered at a certain point of interest, depending on the orientation that has been selected there.. Every orientation and position within the image is handled in this way. The diameter of the window is 20σ. The window is then divided into smaller 4 × 4 = 16 blocks at regular intervals using the spatial knowledge of the features. The wavelet reply is calculated in a standard 5 × 5 grid in each sub-region. For simplicity, we use the notation dx and dy to denote the flat and perpendicular Haar wavelet responses, respectively. The terms “horizontal” and “vertical” in this paper refer to the designated opinion of attention, as described below. Determining the direction A direction is chosen at the point of interest to achieve invariance to image rotation. First, the Haar wavelet response is computed along the X and Y directions within a 6σ circle radius around the point of interest (where σ is the scale at which the point of interest is detected). The sampling step is equal to σ and is scale-dependent. Its size is fixed to 4σ side length and, like other wavelets, is scale-dependent. It enables a fast filtering process by reusing computations using integral maps. Lone six processes are obligatory to compute the reply at each scale on the X and Y axes, as it belongs to the Haar wavelet template group. The wavelet responses are computed and Gaussian filtered using a weighting function and can be represented as coordinate points in space. The abscissa represents the strength of the response in the horizontal direction and the ordinate represents the strength of the response in the vertical direction. The way container be got by summing the replies inside a descending subdivision window of stage scope π/18 and radian π/3, as shown in Figure 1.5. The local direction vector is created by adding the horizontal and vertical responses within the window independently. The direction of the point of interest is given by the longest vector in all windows. The size of the sliding window is critical. It is a parameter that must be chosen carefully. Large sizes usually result in insignificant vector maxima, while small sizes usually reflect dominant gradients. Descriptors created from the sum of Haar wavelet responses. We will first compute the descriptor by tiling a square region centered on our point of interest in the direction specified in the previous section (so that we are also rotationally invariant). We will use a window size of 20σ. These squares are regularly divided into 4 × 4 = 16 sub-squares, showing important local and spatial information. Please compute the hair wavelet for each sub-region. Responses are made in a fixed 5 × 5 grid. To simplify notation, we denote the horizontal Haar wavelet response as dx and the vertical Haar wavelet response as dy. Note that "vertical" and "horizontal" mention to the designated opinion of attention. (Fig 1.6).

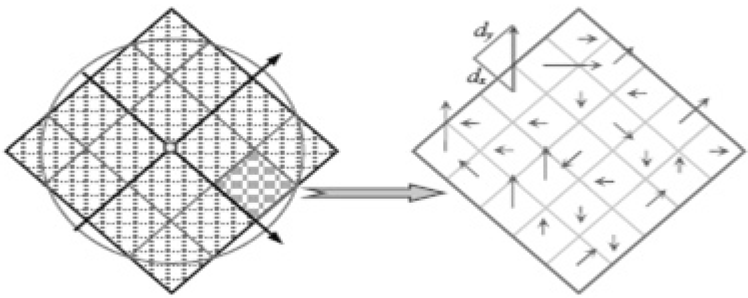

Fig 1.6 : Four-sided area around opinion of attention

Review of Reconstruction Theory in Visual Perception Glitches with Rebuilding Philosophy Rendering to Marr's philosophy, "The fundamental purpose of processing is to recover the scene from visual stimuli and incorporate it into the representation. Representation is the core idea shared by all visual activities or professions. There needs to be a representation that can enable diversity of recovery if the vision system is to recover the scene's features, including the object's surface reflectance, movement direction and speed, surface structure, and so on. The strong theory states that various occupations are built on the same data structure, share a same conceptual core, and go through the same understanding process. Marr provided evidence in support of his theory that the representations that comprise the visual People can deduce the inside world from a multitude of hints. If the creation of such a unified representation is taken to be the ultimate goal of all visual input processing and decision-making, then vision may be thought of as a reconstructive process that begins with scene interpretation of information. Since reconstruction does not directly contribute to interpretation, the reconstruction-based approach can have the widest circle. Obtaining the representation solely through reconstruction from the source image is another challenge. Recovering good scene representations from original photos in computer vision is tremendously computationally expensive.. In fact, a growing body of research in biological vision tends to corroborate alternative representation ideas. Lastly, there are conceptual issues with the reconstruction theory. The theoretical assumption that reconstruction is possible for every representation task is the root of the issue. Setting aside the issue of whether reconstruction is technically feasible, one may first wonder if it is worthwhile to search for a representation that is universally consistent. Given that the most effective representation will be whichever performs the task the best; a representation with universal homogeneity is not always required. Information processing theory actually states that choosing the appropriate representation for a given computer issue is crucial. Marr also mentioned this in Foundations and Advanced

Methodologies of 3D Computer Vision. page16.17 algorithms that work together. Cooperative algorithms were among the first techniques for disparity calculation, and they were influenced by computer models of human stereo vision. (e.g., Dev, 1974; Marr and Poggio, 1976; Marroquin, 1983; Szeliski and Hinton, 1985; Zitnick and Kanade, 2000). These algorithms, which generally behave similarly to global optimization algorithms, update disparity estimates step-by-step through nonlinear operations. Actually, a global function that is being minimized for some of these algorithms can be stated explicitly [Scharstein and Szeliski 1998]. Gradual, coarse-to-fine warping. The majority of the most advanced algorithms available today initially count all potential matches at all potential discrepancies before choosing, by some method, the best set of matches. Sometimes, techniques derived from classical (infinitesimal) optical flow computation yield faster results. Here, disparity estimations are updated incrementally as each image is successively warped, until a decent registration is achieved. The most typical implementation of these techniques is inside a hierarchical coarse-to-fine refining framework (Quam, 1984; Bergen, Anandan, Hanna, Hingorani, 1992; Barron, Fleet, Beauchemin, 1994; Szeliski, Coughlan, 1997).

Programming that changes with time Dynamic programming gives rise to a variant of the global optimization algorithm. For typical classes of smoothness functions, 2D optimization of Equation is known to be NP-hard [59], but for independent scanlines, dynamic programming can find the global minimum in polynomial time. The first use of dynamic programming was in sparse, edge-based stereo vision techniques. The problem of dense (intensity-based) scanline matching has been the focus of more recent methods. Belhumeur (1996), Geiger, Ladendorf, and Yuille (1992), Cox, Hingorani, Rao, and others (1996), Bobick and Intille (1999), and Birchfield and Tomasi (1999) are instances of this. These methods operate by calculating the least-cost path via a horizontal slice of the DSI, or through the matrix of all pairwise matching costs between two related scanlines. Utilizing plane-based approaches, partial occlusionIt is feasible to directly estimate homographies between distinct planes in situations rich in planar structures, such as those containing architecture and specific manufactured goods like furniture, by applying either feature-based or intensity-based techniques. Essentially, this data may be utilized to concurrently deduce the camera positions and the plane equations, meaning that plane-based structure can be calculated from motion. Luong and Faugeras (1996) provided concrete evidence of recovering a fundamental matrix from two or more homographies. by the use of least squares and algebraic manipulation. Regretfully, this method is frequently not particularly reliable because meaningful re-projection mistakes are not directly correlated with the generated algebraic errors (Szeliski and Torr 1998). The homo-graphics' sites are, hallucinated in order to guarantee that a few virtual point correspondences are displayed and incorporated into a typical structure from motion algorithm. Using full bundle adjustment, which includes explicit plane equations and extra constraints to make reconstructed co-planar features lie precisely on their respective planes, may be an even more effective method. (A logical approach to this would be to utilize 2D in-plane parameterizations for the remaining points and to define a coordinate frame for each plane, for example, at one of the feature points.) The mechanism that Han Shum created."

Methodology Research and Application of Image Engineering: 28 years ago saw the launch of image engineering (IE), which sparked the start of an extensive review series devoted to the statistical classification of image engineering literature. Every pertinent study that was published in 15 different publications was methodically reviewed for this review. The literature review divided the research into four main sections: image processing, image analysis, and image understanding. This is because image engineering is a smooth combination of these three interrelated but distinct fields, as well as their practical engineering applications. comprehending images and using technology. Based on the substance and emphasis of the research, each of these categories was further subdivided into more focused groups. This classification, which comprises 25,970 research articles published over the previous 18 years and highlights important trends and research areas in the field of image engineering, is condensed in. The data emphasizes the most popular fields of study and how they are used.

Review of Reconstruction Theory in Visual Perception: The first step in this research is to evaluate David Marr's seminal theory of visual perception. Marr claims that the main job of the visual system is to create an internal, coherent representation of the external scene by reconstructing it from the visual stimuli it receives. According to his idea, this reconstruction process happens in a sequential manner, allowing the visual system to recognize crucial aspects of the image such surface reflectance, the direction and speed of moving objects, as well as surface patterns. Marr's hypothesis has been under increasing examination despite its enormous effect, as current research in neuropsychology and psychophysics suggests that visual processing is not strictly linear. Rather, it makes use of intricate parallel processing systems that combine several data sources at once. This study reevaluates Marr's reconstruction theory and points out its flaws, particularly in terms of elucidating the entirety of human visual perception in terms of distinguishing and categorizing complex visual scenes.

Classification of Image Engineering Literature: This study conducted a systematic review to categorize the vast and diverse literature on image engineering (which combines image processing, image analysis, and image understanding) and its useful applications across a range of technological fields. The literature was categorized into four main areas: copy dispensation, copy examination, image understanding, and technical requests to address broad problems in the discipline. Within these categories, subcategories were created to better represent the specific focus of each research. For example, research in image processing was categorized into categories such as segmentation, enhancement, and filtering, and image analysis was categorized into categories such as feature extraction, object recognition, and pattern recognition. This systematic review included more than 25,970 research publications from the past 18 years, providing a comprehensive overview of research trends and priorities in imaging technology. This categorization emphasizes the broad range of research and the connections between multiple subfields within the discipline.

Model Design and Evaluation: Based on theoretical understandings, a strong model was created to assess the efficiency of different visual processing methods. A crowd sourced dataset that was selected for its diversity and fit with the study's goals was used to train this model. Many photographs in the collection had been annotated to detect qualities like tappability, or components of a user interface that users believe to be trappable. The dataset was divided into ten sections and a tenfold cross-validation procedure was applied to guarantee the model's resilience and dependability. 10% of the data was used for validation and 90% of the data was utilized for training in each iteration. 100,000 rounds of this method were done to make sure the model was accurate and stable. The performance of the model was evaluated using precision and recall measures, which are widely used in machine learning and information retrieval. The results showed that it was highly effective in properly predicting which elements users will interact with. The average precision was 90.2% (SD: 0.3%), and the average recall was 87.0% (SD: 1.6%). would perceive as trappable.

Results Analysis: A thorough analysis was carried out to compare the model's predictions with the "Click ability" attribute—a prevalent element in mobile interface design—found in the view hierarchy in order to obtain a deeper understanding of the model's efficacy. A confusion matrix was used in the evaluation process to give a thorough understanding of the model's accuracy across several classes. The model greatly increased recall, demonstrating an improvement of over 7%, and improved precision marginally when compared to the clickable property, according to the data. This implies that even in cases when items aren't clearly labeled as trappable in the display hierarchy, the model is better at detecting elements that consumers consider to be trappable. Additionally, the minority class was up sampled in order to establish a balanced dataset, which addressed the imbalance in the dataset—where trappable components were more frequent than non-trappable ones. The model's precision and recall rates were 82% and 84% for trappable elements and 81% and 86% for non-trappable elements in tests conducted with this balanced dataset. These results show that even in complicated and variable data settings, the model can reliably detect trappable parts.

Critical Discussion: The study's conclusions provide a substantial contribution to the current discussion concerning the applicability of reconstruction theories in visual perception. The findings suggest that while reconstruction-based methods provide a well-organized framework for visual representation, they might not be entirely sufficient for tasks requiring scene recognition and classification. This is especially true for increasingly complicated visual tasks, because Marr's sequential and hierarchical approach conflicts with the simultaneous and integrative processes seen in human vision. From the standpoint of computer vision the study supports the case for alternative theories of visual representation by highlighting the inherent difficulties in reconstructing scene representations from source photographs. The study highlights the significance of choosing appropriate representations customized for particular computational tasks instead than following a generic, one-size-fits-all strategy.

Conclusion

This Within the field of computer vision, research has covered a wide range of subjects, from fundamental concepts like picture generation and preprocessing to more sophisticated methods like semantic analysis and three-dimensional reconstruction. The research examined diverse techniques for augmenting, dividing, and evaluating images in order to derive significant data, encompassing motion evaluation, object observation, and image creation. Even with substantial advancements in these areas, humans and computers still perceive visual information quite differently, and this disparity is expected to increase in the near future. The study also looked into a wide range of mathematical techniques, such as linear algebra, least squares optimization, three-dimensional and projective geometry, variational methods, and continuous mathematics like signal processing. It also looked at computer science and discrete mathematics subjects like graph algorithms, combinatorial optimization, and database retrieval methods. Concurrently, the study focused on how incorporating a robot's state awareness can improve how well it interprets visual information, especially for non-egocentric activities. Three complimentary robotic use cases were demonstrated, each tackling unique obstacles in the estimate of three-dimensional shapes. The results showed that consistent increases in spatial perception ability across different settings can be achieved by adding the state of the robot into visual deep learning models. Similar to the difficult subject of automotive engineering, computer vision is wide and multifaceted, which emphasizes the need for a multidisciplinary approach. Through the incorporation of many methodologies and theoretical frameworks, The purpose of this research is to improve our knowledge of visual perception and support the creation of increasingly sophisticated robotic systems. Even though computer vision is still a very difficult field to grasp, research is still being driven by the wide range of interesting issues it poses. The goal is to improve the field and offer significant insights through this work.

References

1."Robotics, Vision and Control: Fundamental Algorithms in MATLAB" by Peter Corke (2011), • dspace.agu.edu.vn

2.Computer Vision Algorithms and Applications by Richard Szelisk,(2010) .

3.3D Computer VisionFoundations and Advanced Methodologies (2024).Springer.

4. M. Rozemund, (2006) ,"The Nature of the Mind," in The Blackwell Guide to Descartes' Meditations, S. Gaukroger, John Wiley & Sons,

5. R. S. Geiger *et al.,* (Jan. 2020), "Garbage In, Garbage Out?," in Proceedings of the 2020 Conference onFairness,Accountability,andTransparency,,pp.325–336.doi:10.1145/3351095.3372862.

6. G. Andrews, (accessed Apr. 13, 2022), "What Is Synthetic Data?" Carrushome.com, 2022. https://www.carrushome.com/en/what-is-synthetic-data/.

7. Scikit, (accessed Apr. 13, 2022), "3.3. Metrics and scoring: quantifying the quality of predictions — scikit-learn 1.0.2 documentation." Sicikit Learn, 2022. https://scikit-learn.org/stable/modules/model_evaluation.html.

8. Airbus, "Can Synthetic Data Really Improve Algorithm Accuracy?" (May 20, 2021), https://www.intelligence-airbusds.com/newsroom/news/can-synthetic-data-really-improve-algorithm-accuracy/, (accessed Apr. 13, 2022).spiepr Par10

9. (S. Ramos *et al.*, (Nov. 2021), "RLDS: an Ecosystem to Generate, Share and Use Datasets in Reinforcement Learning,", Accessed: Apr. 13, 2022. [Online]. Available: https://github.com/deepmind/envlogger.

10. N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, (May 6, 2021) ,'Everyone Wants To Do The Model Work, Not The Data Work': Data Cascades in High-Stakes AI, in Proceedings of the (2021) CHI Conference on Human Factors in Computing Systems, , pp. 1–15, doi: 10.1145/3411764.3445518.spiepr Par109

11. G. Press, (accessed Apr. 13, 2022), "Andrew Ng Launches A Campaign For Data-Centric AI," Forbes, Jun. 16, 2021. https://www.forbes.com/sites/gilpress/2021/06/16/andrew-ng-launches-a-campaign-for-data-centric-ai/?sh=d556d9d74f57,

12. K. Singh, (accessed Apr. 13, 2022), "Synthetic Data — key benefits, types, generation methods, and challenges," Towards Data Science, May 12, 2021. https://towardsdatascience.com/synthetic-data-key-benefits-types-generation-methods-and-challenges-11b0ad304b55.

13. (B. Ding *et al.,* (2020), "DAGA: Data Augmentation with a Generation Approach for Low-resource Tagging Tasks,". [Online]. Available: https://ntunlpsg.

14. MRP, (accessed Jul. 02, 2022), "Natural Language Processing: Providing Structure To The Complexity OfLanguage."Jul.30,2018.https://www.mrpfd.com/resources/naturallanguageprocessing/

15. S. I. Nikolenko, (2021(Synthetic Data for Deep Learning, vol. 174. Cham:Springer International Publishing, , doi: 10.1007/978-3-030-75178-4.

16. S. Colaner, (accessed Apr. 13, 2022), “Why Unity claims synthetic data sets can improve computer vision models,” VentureBeat, Jul. 17, 2022. https://venturebeat.com/2020/07/17/why-unity-claims-synthetic-data-sets-can-improve-computer-vision-models/.

17.(Liu X S, Li, A H, Deng Z J, *et al.,* (2021), Advances in three-dimensional imaging technologies based on single-camera stereo vision. Laser and Optoelectronics Progress, 58(24): 9-29.

18. Zhang Y-J., (2017), Image Engineering, Vol.3: image understanding. De Gruyter, Germany.

19. Shapiro L, Stockman G., (2001), Computer Vision. Prentice Hall, London.

20. ElSheikh A, Abu-Nabah B A, Hamdan, M O*, et al.,* (2023), Infrared camera geometric calibration: A review and a precise thermal radiation checkerboard target. Sensors, , 23(7): #3479 (DOI: 10.3390/s23073479).

20.Tsai R. Y., (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the shelf TV camera and lenses. Journal of Robotics and Automation 3(4): 323-344.

21.Ma S D, Zhang Z Y (1998), Computer Vision: Fundamentals of Computational Theory and Algorithms. Science Press, Beijing.

22.Li Y, Yan Y C., (2021), A novel calibration method for active vision system based on array of concentric circles. Acta Electronica Sinica 49(3): 536-541.

23.Hartly R, Zisserman A., (2004), Multiple View Geometry in Computer Vision, 2nd ED. Cambridge University Press, UK.

24.Kim J-S, Gurdjos P, Kweon, I-S., (2005), Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. IEEE Transaction on Pattern Analysis and Machine Intelligence 25(4): 78-81.

25.Civera J, Bueno D R, Davison A J, *et al.,* (2009), Camera self-calibration for sequential Bayesian structure from motion. Proceedings of International Conference on Robotics and Automation 403-408.

25.Daniilidis K., (1999), Hand-eye calibration using dual quaternions. The International Journal of Robotics Research 18(3): 286-298.

26.Zhang Z Y (2004), Camera calibration with one-dimensional objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(7): 892-899.

27.Liao W L, Zhao H Q, Yan J C (2021), Online extrinsic camera calibration based on high-definition map matching on public roadway. Journal of Image and Graphics 26(1): 208-217.

28.Triggs B, McLauchlan P F, Hartley R I, *et al*., (1999), Bundle adjustment—a modern synthesis. Proceedings of the International Workshop on Vision Algorithms: Theory and Practice 298-372

29.Computer Vision: Three-dimensional Reconstruction Techniques, (2022).

30.Artificial Intelligence for Human Computer Interaction :A Modern Approach, (2021).