CHAPTER 6

AUGMENTED REALITY-BASED OBJECT DETECTION AND RENDERING USING VISUAL SLAM-BASED DEEP LEARNING TECHNIQUE

6.1 INTRODUCTION

In recent times, there seems to be a peak in the use of Visual SLAM – Simultaneous location as well as Mapping in the world of autonomous driving, Virtual reality and Augmented Reality. However, these features of the algorithm of visual SLAM suffer from the problems of translation and sensitivity of light. This work has offered a technique for the settings of augmented reality and visualization of digital data that are derived from such objects. Markers play a major role in present technologies of Augmented Reality. This virtual environment is only used within the restricted space of marker and within the virtual objects that are placed and tracked in 3 dimensions within the areas as well.

But still, it is quite challenging to do this on cellphone that has only limited resources while keeping a lower latency and proper accuracy. Together employing the technique CLAHE and offer a photograph or image processing component that can locally enhance the contrast in pictures and also can extract the additional information of feature. To do multi-user SLAM with lower latency and also excellent productivity, this work suggest an appropriate scheme of partitioning of client server.

This work uses a simultaneous localization and mapping for creating a cloud points that are mapping to the air and then it employs that map for identifying the objects in built-up broad environment of AR in an instantaneous manner. Consequently, there is more freedom in how the space can be used and it useful for applications related to augmented reality like display of physical items by precisely identifying and pinpointing their exact position in the physical or real world. A real time, strong, visual technology of SLAM based on features with lesser drift and higher accuracy is introduced for strengthening the system that was used. These points can be summarized as,

Challenges:

Feature-based visual SLAM suffers when it comes to translation and light sensitivity.

Proposed Technique:

Object recognition in AR and visualization of digital information that are derived from the objects.

Marker Limitations:

Current technologies of Augmented Reality are heavily on markers, that restricts the placement of virtual objects and tracking to marker-defined spaces.

Cell phone Constraints:

Difficulties in achieving lower latency and higher accuracy on cell phones that has limited resources.

CLAHE Technique:

Uses the Contrast Limited Adaptive Histogram Equalization (CLAHE) for pre-processing images for improving the contrast and extraction of feature.

Scheme:

Recommends a partitioning scheme for multi-user SLAM for achieving low latency and higher productivity.

SLAM Implementation:

Creates a point cloud map of environment for real-time identification of objects in AR. The technology of 3D point clouds is used in many areas like Unmanned aerial vehicles, visualization with high precision, robotic hand holding, Virtual reality and augmented reality, mapping etc. To be specific, many research works have focused on recognition of objects by the process of analysing the cloud points directly.

But many of the researchers have missed to take in account of the uneven amounts of points of cloud on the surface of objects and the absence of data of features. Nevertheless, it is important to consider the possible drawback and different results generated by several different frameworks of augmented reality when developing many immersive applications.

6.2 PROBLEMS FACED BY EXISTING MODELS

Some of the existing models have did a comparison of Multi factor of 2 basic AR frameworks they are ARKit and ARCore with the aim of evaluating their performance

in several environments of computing. In the applications of computer vision, handshaking and detecting the parts like estimating the pose of hand, gesture recognition in busy environments in spite of relevance. An ideal source of virtual audio for a fully immersive AR would be the one which cannot be differentiated from the real things. For evaluating the quality, the conditions of co-immersion that includes situations with any digital or physical element combination. The existing models and taxonomy of AR is shown in ***Figure 6.1.***

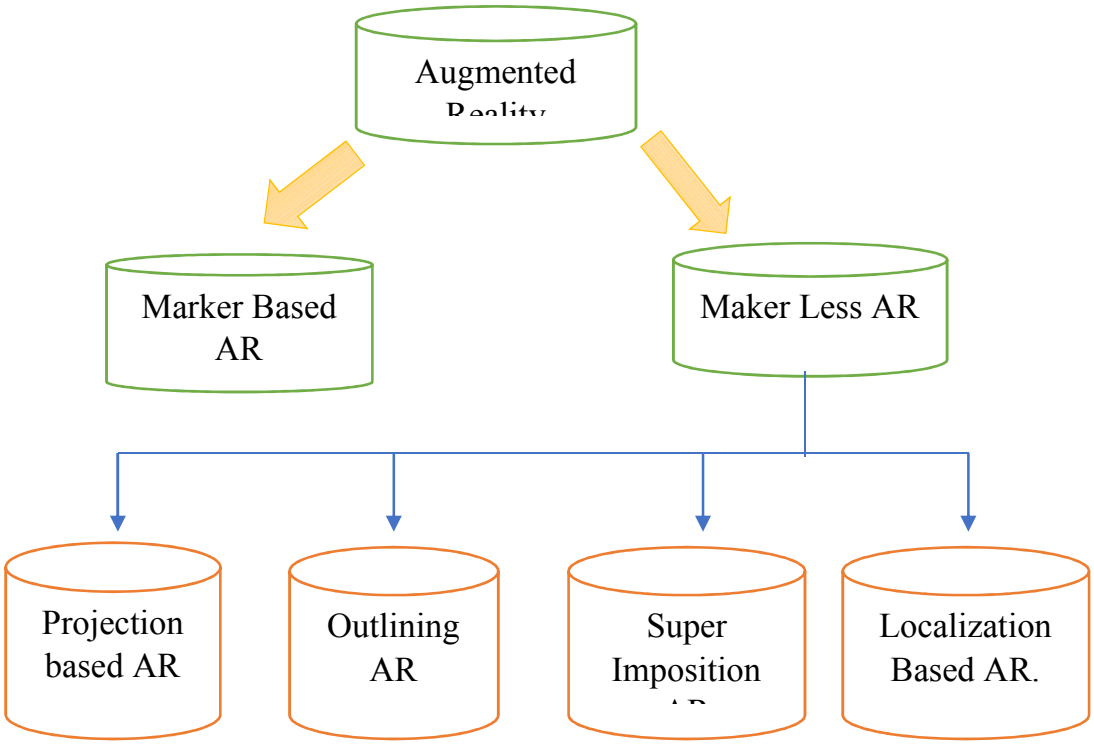

Figure 6.1 AR Taxonomy

For these reasons it is suggested the use of AAV for audio improved virtuality that stands for fully simulated setting which combines the audio of real world with the sounds that are generated artificially. It is recommended the use of slam systems that can help in measuring the posture of the mobile robots and will also rebuild the maps of the environments. Assuming that a stable workplace is made mostly with the help of the SLAM systems.

Nevertheless, the SLAM model's precision and resilience are compromised by the presence of several moving this of the real world in case of situations of real world. But in the case of robotics for the process of exploring the territories that are unchartered, the model SLAM plays very significant role.

The models of visual SLAM also have major roles in finding the object or persons in webcams as these models are inexpensive and they also offer rich information of picture as in contrast to the other models like the LiDARs. Conventional techniques of visual SLAM have proven to be effective in the static settings, but these models have some challenges when it comes to applications of robotics based on real world while facing with some dynamic situations.

If there is a proper assessment of existing model like the integration of semantic division into the conventional algorithms. There is major development towards a more robust visual algorithm that is SLAM.

There are 3 modules in this work shown in ***Figure 6.2***, they are

- Object detection
- VSLAM
- Mobile AR Application

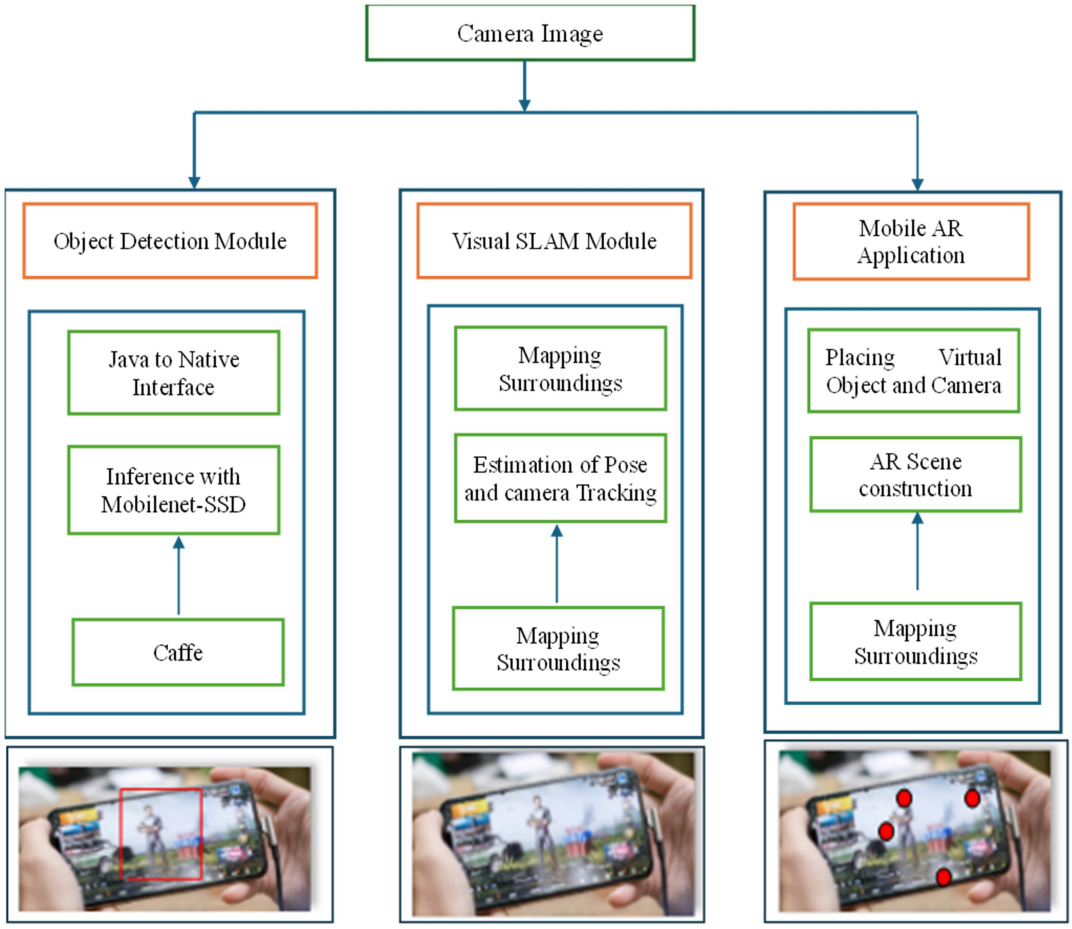

Figure 6.2 Object Detection in Mobile AR

This model performs well in static scenes but in case of changing or dynamic scenes. This work introduces a way more moving objects for avoiding the accurate tracking and also results. Further, the traditional SLAM model uses sparse maps which are not enough for intricate tasks like navigation.

Especially for more complex automobile components, use of additive manufacturing enables significant benefits. As the use of additive manufacturing is an instance of production with precision, it is important that several parts of the machinery of production is placed in some correct places to ensure accuracy.

SLAM is one of the potential and useful tools to do such works. With the use of visual based SLAM, robots that moves may do some positional tasks in settings which are unfamiliar while creating dense and sparse features of map at the same time. But the traditional vSLAM only works well in models of static scenes and it may not take the objects of real world into account.

As it cannot extract the seldom data from dense and sparse features of maps. One of the major issues in the area of AUVs in the control of navigation. The technique of simultaneous localization and mapping has conventionally used in prior works for addressing such issues. There are many SLAMS based model that are suggested by many different works, but still most of them have failed to overcome challenges associated to semantic details in scenarios that are visually challenging. Especially for more complex components of automobiles, additive manufacturing gives significant advantages.

The SLAM is a viable option. While considering the dynamic nature of settings of additive information this work has developed a robust technique of SLAM for the process of monitoring results of production of AI that is augmented with the models of Deep Learning. The Current techniques for recognition of visuals objects to have complete datasets that are kept in a single central location and approaches of artificial intelligence are necessary for the applications. As a consequence, transfer of data and robust overheads are considerable.

One of the significant approaches of deep learning gets over such restrictions is the Federated Learning that allows the users to train a model collaboratively by using the data that are processed on their devices. Server acting as global method, receives the weights that are updated.

One significant approach of deep learning that gets over such restriction is federated learning that lets the users training a method collaboratively with the help of the data processed on devices. The server that acts as a global method will receive the weights that are updated from each of the local units in all cycles. Each of the gadgets executes the processing separately and after that the local method is altered with aggregated weights.

By describing the pairwise relationship of various channels in convolution maps of feature an illustration is built around the matrix of covariance is proven to be efficient for categorization of pictures. The confounding effect happens when a third channel correlates with both of the target channels that renders the pairwise association inaccurately. Instead, this work will estimate the partial correlation in this instance that neglects the muddy effect. But still, the estimation of covariance is the only model which can successfully measure the partial correlation by solving the problem of optimization with the help of symmetrical positive some matrix. But their implementation into CNN is not yet resolved.

6.3 OVERVIEW OF 3D SPATIAL MARKER USING VSLAM

- The proposed work combines elements of two popular methods of AR: they are the techniques of three-dimensional space marker and the technique of object marker.
- Additionally, the drawbacks of each of the approaches of marking is what defines the proposed solution. Not just that, but it's possible to automatically determine a pose of an object in 3-dimensional space, thus setting the location of visible object's needs lesser labour from human. The recommended system's input tool is RGB-D digicam.
- With use of advances of technology of SLAM (simultaneously localization as well as mapping), spatial markers are produced. It is possible for constructing a setting of augmented reality with the help of SLAM because it will simultaneously create a marker and monitor real-time camera’s position.
- Next to the space of development, units of object are located and are recognized with help of technologies of DNN using the 2D picture learning. The mechanism behind automatically executed visualization object registrations is called as ICP (Iterative Nearest Point).

6.4 STIMULATIVE IMPLEMENTATION OF 3D OBJECT WITH REAL-WORLD SCENARIO

The proposed model is a novel framework based on augmented reality that can be seen as 3 different modules they are

- Pre-Processing
- Visual SLAM
- Registering.

For the process of determining the location of camera, the modules of Visual SLAM use the process of monitoring for analysing each of the frames and to determine when an additional frame should be inserted. The module of visual SLAM calls the module of registration while initializing 3D map after creating the maps successfully. A dense map is created with the help of merging the points of cloud with the data of RGBD with the help of the posture that is computed previously. For the process of determining the identity of object and to obtain the change matrix, we are dependent on the model of 3Ds.

After that the location of camera and pose are transformed to the system of OpenGL positioning beneath the matrices of Model View with the help of procedure in conjunction to the visual SLAM. For completing the process of AR, the final step will be alignment of simulated 3D object with real environment.

6.4.1 Pre-Processing Module

In this model the input that are the binocular or stereo sequences is subjected to the process of enhancement of picture as a primary step. In this the process of histogram equalization boots the quality especially the contrast. A higher brightness level in photos of input results in a better local accuracy.

When it comes to processing of images, histogram equalization brings the concept of histogram from the statistics for depicting the grey levels of different distribution shows a different quality of image. Consequently, the gray histogram serves as a foundational for the enhancement of images processing the reflecting the picture quality and outline. Gaining greater details of feature while over enlarging the clutter is achieved with the help of the CLAHE model for selectively improving the contrast in picture.

In a picture of grayscale, the gray histogram shows how often each of the levels of gray appears and how may pixels will cover that level. The most popular technique for enhancing the brightness of picture is histogram equalization that is easier to implement as well as produces better results.

The distribution of pixel across different gray levels in a digital image can be obtained by counting the frequency of occurrence of each intensity value. A common way to visualize the distribution can be done through the construction of grayscale histogram. Here the pixel intensity are values are plotted against their corresponding frequency of occurrence. The layout of the system is shown in ***Figure 6.3***

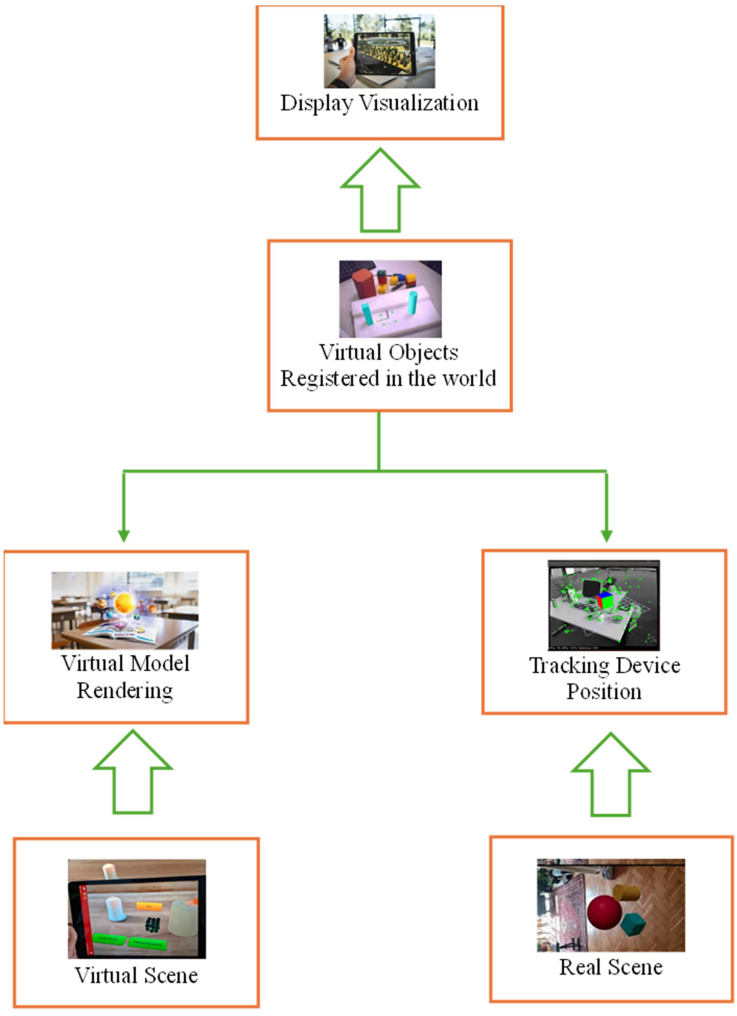

Figure 6.3 Stimulation of 3D Objects

The gray distribution of the image is described as

$P(r_k) = \frac{n_k}{n} k = 0,1, \ldots L-1$ (6.1)

In this

$\sum_{k=0}^{L-1} P(r_k) = 1$ (6.2)

In contrast to this, there are 3 problems with the concept.

- A gray distribution shows only you how often all the gray appears in digital images—that is, what percentage of pixel of image have a given value of gray—but it will not tell about where in that picture those pixels are located physically.
- The shape of the histogram will be same for all of the photos with the identical distribution of frequency of the occurrences of gray scale because a histogram will lack in one location.
- If a picture is sub categorized as many pictures, the cumulative histogram for combined images is same as sum of individual image's histogram since these shaded histograms represents statistical significance of every gray level occurrence.

To summarize, the major process of this module in the below section and in flowchart ***Figure 6.4.*** Initially preliminary improvement of Picture is carried out by reducing unwanted artifacts, random variations in the image. Then inputs are Stereo sequences. The image of input is subjected to undergo enhancement of picture. Next the model of histogram equalization is used for improving the quality of image, especially contrast. Histogram equalization improves contrast by adjusting the distribution of gray levels. Here, the results in better local accuracy with improved brightness. Next, Gray Histogram depicts gray level distribution in an image. It acts as a tool for evaluating the quality of image and guides improvement. Then, the CLAHE Method applies contrast-limited adaptive histogram equalization (CLAHE) to enhance feature details. Helps to improve contrast selectively and manage clutter.

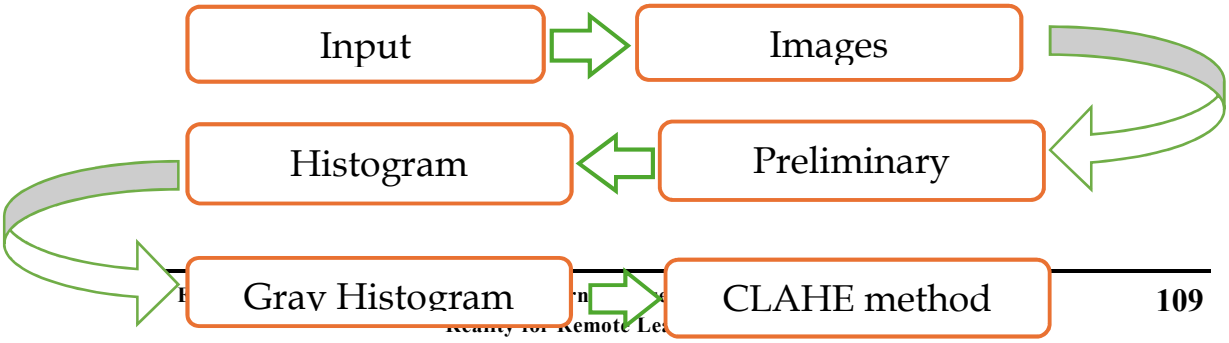

Figure 6.4 Steps involved in Pre- processing Module

Characteristics of Histogram:

Shows frequency of each of the gray level and number of pixels at each of the level. Simple and efficient for enhancing the brightness of the images.

Limitations:

- Does not reveal the physical location of gray levels within image.
- Histogram shape is consistent for images with same frequency of distribution of gray level.

For subdivided images, cumulative histogram is sum of individual histograms.

6.4.2. Visual SLAM Module

Researchers are interested in this Visual Simultaneous Mapping and Localization (VSLAM) in the time period of 1990s, where it was actually developed with the help of conventional process of visual analysis models and as well as the methods of recognition. In recent times it is built with the help of models based on deep learning while there is some long way to go in completely implementing these approaches of VSLAM, recent studies of deep learning have shown some encouraging results for the use cases like self-driving, navigations, robots for assistance, VR/AR and as well as prediction of position.

The difficulty in the process of segmenting and tracking in videos while the same time has made the identification of object in optical SLAM in spotlight. Our novel process of segmentation of video by an instance method, Mask RCNN generates the features vectors from a recurrent fully convolutional monitoring branching and implements tracking via accurate matching of feature vectors.

In this work, we have presented a new iterative fully convolutional trackers that, as part of tracking manipulate, creates a feature vectors for all the case within a box of limits. This vector is then used for process of instance categorization assignment. Several core concepts that are often thought to be a separate tasks of computer vision are really part of architecture of VSLAM. Traditional geometric approaches are classified as 2 groups they are shown in below ***Figure 6.5.***

- Optimization based methods of VSLAM
- Filtering based methods of VSLAM

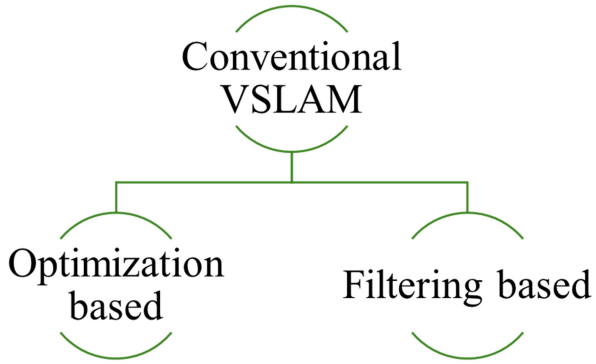

Figure 6.5 Types of Conventional VSLAM

Optimization based methods of VSLAM offers better accuracy and filtering that are based on techniques of VSLAM, that are first extensively researched due to their affordable cost of computing. The classical geometric techniques of VSLAM are quite efficient in environmental mapping as well as translation they do have many drawbacks like being sensitive changes regarding illumination, weather, seasons and being non-invariant to scale between others.

For the process of matching the feature connections in different pictures or the frames, the major features-based models of VSLAM uses the robust descriptors along with extensive features of image. There is performance hit in low-textured, unstructured or settings of motion blurred while using the display of fixed features. Crafted extraction of features is not important for models of deep learning. Thus, the semantic segmentation has replaced the jobs of old VSLAM for the process of extraction of features. And the process is depicted in ***Figure 6.6.***

The major points are highlighted below.

VSLAM:

- Developed in 1990s with the help of conventional methods analysis of visual and recognition.
- Recently these have been advanced using models of deep learning.

Used in:

- Deep learning has shown reliable results in self-driving-based cars, navigation, robotics, VR/AR, and prediction of positions.

Challenges:

- Difficulty in segmentation and tracking in videos in a simultaneous way.
- Identification of object in optical SLAM

Techniques:

- Mask RCNN:

  Generates feature vectors for the process of segmenting the videos and tracking.
- Iterative convolutional trackers are used in creating feature vectors for categorization.

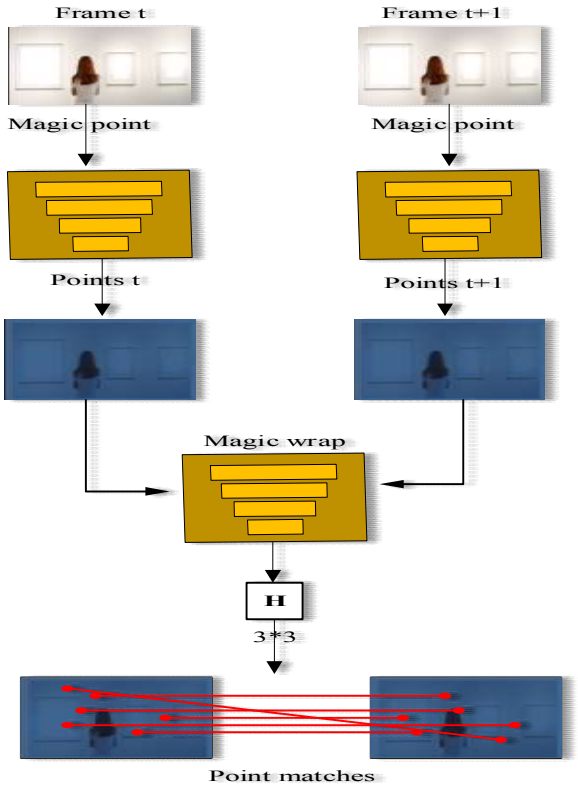

Figure 6.6 Feature Point Matching using VSLAM

Conventional Approaches:

- Optimization-Based: Provides better accuracy
- Filtering-Based:They are preferred for their lower cost of computational.
- But both of these methods have limitations, like sensitivity to changes in lighting, weather and scale.

Feature-Based Approaches:

- They use robust descriptors & many image features.

- Issues in performance arises in settings that are low-textured, unstructured, or motion-blurred.

Advancements in Deep Learning:

- Semantic segmentation has replaced conventional extraction of features in VSLAM

6.4.3 Object Detection and Registration Module

For the process of generating the virtual objects accurately, first know about how everything combines in an actual scene. If not, it will result in errors while trying to merge them with actual things.

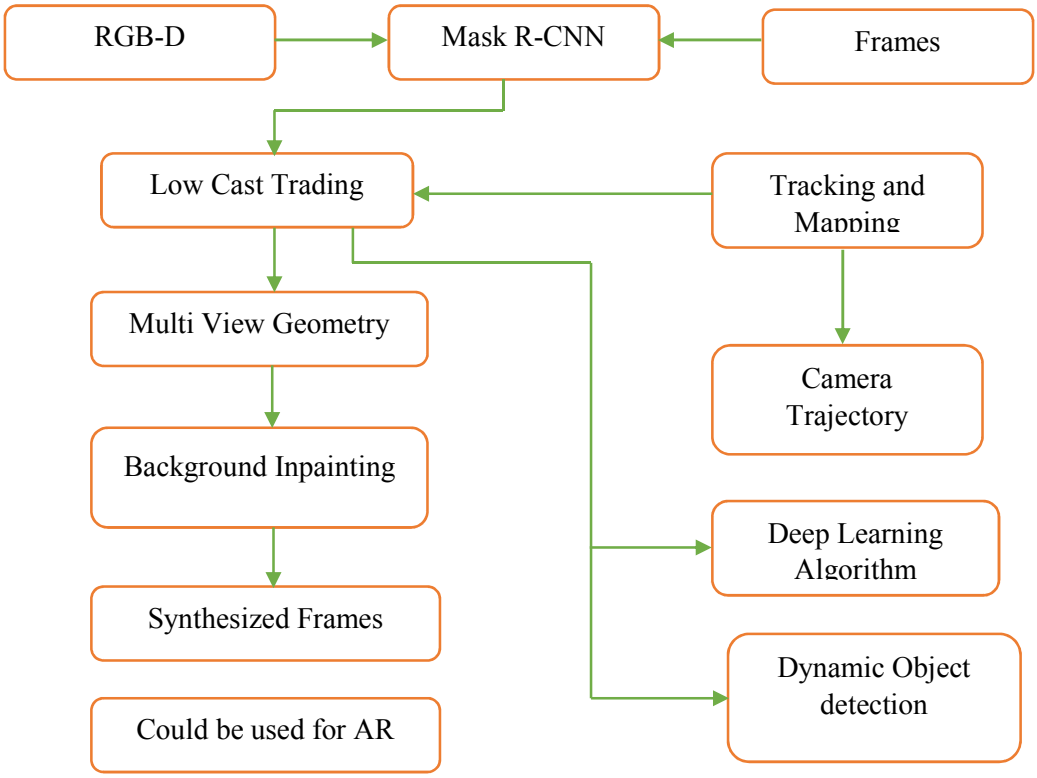

Figure 6.7 Mask RCNN based Object Detection and Registration

For example, in the scenarios of real world, techniques for detection of objects if it must generate a digital wall then the first step will be to find a wall in environment. This is called as step of identification of plane. This process of detection of objects using MRCNN is shown in ***Figure 6.7.*** After completing the step of plane identification, the edge location of the scene should be determined, they can be determined when 2 surfaces

intersect, a process called as recognition of edge, is where the edges of the scene will be detected.

Metric IoU:

One of the commonly used metrics for evaluating the accuracy of detection of items in any dataset is the score of IoU – Intersection over union. This metric is used in evaluating the accuracy of detection of items. This metric can be used for finding out the matching between the predicted and real price. That is, it tells about how near or relevant the predicted data is to the actual outcome. The increase in value is because of the increase in correlation.

As a common rule, it is said that if the IoU is greater than 0.5 then it is considered as excellent result. The mask RCNN will perform well in case of any obstacles and it can identify the objects as small, medium or big that is shown in table.

MRCNN:

The MRCNN model is a versatile neural network model. By the process of including several modelling branches perform the following.

- Models like instance segmentation
- Identification of object
- Classification of objects.

The process of Mask RCNN is given below as:

Step 1: Feature extraction

The features are extracted after completing the preliminary processing of the filtering of photographs and denoising as well.

Input source: The pictures that are captured by choosing robot will be sent via multi scale pyramidal feature extraction approach of network that results in producing the matching features maps.

Step 2: Return on Investment

For all the location in the feature maps there will be an established number of ROI that is intended to be found

Step 3: Network for regions' proposals

The RPN – Regional Proposition Network will receive candidate ROI, it will process it with the help of the box regression and categorization of foreground and background, and then it will remove any of the ROIs that will not include the objects.

Step 4:Bounding Box Alignment

The bounding boxes are aligned with the aim of achieving segmentation at the level of pixels, the matching of bounding box is done with the help of an approach of linear interpolation.

Step 5:

Classification and regression Classification of the remaining candidate ROI, the box of regression, however, as well as mask generation are all done.

For instance, the Mask RCNN offers

- Higher level of precision for detection
- High segmentation of accuracy of instances
- Multi-scale pyramidal feature mining

The network of multi scale pyramidal feature mining is built upon the operation of ROI alignment, ROI network of region proposition and the networks of multi scale pyramidal extraction of information. The whole process of MRCNN is shown in ***figure 5.8*** The search and recognition are based on the neural network foundation with the design of convolution. After the transmission of feature maps to the systems of suggestion of area, 9 potential ROIs are selected for all the anchoring position of map.

The first thread will predict the central secure point of ROI, the second will use a 2 class Softmax operate for deciding if the applicant of ROI is in background or in the foreground. The output is split as 2 strands they are

- Categorization thread
- Estimation Thread

The x, y, w, and h coordinates of a possible region of ROI. With the help of the previous works done by lens concerning registration and location it will be able to create a 3 D representation of real scene with the help of the RGB-D sequence of photo and can rebuild it (***Figure 6.8***).

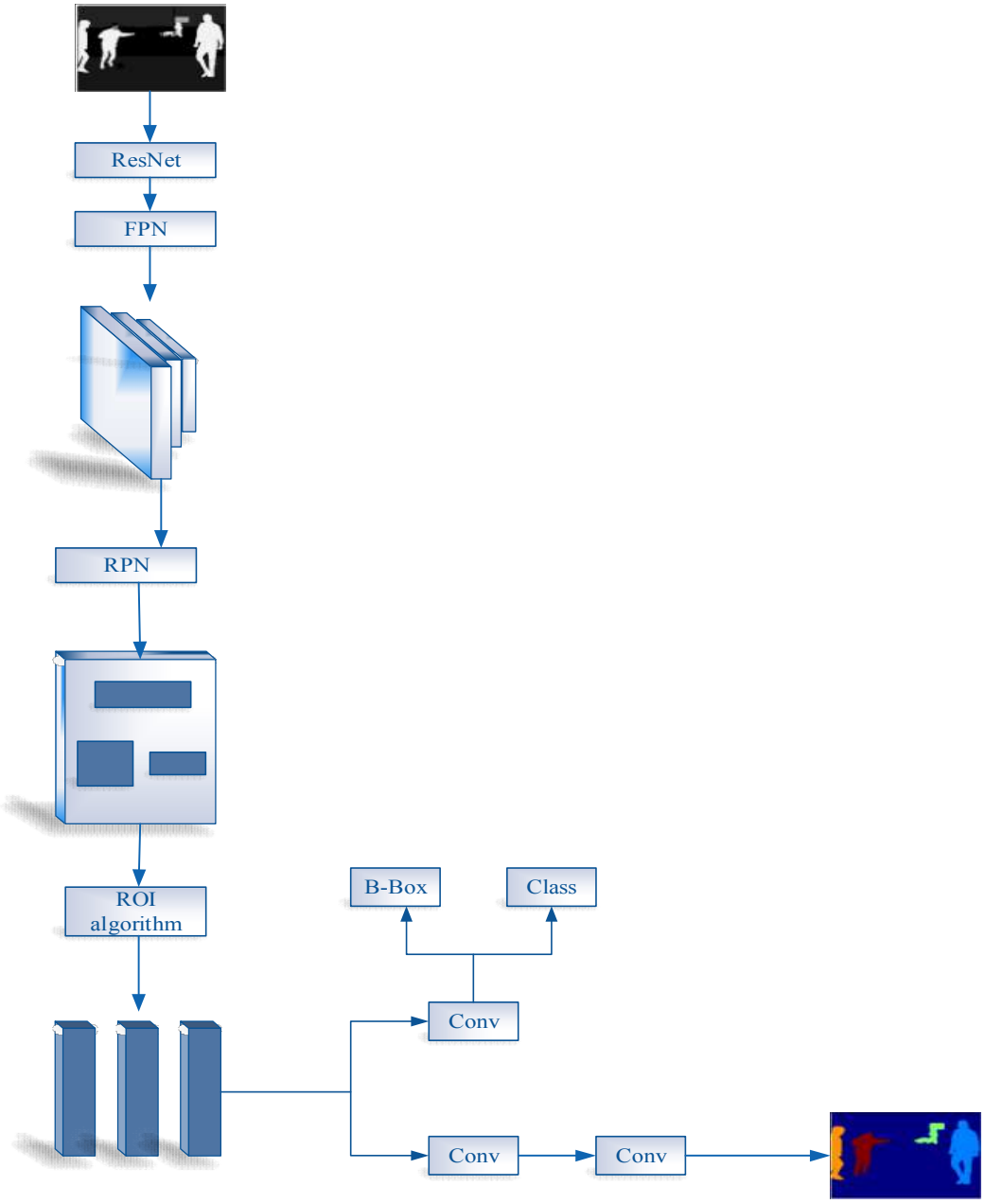

Figure 6.8. Mask RCNN Procedure

Coordination of superimposition of each of the frame will allow for a 3D reconstruction to be done. The first step is the identification of 2 neighbouring sequences of picture and can refer to them as N and N+1. The approach of extraction of feature while applying to the sequence of frames of Nth and (N+1)th, discovers a solution to the issue of PnP. After that find the relative intrinsic variables of sequence of picture from the frame N

and (N+1). For the process of doing a 3d reconstruction, the formula given below is used for the process of calculating the exterior matrices of camera,

$$W_{col} = \begin{bmatrix} r_{11} & r_{21} & r_{31} & t_1 \\ r_{12} & r_{22} & r_{32} & t_2 \\ r_{13} & r_{23} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.3)$$

(X, Y, Z) are coordinates in 3d of RGB-D video stream. Reconstruction of scene as a 3D representation is possible with the help of the recursive repetition. Reconstruction for a single frame that is not optimized is shown in figure

At last, the virtual object is to brought into the physical reality by subjected to a sequence of changes to its structure of coordination that includes the transition between world, camera, crops and the system of screen coordinates. The 5th equation offers a left to right description of sequences of transformation that is a translation matrix and rotation matrix R(3x3) they convert external coordinate system into the coordinate system of camera that is $T_{3\times1}$.

These matrices are built with the help of camera coordinates and data about planes that are identified. Next to that the coordinate system of screen is converted from the system coordinate of camera that is (u, v) by central length (f_x, f_y) and the principal point(d_x, d_y). The camera's calibrating gets such values. At last, screen will be a real-world register for virtual objects.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & d_x & 0 \\ 0 & f_y & d_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.4)$$

To summarize

Challenges in generation of Virtual Object:

- Accurate placement of virtual object will be based on understanding actual scene.
- Example: Placement of digital screen on a wall to place a digital screen on it includes identification of plane and recognition of edges

Metric of Item Detection:

- Intersection over Union (IoU) helps in measuring accuracy of prediction
- IoU score more than 0.5 is said to be a better score
- Mask RCNN efficient in detecting objects as small, medium, and large

Mask RCNN Process:

- Step 1: Feature Extraction
  - Process the images of input with the help of multi-scale pyramidal extraction of feature.
- Step 2: ROI Identification
  - Establish and identify ROIs from the maps of feature.
- Step 3: ROI Proposal Network (RPN)
  - Process ROIs with boundary box regression and categorization of objects
- Step 4: Bounding Box Alignment
  - Match bounding boxes with the method of linear interpolation for segmentation based on pixel-level.
- Step 5: Classification and Regression
  - Classify remaining ROIs, perform box regression, and generate masks.

Reconstruction of 3D scenes:

- Image Sequences:
  - Use RGB-D photo sequences for creating a 3D scene.
- Feature Extraction and PnP Solution:
  - Solve the problem of Perspective-n-Point (PnP) for neighbouring frames.
- Calculate Camera Matrices:
  - Calculate external matrices for 3D based reconstruction.

Virtual to Physical Integration:

- Coordinate Transformation:
  - Convert the coordinates of virtual items via matrices of translation and rotation.
- Screen Coordinate System:

- Map the coordinates of camera coordinates to the screen coordinates with the help of calibration data.

Outcome:

- Achieving accurate placement of virtual objects in the real world by aligning the system of coordinates and performs required transformations.

6.5 RESULTS & DISCUSSION

This work has presented a system which will detect the items in virtual space and estimates its position that simultaneously allows the camera monitoring in environments of real world. The model can be accessed from top to bottom using the dataset of ADVIO by a setup test in the changing situations between the other complex datasets for ensuring its resilience and accuracy in scenarios that are dynamic.

The studies are done with the help of laptops which as 2 GB memory of graphics for components of semantic segmentation, the NVIDIA GTX860M GPU, 16 GB of RAM and the Intel Core i7-4710MQ CPU (4-core 2.5GHz) is used.

The examples of some actual spaces of indoor are malls and museums and they can be intricate and ever changing. Most of the open source computations of SLAM are not that appropriate for the situations that are dynamic as they are all based on static settings assumption. This work has used the ADVIO DATASET that is intended to identification variations in existing model for evaluating the performance of our system completely.

Actual world machine vision standard sets for the visual inertial optical are designed by the dataset ADVIO which are diverse and complex. This collection involves 23 sequences that are captured in several settings of indoor and outdoor with the help of the several devices of mobile that includes iPhone, Pixel Android Phone and the Google Tango as well.

In addition to such ground truth direction, RGB and IMU data are present as well. For the process of doing a proper comparison evaluate the effectiveness of the proposed technique with the help of the metrics given below they are,

- Root Mean Square Error (RMSE)
- Standard Deviation of Relative Pose Error (APE)

6.5.1 Result of Tracking Time

a. Phase 1 Number of data 300 to 500

In the first phase of evaluation, the number of data taken varies between 300 to 500 is shown in ***Figure 6.9***.

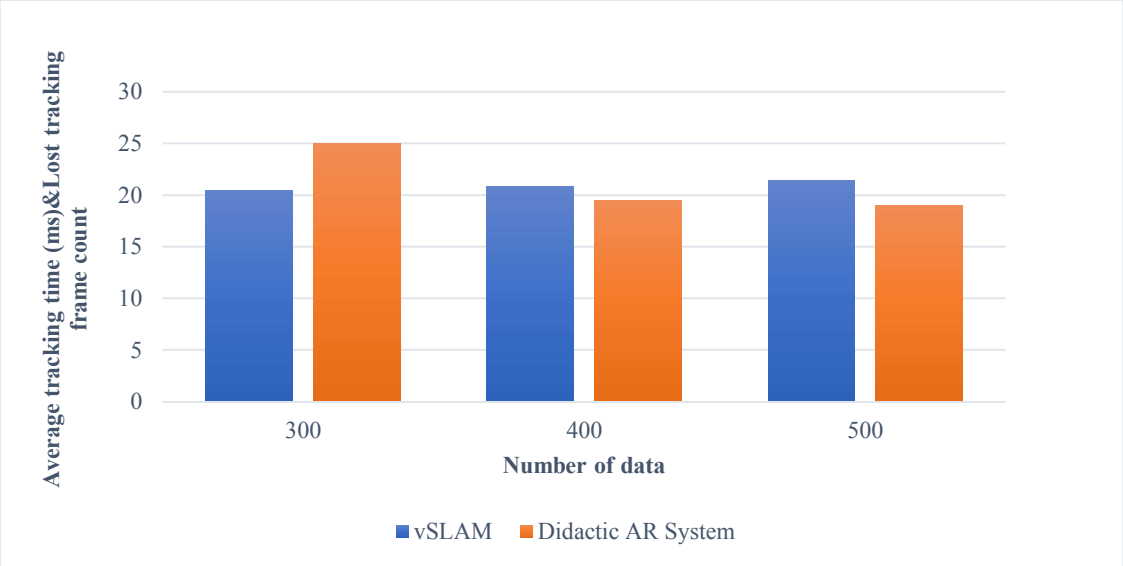

Figure 6.9 Tracking time for 500 data

Now investigate time taken for the model for the process of tracking based on the number of data. Tests were conducted as 3 different stages, in the first stage, the minimum number of data taken were 300 and the maximum is 500. For this, the average tracking time of the proposed deep learning based VSALM model with the conventional VSLAM model.

b. Phase two – number of data minimum 600 and maximum 1000

In this second step of evaluation, the number of data are taken as 600 to 1000 is shown in ***Figure 6.10.***

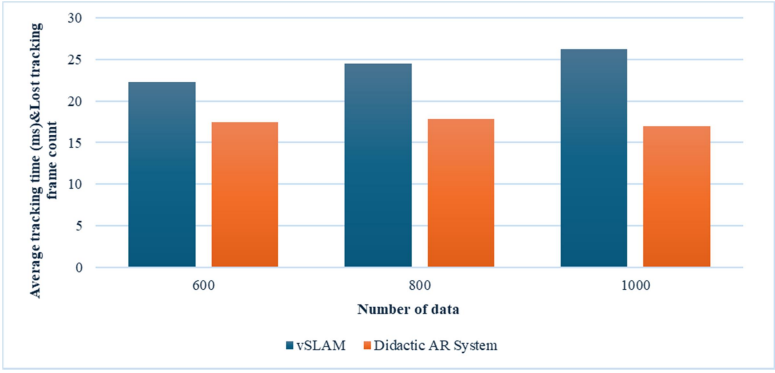

Figure 6.10 Tracking time for 1000 number of features

c. Phase Three evaluation – number of data taken is between 1250 and 1500 (***Figure 6.11***)

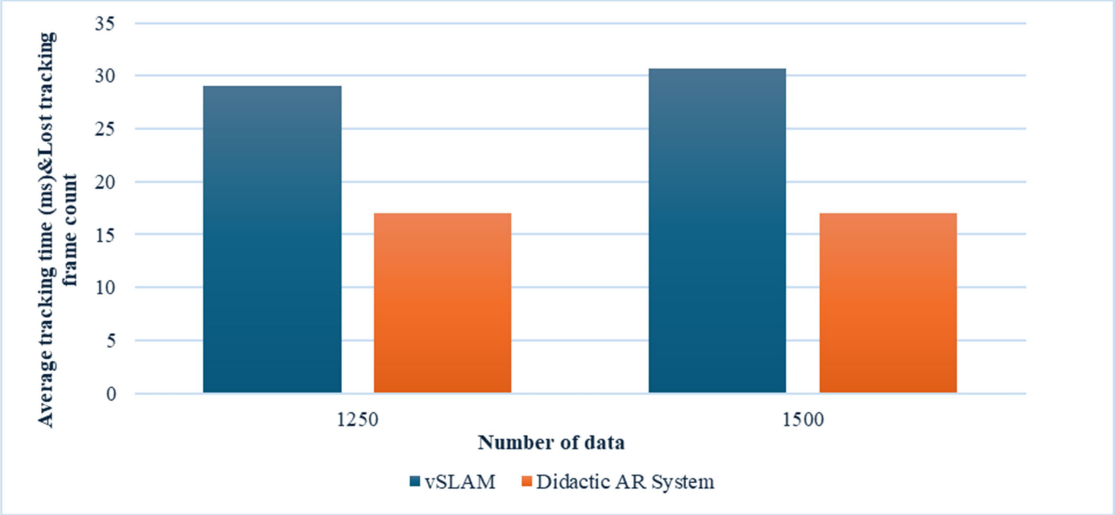

Figure 6.11 Tracking time for 1500 number of features

Now in the third step of evaluation the proposed model is evaluated with more than 1000 data between 1250 and 1500 data. The results are shown in figure 6.11. The result of tracking time is shown in ***Table 6.1.***

Table 6.1 Tracking time comparison

Number of data	VSLAM	Proposed Model
300	20.5	25
400	20.8	19.5
500	21.4	19
600	22.3	17.5
800	24.5	17.8
1000	26.2	17
1250	29.1	17
1500	30.7	17

So far, the results of the tracking time based on number of data were displayed. Now, investigate the results of the performance metrics. They are

- RMSE
- APE

a. APE:

Table 6.2 APE performance Comparison

Iteration	APE of ORB	APE of ORB-BREAK	APE of FAST_ORB	APE of CNN	APE of Didactic AR System
2000	15	5	5	3	1.5
4000	15	5	5	3	1.5
6000	15	5	5	3	1.5
8000	15	5	5	3	1.5
10000	15	5	5	3	1.5

The results of APE based performance for registration is calculated and it is compared to the other existing models like the ORB, ORB-Break, Factor, CNN to the proposed model which is shown in figure. As the error rate of the proposed model is much lower than the other models which proves the efficiency of the proposed model. The graphical representation of these results is shown in ***Figure 6.12 and Table 6.2***

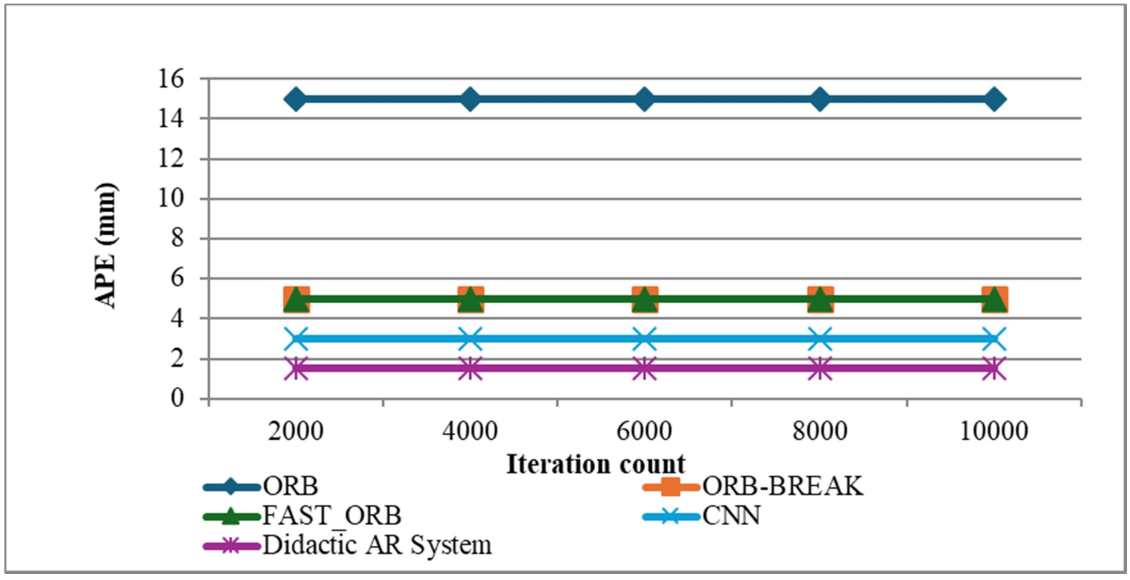

Figure 6.12. RPE Performance for Registration

b. RMSE:

The results of RMSE based performance for registration is calculated and it is compared to the other existing models like the Oriented fast and rotated brief (ORB), ORB-Break, Fast-ORB, Convolution neural network (CNN) the proposed model that is the deep learning based which is shown in ***Figure 6.13 and Table 6.3***

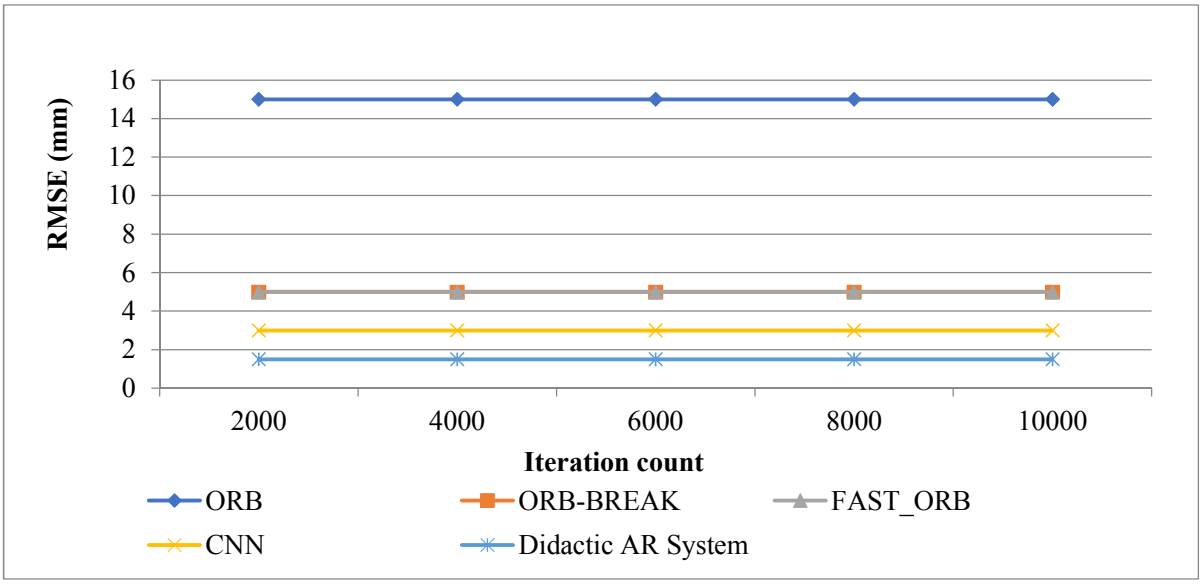

Figure 6.13 RMSE Performance for Registration

Table 6.3 RMSE performance comparison

Iteration	APE of ORB	APE of ORB-BREAK	APE of FAST_ORB	APE of CNN	APE of Didactic AR System
2000	15	5	5	3	1.5
4000	15	5	5	3	1.5
6000	15	5	5	3	1.5
8000	15	5	5	3	1.5
10000	15	5	5	3	1.5

In this, it has presented the virtual visual data with the help of various colors that indicates the results of detecting objects. All the items were identified, and an excess of colored virtual visual data were superimposed over physical environment. It was seen that the dimensions of boundaries were properly encircled the objects based on rotation and location that was verified each of the results.

According to the orientation of placement, precise tracking is maintained regardless of movement of the camera. The end result will be the mapping and in the mode of localization, monitoring of virtual visual based data that functions correctly. It is also feasible for verifying the tracking and also the overlay of virtual visual information on the objects.

In this work we have evaluated the methods of GPU acceleration against the server-side standard ORB-SLAM for the process of knowing how long it has taken for tracking the objects. In figure 8, the time was shown that was spent for the process of monitoring in graphical representation showing the number of steps involved. These steps include the extraction of features, prediction of posture, feature matching and the search localized point as well.

While contrasted to the standard ORB-SLAM, the recommended model helps in reducing the local time that is spent on tracking by 25-50%. While taking the binocular dataset, the SLAM-share will reduce the overall latency of tracking to around 40% and while using the dataset of stereo images, it has gone down by more than 50%. Thus, it is evident that the SLAM share is able to attain better efficiency in real time with rate of frame lesser than 33 milliseconds.

The data of visual that is synthesized and simulated at precise object location during real time and performs normal monitoring as per camera's position in real world are the ability of the introduced system. But still with several obstructions it is difficult to function properly as it is dependent on the inputs of visual. The effect is more on ICPs which process the data from the cloud points. While preparing a virtual visual data for the process of synthesis, ICP have to make sure that the accuracy is maintained with regard to placement objects. For such reasons, proper positioning of synthesized objects can be a difficult task when they are partially or completely covered by other things.

In this work, an app was offered which was used to follow a camera in real world and uses the estimates of identification of object for the process of displaying the visual data from a virtual environment. With the help of the data of image of RGB, the system will be able to monitor the things in 3D, identifies them in 2 dimensions and will measure their poses in 3D all this happens when identifying the objects in 3D and also their positions as well. Since there are 2 coordinates that must be shared for connection between virtual and real spaced, method of tracking of camera based on simultaneous localization and mapping the SLAM is used. One thing is the smallest controllable entity is the taken space. Making a geo map using SLAM, but still, makes unfeasible for recognition of item and estimation of position. The deep learning-based object recognition will help in solving such issues. And finally, the method of ICP uses the actual location of object along with rotation and orientation for improving the virtual visual information.

6.6. CONCLUSION

The major issues faced by the augmented reality systems are related to accurately viewing the physical environment and extension of actual world scene area. For solving these issues, a model to actual world 3D recognition of objects, tracking and augmentation which decreases the effect of augmented reality's issues of expanding space is given in this work. The introduced system has the ability to be a technology which enables virtual objects in one of the locations to be interacted with the real-world environment in real time. Considering that this model is successful to address the problems in estimation of position and occlusion of moving items that are complementary to this work the AR is expected to become accessible through confluence with IoT. By superimposing the digital images into the physical spaces, the AR will allow the users for manipulation. As a method of operating it is expected to offer customers with an intuitive interface which improves immersion and realism. Applications in several domains like AI, health care, gaming, education, military and entertainment as well can be improved by the sensory based advantages and technological advancements in several technologies for tracking. One of the obvious use cases for AR based equipment for tracking is creation of dataset that need learning of 3D objects in AI.